

KATHOLIEKE UNIVERSITEIT LEUVEN  
HOGER INSTITUUT VOOR WIJSBEGEERTE



---

KATHOLIEKE  
UNIVERSITEIT  
LEUVEN

# **RATIONALITY AND INSTITUTIONS**

An Inquiry into the Normative Implications  
of Rational Choice Theory

Promotor:

Prof. Dr. Antoon Vandevelde

Proefschrift aangeboden tot het  
behalen van de graad van  
Doctor in de Wijsbegeerte door

Bart Engelen

Leuven, augustus 2007



## ACKNOWLEDGEMENTS

---

The work that lies before you is the result of four years of reading, thinking and writing on rationality and institutions. In the end, the work involved is performed by a single man in an often all too lonesome job. Therefore, my original intention was to thank nobody except the Research Foundation – Flanders (FWO) – and thus indirectly the average taxpayer – for their financial support. Now, I realize that I couldn't have been more wrong. Since I want to grab the opportunity to return the favor, I am pleased to acknowledge my thanks to the people whose support throughout the years helped me to successfully complete this PhD.

First and foremost, I am truly thankful to my promoter, Prof. Dr. Toon Vandevelde, who was there from the very beginning to the finish of the entire process of writing this PhD. His unrelenting efforts to read earlier versions of my texts and to provide helpful comments inspired me intellectually and stimulated me to persevere and bring this work to a finish. Without his continuous support, this project would not have been possible at all.

Second, I want to thank the members of my doctoral commission, Prof. Dr. Stefaan Cuypers and Prof. Dr. Luc Van Liedekerke. Their critical remarks on different parts of this dissertation certainly helped me to improve the whole in substantial ways. Moreover, they gave me the opportunity to experience how pleasurable academic collaboration can be. I praise myself lucky for being able to assist Prof. Cuypers in his teaching activities and for working in a flourishing research centre populated by Prof. Van Liedekerke. In addition, I want to thank Prof. Dr. Erik Schokkaert, whose critical remarks in various seminars were a systematic reminder to take the field of economics – and the people studying it – seriously. Finally, I owe a word of thanks to Prof. Dr. Shaun Hargreaves Heap for taking the time and trouble to serve on the jury for this dissertation's public defense.

Third, I want to thank my colleagues at the Institute of Philosophy and, more specifically, the Centre for Economics and Ethics. In this respect, I am especially grateful to John Alexander, Yvonne Denier, Helder De Schutter, Sylvie Loriaux, Stijn Neuteleers, Thomas Nys, Joris Van Damme and Ann Van Espen who all contributed to this project in their own way. Most of them commented extensively on several of my texts. Additionally, they were always in for a hearty discussion in the local pub. Above all, they helped to establish a highly agreeable environment in which working was both efficient and fun.

Fourth, because all work and no play makes Jack a dull boy, I also want to thank each and every one of my friends back home. Too numerous to mention each one of them personally, they probably know who they are. I think I would not have been writing these acknowledgements, if it wasn't for the many hours I have spent with my friends, having a drink, partying or just chatting away. As such, they provided me with plenty of occasions to keep the delicate equilibrium between labor and leisure in balance.

My last but also most precious words of thanks go out to my parents for supporting me in every possible way. Even though things weren't always easy the last few years, they never stopped believing in me. Thank you.

Bart Engelen

Leuven, August 12<sup>th</sup>, 2007.

# CONTENTS

---

<b>Acknowledgements</b>	<b>v</b>
<b>Contents</b>	<b>vii</b>
<b>General introduction</b>	<b>1</b>
1. Remarks regarding content	1
2. Remarks regarding style and methodology	3
<i><b>PART I: RATIONALITY</b></i>	
<b>Introduction to Part I</b>	<b>9</b>
1. Why rationality?	9
2. Methodological individualism	11
<b>Chapter 1: The economic conception of rationality</b>	<b>15</b>
1. Economic rationality and its failures	15
1.1. The level of actions and the relation with their reasons	16
1.1.1. Maximization	16
1.1.2. Instrumentality	16
1.1.3. Intentionality	19
1.2. The level of the reasons themselves: beliefs	20
1.2.1. Consistency	20
1.2.2. Perfect rationality	21
1.3. The level of the reasons themselves: preferences	22
1.3.1. Consistency	22
1.3.2. Transitivity	23
1.3.3. Completeness	23
1.3.4. Continuity	23
1.3.5. Exogeneity	24
1.3.6. Egoism	25
1.3.6.1. Self-regarding preferences	26
1.3.6.2. Other-regarding preferences	29
1.3.6.3. Reducing other-regarding preferences to egoism	30
1.3.6.4. Process-regarding preferences	33
1.3.6.5. Reducing process-regarding preferences to egoism	34
1.3.7. Egocentrism	34
2. The role of economic rationality within the social sciences	37

<b>Chapter 2: Alternative conceptions of rationality</b>	<b>45</b>
1. Minimal rationality	45
1.1. The level of actions and the relation with their reasons	45
1.1.1. Maximization	45
1.1.2. Instrumentality	47
1.1.3. Intentionality	48
1.2. The level of the reasons themselves: beliefs	51
1.2.1. Consistency	51
1.2.2. Perfect rationality	51
1.3. The level of the reasons themselves: preferences	52
1.3.1. Consistency	52
1.3.2. Transitivity	52
1.3.3. Completeness	53
1.3.4. Continuity	53
1.3.5. Exogeneity	53
1.3.6. Egoism	54
1.3.7. Egocentrism	54
2. Broad rationality	54
2.1. The level of actions and the relation with their reasons	55
2.2. The level of the reasons themselves: beliefs	55
2.3. The level of the reasons themselves: preferences	57
3. Expressive rationality	58
3.1. The level of actions and the relation with their reasons	59
3.1.1. Non-instrumentality	59
3.1.2. Intentionality	60
3.2. The level of the reasons themselves: beliefs	62
3.3. The level of the reasons themselves: preferences	62
3.3.1. Consistency	67
3.3.2. Transitivity	68
3.3.3. Completeness	68
3.3.4. Continuity	68
3.3.5. Exogeneity	70
3.3.6. Egoism	72
3.3.7. Egocentrism	73
<b>Conclusion of Part I</b>	<b>75</b>
1. Schematic overview	75
2. The relation between different conceptions of rationality	77

***PART II: THE RATIONALITY OF VOTING DECISIONS***

<b>Introduction to Part II</b>	<b>83</b>
<b>Chapter 3: Economic rationality and voting decisions</b>	<b>85</b>
1. Economic rationality and voting decisions	85
1.1. Economic rationality and the decision whether or not to vote	85
1.1.1. The calculus of voting	85
1.1.2. The paradox of voting	86
1.1.3. Amending the calculus of voting	88
1.1.4. Problems with amending the calculus of voting	91
1.2. Economic rationality and the decision whom to vote for	92
1.2.1. The calculus of voting	92
1.2.2. Another paradox of voting?	93
2. Reconciling rationality and democracy	96
<b>Chapter 4: Alternative conceptions of rationality and voting decisions</b>	<b>99</b>
1. Minimal rationality and voting decisions	99
1.1. Minimal rationality and the decision whether or not to vote	99
1.2. Minimal rationality and the decision whom to vote for	100
2. Broad rationality and voting decisions	100
2.1. Broad rationality and the decision whether or not to vote	100
2.2. Broad rationality and the decision whom to vote for	102
3. Expressive rationality and voting decisions	102
3.1. Why an expressive account is needed to solve the paradox of voting	103
3.2. Expressive rationality and the decision whether or not to vote	104
3.3. Expressive rationality and the decision whom to vote for	107
3.4. The distinction between the decisions why and how to vote	108
3.5. Expressive rationality and cheering	109
<b>Conclusion of Part II</b>	<b>113</b>

***PART III: THE RATIONALITY OF INSTITUTIONAL DESIGN***

<b>Introduction to Part III</b>	<b>119</b>
1. Why study institutions from the perspective of rationality?	119
3. In search of a realistic utopia	121
4. Introducing Public Choice theory and Analytical Marxism	123

<b>Chapter 5: James Buchanan's theory of constitutional choice</b>	<b>125</b>
1. Why study James Buchanan?	125
1.1. Who is James Buchanan?	125
1.2. Why study James Buchanan?	128
1.3. Why not study another political economist or political philosopher?	130
2. Buchanan's intellectual background	131
3. Buchanan's theoretical assumptions	135
3.1. Methodological individualism	135
3.2. The Homo Economicus model	137
3.2.1. Instrumentality	137
3.2.2. Egoism	186
3.2.3. Exogeneity	139
4. Normative implications of Buchanan's theoretical assumptions	139
4.1. Normative implications of methodological individualism	140
4.1.1. Buchanan's theory of constitutional choice	140
4.1.2. Buchanan's view of the market	143
4.1.3. Buchanan's view of the state	144
4.1.3.1. Minimal protective and productive state	144
4.1.3.2. Contractarianism	145
4.1.3.3. Constitutional democracy	146
4.1.3.4. Unanimity and near-unanimity	148
4.1.4. Buchanan's proposals to downsize the state	151
4.2. Normative implications of the Homo Economicus model	152
4.2.1. Normative implications of instrumentality	152
4.2.2. Normative implications of egoism	154
4.2.2.1. Political failures	154
4.2.2.2. The state as Leviathan	156
4.2.2.3. The need for constitutional constraints on the state	157
4.2.3. Normative implications of exogeneity	157
5. Criticizing and complementing Buchanan's theoretical assumptions	158
5.1. Criticizing and complementing methodological individualism	158
5.2. Criticizing and complementing the Homo Economicus model	159
5.2.1. Criticizing and complementing instrumentality	160
5.2.2. Criticizing and complementing egoism	160
5.2.3. Criticizing and complementing exogeneity	164
6. Normative implications of criticizing and complementing Buchanan's theoretical assumptions	167
6.1. Normative implications of criticizing and complementing methodological and normative individualism	167
6.2. Normative implications of criticizing and complementing the Homo Economicus model	167
6.2.1. Normative implications of criticizing and complementing instrumentality	167



6.2.2. Normative implications of criticizing and complementing egoism	170
6.2.3. Normative implications of criticizing and complementing exogeneity	171
7. Conclusion	177
7.1. Buchanan as an economist	177
7.2. Buchanan as a political economist and a political philosopher	179
7.3. Buchanan in search of a better society	180
7.4. Summing things up	182
 <b>Chapter 6: The alternative of Samuel Bowles and Herbert Gintis</b>	 <b>185</b>
1. Why study Samuel Bowles and Herbert Gintis?	185
1.1. Who are Samuel Bowles and Herbert Gintis?	185
1.2. Why study Samuel Bowles and Herbert Gintis?	186
2. Bowles' and Gintis' intellectual background	187
3. Bowles' and Gintis' theoretical assumptions	191
3.1. The motivational dichotomy	192
3.1.1. Self-regarding preferences	192
3.1.2. Other-regarding preferences	193
3.2. Beyond the motivational dichotomy	194
3.2.1. Reciprocity	194
3.2.2. Strong reciprocity: definition	195
3.2.3. Strong reciprocity: empirical evidence	196
3.3. Standard accounts of the evolutionary origins of prosocial actions	197
3.3.1. Kin selection	198
3.3.2. Direct reciprocity	199
3.3.3. Reciprocal altruism	199
3.3.4. Indirect reciprocity	200
3.3.5. Strong reciprocity as a maladaptive trait	201
3.4. The evolutionary origins of strong reciprocity	202
3.4.1. Cultural group selection	203
3.4.2. Multilevel selection	205
3.4.3. The role of norms and emotions	206
3.5. Complementing Homo Economicus with Homo Reciprocans	209
4. Normative implications of Bowles' and Gintis' theoretical assumptions	211
4.1. The institutional dichotomy	211
4.1.1. Markets	212
4.1.2. States	212
4.2. Beyond the institutional dichotomy	213
4.2.1. Egoistic preferences and their amendments	214
4.2.2. Exogenous preferences and their amendments	216
4.3. Normative implications at the level of the state	217
4.3.1. Egoistic preferences and their amendments	219
4.3.2. Exogenous preferences and their amendments	220

4.3.2.1. Crowding effects: definition	221
4.3.2.2. Crowding effects: evidence	221
4.3.2.3. Crowding effects: psychological mechanisms	222
4.3.2.4. Crowding effects: normative repercussions	223
4.3.2.5. Crowding effects: theoretical relevance	224
4.4. Normative implications at the level of the market	226
4.4.1. Egoistic preferences and their amendments	226
4.4.2. Exogenous preferences and their amendments	226
5. Conclusion	229
5.1. Summing things up at the individual level	229
5.2. Summing things up at the institutional level	230
5.3. Bringing together the individual and the institutional level	232
<b>Conclusion of Part III</b>	<b>235</b>
1. Conclusions at the explanatory level	235
2. Conclusions at the normative level	238
<b>General conclusion</b>	<b>241</b>
1. Avoiding a pessimistic determinism	241
2. The value and limitations of this research project	245
<b>Bibliography</b>	<b>247</b>





# GENERAL INTRODUCTION

---

*“Is life just a game where we make up the rules while we are searching for something to say? Or are we just simply spirally coils of self-replicating DNA?”*

(Eric Idle, *The Meaning of Life Song*, in: *The Meaning of Life*, 1983)

## 1. Remarks regarding content

This dissertation’s opening quote is from a song in which Eric Idle from Monty Python’s Flying Circus ponders on the meaning of life<sup>1</sup>. It suggests some of the central issues that I will analyze in this dissertation. As such, it will regularly pop up throughout the text. Its first line nicely corresponds to a specific liberal view of individuals and societies. Social life is a game where the players themselves can choose the rules that govern their actions. Authors who defend this view nicely capture the ability of human beings to act freely within these rules and even intentionally change them. However, they seem to neglect that human nature is to some extent fixed. In their attempt to design desirable rules and institutions, they systematically assume people to be instrumentally motivated to satisfy their egoistic and given preferences. In contrast, the quote’s second line stresses the need to take human beings as they have evolved through the processes of natural selections. I will show how this emphasis on their evolutionary origins leads to a different stance in issues of institutional design.

As always, it is a combination of both thoughts that shows the route to follow. While people can to some extent design and reform the rules and institutions that guide their actions, they should not try to do so without taking into account their own nature. Nevertheless, even though people are made up of “spirally coils of self-replicating DNA”, they are not completely determined by their genetic constitution. After all, they have a unique ability to reflect upon their own nature. The underlying claim of this combination of both thoughts is

---

<sup>1</sup> By the way, the true meaning of life – which has intrigued philosophers for over twenty-five centuries – is actually quite simple according to Monty Python: “well, it’s nothing very special. Try and be nice to people, avoid eating fat, read a good book every now and then, get some walking in, and try and live together in peace and harmony with people of all creeds and nations” (Michael Palin, in: *The Meaning of Life*, 1983).

that one has to take people as they are when thinking up good rules and institutions. If one wants to “make up the rules” of the game of life, one has to start from a realistic view of its players.

In this respect, I will focus on the concept of rationality and argue that this can be conceived in a number of ways that are largely neglected in mainstream academic debates. While it is a quality that has evolved in the process of natural selection, it enables humans to distance themselves from mere animality (Nozick 1993: 181). In this respect, it is central to the essence of what it means to be human. However, it does not fit the economic conception of rationality, which is based on a narrow model of people as egoistically and instrumentally motivated. As a result, this dominant conception of rationality proves inapt not only for explanatory but also for normative purposes. To think up desirable institutions, one needs to start from a more realistic view of people as they have been formed by the processes of natural selection.

Because the notion of rationality encompasses a wide array of approaches, models and theories, I deem it useful to develop in the first part of this dissertation a systematic account of its different conceptions. The main aim here is to clearly define the relevant terms, models and assumptions. This way, I hope to avoid the rather vague terminology that has characterized rational choice theory over the last decades<sup>2</sup>. After stressing the importance of rationality for the study of human behavior, I focus on its standard, economic conception and its problems (first chapter). Next, I stress the relevance and value of a number of its alternative conceptions (second chapter).

In the second part, I apply these models and assumptions to voting decisions. Here, I show that the conventional economic conception of rationality is grossly inapt to explain why and how people vote in large-scaled elections (third chapter). In addition, I argue that the alternative conceptions of rationality are needed to solve some of the most fundamental problems that the economic one encounters (fourth chapter). In doing so, I show explicitly that each of these different conceptions has its value. Instead of replacing the economic one with its alternatives, I show that they can and should mutually complement each other.

---

<sup>2</sup> In the first two chapters of this dissertation, I will clearly define the terms rationality and rational choice theory. For now, suffice to say that I see no reason to distinguish between both terms, which is why I use them interchangeably in this dissertation and its title. I thus do not follow the strategy of Pettit (1993: 282) who stresses that “rational choice theory should be distinguished from the abstract theory of rationality”. Whereas he limits rational choice theory to what I label the economic conception of rationality (Pettit 1993: 272), I stress the importance of a number of alternative conceptions of what it means for a choice to be rational.

In the third part, I show how the theoretical assumptions and models lead to specific normative views on the basic institutional structure of modern societies. I aim to develop the normative implications of the different conceptions of rationality by focusing on two areas of research. The first is that of Public Choice theory. Through a detailed analysis of the work of James Buchanan, I show that these theorists in this field defend a specific stance on issues surrounding institutional design on the basis of the economic conception of rationality (fifth chapter). The second research area is that of Analytical Marxism. Here, I show how authors like Samuel Bowles and Herbert Gintis amend the assumptions of the economic conception of rationality and, as a result, come to support quite different normative conclusions (sixth chapter). In general, I want to argue that the goal of designing a desirable basic institutional structure for modern societies is inevitably based on the ways in which one models the motivations and actions of individuals. Finally, I try to formulate some general conclusions and point out what I believe is the value of this research project.

The central insights of this dissertation have already been published in a number of articles. The different conceptions of rationality in the first two chapters are discussed in Engelen (2005). The third and fourth chapter in which I apply these conceptions to the decisions whether and how to vote draw heavily on Engelen (2006). The main argument in the introduction to Part III has already been published in Engelen (2007b), which forms the most concise expression of my general line of reasoning. In addition, a more condensed version of the fifth chapter on James Buchanan's theory of constitutional choice, has been published in Engelen (2007c). The same goes for the sixth chapter on Samuel Bowles and Herbert Gintis, a part of which has been published in Engelen (2008b). While not addressing the main themes of this dissertation, the other articles (Engelen 2007a; 2007d; 2008a; Engelen & Nys 2008) all address some of its topics in one way or another.

## 2. Remarks regarding style and methodology

Before going more fully into the heart of the matter, I want to make some remarks regarding the style and methodology of the research I have conducted and the text that has resulted from it. First, this dissertation is written completely in English for the simple reason that this is the dominant language in the literature that I have studied. In order to contribute anything to the academic debates as they are held nowadays, one has no choice but to read, speak and write this scientific 'lingua franca'. Since I am not a native speaker, I

want to apologize for whatever flaws or errors remain in this text. In addition, I want to make clear that I make exclusive use of male terms (he, him, his) in referring to individuals, citizens, voters, et cetera. Of course, this is done purely for reasons of simplicity.

Second, I refer extensively to the works of the authors that I have studied. I do so in a detailed and meticulous manner, primarily because it enables readers to check what the authors themselves have to say on the issues at hand<sup>3</sup>. In my view, literature study – which is the only method used for this dissertation – is at its best when it stays close to the literature itself. This also suggests why I make extensive use of quotes. Whenever I think that others have formulated arguments or insights as clearly and succinctly as possible, I deem it best to quote them literally instead of simply paraphrasing them. As the attentive reader will notice, however, I do not simply repeat what has been written elsewhere already. It is my humble hope that the originality of this text lies primarily in bringing together different arguments and insights in previously unseen ways.

Third, I should perhaps explain why this PhD is completed at the Institute of Philosophy and not the Faculty of Economics or the Faculty of Social Sciences for that matter. After all, the bulk of the topics and literature analyzed here do not belong to philosophy in its strict sense. This is already clear if one throws a quick glance at the main parts of this dissertation. The first part largely centers round the conventional model in economics, namely that of the *Homo Economicus*. Since this also forms the starting point of my analysis of voting decisions, the same holds for the second part. One could also argue that my focus on institutional design in the third part is not particularly philosophical at all. This also shows in the fact that the main proponents in this dissertation's story line can hardly be defined as full-blooded philosophers. Whereas authors like Amartya Sen and Jon Elster (Chapter I) are situated respectively within economics and the social sciences, Public Choice theorists like Anthony Downs, Gordon Tullock (third chapter) and James Buchanan (fifth chapter) are (political) economists. In addition, authors like Samuel Bowles and Herbert Gintis (sixth chapter) would not consider themselves to be philosophers.

Nevertheless, I want to argue that this dissertation is a philosophical one in the end. The proposed amendments to the economic conception of rationality and

---

<sup>3</sup> I do not believe in the explanation of psychoanalysts that this is just a way of hiding my insecurity. Of course, I acknowledge that my rhetorical skills are no match for theirs: "it is like what happens if I say that Freudian theory is obvious nonsense: I am told that I only say that because of my relation with my mother. I protest that my mother had nothing to do with it. "See!" says my Freudian friend and walks away a sure winner" (Solow 1995: 241).



the minimal, broad and expressive conceptions that arise from them are distinctively philosophical in nature. In this respect, I make extensive use of insights from philosophers like Donald Davidson, Harry Frankfurt and Philip Pettit. While I refer to research within economics, psychology and the social sciences, the emphasis always lies on the conceptual and explanatory level rather than the empirical one. In addition, this dissertation's primary aim is a normative one, namely to think about the ideal institutional structure of modern societies from a rational choice perspective. As such, it is clearly situated within moral and political philosophy. In my view, this dissertation is therefore best characterized as multi-disciplinary, touching upon issues within economics, psychology, evolutionary theory, social sciences, political theory and – last but not least – moral and political philosophy.



# PART I

## RATIONALITY



# INTRODUCTION TO PART I

---

*“Man is a rational animal – so at least I have been told. Throughout a long life, I have looked diligently for evidence in favor of this statement, but so far I have not had the good fortune to come across it”*

(Russell 1950: 71)

## 1. Why rationality?

Rationality is a central concept within the analytical tradition in philosophy and the social sciences of the past century. But what is its exact meaning? While the economic conception is dominant within the social sciences, I aim to show the relevance of alternative conceptions. In my view, these are highly needed to understand what it means for humans to be rational. There are two interrelated reasons why I believe that this is a subject worth studying.

First, rationality is imbedded in the ways in which people understand each other. As soon as one knows what an individual's reasons are, one understands his actions. In everyday life, people unavoidably presuppose that the others with whom they interact are rational: “in our dealings with other people, we are compelled to treat them as, by and large, rational (...). To understand others, we must assume that, by and large, they have rational desires and beliefs and act rationally upon them” (Elster 1989b: 28). Without this presupposition, one would be at a total loss when communicating and interacting with others (Davidson 1980: 290; Davidson 1982: 301-303; Elster 1999: 141). In this respect, Elster (1979: 153) argues that “there is a general *presumption of rationality* in human affairs”. Donald Davidson (1980: 221) is known for his claim that “we cannot intelligibly attribute any propositional attitude to an agent except within the framework of a viable theory of his beliefs, desires, intentions, and decisions (...). Crediting people with a large degree of consistency (...) is unavoidable if we are to be in a position to accuse them meaningfully of error and some degree of irrationality”. If one does not presume that others generally are and want to be rational, one cannot accuse them of violating the requirements of rationality: “only a rational creature can be irrational” (Davidson 1982: 289).

The fact that people inevitably frame the actions of others in terms of intentions, beliefs, desires and reasons is captured in what has been labeled ‘folk psychology’. One cannot engage in social interactions without supposing that others are generally rational. In this respect, I want to argue that the different conceptions of rationality that I will outline are essentially a formalization of this folk psychology. This also explains the intuitive attractiveness of the concept of rationality and provides the main reason why I believe that social scientists and philosophers aiming to explain human actions would do well starting from this concept.

Second, people generally want to be rational and try to live up to the criteria that need to be fulfilled in order to be called rational: “we care about rationality because we want to be rational and want to know what rationality requires us to do” (Elster 1993: 189). To show this, one can refer to the fact that nobody is proud of his irrational behavior. Even though people will never be perfectly rational beings, they fundamentally conceive of themselves as rational. Rationality belongs to their deepest identity, what it means to be human. It is part of the self-image and the specific nature and status of the human species (Nozick 1993: xii). This does not mean that people are always rational, but that they have the capacity to be rational and that it lies in their nature to actualize this potentiality. Since people consider themselves to be rational, they want to express this in their actions. This is also why rationality plays such a central role in the way people understand themselves and others.

To the degree that they fulfill the requirements specified in the normative ideal of rationality, they will actually be rational. In this respect, I want to stress that I understand rationality primarily in its descriptive and explanatory guise. In contrast, Elster (1986: 1) believes that rationality should primarily be interpreted as a normative notion that tells people that they should be rational and prescribes what this entails. In my view, the normative and explanatory variants of the notion of rationality belong together. People *are* rational insofar as they follow the *prescriptions* of rationality. As John Harsanyi argues, “rationality is a *normative* concept: it points to what we *should* do in order to attain a given end or objective. But, even at a common-sense level, this concept of rationality does have important *positive* (non-normative) application: it is used for *explanation*, for *prediction*, and even for mere *description*, of human behavior. Indeed the assumption that a given person has acted or will act rationally often has very considerable explanatory and predictive power” (Moser 1990: 272). Rationality thus forms an ideal that people care about: “rational choice theory is far more than a technical tool for explaining behavior. It is also, and very importantly, a way of coming to grips with ourselves – not only what we should do, but even what we should be” (Elster 1993: 179).

All this, of course, does not imply that humans are rational all of the time. Like other animals, they are subjected to physical and physiological processes that fall completely beyond the domain of rationality. The fact that rationality is a normative ideal already suggests that people sometimes fail to realize it. It explains why they do not like to be accused of irrationality and employ indirect strategies to guard themselves against their irrational tendencies. I am thus happy to acknowledge that a whole range of phenomena is properly labeled irrational. Indeed, it would be troublesome for the concept of rationality if this were not the case. While interesting in their own right, however, these phenomena do not form the main focus of this dissertation. The fact that people are not rational all of the time, by no means implies that the concept of rationality should be discarded altogether within the social sciences. As I have suggested, it provides the necessary background against which the label of irrationality makes sense. An additional problem is that there is no overarching theory to explain when and why people are irrational. Elster (1989b: 35; 1986: 26-27) argues in this respect that rationality will remain the central paradigm within the social sciences, as long as its alternatives do not make up a simple, unified theory.

In trying to understand what it means to be human, I focus on the fact that most people conceive of themselves as rational beings. In this first part of this dissertation, I distinguish between different conceptions of rationality and show that it is not possible, and certainly not desirable, to reduce the concept of rationality to one of its meanings. In this respect, one of the aims of this dissertation is to end Bertrand Russell's quest that is expressed in this introduction's opening quote. This emphasis on rationality as an essential characteristic of humankind dates back to Aristotle, who refers to man as "animal with sapience" (*zoon logon echoon*). In Porphyry's tree, which was based on this view and remained influential throughout Western medieval philosophy, humans are defined as rational animals, which is still present in the notions of "Homo Sapiens" and "animal rationale". This definition returns during modern Enlightenment, most notably in the work of Immanuel Kant who stresses the rational capacities of humans to think for themselves.

## 2. Methodological individualism

This dissertation starts from the principle of methodological individualism, according to which social phenomena are best explained by referring to individuals and their actions. In its ontological sense, individualism is quite straightforward. Social phenomena do not exist independently from

individuals. It is simply impossible and inconceivable to encounter social facts like institutions. A society and its institutional structure only exist through the citizens that devise, uphold and embody them. From ontological individualism, one can conclude that individuals should form the center of explanations within the social sciences: “due to the biological separation of persons and the location of thought in the individual mind, action is a capacity of individuals and only individuals. The explanation of social action is the essential task of social science, from which it follows that individuals are the logical starting point of social theory” (Mayer 1994: 19). This, of course, leads to individualism in its methodological sense, which claims that one best understands social phenomena as resulting from the interactions between individuals.

In this respect, I deem it useful to refer to a number of distinctions made by Philippe Van Parijs (1990: 49-52), who defines methodological individualism (1) as the claim that social phenomena can only be explained as consequences of individual actions. This is to be distinguished from methodological intentionalism (2) that stipulates that these actions must be explained on the basis of individual reasons and methodological rationalism (3) that states that all actions must be explained on the basis of good reasons. However, the intentional images of individuals implied here are still more general than that of methodological utilitarianism (4) that claims that actions should be explained as means to serve the individual’s self-interest. Each of these methodological principles must be understood as prescribing a methodology to be employed in the social sciences. The restrictions these principles put on the scope of valid research within the social sciences are increasingly stringent.

Methodological individualism (1) is the least stringent principle, since it merely stipulates that valid social science explanations should start from individual actions. It excludes only those theories that explain social phenomena through the unmediated impact of supra-individual factors<sup>4</sup>. Methodological individualism is, however, perfectly able to incorporate the influence of social norms and institutions, insofar as these are understood as constraining or determining individual choices. Methodological intentionalism (2) and rationalism (3) are more stringent in that they exclude theories that refer to social phenomena as directing individual behavior in some unintentional way. While such theories fulfill the conditions of methodological individualism – the

---

<sup>4</sup> This also shows why I will not focus on what might be called ‘systemic’ or ‘evolutionary’ conceptions of rationality, according to which certain individuals with specific behaviors will be selected through the working of a number of evolutionary mechanisms. As will become clear later on, I surely do not want to deny the existence of such mechanisms. However, I do not believe that such processes should be framed in terms of rationality. In my view, this term should be reserved for individuals and their behavior.



individual remains the main explanatory variable – they do not live up to the criterion of intentionality, according to which individuals act purposefully. Methodological utilitarianism goes even further and excludes any theory that refers to factors other than the self-interest of individuals.

As will become clearer later on, this opens up a number of routes one can follow when trying to understand in what sense individuals make up a society. Whereas some theorists stress the need to model individuals in broad, intentional terms, others will stress the need to think of them as egoistically motivated. Since I believe that rationality has, or at least should have, a prominent place, I support not only methodological individualism (1) but methodological intentionalism (2) and rationalism (3) as well. In order to explain social phenomena, one best refers to individual intentions and reasons to explain what goes on in society. I believe that this results from the abovementioned fact that people interacting with each other on a daily basis inevitably interpret the other's actions in terms of intentionality and rationality. Without this basic folk psychology, it is impossible to understand any social, cultural, economic or political interaction between fellow citizens. As I will show more fully later on, I believe that methodological utilitarianism is one bridge too far, since it excludes a number of valid and relevant models within the social sciences.

One could add an extra category and label it methodological instrumentalism. This principle stresses that individuals should be modeled as being exclusively instrumentally motivated. Individuals are modeled as seeking to achieve their goals, whatever these are. On the one hand, it is less restrictive than methodological utilitarianism (4), because it does not necessarily limit the range of goals to that of utility-maximization. On the other hand, it is more restrictive than methodological intentionalism (2) and rationalism (3), because it requires that theories should restrict their scope to instrumental behavior. As will become clearer later on, an individual may have good reasons to do something intentionally (2 and 3) without aiming at the realization of some goal (4). Intentionality literally means to be directed at something, which is not necessarily to be equated with some goal. Methodological intentionalism (directed at reasons) is thus broader than methodological rationalism (directed at good reasons), which is still broader than methodological instrumentalism (directed at goals, which are only one type of good reasons). The fact that rationality and intentionality (aimed at reasons) do not have to be understood exclusively in an instrumental sense (directed at goals) will be the main argument in favor of developing an alternative, non-instrumental (expressive) conception of rationality.

After analyzing the role of institutions in society and their impact on individuals living there, I will introduce a few nuances with respect to this postulate of methodological individualism. Nevertheless, let us first focus on individuals and what it means for them to be rational.

# CHAPTER 1

## THE ECONOMIC CONCEPTION OF RATIONALITY

---

*“It may not be quite as absurd to argue that people always actually do maximize their self-interest, as it is to argue that rationality must invariably demand maximization of self-interest. Universal selfishness as actuality may well be false, but universal selfishness as a requirement of rationality is patently absurd”*

(Sen 1987: 16)

### 1. The economic conception of rationality and its failures

In order to analyze what the concept of rationality encompasses, I start by investigating how economists have developed it. After all, it is the economic conception of rationality, as it is expressed in the Homo Economicus model, that has become the dominant one within the social sciences<sup>5</sup>. In order to understand what economic rationality entails, I want to distinguish between the level of the relation between actions and their reasons and the level of the reasons themselves (Elster 1989a: 31). At each level, I will try to show that people do not always act as *Homines Economici*. As a matter of fact, they tend to violate almost every condition of economic rationality. I will thus analyze the different ways in which the Homo Economicus model fails to describe and explain individual behavior.

---

<sup>5</sup> Next to indirect support for this claim, like the vast amount of articles and books that focus exclusively on the Homo Economicus model when talking about rationality, I want to refer to Thomas Christiano (2004: 123) who argues that “mainstream rational choice theory, as I shall conceive of it, adheres strictly to the thesis of *homo economicus* (...). The mainstream approach is still the dominant approach to rational choice theory, though the revisionist approach has acquired many allies over the years”. The notion of ‘Homo Economicus’ or ‘economic man’ was implicitly present in the works of classical economists like Adam Smith (1776). Even though it was first articulated in detail by John Stuart Mill (1836), the term itself was coined for the first time only by John Kells Ingram (1888) (Engelen 2005: 69).

### 1.1. The level of actions and the relation with their reasons

In general, one can define the Homo Economicus as an isolated individual who in all circumstances aims to serve his self-interest by choosing freely between alternative actions of which he has calculated all relevant costs and benefits.

#### 1.1.1. *Maximization*

Its first characteristic is that of maximization or optimization. Economists typically construct a preference function or ordering and argue that the individual chooses the action that satisfies his highest-ranked preference: “there are two principal species of economic man: the consumer and the entrepreneur (...). The former wishes to maximize his utility, which is a known function of the goods and services he consumes; the latter wishes to maximize his profit” (Simon 1957: 197). The maximization of preference satisfaction is often considered to be the core of the economic conception of rationality (Slote 1989: 1).

However, individuals do not always maximize or optimize. In his search for a more realistic conception of rationality, Herbert Simon (1957: 205) stresses that rational people do not always aim to find the best course of action but are satisfied with what is “good enough”. The fact that satisficing – settling for less than the very best – is characteristic of human actions leads Simon to question whether maximization is necessary or sufficient to speak of rational action. Such satisficing behavior is so widespread that it is deeply ingrained in our everyday intuitions about the behavior of others. As Michael Slote rightly argues, it is perfectly possible for individuals to have good reasons to choose what is less than best for themselves: “our ordinary conception of moderation allows it to be rational, or not irrational, for an individual modest or moderate in his desires to choose a given course of action while knowing that he could do better for himself by another” (Moser 1990: 471). Somewhat paradoxically put, one can deem it better for oneself to do what is less good for oneself.

#### 1.1.2. *Instrumentality*

A second characteristic of the Homo Economicus is its instrumental motivation. Individuals are modeled as weighing off the expected costs and benefits of actions and choosing the action that they consider to be the best

means to reach a certain goal, namely the maximal satisfaction of their preferences. Economic rationality is thus completely consequentialist in nature: an action is rational if the agent thinks it has the best consequences (Elster 1991: 109). In this respect, one can speak of forward-looking or prospective rationality. The only thing that counts is what the action brings about, not how it has come about or how it is performed. Individuals are assumed to look at the outcomes of possible alternatives and choose the action they deem best given their ends on the one hand and the available means on the other hand (Pettit 2002: 207).

Here too, I want to stress that individuals do not always act instrumentally. In this respect, the *Homo Economicus* model is challenged by the *Homo Sociologicus* model (Boudon 1993: 5; Elster 1989a: 113; Elster 1991: 109). According to this non-consequentialist model, individuals base their actions on socially and culturally inherited norms, without taking into account the future consequences of their actions. In contrast with the forward-looking or prospective attitude of the *Homo Economicus*, the *Homo Sociologicus* is characterized by a backward-looking or retrospective attitude (Elster 1991: 109; Vandeveldt 1994: 95). Because of its non-instrumental character, norm-guided behavior is clearly economically irrational.

According to Robert Sugden, a norm is best defined as “any regularity in behavior within a given community which is generally expected, not only in the empirical sense, but also in the normative sense. That is, conformity with the regularity elicits sentiments of approval while failure to conform elicits disapproval” (Ben-Ner & Putterman 1998: 79). Norms can be cultural, political or moral – depending on the behavior they prescribe – but they are always social, since they are shared by the members of the same group. Sugden’s emphasis on the enforcement of norms through social sanctions perfectly fits the conventional strategy of economists to understand norm-guided behavior in purely calculating and instrumental terms. According to Robert Frank, such attempts are unconvincing, because they do not take into account the fact “that people often follow norms even when external sanctions are not a credible threat” (Ben-Ner & Putterman 1998: 283). Through a process of internalization, norms have an additional impact on behavior that is more internal. Norms can thus constitute people’s goals: “internalization moves norms from constraints that one can treat instrumentally toward maximizing well-being to norms that are then valued as ends rather than means” (Bowles & Gintis 2003: 440). Norms can thus define preferences, rather than that they form means towards or constraints upon their realization (Gintis 2003: 156). Sociologist James Coleman (1990: 242) takes all this into account by defining norms as socially shared guidelines that specify “what actions are regarded by a set of persons as proper or correct, or improper or incorrect”. These

behavioral criteria are transmitted across generations through education, socialization and social learning.

From a large survey of randomly selected cultures, Sober and Wilson (1980: 166) conclude “that human behavior is strongly regulated by social norms in most cultures around the world and presumably throughout human history”. It is best understood as the expression of specific emotions, triggered by the fact that one has internalized certain criteria of what kinds of behavior is appropriate within one’s group. Norm-guided behavior is thus typically motivated by emotions like shame, which mutually support more formal ways of enforcing norms through external sanctions: “where the emotion of shame is common, punishment of antisocial actions will be particularly effective and, as a result seldom used” (Bowles & Gintis 2003: 439). While strong social norms reduce the need of institutions enforcing them, they thus complement – rather than substitute – each other: “it is widely held that strong social norms reduce the burden on law enforcement; that laws supported by social norms are likely to be significantly more enforceable; and that laws that are formulated in ways that are congruent with social norms are much more likely to be enacted than laws that offend such norms” (Etzioni 2000: 159).

It has been argued that norm-guided behavior can still be understood as instrumentally motivated towards the maximal satisfaction of one’s preferences. However, I want to argue that it forms an inadequate account of norm-guided and dutiful behavior. First, people who act upon a sense of duty are concerned neither with circumstances nor with outcomes. They act as they do, regardless whether others do so as well. Their main motivation is their firm belief that this is the right thing to do and not that this is the best means to achieve a certain goal. While it may be useful to distinguish between norm-guided and dutiful behavior, I want to argue here that duties can rightly be understood as internalized moral norms. In contrast with the forward-looking nature of instrumental actions, dutiful and norm-guided behavior is thus typically backward-looking. Second, people who act upon an internalized norm of fairness are concerned not with outcomes, but with circumstances. They help others or contribute to the public good, only if others do their share as well. Fair friends all want to do the dishes, even if this leads to an overcrowded kitchen and thus to an inefficient way of achieving the desirable outcome (Elster 1989a: 56-57). Clearly, both kinds of norm-guided behavior are non-instrumental in nature.

### 1.1.3. *Intentionality*

A third characteristic of the Homo Economicus is that of intentionality. An individual acts intentionally if he has reasons for his actions and if these reasons cause his actions in appropriate ways (Davidson 1980: 3-19). If one wants to understand why the individual at hand acts like he does, one should refer to his reasons<sup>6</sup>. Reasons constitute the purpose of his actions. I thus agree with John Ferejohn and Debra Satz (1996: 80-82) who argue that good explanations within the social sciences should be based on Davidson's charity principle according to which individuals in general attribute intentionality to others. In this respect, I also sympathize with their view that rational choice theory should be recast on the basis of folk psychology, which centers round this notion of intentionality (Pettit 2002: 195).

Nevertheless, individuals do not always act intentionally. According to Raymond Boudon, the main distinction between the Homo Economicus and Homo Sociologicus model is that the former stresses intentionality, whereas the latter does not. In the latter, behavior is assumed to be "basically causal (behaviour has to be explained by causes rather than reasons) and in that sense irrational or rather arational (no reasons, but causes)" (Boudon 1993: 5). As will become clear later on, I want to understand norm-guided behavior as intentional nonetheless.

To be sure, I fully acknowledge that some behaviors can no longer be called intentional. First, there are bodily movements that lie completely beyond the individual's control. Someone who trips over a stone is not acting intentionally. Second, some behaviors, like ticks, are performed on an entirely automatic basis, without any reason whatsoever. These two sorts of behavior are completely unintentional and arational in nature. In such cases, the individual is not motivated by reasons at all. Third, there is the phenomenon of weakness of the will or 'akrasia', which constitutes the paradigmatic example of irrationality at the level of the relation between an individual's actions and his reasons. A weak-willed individual fails to act according to what he believes is best and prefers most. An addict, for example, may believe that drinking is bad and

---

<sup>6</sup> For reasons that will become clear, I postpone a more precise definition of what exactly is meant by a reason. For now, it should be clear that I do not agree with Elster (1986: 12-13; 1989b: 3-4), who understands intentionality in exclusively instrumental terms. In his view, an action is intentional when it is aimed at the achievement of a goal: "reasons are reasons for the action if it is a way of realizing the desire (given the beliefs)" (Elster 1983: 3). In contrast, I follow Davidson's definition of intentionality, which allows me to distinguish more clearly between the instrumental and the intentional aspects of economically rational actions.

prefer to quit, but is nonetheless overpowered by a craving for alcohol<sup>7</sup>. Here, the causal force of the visceral urge to drink turns out to be stronger than that of the reasonably deliberated intention not to drink. Such an addict thus acknowledges that he has good reasons not to drink but nevertheless fails to translate these into action. His drinking behavior results from a loss of control over his own actions, which leads him to act against his own better judgment (Davidson 1980: 21-22, 42; Davidson 1982; 294-297; Elster 1979: 174-175). While the latter provides the better reason, it is bypassed by a causal force that does not count as a good reason itself. The resulting behavior is therefore obviously irrational.

## 1.2. The level of the reasons themselves: beliefs

Here, I distinguish between two sorts of reasons, namely beliefs and preferences, which is the term economists employ to model desires. Since the rationality of actions is dependent on such beliefs and preferences, Robert Audi rightly argues that it is necessary to “explore what constitutes the rationality of motivational and cognitive elements, and how it bears on that of actions based them” (Moser 1990: 416). At this level of the reasons one can have for one’s actions, the economic conception of rationality indeed adds further requirements. Most economists do not specify explicitly which criteria are to be satisfied in order to speak of rational beliefs. Nevertheless, I try to lay bare some implicit assumptions and show that these are often inappropriate.

### 1.2.1. *Consistency*

First, economists systematically stress the need for beliefs to be consistent with each other. A set of beliefs is internally consistent if they can all be true and believed at the same time (Elster 1983: 4-5, 61). Nevertheless, a lot of people often have inconsistent beliefs that cannot be true together at the same time. Take the example of an addict who is convinced that alcohol is bad for one’s health, but also believe that one (more) drink will not harm anybody. While inconsistencies often continue to exist, because they belong to different domains of one’s life, such individuals can rightly be called irrational. When

---

<sup>7</sup> For those who are interested in the extent to which phenomena like addiction and weakness of the will pose a threat to the rationality of individuals, I refer to my chapter in the book ‘Metallica and Philosophy’ (Engelen 2007a: 33-35). Here, I analyze more fully how an addiction leads to irrational actions, beliefs and preferences.



such inconsistencies are made explicit, the individual at hand will usually adjust them (Elster 1979: 150-151).

### *1.2.2. Perfect rationality*

Second, economists employ more demanding requirements concerning the quality and quantity of information on which to ground one's decisions. The Homo Economicus is typically assumed to possess perfect information, which means that he possesses all the relevant information of the alternative actions and their consequences. However, most decisions – if not all – are based on imperfect information. Often, people gather only limited information about the available options. In addition, the Homo Economicus has the cognitive capacities required to perform whatever calculations are necessary to select the best of these alternatives. In this respect, one can speak of “perfect rationality” (Simon 1957: 199). Herbert Simon (1957: 241) summarizes all this as follows: “traditional economic theory postulates an “economic man,” who (...) is assumed to have knowledge of the relevant aspects of his environment which, if not absolutely complete, is at least impressively clear and voluminous. He is assumed also to have (...) a skill in computation that enables him to calculate, for the alternative courses of action that are available to him, which of these will permit him to reach the highest attainable point on his preference scale”. Here too, people are not rational in the perfect sense of the word. They simply lack the cognitive capacities to calculate exactly which consequences will occur after performing specific actions. Studies in the field of experimental economics have shown that people generally do not base their actions on such heroic calculations (Vandevelde 1994: 97).

Simon (1957: 202) thus rightly criticizes the all too demanding cognitive requirements of the Homo Economicus model, which is wrong “in seeking to erect a theory of human choice on the unrealistic assumption of virtual omniscience and unlimited computational power”. His own model of bounded rationality is explicitly based on a more realistic view of the limited cognitive capacities of humans (Simon 1957: 241-243, 273-278; Simon 1990: 200). In Simon's own words, it “describes the way people, in fact, make decisions and solve problems. The theory has an increasingly firm empirical base as a description of human behavior” (Simon 1990: 200). He stresses that choices are always made in a context that is perceived by individuals in a specific way. Individuals always frame their choices by concentrating on specific aspects of the options, which ensures that the available information is well-organized (Tversky & Kahneman 1981: 453). Because of their limited capacities to deal with the complexities of the world they live in, they tend to rely on the rules

and understandings on how to act that are imbedded in culturally inherited norms, traditions and habits (Hargreaves Heap 1989: 157-160).

### 1.3. The level of the reasons themselves: preferences

Economists are more explicit about their requirements to speak of rational preferences. As Robert Sugden rightly argues, “neoclassical economics is based on very restrictive assumptions about the nature of preferences” (Ben-Ner & Putterman 1998: 75). The Homo Economicus is assumed to have consistent, transitive, complete, continuous, exogenously given and egoistic preferences (Elster 1982: 122; Vandeveld 1994: 89-92).

#### 1.3.1. Consistency

The first requirement is that, like beliefs, preferences should be consistent. A rational individual always prefers *a* over *b* or *b* over *a*, but never both at the same time. However, people often have inconsistent preferences. An addict, for example, may want to quit but still long for another drink. In addition, people often have different preferences when the available options are presented in different ways. According to Tversky and Kahneman (1981: 457), the problem here lies in an inconsistency in the set of preferences. Take the case in which consumers have to pay extra if they use a credit card or receive a discount if they pay cash. Since the alternatives remain the same, different ways of framing them should have no impact on preferences. Yet, experiments show that consumers choose to pay cash more often in the first than in the second case<sup>8</sup>. This is clearly irrational, since “rational choice requires that the preference between options should not reverse with changes of frame” (Tversky & Kahneman 1981: 453).

---

<sup>8</sup> According to Elster, the problem lies not with the preferences themselves, but with the causal influence of irrelevant changes in the choice situation on the preferences at hand: “*preference change by framing* occurs when the relative attractiveness of options changes when the choice situation is reframed in a way that rationally should make no difference” (Elster 1983: 25). I will explain this view more fully, when discussing Elster’s broad conception of rationality.

### 1.3.2. *Transitivity*

The second, related requirement is that a preference function should be transitive. A rational individual always prefers  $a$  over  $c$  if he, at one moment in time, prefers  $a$  over  $b$  and  $b$  over  $c$ . Nevertheless, preferences sometimes turn out to be intransitive. The problem here is that it is impossible to let an individual choose between  $a$  and  $b$ ,  $b$  and  $c$  and  $a$  and  $c$  at the same time. One can, for example, prefer  $a$  over  $b$  and  $b$  over  $c$ , but ultimately prefer  $c$  over  $a$ , because one has gained additional information about one of the alternatives in the meantime<sup>9</sup>.

### 1.3.3. *Completeness*

The third requirement is that a preference function should be complete. A rational individual is always able to choose between  $a$  and  $b$ . Either he prefers  $a$  over  $b$ , or he prefers  $b$  over  $a$ , or he is indifferent. This is not always the case either. Preference functions are sometimes incomplete in the sense that the individual cannot say which of the available alternatives he prefers. A well-known example is that of Buridan's ass starving because he cannot make up his mind on which stack of hay to prefer (Vandevelde 1994: 91-92). Similarly, the choice to kill either my daughter or my son can be so bewildering that I simply cannot decide<sup>10</sup>.

### 1.3.4. *Continuity*

The fourth requirement is that a preference function should be continuous, which means that a rational individual is always able to weigh off different options against each other. This implies that an increase in one good should

---

<sup>9</sup> Someone who receives additional information will typically adapt his preference ranking accordingly. If he does not, this will lead to intransitivity and thus irrationality. A girl who wants to get to know a boy initially prefers a date with him ( $b$ ) to a lonely evening on her room ( $c$ ). Moreover, she would even like to invite the boy straight to her room ( $a > b$ ). But when she finds out that the boy is a drug addict, she will prefer  $c$  over  $a$ , which leads to intransitivity. However, this is only so if she sticks to her other preferences. In contrast, she can stay rational by adjusting these preferences, for example by preferring a lonely evening over company of the boy ( $c > b$ ) or by preferring a date in public over an evening with the boy in her room ( $b > a$ ).

<sup>10</sup> This example is, of course, inspired by the novel 'Sophie's Choice' (Styron 1979).

always be able to compensate a decrease in another good. This requirement allows economists to construct a single, comprehensive, one-dimensional function that orders all of the individual's preferences (Moser 1990: 2, 195, 275).

However, this does not always do justice to the truth. Preference functions are often not continuous. Individuals with lexicographic preferences do not accept less of a specific good, even if this is compensated by much more of another good. Certain things, like duties, moral principles or love, fall beyond the calculating reasoning in which everything has its price. One could even argue that what is essential to humans is the extent to which they can distance themselves from such calculations (Vandevelde 1994: 92-93). In this respect, Philip Pettit (2002: 189) distinguishes between a rational system, which systematically and automatically calculates which option is optimal, and a reasoning system, which is characterized exactly by this capacity to reflect: "when rational explanation assumes an interpretative or hermeneutic form (...), then it casts the person as a reasoning or ratiocinative subject, not merely as a rational system. The rational system – the ideal subject of decision theory – may realize its rationality on the basis of a purely sub-personal mode of organization and attunement; it need not have what we would describe as a mental life. The ratiocinative system – the sort of system that our species implicitly or explicitly typifies – may be a more or less rational system in this sense but it is also something else besides; it is a rational system that attains rationality, to the extent that it does, on the basis of attention to reasons and what reasons require". Whereas the first is properly explained in exclusively causal terms (*Erklären*), explanations of the latter's behavior crucially involve some form of interpretation (*Verstehen*). As one can see here, Pettit criticizes the economic conception of rationality as missing the very essence of what it is to be human. I will go into this issue more fully in the next chapter.

### 1.3.5. *Exogeneity*

The fifth assumption economists employ is that preferences are exogenously given (Ben-Ner & Putterman 1998: 3, 75). There is a widespread tendency among economists to consider preferences as the starting point for their explanations, without asking how these come into being and are transformed. According to Amitai Etzioni (2000: 166), the assumption "that preferences are fixed (...) is still very widely held by neoclassical economists". It is based on their basic view of individuals as isolated, free individuals who are to judge for themselves what they want and how to get it. As such, this assumption is grounded on the fundamental notions of freedom of choice and consumer

sovereignty (Miller 1990: 146-147). If preferences themselves change under the influence of all sorts of factors and processes lie beyond the individual's control, the latter's behavior seems to be unfree and thus hardly rational.

This requirement is problematic in quite a few instances, since preferences change endogenously through a number of mechanisms. Even though economists tend to shy away from issues concerning the formation of preferences, research in this respect can complement rather than substitute economic models. I thus deem it useful to analyze more fully how individuals adapt their preferences and how “the structure of everyday social interactions affects both beliefs and preferences” (Henrich et al. 2004: 47).

Norms, for example, have a clear impact on individual preferences. However, they have often been analyzed as part of the environment an individual faces when making a decision. They constrain the available options or at least influence their costs and benefits by rewarding some and punishing other options. However, as Etzioni (2000: 161) rightly remarks, norms have a more internal impact, next to such an external impact: “social norms are not merely a part of the actors' environment but also affect their intrinsic predispositions”<sup>11</sup>. The problem most neoclassical economists have in thinking this more internal impact of norms on preferences is that they “tend to assume that preferences are given and fixed” (Etzioni 2000: 166).

### 1.3.6. *Egoism*

The sixth requirement is that preferences should be egoistic. While controversial, this is often regarded as the pre-eminent characteristic of the Homo Economicus model. John Stuart Mill (1836: 321) refers to egoism as the defining characteristic of economics, which analyzes the individual “as a being who desires to possess wealth, and who is capable of judging of the comparative efficacy of means for obtaining that end. It predicts only such of the phenomena of the social state as take place in consequence of the pursuit of wealth. It makes entire abstraction of every other human passion or

---

<sup>11</sup> This insight has a number of important repercussions: “the significance of the distinction between treating social norms as part of the actor's environment, affecting costs and constraints, and treating them as factors that shape the actor's predispositions stands out in several important respects (...). The contrast is apparent in the levels of compliance with social norms achieved, the level of social order sustained, and the relative costs of enforcing norms” (Etzioni 2000: 163). If people follow norms out of fear for social sanctions, the level of adherence will be lower than it is when people have internalized these norms. Later on, I will go deeper into the normative implications of these issues.

motive”. As Francis Ysidro Edgeworth (1881: 16) notes, “the first principle of economics is that every agent is *only* actuated by self-interest”. Both authors agree that individuals might be motivated by considerations beyond self-interest, but stress that the contribution of economics lies in explaining social phenomena by abstracting from such motivations.

According to Amartya Sen (1977: 317; 1987: 1), the assumption of egoism forms an integral part of economic models throughout the twentieth century. In his view, the equation of rationality with self-interest maximization has a long tradition: “the self-interest interpretation of rationality goes back a long way, and it has been one of the central features of mainline economic theorizing for several centuries” (Sen 1987: 15). Sen argues that “many economic models tend to proceed as if the assumption of universal pursuit of self-interest is the only motivation that can be legitimately presumed in serious economic analysis” (Ben-Ner & Putterman 1998: vii). In his view, “the world is made to fit this momentous assumption, rather than the assumption being made to fit the world” (Ben-Ner & Putterman 1998: viii).

Richard Thaler (1994: 8) agrees that “much economic analysis – and virtually all game theory – starts with the assumption that people are both rational and selfish”. According to David Schmitz (1993: 54), *Homo Economicus* is a combination of the assumptions of instrumentality and egoism: “the result is the *Homo economicus* model of rational agency. The *Homo economicus* model leaves no room for altruism”. It assumes pure self-regard, which implies that individuals “care about no one’s welfare other than their own and recognize no constraints on their pursuits beyond those imposed on them by circumstances” (Schmitz 1993: 52). The conventional *Homo Economicus* model thus arises as soon as one combines instrumentality and maximization with the claim that preferences are typically self-regarding in nature: “the conventional instrumentalist conception of rational choice sometimes is combined with a substantive assumption of mutual unconcern (the assumption that rational agents are immediately concerned with no one’s welfare but their own). This combination produces a model of rational agency that has become notorious in the social sciences: *Homo economicus*. By hypothesis, *Homo economicus* is purely self-regarding; she cares about others only because and insofar as responding to the actions of others is a means to her own welfare” (Schmitz 1994: 250).

#### 1.3.6.1. Self-regarding preferences

Pettit (2002: 223) argues that “the main question here is how far economists cast human beings as egocentric in their desires”. He distinguishes between

different variants of this assumption, which he believes are all implicitly present in economic models. According to its most demanding interpretation, egoistic literally means selfish, self-interested or self-regarding. In this view, the Homo Economicus is only concerned with the maximization of his self-interest, often understood in hedonistic and materialistic terms. Individuals are assumed to seek wealth, strive for pleasure, avoid pain and prefer more rather than less goods (Mill 1836: 322; Vromen 1998: 24-25).

A less stringent interpretation states that egoistic considerations will outweigh other considerations when both come into conflict. According to Pettit (2002: 226), economics “is committed to the assumption that people’s self-regarding desires are generally stronger than their other-regarding ones (...). The assumption that people are relatively self-regarding in their desires shows up in the fact that economists tend only to invoke relatively self-regarding desires in their explanations and predictions”. This assumption is not as stringent as the one above, because it allows economists to recognize the presence of such considerations, even though they are less powerful than self-regarding ones. Even less stringent is “what Wicksteed called ‘non-tuism’: participants in market transactions are assumed to be concerned about the extent of their own holdings (for whatever ultimate reason), but to have no direct concern for the welfare of their contractual partners” (Miller 1990: 152). Philip Wicksteed (1910: 158) indeed stressed that he wanted to study “that complex of relations into which we enter with other people, and lend ourselves or our resources to the furtherance of their purposes, as an indirect means of furthering our own”. The interests of others may come into focus, but only if serving them serves one’s own interests<sup>12</sup>.

Together with prominent authors like Sen, Thaler, Schmitz and Pettit, I believe that the assumption of egoism in its stringent interpretation is still rightly considered as an integral part of the Homo Economicus model. Even though it is a theoretical abstraction that may not be defended by all economists in all of its aspects, it nevertheless lies at the basis of the bulk of economic handbooks, explanations and models. I thus feel confident in claiming that a lot of economists still implicitly or explicitly rely on the

---

<sup>12</sup> According to Thomas Christiano (2004: 122-123), this variant does not diverge much in practice from the more stringent interpretation: “for the most part, rational choice theorists think of the mainstream approach as committed to the thesis that agents are self-interested (...). But some theorists have used the idea that the agents are “nontuistic” in the sense that they are not inherently concerned with the interests of those they strategically interact with (...). The differences between these two types of motivations are important, but the mainstream approach treats the nontuistic motivation as if it operated in the same way as the purely interested motivation that takes little or no account of the interests of others”.

assumption that “agents care only about what they personally gain and lose through an interaction” (Gintis et al. 2005: 6). As I will show more fully later on, it is this assumption that forms the basis of the centerpiece of modern-day economics, namely Adam Smith’s invisible hand mechanism, according to which egoistic individuals will reach socially desirable outcomes in market settings<sup>13</sup>.

However, the simple fact that people manage quite well to live together already suggests that most of them do not act in purely opportunistic ways: “under strictly self-interested motivation (...) human societies as we know them could hardly exist (...). Everyone would cheat whenever it was safe to do so; hence everyone would also have to be continually concerned with being cheated” (Margolis 1984: 121). Conventional economists “typically assume that people’s preferences are ‘self-regarding’ and ‘outcome oriented’. In other words, people want stuff for themselves, and care only about their personal costs in getting what they want. There is much evidence, both from within economics, and from other disciplines, that this view of human nature misses a lot: people care both about other people, and about how social transactions occur – not just the outcomes” (Henrich et. al. 2004: 1).

I thus agree that there is need for “economists, anthropologists, psychologists, and behavioral scientists to develop systematically richer models of preferences according to which people take account of the effects of their actions not only on themselves, but on others as well, and in which the processes determining outcomes matter as well as the outcomes themselves” (Henrich et. al. 2004: 1-2). Next to the self-regarding preferences on which economists typically focus, individuals thus have other- and process-regarding preferences (Ben-Ner & Putterman 1998: 7, 22-23, 283-285). People care not only about the outcomes of social interactions on their own well-being, but also about the well-being of others and about the way these interactions themselves proceed: “*self-regarding preferences* are the essence of *homo economicus*. *Other-regarding preferences* reflect concern either for the overall well-being or for certain of the activities or outcomes of other individuals. They have both altruistic or benevolent and envious or malevolent varieties (...). Individuals care about the manner in which they themselves and others behave, including the ways in which they attain outcomes of interest. These are *process-regarding preferences*” (Ben-Ner & Putterman 1998: 20). Experimental evidence confirms that actions are often based on such other- and process-regarding preferences (Ben-Ner &

---

<sup>13</sup> To be sure, Smith also focused on the role of other considerations in human behavior. Nevertheless, these writings “have become neglected as these considerations have themselves become unfashionable in economics” (Sen 1987: 28).



Putterman 1998: 11). In what follows, I will analyze more fully both sorts of preferences.

### 1.3.6.2. Other-regarding preferences

People with other-regarding preferences want to serve another's interests rather than one's own. David Miller (1990: 101) aptly defines altruism as a concern for the interests of others, which motivates one to act on their behalf. He distinguishes between the calculating altruist who stops giving as soon as someone else gives, the reciprocal altruist who only gives when others do their fair share as well, the conscientious altruist who gives out of a sense of duty, no matter what others do, and the superconscientious altruist who fulfills not only his own duty but also that of others (Miller 1990: 111-117). Whereas the first primarily aims to bring about particular consequences (future) and the second focuses on the circumstances in which he gives (present), the third and the fourth give because they have internalized the norm that this is the right thing to do (past).

Schmidtz (1993: 52-53) distinguishes between two forms of other-regarding preferences. First, people are often motivated by concern, which means that they care about the welfare of others. Second, they are often motivated by respect, which means that they adhere to self-imposed constraints on what they may do to others when pursuing their own goals.

Sen (1995: 2-3) also stresses the fact that people often incorporate the interests of others in their considerations. If altruistic actions serve not only the interests of others, but also one's own interests, he labels them acts of sympathy (Sen 1977: 326). Crossing a beggar, it is in my interest to give him money if I believe that this will make me feel good. Some altruists do indeed derive satisfaction from the act of giving or helping itself. In essence, this is still egoistic and thus economically rational (Margolis 1984: 21). Sympathy is to be distinguished from commitment, which entails helping others, even if this goes against one's self-interest (Sen 1977: 326-333). Crucial here is that this can no longer be reduced to a feeling of pleasure or satisfaction.

Similar motivations are "morale, civic virtue, social capital, trust, and intrinsic motivation" (Ben-Ner & Putterman 1998: 439). The Homo Economicus model fails to incorporate them, because it focuses exclusively on extrinsic motivation through rewards and punishments. In contrast, intrinsic motivation is crucial to understand people who do something simply because they enjoy it, because they feel it is their duty or because they owe this much to the ones they

care about. Later on, I will go into the importance of intrinsic motivation more fully.

The fact that people care about others does not immediately fit the narrow Homo Economicus model: “it is commonplace to point out that the Homo economicus model (...) does not accurately describe human agents. Like Homo economicus, we have preferences, but unlike Homo economicus, we have preferences directly relating to the welfare of others. That is the obvious and much celebrated difference between Homo economicus and us” (Schmidtz 1994: 150). The problem lies in the fact that conventional economists have equated rationality with self-interest. Sen (1987: 12) aptly summarizes their two-fold strategy as follows: “(1) identifying actual behaviour with rational behaviour; and (2) specifying the nature of rational behaviour in rather narrow terms (...). There are two predominant methods of defining rationality of behaviour in mainline economic theory. One is to see rationality as internal *consistency* of choice, and the other is to identify rationality with *maximization of self-interest*”.

As this chapter’s opening quote by Sen stresses, the problem lies primarily with the second step. There is no need to assume that rationality presupposes narrowly self-interested actions or preferences. While I do not want to deny that rationality can be defined in self-interested terms, I deem it important to point out that this does not correspond to a more intuitive and common meaning of the word: “unless you are prepared to argue that, by definition, social motivation must be irrational unless it can be formulated in terms of self-interest, there appears to be something wrong with the conventional definition” (Margolis 1984: 16). The main strategy that I will follow in this dissertation is thus not to do away with the notion of rationality altogether but to propose alternative conceptions of what it entails. This way, I hope to show that altruism is not necessarily irrational. After all, as Schmitz (1993: 54) rightly notes, one can have perfectly good “reasons for altruism (...). What we want is an account according to which it is rational for us to have concern and respect for others”. Before developing my own thoughts on what these reasons for altruism might look like in the next chapter, I first want to show how economists try to conceptualize them and which problems their strategy encounters.

#### 1.3.6.3. Reducing other-regarding preferences to egoism

In spite of these problems, economists typically hold on to the stringent interpretation of egoism: “there is a strong tendency among economists and

philosophers to (...) reduce apparently altruistic behaviour to ultra-subtle forms of self-interest” (Elster 1982: 123). They develop several strategies to frame altruism in self-interested terms. The most straightforward strategy postulates a ‘warm glow’ effect and argues that altruistic actions always produce a feeling of satisfaction or help sooth one’s conscience. Kenneth Boulding, for example, stresses that “if my perception of an increase in your welfare makes me feel better off, I am benevolent” (Koslowski: 1985: 101). This strategy thus simply defines an individual’s self-interest as encompassing all sorts of motivations (Van Liedekerke 2000: 108).

Broadly speaking, there are two problems with this strategy. First, such a linguistic move turns the notion of self-interest into a completely empty “passe-partout” (Ben-Ner & Putterman 1998: 156). This makes it hard to falsify or verify experimentally. Because this strategy of systematically postulating a warm glow effect is ‘ad hoc’, its scientific legitimacy is troublesome. After all, it is simply impossible to (dis)prove the existence of such psychic benefits each and every time someone acts altruistically. While “it is possible to define a person’s interests in such a way that no matter what he does he can be seen to be furthering his own interests in every isolated act of choice” (Sen 1977: 322), one should try to explain actions rather than simply rationalize them. In this respect, I propose to hold on to a narrow definition of self-interest in terms of personal benefits. While these can be materialistic, psychic or symbolic, they ultimately refer to the well-being of the individual itself. In this respect, one can argue that an action that is beneficial to someone else but performed with the primary purpose of gaining some personal benefit is not genuinely altruistic in nature. Even if such benefits arise, they do not necessarily form the only or even the main drive behind the action. The fact that one helps others because it makes one feel good does not necessarily mean that one does so *only* because it makes one feel good (Sober & Wilson 1998: 244-247). Moreover, if I give to someone I love, I might take pleasure in his pleasure, but this is simply a by-product and not the primary motive (Elster 1989a: 53). As Schmitz (1994: 250) aptly puts it, “we reap psychic rewards from helping others, of course, but this fact does not begin to suggest that our real objective is psychic benefit rather than other people’s welfare. On the contrary, there can be no psychic reward for helping others unless we care about others”. The satisfaction of preferences thus cannot be equated with the feeling of satisfaction this can but not necessarily does induce.

Second, one can refer to empirical evidence that supports the common sense view that people sometimes want to help others, even if this does not cause them to feel happy themselves. Altruistic individuals are primarily motivated out of a concern for others, even at one’s own cost (Walsh 1996: 167). In this respect, Jane Mansbridge refers to interviews with people stressing that they act

altruistically without any concern for personal benefit whatsoever (Ben-Ner & Putterman 1998: 151). In his experimental psychological research, Daniel Batson convincingly shows that some acts of altruism cannot be explained by a desire to avoid negative emotions (people tend not to turn their back from misery), a desire for approval (people help even if nobody is watching) or a desire for shared joy (people help even if they do not know whether the other enjoys it) (Ben-Ner & Putterman 1998: 158-159). In conclusion, there is no scientific evidence whatsoever that self-interest is the exclusive or even the main concern of individuals. Common sense tells us that people are typically motivated by a complex combination of different considerations (Ben-Ner & Putterman 1998: 5, 163). According to Sen (1987: 19), “the real issue is whether there is a plurality of motivations, or whether self-interest *alone* drives human beings”. It should by now be clear that the first view is the right one. While egoistic and other considerations can be closely related, this does not mean that the latter can systematically be reduced to the former (Sen 1987: 43-44).

I believe these problems show that people can have reasons for being altruistic that go beyond a self-interested cost-benefit analysis. In this respect, I want to refer to Schmidtz (1993: 65), who starts off by arguing that it is in one’s self-interest to care about things besides one’s own interests: “one crucial way to nurture self-regard is to nurture concerns that give us more to live for than we have if we care only about ourselves”. If one has time and means to go beyond the purely egoistic concern for survival, it is in one’s interest to belong to a community, to care about and to be important for others. It may thus be economically rational to develop altruistic preferences and take on engagements and commitments. This might induce one to interpret Schmidtz as following the abovementioned strategy of reducing altruism to egoism. However, he goes on to stress that motivations like concern and respect for others – and one might add sympathy and commitment – can in the end induce people to go directly against their self-interest: “out of self-regard, we nurture the enrichment of our utility functions so that they come to incorporate other-regarding preferences. As these preferences become part of the function, they acquire a certain autonomy, become more than mere means to previously given ends” (Schmidtz 1993: 58). While other-regarding motivations may arise from self-regard, they ultimately outgrow the narrow Homo Economicus model: “the concern and respect for others that reflective self-regard leads most of us to nurture may be of an entirely wholehearted and uncalculating kind” (Schmidtz 1993: 67). In the next chapter, I will try to show why this line of reasoning leads me to complement the economic conception of rationality with alternative ones.

#### 1.3.6.4. Process-regarding preferences

People are often not so much concerned about the outcome of social interactions, as they are about the process of this outcome itself. To explain what this entails, I want to focus on what has become known as the Ultimatum game. Here, a subject is given an amount of money that he has to divide between himself (the proposer) and the other (the responder) (Thaler 1994: 22-25). If the responder accepts the offer, he gets the proposed portion and the proposer gets to keep the rest. If the responder rejects the offer, both get nothing. If both were egoistic, one would expect the responder to accept any offer larger than zero. Anticipating this, the proposer would propose the smallest offer possible. In numerous experiments conducted throughout the world, however, “*only a minority of agents behave in a self-regarding manner*” (Gintis et al. 2005: 12). Instead, “the vast majority of proposers offer between 40% and 50% of the pie and offers lower than 30% of the pie are often rejected” (Bowles & Gintis 2000: 42).

In Public Goods and Prisoner’s Dilemma games, where the threat of freeriders is present as well, the same results hold. Only a third of all people act in a purely self-interested way, which is to contribute nothing at all (Henrich et. al. 2004: 66-67). Such experimental evidence clearly suggests that a lot of people often do not act opportunistically: “many people deviate from purely self-interested behavior in a reciprocal manner. Reciprocity means that in response to friendly actions, people are frequently much nicer and much more cooperative than predicted by the self-interest model; conversely, in response to hostile actions they are frequently much more nasty and even brutal” (Fehr & Gächter 2000: 159).

When internalized, norms like reciprocity and fairness induce preferences that are aimed at the process of the interaction itself and that can outweigh a strictly egoistic calculus. Even though this can be individually costly, a majority of individuals are prepared to follow such norms and punish those who violate such norms. Fairness and reciprocity have therefore been interpreted as the main sources for the observed “large and consistent deviations from the predictions of the textbook representation of *Homo economicus*” (Henrich et al. 2004: 8).

### 1.3.6.5. Reducing process-regarding preferences to egoism

As with other-regarding preferences, the main strategy of economists is to reduce norm-guided reciprocal and fair behavior to underlying egoistic calculations: “neoclassical economists (...) try to obviate the need to modify their basic paradigms by arguing that when people abide by norms for what seems like intrinsic predispositions they actually have extrinsic motivations, such as aiming to please their friends or acquiring prestige” (Etzioni 2000: 164). This reductionist strategy fails to capture the essence because of the same problems. First, acting in a reciprocal or fair way does not always give one satisfaction. Second, even if this is the case, it does not necessarily form one’s primary motivation. Such behavior can thus not be reduced to egoistic motivations: “a person who is motivated by the warm glow that comes from having done one’s duty is not acting out of duty but engaging in narcissistic role playing” (Elster 1989a: 53).

In this respect, I want to refer to the French sociologist Marcel Mauss, who argues that gifts always create an obligation to reciprocate and are thus never truly free. This may lead one to conclude that all forms of altruism (giving) and reciprocity (giving back) can be reduced to pure self-interest. The problem with Mauss’ view is that it ignores the personal experience of the parties at hand. After all, from an internal point of view, the spontaneous and non-committing nature of gifts is crucial. As Alain Testart rightly argues, genuine gifts indeed have no strings attached at all. In fact, a donation stops being a gift as soon as a corresponding return is obliged: “to give, is to hand over something to somebody *free of charge*. To give, is not to seek payment, it is even more or less the opposite” (Testart 1998: 97). Later on, I will argue more fully why I defend Testart’s internal perspective against Mauss’ external perspective.

### 1.3.7. *Egocentrism*

The seventh and final requirement is actually the least demanding interpretation of the assumption of egoism. It stipulates that the individual is the final judge in evaluating and choosing an alternative from the available alternatives. This variant can be termed “egocentrism” or “egotism” (Dowding & King 1998: 12-14). It has led to a less stringent Homo Economicus model, in which preferences should only be consistent, transitive, complete and continuous (Sen 1977: 323). As a matter of fact, one can distinguish between classical and neoclassical economists who respectively defend the assumptions

of egoism and egocentrism. In this respect, I am happy to admit that most economists no longer explicitly insert substantial requirements into their models (Vromen 1998: 32). This has allowed them to include other than purely materialistic considerations in preference functions, like the desire for recognition, respect and love. Whether the fulfillment of one's preferences provides pleasure or benefits is irrelevant in this view, as long as the individual himself genuinely holds these (Moser 1990: 1, 191).

The assumption of egocentrism simply stipulates that people try to maximize the satisfaction of their preferences, without stipulating what these preferences are about (Vromen 1998: 24-27). In this respect, one can speak of a purely formal, subjective and instrumental conception of rationality: "to be rational in the instrumental sense is to be committed to serving preferences of oneself, but one may or may not be committed to serving preferences regarding oneself" (Schmidtz 1993: 54). In this broader view, there is nothing irrational about preferring to serve the interests of others over one's own interests. In essence, this assumption of egocentrism does not go much further than the requirement of intentionality and comes close to John Broome's view, according to which a rational individual simply maximizes what he deems good (Broome 1995: 136-139). To be sure, one's conception of the good does not always correspond with one's interests<sup>14</sup>.

In his path-breaking essay on economic methodology, Lionel Robbins already moved away substantially from the narrow, classical interpretation of the Homo Economicus model towards the broader, neoclassical interpretation. Whereas John Stuart Mill still referred to egoistic motives, he no longer did so: "economic subjects can be pure egoists, pure altruists, pure ascetics, pure sensualists or – what is much more likely – mixed bundles of all these impulses" (Robbins 1935: 95). Some authors argue that all the abovementioned evidence showing the pervasiveness of other- and process-regarding preferences has facilitated this shift from classical economics (and its assumption of egoism) to neoclassical economics (and its assumption of egocentrism) (Henrich et al. 2004: 9).

Here, I want to argue that other- and process-regarding preferences are perfectly compatible with the assumption of egocentrism, as long as the individual himself is genuinely motivated by such preferences. Sen's notion of sympathy and Schmidtz's notion of concern are easily interpreted within the

---

<sup>14</sup> Sometimes, the same thought is formulated in different terms. John Rawls (1993: 51), for example, distinguishes between interests and benefits: "every interest is an interest of a self (agent), but not every interest is in benefit of the self that has it". While framed in different terms, this perfectly corresponds to my distinction between preferences (which Rawls terms 'interests') and interests (which Rawls terms 'benefits').

broad model of individuals as aiming to satisfy their preferences. After all, the welfare of others systematically enters the picture through their own preferences (Schmidtz 199: 52). The same holds for less benign other-regarding preferences like jealousy and envy (Elster 1982: 123). It can be argued that jealous and envious individuals evaluate the situation from another's point of view rather than their own and that they are moved by something alien to themselves (Van Liedekerke 2000: 107). In some interpersonal relations, the bond can be so powerful that it seems inadequate to claim that the individual at hand is sovereign in making his own decisions (Ben-Ner & Putterman 1998: 172). Pettit (2002: 230) seems to suggest something similar when referring to "concerns that dramatically transcend the boundaries of the self". Nevertheless, I want to understand all of these phenomena as motivations that genuinely belong to the individual himself. It remains possible to say that someone acting out of sympathy, concern, jealousy or envy still does what he himself prefers.

In my view, the same holds for the phenomena of commitment and respect. In Sen's view, a committed person is not primarily motivated by his own goals but by those of others with whom he identifies. This may suggest that commitment violates the assumption of egocentrism. However, Hans Bernhard Schmid (2005: 56) rightly shows that the process of identification leads one to endorse these goals oneself: "one cannot pursue the other's goals without making these goals one's own". This way, commitment can be understood within an egocentric framework, namely as providing a reason to act against one's own self-interest but not against one's own goals and intentions. The thought that committed action can be based on good reasons while being not-egoistic and not-instrumental in nature is defended by Sen (2005: 5) himself in a recent article: "by living in a society, one develops possible reasons for considering other people's goals as well, which takes one beyond an exclusive concentration on one's own goals, not to mention the single-minded pursuit of one's own self interest. The recognition of other people's goals may be a part of rational thought".

In a similar vein, it has been argued that norms cannot be studied within a rational choice approach (Coleman 1990: 292). However, I do not see why this should be the case. To understand the impact of norms on behavior, one has to refer to processes of socialization and internalization, through which the individual comes to identify with these norms (Elster 1989a: 119). This way, norm-guided behavior clearly fits the egocentric assumption. After all, the individual remains the final judge in choosing whether or not to respect some norm. This also suggests that norm-guided behavior, in my view, also fits the assumption of methodological individualism. Since norms do not exist as supra-individual entities (ontological individualism), they cannot have an



impact on social life, independently of their impact on individuals (methodological individualism). In this respect, I agree with Elster's definition of norms as "emotional and behavioral propensities of individuals, not supra-individual entities that constrain behavior" (Elster 1991: 113-114).

A number of authors suggest that the assumption of egocentrism can be violated, since individuals can act against their preferences. Jane Mansbridge, for example, defines duty as "the capacity to act against one's preferences on the grounds that one ought so to act" (Ben-Ner & Putterman 1998: 155). The underlying thought is that preferences are but one of multiple criteria to evaluate choice options. As Timur Kuran argues, these are not necessarily in congruence with each other: "in addition to the preferences by which we choose among our options, I am suggesting, we have values that evaluate what we want and actually do" (Ben-Ner & Putterman 1998: 232). However, I do not think that such a notion of counterpreferential choice is instructive<sup>15</sup>. Instead, I want to define preferences in such a way that they encompass considerations that are other-regarding or process-regarding. In this sense, I thus want to conclude that the assumption of egocentrism is not violated in any of the abovementioned cases.

## 2. The role of economic rationality within the social sciences

One can ask on which grounds I am criticizing the economic conception of rationality. Until now, I have referred in a somewhat loose manner to an intuitive and common sense notion of rationality. Essentially, what I have tried to argue for is that the Homo Economicus model fails as a general account of individual motivation and behavior because of the unrealistic character of its central assumptions. Its requirements of maximization, instrumentality, perfect rationality, complete, continuous, egoistic and exogenously given preferences are so demanding that no individual can actually live up to them. Within the philosophy of economics, however, it has been argued that this is no reason to discard the model. After all, scientific models always and inevitably abstract from particular complexities of the concrete world. What matters is not to what extent a model's assumptions are realistic, but to what extent they are instrumental in predicting events (Caldwell 1982: 173-178).

---

<sup>15</sup> Schmitz (1993: 59) also expresses his doubts whether this notion is useful in analyzing commitment and respect (which is a specific form of commitment, namely commitment to principles that specify what one may and may not do to others).

Milton Friedman (1953: 8-9) is widely known for defending this kind of instrumentalism, according to which each “theory is to be judged by its predictive power for the class of phenomena which it is intended to “explain.” (...). The only relevant test of the validity of a hypothesis is comparison of its predictions with experience”. Because experiments are difficult to perform in the social sciences, it may be tempting to test a model’s validity by focusing on its assumptions rather than its implications. Assumptions should always be valued with respect to the predictions they engender: “the relevant question to ask about the “assumptions” of a theory is not whether they are descriptively “realistic,” for they never are, but whether they are sufficiently good approximations for the purpose in hand. And this question can be answered only by seeing whether the theory works, which means whether it yields sufficiently accurate predictions” (Friedman 1953: 15). In this spirit, a lot of economists deem it irrelevant whether or not their image of individuals corresponds to reality (Sen 1977: 319-324).

To better understand this kind of instrumentalism, it may be instructive to consider Friedman’s example of the leaves of a tree. He stresses the validity of “the hypothesis that the leaves are positioned as if each leaf deliberately sought to maximize the amount of sunlight it receives” (Friedman 1953: 19). While clearly unrealistic, this hypothesis allows one to explain a wide range of characteristics of the leaves. Despite its unrealistic assumptions, its implications actually conform to actual phenomena. A model is a theoretical idealization that is inevitably simpler than the real world. The very meaning of assumptions lies in focusing on specific aspects of the complex world while abstracting from other aspects. Such a hypothesis is best understood as “stating that the phenomena it is desired to predict behave in the world of observation *as if* they occurred in a hypothetical and highly simplified world containing only the forces that the hypothesis asserts to be important” (Friedman 1953: 40).

The criticism of the economic conception of rationality as providing an all too narrow and thus inadequate view of man can thus be answered by arguing that economists – like other scientists – choose to abstract from certain complexities of the actual world (Robbins 1935: 94). To be sure, the Homo Economicus model does not represent real human beings in all of their particularities. Instead, it is a theoretical construction designed for explanatory and predictive purposes. If one would take into account all the complexities that make up actual individuals, these explanations and predictions would lose their desired exactness. One should not try to provide a fuller and more realistic view of individuals, since scientific insights only arise thanks to some level of abstraction.

There are two ways of criticizing this instrumentalist defense of the Homo Economicus model<sup>16</sup>. First, one can argue that economics has a poor record in predicting individual behavior. Sen (1987: 79), for example, stresses that framing all behavior in terms of self-interest “is hard to justify on grounds of predictive usefulness, and it also seems to have rather dubious empirical support”. As I have tried to show, the economic conception of rationality grossly fails to predict how actual people behave in a number of instances. Second, one can argue that the realism of the assumptions underlying the Homo Economicus model does matter. The main goal of science is not to predict but to explain phenomena. That is why each scientific model aims to represent (something in) the real world. While it is hard to decide in a definitive way whether the model is true or false, it is in fact either true or false. If one wants to be an instrumentalist like Friedman, one has to accept that economists are only trying to predict phenomena and cannot provide any insight into the nature of individual actions and their social consequences (Caldwell 1982: 183-186). In this respect, I believe that the aspects of human behavior that cannot be understood as economically rational are indispensable to understand what it is to be human. In my view, scientific models that make abstraction of the characteristic core of their objects are thoroughly deficient. The view that some degree of realism is an important criterion to evaluate them does not imply that they should completely reflect the actual – and indeed highly complex – world (Caldwell 1982: 39).

This methodological discussion has a philosophical background that needs to be made explicit. In this respect, I want to refer to Bernard Williams’ well-known distinction between “internal and external reasons” (Moser 1990: 387). Williams aims to analyze what it means when one says that an individual has a reason to perform an action. According to the internal interpretation, the individual at hand actually has a motive that drives him to perform this action. According to the external interpretation, this is not necessarily the case. The latter corresponds to Friedman’s instrumentalist view, in which the individual may – but does not necessarily – have the reasons that would drive him to act in the same way *as* he would do *if* he had these reasons.

According to Williams, the internal interpretation is the only valid one when one tries to explain individual behavior as rational. If something is a reason for an action, then it must figure in an explanation of this action and, consequently, it must motivate the individual at hand to perform this action. The external interpretation neglects the question which motives the individual

---

<sup>16</sup> In my fifth chapter on James Buchanan’s theory of constitutional choice, I criticize Buchanan’s way of defending a specific form of this instrumentalist methodology. For now, I hope two general remarks suffice to point out the main problems.

actually has. As such, it can never provide an adequate explanation of his action. In Williams' own words, "if something can be a reason for action, then it could be someone's reason for acting on a particular occasion, and it would then figure in an explanation of that action. Now no external reason statement could by itself offer an explanation of anyone's action (...). The whole point of external reason statements is that they can be true independently of the agent's motivations. But nothing can explain an agent's (intentional) actions except something that motivates him so to act" (Moser 1990: 392). Of course, an external reason can become internal when an individual becomes convinced that it is a good reason to act on. However, as long as the individual himself does not endorse it, it cannot explain his actions.

This implies that Friedman's methodology is inappropriate within the social sciences. If one is trying to explain an individual's behavior, one should refer to his actual set of motives. If an individual is driven by an internalized norm, one best explains his behavior by referring to this fact, just like one best explains the position of the tree's leaves by referring to what actually goes on inside (namely a process of photosynthesis through which plants extract energy from sunlight to produce adenosine triphosphate)<sup>17</sup>.

To explain an individual's actions, one should not just provide a story of which the outcome corresponds to these actions, but one should show what causes him to act in such a way. In order to understand him as rational, one must demonstrate that he performs his actions because of the reasons he has and not because some accidental cause happens to produce this behavior (Elster 1986: 13). While an individual can do what is best for him without even knowing it, this does not count as rational, since he does not perform this action for this reason. From the fact that there might exist some reason for a particular action (external interpretation) one may not conclude that the individual actually has performed this action *for* that reason (internal interpretation). Doing the (objectively) 'rational' kind of thing does not entail that the individual is (subjectively) rational when doing it. As Robert Audi rightly stresses, "a particular action should be considered rational in virtue of a set of beliefs and wants expressing reasons for it, only if these wants and beliefs

---

<sup>17</sup> To rephrase this thought, I focus on the internal or first-person perspective, which means that I aim to explain behavior by referring to the considerations, motives and reasons that motivate the individual at hand. In contrast, evolutionary theorists for example, underline the external or third-person perspective, arguing that behavior always unconsciously serves the individual's fitness in the struggle for life. Similarly, game theorists tend to argue that behavior always benefits the individual that performs it, regardless whether or not this is deliberately aimed at. I acknowledge that actions might, in fact, be *in* the self-interest of an individual (external), even though they are not performed *out* of self-interest (internal).

play a role in generating or sustaining it” (Moser 1990: 427). In his view, external reasons statements are thoroughly misleading because they cut “rational actions off from the system of propositional attitudes from which, by virtue of an explanatory connection, they derive their rationality” (Moser 1990: 445). Since rationality consists in the inherent link between an individual’s actions and his reasons for performing them, the external interpretation of reasons is simply incoherent (Moser 1990: 395).

This also relates to the discussion on the distinction between reasons and causes. Whereas unintentional behavior (falling to the ground) is explained by referring to what causes it (a stone over which one trips), intentional behavior is explained adequately only if one sees what its purpose is. This distinction underlies the dichotomy between (natural) sciences that aim to explain the causal history of events (*Erklären*) and (social) sciences that aim to understand the reasons behind particular events or behavior (*Verstehen*) (Pettit 2002: 159). In my view, however, this dichotomy is misleading in that there is a causal aspect inherent in the latter sort of explanation as well. As Davidson (1982: 293) rightly argues, “the reasons an agent has for acting must, if they are to explain the action, be the reasons on which he acted; the reasons must have played a causal role in the occurrence of the action (...). Since beliefs and desires are causes of the actions for which they are reasons, reason explanations include an essential causal element”. If such accounts are to explain actions, it is not sufficient to demonstrate that the individual has a reason for them and performs them, one should also show that it was this reason why he did it: “central to the relation between a reason and an action it explains is the idea that the agent performed the action because he had the reason” (Davidson 1980: 9).

In the end, one may wonder where the value of economics lies as it is currently performed in academic circles around the world<sup>18</sup>. To be sure, I have no quarrel with economists who accept that their model provides only a partial explanation. The problem, however, arises as soon as they treat their model as the only valid one and start to argue that all people in all circumstances act as their model predicts. This is what happened in the transition from classical to neoclassical economics. Economists like Edgeworth (1881: 17) applied their models only to specific domains, like the conduct of war and the signing of contracts, where an egoistic cost-benefit calculus is a plausible motive. Since then, however, economists increasingly applied their models to actions outside these traditional domains. This so-called “economic imperialism” (Tullock

---

<sup>18</sup> While I happily acknowledge and applaud the fact that (some) insights seem to gain more attention amongst (some) economists, I want to focus in this dissertation on possible ways of criticizing and complementing economics as it is currently practiced.

1972) opens up possibilities for criticizing the adequacy of the employed assumptions. While it may be so that all people are economically rational some of the time and some people may even be economically rational all of the time, it is certainly not so that all people are economically rational all of the time. While consumption decisions are indeed likely to be egoistically and instrumentally motivated, other decisions are more likely to be based on other considerations. In the next part of this dissertation, for example, I will show that alternative conceptions of rationality are especially relevant when analyzing the rationality of decisions in an electoral context.

Instead of limiting the economist's object of study, Pettit aims to reconcile the economic model of human motivation (the economic mind in which egoistic considerations are relatively stronger than other ones) with the more everyday, folk psychological view (the common mind in which other considerations can be stronger). As I have shown extensively, this is not evident given the fact that "the economic explanation of individual behavior (...) flies in the face of common sense; it conflicts with our ordinary assumptions about how we each feel and think in most situations" (Pettit 2002: 222). This clearly suggests that either common sense or economics is wrong. Nevertheless, Pettit argues that it is possible to reconcile both perspectives.

In his view, the egoistic *Homo Economicus* is virtually present within people in that it refers to the limits to what they are prepared to do (Pettit 2002: 168). While they may generally be moved by non-egoistic dispositions under a more or less automatic, cultural pilot, their egoistic considerations can come into play as soon as the costs of some practice become greater: "at any point where a decision is liable to cost them dearly in self-regarding terms, the alarm bells ring and prompt them to consider personal advantage" (Pettit 2002: 236). Individuals can then decide whether to listen to or ignore the alarm bells of the *Homo Economicus*. Pettit's reconciliation allows him to apply his model outside the traditional area of market behavior: "the assumptions that economists make about the human mind, in particular about human motivation, can be rendered consistent with the assumptions of commonplace, everyday thinking. And it shows that, so interpreted, the assumptions motivate a promising and indeed developing programme for economic explanation: and explanation, not just in the traditional areas of market behaviour, but across the social world more generally" (Pettit 2002: 241).

If one wants to criticize the unrealistic character of the central assumptions of the economic conception of rationality, one should provide an alternative (Friedman 1953: 31). While I do not aim to present a unified body of alternative models, I hope to show in the next chapter that a more pluralistic approach in the social sciences is desirable. However, while the imperialistic

tendency of economics is to be opposed, I see no help coming from a wide range of disciplines that perform their own research in a fragmented way. That is why I continue to focus on the notion of rationality and show the plausibility and desirability of a number of alternative conceptions of what this notion exactly means.





## CHAPTER 2

### ALTERNATIVE CONCEPTIONS OF RATIONALITY

---

*“We believe that there is a real and extensive domain where economic man is a useful, analytically powerful abstraction (...). What we cannot accept, however, is that there is nothing ‘beyond economic man’ (...). What we see in the ‘beyond’ is the essential and distinctive human”*

(Lutz, in: Koslowski 1985: 116)

#### 1. Minimal rationality

All of the abovementioned failures of the Homo Economicus model to explain actual behavior lead me to conclude that it simply cannot serve as a general and exclusive conception of rationality. Instead of dismissing it in its totality, however, I want to arrive at alternative conceptions by amending some of its requirements. In what follows, I want to loosen the requirements of the economic conception of rationality that are too demanding. This way, I want to arrive at a less restrictive and thus more general conception that stipulates the minimal conditions to speak of rationality. In what follows, I will therefore analyze to what extent the phenomena that are economically irrational are also irrational in a more intuitive sense of the word. Whenever this is not the case, I will eliminate the requirement from this minimal conception of rationality. Here too, I will distinguish between the level of the relation between actions and their reasons and the level of the reasons themselves.

##### 1.1. The level of actions and the relation with their reasons

###### 1.1.1. Maximization

First, I want to argue that the requirement of maximization is too strict to be retained in a minimal conception of rationality. The definition of rationality as the maximization of preference satisfaction does not fit the intuitive and everyday use of the word (Van Liedekerke 2002: 111). Michael Slote (1989: 1)

bases his criticism of this assumption on what he labels an “ordinary intuitive understanding of practical rationality”. He argues that maximization is not part of it: “our common-sense intuitions about rationality treat some actions as rationally acceptable even though their consequences are less than the circumstantially best possible for their agents” (Slote 1989: 30). Some form of satisficing or moderation is not irrational, especially given the limited human capacities to process large amounts of complex information (Slote 1989: 11-12). Slote (1989: 65) argues that individuals who systematically try to maximize the satisfaction of their preferences would be labeled irrational by most standards: “something seems to be inherently irrational about trying to do better when things are already satisfactory in every way”. Consequently, he concludes that the maximization requirement cannot function as a necessary or sufficient condition to speak of rationality (Slote 1989: 47).

Jon Elster mentions three cases where a strategy of satisficing is inevitable, namely complete indifference (in which case the individual deems no specific option optimal), complete uncertainty (in which case the individual does not know which option to deem optimal) and some strategic settings (in which case there is no single solution that is deemed optimal by all parties involved): “these provide the *special argument for satisficing* (...). When the course of action ‘than which none better’ is not defined, one will have to go for something that is *good enough* or satisfactory, rather than optimal” (Elster 1983: 14). In addition, Elster argues that there is an aspect of satisficing inherent in each choice. The fact that one does not know beforehand how much information to collect before taking a decision provides the *general argument for satisficing*. While it is clearly not rational to base one’s actions on no information whatsoever, the same is true for the strategy of endlessly collecting information. At a certain point, one simply has to make a decision: “a rational man will make a choice, rather than procrastinate indefinitely” (Elster 1986: 19). For example, “Buridan’s ass, which died of hunger being unable to decide which of the two haystacks in front of it happened to be superior, could have rationally chosen either of the haystacks, since it had good reason for choosing either rather than starving to death” (Sen 1987: 68). Because rationality clearly does not necessarily presuppose maximization, I propose to eliminate the latter from the minimal conception of rationality.

In my view, it is crucial to stress that the individual’s goals are defined by himself so that it is perfectly possible for him to be happy with a suboptimal level of preference satisfaction<sup>19</sup>. Of course, the issue then arises which level a

---

<sup>19</sup> This also shows that the minimal conception of rationality is still subjective in nature. It is the individual himself who chooses the option that he deems desirable in the light of his own beliefs and preferences.

rational individual will satisfactory. In this respect, satisficing theories tend to be become rather ‘ad hoc’. Whatever behavior is observed, they can refer to some aspiration level that corresponds to it. Until more psychological research is done in this respect, such theories only describe the behavior at hand instead of actually explaining it. Even though he wants to move away from an all too demanding conception of rationality, Elster (1986: 26) argues that the latter still has the advantage of being more general and more simple than its alternatives: “neoclassical economics will be dethroned if and when satisficing theory and psychology join forces to produce a simple and robust explanation of aspirations levels (...). Until this happens, the continued dominance of neoclassical theory is ensured by the fact that one cannot beat something with nothing”. In this respect, I believe Simon’s satisficing theory is best interpreted as complementing the conventional theory when the latter fails rather than replacing it altogether (Elster 1989b: 29).

### *1.1.2. Instrumentality*

Second, I want to move away from the instrumental character of economic rationality. In my view, minimal rationality of actions requires only that they are based on what the individual at hand believes are good reasons. Here, I want to argue that the notion of a reason is broader than that of a goal. The economic focus on actions that are instrumentally aimed at the achievement of some goal leads to an all too narrow conception of rationality: “it is a thin view of rationality that reduces all rational action to instrumental action” (Miller 1990: 268). Robert Nozick (1993: 133) rightfully argues that instrumental conception of rationality is a powerful one, but questions whether it is the only one: “instrumental rationality is the base state. The question is whether it is the *whole* of rationality”.

In this respect, I take on Donald Davidson’s broad definition of a reason as “some sort of pro attitude toward actions of a certain kind” (Davidson 1980: 4). He mentions different sorts of reasons, like “desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values in so far as these can be interpreted as attitudes of an agent directed toward actions of a certain kind” (Davidson 1980: 4). According to Thomas Scanlon (1998: 23), rational actions are based on considerations that the individual himself judges to be reasons worth acting on. In his view as well, “to take there to be a reason for something is just to see some consideration as counting in favor of it” (Scanlon 1998: 50).

For an action to be rational in the minimal sense, it should not always be a means towards the realization of some goal. While this may count as a reason, the latter encompasses much more: “a reason rationalizes an action only if it leads us to see something the agent saw, or thought he saw, in his action – some feature, consequence, or aspect of the action the agent wanted, desired, prized held dear, thought dutiful, beneficial, obligatory, or agreeable” (Davidson 1980: 3). Later on, I will go into the issue how to analyze non-instrumental reasons in a coherent way.

Raymond Boudon also argues that actions based on reasons should be labeled rational, even though they are not aimed at achieving some goal. He disconnects rationality from instrumentality as well: “rationality should also be defined for unpurposeful actions such as believing in a theory. On the whole rationality cannot be applied only to purposive actions” (Boudon 1993: 10). In his “synthetic theory of rationality” (Boudon 1993: 18), he aims to apply rationality to both purposive and non-purposive actions, since both types of action can be based on good reasons: “a *purposive* action is rational iff (if and only if) X has good reasons for doing Z. Unpurposive actions are rational iff X has good reasons for being convinced by a theory T, to feel indignation when such and such an act is committed, etc.” (Boudon 1993: 10). In my view, these remarks are best understood as expanding the notion of rationality in two ways. First, they show the necessity to define the rationality of reasons in addition to that of actions. Second, they show that a purely instrumental framework is too narrow to define rationality. After all, the rationality of one’s beliefs (in a theory, for example) does not depend on the question whether these help achieve some goal.

### 1.1.3. *Intentionality*

Third, Boudon’s use of the notion ‘purpose’ raises the question how minimal rationality relates to intentionality. While it might suggest that he is arguing against intentionality as a criterion of rationality, I contend that believing in a theory is clearly not instrumental but certainly intentional in nature. Beliefs are aimed at something – namely the truth of a theory – but not in the way means are aimed at ends. I thus believe that Boudon is wrong when he calls believing in a theory not only “non-instrumental” but also “unintentional” (Boudon 1993: 5). Whereas intentions are indeed situated at the level of the relation between actions and their reasons, intentionality can also be predicated from reasons themselves. Both beliefs and preferences are propositional attitudes that are about something and that refer to an intentional object.

It is important to see that unintentional behavior, like tripping over a stone, is arational rather than irrational. To explain such behavior, one needs to refer to causes other than his beliefs and preferences. Bodily movements that fall completely outside the domain of intentionality and rationality should thus be understood as unintentional and arational. In contrast, other actions are still rightly labeled intentional and minimally rational, even though they are based on certain habits and traditions. While visiting one's parents every Sunday afternoon is a habit that is performed almost automatically without much deliberation, one may still have good reasons for doing so (Moser 1990: 444). In this respect, Robert Nozick argues that actions (and beliefs for that matter) are rational if they are responsive to reasons for and against them. However, this does not imply some purely calculating attitude: "responsiveness to reasons does not require explicit consideration of them. Rationality can be modest and choose to step aside sometimes or even, in some types of circumstances, almost always" (Nozick 1993: 106). While norm-guided actions are clearly not instrumentally motivated, they are still intentional<sup>20</sup>. The minimal conception of rationality thus allows one to understand such non-instrumental behavior as rational. In my view, internalized norms can thus be conceived as part of an individual's set of reasons.

In this respect, I deem it useful to distinguish between different forms of norm-guided behavior. First, there is norm-guided behavior as it appears in the Homo Sociologicus model. Such unconscious and purely mechanical behavior cannot be called rational in any sense of the word. It is caused by all kinds of impulses without these ever constituting a reason for his actions. Second, there is non-instrumental behavior that the individual performs for good reasons<sup>21</sup>. One's reasons for following norms do not necessarily consist of goals one wants to achieve. Following the norm to tell the truth, for example, can be called rational in the everyday meaning of the word if this norm is internalized and thus forms the basis of a preference. In this respect, it would be too simplistic and downright incorrect to systematically understand each instance

---

<sup>20</sup> In some cases, these reasons even fit the economic conception of rationality. As I will explain more fully in the next chapter, such actions can still be called irrational in another sense of the word if they violate more demanding requirements of other conceptions of rationality. If they do not allow the individual to maximize the satisfaction of his egoistic preferences, for example, such behavior is rightly called economically irrational.

<sup>21</sup> One can, for example, decide to follow norms of etiquette, because one wants to avoid social sanctions connected to the violation of these norms. However, norm-guided behavior is often not motivated by such a calculus. In addition, one can question whether such economically rational behavior can still be called norm-guided. Since one's actions are in accordance with the norm but are not caused by a respect for the norm, the connection with the norm is rather arbitrary. In this case, a theory referring to norms does not form an adequate explanation of the resulting behavior.

of norm-guided behavior as irrational. Yet, this is exactly what happens if one understands the distinction between rational and norm-guided behavior as a dichotomy between conscious, intentional behavior and unconscious, mechanistic behavior.

The way in which Boudon reduces norm-guided behavior to the blind, mechanical behavior of the *Homo Sociologicus* and juxtaposes it to the rational behavior of the *Homo Economicus* is overly simplifying. Moreover, it leads to an all too simple equation of norm-guided behavior with irrationality: In this respect, I agree with Antoon Vandewilde (1993: 85), who argues that “the distinction between the *Homo oeconomicus* and the *Homo sociologicus* (...) is presented as one between conscious intentions on one hand and unconscious drives or causes on the other hand. In fact this characterization of the *Homo sociologicus* makes it a much too easy target for criticism”. In contrast, I want to keep middle ground between a dichotomous distinction between a norm-guided *Homo Sociologicus* and a rational *Homo Economicus* on the one hand and a complete rationalization of norms on the other hand. The latter strategy has often been employed by economists. However, their attempts to explain the existence and functioning of norms in a completely instrumental framework have failed, for the simple reason that norms indeed form a motivation independent from economically rational considerations. Instead of understanding norms as supra-individual entities that drive individual actions, I want to conceive of them as emotional and psychological propensities within the individual itself (Elster 1991: 113-114). This is exactly where the value of a minimal conception of rationality lies. It allows one to understand norm-guided behavior within a methodologically individualist framework, while it has traditionally been understood as lying outside, beyond and opposite to this framework.

To show that norm-guided behavior can still be intentional in nature, I want to refer to distinguish between the following of a rule – or a norm, for that matter – and conforming to it: “the notion of following a rule, as it is conceived here, involves an important element over and beyond that of conforming to a rule. The conformity must be intentional, being something that is achieved, at least in part, on the basis of belief and desire. To follow a rule is to conform to it, but the act of conforming (...) must be intentional. It must be explicable, in the appropriate way, by the agent’s beliefs and desires” (Pettit 2002: 27). Whenever I talk about norm-guided behavior, I am thus referring to people following norms.

## 1.2. The level of the reasons themselves: beliefs

I now want to focus on the level of the reasons themselves, and more specifically on the beliefs.

### 1.2.1. *Consistency*

First, an individual with inconsistent beliefs is called irrational, even in the common sense of the word: “if someone asserts ‘ $p$  and not- $p$ ’ we cannot even grasp what he has in mind” (Elster 1979: 151). Therefore, I will hold on to this requirement in the minimal conception of rationality.

### 1.2.2. *Perfect rationality*

Second, the requirements of perfect rationality are too demanding to fit such a minimal conception. As I have already suggested, it is often rational for people to make a decision on the information available to them at the time. Even though their beliefs are far from perfectly informed, they inevitably form the basis of their actions. In addition, an individual is not necessarily irrational if he makes a choice without performing detailed calculations about the probabilities of all possible consequences of all alternative options. Simon’s bounded conception of rationality shows that rationality cannot be reduced to a theory of idealized choice situations, like in the economic conception of rationality. His conception is based on an explicit recognition of the limits of human cognition (Simon 1957: 2, 158).

Robbins (1935: 93-94) already argued that these requirements at the cognitive level are too stringent to be fulfilled by actual individuals. These assumptions are to be understood as theoretical idealizations of what rationality ideally entails at the level of beliefs (Pettit 2002: 196; Vromen 1998: 28). Since then, however, most economists have neglected Robbins’ methodological warnings. In contrast, I want to take heed of his remarks by eliminating these requirements in the minimal conception of rationality. Since I retain only the requirement of consistency, minimal rationality at the level of beliefs is very formal indeed. The only difference with Robbins is that he continues to frame this in terms of the Homo Economicus model, while I distinguish the economic from the minimal conception of rationality in order to get both as conceptually clear as possible.

### 1.3. The level of the reasons themselves: preferences

The same is true at the level of preferences, where I want to hold on to the conditions of consistency, transitivity and egocentrism. In my view, individuals violating these requirements are rightly called irrational.

#### 1.3.1. *Consistency*

First, I want to distinguish between weak and strong inconsistency. Weak inconsistency refers to a situation in which an individual has inconsistent preferences at different moments in time. Since he does not violate the consistency requirement as set out above – never to prefer  $a$  over  $b$  and  $b$  over  $a$  at the same time – I propose not to call him irrational. Strong inconsistency, however, means that a single individual prefers  $a$  over  $b$  and  $b$  over  $a$  at the same time, like the addict who wishes to quit but still longs for another drink. As long as he does not make up his mind in some definitive way, he is rightly called irrational. I therefore suggest maintaining this requirement within a minimal conception of rationality. This is in line with Sen (1977: 323) who contrasts the Homo Economicus with a rational individual, of whom he only demands that his preferences are consistent.

#### 1.3.2. *Transitivity*

Second, I want to argue that a minimally rational individual has not only a consistent, but also a transitive preference function (Elster 1982: 122; Elster 1983: 10; Margolis 1984: 6; Moser 1990: 410). An intransitive preference function is rightly considered as irrational, because it leads to “the phenomenon of *improving oneself to death*” (Elster 1983: 26). If an individual prefers  $a$  over  $b$ , he will be prepared to give money to exchange  $b$  for  $a$ . If he also prefers  $c$  over  $a$  and  $b$  over  $c$ , this also holds for exchanging  $a$  for  $c$  and  $c$  for  $b$ . This way, he will end up holding the same goods than in the initial situation ( $b$ ) but with less money than before. If he does not adjust his preferences, he is rightly labeled irrational.



### *1.3.3. Completeness*

Next, I want to argue that the requirements of completeness, continuity, exogeneity and egoism are too stringent for a minimal conception of rationality by showing that it is perfectly legitimate to call incomplete, lexicographic, endogenously changing and altruistic preferences rational as well. An individual can have good reasons for refusing to make a choice between killing his daughter or his son. While this means that one's preference function is incomplete, it is clearly not irrational (Vandeveldt 1994: 92). Therefore, I want to eliminate the requirement that individuals always have to be able to choose between two alternatives.

### *1.3.4. Continuity*

The individual whose love or moral principles are not for sale has lexicographic preferences. While this is economically irrational – he does not grab the opportunity to make a lot of easy money – common sense would not label him irrational. In fact, it is exactly this ability to go beyond the immediate experience of utility that is typically human. Therefore, I want to eliminate the requirement of continuity as well. Later on, I will analyze more fully this distinctively human capacity to distance oneself from one's immediate impulses.

### *1.3.5. Exogeneity*

The fact that preferences can change endogenously in the process of making a decision does not render them irrational. In my view, accounts of how preferences arise and change are perfectly compatible with the notion of rationality. People often upgrade or downgrade some of their preferences. Moreover, this is sometimes done for the sake of rationality, as in the case with the addict who avoids situations in which alcohol is available in order to get rid of his preference for drinking. In addition, an adequate conception of rationality should take into account that, through their actions, individuals can acquire relevant information that induces them to change their preference ordering. While the assumption of completely fixed and thus invariable preferences is too strict, its complete opposite of extremely volatile preferences poses problems as well. Someone whose preferences continuously shift is likely

to be labeled irrational in a more everyday sense of the word. In this respect, I want to defend the requirement of more or less stable preferences, which can change but only with good reasons. Later on, I will try to show what exactly this entails.

### *1.3.6. Egoism*

As I have already shown extensively, other-regarding and process-regarding preferences can be rational as well. According to common sense, there is clearly nothing irrational about wanting to serve another's interests or to insist on fairness and reciprocity in social interactions. The requirement of egoism is thus too stringent and should be eliminated in a minimal conception of rationality. In this respect, a distinction can be made between economic models, which require that people act egoistically, and rational choice models, which no longer do so.

### *1.3.7. Egocentrism*

To conclude, I want to hold on to the assumption of egocentrism, because this is based on the underlying assumption of methodological individualism. According to the minimal conception of rationality, actions are to be understood as the result of an intentional choice between a limited set of possible alternatives on the basis of a subjective, consistent and transitive preference function. According to the assumption of egocentrism, they are to be understood as resulting from the individual's evaluation and judgment of the available alternatives. Since they all place the individual central, the concepts of methodological individualism, rationality and egocentrism are closely connected to each other.

## 2. Broad rationality

Both the economic and the minimal conception of rationality deal with their own problems. Whereas the economic conception is too demanding to be a plausible account of what rationality entails, the minimal conception is too formal. Someone who wants to commit a murder and believes in voodoo is minimally rational if he sticks needles into a puppet (Elster 1983: 3). This, however, goes directly against the grain of everyday intuitions, according to

which something is wrong with such beliefs (and perhaps also with such preferences). Instead of using vague intuitions as the ultimate bedrock on which to ground more demanding conceptions of rationality, I aim to base them on explicit, philosophical grounds.

By adding more requirements, I want to give more substance to the concept of rationality. The resulting conceptions limit the range of phenomena that can be understood as rational but not in the same direction as the economic conception. While this reduces the explanatory scope of the concept, it enhances its explanatory power (Moser 1990: 421). This way, it avoids the danger of turning rationality into a completely empty notion that can rationalize each action by referring to reasons that justify that action. In what follows, I will first develop a broad conception of rationality, which is primarily based on remarks from Jon Elster.

### **2.1. The level of actions and the relation with their reasons**

The basic thrust of moving from a minimal to a broad conception of rationality is that actions should no longer be based on mere reasons but on good reasons. The extra conditions imposed by the broad conception of rationality are thus situated primarily at the level of the reasons themselves. The irrationality of a person sticking needles in a puppet is located at the level of his beliefs (it is irrational to believe that one can kill someone this way) and perhaps his preferences (it is irrational to want to kill someone). When somebody acts on the basis of irrational beliefs or preferences, the resulting actions are irrational as well, but the source of this irrationality lies deeper. At the level of the relation between the actions and their reasons, I will thus hold on to the requirement of intentionality, which was the core of the minimal conception of rationality.

### **2.2. The level of the reasons themselves: beliefs**

People are not just ‘automata’ that select the action that best satisfies their preferences according to their beliefs, they are also concerned whether their preferences and beliefs are justified. To evaluate whether this is the case, one can refer to the ways these arise and change. In this respect, Elster (1983: 15-26) stipulates that beliefs should be well-informed and that preferences should be autonomous. The main source of irrationality at this level is that of irrelevant causes bypassing reasons. The resulting beliefs and preferences “are

shown to be irrational if they are explained by mental causes that are not reasons for them” (Davidson 1982: 299). The broad conception of rationality thus leaves the assumption of exogenously given reasons, which dominates the economic conception of rationality. Instead, it analyzes whether the formation process of reasons satisfies certain standards.

Boudon (1993: 5) argues that it is crucial to construct a conception of rationality that can be applied to beliefs, since all actions are ultimately based on beliefs about the way the world works. As I have suggested, one needs specific requirements to speak of rational beliefs, because believing in a theory cannot be explained instrumentally. Even though beliefs are not instrumentally formed, they can be based on good reasons. The broad conception of rationality thus refers to the central notion of reasons as well. It stresses that beliefs are rational if they are supported by reasons that make them credible and arise from processes that reliably produce true beliefs (Nozick 1993: 64, 176).

To develop all this more fully, I want to refer to the work of Jon Elster. He argues that the information and cognitive processes available to the individual are crucial to speak of rational beliefs. Instead of requiring these to be perfect – as the economic conception does – Elster (1983: 16; 1986: 13-14) argues that beliefs should be properly grounded in and derived from the available information. A belief is rational in the broad sense when it is well-informed and based on a sound judgment of the available information. Beliefs that go against the available evidence are clearly irrational. To get a better grip on what this exactly means, it is useful to proceed ‘*ex negativo*’. To summarize, beliefs are broadly irrational when they “are generated by *belief-irrelevant* causal processes, i.e. causal processes where the causes include not only *reasons*” (Elster 1979: 130). One can distinguish between two categories of broadly irrational beliefs.

A first category consists of beliefs arising from cognitive defects. Evidence from experimental economics convincingly shows that people rely on heuristics to assess probabilities. While these are quite useful in general, they sometimes lead to cognitive errors (Moser 1990: 171). Take the case where people are asked to estimate the odds that a Third World War will break out in the next year. If the options are ranked from zero to hundred percent, the average answer will be lower compared to when one ranks them from hundred to zero percent (Elster 1989a: 34-35). Here, one can speak of an unsound judgment and thus of a broadly irrational belief. Another factor that can cause a cognitive bias in belief formation processes is the tendency of people to attach more importance to concrete than to abstract information. Unemployed people, for example, tend to overestimate the unemployment rate, while employed people tend to underestimate it (Nisbett & Ross 1980: 19). Vivid

information, which is emotionally laden or close in time and space, is likely to have a greater weight on the formation of beliefs (Nisbett & Ross 1980: 43-51). Here too, there is a lack of sound judgment, which can be defined as the capacity to process the available information without attaching excessive importance to a particular part of it (Elster 1983: 16). These examples show that one can be rational in the minimal sense – as long as beliefs are consistent with other beliefs – but irrational in the broad sense.

A second category consists of beliefs that are improperly influenced by volitional urges. Here, the best example is that of wishful thinking, which means that individuals believe things to be as they want them to be (Elster 1983: 148; Elster 1989a: 37-38). Since the available information is distorted by an irrelevant motivational bias, such beliefs are rightly called irrational. This defect in the belief formation process forms the main culprit of an individual's irrationality.

### **2.3. The level of the reasons themselves: preferences**

A broad conception of rationality also adds requirements at the level of preferences. Next to the formal conditions of consistency, transitivity and egocentrism, it stresses that preferences should be autonomously formed. Here too, Elster (1983: 22) proceeds to define autonomy 'ex negativo' by referring to those preferences that are surely heteronomous and thus irrational in the broad sense of the word. Heteronomous preferences are formed by processes over which the person has no control: "in such cases the individual is in the grip of causal forces operating 'behind his back' and governing his preferences in a way that he does not himself understand" (Elster 1979: 148). This is clearly irrational: "one cannot be rational if one is the plaything of psychic processes that, unbeknownst to oneself, shape one's desires and values" (Elster 1989b: 6).

I want to distinguish between two kinds of distorting influences. First, there are extra-mental circumstances that can have an impact on an individual's preferences beyond his control. Examples are adaptive preferences (I prefer something simply because it is available) and counteradaptive preferences (I prefer something because it is not available). The exemplary case of the former is La Fontaine's story of the fox that no longer prefers the grapes – which he believes to be sour – simply because they lie outside his reach (Elster 1983: 109-140). The logic of the latter is captured in the saying that "the grass is always greener at the other side of the fence" (Elster 1989a: 8-9). Other examples are conformism (I prefer something simply because others do so as well), anti-conformism (I prefer something simply because others do not),

novelty (I prefer something simply because it is new to me) and inertia (I prefer something simply because it was already there for me) (Elster 1989b: 6).

Second, there are mental forces that have an impact on an individual's preferences beyond his control. Here too, one can distinguish between a cognitive and a motivational bias. First, empirical findings from experimental economists show that changes in the way equivalent options are framed cause preferences to shift (Tversky & Kahneman 1981: 453). Second, preferences can be distorted by underlying volitional forces or urges. For example, someone who wants to quit smoking can be overwhelmed by a visceral craving for nicotine. The irrationality of his longing for a smoke lies in the fact that his preference ordering is influenced by unconscious and irresistible drifts and thus changes without any reason whatsoever.

### 3. Expressive rationality

I want to argue that Elster's broad conception of rationality is problematic, since actions, beliefs and preferences that arise heteronomously are not necessarily irrational. Even though most actions, beliefs and preferences are formed under the influence of the individual's milieu and context, this does not make them irrational. In this respect, I want to make use of the insights of Harry Frankfurt, who argues that an individual can be autonomous, even if he does not act on the basis of consciously, intentionally or deliberately formed reasons. As long as the individual identifies with his reasons and thus regards them as genuinely his own, he can be considered autonomous and thus rational.

In one passage, Elster (1986: 15) seems to agree with Frankfurt's view: "acting rationally (...) implies that the beliefs and desires behind the action have a causal history with which we can identify ourselves". While I will go into the differences between Elster and Frankfurt in greater detail later, I want to remark here that Elster claims that rational individuals should identify with the causal processes that lead to the formation of reasons, whereas Frankfurt stresses that they should identify with the reasons themselves, regardless how they are formed. Instead of incorporating Frankfurt's insights into Elster's broad conception of rationality, I will treat them as constituting an alternative and divergent conception of rationality. In contrast with Elster, who does not stipulate any further requirements at the level of the relation between actions and their reasons, Frankfurt's views allow to do exactly this.

### 3.1. The level of actions and the relation with their reasons

Actions are expressively rational when they express the identity of the individual at hand (Hargreaves Heap 1989: 148-152). They can be understood as the authentic articulation of the way the individual perceives himself.

#### 3.1.1. *Non-instrumentality*

To explain this more fully and stress the contrast with economically rational actions, I want to stress that expressive actions are not means to achieve some goal, but actions of which the performance is a goal in itself. Take Martin Luther's famous quote: "hier stehe ich, ich kann nicht anders"<sup>22</sup>. In response to the question of why he does what he does, he would probably have referred to what he considered the most fundamental aspects of his identity. The fact that he is who he is gives him a good reason for doing what he does. It is the symbolic meaning of the action itself that counts, not its consequences. While people are often not instrumentally motivated, this does not imply that they are necessarily irrational: "one way we are not simply instrumentally rational is in caring about symbolic meanings, apart from what they cause or produce" (Nozick 1993: 139). In this sense, I believe that actions can rightly be called rational in an expressive sense, even though they are clearly economically irrational.

The expressive character of such actions cannot be understood within the instrumental framework of economic rationality: "what is missing in the means-to-given-ends vocabulary of instrumental rationality is the open-ended nature of action. It is a process sense of rationality or a 'doing' rather than an 'achieving' model of rationality which we need. One final way of describing the open-ended nature of some actions, which brings out why it cannot be reduced to instrumental rationality, is to say that expressively rational acts are ends in themselves. Acts which are ends in themselves are ones which are valued for the 'doing' rather than the 'achieving'" (Hargreaves Heap 1989: 173-174). Intrinsically valued actions have no goal outside the performance itself, they do not fit within an instrumental framework. According to Michael Taylor (1988: 85-86), the instrumental character of the economic conception of rationality has two consequences: "first, that the pleasure or benefits of any kind which are got in the process of *doing* the action, as opposed to the value of the

---

<sup>22</sup> Luther in a speech in Worms on April 18, 1521.

*consequences* of the actions, must be unimportant; and second, that expressive motivations – the desire to be “true to one’s self”, to act consistently with one’s deeply held commitments, and so on – play no important role”.

The distinction between instrumental and expressive considerations thus largely corresponds to what psychologists have labeled extrinsic and intrinsic motivation. If an action is extrinsically motivated, it is performed for the sake of some kind of reward that is external to the action itself. Take an individual who performs his job only because of the paycheck he receives at the end of the month. His main motive for performing his action is the fact that it achieves some goal. In contrast, if an action is intrinsically motivated, it is performed for its own sake: “one is said to be intrinsically motivated to perform an activity when he receives no apparent rewards except the activity itself” (Deci 1971: 105). Most people who choose to read a novel, for example, do not aim to achieve some kind of goal or reward that is external to the reading of the book in itself.

Quite a few consumer decisions, which form the preeminent domain of economists, are best explained in expressive terms (Hargreaves Heap 1989: 160-163). The way people choose their clothes, for example, is not guided in the first place by an instrumental reasoning how best to protect themselves against wind and rain. It should rather be understood as a non-verbal form of communication and thus as an expression of who they are. While the fact that they wear clothes is best understood in instrumental terms, the fact that they wear these clothes rather than others is best understood in expressive terms. Both conceptions can give more or less valid but essentially different descriptions of the same action. In addition, it is perfectly possible for a single individual to act instrumentally at one moment in time, but expressively at another moment. This also leaves open the possibility of some individual acting instrumentally and others acting expressively in a particular situation. All this shows that the expressive conception of rationality complements the instrumental one without replacing or undermining it (Hargreaves Heap 1989: 172).

### *3.1.2. Intentionality*

As these examples suggest, expressive actions are still intentional in nature. While the requirements of maximization and instrumentality no longer hold in the expressive conception of rationality, the characterization of rationality in subjective and intentional terms thus remains in place. This allows me to come back to the relation between rationality and norms. In its economic sense,



rationality is inapt to explain norm-guided behavior. Since non-instrumental considerations can form good reasons as well, the expressive conception of rationality allows one to further specify the thought that non-instrumental behavior can also be rational. After all, norms often help constitute a person's identity: "social norms help people form (and re-form) the self, by profoundly influencing their identities, their worldviews, their views of themselves, the projects they undertake, and thus the people they seek to become" (Etzioni 2000: 163).

A person who identifies with a particular group expresses this by following the norms and rules shared by its members. Such a community can consist of humanity as a whole (as a human, I follow the norm not to kill others), a society (as a democratic citizen, I follow the norm that voting is a good thing to do) or a smaller group (as a friend, I follow the norm that others should be helped to whatever extent possible). The motivating force of such norms can only be understood by referring to the fact that the individual identifies with the group at hand. This allows one to understand norm-guided behavior as expressively rational: "if norms shape people's preferences, they will tend to abide by these norms because such adherence is a source of intrinsic affirmation" (Etzioni 2000: 163). Norms thus do not only affect the costs and benefits attached to the choice alternatives, but can also become part of his motivational array. This way, a view of the individual arises that lies beyond the dichotomy between the *Homo Economicus* and the *Homo Sociologicus*. He is not exclusively motivated by instrumental considerations how best to serve his interests, but neither is he completely overwhelmed by social forces.

The rationality of instrumental acts depends on the extent to which they help achieve the desired goal. The rationality of expressive acts depends on the extent to which they express the reasons with which the individual identifies. The notion of identification is crucial, since it allows one to understand a wide range of phenomena that cannot be grasped within the economic conception of rationality. Relevant in this respect are issues that are often heavily and emotionally debated (like nationalism, religion, racism, gender, and so on) or more everyday phenomena (like empathy with and caring for one's friends or one's favorite sports team). Identification can thus be defined as "regarding one's self as part of a whole (...) and caring about the fate of that whole (...). This same pattern of thinking is present in deeper and more pervasive types of identification – with family, clan, ethnicity, nationality, and religion. The "I" is defined by relation to a "we." Human beings don't simply belong to groups; they identify with them" (Sober & Wilson 1998: 233).

### 3.2. The level of the reasons themselves: beliefs

The broad conception of rationality postulates that well-informed beliefs and autonomous preferences count as good reasons. The expressive conception stresses that good reasons are fundamentally connected to one's personality. If an individual acts on the basis of reasons with which he identifies, his actions are authentic and thus expressively rational. Some beliefs are an essential part of an individual's deeply-rooted identity. If he acts on the basis of such beliefs, he expresses his identity and is thus expressively rational. The source of such rational actions is ultimately situated at the relation between his beliefs and his deeply-rooted identity. As the example of Luther shows, one can have good reasons for having certain convictions and for basing one's actions on them. In this respect, I want to call such beliefs – and the resulting actions – expressively rational.

Since it is impossible to authentically believe in a proposition and its negation, the requirement of consistency remains valid in the expressive conception of rationality. In this respect, I also want to stress that beliefs are not to be treated as exogenously given. While their formation is essentially a byproduct that cannot be attained in a deliberate and intentional way (Elster 1979: 49-50), individuals can reflect upon and revise their beliefs. I will go into this thought more fully when discussing the expressive rationality of preferences. First, I will give some general suggestions what this entails. Next, I will give an overview to what extent the requirements of economic rationality still hold in the expressive conception of rationality.

### 3.3. The level of the reasons themselves: preferences

Like actions and beliefs, preferences are expressively rational if they are connected to an individual's deepest sense of who he is. Rationality not only concerns decisions how to attain certain ends, but also which ends to attain<sup>23</sup>. Since these can and should undergo rational scrutiny as well, the question arises how one can rationally choose the goals that provide one with something to

---

<sup>23</sup> Elliott Sober and David Sloan Wilson (1998: 280-281) frame this issue in what they label substantive rationality: "philosophers disagree about whether "rationality" refers just to the narrow ability to choose efficient means to achieve whatever ends one might have (sometimes called "instrumental" rationality) or should be understood in a fuller sense to mean that the ends themselves are morally defensible ("substantive" rationality)". Nozick (1993: 163) also speaks of the "substantive rationality" of desires and ends.

live for. Economists have the tendency to neglect such crucial issues: “economic imperialism tends to reduce all acts of choice and the selection of goals to allocative problems of selecting means for ends” (Koslowski 1985: 8). As Nozick (1993: 139) argues, the rationality of these very ends cannot be framed in purely instrumental terms: “something is instrumentally rational with respect to given goals, ends, desires, and utilities when it is causally effective in realizing or satisfying these. But the notion of instrumental rationality gives us no way to evaluate the rationality of these goals, ends, and desires themselves”.

John Harsanyi agrees that the basic problem of the economic conception of rationality “lies in the fact that it restricts rational behavior to a choice among alternative *means* to a given end, and fails to include a rational choice among alternative *ends*. Therefore, it cannot explain why a given person may shift from one end to another” (Moser 1990: 275). Sen (1987: 13-14) also believes that it is this capacity that is missing in both the economic and minimal conception of rationality, stressing that these should “be supplemented by rationality requirements on the nature of the reflection regarding what one should want, value, or aim at”.

An expressively rational person can be said to possess his preferences rather than that his preferences possess him: “it is not enough for an individual to have a set of preferences; it is important that those desires be his or hers in a more active sense. They should belong to the individual because he/she has come through reflection to hold them as his/her own” (Hargreaves Heap 1989: 148-149). This thought complements Elster’s broad conception of rationality, which centers round the notion of autonomy. It shows that rational people want to be autonomous, not because it helps them realize their goals, but because they want their actions to say something of who they are. In order to qualify as rational, people should be autonomous, which means that they should use standards and criteria that they have critically scrutinized and regards as authentically their own (Taylor 1988: 87). The good reasons an autonomous person has for his actions are derived from his character, his identity, his real self, what and who he essentially is. The crucial issue is that expressing all this is a perfectly rational way of acting in accordance with them: “action is rational if it manifests attitudes, values, or principles that it would be inconsistent in a person, under appropriate conditions, not to give expression to, given the character that he is generally content to acknowledge as his own” (Benn 1979: 294).

Expressively rational actions are based on the fundamental things that constitute a person’s identity: “it is important for us to guide our activity in accordance with the requirements of what we care about” (Frankfurt 1999: 84). However, this should not lead one to simply take these cares and the identity

that is based on them as rock bottom to explain expressive actions: “without an explanation of how such commitments and identifications are made, this explanation of actions and their interactions is incomplete” (Taylor 1988: 89). In this respect, it is crucial to see that human beings have the distinctive ability to reflect upon, criticize and revise their own ends, goals and preferences. In contrast with Schmidtz (1993: 55), who speaks of a reflective conception of rationality, I deem it useful to incorporate this within an expressive conception of rationality.

This reflective capacity cannot be captured within the economic conception of rationality: “whereas *Homo economicus* deliberates only about alternative means of achieving stipulated ends, we also deliberate about ends themselves. We sometimes stop to wonder whether an end like maximizing profit is worth having. We have self-regarding ends, to be sure, but they are not given to us in the same way they are given to *Homo economicus*. On the contrary, we shape ourselves and our ends as we go. We are the outcomes as well as the makers of our choices (...). Thus, *Homo economicus* is a poor model of rational choice even when self-interest is all that matters, for even then there is a crucial difference between *Homo economicus* and beings like us. The difference is this: we need to worry about our goals in a way that *Homo economicus* does not” (Schmidtz 1994: 251).

Schmidtz (1994: 227) distinguishes between ends that are instrumental to some further end and final ends that are intrinsically valued. If I want to go running, this may be an instrumental end for the sake of acquiring a healthy condition, but it may also be a final end, if I want to go running just for the sake of being out there running. Like Frankfurt, Schmidtz (1994: 233) stresses that final ends are not always deliberately chosen: “some people are simply gripped by particular final ends. They do not think of the process of acquiring those ends as a choice process”. Such ends serve as a normative framework that allows people to evaluate the ends they can acquire by choice. Against this more or less fixed background, some ends are perceived as more valuable than others. Schmidtz focuses more extensively on the end of having something to live for, which gives people reason to adopt all kinds of goals and reflect whether these are worth striving for. It makes them feel that they are doing something valuable, relevant and important. While unchosen ends can, of course, never be chosen rationally, they can be called rational if adopting them gives the person at hand something to live for (Schmidtz 1994: 241). This does not imply that people adopt ends merely in order to have something to live for.

In this respect, I want to go deeper into Harry Frankfurt’s anthropological views (Frankfurt 1988: 11-25, 47-68, 159-190; Frankfurt 1999: 95-107). According to Frankfurt, what is typically human is not that one has preferences

upon which people base their decisions – other animal species exhibit this form of intentionality as well – but that people can identify with certain preferences and distance themselves from others. A dissertation on the rationality of individuals that fails to take into account the fact that humans are not the mechanistic type of being most economists take them to be would be grossly inadequate.

This capacity to distance oneself from the motivational impulses one experiences cannot be understood within the economic conception of rationality, which focuses on a consistent, complete and continuous ranking of preferences that are weighed off against each other. In Frankfurt's hierarchical anthropology, which provides a fuller, richer and more realistic image of individuals, people can have a preference (of the second order) to be motivated by particular preferences (of the first order). If one identifies at a higher level with an already existing preference, one's entire preference function will be reorganized so that the latter becomes the most important one. Crucial is that such higher-order preferences constitute one's identity, while first-order preferences consist of more or less arbitrary tendencies. Accordingly, someone acts expressively rational when he acts on the basis of his meta-preferences.

This also allows one to understand how people can experience some of their preferences as alien to themselves. An addict who wants to quit smoking but still longs for a cigarette can go two ways (Engelen 2007a: 37-38). If he identifies with his desire to smoke, he becomes a willing addict who acts rationally in all meanings of the word when he smokes. However, if he identifies with his preference to quit, he becomes an unwilling addict and starts to experience the preference to smoke as alien, inauthentic and thus not really his own. If he decides to light a cigarette nonetheless, this may be economically rational, but surely expressively irrational: "what motivates his action is a desire by which, given the alternatives he confronts, he does not want to be moved to act. There is a conflict within him, between a first-order desire to do what he actually does and a second-order volition that this first-order desire not be effective in determining his action (...). In the conflict from which his action emerged, he was defeated by a force with which, although it issued from inside of him, he did not identify himself (...). In this respect it is alien to him" (Frankfurt 1988: 48).

Like Elster, Frankfurt points at the importance of factors that underlie the one-dimensional Homo Economicus and his smoothly ordered preference function. As mentioned above, Elster's broad and Frankfurt's expressive conception of rational preferences are closely connected. An individual acts rationally if his motivations consist of autonomous preferences and well-informed beliefs (broad rationality) that the individual expresses through his

actions (expressive rationality). He acts irrationally if he is moved by heteronomous preferences (broad irrationality), which push the individual without any process of internalization or identification (expressive irrationality). Yet, there is a crucial difference between both. While Elster stresses the formation process of preferences, Frankfurt focuses on the extent to which an individual identifies with the resulting preferences or not (Frankfurt 1988: 54, 170; Frankfurt 1999: 101-102). In what follows, I want to show that Frankfurt's approach has a number of advantages over Elster's.

First, it allows one to understand preferences as authentically one's own, even though they have arisen independently from one's will and choice. According to Frankfurt, it does not matter whether one's preference is formed autonomously but whether one has made it constitutive of his will. This way, one can act freely and autonomously on the basis of a preference that has come about heteronomously. For willing addicts who identify with their preference to smoke, it is completely rational to smoke, even if this preference has arisen heteronomously. After all, it expresses how they think of themselves. For unwilling addicts, who identify with the preference to quit smoking, it is irrational to smoke. Whereas Elster argues that this irrationality arises from the fact that a blind motivational urge for cigarettes overwhelms one's other motives, Frankfurt (1999: 136) argues that it is not the blindness or the irresistibility of this urge that matters but the fact that one has not identified with it. While Elster suggests the right direction, Frankfurt more plausibly locates the exact source of this irrationality.

Second, Frankfurt's views allow one to move further away from an instrumental conception of rationality than Elster's broad conception does. In contrast to the latter, the first suggests that people do not want to be autonomous in order to achieve their goals better. Instead, they simply want to express themselves through their preferences and actions, since they want their lives to be meaningful. This clearly suggests that the instrumental framework of goals and means is too narrow (Frankfurt 1999: 82).

Third, Frankfurt's anthropology is richer than Elster's, who stresses only first-order preferences that have or have not arisen autonomously. In contrast, Frankfurt (1999: 108-116) shows the relevance of a person's underlying identity. In this respect, Frankfurt speaks of "the things we care about most and with which, accordingly, we are most closely identified" (Frankfurt 1988: 91). The use of the passive suggests that the process of identification is not to be interpreted in a completely voluntaristic way. Nevertheless, individuals sometimes have to make up their mind and decide to identify with some preference and distance oneself from its opposite.

Even though an individual's most fundamental cares are not typically the object of a free choice, Frankfurt nevertheless stresses that individuals can, to a certain extent, choose them: "when someone undertakes to choose his final ends, he is proposing to identify the goals and the values that will most basically guide and constrain his conduct. In other words, he is seeking to resolve the question of how he should live" (Frankfurt 1999: 91). This forms a constitutive moment in the construction of his identity. While one cannot but care about certain things, this forms no obstacle to but is a necessary condition to lead the life of a rational human being. After all, if these things were not given in some sense, one would have no basis whatsoever to adopt life projects (Frankfurt 1988: 94). As a result, a more substantial conception of rationality than the minimal one arises: "the will of a rational agent need not be, then, empty or devoid of substantial character. It is not necessarily altogether formal and contentless" (Frankfurt 1988: 190). In this respect, Frankfurt speaks of "fundamental rationality" (Frankfurt 1988: 188). Schmitz (1993: 68) connects this to an expressive conception of rationality: "being rational in a full-blooded sense – developing oneself in a way that is true to oneself – is a profoundly demanding activity". In other words, "we identify ourselves largely in terms of what we do, and therefore individual rationality requires us to do things that can ground the kind of self-conception we would like to have" (Schmitz 1993: 64).

In what follows, I will briefly run over the different requirements of the economic conception of rationality at the level of preferences and try to show how the expressive conception criticizes and complements them.

### 3.3.1. *Consistency*

The first requirement is that of consistency. On the basis of Frankfurt's remarks, one can argue that inconsistency is not really problematic. Someone with inconsistent preferences can still be rational as long as he does not remain indecisive towards this inner ambivalence. Even if he identifies with one of his preferences, the opposite preference can continue to exist. Frankfurt (1988: 172) stresses that inconsistent preferences often continue to exist, even after one has identified with one of them: "quite possibly, the conflict between the two desires will remain as virulent as before (...). The conflict between the *desires* is in this way transformed into a conflict between *one* of them and the *person* who has identified himself with its rival". An individual is thus expressively irrational not if he has inconsistent preferences, but if he fails to identify with one of these preferences or if he acts on the basis of preferences with which he has not identified. Irrationality thus consists in indecisiveness

rather than inconsistency. When a person remains indecisive which preference to identify with, his unity is at stake (Frankfurt 1999: 99-100). Such an individual is called a wanton. He is continuously led by arbitrary urges without ever asking himself whether he actually wants to be moved by them: “not only does he pursue whatever course of action he is most strongly inclined to pursue, but he does not care which of his inclinations is the strongest” (Frankfurt 1988: 17). A wanton has no higher-order preferences and thus “no identity apart from his first-order desires” (Frankfurt 1988: 18).

### *3.3.2. Transitivity*

The second requirement is that of transitivity. Since this is a further translation of the requirement of consistency, the same remarks hold here as well. In addition, one can argue that such purely formal conditions are not what rational individuals care about.

### *3.3.3. Completeness*

The third requirement is that of completeness. Here, the economic and expressive conceptions of rationality coincide. If one wants to be called rational, one should choose between two options or decide that one is indifferent between them. Indecisiveness is irrational in both conceptions. While economists emphasize the importance of one-dimensional preference functions for modeling purposes, authors like Frankfurt show the underlying rationale of this condition. If a person does not identify with one of his preferences, he becomes a slave of his arbitrary urges and no longer employs the uniquely and distinctively human capacity to reflect upon his motives.

### *3.3.4. Continuity*

The fourth requirement is that of continuity. In this respect, the economic and expressive conceptions of rationality diverge widely. This condition, which stresses that the individual should always be able to weigh off different options, turns out to be grossly inapt in an expressive conception of rationality. While clearly economically irrational, lexicographic preferences fit the expressive framework perfectly. People often find it inconceivable to act against what they care about. They refuse to accept a decrease in one of these goods in exchange



for an increase in another good. Some things are so closely connected to one's identity, that it is inconceivable not to honor them in one's actions. Exactly like rationality implies logical constraints (one cannot believe both  $p$  and not- $p$ ), there are limits on what a rational person can prefer.

Frankfurt's hierarchical anthropology shows that some preferences are more closely connected to one's deeply-rooted identity than others. Arguing that all goods can be substituted by other goods, economists fail to understand that some preferences are more fundamental than others and thus that preferences cannot be merged into an overall, one-dimensional preference function. In his evaluation of the main shortcomings of the Homo Economicus model, Mark Lutz argues that certain preferences "resist being calculated, quantitatively compared and balanced against each other. So it appears that the economists' treatment of utility as a one-dimensional abstract concept is, for purposes of describing real human behavior, a highly misleading one" (Koslowski 1985: 99). Next to basic material needs, which are so fundamental for one's survival that they cannot be traded off, Lutz refers to moral motivations as an example of lexicographic preferences. He situates these at the level of a higher 'self' which conflicts with a lower 'ego' that is exclusively concerned only about the satisfaction of basic desires (Koslowski 1985: 104-107).

Since the Homo Economicus is nothing more than a bundle of preferences, Lutz properly criticizes it "as lacking the distinctly human qualities and so falling short of being the image of a genuine human" (Koslowski 1985: 103). This is why he has been called a moron and a fool: "the purely economic man is indeed close to being a social moron. Economic theory has been much preoccupied by this rational fool decked in the glory of his *one* all-purpose preference ordering" (Sen 1977: 336). The reduction of rationality to its economic conception thus results in an all too narrow view of individuals, which inhibits a thorough understanding of human rationality. People are more than a one-dimensional bundle of mutually substitutable preferences.

The main purpose of the expressive conception of rationality is thus to provide a richer view of people<sup>24</sup>. Introducing higher-order and lexicographic

---

<sup>24</sup> Mueller (1993: 512) subscribes to my argument in favor of an image that is richer than the narrow one of Public Choice theorists: "too often rational, self-interested behaviour is modeled as the maximization of (...) an individual's wealth. Although such stark simplifications often do not cause trouble, they do sometimes lead to inaccurate predictions (...). Too often economists (...) reject any attempt to broaden the behavioural or institutional premises of their models on the grounds that such broadening will make their theory ad hoc; they then continue to accept the theory despite a very weak correspondence between its predictions and what is observed. More accurate predictions would be possible if we were to build our models on a richer behavioural base".

preferences, it better captures what the life of a rational person is all about. As this chapter's opening quote by Mark Lutz suggests, the Homo Economicus model fails to capture some of the aspects of rationality that are crucial in order to understand what it means to be human. Whereas the economic conception of rationality cannot capture this essence, the expressive one can.

### *3.3.5. Exogeneity*

As should already be clear, the fifth requirement – that of exogenously given preferences – should be dropped in the expressive conception as well. This assumption is completely inadequate if one wants to take into account the fact that people can reflect upon, revise and even indirectly transform their preferences. The need for a radically different model of human beings in order to incorporate this thought is expressed by John Plamenatz (1966: 172-173), who argues that the conventional view can stress “that men are as much creatures as creators of their social environment and still fall far short of the truth in describing what is involved in being a social creature. Though we allow that men's needs and ambitions differ greatly from society to society, and also that as society changes so too do the wants and interests of its members, we can still treat man as if he were no more than a subject of desires intent on satisfying as many of them as possible (...). The inadequacy of this account lies, not only in its hedonism and the strong (though sometimes unconscious) bias towards egoism, but in the quite unreal conception of what a moral and social being is (...). Man (...) is a self-conscious, self-communing, animal (...) who wants to be one kind of person rather than another and to live one kind of life rather than another”.

Expressively rational persons not only think of their actions as expressing their identity but also realize that they have an influence on their identity. It is thus to some extent possible for a person to choose which life to live (Frankfurt 1999: 91). After all, his actions partly determine the kind of person that he will become: “adopting one final end may lead a person to become engaged in a network of feeling, emotion, thought, and action that differs very considerably from the network in which he would be engaged if he were to adopt another. The life he will have if he pursues the one may therefore be much richer in meaningful activity, and in overall desirability, than the life he will have if he pursues the other” (Frankfurt 1999: 86-87). In contrast with the static model in which preferences are exogenously given, a more dynamic model is needed to incorporate the effects of one's choices on one's own personality and preferences: “the quality of our lives is a function not only of what we get, but also of what we are. And what we are, no less than what we get, depends on

what we choose (...). Reflectively rational agents understand that their preferences can and will change over time” (Schmidtz 1993: 55).

Preferences and identities are thus not only formed by an individual’s environment, but also by his own actions. While they are not under his direct control, they are influenced by the choices over which he does have control (Schmidtz 1990: 60). This is certainly the case for public choices. These have an impact on the institutional context, which in its turn forms the citizens, their beliefs, preferences and personalities. I will analyze this interplay between individual actions and social institutions more fully in the third part of this dissertation. However, this same is true for private choices, such as consumer behavior: “because economic activity is so much of what we do, it has a powerful influence on who we are” (Bowles & Gintis 1986: 127). In their criticism of the economic conception of rationality, Samuel Bowles and Herbert Gintis (1986: 21) focus on the twofold relation between an individual’s preferences and his actions: “the individual is socially constituted in such a way that preferences and action are mutually determining, and hence that preferences are formed through choice”. Individuals do not act only to get certain things, but also form their preferences and their personalities while doing so. People become who they are partly through their own actions. Intentional actions are still aimed at fulfilling one’s preferences (instrumental rationality), but one should realize that these preferences themselves are formed by such actions (expressive rationality): “individuals and groups (...) act not merely to get but to become” (Bowles & Gintis 1986: 22). This way, the endogenous formation of preferences can be incorporated in a conception of rationality.

It is even possible to consciously and intentionally try to change one’s own preference ordering. In what can be labeled character planning, one deliberately aims to form particular preferences that one does not yet have (Elster 1979: 177-179; Elster 1983: 117-118). This is rational if one has a meta-preference that provides a reason to develop strategies aimed at the formation of one’s character. Take a person who has a tendency for laziness (first-order preferences) but wants to become a sportive type (second-order preferences). In order to create such a personality, he can enroll for sports happenings or take a gym subscription. Such attempts to become someone else than one currently is are completely incomprehensible within the economic conception of rationality but can be interpreted within the expressive conception.

To be sure, I do not want to defend the view that preferences are expressively rational only if they are deliberately formed. After all, the formation of one’s personality can only be achieved indirectly. The effects of one’s actions on one’s identity and preferences are essentially by-products and can thus never be

realized in a direct, intentional, conscious and deliberate way (Elster 1983: 43-71). Nevertheless, the fact that individuals act primarily to fulfill or express their preferences, does not exclude the possibility that they take into account the side effects of their actions on their own preferences. Since most preferences are not consciously and intentionally formed, it would be preposterous to label them irrational. This is exactly why I refer to Frankfurt as an alternative to Elster in this dissertation on rationality.

A good example is the role work fulfills in the life of individuals. Economists analyze work as means towards the realization of some goal, like that of receiving an income. Calculating the expected costs (loss of time) and benefits (increased income), the economically rational individual decides how much and which work to perform. This purely instrumental view does not take into account that work plays a crucial role in the self-realization and self-actualization of the individual at hand: “we realize ourselves more in production than in consumption” (Koslowski 1985: 54). People do not choose their work purely on the basis of cost-benefit calculations, but take into account to whether it allows them to express who they are and how it affects their identity (Koslowski 1985: 112; Sen 1977: 333-334). Conceiving of one’s work as interesting and meaningful is crucial when one wants to lead a good life. Who I am is closely connected to what I do and the work I perform constitutes a large part of what I do. When applying for a job, most people will take into account to what extent they can identify with the role attached to it. Such crucial issues clearly do not fit the narrowly instrumental framework of economic rationality.

### 3.3.6. *Egoism*

The sixth requirement is that of egoism. Here too, the economic and expressive conceptions of rationality clearly diverge. Whereas the Homo Economicus is egoistic, an expressively rational person often is not. Even though economists typically try to incorporate substantial aspects of people’s lives in some notion of enlightened self-interest, expressive considerations can hardly ever be reduced to pure self-interest. They are more or less disconnected from the aim of serving one’s purely private interests and connected to the image one wants to have (for others but also for oneself) (Engelen 2005: 65). People define themselves more through the things they care about rather than through their self-interested goals. Expressing the identity that is constituted in this way is a perfectly rational thing to do: “action is rational (...) if it is consistent with these autonomously derived *commitments* to values, principles and ideals the individual cares about” (Taylor 1988: 87).

Like the example of Luther suggests, people live for their ideals, principles and loved ones: “a person who does not have commitments has little with which to identify himself. What we are is in large part what we stand for. We think of having to make a stand on behalf of our ideals or on behalf of our loved ones as frightening and painful, and it often is. Yet to make a stand for what we think is right is one of the most self-defining things we can do” (Schmidtz 1993: 62). Focusing exclusively on the level of the individual’s ‘ego’ and his self-interested concerns, economists inevitably leave out crucial aspects in human motivation, like moral commitments, personal engagements, and so on. Phenomena like trust and personal integrity are crucial in everyday social life: “when we trust we act on the assumption that the other is going to honor his or her commitment and to forego self-interest (...). Economic man, incapable of anything beyond self-interest, is also incapable of integrity and, as a result, of trustworthiness and a capacity to trust others” (Koslowski 1985: 114). By focusing on egoism, economists start from an abstraction that can be useful for scientific research, but neglects what is characteristic of humans. That is why their reduction of the individual to a one-dimensional utility-maximizing being is to be criticized and complemented.

### 3.3.7. *Egocentrism*

One could argue that the seventh assumption of egocentrism is under pressure here as well. As I have suggested, there is something about an individual’s cares that transcends him. The individual is not fully in control, since the objects of his cares are to a large extent arbitrary and contingent. However, I believe it would be wrong to conclude from this that the assumption of egocentrism should be done away with. Even though an individual does not freely choose all of his final ends, the latter still remain *his* ends. It is always the individual himself who cares about certain things and these things are only relevant insofar as they constitute the identity of individuals. The fact that the expressive conception of rationality is based on more or less contingent objects of care that are not always freely chosen does not imply that one should give up the individual as the final judge in evaluating and choosing from the available alternatives.



## CONCLUSION OF PART I

---

*“It is theoretically unsatisfying to end the investigation into the nature of rationality with the pluralism of rationalities, since this viewpoint does not fulfill the reason’s need for unity. A plurality of rationalities is, however, more apt to save the phenomena than monistic economic rationality”*

(Koslowski 1985: 9-10)

### 1. Schematic overview

In order to avoid that one no longer sees the forest for the trees, I want to summarize schematically the main line of reasoning of the first part of this dissertation. The following scheme presents the requirements of the different conceptions of rationality at the different levels I have distinguished.

	Actions	Beliefs	Preferences
Economic rationality	maximization instrumentality intentionality	consistency perfect rationality	consistency transitivity completeness continuity exogeneity egoism egocentrism
Minimal rationality	intentionality	consistency	consistency transitivity egocentrism
Broad rationality	intentionality	consistency judgment	consistency transitivity egocentrism autonomy
Expressive rationality	non-instrumentality intentionality	consistency authenticity	consistency transitivity completeness egocentrism authenticity

The main strategy in the first chapter of this dissertation was to start from the dominant conception of rationality within the social sciences, namely the economic one, and strip it from the requirements that are too demanding to speak of rationality in a more intuitive and common sense. When explaining individual behavior, it is wise to start by trying to understand it as economically rational. If this does not work, one should try to understand it as minimally rational. If it falls beyond this conception of rationality as well, one should still try to explain it as intentional (Elster 1983: 10). While the economic conception of rationality includes all requirements – egoistic rational intentional action – the minimal conception of rationality is less demanding – consistent intentional action – and the requirement of intentionality is even less strict, stipulating only that one's actions should be based on reasons. While the minimal conception of rationality is more demanding than intentionality (because of the requirement of consistent reasons) and less demanding than economic rationality (Van Liedekerke 2002: 116), it encompasses only a limited number of conditions. As such, it remains very formal: “consistency, in fact, is what rationality in the thin sense is all about: consistency within the belief system; consistency within the system of desires; and consistency between beliefs and desires on the one hand and the action for which they are reasons on the other hand” (Elster 1983: 1).

Stripping down the economic conception of rationality to a minimal one, I have broadened the range of phenomena that can be understood as rational. The danger is that such a conception lacks analytical power and explanatory strength and thus turns into “a model that is pliable enough to explain anything and hence explains nothing” (Margolis 1984: 106). To avoid that rationality becomes such an empty notion, I have proposed more demanding requirements in the second chapter, where I have developed the broad and expressive conceptions of broad rationality. I believe it is important to ground these alternative conceptions in empirical research in order to avoid this ‘ad hoc’ criticism. Here, I side with Ben-Ner and Putterman (1998: 19) who aim “to characterize individual preferences and behavior richly enough to conform with commonsense observations and with the findings of social-scientific research (including psychology and the social research of biologists), and parsimoniously enough to be useful for systematic inquiry”. That is why I have referred to insights from experimental economists, psychologists and sociologists in order to develop these still largely neglected conceptions of rationality. While the minimal conception is based on the central notion of reasons, these more substantial conceptions conceptualize – each in their own way – what exactly good reasons consist of.



## 2. The relation between different conceptions of rationality

The first part of this dissertation can be interpreted as a return to authors like Adam Smith and Max Weber, who had a richer view of individual motivation than conventional economists nowadays. Whereas Smith explicitly acknowledged the relevance of other than purely self-regarding preferences, Weber distinguished between different types of rational action. First, there is instrumentally rational action, which “involves rational consideration of alternative means to the end, of the relations of the end to the secondary consequences, and finally of the relative importance of different possible ends” (Weber 1914: 26). This clearly corresponds to the economic conception of rationality. Second, there is value-rational action, which is not instrumentally aimed at achieving some goal but which is performed for its own sake. Weber (1914: 25) refers in this respect to “the actions of persons who, regardless of possible cost to themselves, act to put into practice their convictions of what seems to them to be required by duty, honor (...), personal loyalty, or the importance of some “cause” no matter in what it consists”. This corresponds to the expressive conception of rationality. Finally, Weber refers to affectual action, which is based on the individual’s emotions, and habitual action, which is based on some form of ingrained habit. According to Weber, what distinguishes the first two types of action – and what makes these more rational than the latter – is the degree in which these are self-consciously performed and planned.

Like Weber, I also defend a definition of rationality based on the notion of reasons. This implies a criticism of the exclusively instrumental conception of rationality in which “rationality cannot be anything else but an attribute of the relation means/ends (...). But nobody needs to accept the postulate that the attribute “rational” can be applied exclusively to this relation. Weber starts from a simpler, more natural and powerful assumption (...) when he says that a given behaviour is rational as soon as it is inspired by reasons” (Boudon 1993: 8). The latter corresponds to the minimal conception of rationality. The fact that the individual at hand has reasons for doing, believing or preferring something is enough to call him rational in a minimal sense. These reasons may consist of goals that he wants to realize, but they can also refer to certain aspects of his identity or to socially shared criteria on how to behave.

Weber might have argued that I stretch the meaning of rationality too much by understanding expressive and norm-guided behavior – with their affectual and habitual characteristics – as rational. A multiplication of rationality conceptions may indeed seem ‘ad hoc’ in that it allows one to change the meaning of

rationality depending on the phenomenon to be explained. Consequently, the purported meaning of rationality can be said to lack unity. I think this criticism is mistaken. The different conceptions of rationality are not constructed to make up a single, unified theory. Instead, they are to be understood as alternative models that complement each other. A particular action can be rational in one specific sense of the word but irrational in another sense. Wearing a particular sweater can be instrumentally rational (if it protects one against the cold) but expressively irrational (if it does not express how one perceives oneself). The very fact that it remains possible to speak of irrationality shows that rationality has not become a 'passe-partout' that explains everything. As this conclusion's opening quote by Peter Koslowski suggests, distinguishing different conceptions of rationality has both advantages and disadvantages. While it lessens predictive force, it provides for a richer and more adequate view of individuals.

While it is thus crucial to respect the multitude of meanings of rationality, I also want to stress the unifying capacity of the minimal conception of rationality. Not only does it closely correspond to common sense, it also functions as an umbrella, encompassing the more specific conceptions under one general heading. This allows one to address "a fundamental problem within the discussion between philosophy and economics (...). It is a question of reconciling diverse conceptions of rationality and of constructing, or rather reconstructing, one single concept of rationality" (Koslowski 1985: 141).

To be sure, the instrumental and expressive conceptions of rationality are closely connected, since goal-oriented actions are crucial in anyone's view of a meaningful life: "without the goal-directed activity that is the locus of instrumental value, we would lack the indispensably foundational sense that we have of ourselves as rational agents (...). The activity in which we engage (...) is valuable to us not only for the sake of its products or goals. It is also valuable to us in itself, because it is inherently important to us to have something useful to do" (Frankfurt 1999: 178). Without instrumental actions, life would simply have no meaning: "living without goals or purposes is living with nothing to do. Without ends, there are no means" (Frankfurt 1999: 84). In this respect, Frankfurt argues that the strictly instrumental distinction between ends and means has a blind spot: "in my opinion, however, the framework is too narrow and too rigid. It cannot accommodate a realistic appreciation of the issues we confront when we try to decide upon a good way of living. Our conception of the relationship between means and final ends has to be more spacious and more supple. Otherwise, it will impede our efforts to develop a comprehensive and authentic representation of what actually concerns us when we concern about how we should live" (Frankfurt 1999: 82-83). What we deem important in our lives cannot be grasped in an instrumental framework.

There is also a close relation between the expressive and the broad conceptions of rational preferences. After all, heteronomous and thus broadly irrational preferences are typically inauthentic and thus expressively irrational as well. Both Elster's broad conception as well as Frankfurt's and Hargreaves Heap's expressive conception of rationality center round the notion of autonomy.

Now, it is crucial to see that the different conceptions of rationality are best understood as complementing rather than substituting, replacing or undermining each other. Take the relation between the instrumental and expressive conception of rationality. In this respect, I want to defend a model of individuals as motivated by different considerations of which the relative weight depends on the context. Instrumental considerations dominate in settings where individuals can directly realize their most preferred outcome, like markets. In contrast, expressive considerations tend to dominate in settings where individual behavior has little or no actual consequences, like large-scale elections. Here, no individual can directly choose the outcome he prefers, since all he can do is express his preference, while the resulting outcome emerges but is not chosen by anybody (Eusepi & Hamlin 2004: 125-126). Because such circumstances do not allow one to achieve one's goals, it makes no sense at all to act in an instrumental way. Consequently, rational people will search for other ways to act on the basis of one's values, principles and preferences. While some actions are best understood as attempting to satisfy one's preferences (these can be labeled instrumental actions), others are best understood as expressing one's identity (these can be labeled expressive actions). As I have already suggested, some actions, like the wearing of a particular piece of clothing, are best understood as motivated by both instrumental and expressive considerations.

Crucial in this respect is that rationality can and should not be reduced to one of its conceptions. Perhaps the main point of the first part of this dissertation is to provide an antidote for such reductionist strategies, of which the economic one is the most renowned. As Frankfurt (1999: 86) stresses, the expressive aspects of rational actions can and should not be reduced to some sort of instrumental reasoning: "it is not essential that the activity he devotes to the things he cares about be successful. The extent to which a life is meaningful depends less upon how much it accomplishes than upon how it is lived. What counts primarily is the extent to which the person cares about the final ends at which he aims". While instrumental rationality is important (actions are valued less if they are not successful in accomplishing what they are aimed at), it has to be complemented by broad rationality (actions often spring from a desire for autonomy and judgment) and expressive rationality (actions often concern the things one cares about and are thus often performed for their own sake and valued intrinsically).

By pointing out the plausibility and relevance of different conceptions of rationality, I hope to have shown that rationality is more widely applicable in the social sciences than its exclusively economic conception allows for. In this sense, I believe that the more realistic character of the alternative conceptions is a welcome improvement in any respect. More work is needed to develop and empirically test a general theory to explain under which conditions different sorts of considerations come to dominate. An important task for the social sciences is to answer the question when the economic conception of rationality is the most relevant one and when its alternatives are more plausible. In the following part, I will develop this line of reasoning in greater detail.

## PART II

# THE RATIONALITY OF VOTING DECISIONS



## INTRODUCTION TO PART II

---

*“Electoral participation is the most widespread form of political participation and, for a great number of citizens, the only one they exercise on a regular basis”*

(Perea 2002: 645)

In this second part of the dissertation, I want to apply the insights gained in the first part to the domain of politics. Starting from the assumption of methodological individualism, elections seem to be a logical place to start. As this introduction’s opening quote by Eva Perea suggests, they form the pre-eminent occasion for most citizens to participate in politics. Because direct participation in policy measures is practically impossible in large societies, modern democracies have to resort to some type of representation. Since democratic policy decisions should be made by a publicly elected government, voting is the most prominent and widespread act individual citizens perform in the public domain.

In what follows, I want to distinguish between two voting decisions. The first is *whether* one will vote. Will one take the trouble of showing up at the polling station to register one’s vote? The second is *how* one will vote. Once inside the voting booth, which party or politician will one vote for? I will focus on these decisions in a context where attendance at elections is voluntary and abstention is thus not formally sanctioned. In the third chapter, I will analyze how these decisions are modeled in the economic conception of rationality and formulate a number of critical remarks. In the fourth chapter, I will show how alternative conceptions of rationality solve these problems and are thus highly needed to understand why and how people make voting decisions.

Here, I am gradually moving into normative terrain. After all, democratic politics is the domain where citizens express what they believe to be a desirable society. Whereas the first part of this dissertation remains on purely explanatory terrain and thus does not engage in normative issues, the third explicitly aims to answer the normative question of how to improve society. This second part forms a bridge between both. Stressing that people should be analyzed as rational in their public role as citizens, it aims to analyze how they choose the representatives that are to implement what they believe to be a desirable society.





## CHAPTER 3

### ECONOMIC RATIONALITY AND VOTING DECISIONS

---

*“Persons’ voting behavior may have many explanations, but one that must usually have relatively little weight is the intention to produce a favored outcome. The larger the electorate, the less important a motivation to produce outcomes becomes”*

(Brennan & Lomasky 1989: 46)

#### 1. Economic rationality and voting decisions

In order to show how the political domain can be analyzed from the perspective of the economic conception of rationality, I want to refer to the field of Public Choice theory<sup>25</sup>. This is aptly defined as “the application of economics to political science” (Mueller 2003: 1). Public Choice theorists thus aim to analyze political phenomena by modeling individuals as *Homines Economici* in both their private and public roles: “the salient feature of all of public choice is its use of the *homo economicus* assumption (...). The bulk of the work in public choice employs this assumption to explain and predict behavior in a political context” (Mueller 1993: 405). As I have already suggested, I will focus more specifically on the decisions whether or not to vote and whom to vote for.

##### 1.1. Economic rationality and the decision whether or not to vote

###### 1.1.1. *The calculus of voting*

The standard Public Choice analysis of the decision whether or not to vote was formulated first by Anthony Downs (1957: 260-276) and was later formalized by Gordon Tullock (1967: 110-114). They argue that economically rational

---

<sup>25</sup> I will go into Public Choice theory more fully in the fifth chapter, where I analyze extensively the work of James M. Buchanan, perhaps its most important and best-known defender.

citizens vote only if its expected costs ( $C$  for costs) do not exceed its expected benefits ( $B$  for benefits). The latter, however, only arise if their vote has an impact on the electoral result, which depends on the extremely low probability of a single vote being decisive ( $P$  for probability). Because the resulting condition to vote ( $PB > C$ ) is almost never met, these authors argue that economically rational citizens systematically decide to abstain. Downs (1957: 260) comes to this conclusion of rational abstention by assuming “that every rational man decides whether to vote just as he makes all other decisions: if the returns outweigh the costs, he votes; if not, he abstains”.

While all Public Choice theorists focus on such a cost-benefit calculus, they do not agree on its nature. According to some, citizens are to be modeled as trying to further their self-interest, further qualified as their subjective well-being or material wealth. In this narrow view, citizens vote if they think this will benefit them in self-interested terms. According to others, it is not necessary to assume that individuals act egoistically (Fiorina 1997: 408-409). This broader view still refers to the consequences that citizens want to bring about. It stresses the instrumental motivation of citizens going out to vote in order to achieve their goals, whatever these are. If one wants some specific candidate to win the elections – regardless whether this will benefit oneself – and one believes that one’s vote will help realize this, it is rational to go out and vote. To be sure, the assumption of egoism remains dominant among Public Choice theorists, since it allows them to make specific predictions. Nevertheless, I will focus mainly on the broader view, not only because it is intuitively more plausible, but also because any counterargument against it automatically applies to the narrower view as well.

### *1.1.2. The paradox of voting*

Within the economic conception of rationality, it is thus hard to grasp why any rational citizen would ever vote. The crux lies in the fact that the impact of a single vote on the electoral outcome ( $P$ ) is infinitesimal<sup>26</sup>. Despite the prediction that rational individuals will always decide to abstain, lots of citizens still vote. This has become known as the “the paradox of voting” (Blais 2000: 2). Even though one would expect it to have a comparative advantage in explaining voting decisions, Public Choice theory is blatantly falsified here (Margolis 1984: 3, 12). Its failed attempt to explain why people vote is even

---

<sup>26</sup> Sometimes the electoral outcome is already known before citizens decide whether or not to vote, as was the case for citizens living in the West Coast states during the Nixon landslides in the 1980s (Brennan & Lomasky 1993: 35).

renowned for being “the Achilles’ heel of rational choice theory in political science” (Aldrich 1997: 373).

To avoid misunderstanding, I want to stress that voting itself can hardly be called a paradox. If a paradox is “a tenet contrary to received opinion” (<http://m-w.com>), it is not so much voting that is paradoxical but the theory that no rational citizen ever votes. It is only within an exclusively instrumental account of human behavior that voting becomes a mystery. However, Alan Carling (1998: 21) is right in pointing out that even this is no genuine paradox, which is properly defined as “an argument that apparently derives self-contradictory conclusions by valid deduction from acceptable premises” (<http://m-w.com>).

As this chapter’s opening quote by Geoffrey Brennan and Loren Lomasky suggests, any explanation of voting behavior as an attempt to bring about preferred goals is highly dubious. Because the influence of a single vote on the electoral outcome is infinitesimal, any instrumentally motivated individual will decide to abstain. A single individual’s decision to go out and vote cannot be understood as an attempt to bring about the consequences he prefers. It is completely irrelevant whether his goals are defined narrowly (in egoistic terms) or more broadly (in terms of ends, whatever these are): “whether one favors selfish or selfless ends, virtually any activity in pursuit of those ends would be more effective than the time spent on voting” (Friedman 1996: 6).

In addition, even if one’s vote influences the electoral outcome, it will only select one of many political representatives. The problem then arises that one does not know for sure whether this politician will select the policy measures that one prefers. And even if he does, his choices are likely to have only a small impact on the state of affairs. The basic point is that “citizens voting in mass elections neither bear the full consequences of their decision nor have much impact on the outcomes. The combination of these two features of large elections means that any conception of voting behavior that is instrumental in nature is highly suspect” (Fiorina 1997: 403).

This situation resembles a classic Prisoner’s Dilemma in which it is rational for every individual to freeride and profit from the contributions of others to the public good. Even though all citizens want the democratic system to continue, the instrumentally motivated ones will give in to the incentive to abstain at elections. Since contributing to the public good of democracy is individually costly but not strictly needed, economically rational citizens will be tempted to freeride (Bowles & Gintis 1986: 138). While the abovementioned analysis assumes a fixed environment of high turnout, this line of reasoning suggests that the decision to vote is a strategic rather than a parametric one. Carling (1998: 21-24) has argued that this gives rise to a genuine paradox of voting.

Whether it is rational to vote depends on the decisions of one's fellow citizens. If everybody deems it rational to abstain, turnout drops to zero. This increases  $P$  drastically and makes voting the rational thing to do. If, however, everybody thinks this way, everybody will rationally decide to vote, resulting in the initial situation of high turnout. The conclusion that nobody votes if everybody votes and vice versa then forms a genuine paradox.

### *1.1.3. Amending the calculus of voting*

To disentangle these mysteries and explain why so many citizens vote, several strategies have been deployed (Blais 2000: 3-10; Dowding 2005: 442-453). In search of alternative explanations of the decision to vote, most of these stress that rationality is wholly subjective in nature, thereby allowing for individual differences in aims, beliefs and preferences (Carling 1998: 29).

A first strategy stresses that it does not matter what the objective probability of a single vote being decisive is, but how citizens assess this chance. This way of explaining turnout by increasing  $P$  assumes that individuals systematically overestimate the impact of their vote on the electoral outcome. One can rationally decide to vote on the basis of the false belief that  $P$  is quite large (Riker & Ordeshook 1968: 38-39). Some empirical studies indeed find that "many people are prone to overestimate  $P$ " (Blais 2000: 81). Nevertheless, this must not distract from the fact that "on any reading, the probability of any one voter's being decisive (or more generally the extent of any individual voter's influence on electoral outcomes) is bound to be small" (Brennan & Lomasky 1993: 73). As a matter of fact, most of the available empirical evidence shows that  $P$  has only a weak impact on the decision to vote (Aldrich 1997: 387-389; Brennan & Lomasky 1993: 120; Mueller 2003: 309-312, 319)<sup>27</sup>. All this is reassuring with respect to both the intelligence and the rationality of voters (Mueller 2003: 329). It would indeed be odd to assume that a majority of citizens is wholly incapable of making fairly simple statistical estimates.

A second strategy stresses that voting is often perceived as taking very little trouble (Blais 2000: 87). However, this strategy of minimizing the expected costs ( $C$ ) is doomed to fail as long as  $P$  remains infinitesimally small. The standard argument is that even a small cost ( $C$ ) will discourage a rational citizen to vote, since the expected benefits ( $PB$ ) are negligibly small (Barry 1970: 14-15).

---

<sup>27</sup> The most straightforward proof can be found in questionnaires: "when asked why they vote few people cite the probability that their vote will be decisive" (Dowding 2005: 448).

A third strategy increases  $B$  by stressing the importance of elections. Some citizens believe that there is a small chance of their vote bringing about enormous benefits, either for themselves or for society as a whole. Downs (1957: 261-262), for example, argues that people vote in order to uphold the democratic system they live in: “if no one votes, then the system collapses because no government is chosen. We assume that the citizens of a democracy subscribe to its principles and therefore derive benefits from its continuance; hence they do not want it to collapse”. However, incorporating a wide range of benefits to raise  $B$  does not help much as long as  $P$  remains negligibly small. While this strategy drops the assumption of egoism, it continues to explain the decision to vote in instrumental terms. This fails because of the simple fact that a single vote has no impact whatsoever on the way a democracy functions. The collective action problem remains that each individual may consider democratic government to be a public good, but choose not contribute to it.

The fourth strategy to solve the paradox of voting is the most promising one. In line with William Riker and Peter Ordeshook (1968), it has been argued that individuals vote if they experience satisfaction from the act of voting itself. Voting is deemed pleasurable, no matter what the electoral result is and thus regardless of whether one’s preferred candidate is elected. In other words, voters perceive the trouble of going out to vote as a benefit<sup>28</sup>. Because a single vote is virtually unable to influence the electoral result, rational individuals will largely disregard such consequences and will only take into account the benefits and costs of the act of voting itself. Next to the instrumental benefits of voting ( $PB$ ), one must thus incorporate the expressive benefits from the act of voting itself ( $D$ ) (Carling 1998: 27). Often, these benefits arise from fulfilling what people consider their duty. This argument adds an extra term ( $D$  for duty) to the comparison so that the resulting condition to vote ( $PB + D > C$ ) is met more easily. At first sight, this seems to be an empirically promising move: “the quantitative importance of  $D$  is much greater than that of either  $P$  or  $B$ ” (Mueller 2003: 311).

Several authors follow this line of reasoning. First, Riker and Ordeshook (1968: 28) argue that expressing one’s vote can bring about various sorts of satisfaction: “1. the satisfaction from compliance with the ethic of voting (...) 2. the satisfaction from affirming allegiance to the political system (...) 3. the satisfaction from affirming a partisan preference (...) 4. the satisfaction of deciding, going to the polls”. Second, Morris Fiorina (1976: 395) distinguishes between partisan and independent citizens, the first of whom experience

---

<sup>28</sup> One could argue that this strategy lowers the cost of voting below zero, since voting is experienced not as a cost, but as a benefit. In the end, this is a purely terminological question. The underlying thought remains the same.

“utility or disutility of satisfying or violating one’s party allegiance”. Arguing that *D* plays a role for some citizens but not for others, he is able to explain the intermediate turnout levels that characterize elections in modern democracies with voluntary voting. Third, Geoffrey Brennan and Loren Lomasky (1993: 33) argue that “revealing a preference is a direct consumption activity, yielding benefits to the individual in and of itself”. This allows them to distinguish between instrumental voters who – like investors – seek to bring about something that benefits them and expressive voters who – like consumers – gain utility from the act itself (Ferejohn & Fiorina 1974: 526)<sup>29</sup>. Fourth, Alexander Schuessler (2000a) applies his “logic of expressive choice” to voter decisions, arguing that voting itself is thought of as agreeable. Waiting in line to vote is not perceived as an impediment to be overcome, but as an additional benefit of voting (Schuessler 2000a: 25, 56). Even though Schuessler (2000b: 103-105) explicitly wants to move away from Public Choice theory, he continues to define the expressive aspects of voting wholly in terms of benefits.

Stressing that voters experience satisfaction from the very act of expressing their preferences, this account has become known as the “expressive voter hypothesis” (Mueller 2003: 320-323). However, it should be clear that it stays squarely within the instrumental framework of Public Choice theory. It understands the decision to vote as aimed towards a goal external to the voting act itself, namely that of experiencing “satisfaction” (Riker & Ordeshook 1968: 28), “intrinsic benefits” (Brennan & Lomasky 1993: 61), “psychic rewards” (Overbye 1995: 372), “utility gain” (Mueller 2003: 320) or “expressive returns” (Schuessler 2000: 107).

Some authors instrumentally rationalize the decision to vote in a more indirect way. First, Einar Overbye argues that people vote in order to build a reputation that will benefit them in the long run. Since expressing their identity consists of sending signals to fellow citizens, people will take into account the possible impact of their acts on their social capital: “voting may be regarded as a rational investment decision: not an investment in a particular electoral outcome, but in a type of reputation which the individual is interested in maintaining when carrying out his/her everyday activities” (Overbye 1995: 369). Second, Schuessler (2000b: 88) argues that “voting is a means to express political beliefs

---

<sup>29</sup> Authors like Brennan, Buchanan and Lomasky stress that the electoral context is characterized by the complete absence of a connection between the individual’s action and the resulting outcome. In this respect, it cannot be understood as analogous with consumption. While citizens decide to vote irrespective of whether they actually get the outcome they prefer, consumers only pay the price if they know they will receive the desired product (Buchanan 1999: XIII, 163; Brennan & Lomasky 1989: 44). I will go into the repercussions of this thought when arguing in favor of a completely non-instrumental account of voting decisions.

and preferences and, in doing so, to establish or reaffirm their own political identity”. He distinguishes “between strict electoral-outcome-oriented rationality and expressively focused motivation which targets an outcome other than the electoral result” (Schuessler 2000b: 116). Once more, it is clear that these accounts continue to analyze voting as means towards the realization of a goal external to the act of voting itself.

#### *1.1.4. Problems with amending the calculus of voting*

Since the decision to vote has virtually no impact on the outside world, it is tempting to explain it in terms of purely psychic benefits. However, I want to claim that this does not adequately explain why so many citizens vote. In my view, there are several problems with such an instrumental rationalization of the decision to vote. Consequently, I agree with Overbye (1995: 371) who stresses that none of the proposed explanations has been able to convince the critics so far.

First, one may legitimately ask what exactly is so fun about voting. Would a hedonistic citizen not be better off enjoying his day as he usually does, rather than spending it in a queue before the voting booth? If voters thoroughly enjoy expressing their political opinions, why do they not massively join political discussion groups? In addition, a civic sense of duty simply cannot be understood in purely hedonistic terms. As moral philosophers know for quite a while now, people who feel they have to obey a duty will often do so even if they derive no satisfaction from it (Blais 2000: 93; Boudon 1997: 222). They simply do not weigh the benefits of fulfilling one’s duty (*D*) against the costs of doing so (*C*). They vote because they feel they have to, not because they like doing so<sup>30</sup>.

Second, this account does not take seriously the thought that an activity can be done purely for its own sake. Because it continues to compare the benefits and costs of voting, it basically suggests that voting is not done for its own sake (Ferejohn & Fiorina 1974: 525). If voting is valued intrinsically, it is not to be understood as a means towards some extrinsic goal, like the satisfaction derived from it. All of the abovementioned accounts lump together the reasons

---

<sup>30</sup> This objection essentially argues that “voters obviously do not consciously impute the above type of complicated calculations before they decide whether or not to vote” (Overbye 1995: 381). The typical defense is that people do not necessarily maximize their utility in a deliberate manner, but act *as if* they do (Ferejohn & Satz 1994: 76). I have already shown why I believe this strategy to be fundamentally wrongheaded if one aims to explain actions in terms of reasons that cause these actions to be performed.

for voting into a single calculus of costs and benefits. In doing so, “they obscure the distinction between acting for the sake of some expected advantage to which the action is believed to be a means, and acting for the sake of a principle or from duty – between what Max Weber called *Zweckrationalität* and *Wertrationalität*” (Benn 1979: 299). In this respect, I argue in favor of distinction between instrumental and expressive considerations that is as conceptually clear as possible.

Third, and most importantly, this account systematically posits a preference or taste for voting among voters. Simply rephrasing the issue, this comes down to a tautological explanation with no real explanatory power (Blais 2000: 9-10; Boudon 1997: 221; Mueller 2003: 3306; Overbye 1995: 372; Schuessler 2000a: 47). In the end, all it says is that citizens vote if they like to vote and abstain if they do not like to vote. Such an account does not explain the observed behavior in terms of a preference for voting, but derives the presence of the latter from the observed behavior. Systematically postulating that the expression of one’s vote gives satisfaction essentially begs the question. It is an ‘ad hoc’ hypothesis which lacks predictive content and explanatory power: “the very fact that the conventional theory is driven to rely on psychic income raises the question of whether it has anything very interesting to say about psychic income, or whether it merely treats psychic income as a residual category of motivation to be invoked when the theory otherwise gets into trouble” (Margolis 1984: 87). Since it cannot be falsified, its scientific validity is highly questionable. Psychic benefits should not be invoked as the explanatory basis, but should themselves be accounted for (Margolis 1984: 60). In order to avoid that such explanations become completely trivial, one should show why “some people have this kind of motivation more strongly than others” (Barry 1970: 16). I will try to do exactly this in the next chapter.

## **1.2. Economic rationality and the decision whom to vote for**

### *1.2.1. The calculus of voting*

To understand how Public Choice theorists explain the decision whom to vote for, one simply needs to extend the analysis presented above. Basically, citizens are expected to vote for the party or politician that they think will benefit them most. Once more, the seminal formulation can be found in Downs (1957: 36) who assumed “that each citizen casts his vote for the party he believes will provide him with more benefits than any other”. In his view, voters rank the alternatives by comparing “the different “streams of utility” to be derived from



the policies promised by each candidate” (Mueller 2003: 304). Economically rational citizens will try to maximize the benefits they expect to receive from the policy measures of the next elected government. This is essentially what Public Choice theorists have tried to capture in the *B*-term of the voting calculus.

There is some empirical support for the claim that people vote for the party that promotes their interests. On the whole, low-income employees tend to vote for a left-wing party that favors income redistribution. In contrast, high-income employees tend to vote for a right-wing party that favors tax deductions and social security cut backs. However, this so-called pocketbook voting is heavily disputed (Fiorina 1997: 407). That is why Public Choice theorists increasingly define interest in a broader sense, for example by including long-term and collective considerations. The argument then goes that citizens will vote for the party they think will help them most in realizing their goals, whatever these may be. This still fits the economic conception of rationality: “though material self-interest is often a maintained hypothesis in public choice research, the rational choice approach admits a much broader range of preferences” (Fiorina 1997: 408). Such so-called sociotropic voting – voting to realize some public good – has been confirmed by empirical studies (Mueller 2003: 460-461). A rich self-employed business man may vote for a party that favors high taxes if he cares more about the realization of some public good than about the maximization of his own income. This broader, instrumental model is able to explain an individual’s decision to vote against his own narrowly defined self-interest (Mueller 2003: 323).

### *1.2.2. Another paradox of voting?*

Here, I want to argue that proponents of the economic conception of rationality cannot really defend these views. After all, the problem remains that instrumentally motivated citizens will not even take the trouble to show up at the polling station. If one consistently applies this instrumental reasoning, the vast majority of such citizens will decide to abstain and will thus not even have to make up their minds which party or politician to vote for. The few that wrongly believe that their vote will make a difference will probably indeed vote in their interest but they will make up only a small part of the entire voting population.

In contrast, citizens who vote because of the pleasure they derive from the act of voting itself do not really care about whom to vote for. Since it is all about showing up at the polling station, having a chat with their fellow citizens and

enjoying the pleasant feeling of living up to social expectations, it is completely irrelevant for them what to do inside the voting booth. From the perspective of the economic conception of rationality, one can thus expect ordinary citizens to be rather apathetic and ignorant with regard to politics. Since their vote decides nothing, they have no incentive to become engaged in or informed about politics. Stressing the costs of acquiring politically relevant information, Downs argues that “ignorance is not the result of mere apathy; rather it stems from the great cost of obtaining enough information to exert effective influence” (Downs 1957: 253). For indifferent citizens, it may thus be rational to vote at random (Downs 1957: 274-276). In this respect, one can refer to phenomena like donkey voting (after the game in which a blindfolded child pins a tail on a donkey), invalid voting, voting for the candidate on the top left of the ballot or voting for a candidate with a funny name. Studies have shown that, depending on the nature of the electoral system, such votes may rise up to 5% of all the votes cast (Orr 2002: 575)<sup>31</sup>. However, this implies that 95% of the voters apparently have a pretty firm idea whom to vote for. The prediction that a majority of the voters does not know whom to vote for is thus falsified by both common sense and empirical research.

It has been argued that political apathy and ignorance are currently growing. This might suggest that more and more citizens start to reason in economically rational ways. It might also explain decreasing turnout levels throughout the world. Ever more citizens seem to understand that the freeriding option (abstention) is superior to political participation (voting) in any cost-benefit analysis. People no longer perceive voting as being in their self-interest, since there is nothing in it for them. If politicians do not hear my voice, the argument goes, why bother to express it? While this suggests that the economic conception of rationality is not without explanatory force, it also shows that it remains inadequate to explain the basic fact that a lot of citizens still vote in systems where voting is voluntary.

Brennan and Buchanan argue that the insight that a single vote has only negligible consequences induces citizens not to take these into account inside the voting booth (Buchanan 1999: XIII, 162-164). As Loren Lomasky (1992) argues, the specific nature of the electoral context induces citizens to vote irresponsibly: “because voters are unlikely to “get anything” from the direction of their votes, they have reason to indulge incentives that are not oriented

---

<sup>31</sup> Notice that this study was performed in Australia, where attendance at elections is compulsory there. Such an institutional measure raises turnout by increasing the costs of abstention. This tends to prove an efficient way of solving the freeriding problem that arises if a lot of people act in economically rational ways. In doing so, however, it is likely to draw indifferent citizens to the polls who might otherwise had chosen to abstain. In systems of voluntary voting, therefore, the number of donkey votes will be even smaller.

toward outcomes (...). Since how I vote is virtually certain not to make any difference whatsoever to political outcomes, I'm freed to indulge any whimsical, voyeuristic, half-baked impulse that happens to come into my head". Since the causal link between an individual's decision and the final outcome is mediated by an aggregation procedure, he is less likely to think about the consequences of his choice. After all, the outcome will arise, regardless whether he contributes to it or not: "the single individual loses the sense of decision-making responsibility that is inherent in private choice. Secure in the knowledge that, regardless of his own action, social or collective decisions affecting him will be made, the individual is offered a greater opportunity either to abstain altogether from making a positive choice or to choose without having considered the alternatives carefully" (Buchanan 1999: III, 38).

As I have already suggested, the basic problem with this hypothesis of uninterested and ignorant voters is that it assumes that people do turn up to vote. However, if they actually are instrumentally motivated, they will not take the trouble of doing so (Margolis 1984: 65). For now, suffice to say that the same line of reasoning opens up a whole range of possible motivations. The fact that people cannot benefit themselves in an electoral context might as well induce them to act on nobler motives. As I will show more fully in the next chapter, it may as well lead them to base their decision on purely principled considerations (Mueller 2003: 322-325).

There is something paradoxical about the way Public Choice theorists try to explain how people vote. The main difficulty they face is that the problem of explaining turnout comes back with full force. To find a way out of this conundrum, Public Choice theory often relies on a combination of both lines of thought, arguing that people vote because they like doing so (*D*), but change gears when deciding whom to vote for. Once inside the voting booth, they vote in accordance with their interests (*B*), bearing in mind that one never knows whether one's vote might be decisive after all (*P*). Even though it will probably make no difference, I will cast my vote in my interest, because it might just help and will never hurt to achieve the electoral outcome I prefer (Aldrich 1997: 375-376).

Public Choice accounts of the decision whether to vote thus typically refer to the intrinsic benefits voting yields, while Public Choice accounts of the decision how to vote typically refer to its instrumental benefits. As Mueller (2003: 329) rightly stresses, the view that "some individuals get utility from expressing their preferences for a particular candidate through the act of voting (...) provides an explanation for *why* a person votes, but not for how she votes". Similarly, Merrill and Grofman (1999: 164) present support for Public

Choice accounts of how people vote, while explicitly abstracting from the question of why people vote.

This division of tasks suggests that there is no theory that aims to explain both the decision whether or not to vote and the decision whom to vote for on the basis of the same reasoning. Most explanations why citizens vote refer to psychological characteristics such as political interest, political efficacy and a sense of duty. In contrast, most explanations of how they vote refer to completely different factors such as self-interest, candidate qualities, campaign issues and party identification (Fiorina 1976: 391). Relying on two wholly different models, Public Choice theory goes against its very own assumption of behavioral symmetry, according to which individuals typically act the same in different spheres of life (Buchanan 1999: X, 56-57). Whether they are making private or public decisions does not matter; they should be analyzed as *Homines Economici* in every domain of life. After all, the individual remains essentially “the same man in the supermarket and in the voting booth” (Tullock 1976: 5). As I will show more fully in my analysis of James Buchanan’s work, Public Choice theorists heavily criticize the tendency to postulate different motivations for different decisions. Nevertheless, this is exactly what they are forced to do here. Their attempt to consistently apply the *Homo Economicus* model to different voting decisions thus fails.

## 2. Reconciling rationality and democracy

It can be argued that political philosophers face a huge dilemma if they – like myself – value both the rationality of individuals and the democratic nature of institutions. After all, both seem to be at odds with each other. While the participation of a vast majority of individuals is needed for a democracy to function well, it seems to be irrational to participate in mass political action: “where goals are at stake that require the participation of large numbers of people, any given individual’s participation is going to make no perceptible difference to the outcome” (Benn 1979: 292). This may lead one to conclude that a democracy of rational individuals could never survive. The dilemma that Public Choice theorists face is clear. If they stress the importance of the economic rationality of people, they seem to neglect the importance of democratic politics. And if they stress the value of the latter, they seem to give up their basic model of individuals.

There are two ways out of this conundrum. First, one can conclude that participation in a political is indeed irrational. However, this effectively degrades one’s valuation of both democratic politics. Here, one can refer to

Buchanan (1999: I, 85) who argues that “market choice tends to embody greater rationality *in individual behavior* than does voting choice”. As I have shown, the connection between individual preferences and social outcomes – which is present in market settings but absent in electoral contexts – leads him to favor the market over the state. Second, and this is what I will argue for in the next chapter, one can try to show that political participation can be understood as rational in a different sense than the conventional one. In this respect, I want to show in what ways amending the most problematic assumptions of the economic conception of rationality helps solve the paradox of voting. This way, it becomes possible to reconcile both rationality and democracy.

The paradox of voting crops up when one tries to explain the decision to vote in an exclusively instrumental framework: “the voting paradox arises not because there is a problem when people vote, but because of the flawed notion of rationality held by public choice” (Pressman 2006: 251). Providing a fuller, deeper and more adequate view of the way in which citizens make voting decisions, I believe alternative conceptions of rationality make the paradox vanish into thin air. They allow one to explain voting decisions as rational (the voter judges there to be good reasons for his decisions) without interpreting them as instrumentally motivated. Even though I will rely on suggestions from Fiorina, Brennan, Lomasky and Schuessler, I will not follow their strategy of reducing expressive aspects of voting to instrumental benefits.

If one drops the assumption that people are instrumentally motivated to satisfy their egoistic and exogenous preferences, it becomes clear that voting is not a paradox but rather a rational response to the specific context of large-scaled elections. While it is rational to try and fulfill one’s preferences in situations where this is possible, it is rational simply to express them whenever one cannot have a substantial impact on the final outcome. As I will show more fully in the conclusion to this second part, this does not imply that the economic rationality is superfluous or worthless.



## CHAPTER 4

### ALTERNATIVE CONCEPTIONS OF RATIONALITY AND VOTING DECISIONS

---

*“Rational choice modellers will be able to explain voting and nonvoting as soon as they solve the problem of why people salute the flag when they know that nobody is looking”*

(Kramer, in: Grofman 1983: 57)

#### 1. Minimal rationality and voting decisions

The economic conception of rationality fails to grasp why and how people make voting decisions the way they do. It analyzes voting decisions as investment or consumption acts. While the first model predicts abstention ( $PB < C$ ), the second stresses that people vote for the pleasure of doing so ( $D > C$ ) without actually caring about the electoral result. Because both do not correspond to one's intuitive image of voters, I want to stress the relevance of alternative conceptions of rationality. This way, I avoid banning voting decisions “to the mysterious and inexplicable world of the irrational” (Riker & Ordeshook 1968: 25). As I have shown, I believe voters should be called rational in its minimal sense as long as they act on the basis of what they believe to be good reasons.

##### 1.1. Minimal rationality and the decision whether or not to vote

Since the paradox of voting arises from the instrumental character of the Homo Economicus model, it is crucial to drop exactly this assumption when trying to explain voting decisions (Aldrich 1997: 374; Brennan & Hamlin 1998: 167; Ferejohn & Fiorina 1974: 535; Schuessler 2000a: 6-8). The minimal conception of rationality broadens the range of possible reasons that motivate people to vote. I want to argue that people can have reasons to go out and vote that cannot be reduced to goals: “even without an extensive psychological examination of voters, there is some empirical evidence that desire to affect

outcomes is not the only or primary motive for voting (...). We maintain that voters are not predominantly irrational, and thus they vote as they do for reasons *that have little to do with an intention to affect outcomes*” (Brennan & Lomasky 1989: 49). In what follows, I will go into the nature of such reasons more fully. Here, I only want to broaden the notion of rationality beyond its purely instrumental meaning.

### 1.2. Minimal rationality and the decision whom to vote for

The same reasoning can be applied to the decision whom to vote for. There is a wide range of reasons that motivate citizens to vote as they do. The economic conception of rationality focuses exclusively on instrumental considerations that hinge upon the electoral outcome. It assumes that voters base their decisions exclusively on the expected policy outcomes of the electoral process (Brennan & Hamlin 1998: 164-165). However, people vote on the basis of a much wider range of considerations: “the voter may identify with the candidate’s moral character, good looks or ethnic origin or with the candidate’s or party’s general ideology” (Brennan & Hamlin 1998: 156). Such aspects, which clearly come to the fore in surveys and polls, cannot be incorporated within a purely instrumental framework. Later on, I will show that an expressive account performs much better in this respect.

## 2. Broad rationality and voting decisions

The broad conception of rationality focuses on the ways in which people’s reasons for voting decisions are formed. It stresses that beliefs should be well-informed and preferences should be autonomous in order to be broadly rational. As I have shown earlier on, it is useful to focus on the cases where voting decisions are clearly irrational in the broad sense of the word.

### 2.1. Broad rationality and the decision whether or not to vote

First, one can argue that people who vote because they believe they actually have an impact on the electoral outcome are broadly irrational. After all, their estimate of this impact ( $P$ ) blatantly disregards common sense surrounding statistics. It may be the case that people think their vote is crucial because they want it to be so, but such a case of wishful thinking is clearly irrational. As I



have shown, empirical research shows that only a minority of citizens overestimate the impact of their vote in such a degree that it motivates them to go out and vote. The underlying reasoning, which is based on the plausible assumption that others think like they do, often goes as follows. If I decide to abstain, others are likely to do so as well, thereby contributing to the downfall of the democratic system as such. However, if I decide to vote, others are likely to do so as well, thereby upholding democracy. Here, I want to leave open for interpretation whether this line of reasoning is best framed as increasing the expected impact of one's vote (*P*) or as raising the stakes of one's vote (*B*). What matters is that, while such beliefs make it minimally rational to go out and vote, they clearly go against widely available information surrounding the choice situation at hand. Labeling them broadly irrational makes explicit the intuitive thought that there is something wrong with this kind of reasoning.

Second, one can argue that preferences for voting are sometimes heteronomous in that they are formed through indoctrination or intimidation. In dictatorial regimes, for example, citizens go out and vote because they are heavily influenced by propaganda.

Third, one could argue that a preference for voting is broadly irrational if it does not arise from rational deliberation. Marquette and Hinckley (1988: 59) empirically show that the most important factor to explain why people vote is their remembrance of their previous decision: "the best predictor of turnout is previous voting history". Empirical results do indeed show that voting often "is an unreflective and habitual act" (Blais & Young 1999: 53). However, this does not mean that it is completely arational or irrational, since one can have good reasons to vote, even if these are not fully and consciously deliberated.

The assumption of exogenously given preferences forms an important limitation of Public Choice theory. Samuel Bowles and Herbert Gintis (1986: 138-139) for example, explain the decision to vote by amending this assumption: "the most fruitful method of explaining such behavior, we maintain, involves abandoning the instrumental conception of action itself and in particular jettisoning the notion of preconstituted or exogenously given preferences (...). One does not merely register one's preferences in (...) voting; rather, one constitutes and reaffirms oneself as (...) a good citizen". The fact that preferences can change endogenously is thus crucial when trying to explain voting decisions (Mueller 2003: 324-332).

## 2.2. Broad rationality and the decision whom to vote for

A lot of voters vote as they always have done, which is heavily influenced by their upbringing, education and peers. This is already intuitively plausible: if one grows up in a worker's family with a father who is a trade unionist and one's friends all criticize libertarian attempts to cut back pension schemes, one is simply more likely to vote left-wing. While it is not always fully deliberated, all this can provide one with good reasons to vote as one does.

Here too, it is useful to focus 'ex negativo' on broadly irrational beliefs and preferences. The former are ill-informed in that they go against available evidence. If people vote on a socialist party because they believe this will help reduce state intervention, they can be called broadly irrational. The same conclusion holds if they are indoctrinated by socialist propaganda.

The fact that preferences can change endogenously can be relevant in explaining the decision whom to vote for as well. By voting socialist, people (hope to) become a good socialist. While a possible impact on their identity is actually a by-product, this aspect cannot be ignored if one wants a full account of voting decisions. As I have stressed earlier on, however, such a conscious formation of one's reasons is not required to speak of broad rationality, since it cannot be achieved in a deliberate and intentional way.

## 3. Expressive rationality and voting decisions

Whereas the broad conception of rationality focuses on the formation process of beliefs and preferences, the expressive conception stresses that the individual at hand should identify with his reasons and regard them as authentically his own. Whether his reasons stem from socialization (education, family, friends), indoctrination (media, propaganda) or rational deliberation is largely irrelevant. What matters is that there is a certain level of psychological and emotional involvement: "the motivation that guides individuals' participation in these activities is one of expressive attachment: through their participation, these voters (...) express who they are, and they attach to a collective that they feel is like them" (Schuessler 2000a: ix).

In contrast with the abovementioned expressive accounts of voting that remain within an instrumental framework, I want to provide a completely non-instrumental conception of expressive rationality. In my view, voting can be understood as expressively rational if it is "undertaken for non-consequentialist

reasons” (Benn 1979: 304). An expressively rational voter simply wants to express what kind of a person he is and what he values highly in life: “individuals do not necessarily participate in collective action in order to produce outcomes, but instead often do so in order to express who they are” (Schuessler 2000a: 5). In what follows, I first spell out more fully why such a non-instrumental characterization of expressive considerations is needed to explain voting decisions as most rational citizens make them.

### **3.1. Why an expressive account is needed to solve the paradox of voting**

As Barry (1970: 16) stresses, one should not simply postulate a taste or preference for voting if one really wants to explain why people vote, but find out why some citizens have this motivation more strongly than others. This search for “deeper reasons” (Dowding 2005: 453) is needed to avoid the ‘ad hoc’ character of Public Choice theories amending the calculus of voting. In what follows, I will therefore analyze some of the variables that underlie an individual’s motivation to vote, like social expectations, moral principles and democratic norms according to which voting is a good thing to do. In contrast with postulating that one votes because one prefers to, there are quite a lot of strings attached to the claim that one votes because one cares about these things. After all, such cares typically guide more decisions than those concerning whether and how to vote.

In this respect, I want to argue that it is rational in an electoral context to express one’s identity, regardless of the consequences of doing so. After all, it is a way of bringing one’s actions in accordance with one’s cares (Boudon 1997: 222; Schuessler 2000a: ix-x; Schuessler 2000b: 88). The purpose of voting is simply to express one’s identity, nothing more (one does not want to reach some goal in the outside world) and nothing less (one does not mindlessly react to environmental stimuli). As Amartya Sen (1977: 333) notices, voters may be “guided not so much by maximization of expected utility, but something much simpler, viz, just a desire to record one’s true preference”.

The main argument is that the electoral context induces rational people to base their decisions on non-instrumental considerations, because it breaks any connection whatsoever between the individual’s action and the resulting outcome. Whereas private choices have a direct impact on my situation (and mine alone), public choices have almost no impact on my well-being. Whereas it is rational to base the former primarily on self-interested considerations, it is rational to base the latter on other considerations. However, this dichotomy should not be interpreted too strictly (Buchanan 1999: I, 66-70; III, 17-30). The

difference between both is gradual rather than absolute in nature. While people typically take into account all sorts of considerations, their relative weight depends on the choice situation at hand.

While authors like Buchanan, Brennan and Lomasky rightly stress that this difference between private and public choices provides the decisive reason to go beyond conventional economics in explaining human actions, I believe they do not go far enough (Engelen 2008a). After all, the electoral context provides a setting where rational individuals will not act instrumentally because “they don’t have reason to care greatly about the consequences of their acts” (Ferejohn & Satz 1996: 78).

In such a setting, rationality does not require standard instrumental cost-benefit analysis (Brennan & Lomasky 1993: 21, 30). Rather, it recommends one to discount the possible consequences of one’s vote. Since a potential voter knows that he will not necessarily get what he chooses, he will not, in the first place, vote in order to get what he wants. If he knows that he may just as well not get what he votes for, it is rational for him not to vote in order to get it. I thus want to assume that citizens in general realize that – whatever they decide to do – they will have no significant impact on the outcome (Brennan & Lomasky 1993: 171; Mueller 2003: 319).

### **3.2. Expressive rationality and the decision whether or not to vote**

If most citizens understand that they cannot influence the electoral outcome – as I believe they do – they will base their decision whether or not to vote on other than instrumental considerations. In this respect, I want to argue that “turnout should be considered a decision, a perfectly rational decision, to express one’s preferences” (Aldrich 1997: 390).

To make clear the difference between instrumental and expressive considerations in favor of voting, I want to refer to a number of other distinctions. First, instrumental considerations always refer to the future consequences that individual citizens want to achieve: I vote because I think my vote will influence policy goals as I prefer them (*PB*) or because I will experience satisfaction from the act of voting itself (*D*). In contrast, expressive considerations refer to existing norms, traditions and values to which citizens are committed. In this respect, voting is not so much prospective as it is retrospective in nature. Whereas instrumental actions are thus typically forward-looking (prospective), expressive actions are predominantly backward-looking (retrospective). As I have already suggested, a lot of citizens vote

because they have always done so in the past<sup>32</sup>. Second, whereas instrumental actions are hypothetical in nature (I vote if I want to achieve my goals), expressive actions are typically categorical (I vote, full stop). Third, whereas instrumental actions are valued and performed for the sake of their consequences, expressive actions are valued and performed regardless of their consequences (I vote for the sake of voting itself). Of course, these distinctions tend to overlap each other to a certain extent.

The decision whether or not to vote depends upon the commitments, engagements and loyalties that constitute a person's identity. One might argue that social expectations and the worries to live up to conventions can hardly be labeled fundamental things to care about. For such expectations to have motivational force, however, an individual has to identify himself with a certain community and its accompanying values, norms and traditions. There are several interrelated ways to rephrase these general statements. In doing so, it will become clear that the expressive account is more fundamental than and cannot be reduced to simply positing a preference for voting.

First, a large part of the electorate cares about democracy itself. Downs (1957: 267) stresses that this can provide an incentive to vote. However, whereas Downs models this motivation as a rational consideration about the long-term benefits of democracy (hypothetical)<sup>33</sup>, I want to interpret this as a sense of social responsibility (categorical). André Blais (2000: 143) formulates it aptly: "like many of my fellow citizens, I feel that I must act in accordance with the principle I believe in. As I think of myself as a democrat, it would be incongruous not to vote. I vote, then, because I want to be consistent with my principles". A lot of voters think of themselves as good citizens and want to express this part of their identity. Elections provide an excellent opportunity to do exactly this.

Second, a lot of citizens experience a strong sense of a civic duty to vote. About half of the electorate decides to vote "out of a sense of moral obligation, because they believe it would be wrong not to vote; they do not calculate benefits and costs" (Blais 2000: 137). Empirical research shows that such a sense of "duty is the overriding motivation for about half of those who vote and a clear majority of regular voters. Those with a strong sense of duty almost always vote" (Blais 2000: 112). Indeed, the number of citizens that decide to

---

<sup>32</sup> This does not refer to the claim that voting decisions are primarily based on retrospective judgments of party performance, which is situated wholly within the instrumental account in which decisions are completely goal-oriented.

<sup>33</sup> Downs thinks that citizens would in the end give priority to the short-term advantages of abstention over the long-term advantages of upholding democracy, which is why his instrumental model predicts (near-)universal abstention.

vote range from 13% among those with a low sense of duty to 85% of those with high sense of duty (Campbell et al. 1960: 105-106). The former tend to give greater weight to instrumental cost-benefit comparisons: “*B*, *P*, and *C* do a much better job of explaining the vote among those with a weak sense of duty” (Blais 2000: 102).

As I have shown, most authors try to incorporate this sense of duty in the calculus of voting by adding a *D*-term. In doing so, they have to assume that people systematically experience satisfaction from fulfilling their duties. Here, I want to argue that the basic reason for living up to one’s obligations is not the pleasure one may derive from it. Whatever the costs, benefits and consequences of voting are, they do not form the primary reason for performing it. Instead of instrumentally rationalizing this sense of duty, I simply want to take it at face value. Acting in accordance with one’s principles is perfectly rational as long as one identifies with them and takes them to constitute good reasons to act upon (Boudon 1997: 223-224). In this respect, I want to understand the civic sense of duty as “the main reason to vote” (Blais 2000: 104-112). Even though it is not necessarily instrumentally rational, it perfectly fits the requirements of expressive rationality.

Third, voting is often deemed intrinsically valuable. It is not perceived as a means to produce some outcome, but as an act of which the performance itself counts. It is the action itself rather than its consequences that matters. Voters are mainly “oriented toward purely expressive or symbolic action, action that is undertaken for its own sake rather than to bring about particular consequences” (Brennan & Lomasky 1993: 25). Here, it is important to distinguish between voting as intrinsically valuable and voting as intrinsically pleasurable. As I have already stressed, I believe that the abovementioned accounts systematically incorporate the former within the Public Choice framework of the latter. Mostly, this is done by stating that “evidently voters take some psychic satisfaction in expressing their preferences” (Fiorina 1997: 403).

Fourth, a lot of citizens care about a particular ideology, party or politician. Those who consider themselves to be good republicans want to express this by going out to vote on a republican party or politician. This does not mean that they think their vote will contribute to its victory. It is a symbolic act, expressing the way they see themselves. The stronger this so-called sense of party identification, the more likely it is that one will go out to vote. Because this also addresses the question how one votes, I will go deeper into it in what follows.

### 3.3. Expressive rationality and the decision whom to vote for

The argument that the electoral context makes it rational to act on the basis of non-instrumental considerations applies to decisions inside the voting booth as well: “in the market the agent is *decisive* (...). The chooser actually gets what he chooses (...). At the ballot box, in particular contrast, the agent is nondecisive (...). Whether option *A* or option *B* actually emerges as the electoral outcome is a matter not of how I vote, but of how everyone else does. Electoral outcome is *detached* from electoral ‘choice’ for each voter in a crucial way” (Brennan & Lomasky 1993: 15).

In contrast with what Public Choice theorists are forced to assume and as I will explain more fully later on, I want to argue that the considerations that motivate the decisions whether and how to vote are basically the same. If one decides to turn up because of expressive considerations, these will also dominate the decision whom to vote for. As I have suggested, a strong sense of party identification provides an incentive both to show up and to vote in a specific way. Empirical research has shown that this “sense of attachment with one party or the other” (Campbell et al., 1960: 121) is among the most important factors in explaining how people vote. If people identify with a particular party or candidate, it is rational to express their support for it by checking its box on the ballot. Moreover, citizens can have an affectionate attitude towards a particular candidate. This is not only the case when they know a politician personally, but also when a politician holds beliefs, values and ideologies similar to their own. In such cases, it is expressively rational for citizens to express support for these politicians and the things they stand for.

One could argue that this line of reasoning does not apply to citizens who regard voting as intrinsically valuable or who experience a sense of duty to vote (Blais 2000: 111). Since they are mainly concerned with the act of voting itself, they have no clear preference about whom to vote for. While this may be an accurate description of some voters, I think it is too far-fetched to form a generalized account. The fact that they care about democratic values suggests that they are not so indifferent that they will vote completely at random (Fiorina 1997: 403). Even though their reasons for showing up do not provide a straightforward manual to guide their decision how to vote, they at least indicate a direction. Their impact inside the voting booth goes beyond the calculus of the potential benefits of one’s vote, for example by eliminating some candidates as inconsistent with their basic commitments.

### 3.4. The distinction between the decisions why and how to vote

While not always made explicit, the distinction between the decisions why and how to vote is central in Public Choice theories. In the end, I believe that this distinction inhibits a proper account of the decision process that most rational citizens go through and thus tends to confound matters only further (Campbell et al. 1960: 89). In my view, it has contributed to the renowned paradox of voting. An instrumentally motivated citizen will indeed first calculate whether or not going out to vote is worth it. Only if voting is deemed beneficial, does he have to make up his mind whom to vote for. In contrast, I want to argue that the decision how to vote often precedes the one whether or not to vote. Expressive voters often know how to vote before deciding whether or not to vote. A devoted republican will vote republican, without first considering in some distinct moment of choice whether or not to vote, simply because he considers it a unique opportunity to express this aspect of his identity. His decision to vote is secondary to his decision whom to vote for. Even if he actually ponders whether or not to vote, his mind is already made up about whom to vote for.

If one focuses on the expressive aspects that guide voting decisions, it becomes clear that the distinction between whether and how to vote is often irrelevant or even misleading. Most of the time, both issues do not present themselves independently from each other (Brennan & Hamlin 1998: 172). The expressive conception of rationality thus allows one to analyze both decisions as parts of the same overall reasoning: “turnout and candidate choice are not necessarily two separate decisions, but rather a joint decision based on the same sorts of factors” (Marquette & Hinckley 1988: 57). Among these factors are one’s party identification and one’s perception of differences between the candidates. As these fade, more and more citizens feel that it does not really matter who gets elected and start to question whether voting is worth the effort at all (Aldrich 1997: 386). The fact that people without a clear idea whom to vote for tend to abstain more often shows that the decision whether and how to vote are strongly related<sup>34</sup>. In my view, stressing the importance of the issue of how citizens vote (partly) solves the paradox of why they vote.

---

<sup>34</sup> Downs formulates this thought in purely instrumental terms: “men who are indifferent about who wins have nothing to gain from voting, so they abstain” (Downs 1957: 262). He argues that there should be a considerable difference between the utility they will experience if different parties gain power. If this so-called “expected party differential” (Downs 1957: 39) is zero, they will rationally decide to abstain. In contrast with Downs’ claim that “rational men are not interested in policies *per se* but in their own utility



### 3.5. Expressive rationality and cheering

To further illustrate the expressive rationality of voting, it can be compared to cheering for one's favorite sports team (Mueller 2003: 320-32). Cheering, like voting, is intended not to influence the outcome of the contest, but to express support for one of the competitors. As Brennan and Buchanan put it, "although spectators care about the outcome, they do not act to determine it, and they do not conceive themselves to be so acting" (Buchanan 1999: XIII, 154). Like voters, they are aware that their actions do not influence the outcome. Consequently, it is rational for them to put aside cost-benefit calculations and act in non-instrumental ways.

Here too, Public Choice theorists hurry to argue that cheering is instrumentally rational, because it is a cost-effective way of acting according to one's preferences: "the actor obtains personal pleasure from the act (...). A fan's cheering is rewarded if his team wins; most fans cheer for the home team. Winning home teams provide more positive reinforcement for their supporters. Winning home teams tend to have higher attendance levels and more vocal fans than do losing teams" (Mueller 2003: 328). There is thus no fun if there is no cheering: "the sports fan's expression of team support is required for him to enjoy his participation" (Schuessler 2000a: 46). Cheering is perceived not as a cost, but as a benefit, since it is agreeable in itself. This line of reasoning corresponds to the abovementioned strategy of instrumentally rationalizing voting decisions: "expressing support for an outcome can be much less costly than actually bringing about that outcome (...). In a democracy, one important way in which one can express support for an outcome is by voting for it. When the number of voters is even moderately large, voting is a low cost way to give vent to sentiments that would, in other contexts, lie dormant" (Brennan & Lomasky 1989: 51).

In contrast, I want to argue that cheering, like voting, does not always bring about a sense of joy. Fans do not cheer because it brings them cheer. I thus want to go against the tendency to reduce the act of cheering (a cheer as an encouragement for one's favorite team) to the pleasure this may produce (a cheer as a happy feeling). In my opinion, this attempt to explain cheering instrumentally leads to the paradoxical prediction that nobody will cheer for a losing team. It cannot explain why fans continue to cheer if their favorite team is losing or why some people tend to support the underdog. The bottom line is that people do not cheer in order to attain, but to express the pleasure they feel

---

incomes" (Downs 1957: 42), I want to understand all this in ideological and thus expressive terms.

when their team wins. This basically comes down to the point I have made previously, namely that systematically postulating that individuals enjoy expressing their preferences does not explain anything. Not only does it provide a tautological and circular account immune to empirical criticism (I cheer because I like to cheer and I like to cheer because I enjoy doing so), it is also counterintuitive. People cheer because they care about their team, regardless whether this brings pleasure or frustration.

To be sure, I do not claim that cheering and voting are the same, only that they are analogous. Voting is not cheering, it is *like* cheering. While this analogy – like every analogy – fails in some respects, it shows that both actions have something in common, namely that they both belong to the category of expressive actions. According to Pressman (2006: 244-246), the analogy fails because cheering is a secondary act which costs almost nothing, since people attending a match primarily pay to enjoy the match itself. In contrast, voters take time and trouble to vote, even though this has no purpose beyond supporting their favorite candidate. However, it is perfectly possible that fans may not enjoy the match at all, just as much as voters may enjoy the act of voting itself. In addition, contrary to what Pressman (2006: 249) thinks, drawing an analogy between cheering and voting does not imply that an increase in the first automatically implies an increase in the latter as well. If people's identities shift, the actions that express them will change accordingly. This perfectly accounts for the recent increase in all sorts of cheering behavior occurring simultaneously with falling turnout levels. One could argue that people increasingly construct their identities on the basis of personal commitments, without referring to political ideologies. In this light, the covariance between more cheering and less voting can be explained by a general shift in the way citizens conceive of themselves. While personal commitments may alter, the tendency to express them does not seem to be declining or increasing over time.

In this respect, I want to refer to this chapter's opening quote by Gerald Kramer. He seems to suggest that voting – like cheering or saluting the flag on one's own – is close to absurdity or at least irrationality<sup>35</sup>. In contrast with Kramer, I believe that the remarkable similarities between cheering and voting show that both are rational responses to a particular kind of situation. Both are ways of expressing particular aspects of one's identity. If one cares about some team, one will express this by applauding at their performances. Similarly, if one cares about a specific political candidate, one will express this by going out

---

<sup>35</sup> While Kramer focuses explicitly on the private nature of both actions (nobody sees whether or not one really votes inside the voting booth), I believe that going out to vote is a very visible act.

to vote for him. I think it is not necessarily irrational to express one's loyalty to a team, a politician or a political party even if nobody is watching (Schuessler 2000a: 15). After all, cheering on one's own – which is not as unusual as Kramer may think it is (Buchanan 1999: XIII, 154) – is not aimed at some goal external to the act itself, even not the goal of letting others know what one stands for. Fans cheer simply to express their involvement and honor certain fundamental aspects of their identity<sup>36</sup>. The analogy thus brings to the fore that both voting and cheering are based on good reasons that cannot be understood in instrumental terms.

The analogy also further clarifies that I do not randomly shift from an instrumental to an expressive conception of rationality. Situations that induce people to cheer for their favorite sports team are characterized by the impossibility of contributing anything to the event itself. Jumping on the field during a sports match is impossible, inappropriate and probably grossly inefficient. After all, one can expect the athletes to do better than oneself. Because fans are aware that they cannot influence the outcome of the match, it is rational for them not to try and do so. The same holds for expressive voters. Because elections do not allow them to promote their values (they cannot ensure the victory of their favorite party or candidate on their own), they have to search for other ways to bring their actions in accordance with their principles and commitments (Aldrich 1997: 390).

In this respect, I want to argue that honoring and promoting are two possible ways of translating one's cares into action (Pettit 1991: 230-231). A citizen who votes is honoring what he cares about rather than promoting it. After all, the latter would require him to try and persuade two other citizens to vote in the way he does. An attitude of honoring means that, while personally committed to certain things, he does not necessarily want others to uphold them as well. Since he is satisfied with living by his own standards and does not want to impose them on others, he lets his own actions exemplify his values. If a citizen who cares about democracy would not participate in elections, he risks being charged with hypocrisy and not having the concerns he claims to have. Expressing one's principles is often the only way of remaining true to oneself, even if it is not the most efficient means to promote one's cares.

The analogy also further clarifies that the distinction between whether and how to vote is often irrelevant from an expressive point of view. After all, a fan does not first make up his mind whether or not to go to the game (1), after which he has to decide whether to cheer or not (2) and for which team to cheer

---

<sup>36</sup> The fact that fans may cheer louder when watching a game with others can be made consistent with the expressive conception of rationality, because the tendency to manifest one's identity is typically larger if there is a public to interpret one's actions.

(3). Rather, it is because he is involved with a particular team (3) that he will cheer to express his support (2) and that he may even take the trouble of going to the game (1). Likewise, expressive voters have already acquired a certain party identification (3), which they express by supporting their favorite party (2), even before having decided whether or not to go out and vote (1).

The fact that voting decisions are based on good, non-instrumental reasons thus shows that they can be perfectly rational in the expressive sense of the word. In this respect, I completely agree with Brennan and Hamlin (1998: 167) whose “first and most basic point to stress is that voters are rational: whether they vote instrumentally or expressively in any particular situation, they do so as a rational response to that situation”. Perhaps an explanation of voting decisions as expressions of the things citizens care about is not a very sophisticated one. However, this poses no serious problem, since these decisions themselves are not that complex either (Aldrich 1997: 390).

## CONCLUSION OF PART II

---

*“The elements of the calculus of voting play only at the margin, and only among a fraction of the electorate”*

(Blais 2000: 139)

As I have suggested, I believe that the expressive conception of rationality is able to bring together both rationality and democracy without degrading either one of them. In my view, it is pre-eminently suitable to understand the importance of democracy and citizen participation. It connects rationality to one's deeply-rooted identity and shows that voting is a profound expression of the fundamental values, principles and commitments that constitute one's identity. To be sure, the different conceptions of rationality complement rather than substitute each other. In what follows, I want to analyze more fully how this comes to the fore when studying voting decisions.

First, it may happen that one citizen votes out of expressive considerations while another abstains because of instrumental considerations. Since both have good reasons for their acts, they can both be called rational in a minimal sense. However, one can still call the former instrumentally irrational. Moreover, if the latter actually cares about democracy, he is expressively irrational if he abstains. Second, a single individual typically regards both expressive and instrumental considerations as legitimate reasons on which he can decide to act (Blais 2000: 126; Fiorina 1976: 393). My portrayal of citizens who are motivated either purely instrumentally or purely expressively refers, of course, to ideal-type descriptions that form the ends of a continuum of citizens with mixed motives. Each citizen thus has to weigh off his inclination to act upon his principles against the costs of doing so. Whereas some decide to act in a principled way, others tend to let the costs dominate more easily.

In order to further illustrate the relation between instrumental and expressive considerations in voting decisions, I want to analyze two specific phenomena. Both seem to suggest that a number of voters overestimate the influence of their vote on the electoral outcome and will base their decision to vote at least partly on instrumental considerations. First, there is the empirical fact that turnout is larger in close elections (Mueller 2003: 314-318). Second, there is the phenomenon of strategic voting, which means that “a voter does not vote his or her true preferences in the hopes that voting for one's second or third

choice will yield better results” (Pressman 2006: 245). Most citizens do not like to see their vote wasted<sup>37</sup> and therefore narrow down their options to candidates who have a shot at winning the election. They may decide to vote for a less preferred candidate who has more chance of winning the election than their most preferred candidate. Someone who, for example, favors Ralph Nader might decide to vote for John Kerry, because he believes that Nader has no chance of winning at all and he deems Kerry a better option than Bush. Driven by a concern about the electoral outcome, these voters are largely instrumentally motivated (Mueller 2003: 321; Pressman 2006: 249). Downs (1957: 145) stresses that it is economically rational to vote in favor of a candidate who has a shot at winning the election: “we define rational behavior in elections as that most efficiently designed to select the government a voter most desires from among those candidates with reasonable chances of actually governing”.

It should be noted that increased turnout and strategic voting in close elections occur only at the margins. The bulk of the electorate votes sincerely and shows up at the polling station even when the electoral result is largely known beforehand. Voting sincerely for a fringe party that has no reasonable chance of winning and going out to vote even if elections are not close are ways of expressing the ideologies, principles and values one believes in. In my view, both phenomena show that instrumental and expressive considerations come into conflict with each other. On the one hand, people tend to vote for the candidate they consider to be closest to their views and values, which shows that they want to express what they care about most (Brennan & Hamlin 1998: 156). On the other hand, they also realize that their vote will be a lost one if there will be a landslide victory or if their preferred candidate has no chance at all. Ultimately, they thus have to weigh off both sorts of considerations against each other.

While more work should be done to develop an empirically supported theory about the relative weight of instrumental and expressive considerations, I have argued that – for most citizens – the former tend to be outweighed by the latter in elections, since individual choices here have only a negligible impact on the final outcome. This implies that instrumental considerations only come into play at the margin (Blais 2000: 10, 81, 137; Brennan & Lomasky 1989: 46; Brennan & Lomasky 1993: 65; Dowding 2005: 443-446). As this conclusion’s opening quote by André Blais suggests, the instrumental conception of rationality has only limited explanatory power, since it works only for those

---

<sup>37</sup> This is also what Ferejohn and Fiorina (1974) pointed at in their minimax regret explanation of the decision to vote (Aldrich 1997: 378-381). In their view, voting is the best way to prevent the regret of having abstained.

citizens with only a weak sense of duty to vote. This is supported by empirical evidence that aggregate turnout levels are only marginally affected by changes in the external conditions on election days, like the weather. Citizens who experience a strong sense of duty tend to vote, regardless of the pleasure, satisfaction or psychic benefit this might provide (Blais 2000: 137). Instead of framing such a duty as part of their overall utility function, I have analyzed it as a non-instrumental reason to vote.

My emphasis on the specific nature of the electoral context ensures that expressive rationality does not function as an ‘ad hoc’ explanation whenever instrumental rationality is deemed inadequate. It stresses that the relative weight of expressive and instrumental considerations depends on the institutional context: “if we imagine a spectrum running from the case in which the chooser is decisive through cases in which the chances of his being decisive are increasingly remote, then the role of expressive relative to instrumental elements in preference revelation increases along that spectrum” (Brennan & Lomasky 1993: 26).





# PART III

## THE RATIONALITY OF INSTITUTIONAL DESIGN



## INTRODUCTION TO PART III

---

*“Perhaps the two most central challenges in social and political theory are: first, to identify the sorts of public institutions that have best claim to be regarded as desirable; and, second, to demonstrate that those institutions are feasible, showing how they can be introduced and maintained among ordinary, unsaintly human beings”*

(Pettit 2002: 275)

In the third and last part of this dissertation, I want to investigate what the normative implications are of the different conceptions of rationality at the institutional level. What does the economic conception of rationality lead to when thinking about the ideal institutional structure of modern societies? Do the alternative conceptions have institutional implications that diverge from this? In my view, criticizing and complementing the narrowly economic conception is not only needed for explanatory purposes, but opens up a number of interesting possibilities at the normative level as well.

### 1. Why study institutions from the perspective of rationality?

Ethics is traditionally defined as the philosophical search for the good life (individual ethics) and the good society (social ethics). Paul Ricoeur combines both aspects in his definition of ethics as “the vision of the ‘good life’, with and for others, in just institutions”<sup>38</sup>. Institutions form an essential part in any study of the ideal society. They can be defined as relatively stable patterns of more or less formal rules, customs, practices and traditions that help organize, regulate, coordinate and structure social interactions between members of a society (Abercrombie, Hill & Turner 1994: 216-217). Despite their characteristic stability, these social and structural phenomena can be changed by deliberate interventions of individuals (Goodin 1996: 24-28).

---

<sup>38</sup> This is my own translation of the original quotation: “la visée de la ‘vie bonne’, avec et pour autrui, dans des institutions justes” (Ricoeur 1990: 202).

According to Claus Offe, institutions are the formalized counterparts of more informal norms. Both embody the expectations people can legitimately have towards others: “institutions establish standards, both normative and cognitive, as to what is to be held to be normal, what must be expected, what can be relied upon, which rights and duties are attached to which positions, and what makes sense in the community or social domain for which an institution is valid. Institutions accomplish a socializing function in that they serve as examples and reminders of how people “ought to” behave and relate to each other, and what they legitimately can expect from each other” (Goodin 1996: 199-200).

While these characterizations still remain very broad, I want to focus more concretely on what Thomas Christiano (2004: 124-125) labels “basic structural institutions”, by which he refers to “the basic institutions of the market, the political system, and the basic informal institutions of society such as the family and religious institutions”. In what follows, I will limit the analysis to three basic institutions, namely markets, states and communities. While these are spheres of social life that can be distinguished quite sharply on a conceptual level, they always overlap in actual practice. As such, they are best understood as ideal types that do not exist in their pure form. Modern societies always consist of a complex mixture of elements that characterize them: “‘market’, ‘state’, and ‘community’ may be thought of as describing three ways in which people may relate to one another (...). As participants in a market, people’s relationships are those of voluntary exchange. Each obtains what he wants by offering some equivalent benefit in return. As citizens of a state, people’s relationships are constituted by formal rules specifying what each is entitled and obliged to do. Goods are allocated by requiring some to provide them and empowering others to receive them. As members of a community, people are related by ties of identity which give rise to informal obligations of mutual aid” (Miller 1990: 17-18).

Rational choice theory can be said to have two principle aims when taking on the topic of institutions. First, it tries to explain and predict the functioning of institutions on a basis of the conception of individuals as rational beings. Second, it tries “to justify and criticize institutional structures by showing that when in place, the institutions are likely to bring about outcomes that are for the common good or not, under the assumption that they operate in the way that the explanatory theory says they do” (Christiano 2004: 122). In what follows, I will address both the explanatory and the normative issue.

First, I want to show how markets, states and communities can be expected to work given the insights on individual rationality from the first part of this dissertation. Second, I want to analyze which constellation of these basic

institutions seems to be desirable from the perspective of rationality. As the subtitle of this dissertation shows, I aim not only to explain individual behavior in rational terms, but also to think about the normative issue of how a society should look like. While the first part of this dissertation is devoted to constructing theoretical models that explain individual behavior in terms of rationality, this third part has an explicitly normative focus.

It has been debated to what extent the social sciences can and should be freed from normative judgments. According to positivists, scientific research should steer clear from the normative level altogether. However, positive and normative issues are inevitably linked to some extent. In this respect, I want to argue that economics and other social sciences should always take heed of its normative implications.

## 2. In search of a realistic utopia

As I will show more fully in the following chapters, normative issues are best decided on the basis of adequate models of individuals. As Milton Friedman (1953: 5) rightly argues, “any policy conclusion necessarily rests on a prediction about the consequences of doing one thing rather than another”. In this respect, I want to argue that the main task of political philosophers is to search for “a realistic utopia” (Rawls 1999: 12). This refers to a desirable constellation of institutions and policies that aim to facilitate interaction among individuals as they actually exist. Such a scheme is realistic in the sense that it could work, given the way real people act. It is utopian in the sense that it consists of an ideal that can be envisaged but is not yet realized<sup>39</sup>. Combining these two thoughts, the notion of a realistic utopia stresses that visions of the ideal institutional structure of society should be based on realistic models of its members. In order to explain what it exactly means, Rawls (1999: 13) refers to Jean-Jacques Rousseau whose self-proclaimed “purpose is to consider if, in political society, there can be any legitimate and sure principle of government, taking men as they are and laws as they might be”<sup>40</sup>. Samuel Bowles and

---

<sup>39</sup> A utopia is properly defined as “a place of ideal perfection especially in laws, government, and social conditions” (<http://m-w.com>). Alternative, it can be defined as “an impractical scheme for social improvement” (<http://m-w.com>). By adding the term realistic, one avoids the implication that a utopia refers to an inherently impossible, unrealizable or at least impractical ideal.

<sup>40</sup> In Rousseau’s own words, one should “inquire whether there can be some legitimate and sure rule of administration in the civil order, taking men as they are and laws as they might be” (Rousseau 1762: 17).

Herbert Gintis (1986: 186) stress that Rousseau already noticed the mutual dependence of individuals and institutions: “he recognizes the manner in which rules make actors and actors make rules”. In the following chapters, I hope to show that this insight leads them to defend a basic institutional structure that diverges from the one defended by James Buchanan, who largely ignores the causal impact of institutions on individuals.

A realistic utopia is thus an institutional scheme that is “compatible with the facts of human nature and social life” (Beitz 2000: 677). What matters is which institutional and policy schemes work best in practice, given the way actual individuals are (Buchanan 1999: XVI, 109). In this sense, the assumptions one makes when modeling individuals should be based on knowledge of the actual motivations of individuals in their public roles (Le Grand 2003: 2, 24).

Crucial in this respect is that human nature as it has evolved is characterized by more or less uniform regularities that are not given in some definite form but nevertheless present some boundaries. A realistic utopia is realistic in the sense that it could possibly work, given these general psychological propensities of human nature (Ben-Ner & Putterman 1998: 37; Cohen 2002: 384). This is why explanatory models are indispensable to the normative purpose discussed here. However, while I agree that empirical research is crucial, political philosophers should try to think up and evaluate different institutional schemes. Their normative views can, in the end, motivate individuals to try and improve the institutional structure of their societies (Goodin 1996: 36-37). Since this is a dissertation in philosophy, I will not argue in favor of specific policy options or institutional reforms. Nevertheless, I hope it contains a number of useful suggestions with regard to the routes one might choose to follow in order to arrive at a better society. After all, this is what political philosophers should be concerned with: “political philosophy provides a long-term goal of political endeavor, and in working toward it gives meaning to what we can do today” (Rawls 1999: 128).

The notion of a realistic utopia combines both of the tasks set out by Philip Pettit in this introduction’s opening quote. While most contemporary political philosophers have focused on the first task of institutional design as some kind of ideal theorizing, they have largely ignored the second task of investigating which design is fit to survive in actual practice. In what follows, I therefore want to take on the issue which basic institutional schemes can be implemented in a modern society full of people that are rational in the meanings distinguished above.

### 3. Introducing Public Choice theory and Analytical Marxism

In order to introduce the following chapters, I want refer to Robert Goodin (1996: 13) who distinguishes between two dominant fields within political thinking. According to the first, individuals are analyzed as economically rational. According to the second field, individuals are analyzed as they are. As should be clear by now, I am convinced that the second strategy is to be preferred, even though I believe this does not imply doing away with the notion of rationality altogether.

I hope to analyze more fully what the institutional implications are of both fields. In the fifth chapter, I will focus on Public Choice theorist James Buchanan as the main proponent of the first field. Public Choice theory essentially consists of two strands, which correspond perfectly with the two main aspects of any study of institutions from a rational choice perspective (Buchanan 1999: XVI, 331; Mueller 1993: 447-485). The first, positive strand analyzes institutions as they function in reality. How do citizens elect their representatives? Which policy decisions will the elected governments take and what does the resulting institutional structure look like? The second, normative strand aims to analyze not what institutions are, but what they should be. Given the knowledge of how economically rational individuals act, how can and should institutions be designed and reformed. Public Choice theory thus explicitly addresses both pillars on which this dissertation is based: rationality and its normative implications (Mueller 2003: 668-674).

In the sixth chapter, I want to show that Analytical Marxists such as Samuel Bowles and Herbert Gintis have a legitimate say in these debates. They too want to improve (the basic institutional scheme of) society by starting from a rational choice perspective. I believe they form a welcome counterpart to Public Choice theorists in that they try to construct more adequate models of individuals that lead to completely divergent views of desirable institutional arrangements.

To be sure, I think both fields make a valuable contribution to the issues at hand. This way, I hope to open up the debate as it is held in contemporary political philosophy. After all, this is completely centered round and dominated by the work of John Rawls. As Robert Nozick claims, “political philosophers now must either work within Rawls’ theory or explain why not” (Nozick 1974: 183). I choose not to address his theory, partly because I believe that there exist already (more than) enough decent and extensive studies of his work. Instead, I focus on authors like Buchanan, Bowles and Gintis whose views provide much

lesser known voices in the debates on a desirable basic institutional structure for modern societies.



## CHAPTER 5

### JAMES BUCHANAN'S THEORY OF CONSTITUTIONAL CHOICE

---

*“It is not that everybody is somehow narrow, self seeking, always out to grab his own interest and trying to increase his own income share. But if we are talking about institutions, that is the best presumption we can make”*

(Buchanan 1999: XVI, 108-109)

#### 1. Why study James Buchanan?

##### 1.1. Who is James Buchanan?

The aim of this dissertation is to answer the question what the institutional structure of society should look like from a rational choice perspective. If one wants to analyze the role of institutions in a society of rational individuals, one cannot ignore the field of Public Choice theory. While this covers a wide range of authors, ideas and topics (Mueller 2003), it is useful to focus on the extensive work of James M. Buchanan<sup>41</sup> rather than give an overview of the whole field. This 1986 Nobel Laureate in Economics is often considered to be the founder and most important proponent of Public Choice theory. After his first publications helped establish the field, he never stopped enlightening his colleagues with inspiring ideas and new insights. Throughout the years, Buchanan has covered a wide range of topics from a number of methodological perspectives. As a result, his extensive body of work is characterized by some ambiguities and inconsistencies as well. Where

---

<sup>41</sup> Throughout this chapter and unless mentioned otherwise, I am referring to *The Collected Works of James M. Buchanan* as they have been edited and published by Liberty Fund (Buchanan 1999). This is by far the most complete collection of Buchanan's writings. It consists of 20 volumes and over 6,500 pages. This collection contains – a few exceptions aside – all of Buchanan's publications. For reasons of simplicity, I attribute the views expressed in co-authored works to Buchanan. This is in no way intended to downgrade the efforts of his collaborators.

necessary, I will investigate these and try to clarify what Buchanan's main point is.

Buchanan consistently applies the concepts and tools of economics to research areas that lie beyond its traditional domain of market relations. In doing so, he has helped establish Public Choice theory, which he defines as the research area that "applies the techniques and analytic apparatus of modern economics to the study of political processes" (X, xvii). Public Choice thus consists of economists using their own conceptual apparatus and methodology in their analysis of the functioning of public institutions (Mueller 2003: 2). In order to see what the public domain looks like from this perspective, I will examine Buchanan's intellectual background (second section), his theoretical assumptions (third section) and the way these result in his theory of constitutional choice (fourth section). Then, I will try to formulate some criticisms (fifth section) and consider their implications for Buchanan's normative views (sixth section). Finally, I will try to draw some general conclusions (seventh section).

I thus aim to show how Buchanan's normative views arise from the specific theoretical assumptions in his economic conception of rationality. Of course, since the different aspects of his work are closely entangled, several assumptions tend to have the same normative implications, just as well as a single assumption can have a number of normative implications. All this inevitably leads to some overlap between and repetition in different parts of this chapter. While I believe that this way of analyzing and evaluating Buchanan's work has its value, it is up to the reader to decide whether it is worthwhile or not.

If Public Choice theory is economics applied to politics, it is important to understand correctly what exactly economics is. In this respect, I immediately want to stress that Buchanan has a particular conception of economics. It is situated squarely within the Austrian tradition of von Mises and von Hayek who stress the subjective nature of choice and thus also of economics, which can be defined as the science of choice. Buchanan proposes a completely subjective definition of cost, which leads him to discard the possibility of determining the efficiency of some choice or institutional structure independently of an individual's judgment (VI, 23-25, 30, 41, 87; XII, 8). In the same vein, Buchanan gives a subjective definition of economics as the "theory of exchange of whatever it is that persons value" (XII, 23). In the language of an economist, what individuals value is what they prefer. Because an individual's preferences do not exist in some objective sense, independent of his choice, it is impossible to derive an objective utility function (XII, 26). This is also why Buchanan defines economics as the study of 'catallaxy' rather than

that of 'maximizing'. Instead of limiting economics to the study of optimally allocating the existing means in order to achieve certain goals or ends (maximizing), Buchanan wants to focus on the process of voluntary trade and exchange (catallaxy) (XII, 115-117, 470; XVI, 32, 60, 240).

Applying the conceptual apparatus of economics to politics, Buchanan has filled up a huge lacuna in theorizing about the role of the public domain in society. Until the foundation of Public Choice theory, the bulk of the literature on institutional design suffered from a lack of explicitly defined theoretical underpinnings about the ways institutions like markets and states function. While classical economists still realized the importance of an institutional framework for societal interactions to run smoothly, neoclassical economists increasingly relied on concepts of the market and the state as ideal-type abstractions. The fact that these were reflected only to a greater or lesser degree in reality receded into the background. Instead of engaging in formal modeling of idealized equilibria, Buchanan wants to renew the attention for the institutional conditions to be fulfilled in order for socially desirable outcomes to arise.

In doing so, Buchanan also explicitly criticizes most political philosophers and their romantic view of the state as benevolent despot (VII, 204; X, 39). They naively assume that governments automatically serve the public good, however this may be defined. Those who conceive of politics as being concerned about some general will, like Jean-Jacques Rousseau (1762: 26), take for granted that public servants try to find out what this entails and how to realize it. Yet even those who think that politics is really about private interests trust the state to serve exactly these interests. Social choice theorists, for example, who aggregate private interests in order to construct a social welfare function, presume that governments systematically take this function as their main objective.

Buchanan deserves credit for making this view explicit and questioning its validity. He urges one to be wary of the goodwill of political representatives and public servants. Simply assuming that they are benevolent and public-spirited is a sign of grave naivety. In this respect, Buchanan wants to revitalize the classical insistence on a theoretically unified model of individuals. Instead of simply assuming a public-spirited nature, Buchanan relies on the model of individuals as *Homines Economici* in both their private and public roles. In his attempt to construe institutions that guarantee social order, he thus wants to avoid assuming that everybody is cooperative by nature.

## 1.2. Why study James Buchanan?

As will become clear, I want to focus on what I believe is Buchanan's main goal, namely to formulate proposals concerning the design and reform of institutional arrangements in order to facilitate social life, while respecting each individual's liberties and interests. Within the game of life, Buchanan thus focuses on defining and adjusting the rules in ways that allow the players to play the game in mutually beneficial ways. I try to interpret Buchanan's oeuvre in this light, making abstraction from themes and topics that do not immediately touch upon this aspect. I fully acknowledge that such a thematic approach largely neglects the evolutions that his work has undergone during the last 45 years. Nevertheless, I believe that it can be valuable, partly because it is also how Buchanan's Collected Works are structured. I aim to present a general overview of his work rather than perform a detailed study of his main works, primarily because I believe that quite a few of his most revealing insights can be found in some of his less well-known articles and monographs. Since not every aspect of his work can be treated in the limited space of this chapter, I realize that this method is a second-best solution. Nevertheless, I hope it helps the reader to understand and appreciate Buchanan's insightful contribution to this dissertation's central topic.

Buchanan has consistently aimed to do what he considers to be the task of every social scientist, which is to propose institutional reforms that facilitate social life. In this light, he has to find a criterion to determine which institutional changes form an improvement and which do not. To answer this meta-ethical question of which principles, values and ideals are suitable in deciding what is morally good, Buchanan argues that the values and preferences of individuals are the only relevant standards. What is morally desirable is to be equated with what is preferred by the individuals themselves. This implies that the public good is whatever individual citizens deem valuable.

The most important value in Buchanan's theory is freedom, which he defines "only in its negative sense: an individual is at liberty or free to carry on an activity if he or she is not coerced from so doing by someone else" (XIV, 219). Buchanan delineates private spheres in which individuals should be left alone without outside interference. Within these "behavioural spheres of action" (XVII, 162), each individual is free to choose his own goals and to try and realize these. Individual freedom is nothing more than the number of options open to the individual (XVI, 387). As I will show later on, this not only implies an instrumental image of individuals, it also leads to a view of institutions as externally constraining the action range of individuals. According to Buchanan, "rules define the private spaces within which each of us can carry on our own

activities" (X, 5). The general thrust of his work is quite basic: if individual freedom is a highly praised value, institutions and rules should be arranged in such a way that these protect the liberties of individuals (XVI, 329).

An important reason for studying Buchanan therefore lies in his emphasis on the importance of institutions in society. As Robert Tollison remarks, "more than anything else, Buchanan's basic insight that rules are important can be said to have driven most of his written work over his career. More than any other single scholar, he is responsible for this emphasis and its emergence in modern economics and political science" (X, xi-xii)<sup>42</sup>. In this respect, Buchanan places himself within the tradition of Adam Smith (X, 15), with whom he shares the view that, in short, "institutions matter" (VIII, 131).

Buchanan's emphasis on the importance of institutions in society is best understood using the analogy with rules of a game (IX, 5; XIV, 271-276 XVI, 23, 103, 339; XVIII, 269). Providing the framework within which individuals interact, rules are necessary if one wants to play a game. They bring predictability and, as a consequence, enable individuals to engage in collective actions. As will become clear, Buchanan has his own take on the Monty Python dilemma from this dissertation's opening quote.

Buchanan focuses on institutional matters, because he believes that improvement is hard to think of when one treats the institutional setting as given. In contrast, improvement can be attained by changing this setting itself (XV, 290-294). Therefore, Buchanan ultimately aims to compare, evaluate and propose alternative institutional arrangements (XVI, 107). One could argue that Buchanan is not a typical economist, since economics is the study of how choices are made within given institutions and rules. Studying how choices are made among institutions and rules, he conceives of himself as a constitutional political economist (XVI, xi, 3-7, 127, 331, 357, 432). Buchanan wants to analyze the role of the political economy, which consists of "the whole set of constraints, or structure, within which individuals act in furtherance of their own objectives" (XVI, 250).

Buchanan thus returns to the classical insistence on studying "the choice of constraints as opposed to the choice within constraints" (XV, 4) which is the main focus of neoclassical economics. Analyzing individual decisions within given rules and institutions, neoclassical economists have largely ignored the

---

<sup>42</sup> Notable contenders are the so-called institutional economists, like Geoffrey Hodgson (1988), Keith Dowding and Desmond King (1998), who stress the importance of institutions in making possible and regulating economic interactions. Following Thorstein Veblen (1909), these economists show that theories that ignore the central role of institutions are inapt to gain insight in the functioning of an economy, since they form the preconditions for its survival.

fact that the latter are formed through individual choices (I, 377-383). Buchanan systematically stresses that society's rules are ultimately made by individuals in democratic discussions. This is also why he puts his hopes for a better society in institutional design and reform (XI, 5; XVI, 147).

### **1.3. Why not study another political economist or political philosopher?**

There are several strands of thought that combine a proper recognition of both the importance of institutions in society with an emphasis on the need for rational individuals to consent to them. Claiming “that legitimate authority of government must derive from the consent of the governed, where the form and content of this consent derives from the idea of contract or mutual agreement” (Cudd 2003), these theories refer to some sort of social contract that is to be agreed upon by all relevant parties. In this field, theorists also think about normative issues – like the desirability of institutions – from a rational choice perspective. They basically argue that institutions are legitimate when they are agreed upon by rational individuals who are expected to consider whether the arrangements at hand will help them to achieve their goals. Like Buchanan: “social contract theories take individuals to be the best judges of their interests and the means to satisfy their desires” (Cudd 2003).

Within this field, one can distinguish between contractarianism and contractualism. The former, which stems from Thomas Hobbes and is defended (amongst others) by David Gauthier and James Buchanan, argues that individuals should primarily be modeled as self-interested beings. In their view, institutions are legitimate insofar as they provide arrangements beneficial to each citizen: “justice, and so a social contract is only possible where there is some possibility of benefit to each individual from cooperation” (Cudd 2003). This strand conceives of justice as mutual advantage. The latter, which stems from Immanuel Kant and is defended (amongst others) by John Rawls, argues that individuals should primarily be modeled as reasonable beings. In their view, institutions are legitimate insofar as they provide arrangements that can be justified to each citizen in a public and reasonable way. This strand conceives of justice as impartiality. As will become clear later on, these differences have implications for the way in which public arrangements are analyzed and evaluated.

In order to defend my choice of Buchanan among these social contract thinkers, I need to argue why I do not focus on the others. While interesting, the work of David Gauthier falls largely beyond the scope of this research project, since it focuses primarily on the connection between rationality and

morality. As a contractarian, Gauthier (1986: 2) defends a “conception of morality as a rational constraint on the pursuit of individual interest”. In search of a rational basis for morality, Gauthier argues that constrained maximizers – individuals who are disposed to comply with mutually advantageous moral constraints – tend to benefit more from social cooperation than straightforward maximizers do (Gauthier 1986: 15). This shows that behavior guided by ethical concerns is on the whole to the agent's benefit. Gauthier therefore argues that all moral principles (prescribing impartial constraints) are rational (prescribing the maximization of one's interests<sup>43</sup>) (Gauthier 1986: 3-4).

Instead of focusing on morality, Buchanan explicitly wants to analyze the connection between rationality and institutions. Nevertheless, Gauthier has something to say about the role institutions should play as well. He, for example, argues “that in a perfectly competitive market, mutual advantage is assured by the unconstrained activity of each individual in pursuit of her own greatest satisfaction” (Gauthier 1986: 13). Gauthier thus reserves a large role for markets, since they are able to attain the most efficient outcome while respecting each individual's freedom. Nevertheless, he also admits that there is need for politics since the world does not function as a perfectly competitive market (Gauthier 1986: 18). In order to move beyond Gauthier's scarce remarks, I deem it useful to focus on the work of Buchanan who analyzes these issues more thoroughly and extensively.

As I have already suggested, I choose to largely ignore the work of John Rawls. Even though I will try to stay away from his theory to the largest extent possible, I will touch upon some of his views, mostly in comparison with those of Buchanan. This way, I hope to show the appeal of Buchanan's views as a way of answering some of the questions Rawls raises.

## 2. Buchanan's intellectual background

If one wants to situate James Buchanan in the history of ideas, one should refer to welfare economics as it has developed since the 1960s (Gravelle & Rees

---

<sup>43</sup> Gauthier remarks that his conception of rationality is a maximizing one, according to which rationality prescribes one to seek the greatest satisfaction of one's own interests. Nevertheless, he quickly adds that this does not assume self-interest: “on the maximizing conception it is not interests in the self, that take oneself as object, but interests of the self, held by oneself as subject, that provide the basis for rational choice and action” (Gauthier 1986: 7). This touches upon the distinction between egoism and egocentrism, which also returns in Buchanan's work.

2004: 293-304; Walsh 1996: 159-163). Indeed, the two fundamental theorems of welfare economics form a major part of Buchanan's general intellectual background. The first states that any equilibrium in a perfectly competitive market is Pareto-efficient, which means "that there is no other feasible allocation which makes some individual better off and no individual worse off" (Gravelle & Rees 2004: 280). However, the problem remains that Pareto-efficient allocations of goods can be widely unequal. The second theorem addresses this problem by stating that any Pareto-efficient allocation can be achieved through a suitable redistribution of the initial endowments of the parties involved. The basic conclusion is that perfectly competitive markets allow consumers to maximize their utility and producers to maximize their income without any form of regulation or intervention. In such an Arrow-Debreu world, efficiency and freedom go hand in hand. The influence of this general line of reasoning on Buchanan will become clear as I more fully analyze his reliance on concepts like freedom, efficiency and Pareto-optimality.

The market has long been praised for its ability to manage in an efficient manner the social interactions between individuals in search of their own goals. This has been the case since Adam Smith defended the market on the basis of its capacity to spontaneously coordinate self-interested actions into socially beneficial outcomes. After all, "it is not from the benevolence of the butcher, the brewer, or the baker, that we can expect our dinner, but from their regard to their own interest" (Smith 1776: Book I, Chapter 2: 26-27). In contrast with the inherently coercive measures of the state, the market enables individuals to freely go about their own business without interfering with each other. One can distinguish between different authors who have defended such views. All of them had a substantial impact on economics in general and on Buchanan in particular.

First, there is the Austrian economist Friedrich August von Hayek. His main argument in defense of the market is that its alternative, namely government intervention, immediately and inevitably puts one on "the road to serfdom" (von Hayek 1944). In his view, the state has inherently totalitarian tendencies, which the market manages to avoid. Another advantage of the market is its capacity to disperse information that is too complex to collect and process in any centralized way. As fellow Austrian Ludwig von Mises (1920: 128-130) already predicted, centrally planned economies inevitably perform badly in this respect. In this light, von Hayek systematically stressed that only free markets are able to efficiently allocate existing resources into the domains where they are most productive. In his view, the economic problem is not, as most economists think, the allocation of scarce resources among competing ends. Rather, it is about how to use the available resources for ends whose importance only individuals know. The basic problem is thus an epistemic one:



because no single individual can take into account all possible factors of the complex economic reality, one has to rely on the impersonal mechanism of the market to transfer and coordinate the relevant information (von Hayek 1944: 48-50, 75-76).

This emphasis on the epistemic advantage of the market over the state has been the primary focus of von Hayek's attention (von Hayek 1945: 524-526). In his view, the market generates a spontaneous order, indirectly caused by the actions of individuals interacting within its structures but never consciously intended by any of them. The fact that the market allows individuals to act free from coercion forms its main advantage: "the competitive system is the only system designed to minimize by decentralization the power exercised by man over man" (von Hayek 1944: 145). In contrast with the market, the state requires deliberate planning and organization of such a kind that, at least in principle, a single individual should be able to understand the whole economic structure. The fact that the capacities of the human mind are limited leads von Hayek to be wary of state intervention.

Von Hayek's defense of the market based on the limited knowledge of its participants is rather unusual in light of the fact that most economists, as I have shown earlier, impose demanding requirements to speak of rationality at the cognitive level. In the same vein, critics of the market have argued that the market is bound to fail – and that therefore government intervention is needed – because real-world individuals cannot possibly live up to the epistemic requirements of perfect rationality. In contrast, von Hayek uses the fact that agents are not perfectly rational as an argument in favor of the market and against centralized planning efforts.

According to von Hayek (1944: 205), government intervention unavoidably leads to the centralization of power in the hands of political elites, which entails that each citizen's freedom is submitted to the arbitrary power of others. That is why governments are inherently less desirable than markets: "planning leads to dictatorship because dictatorship is the most effective instrument of coercion and the enforcement of ideals and, as such, essential if central planning on a large scale is to be possible. The clash between planning and democracy arises simply from the fact that the latter is an obstacle to the suppression of freedom which the direction of economic activity requires" (von Hayek 1944: 70). In the end, von Hayek's main reason for defending the market is that it protects what he considers the most basic value of them all, namely freedom. Since this is to be understood as freedom from coercion, von Hayek (1944: 52, 102-104, 241) favors the impersonal mechanism of the market over the personal power of political elites.

Second, there is the work of Milton Friedman, a free market advocate who – quite like von Hayek – defends the market because it forms the only guarantee to economic, human and political freedom. This emphasis on freedom is already present in his definition of the market as “a mechanism for enabling a complex structure of cooperation to arise as an unintended consequence of Adam Smith’s invisible hand, without any deliberate design. A free private market involves the absence of coercion. People deal with one another voluntarily, not because somebody tells them to or forces them to”<sup>44</sup>. The main thought in Friedman’s work in fields like monetary theory and tax theory is expressed in his motto: the government solution to a problem is usually as bad as the problem itself. He argues that the market is morally justified since it preserves individual dignity by leaving choices to the individual. It allows a society to be organized from the bottom up and not from the top down.

Third, I want to refer to the libertarian views of American philosopher Robert Nozick, whose thoughts are partly based on the work of von Hayek and Friedman. Nozick (1974: 163-164) argues that any distribution of goods is just, insofar as it arises from free exchanges by initially equal participants. As a result, he provides a purely procedural defense of the market. Relying on its invisible hand mechanism, he ends up justifying a minimal state whose only task is to protect basic individual rights (like property rights) and to ensure the enforcement of contracts between free and sovereign individuals (Nozick 1974: 333-334).

As Allen Buchanan states, one can distinguish between several types of arguments for and against the market, namely those based on efficiency, justice and freedom. First, economists often argue that markets always reach an efficient outcome under ideal circumstances (Buchanan 1995: 14). This is the main thrust of the abovementioned theorems of welfare economics and von Hayek’s emphasis on the market’s epistemological advantage over a centralized planning system (Buchanan 1995: 17). However, there is the widely acknowledged problem that actual markets do not always approximate their ideal. In the case of externalities and public goods, market equilibria are not nearly as efficient as they could be (Buchanan 1995: 22). Nevertheless, the burden of proof in this respect lies with the advocates of government intervention, who do not – as market advocates do – have a solid explanation of why their preferred system would systematically produce efficient outcomes (Buchanan 1995: 25-26).

---

<sup>44</sup> Friedman in a speech delivered to the Smith Center at California State University on November 1, 1991. The speech is titled ‘Economic Freedom, Human Freedom, Political Freedom’ and is available online at: <http://www.sbe.csu Hayward.edu/~sbesc/frlect.html>.

Second, economists defend or attack the market from the perspective of justice. Nozick, for example, argues that the market is the only system that does not infringe upon inviolable individual rights, like the basic right to private property (Buchanan 1995: 64-66). The state should only support the market (by enforcing contracts) and complement it when it fails (by correcting for externalities and public good problems) (Buchanan 1995: 71-74). Critics of the market argue that it allows for inequalities that prevent individuals making effective use of their rights, even though these may be equal for all participants in some formal sense (Buchanan 1995: 81-83).

A third sort of argument refers to the value of freedom. As I have shown above, authors like von Hayek and Friedman argue that the market is to be preferred over other institutional arrangements because it does not intervene in the private spheres within which individuals are free to do what they want to. However, this ignores the possible side effects of market relations on third parties (Buchanan 1995: 80). Once more, externalities seem to be the main reason why markets fail. In what follows, I will therefore go deeper into these arguments and the way James Buchanan employs them.

### 3. Buchanan's theoretical assumptions

#### 3.1. Methodological and normative individualism

Following von Hayek, Buchanan stresses that economics should be about individual choices and the consequences thereof and not about the distribution of scarce means to attain certain goals. Economists should concentrate on the processes of exchange and trade in which individuals voluntarily engage in order to achieve mutually beneficial arrangements. In this respect, Buchanan labels himself "a methodological and normative individualist, a radical subjectivist, a contractarian, and a constitutionalist" (XVI, 238). In what follows, I will show how these notions relate to each other.

Buchanan is a self-pronounced defender of methodological individualism, which states that individuals are the necessary starting point for any valid explanation within the social sciences. He systematically tries to explain social phenomena by referring to individuals and their choices (XIII, 16-17). In his view, social sciences should honor this principle, because it reflects the basic insight that individuals are the ultimate and only relevant source of value (I, 203, 457; VII, 163, 210; XVI, 8-9). This points out that Buchanan's main emphasis lies on normative individualism (XVI, 62). He explicitly argues that

the public good “can only be defined in terms of individual evaluations” (III, 35) and is thus nothing more than what “tends to emerge from the voluntary agreement among persons” (XVIII, 322). Hence, individuals and their values are the only proper criterion to evaluate possible states of affairs.

Buchanan’s normative individualism follows from the unshakable belief most economists have in what they call “consumer sovereignty” (XII, 416). This basically states that the individual is always free to decide what he wants. Traditionally, it applies to the individual in his private role as a consumer and equates “what a consumer prefers with what it is preferable for that consumer to have” (Walsh 1996: 188). Buchanan extends this thought to the individual’s public role as a citizen, stressing that the individual is the only and final measure in public issues as well. The underlying thought is that each individual has equal value as a human being (XVI, 19, 24, 89). His main reason for defending individualism is thus not an epistemic but a moral one<sup>45</sup>. He argues that individuals have sovereign power to make choices and hence are also ultimately responsible for them (I, 290-291, 390; III, 172). In the end, the notion of consumer sovereignty, which means that all value comes from individual preferences and never from outside (XV, 363, 369), boils down to the same thing as normative individualism.

It should be clear that Buchanan systematically defends methodological individualism as Philippe Van Parijs defines it: “the individual is the unique unit of consciousness from which all evaluation begins. Note that this conception does not in any way reject the influence of community or society on the individual (...). The presupposition requires only that societal or communitarian influences enter through modifications in the values that are potentially expressed by the individual and not externally” (X, 25-26). While social norms and institutions may exist, they only have an impact insofar as they influence individual behavior. Even though Buchanan does not deny that individuals are influenced by norms, values and traditions, he consistently rejects the idea that these exist as supra-individual sources of evaluation, independent of individuals (I, 391; IV, 241; XI, 24; XIV, 73)<sup>46</sup>. According to

---

<sup>45</sup> Buchanan does not place the individual at the center because the latter knows best what it wants (I, 282-283). Nevertheless, Buchanan argues that there should be a considerable degree of laissez-faire within the basic institutional structure of society, precisely because individuals are the only ones who know which goals they want to realize (XVIII, 274-275). He also states that socialism is deficient primarily because an omniscient and benevolent despot does not satisfy preferences as well as the market (XVIII, 310). Even though Buchanan does not consider these Hayekian arguments to be decisive, one can see here that he subscribes to them nonetheless.

<sup>46</sup> In Buchanan’s view, the false belief in the existence of such supra-individual entities arose when neoclassical economists broadened the maximizing paradigm of economics

Buchanan's normative individualism, there is simply no such thing as the public good, defined independently from individual values (IV, 286, VII, 210; XVII, 161). In this respect, he situates himself in the tradition of von Hayek (1944: 212), who already argued that the basis of morality lies in the individual and his evaluative capacity.

Buchanan believes that social states of affairs do not exist as supra-individual entities. Moreover, they cannot be maximized simply because they cannot be chosen in a direct and deliberate manner (XVI, 248). In Buchanan's view, whatever comes out of freely undertaken interactions is good, even though this result is not deliberately aimed at (XVI, 202). Because outcomes can be evaluated and ranked, but not chosen directly, one has to choose between alternative rules in the hope that these will lead to the desired outcomes (XVI, 248-251). Even though individuals ultimately *evaluate* outcomes, they must *choose* between rules and procedures (XVI, 211). They can make up the rules of the game and choose strategies within the game, but they cannot directly choose its final result (XVI, 419).

### 3.2. The Homo Economicus model

Buchanan makes extensive use of the Homo Economicus model to analyze individuals and their behavior. In order to give a general idea of his interpretation of this model, I will bring into remembrance its most distinctive features.

#### 3.2.1. Instrumentality

As I have already shown, Buchanan believes that economics should drop the assumption of maximization at the level of actions and the relation with their reasons. In contrast, he does hold on to the notion of intentionality. However, since this is not unique to the economic conception of rationality, I will focus here on the assumption of instrumentality. Since it analyzes individuals as aiming to satisfy their preferences, the Homo Economicus model is essentially an instrumental one: "the conceptual apparatus here involves a radical separation between means and ends – between opportunity sets and preferences" (X, 56). Individuals are analyzed as being instrumentally motivated to accomplish certain goals, which are defined by their preferences.

---

from the individual to the social level (XVI, 106-107, 240-241). In what follows, I will try to show how Buchanan's catallactic paradigm aims to correct for this fault.

### 3.2.2. *Egoism*

Since Buchanan models the individual as motivated towards the satisfaction of his preferences, the latter form the central explanatory variable in his theory. In what follows, I will focus on the two most characteristic assumptions in the economic conception of rationality at the level of reasons, namely that of egoism and exogeneity. Like most economists, Buchanan stresses that the Homo Economicus model focuses on egoistic preferences and analyzes individuals “as seeking to further their own self-interest, narrowly defined in terms of measured net wealth position” (XIII, 21). This view, which has predominated economics since John Stuart Mill, implies that individuals systematically try to improve their own well-being.

Buchanan argues that the conventional Homo Economicus model goes much further than the assumption of methodological individualism, since it assumes that people are only concerned about their narrowly defined self-interest (XVI, 10). In this respect, it seems plausible to treat Buchanan as a methodological utilitarian. However, he is not completely unambiguous about the status of the assumption of egoism in this model. Some of his passages suggest that he takes on a different position. From time to time, Buchanan explicitly states that the Homo Economicus is not necessarily a full-blooded egoist (III, 3). Here, he seems to favor the broader, neoclassical variant of the Homo Economicus to the narrower, classical variant. While the latter analyzes individuals primarily as being out to acquire material wealth, the former understands individuals as trying to realize their goals, whatever these may be. For example, Buchanan explicitly argues that, “for the purposes of constitutional design, (...) no specification of arguments in *Homo economicus*’ utility function is required” (XII, 82). Claiming that economic self-interest is not always the same as material utility (I, 460), Buchanan considers his theory to be economic, but not “in the restricted sense that denotes a pre-eminence of materialist motivation in man” (XVI, 385). These remarks are consistent with Buchanan’s individualism, according to which preferences are not necessarily egoistic (I, 48-49, 391).

In Buchanan’s view, not all individuals always act egoistically, but at least a substantial part of them do so in a substantial number of situations (XV, 26; XVI, 200). Interpreting the Homo Economicus model as a partial one, Buchanan thus seems to limit the scope of phenomena it is apt to explain. Here too, one can see that he does not nicely fit the label of a methodological utilitarian, since he does not deny the explanatory force of theories that stress motivations other than narrowly defined self-interest. However, as I will show later on, Buchanan does not go the whole nine yards and continues to analyze actions as means towards the fulfillment of narrowly self-interested goals.

### 3.2.3. *Exogeneity*

Economists – and Buchanan is no exception in this respect – take preferences as the starting point for their explanations (X, 56). Refusing to investigate how they arise and develop in time, they want to take preferences at face value: “along with the economist and other social scientists, the political theorist should take his human actors as he finds them” (III, 309). In general, this way of analyzing preferences as exogenously given leads to an atomistic conception of the individual as an entity, already formed in all of its relevant aspects and largely detached from its social, cultural and institutional environment.

## 4. Normative implications of Buchanan's theoretical assumptions

So far I have tried to focus exclusively on the explanatory and conceptual aspects of Buchanan's work. In what follows, however, I want to analyze his views of the good society and the role basic institutions should play in it. As Buchanan himself rightly argues, “there are important normative implications to be derived from the public choice perspective on politics, implications that, in their turn, carry with them an approach to institutional reform” (XIII, 19). In the end, his aim is not merely to analyze how things are, but to suggest how they should be and in what ways they should be changed: “the only purpose of science is its ultimate assistance in the development of normative propositions. We seek to learn how the world works in order to make it work “better,” to “improve” things” (III, 306). In this respect, one could even argue that Buchanan agrees with Karl Marx's eleventh thesis on Feuerbach, according to which it is the task of philosophers to change the world rather than to interpret it<sup>47</sup>. In Buchanan's view, understanding the world is useful mainly because it ultimately serves attempts to improve the world<sup>48</sup>.

---

<sup>47</sup> In Marx's words: “the philosophers have only *interpreted* the world in various ways; the point is to *change* it”. This is the English translation of the original quote: “die Philosophen haben die Welt nur verschieden *interpretirt*, es kommt drauf an sie zu *verändern*” (Marx 1845: 15).

<sup>48</sup> Buchanan is fully aware that one's evaluation of public arrangements depends heavily on the way one understands them. One's analysis of the ways markets and governments function largely determines one's normative opinions about the role these should fulfill in a society: “in a certain sense, the ought is derived from the presumed is” (XVI, 245).

#### 4.1. Normative implications of methodological and normative individualism

The assumptions of methodological and normative individualism immediately imply a specific stance on normative issues. It has been argued that they form the crux upon which the whole of welfare economics is built, including its two fundamental theorems and the resulting appraisal of competitive markets. Instead of exploring this general argument, I want to go deeper into Buchanan's own remarks and distinguish between different repercussions of his individualist perspective.

##### 4.1.1. *Buchanan's theory of constitutional choice*

In order to think about the ideal institutional structure of modern societies, Buchanan develops his theory of constitutional choice, according to which an ideal structure is one upon which all individuals rationally agree. Instead of relying on supra-individual norms to justify certain institutions and rules, he refers to the principle that all individual contractors should freely agree to the latter (I, 370; XVI, 363). This way, Buchanan is able to stay within the confines of his individualism (XV, 486-487).

Buchanan defines the constitution as the agreed on set of rules within which the game of politics can and should be played (XVI, 43-45). Its main task is to restrict post-constitutional politics in such a way that the freedom of all citizens is optimally protected (XV, 459). It is in everyone's interest to set up a basic set of rules to regulate social life (XVI, 179; XVIII, 178). Even though rules may have negative effects in the short run – they curb individual freedom – they can be freely chosen for their positive effects in the long run. The fact that they bring about overall benefits is why individuals will approve of coercive rules at the constitutional level (XVI, 73). The more the issue is framed in a general way, the easier it becomes to attain unanimous consent (III, 249-250; IV, 294). Buchanan thus strongly stresses that agreement is more easily attainable at the constitutional than at the post-constitutional level (XVI, 63, 74, 307-308). Whereas constitutional choices are typically situated in the long term, post-constitutional policy choices are typically situated in the short term.

Buchanan's analysis of the constitutional level closely resembles Rawls' Original Position. Rawls also wants to think up a desirable institutional structure by referring only to the views and values of individual participants in the social scheme. In both theories, it is the players themselves who on an



equal basis agree on the rules of the game (XVI, 38, 327, 366). Additionally, both authors rely on some kind of constitutional veil, which inhibits the individual's knowledge about his future preferences in order to avoid reliance on criteria external to the individuals themselves. However, there are several important differences.

First, Rawls stresses that individuals behind the constitutional veil are completely ignorant of the impact of their decisions on their own well-being. After all, they know nothing of their future position in society. In contrast, Buchanan stresses that individuals are not wholly ignorant but only uncertain about the impact of their choices among alternative sets of rules on their own position.

Second, Buchanan stresses the purely procedural character of the constitutional level, whereas Rawls deduces more substantive results from it (I, 146, 308; IV, 295; XIV, 215-216, 503; XVI, 22, 74; XVIII, 181). In Buchanan's view, constitutional rules simply define the procedures of collective decision-making. Accordingly, he does not claim any insight into the content of the resulting outcomes. The basic thought is that outcomes are legitimate if all participants agree on them, or at least on the process through which they come about. This emphasis on the procedural aspect of rules does not imply that Buchanan has nothing to say about the outcomes of constitutional choices. After all, he criticizes Rawls for not thinking through his own claim that freedom has lexicographic priority over efficiency and productivity considerations. According to Buchanan, this would automatically imply a much larger role for the market than Rawls himself defends (I, 302). Moreover, when evaluating rules, their outcomes are the main thing to take into account. Rules are not valued intrinsically, but only in relation to the outcomes they produce (XVI, 49, 159). However, since it is impossible to choose in a deliberate and direct way particular social outcomes, constitutional choice has to focus on reforming the procedures in such a way that the resulting outcomes evolve in a desirable direction. This way, Buchanan is able to think of rules and institutions in an individualistic framework.

Third, even though both authors use the constitutional level as a device to think up and evaluate different institutional schemes (XI, 192), they interpret its status differently. While Rawls explicitly thinks of it as a hypothetical construct, Buchanan considers it to be an aspect of actual political decision-making. Whereas Rawls argues that individuals in the Original Position could have chosen the rules that exist here and now, Buchanan stresses that individuals here and now have to obey the rules that would be the result of such a constitutional choice (XVI, 91). While Rawls' veil of ignorance is thus a hypothetical thought experiment designed in order to construct principles of

justice, Buchanan's veil of uncertainty aims to describe and guide the actual reality of constitutional choices. Accordingly, Buchanan models individuals at the constitutional level in a realistic way in that they are supposed to know something about the future (they are not completely ignorant) but do not have perfect foresight (they are uncertain) (XVI, 327).

Since this touches upon the crux of his views, I want to explore Buchanan's theory of constitutional choice still further by analyzing the criteria upon which such choices are made. First, there is the interest-component, which relates the expected outcomes to one's interests (XVI, 128). Second, there is the theory-component, which encompasses one's expectations about the outcomes that result from the chosen rules. Both components are important: "how a person chooses among potential alternatives is not only a matter of "what he wants" but also of "what he believes"" (XVI, 157). Two individuals with the same interests may disagree because of different expectations about the outcome. The same is true for two individuals with similar expectations but different interests. Buchanan thinks the theory-component is especially important in making constitutional choices. In his view, it explains why politicians discuss on the basis of arguments and not preferences (XVI, 157-158, 171). Accordingly, Buchanan believes that – as a political economist – he can and should contribute to the constitutional choice process by gaining and dispersing knowledge on the ways rules and institutions function (XVI, 169). As I have suggested, political economists should ultimately inform institutional design and reform (XVI, 353) and help citizens in their search for rules that enable them to better realize their goals (I, 466-467).

The context of constitutional choice places citizens in a framework that constrains their narrowly defined self-interest<sup>49</sup>. If they understand that determining general rules has long-standing consequences, they will be led by a prudent consideration of their self-interest. Sensible egoism will incorporate long-term concerns as well, which implies a move towards impartiality (III, 96). This allows one to understand the exact role of the constitutional veil in allowing individuals to reach a compromise between their interests (XVI, 166, 171). Contractarian theorists systematically argue that such a veil facilitates agreement by ensuring that the private interests of individuals are shared to a larger extent. Buchanan argues that the veil can be lifted to whatever degree one wishes. As it is lifted, individuals become more fully aware of their interests and agreement becomes harder to attain (XVI, 161-162). It can be

---

<sup>49</sup> In this sense, it resembles some kind of self-binding strategy (VIII, 163). Since the constitution limits everyday democratic decision-making, it might imply that the electorate does not get everything it wants (VIII, 191). Later on, I will go deeper into Buchanan's relation to public opinion.

made less transparent by making rules more general and stable in the long term. Ultimately, when the veil is completely opaque, it guarantees that every individual has the same interest in making constitutional choices (XVI, 129-130, 187). Whatever conflict remains is to be attributed to the divergences between the theories and expectations of citizens.

According to Buchanan, different strands of thought focus on different aspects of constitutional choice. First, social contract theorists – like Rawls and Buchanan – aim to converge the interests of citizens by introducing some sort of veil. This guarantees unity in the interest-component. Second, deliberative democrats – like Jürgen Habermas and James Fishkin – focus more on the convergence of the expectations of citizens. Through dialogue, individuals are able to discover the workings of different sets of rules and agree that a certain set of rules is just and fair for everyone (XVI, 131, 166-167). Rational discourse thus guarantees more unity within the theory-component. Whereas the former aim at agreement as compromise, the latter aim at agreement as truth-judgment. In the end, both traditions and strategies can complement each other (XVI, 156). Agreement on alternative institutional schemes becomes more likely as individuals know more about their general functioning and know less about the impact on their own situation (XVI, 176).

In contrast with Rawls, Buchanan does not require perfect and certain knowledge of the general workings of institutions. He realizes that individual expectations about institutional mechanisms can diverge widely (XVI, 287-289). First, among the obstacles on the road to agreement are the cognitive limitations of the human mind to grasp the complexity of institutional schemes. After all, it is hard to predict what the impact of rules will be in the long run. Second, there are motivational limitations of individuals to inform themselves of the ways different rules function (XVI, 131-132). Since the veil of uncertainty prohibits one from knowing what one's interests are, one has no incentive whatsoever to invest in such information (XVI, 187-188)<sup>50</sup>.

#### *4.1.2. Buchanan's view of the market*

Buchanan's individualism justifies a libertarian ideal: "the recognition of the necessary relativism and individualism of values (...) makes the libertarian

---

<sup>50</sup> To solve these problems, citizens might decide to select experts and entitle them to make decisions in their name. However, Buchanan argues that the problem remains that one has to become informed about (the views of) these experts (XVI, 136-138). Nevertheless, he believes that individuals act in their own interests when transferring their say in these matters to authoritative experts (XVI, 190-191).

social order meaningful” (XVII, 170). This means that the market is preferred to the state, because it spontaneously coordinates the interaction of individuals trying to realize their own goals. Buchanan thus endorses Adam Smith’s thought that markets have the unique capacity to coordinate egoistic individual actions (XVI, 443-444; XVIII, 266-267). It is best known as the “principle of the invisible hand, perhaps the major intellectual discovery in the whole history of economics, and upon which the normative precept of *laissez-faire* was constructed and defended” (XVIII, 311-312). According to Buchanan, this “central principle of classical political economy remains untouchable” (XVII, 243). It leads him to regard markets as the ideal way to guarantee social order while ensuring that individuals can maximally enjoy their freedom (XIII, 19). Buchanan thus praises markets for their ability to channel egoistic actions (private interests) into socially desirable results (public interest) (XII, 83-84; XV, 302-303).

Wholly in line with authors like von Hayek and Friedman, Buchanan thus favors the market over an inherently coercive government because it enables each and every individual to exercise his freedom to the largest extent possible (XII, 117; XV, 471; XVIII, 201-203). I will go deeper into this issue when discussing the normative implications of Buchanan’s Homo Economicus model.

#### 4.1.3. *Buchanan’s view of the state*

Nevertheless, Buchanan stresses that freedom is possible thanks to, and not in spite of, the governmental rules that are agreed upon by all citizens. They facilitate social life and allow people to interact without ending up in complete Hobbesian anarchy. According to Buchanan, freedom cannot exist in such a “warre of every man against every man” (Hobbes 1651: 188). It exists only within and because of an institutional context that defines and enforces individual rights (XVIII, 174-175, 205).

##### 4.1.3.1. Minimal protective and productive state

That is why Buchanan stresses the need for a protective state that does exactly this (XVI, 260). Governments and the rule of law they uphold protect citizens from arbitrary interference in their private spheres of action and demand respect for individual rights. In short, the protective state formulates and enforces a basic set of rules necessary to allow individuals to realize their goals

(XVI, 246; XVIII, 333). In Buchanan's view, such a basic politico-legal framework is the most important public good of all (XV, 25). Considering it is a necessary condition to make markets work properly (XV, 283-288), he believes that each government is better than no government (VII, 223-224; XVII, 273; Buchanan 1986: 139)<sup>51</sup>. He believes that a completely libertarian system of laissez-faire, where rules and constraints are wholly absent and where every political action is illegitimate, is a romantic fiction (XVI, 244-245).

Of course, the question remains how much government is best (XV, 3-4). Different forms of political organization can be analyzed as ranging from complete anarchy (all is decided by freely interacting individuals) to complete totalitarianism (all is decided by a centralized government without any concern for its citizens). As the title of one of his best-known books indicates, Buchanan situates the ideal society somewhere in between: "The Limits of Liberty: Between Anarchy and Leviathan" (VII). Later on, I will try to show that his ideal is closer to the former than to the latter (XVIII, 167-169).

In addition, the state also has a legitimate productive task, namely to produce, allocate and distribute those public goods that are agreed upon by its citizens. As I will show more fully later on, this emphasis on unanimity strictly limits the boundaries of legitimate collective decision-making.

#### 4.1.3.2. Contractarianism

Buchanan believes that governments should respect the sovereignty of each individual to determine what is good to the largest extent possible (XVIII, 89). This makes him a contractarian, which states that legitimate institutions should ultimately be based on the consent of citizens (XVI, 21-23, 41; XIV, 304). While this is implicit in voluntary market exchanges, it has to be made explicit at the political level. To understand politics as arising from voluntary exchanges between rational individuals, Buchanan makes use of the concept of a social contract (V, 79-80; IX, 4, 7).

---

<sup>51</sup> Even though this is often forgotten, Buchanan's libertarian precursors also understand the desirability of a basic politico-legal framework. Von Hayek (1944: 39), for example, acknowledges the need for a state in order to allow markets to function in a competitive way: "in no system that could be rationally defended would the state just do nothing. An effective competitive system needs an intelligently designed and continuously adjusted legal framework as much as any other". In this respect, he stresses the importance of the rule of law: "within the known rules of the game the individual is free to pursue his personal ends and desires, certain that the powers of government will not be used deliberately to frustrate his efforts" (von Hayek 1944: 73).

The purpose of Buchanan's contractarianism is not to explain how institutions have come about, but to evaluate proposals how to design and reform them<sup>52</sup>. It provides a justification of institutions by tracing them back to the agreement reached in rational discussions among free and equal citizens (XVIII, 386). Since coercion should ultimately spring from consent, Buchanan considers agreement among citizens as the only proper criterion to evaluate rules and institutions (I, 463, III, 248; IV, 248; IX, 15; XVI, 328; XVIII, 179). This immediately implies strict limits on the scope of legitimate decision-making: "beyond agreement there is simply no place for the contractarian to go" (XVIII, 181).

In the end, it is the application of the catallactic perspective of economics to politics that leads to a contractarian view of the state (XVI, 62, 68-69). In Buchanan's view, institutions are to be valued not to the degree in which they are efficient in some objective sense, but in which they embody the contractarian principle of agreement (I, 263-277)<sup>53</sup>. Buchanan's constitutional political economy combines the classical political economy of Adam Smith (rational individuals trade freely) with the contractarian political philosophy of Thomas Hobbes (rational individuals agree freely to a coercive government) (I, 387-389). In Buchanan's exchange paradigm, politics is nothing more than a complex process of exchanging individuals who want to reach their goals collectively. This allows him to analyze the state within a methodologically individualist framework (I, 122-123; XVI, 215). Both Buchanan's appraisal of the market and his contractarian view of politics thus find their origin in his insistence on individuals as the only relevant sources of value.

#### 4.1.3.3. Constitutional democracy

In his attempt to avoid referring to some public good independent of individual values, Buchanan stresses the need for democratic elections, since they ensure that policy measures reflect the wishes of the citizens (I, 93, 392;

---

<sup>52</sup> According to Buchanan, the social contract in its historical guise is a myth (XVI, 89). In his view, contractarianism only claims that citizens could have agreed on constitutional limits of government. It does not claim that such a social contract was ever drawn up or signed (XVI, 21).

<sup>53</sup> In his catallactic perspective on economics, Buchanan thus gives precedence to freedom over efficiency. In this respect, he also argues that private property has to be defended because it maximizes the former instead of the latter (XVIII, 216). Enabling the individual to engage in mutually beneficial economic relations, private property enhances individual freedom. Even though it is not fully efficient – why does each family have its own lawnmower? – it reduces mutual economic dependency (XVIII, 240-246).

IV, 285; XV, 353, 380; XVI, 9). He defends democracy as the only regime that considers each individual's values as equally worthy of consideration. He even regards the value of democracy as more important than that of equality and justice (XIV, 37).

Buchanan defines democracy "as a means of allowing individuals to express their own values, the only values that exist" (XVII, 270). In his view, democracy is nothing more than the aggregation of expressed preferences (IV, 143). Within the economic conception of rationality, it can indeed be expected that democratic politics is reduced largely to the aggregation of preferences into a collective preference ordering that has to be maximally satisfied. Nevertheless, Buchanan does not favor democracy in its simple majoritarian form. Aiming to minimize coercion, he not only opposes dictatorship, but is also aware of the dangers of majoritarian democracy, which allows a majority to discriminate a minority to its own advantage (XI, 38, 120; XVI, 39, 183). This way, it systematically tends to violate the liberties and interests of a minority as well (I, 423; III, 148; V, 150). Any majority rule inevitably implies that some individuals are able to use force against others (IX, 9-10). Since it allows for collective decisions on which certain citizens disagree, it inevitably constrains the liberties and interests of these individuals (XVI, 393) and thus no longer guarantees Pareto-optimality (IV, 289).

In addition, discriminatory policy measures are costly because they induce rent-seeking (XI, 123). It motivates individuals and interest groups to invest part of their resources in attempts to influence politicians to change the policies and laws to their advantage. This directs governmental decisions towards piecemeal interference in specific sectors and domains, which in turn leads to an inefficient use of time and money (XVI, 258, 341, 437). Through the creation of budget deficits and inflation, it eventually turns politics into a negative-sum game (XVI, 272, 424-426). In the end, it will result in complete arbitrariness, which will cause inefficiency to rise and the growth of the economy to slow down (XIV, 310-315). In short, rent-seeking in politics causes special interests to dominate the interest of the electorate as a whole (XVI, 268).

To avoid the exploitative discrimination that inevitably results from unconstrained majoritarianism, Buchanan urges that democracy has to be constitutional (XVIII, 259). A strong constitution that delineates, protects and enforces individual rights is needed to avoid such discrimination (XVI, 180). In this respect, Buchanan favors a constitutional democracy, stressing "that both words are important, with *constitutional* taking precedence over *democracy*" (XVI, 226). Because any democratic government has the tendency to favor one group over another, the constitution constraining it ought to have priority (XVI, 448).

Buchanan's emphasis on constitutional politics is completely in line with his individualism and gives more plausibility to his contractarianism (I, 50, 386; XIV, 24; XVIII, 189). Because it is easier to agree on a set of general rules than on a set of particular values, preferences or opinions, constitutional politics is more likely to provide win-win opportunities and thus to resemble market interactions (positive sum games). In contrast, post-constitutional politics tends to be rigged with conflict and therefore often results in win-lose (zero sum games) or even lose-lose situations (negative sum games) (I, 384-386; XVI, 249). As I have suggested, agreement is thus more easily attainable at the constitutional than at the post-constitutional level (I, 23, 238-239, 462-464; IV, 220-221; IX, 15; XV, 486; XVI, 83, 102-103). If certain rules are anchored constitutionally, there is less room for rent-seeking and conflicts between majorities and minorities (XI, 170-174). By default, constitutional politics is more inclusive than post-constitutional politics (XVI, 379).

#### 4.1.3.4. Unanimity and near-unanimity

Buchanan argues that the only grounds for evaluating institutional reforms are actual agreement, consensus and unanimity (which he associates with Vilfredo Pareto and Knut Wicksell) and hypothetical agreement in the Original Position (which he associates with John Rawls) (I, 430-432). Fully aware that unanimous agreement is hard to attain, Buchanan focuses on the constitutional level: "agreement seems more likely on general rules for collective choice than on the later choices to be made within the confines of certain agreed-on rules" (III, 78).

Buchanan defends unanimity as an actual decision-making rule in order to avoid the discrimination of majoritarian politics (I, 147-148, 205-206, XIV, 260; XVI, 220-221; XVIII, 322-323, 351). Giving each individual veto power, the unanimity rule effectively prohibits one group of individuals from forcing costs on another group through collective decisions (III, 201). Buchanan's claim that all governmental arrangements should be justified by unanimous consent solves the problem of majoritarian discrimination, since it ensures that everybody's liberties and interests are respected. If everybody agrees, the state does not have to coerce any of its citizens (von Hayek 1944: 60-69).

As such, the unanimity rule closes the gap between private interests and the public interest (XVI, 29, 74, 409). It guarantees that politics consists of Pareto-optimal rules, as it ideally should (I, 192, 229; III, 189). Indeed, the concepts of Pareto-optimality and Pareto-improvement stress each individual's voluntary agreement as crucial when evaluating a situation. Because they refer to



whatever individuals themselves deem best, they are based on minimal moral presuppositions. Good is whatever emerges from procedures in which each individual chooses freely (I, 244-245; V, 183).

Because individual values are the only ones that exist, unanimity serves as a conceptual test for evaluating possible reform proposals: "the Wicksellian test of conceptual unanimity offers the only defensible normative criterion for evaluating reform proposals" (XVI, 451). This directly results from Buchanan's individualist stance that individual values are the only ones that count (XVI, 10, 113). His focus on unanimity is a direct result of his application of economic tools to the sphere of politics: "market decisions are comparable to political decisions only when unanimity is present" (I, 102). Only if every party agrees, can there be voluntary exchange relations in politics similar to those in the market (XVI, 71).

Additionally, the unanimity rule is able to solve the problem of freeriding. With respect to the provision of public goods, every individual faces the incentive to freeride on the contributions of others (V, 10, 25). In order to avoid this, one has to design institutions that involve everybody in the agreement and make sure that enough individuals contribute. Requiring the voluntary consent of each and every citizen, the unanimity rule ensures that any individual's abstention leads to the breakdown of the public good as such. This way, it makes sure that its provision is linked to individual contributions (IV, 114-117; V, 83-88).

However, because of its large negotiation costs, the unanimity rule is of limited practical use (XVI, 318-319). According to Buchanan, deviations are legitimate if the decision costs involved are extremely high (XV, 462-463). Even though unanimity remains the ideal, less-than-unanimity rules are necessary to avoid complete conservatism (I, 226-227; V, 90-93, 146-147). Citizens can unanimously agree to install such rules in order to increase the efficiency of collective decision-making: "if general agreement is too costly and time-consuming, there is some rational basis for having political decision structures that do not require general agreement" (XVI, 319). According to Buchanan, one thus has to weigh the benefits against the costs of different sets of rules. The large decision-making costs of attaining unanimity may outweigh the costs of violating liberties (III, 93).

In this respect, Buchanan argues that there is nothing inherently desirable about simple majority (III, 127, 171). He proposes using different decision rules for different decisions (V, 148). He insists on unanimity with respect to constitutional issues but grants that less far-reaching matters can be handled with less-than-unanimity rules (XVI, 122). Of course, the more inclusive the decision-making rule is, the more its resulting decisions tend to serve

everyone's interests (III, 226). That is why Buchanan favors more inclusive rules than are currently present in actual politics (XVI, 223). He severely restricts the scope of less-than-unanimity rules, which are justified only if they increase overall efficiency<sup>54</sup>. Making it harder to change existing rules, unanimity increases their stability and thereby decreases rent-seeking (XVI, 270-273). In this respect, Buchanan calls himself a conservative: he thinks rules should have some stability in order to generate a desirable degree of predictability (XVI, 314-315)<sup>55</sup>.

In order to resolve the problem of majoritarian discrimination, Buchanan thus defends (near-)unanimity in constitutional politics. This way, political decisions automatically serve the goals of all citizens (XVIII, 110, 117). A simple majority rule will always generate (ex post) inequality, which violates the central principle of contractarianism and democracy, namely political equality (XVI, 219-220). If all political decisions are made with a rule of near-unanimity, the amount of public goods provided by the government and the amount of taxes levied by it will be quite minimal (V, 151-155). Hence, Buchanan favors a strong constitutional politics (where unanimous agreement is required) in combination with a minimal post-constitutional politics (where unanimity is no longer needed). Everything that is decided in such a state can be called legitimate (XI, 25).

As governments expand, they are increasingly likely to run into objections from some of their citizens. Starting from complete anarchy and construing an ever larger government, Buchanan argues that the size of government should be limited to the degree that the individual most critical of government will agree (V, 127). Once more, this thought is already present in the work of von Hayek: "common action is thus limited to the fields where people agree on common ends" (von Hayek 1944: 60). In this respect, Buchanan proposes several specific measures to reduce the state to a minimum.

---

<sup>54</sup> Buchanan even argues that this increases rather than constrains individual freedom: "when he adopts a rule and insures its enforcement, the individual is exercising his freedom, at a more comprehensive planning stage of choice, only through restricting his own freedom in subsequent potential choice situations" (VII, 190). Analogously, citizens agree at the constitutional level to refrain from certain actions in order to allow each other to interact without fear of domination.

<sup>55</sup> In the extreme, lobbying would be completely abolished if rules could not be changed (XVI, 323-324). Buchanan only uses this thought to illustrate his point, since he does not favor such complete conservatism. In contrast, he strongly believes in the perfectibility of society and its basic institutions and rules.

#### 4.1.4. *Buchanan's proposals to downsize the state*

First, he strongly stresses the need for a constitutional principle of generality, according to which “all citizens must be made to play by the same rules, and all politics must be non-discriminatory in its application” (I, 428). Essentially, it requires that political decisions must always apply equally to every citizen (XVI, 151, 275). Making discriminatory policy measures impossible, it effectively prohibits rent-seeking (XI, 51-54). Buchanan wants to move away from a politics by interest (a majoritarian democracy with rent-seeking and discrimination) towards a politics by principle (a constitutional democracy without rent-seeking and discrimination) (XI, 205)<sup>56</sup>. This would strongly curtail the extent of redistributive transfers, which are discriminatory by nature (III, 195-198; XIV, 42). This thought is already present in the work of von Hayek (1944: 78-83) whose emphasis on the need for the complete absence of discrimination leads him to put restrictions on the possibility of governments to enforce redistributive measures. Redistributive policy measures can only be legitimate insofar as they are based on altruistic preferences (I, 198; IV, 198). Here, redistribution will be voluntary and allows for win-win situations and Pareto-optimal results<sup>57</sup>. Whenever it turns into a win-lose situation, there is no longer a rationale for redistribution (XV, 22, 482, 487).

A second way of downsizing governments is through federalization and decentralization. In this respect, Buchanan argues that “where possible, collective activity should be organized in small rather than large political units” (III, 115). This makes the exit-option at the political level cheaper, thereby serving the ability of individuals to try out and evaluate different institutional schemes (XVI, 142-145, 296). The smaller governments are, the more competition there will be (IX, 197, 216; XVIII, 79, 81-82, 125-133). Next to the possibility of participating in elections (“voice”), this enables citizens to vote with their feet (“exit”) (XV, 35-36). It increases their freedom (XV, 476) and reduces rent-seeking behavior by making exploitation in politics practically impossible (XII, 449-452). Von Hayek (1944: 145) also thinks of decentralization as a way of minimizing governmental power: “to split or

---

<sup>56</sup> Stressing the constitutional level seems to lead Buchanan away from the assumption that everybody always acts egoistically (XIV, 43; XVIII, 286). I will go deeper into this issue later on.

<sup>57</sup> Buchanan argues that egoistic motivations often lead individuals to support redistributive measures as well (XVI, 97). Individuals who gain financially obviously experience some good. Individuals who lose financially can do so as well. This good can be private (the sensation of a warm glow knowing that they have helped someone) or public (the decrease in robberies or neglected neighborhoods) (XV, 483-485).

decentralize power is necessarily to reduce the absolute amount of power, and the competitive system is the only system designed to minimize by decentralization the power exercised by man over man”.

Another way of controlling government, even though it does not necessarily restrict its size, is to require its budget to be in balance (IX, 236; XIV, 382, 454; XV, 381; XVI, 58). This is needed for the simple reason that “elected politicians like to spend money and they do not like to impose taxes” (XV, 423). As I have already suggested, politicians in majoritarian democracies tend to further the interests of small groups with the money of the large group of taxpayers. According to Buchanan, democratic decision-making is likely to follow Keynes’ advice of increased public spending. This may lead to the creation of public deficits, which lower the perceived cost of public goods (VIII, 113, 144). This provides the perfect excuse for governmental officials to enlarge the bureaucratic apparatus (XV, 456). As a consequence, they will grab every opportunity to enlarge the state while leaving the citizens in the dark about the taxes they pay (XIV, 52).

## 4.2. Normative implications of the Homo Economicus model

Buchanan’s use of the Homo Economicus model leads him to favor the market over the state as the main allocation mechanism: “if individuals are assumed to behave solely in *homo economicus* terms, in all of their decision-taking roles, there would seem, on first argument a strong normative case for the widest possible usage of market and market-like institutions” (Buchanan 1986: 238). I will try to distinguish the arguments Buchanan employs in this respect by referring to the characteristics of the Homo Economicus model that I have outlined above.

### 4.2.1. Normative implications of instrumentality

The instrumental Homo Economicus model leads to an instrumental view of institutions as means to satisfy as much preferences as possible. Even though Buchanan himself claims that he does not want to analyze institutions as means towards the realization of some goal (I, 37-38)<sup>58</sup>, I believe his view of

---

<sup>58</sup> Buchanan himself uses the term “teleological” instead of instrumental (I, 257-258, 458-461). He does not want to analyze the market as a teleological process aimed at the maximization of efficiency, but as a creative process formed by free choices of individuals. According to Buchanan, the market does not maximize anything directly, but

institutions are rightly characterized as instrumental. After all, he explicitly defines them as “means through which individuals may co-operate to secure certain mutually desired ends” (III, 90). The main goal of institutions is to facilitate societal interaction and bring about social order (XI, 13). He thus thinks of the public domain as a cooperative scheme that exists only because it enables citizens to attain mutually beneficial agreements (VII, 88-90). Both the market and the state are instruments constructed by and for individuals in search of their own objectives (XV, 155). If public institutions no longer serve their interests, they will ultimately be replaced. They are nothing more than a means to safeguard each individual's freedom to pursue his own ends without interference of others (VII, 17-18).

Buchanan analyzes politics as a means to realize individual values and interests (XVI, 62). This applies to both levels of politics. At the constitutional level, individuals try to agree on a set of rules, rights and claims in order to create a form of “civil order” (XIV, 238) thanks to which they “can go about their own things in harmony and mutual respect” (XVI, 179). At the post-constitutional level, individuals try to make collective decisions that make everybody better off (VII, 66). At both levels, the main task of the state is to resolve conflicting interests and arrange social interaction in such a way that allows everyone to do what they want to (XVII, 161-162).

In Buchanan's ideal of ordered anarchy, there is a maximal amount of choice and a minimal amount of power of one individual over another (XV, 7-10). To approximate this, Buchanan assigns a large role to the market and a minimal one to the state: “that government is best which governs least” (VII, 117). Buchanan proposes the following division of tasks: “to the extent that markets work, there is no need for the state. Markets allow persons to interact, one with another, in a regime that combines freedom and order, provided only that the state supply the protective legal umbrella” (XVII, 243). The state should not replace the market, but only support and complement it when it fails. Wherever there is a decently functioning market, there is thus no need for a bureaucratic state that is inherently more coercive and costly (XVI, 254-259). Decentralizing power, the market is able to depoliticize certain domains (XVI, 227, 264, 354). Since he believes this will serve everybody's long-term interests, Buchanan strongly defends a constitutional politics of non-intervention (XI, 101-102, 109-111; XVI, 228, 381, 421).

Buchanan's ideal is thus situated between anarchy and Leviathan, be it closer to the former than to the latter (XVI, 27, 444). He argues that this will result from

---

simply allows individuals to satisfy their preferences (XVIII, 292-309). I maintain that this implies an instrumental view of the market as well, since its main characteristic is its ability to coordinate social interaction into mutually advantageous outcomes (XVI, 112).

a genuine constitutional choice process (XVI, 41). Because of the presence of decision costs, externalities and public goods, Buchanan thinks it is legitimate to move somewhat further – as little as necessary – towards a Leviathan-like state. He thus considers the state to be a necessary evil, wholly in service of the market (XVII, 244)<sup>59</sup>.

#### *4.2.2. Normative implications of egoism*

Buchanan's contention that citizens only resort to public arrangements if they find these profitable (III, 43) not only confirms the abovementioned instrumental view, but also follows from his assumption that citizens are generally motivated by self-interest.

##### 4.2.2.1. Political failures

Buchanan heavily criticizes the bifurcated image of individuals that is implicit in the works of most contemporary political philosophers, who assume that individuals tend to serve their own interests within the market and the public good within the state. Buchanan stresses that such a motivational dichotomy has to be argued for. He believes that motivational differences can arise within different institutional contexts, but suggests that these should be understood not as a strict dichotomy, but as positions along a gradual continuum (I, 68-69). As this chapter's opening quote already suggests and as I will show more fully later on, Buchanan continues to defend behavioral symmetry as the apt starting point for analyzing and evaluating institutional schemes, insisting on egoism as the overriding motivation in all domains of life. Individuals in their public roles as citizens, politicians and public servants are in this respect no different from individuals in their private roles as consumers and producers (IX, 8; XV, 33; XVI, 69). This view is shared by a number of theorists who do not strictly belong to the strand of Public Choice. Take, for example, Howard Margolis (1984: 7), who wonders: "if it is possible to gain insight into aggregate economic phenomena by exploring models that start with idealized individual

---

<sup>59</sup> Buchanan's preference for the market over the state also comes to the fore in an autobiographical piece, where he states that someone who "encounters the solid arguments of economic theory (...) *understands* that (...) markets tend to maximize freedom of persons from political control, that liberty, which has always been his basic value, is best preserved in a regime that allows markets a major role" (XIX, 167-168).

actors choosing rationally, why should not the same methods lead to fruitful results in politics? The short answer is that it does"<sup>60</sup>.

Since Buchanan does not believe that individuals suddenly become public-spirited in a political context, he vigorously criticizes the romantic view of the government as a benevolent despot, which was omnipresent in socialism (I, 180) and Keynesianism (XIV, 471). In fact, it was the orthodox view of the state in most of the works of modern political philosophy (IX, xvii). Instead, Buchanan defends a view of "politics without romance" (XVI, 439). In his view, "the romantic image of politics as the pursuit of public interest has been shattered, perhaps beyond repair" (XVI, 348).

Buchanan argues that governments do not always provide the ideal response to instances of market failure, because they often imply huge decision costs. While market failures provide a necessary condition for governmental intervention, they do not provide a sufficient condition (XV, xi). The plain existence of externalities is not sufficient to legitimize governmental intervention, because individuals can bilaterally decide to enter and enforce a contract (XV, 121, 184, 256). In this respect, Buchanan argues that most political philosophers neglect to show that the state is more efficient than the market when the latter fails (V, 186). Before transferring power to the state, one must first make sure that this is desirable (XI, 65). Because this is not the case in any systematic way, Buchanan stresses the importance of comparing different institutional structures with the hypothetical case in which all possibilities for mutually advantageous trade are exploited (XI, 56).

Buchanan thus deserves credit for pointing out the existence of political failures next to the better-known market failures (XVIII, 278). The former exist within undemocratic politics, which curbs individual freedom (XVI, 50) and within democratic politics, which produces discriminatory results. Even though the state can be a useful supplement to the market, Buchanan stresses that it does not fully get rid of negative externalities (I, 46, 63, 72-73; XI, 148-151). Since both the market and the state have to deal with problems of freeriding and externalities, Buchanan considers them to be second-best solutions (XV, 32).

---

<sup>60</sup> In fact, Margolis favors a more general conception of rationality than the strictly economic one that Buchanan does. He defends a model of individuals as motivated by both egoistic and altruistic motives. As I will show in what follows, I disagree with Margolis' claim that his model resembles that of Buchanan (Margolis 1984: 44).

## 4.2.2.2. The state as Leviathan

In Buchanan's perspective – which dominates Public Choice theory – egoism is the main reason why politics is doomed to fail. He argues that political representatives use whatever discretionary power they have to influence to their own advantage – and not that of the electorate – the laws and policies that are to be implemented (XV, 34; XVI, 107). Even though they may know what the public good consists of, they will not necessarily aim to realize it. Instead, they all try to keep as much as possible of the public resources for themselves (I, 104; IX, 163; XV, 362; XVIII, 285-288). Because it is in the interest of politicians and bureaucrats, governments have a tendency to expand. As the post-constitutional level of politics expands, negative effects arise and grow in number (I, 113-114, 186). After all, small groups with a lot to lose (or gain) have a larger impact on public policy than large groups with only little to gain (or lose) (XV, 396). Sectional interest groups increasingly influence elected politicians and their policies (XVI, 53).

Buchanan argues that politicians, bureaucrats and public servants are “motivated to expand the size and scope of the governmental sector of the economy” (VII, 201). This expansion will ultimately lead to a Leviathan-like state that becomes so large that it invades almost every domain of social life and systematically breaks its legitimate boundaries (VII, 151; XIII, 75). In this respect, Buchanan argues that the rise of the welfare state since the 1950s implies a move away from the ideal society and that the tendency to privatize and downsize public activities since the 1980s is a step in the right direction. He thinks of something like the United States of the 1950s as coming close to his view of the ideal society (III, 297).

Keeping in mind Buchanan's emphasis on the methodologically individualist reduction of collectivities to individuals, his model of the state as a revenue-maximizing Leviathan eventually comes down to the claim that people who work for the state act egoistically. Of course, such a state will systematically be deemed inferior to the market, which does not forcefully intervene in free exchange relations (XVIII, 68-69). Insofar as governments do indeed resemble a Leviathan, there is an obvious need for constitutional constraints on their range of action (VII, 203-204; X, 38; XIV, 178). Only if the state turns out not to be a benevolent despot, does the question arise to what extent it should be constrained (XVI, 108, 368, 445).



## 4.2.2.3. The need for constitutional constraints on the state

According to Buchanan, it is thus the egoism of citizens and politicians that forms an important reason to put constitutional restrictions on political decision-making (IX, 40-41; XIV, 443). Thomas Hobbes (1651: chapter XVIII) argues that this is simply impossible: the state is a Leviathan that cannot be bound, except by itself. In contrast, Buchanan thinks that a Leviathan can be constrained by a constitution that is agreed upon by all people (IX, 32-33; XVI, 47-48, 104-105, 447). The required “constitutional parameters of democracy” (XVI, 229), which Buchanan considers to be the most desirable institutional reform proposal, can be procedural (like unanimity or qualified majority rules) or fiscal in nature (like maximum tax rates) (IX, 182-196; XVI, 56-57, 310-311). Both are able to serve the purpose of restricting the domain of politics to its legitimate boundaries (IX, 9).

In order to avoid or stop the continuous expansion of the state, the constitutional level has to “specify in detail the operation and the limits of the productive state” (VII, 93). It has to keep a watchful eye and make sure that post-constitutional politics stays within its constitutionally agreed-on constraints (XVII, 53, 274). In order to achieve economic growth, social stability and individual freedom, strong constitutional rules have to restrict governmental power to a greater extent than they currently do (XVI, 278, 438-439). Buchanan's argument in favor of such constraints thus heavily relies on the claim that governments – or at least the people working for them – should not be trusted. In this respect, he situates himself in the tradition of David Hume, whom he often quotes: “in constraining any system of government and fixing the several checks and controls of the constitution, every man ought to be supposed a knave and to have no other end, in all his action, than private interest”<sup>61</sup>.

4.2.3. *Normative implications of exogeneity*

Buchanan argues that his entire constitutional project is primarily built on the assumption of exogenously given preferences: “if the individuals' capacities and objectives are given, the only way the pattern of outcomes can be changed is by

---

<sup>61</sup> It is significant that Buchanan systematically quotes Hume as having written that his goal is that of “constraining any system of government” (IX, 42; X, 68; XII, 87; XVI, 11). In contrast, Hume (1741: 42) himself explicitly wanted to address the issue of “contriving any system of government”.

alteration of the rules. And changes in the rules, obversely, will alter the outcomes that emerge from any society of individuals” (X, 19). Buchanan is a rational constructivist who aims to improve the rules within which individuals play the game of life. As Monty Python’s lyrics reveal, this is easier and faster than changing human nature, which cannot be reformed in a deliberate manner (XVI, 16; XVIII, 317-323).

This assumption leads Buchanan to propose institutional reforms without assessing their impact on the individuals and their preferences: “individuals are assumed to seek to maximize their own utility within the limits of the effective constraints imposed on their action. Not bringing the underlying motivational assumptions into question, economists tend, therefore, more or less automatically to think in terms of modifying the set of constraints on individual action” (III, 280). Later on, I will show that this provides an all too narrow view of the role norms, rules and institutions play in social life.

## 5. Criticizing and complementing Buchanan’s theoretical assumptions

### 5.1. Criticizing and complementing methodological and normative individualism

On a few occasions, Buchanan seems to question his own assumption of individualism. In his account of social norms, for example, he suggests going beyond a purely individualist model (X, 162-165). He nevertheless sticks to it and stresses that individuals remain the apt starting point for any explanation within the social sciences. However, this does not mean that Buchanan’s insistence on the assumption of normative individualism is without problems<sup>62</sup>. Considering consensus as “the only defensible normative criterion for evaluating reform proposals” (XVI, 451), Buchanan has to ensure that his own libertarian view of the ideal society is shared by a vast majority of fellow citizens (XVIII, 106, 117). If one takes seriously his normative individualism, it

---

<sup>62</sup> There are quite a few philosophical and moral problems with the moral relativism that seems to flow from the principle that no preference is of more value than any other. One might legitimately object to the moral claim – made by economists like Buchanan – that well-being is the same as the satisfaction of one’s preferences. After all, people often desire things that are bad for them (Broome 1995: 132). For now, I want to lay aside such criticisms that remain external to Buchanan’s project and focus more extensively on criticisms that undermine his theory from within.

is not Buchanan himself but each and every citizen who should, according to his own preferences, decide which tasks institutions ought to take on (XVIII, 72). In this view, there are no reasons for tearing down the state if not everyone agrees to do so. Ultimately, Buchanan is forced to ground his defense of a minimal state in popular opinion (III, 319-321).

This is why Buchanan argues that, "by widespread agreement, the state has become too powerful, too pervasive in its influence over private affairs" (VII, 19) and that preferences are in general evolving "toward individual freedom and away from constraints" (VII, 162). He even claims that everybody agrees with his conclusion that governments have become too large: "most welfare states are overextended: this diagnosis is almost universally accepted" (XI, 209). In other passages, Buchanan states that the growth of welfare states is not in line with the preferences of their citizens (IX, 28-30) and that people will be persuaded by his ideal of a minimal or at least limited state (XVI, 22).

Of course, the problem is that Buchanan's proposal of downsizing the state will inevitably leave some individuals less well-off and will thus not be welcomed by all. This not only applies to public servants but to citizens who benefit from the welfare state and its social security and pension schemes. I therefore refuse to believe that a vast majority of people would agree on the transition from an extensive to a minimal state. Buchanan's attempt to move beyond mere conceptual analysis towards actual reform fails because of his unanimity criterion, which effectively binds the hands of anyone in search of real institutional change. He too heavily relies on the criterion of Pareto-improvement, which is overly stringent, since changes will almost always produce winners and losers.

## **5.2. Criticizing and complementing the Homo Economicus model**

Despite his self-pronounced devotion to the Homo Economicus model, Buchanan from time to time moves away from it. Admitting that it has often been used as a descriptive model and defended on empirical grounds, Buchanan acknowledges the problem that it fails to explain quite a lot of actions (XII, 74-76). This explains his tendency to amend certain aspects of the model itself. Nevertheless, the revisions he proposes are only minor, since he never questions the model's basic structure. If taken seriously, these amendments can have far-reaching consequences. In what follows, I will discuss Buchanan's remarks in this respect and think them through in ways that Buchanan has neglected.

### 5.2.1. *Criticizing and complementing instrumentality*

In order to show the deficiency of an exclusively instrumental conception of rationality, I have already suggested that an expressive conception of rationality is possible and desirable if one wants to account for non-instrumental behavior like the following of norms. Even though Buchanan seems to situate the latter beyond a purely instrumental framework (I, 250), he never really abandons his instrumental conception of rationality (X, 163-165). In Buchanan's view, social norms stipulate what is to be done in a society and thus contribute to some shared sense of the public good. His emphasis on methodological individualism is crucial in this respect. Whether it concerns private or public interests, individuals are always out to further something they value highly.

### 5.2.2. *Criticizing and complementing egoism*

According to Buchanan, egoism is not necessarily a defining characteristic of the Homo Economicus model. He often argues that individuals want to satisfy their preferences without specifying what these are (III, 3). It is not so much the content of their preferences that is important, but the way these are structured. The individual is expected to fulfill whatever preference is ranked highest. In contrast with the narrow assumption of egoism, this assumption of egocentrism is consistent with Buchanan's normative individualism, according to which individuals are sovereign in determining what they value. Similarly, it is in line with his methodological individualism. Even if an individual prefers altruism above egoism, it is always his preference for helping others that provides motivational force and remains the central explanatory variable.

According to Buchanan, the most widespread "argument is that *Homo economicus* offers a better basic model for explaining human behaviour than any comparable alternative. Most modern economists would probably take this position" (X, 58). I blame Buchanan for not distancing himself completely from this argument, which becomes clear in passages where he stresses the need for economics to accurately reflect reality (III, 28-29; VI, 37-38; X, 35-36). Characteristic of Buchanan's ambiguity in this respect is his contention that his "approach starts with the empirical realities of persons as they exist, moral warts and all" (X, xvi).

However, Buchanan's remarks on the status of the assumption of egoism are ambiguous, to say the least. As he self-avowedly admits, "economists have great difficulty in moving beyond the rather simplistic, if powerful, models of

human behaviour grounded in self-interest motivation. We claim no exception to this generalisation about our disciplinary peers. Nonetheless, (...) we must introduce elements that violate the self-interest postulate" (X, 162). As I have already suggested, Buchanan seems to do exactly this by focusing on constitutional choice (XVII, 249). Placing the individual behind a veil of uncertainty, the constitutional context induces him to take into account other than purely egoistic considerations (III, 95).

Nevertheless, Buchanan does not think this through. His vocabulary of individual "interests, whatever these may be" (X, 59) suggests that he wants to allow for a wider array of motivations (whatever these may be), but also that an individual only does something when he expects to gain from it (what is in his interest). While admitting that not every individual is purely egoistic, Buchanan wants to leave "other possible motivations, such as genuine altruism, out of account" (XIII, 73). He also argues that an economic analysis consists in assuming that individuals maximize their utility, which means that they try to attain what they value highest (XV, 303-304). Whereas the first formulation seems to endorse the assumption of egoism, the second only endorses that of egocentrism.

In full awareness that the assumption of egoism amounts to an unrealistic abstraction (XII, 51), Buchanan thus holds on to it nonetheless, primarily because he believes that it is appropriate for his purpose of comparing different institutional alternatives (I, 392; XII, 70-71). In this respect, he strongly insists on the principle of behavioral symmetry. To compare different institutional arrangements, one has to use the same model of individuals that act within each of these schemes. The belief that people act differently under different institutions cannot simply be postulated but has to be argued for (X, 56-57; XII, 77-79). Buchanan wants to see how much of political decision-making can be explained by assuming that all preferences are egoistic (III, 20, 295-298).

As this chapter's opening quote clearly suggests, Buchanan uses the *Homo Economicus* model as a tool to derive "normative propositions about appropriate institutional design" (XII, 76). Together with Geoffrey Brennan, he states that, "like Adam Smith, we believe that *Homo economicus* remains the appropriate model of behaviour in the derivation of normative propositions about the institutions themselves" (XII, 86). Buchanan thus defends this model on conceptual rather than empirical grounds (X, 60-61; VII, 103; Buchanan 1986: 238). The basic argument is "methodological and analytic rather than empirical" (X, 58). It does not imply that men actually behave as the model predicts: "the battle over the empirical status of *Homo economicus* is *not*, in our view, the crucial issue at all" (XII, 86). He deems the empirical record of the

model to be largely irrelevant (XVII, 48-49). Even in a world where individuals are benevolent or public-spirited to some extent, Buchanan would favor institutions based on the assumption that they are not. Assuming egoism is justified even if – or even though – it does not occur in reality (XII, 98-99). Hence, Buchanan believes the Homo Economicus model is appropriate exactly because, and not just in spite of the fact that, it predicts outcomes that are worse than the actual state of affairs (XII, 106)<sup>63</sup>.

Buchanan thus argues that the degree to which individuals do indeed act egoistically is irrelevant when thinking about institutional design and reform. When signing a contract, one does not necessarily believe the other to be a crook. Nevertheless, one acts as if this may well be the case (XII, 79-81). Analogously, in trying to find out what the apt constraints on politics are, one should not naively assume that public servants always serve the public good (VII, 103-104; X, 60-61). Presupposing such a benevolent or public-spirited motivation would simply beg the question. Buchanan argues that it is safest to assume that public servants are egoistic, even though things may not be that bad at all (IX, xxiii). He thus does not consider the assumption of egoism as a depiction of individuals as they actually exist, but as they might exist. For the particular purpose of institutional design and reform, Buchanan fully agrees with Hume that it is better to assume that individuals are egoistic even though this may not correspond to actual reality (IX, 42; XVI, 11-12).

Buchanan's "sceptical or pessimistic view of human nature" (III, 27) represents the individual, not in his most common but in his most malicious aspects, which must be taken into account at all times. He argues that this view is appropriate regardless of whether it truthfully describes real-life individuals. Because one cannot be sure that everyone will act benevolently, one must assume otherwise in order to design appropriate institutional reforms. The counterfactual scenario is the relevant one: "what matters when we are considering alternative "constraints" is not what their consequences would in fact be, but what their consequences would be *if all men were knaves*" (Hausman 1998: 71).

For any criticism on Buchanan's assumption of egoism to have validity, one must show (1) that the empirical issue *is* the crucial one and (2) that the empirical record of the Homo Economicus model is abominable. With respect to (1), I want to stress that Buchanan himself believes that egoism is what

---

<sup>63</sup> In fact, the Homo Economicus model is not necessarily the worst possible scenario (Elster 1989a: 59). Buchanan admits that a society full of jealous or envious people may be even worse than one full of egoists (XII, 107). While I believe this poses an additional problem for Buchanan's conceptual defense of egoism – why assume that the others are egoistic when they might as well be jealous or envious? – I will not go into this here.

actually causes governments to fail. In order to be consistent, he should think of his Homo Economicus model as a somewhat realistic depiction of real-life politics (I, 45). Externalities arise only if individuals do indeed act in their self-interest. Politics should not be constitutionally constrained if public servants would voluntarily act in a public-spirited manner. In the end, Buchanan's argument in favor of constitutional constraints is based on the assumption that governments do in fact strive for budget-maximization and thus indeed consist of egoistic utility-maximizers.

Both Buchanan's analysis of the state as a Leviathan and his model of the individuals as *Homines Economi* are in need of empirical validation (XIV, 68). That is why he gives empirical arguments in favor of them and stresses that revenue-maximization is indeed the dominant motivation of most political actors (XIV, 153-155). He maintains that egoism has an important impact on behavior, even though it is not the only explanatory variable (IV, 195-196). In his view, it is sufficiently present to allow for economics to become a valid science (IV, 169).

Each proposal on the division of tasks between the market and the state depends on the way these function in reality and this, in turn, depends on the ways individuals behave. Buchanan himself argues that normative views should always be backed by empirically adequate underpinnings (II, 115; XV, 37). Ultimately, the validity and legitimacy of Buchanan's constitutional project hinges upon the empirical adequacy of his analyses of existing institutions and the individuals interacting within the latter. If Buchanan takes seriously his own claim that individuals matter, institutional reform should be based on adequate models of them. In this respect, he stresses that political economists should take preferences as they are and not as they should be (I, 194).

Normative issues should therefore not be decided upon the obviously false claim that citizens, public servants and politicians systematically act egoistically. Instead, institutional schemes should be adapted to individuals as they actually exist. If not everyone is a knave, institutions designed for a world full of knaves will not be optimal. The argument is as straightforward as it is simple: "if the outcomes of institutions designed for knaves are much worse than the outcomes designed for actual individuals and the odds of everybody being a knave were low, then it would be foolish to choose the institutions designed for knaves" (Hausman 1998: 74).

All this still assumes that maximal preference satisfaction forms the main goal of politico-legal institutions. One might also argue that a world full of knaves is the worst possible one and then argue in favor of some maximin principle. However, this is not a desirable criterion, since it "could condemn all mankind to stunted lives to protect against a one in a billion chance of a society of

knives” (Hausman 1998: 75). Robert Goodin (1996: 42) agrees: “whether our institutions should be designed with knives or with potential angels primarily in view depends crucially upon our views as to the relative frequency of each in the population, and as to the relative damage that will be done (or good that will be missed) by making one assumption rather than another given those frequencies”. Of course, it might be reasonable to adopt policies on a realistic estimate that some proportion of the population acts in a purely opportunistic manner. It makes sense to lock your door even if you think that only a fraction of all individuals are thieves. This, however, is not Buchanan’s strategy, since he wishes to reform institutions that matter to everyone on the basis of a model that fails as a generalized account of individuals. By taking into account the reactions of only the wholly opportunistic citizens, he risks making everyone worse off.

To summarize the first step in my criticism of Buchanan’s defense of the assumption of universal egoism, I want to stress that – in contrast to what he proclaims – his argument is not purely analytical or conceptual in nature. His arguments in favor of strict constitutional limits to keep post-constitutional politics in check are based on the implicit claim that individuals actually *are* egoistic in their public roles. Otherwise, the state simply would not fail or turn into a Leviathan. If one’s reliance on the assumption of egoism would be purely analytical, it would carry no weight whatsoever in normative discussions. For this assumption to have any relevance, it should at least be somewhat realistic or empirically adequate.

This brings us to the second step (2). Here, I want to refer to this dissertation’s first chapter, in which I have shown that the economic conception of rationality fails to explain individual behavior, not in the least because of its assumption of egoism. In this respect, it is important to see that Buchanan realizes that the egoistic Homo Economicus model does not have an outstanding empirical record. He admits “the possibility and indeed the likelihood of nonself-seeking behaviour by human agents in all institutional settings” (XII, 86). Buchanan even explicitly argues that individuals rarely act as Homines Economici (XII, 18). They often choose against their own self-interest, for example when they are motivated by internalized norms (I, 363; XII, 88-89; XVI, 377).

### *5.2.3. Criticizing and complementing exogeneity*

Buchanan admits that his assumption of exogenously given preferences is problematic as well, since the institutional context has repercussions on the



motivations of the individuals living in it (XII, 118, 434). In their role of consumers of public goods, for example, individuals tend to demand increasingly more of those public goods, since they are freely available. In their role of producers, however, the same individuals have to choose between alternative options to spend the collected tax money. Different institutional contexts lead to different results in that individuals demand more than they want to provide (XV, 347-350)<sup>64</sup>. This is why Buchanan stresses the need to pay attention to the social, cultural and institutional context in which the individual is situated (XVI, 371).

This criticism of the assumption of exogenous preferences is already present in his emphasis on unanimity in politics. After all, this is based on the very possibility of citizens persuading each other in public debates (XVI, 320). It ultimately hinges upon their ability to adapt their preferences and opinions throughout the decision-making process. If these were fixed, agreement would simply be impossible. His view of politics as “the process through which the initial preferences are expressed, discussed, compromised, and, finally, resolved in some fashion” (I, 236) clearly goes against the very claim of exogenously given preferences. He stresses that “the definition of democracy as “government by discussion” implies that individual values can and do change through the process of decision-making” (I, 99). Buchanan even claims that politicians should not take into account the preferences as they are given but as they are formed through rational discussion and public deliberation (I, 207). The assumption of exogenously given preferences leads to a purely aggregative model of democratic politics: “to take preferences as given allows us to recognize democracy’s contribution to the proper aggregation of wants through the counting of votes” (Bowles & Gintis 1986: 123). Here, however, Buchanan seems to move towards what might be called a deliberative model of democratic politics as a process of discussion in which preferences are not only expressed, but also transformed and selected<sup>65</sup>.

---

<sup>64</sup> As a solution for this imbalance, Buchanan suggests that governments should not only decide which public goods to provide, but also how much. If citizens want more than unanimously agreed on, they will have to pay for it themselves (XV, 356).

<sup>65</sup> Different models of democratic politics can be traced back to different conceptions of rationality. Whereas economic rationality leads to an aggregative model, a deliberative model is based on alternative conceptions of rationality. Whereas broad rationality stresses that beliefs should be well-informed and preferences should be autonomous, expressive rationality stresses that the individual should identify with such reasons: “dialogue politics is an alternative to interest aggregation politics. Under interest aggregation politics, actors are irredeemably egoistic and have essentially immobile opinions grounded in fixed interests. The essence of politics is reconciling interests to forge a coherent majority coalition. Dialogue politics, on the other hand, assumes that political actors seek the common good and that opinions can change drastically as a

In some passages, Buchanan seems to go deeper into this issue of endogenously changing preferences, for example when stating that individuals take on different roles in different situations: “persons do behave differently under differing institutions that place them in differing roles because the roles do modify the constraints and may also affect motivation” (Buchanan 1986: 239). He admits that institutional changes can affect individuals in the long run (IV, 175; XVI, 386) and that their preferences can be influenced by their previous choices (X, 85). Just like institutions, rules, values and norms, preferences are molded by one’s previous choices. Yet their malleability is limited, since they are typically quite stable (XVI, 230-231, 415)<sup>66</sup>. That is why Buchanan labels them “relatively absolute absolutes” (I, 445-447; XVIII, 322, 339).

This already suggests that Buchanan never fully explores this line of reasoning. His examples of preference formation (VII, 193) never lead him to genuinely recognize the fact that institutions influence the reasons upon which individuals base their actions. Ultimately, Buchanan continues to treat preferences as fixed at any moment in time. As with the issue of egoism, he is aware of the blind spot this implies, but is reluctant to do anything about it. Accordingly, he continues to analyze norms, rules and laws primarily as external constraints on one’s actions, not as factors that have a more internal influence on one’s preferences (X, 5).

Buchanan does not think through his own thought that the world can transform the individual living in it (X, 25-26). This would lead him to adopt a broader view of the individual and its place in society. Constitutional choice concerns not only the institutional structure one lives in, but also each and every individual’s identity. I will go deeper into this issue in what follows. For now, I want to refer to Geoffrey Brennan’s notion of “compound symmetry” (Brennan & Lomasky 1993: 13-16; Engelen 2008a; Eusepi & Hamlin 2004: 37-40). Whereas Buchanan assumes that individual behavior is perfectly symmetrical – namely egoistically motivated – in different institutional settings, Brennan allows for marginal changes in individual motivation and behavior. In

---

result of communication” (Mayer 1994: 270). Since rationality has instrumental, broad and expressive aspects, democratic politics should not be reduced to either its aggregative or its deliberative characteristics. The claim that the active engagement of citizens in such discussions is needed to speak of genuine democratic politics suggests that the latter should not be reduced to elections.

<sup>66</sup> With respect to values, this allows for a position between absolutism (my values are the only right ones) and relativism (no moral judgment about values is possible) (I, 451-452). With respect to rules and norms, this avoids the complete immobility that is the inevitable result when everything is always up for discussion. While it is possible to change laws at the constitutional level, there is always the current rule of law that treats the existing laws as given when interpreting and enforcing them (XVI, 127; XVIII, 399).

his view, one needs to start from a model of individuals as heterogeneously motivated by a whole gamut of considerations that gain more or less weight in different contexts. As should be clear from my analysis of voting decisions, I support the general thrust of Brennan's argument.

## 6. Normative implications of criticizing and complementing Buchanan's theoretical assumptions

### 6.1. Normative implications of criticizing and complementing methodological and normative individualism

The main problem with Buchanan's individualism is that it forces him to rely on public opinion when defending his own proposals to restrict post-constitutional politics to a bare minimum. Since his proposals to downsize the state fail his own test of unanimous agreement, he should either continue his search for reforms or conclude that the current state of affairs is best (XVII, 245-246). One way of avoiding absolute conservatism might be to relax the requirement of unanimity, for example by relying more on majoritarian democracy. One could argue that institutions and policies in current welfare states have come about in roughly democratic ways. Although they are not agreed upon by all citizens, they can be thought of as second-best solutions. In the imperfect world we live in, the best thing we can do is rely on democratic procedures to produce results that roughly reflect popular opinion. This seems to be in line with Buchanan's procedural conception of democracy, according to which whatever comes out of the democratic process is legitimate.

### 6.2. Normative implications of criticizing and complementing the Homo Economicus model

#### 6.2.1. *Normative implications of criticizing and complementing of instrumentality*

There are two problems an exclusively catallactic view of society has to deal with. The first is that a contract is rarely voluntarily fulfilled (even though mutual benefits could be attained by voluntarily upholding them), because the incentives to freeride induce each party to break it (XVI, 93). The second problem is that it tends to neglect that third parties can undergo negative side

effects of a contract. One could argue that the core of Buchanan's work in mainstream economics lies exactly in his discussion of negative externalities. Such costs occur when an exchange influences the situation of a third party without his permission (XV, 15-16, 110, 215). Both forms of market failure have to be weighed against the market's enormous advantage of regulating social life while leaving intact each individual's freedom (XII, 124-126). Nevertheless, they provide a justification for governments to intervene in market processes (XV, 17).

With respect to the first problem, Buchanan argues that there is a need for a protective state to enforce contracts and existing rules, which form the framework required for markets to function properly (XII, 476-477). Without this, the problem of freeriding would inhibit markets to coordinate interaction in mutually beneficial ways (I, 39, 350; XV, 30). In Buchanan's view, citizens will mutually agree to such a protective state in order to avoid a suboptimal outcome (I, 323-325; XV, 98-99; XVIII, 186-188). Even though the resulting coercive institution may use power, it ultimately receives its legitimacy from the consent of its citizens (XVI, 15; XVIII, 196).

With respect to the second problem, Buchanan stresses the need for a productive state to provide the public goods that markets fail to supply (XVIII, 121-122)<sup>67</sup>. Here too, he argues that citizens will agree on the necessity of eliminating negative externalities (XV, 19, 328) and the need for a coercive government to provide mutually beneficial arrangements (XI, 81-83; XV, 22). After all, negative externalities prevent markets from exploiting all possibilities for trade (XVI, 243) and thus from reaching an optimal equilibrium (XVI, 82-83, 112). The state should compensate those who suffer from externalities through side payments. Compensation allows the parties to attain a Pareto-optimal solution (XV, 120, 132-133) and facilitates consent on proposed reforms (XVI, 116-117, 375, 456).

Buchanan analyzes extensively the problem of public goods, which are characterized by their non-rivalry and non-excludability (IV, 11). A good is public if it is equally available to all individuals. Once it is in place, everybody can enjoy it without prohibiting others to enjoy it as well (V, 48). Because this creates a freerider problem – nobody will produce and maintain such goods if one can make use of them without doing so – markets cannot provide them efficiently. Since the market equilibrium is not Pareto-optimal in such cases, a

---

<sup>67</sup> In one passage, Buchanan argues that his defense of a governmental provision of public goods shows that he does not defend a minimal state (XVIII, 93). As I have suggested however, Buchanan never moves far away from such a minimal state.

coercive state is necessary to provide such public goods (V, 161; IX, 175; XI, 143; XV, 126)<sup>68</sup>.

Buchanan also analyzes more specific forms of freeriding. As I have shown, a first example can be found in the decision to vote. Why should I vote if I can enjoy the benefits of democracy while abstaining as well (IV, 121-124)? A second example occurs when citizens want their government to provide public goods without being willing to pay the required taxes (VIII, 130). Because they impede one's freedom and have a coercive impact on the individual's choices with respect to labor and leisure, Buchanan generally disapproves of taxes (XIV, 219, 327-329; XV, 135). However, he argues that the state can legitimately take what is needed to provide the goods that are agreed upon. Taxes are the flipside of public goods, just like prices are the flipside of consumer goods (XIV, 214). According to Buchanan, taxes up to 10% of a country's gross national product are enough for the state to fulfill its basic protective and productive roles (XIV, 228-234). Of course, this is a lot less than the 30 to 40% that is publicly spent in the United States (XV, 441). In most European welfare states, the total amount of taxes currently even runs up to half of the country's gross national product.

Once more, Buchanan's amendments to his own theory can have far-reaching consequences, if he would properly think them through. His claim that the state should complement the market in its failures allows for more governmental intervention than one might think at first glance. It enables the state to legitimately levy taxes to provide public goods like stability, order and law enforcement (protective state) but also like roads, schools, et cetera (productive state). In this respect, nothing can prevent governments from implementing far-reaching measures if there is enough public support for the public goods at hand. If citizens want clean air but the market is unable to reduce pollution, governments are justified to levy consumption taxes (for example on plastic bottles) or production taxes (for example on companies that do not comply with environmental laws). Both thoroughly intervene in the market process and its allocative and distributive results.

In a similar vein, Buchanan argues that the state should make sure that everyone follows the agreed-on rules and gets "a fair chance to play the game"

---

<sup>68</sup> Buchanan stresses that citizens ought to make sure that the money they pay on taxes is spent on public goods and not on politicians themselves (IX, 170). In this respect, he proposes that governments finance a public good rather than produce it themselves. This way, it can still be provided within the competitive environment of the market (XV, 34-35). Similarly, Buchanan believes that the monopoly of a bureaucratic government over certain areas – like social security or pension schemes – should be broken by allowing private organizations to compete (XV, 415).

(Buchanan 1986: 141). If this aim of equalizing opportunities is to be taken seriously, one can justify a much more generous state than Buchanan defends. The inequalities that arise through market exchanges can be thought of as the result of voluntarily chosen actions – a Nozickian procedural defense of the market holds up to this extent – but are continuously passed on to children who do not deserve this. If one wants to guarantee that every generation entering the socio-economic game receives an equal start, one has to design rules that interfere with market processes continuously and diachronically (Dworkin 2002: 85-88). Once more, one can see that Buchanan does not think through his own suggestions.

Regarding issues of distributive justice, Buchanan (1986: 151) distinguishes between inequalities in natural and social endowments. Although he mentions that one should correct for social bad luck, just like one does for genetic bad luck, he further neglects this possible justification for redistributive measures. An individual's environment functions not only as an external constraint – as Buchanan would argue – but influences the individual's capacities in more internal ways as well. Poor education might inhibit the development of a child's potentialities. Such an argument might justify the rectification of inequalities that are due to social circumstances as well. The thought that social inequalities should be taken into account as well can be applied to an individual's capacities as well as to his preferences, which are continuously formed by one's social and cultural environment (Rawls 1993: 269-270). The bottom line is that the distinction between social and natural endowments, as Buchanan's formulates it, ultimately cannot be upheld.

#### *6.2.2. Normative implications of criticizing and complementing egoism*

There is a fundamental ambiguity in Buchanan's reliance on the assumption of egoism. While he values it highly in the private domain, he wants to restrict it as much as possible in the public domain. The underlying thought is that egoism produces Pareto-optimal results in a competitive market, but systematically disrupts the proper functioning of governments. A first problem with this analysis is that – as examples of market failures show – egoism in private decisions does not always result in Pareto-optimal outcomes. Even if one accepts Buchanan's assumption of egoism, one can thus oppose his predilection for markets over governments. Because one individual's egoistic actions in market behavior can violate another's interests, politico-legal rules have been designed to prevent individuals from harming each other.

A second, more fundamental problem lies with the assumption of egoism itself. Since not all individuals are egoists, the ideal set of institutions and rules for the actual world does not coincide with the ideal one for a world full of knaves. To this, Buchanan might reply that constitutional constraints necessarily follow from any model that does not assume complete benevolence (IX, 226). However, this does not undermine the basic tenet of my argument, namely that the degree of such constraints should depend on the ways actual institutions function and actual individuals behave. If public servants act egoistically and excesses are inevitable at the political level, then constraints should be strict. However, if public servants act benevolently, then there is no need for constraints whatsoever. Indeed, the very idea of constraining government arises only if one no longer assumes complete benevolence on its part. Buchanan himself argues that the need for constitutional constraints is reduced significantly in a world of benevolent and public-spirited individuals: “moral restraint is a substitute for institutional-constitutional restraint, and in a society with more of the former there will be less need for the latter” (III, 303). As always, the truth lies somewhere in between. Excesses should be avoided and citizens need some guarantee that their government does not turn into an ever enlarging Leviathan. However, this does not mean that there is no room left for post-constitutional politics to impose policy measures in the name of the public good.

Of course, one should take into account the possibility of others being knaves when engaging in contracts or when leaving one's house. However, this does not mean that one should assume that all others are knaves, since that would imply that one no longer engages in contracts at all. While it makes sense to close your door, even if (you know that) not everybody is a thief, it is rational not to turn your home into a fortress. Nevertheless, this seems to be the only right thing to do if you actually assume that everybody is out to steal your possessions. Similarly, the fact that a constitution should indeed keep a watchful eye on politicians does not imply that it should confine their range of options to a bare minimum. Safeguard measures must always be proportionate to the extent to which they are needed, which in turn depends on the probability of individuals trying to take advantage of their power.

### *6.2.3. Normative implications of criticizing and complementing exogeneity*

Buchanan acknowledges that his defense of the market fails “if it might be demonstrated that the process itself exerts predictable effects on the character of the participants, effects that are generally deemed to operate to make participants worse than they would be under some plausible institutional

alternative” (XII, 124). Buchanan does away with this criticism by arguing that the disadvantages of the market are smaller than its advantages. In what follows, however, I want to argue that the negative effects of the market on the formation of the individual and his preferences are greater than Buchanan admits.

Buchanan argues that altruism survives better in the market than in a majoritarian democracy, because the latter does not have an exit-option like the former does (XII, 436-440). He defends the market as a mechanism that supports a morality of fairness (XII, 452) and that reinforces internalized moral codes such as honesty, reciprocity and respect for rights (XIII, 73-74). One could also argue that market relations induce and sustain mutual trust since they typically take place in small-scaled, personal and often repeated environments. The flipside of this argument is that cheating is rewarded in anonymous interactions like tax paying.

The basic problem here lies in the fact that Buchanan uses these arguments exclusively in favor of the market and against the state. First, Buchanan essentially neglects that it is the state that upholds rather than erodes existing social norms through the imposition of legal sanctions against misbehavior in social interactions. Second, the public domain is full of mechanisms, like public scrutiny and deliberation, which induce morally praiseworthy behavior. Incentives to be trustworthy and accountable are generally more pervasive in the public than in the private domain.

Third, the impact of institutions depends on the quality of their services. If state officials are seen to be abusing their power to their own advantage, the government will be distrusted, which will induce an egoistic reflex amongst its citizens. In Buchanan’s view of post-constitutional politics, expanding governments are partially responsible for the current decline of social capital, stability and order, since they have disillusioned their citizens. The latter will, as a result, increasingly behave as “moral anarchists” (XVII, 198) who treat others as means to further their own ends. If, however, governments effectively fulfill their tasks, I believe they will prevent rather than facilitate the deterioration of existing social norms. As I will argue more fully later on, I believe this discussion on the perceived legitimacy of public institutions should be decided empirically. In contrast, Buchanan thus relies on a worst case scenario (which is inappropriate) that he applies to the state but not to the market (which is inconsistent). He argues that the actions of self-interested individuals inevitably lead to an ever increasing Leviathan-like state. Even though there is growing dissatisfaction with the government, the bureaucratic apparatus continues to expand because of the egoism of public servants and citizens alike (XVI, 348-349, 397). In the end, Buchanan relies on motivations outside the Homo



Economicus model to start the constitutional revolution he is hoping for (XVIII, 182-184).

In my view, the political context is more likely to induce or support other- and process-regarding considerations than the market. The basic argument is that political decisions affect one's fellow citizens. Since market decisions influence only the parties to the contract (V, 139), the sensitivity for the consequences of one's choices on one's own situation will be large. As a result, market decisions will be primarily based on egoistic considerations (VI, 92). Because the impact of choices in a political context resorts primarily on the collectivity as a whole, one will be led to a greater extent by their consequences on others. As I have shown in my analysis of voting decisions, egoistic considerations tend to move to the background in an electoral context, which does not allow its participants to have a direct impact on the end result, let alone on their own situation<sup>69</sup>.

Buchanan explicitly wants to change the rules of the game in ways that have an impact only in general and in the long term (I, 220-221). In this respect, he favors lagged implementation, which means that the rules agreed upon now will be adopted later (IV, 302). Because of this enlarged time horizon, decisions will be made on a more impartial or "dispassionate" basis (XVI, 403). Egoistic considerations, which tend to dominate short-term decisions, will be replaced by considerations of fairness at the constitutional level and its veil of uncertainty (I, 239, 464-465; IX, 6, 24; XI, 8-9; XVI, 90; XVIII, 411). Because no player knows beforehand what his hand of cards will look like, all players want rules that are fair rather than those that would serve their self-interest (XVI, 45). One can see clearly here how Buchanan admits that a political context can bring about nobler motivations among its participants.

Conversely, I want to argue that it is probably the market and not the state that can be blamed for turning noble citizens into egoists. Because it rewards egoistic behavior, the market tends to bring about market-oriented and thus self-interested individuals. The market is not a neutral mechanism allowing individuals to do whatever they want to – as Buchanan sees it – but inevitably has an impact on their preferences and identities (Hodgson 1988: 173-179). As I will show more fully in the next chapter, it may be so that the measures to downsize government through liberalization and privatization and the

---

<sup>69</sup> As I have shown, Buchanan endorses the thought that one individual's decision in an electoral context has almost no impact on the final result, which is an aggregation of the decisions of all citizens. Therefore, he expects such a context to induce the individual to choose irresponsibly without paying attention to the consequences of this choice (Buchanan 1986: 230-234). However, my main point still holds: an electoral context can turn the individual away from his egoistic impulses. As I have suggested, it may induce citizens to express what they care about, regardless of whether this promotes their well-being.

increased importance of market relations induce more and more self-regarding preferences within people.

As people are placed in a particular context more often, they will identify with its corresponding role more easily. Individuals tend to behave in accordance with the context they are in and learn to act in this way, even if this context is no longer present. While this goes against the grain of Buchanan's claim that "persons do not become saints as they shift roles" (Buchanan 1986: 239), I think it can be made consistent with some of his own remarks. Buchanan, for example, acknowledges that different institutional contexts induce gradually differing motivations (I, 68-69).

To further explain this line of reasoning, I want to refer to two areas of research. The first is that of institutional economics, which criticizes the assumption of exogenous preferences by stressing ways in which preferences are formed by the institutional context (Hodgson 1988: 10-17; Dowding & King 1998: 1-5). It rightly argues that institutions form the main blind spot of conventional economics, which naively assumes that the institutions needed for markets to function properly would somehow arise spontaneously. Institutional economists also question the claim that a good society is one where actual preferences are maximally satisfied. Instead of taking preferences at face value, they tackle the problem of why preferences should be fulfilled. If the satisfaction of one's current preferences may make one worse off (Broome 1995: 132), it should be decided which preferences are to be satisfied and which not (Hausman & McPherson 1995: 263). Dropping the assumption of exogenous preferences thus opens up a range of possibilities at the normative level: "the assumption of exogenous preferences, while often necessary, limits many important policy and normative conclusions derived from the analysis. The reason is that when we write about the satisfaction of wants (...), it is hard to avoid the issue of the origin of the wants in question, and why these wants, as opposed to others which could as well have emerged from different initial conditions, should be satisfied" (Bowles & Gintis 1993: 100).

The second area of research answers the question how this selection should be thought of. Defenders of a deliberative model of democracy argue that individual opinions and preferences should first survive rational scrutiny before they are to be taken into account at the political level. Because participants must justify their position publicly, they are likely to give arguments and reasons that refer to the public interest (impartiality) instead of their private interests (partiality) (Elster 1998: 101-102). Here too, the influence of the context on its participants is not restricted to the short term. Through the psychological mechanism of cognitive dissonance reduction – the tendency of individuals to bring their actions and preferences into correspondence with

their utterances – this civilization of speech ultimately leads to a civilization of minds: “as representatives cannot publicly depart from the principled views they have expressed earlier, they may begin believing in what they say, even though they may not have been holding such views at the time they began expressing them” (Gosseries 2005).

A political context can thus bring nobler motivations to the fore. In general, this leads to a more generous view and positive appreciation of the role of politics. Even though one does not become an angel when placed in a political context, there is something to say in favor of a continuum in which individuals are more public-spirited in public life than in private life. Repeatedly having to make choices at a constitutional level may increase the inclination of individuals to actually become motivated by more impartial concerns. The aim of this section therefore was to combine Buchanan's constitutionalist perspective with the insight that individuals become who they are through the impact of social interactions, rules and institutions.

The fact that individual preferences are subject to change opens up possibilities for social engineering. As I have already shown, Buchanan is searching for “an institutional setting that operates so as to transform private self-interest into behaviour that is profitable to individuals other than the actors” (X, 61). He relies on the market's invisible hand to coordinate egoistic actions into social order. However, Buchanan's formulation leaves open the possibility of transforming egoistic motivations into nobler ones, which in their turn generate mutually beneficial actions and socially desirable outcomes.

Buchanan admits that such outcomes can be attained by changing individual behavior or changing institutions (XV, 306-307). In his view, ethics deals with the former question (what individuals should do), whereas political economy focuses on the latter question (what institutions should look like). Buchanan focuses exclusively on the latter strategy, because institutions can be molded more easily than human nature. However, his strict distinction between changing the individual and changing his institutional environment (III, 309; XVII, 200) is highly simplifying and even misleading (Cohen 2002: 378). Without replacing means for ends – and aim at better citizens as the ultimate policy goal – I want to take into account the fact that institutions affect citizens, quite like individuals change the institutions surrounding them. Both forms of causal influence have to be taken into account when one hopes to improve society. This forms an additional reason to endorse Ricoeur's definition of ethics that encompasses both the individual and the institutional issue.

One has to keep in mind that the institutional context also has an impact on social, cultural and moral norms (XII, 436). According to Buchanan,

institutions function as external constraints on individual behavior, whereas norms function as internal constraints (I, 364; XII, 359, 427-428). Defining the constitution in a way that includes political, legal and moral rules (XVI, 429), he stresses the importance of morality in creating social order. In contrast with Buchanan, who does not go deeper into this issue, because he thinks deliberate reform of norms is hard to achieve (XVI, 435), I think more attention should be paid to the impact of norms on individual preferences.

In Buchanan's view (VII, 168), there is some kind of paradox involved in trying to understand how norms can motivate people. Compliance is largely dependent upon enforcement, which is always costly: "enforcement has two components. First, violations must be discovered and violators identified. Second, punishment should be imposed on violators. Both components involve costs". How then can norms help people out of prisoner's dilemmas if they all have an incentive to freeride on the efforts of others to sanction potential violators? The enforcement of such norms itself constitutes a public good, since it provides benefits for the group. As such, however, it remains vulnerable to freeriding behavior (Boyd & Richerson 2005: 189). Relying exclusively on the egoistic disposition to avoid sanctions therefore shifts the problem of explaining cooperation to a higher level. It is indeed hard to understand punishment in a world where everybody is purely egoistic. In the next chapter, I will show which motivations do induce people to sanction freeriders.

Philip Pettit suggests that the solution lies in the fact that enforcement of norms is not costly at all, since it does not rely on the explicit imposition of sanctions. If one believes that others think ill of those who violate a socially shared norm and think well of those who comply with it, one will have a reason to comply with it. After all, "we care not just about the rebukes and commendations we receive from others but also about whether they take a negative or positive view of what we do" (Pettit 2002: 324). This very same belief is often sufficient for norm-violators to feel ostracized, even if nobody actually expresses his indignation. People thus do not intentionally have to identify violators and sanction them; it is enough for them just to be around. If norms are internalized by a large number of people, there is no need for costly incentive schemes for them to be effective: "social norms are powerful because rewards and punishments can be imposed at low cost" (Sober & Wilson 1998: 183). The mere possibility and threat of punishment are enough to induce potential freeriders to cooperate (Henrich et. al. 2004: 68).

According to Pettit (2002: 340), "norms are regularities such that nearly everyone conforms; nearly everyone approves of nearly anyone else's conforming and disapproves of his deviating; and this pattern of approval

helps to ensure general conformity". He shows that informal norms are compatible with more formal institutions that regulate behavior: "norms are an important species of social institutions, on a par with conventions, customs, laws, and other brands of established regularity (...). Like the other institutions norms reinforce certain patterns of behaviour, but they do so in their own way, by representing those patterns as peculiarly desirable or obligatory" (Pettit 2002: 308).

Buchanan admits that governmental intervention is needed to a lesser extent whenever informal norms are present (XV, 26-27). He stresses that norms are typically advantageous, for example because they can avoid negative externalities by inducing cooperative behavior (I, 357-358). A solid work ethic, for example, leads to greater efficiency and productivity without additional monitoring costs (XII, 374)<sup>70</sup>. If such norms are eroding – as Buchanan thinks they are – they should be replaced by more formal rules. The internal constraints norms imply should be complemented by the external constraints of politico-legal rules (XIV, 487-491). Buchanan does not seem to be aware that he is essentially justifying an expanding government that has to take on more and more functions that informal norms no longer fulfill.

## 7. Conclusion

### 7.1. Buchanan as an economist

If Public Choice is the application of economics to politics, it is important to understand what economics consists of. In this respect, Buchanan distinguishes between different levels of economic analysis (XII, 5-13, 128). The first is what he calls "the logic of choice" (XII, 4), according to which an individual simply chooses what is ranked highest in his preference ordering. Because such a choice is free, the individual's goals are not specified further. Buchanan equates this logic with the empirically empty rational choice theory, according to which utility is maximized and utility is whatever is maximized (XII, 32-33). The agent acts intentionally in order to achieve his goals, whatever these are (XII, 88). As one can clearly see, this does not correspond to the minimal conception of rationality as I have constructed it, since Buchanan immediately narrows it

---

<sup>70</sup> As I will show more fully in the next chapter, this analysis of social norms can lead one to stress the importance of a third form of institution. Next to the anonymous institutions of the market and the state, they emphasize the existence of communities where informal norms govern social life.

down to instrumental rationality. The second level goes even further by adding the assumption of egoism. In this “abstract science of economic behaviour” (XII, 6), free choice is transformed into predictable behavior by assuming only self-interest (XII, 72-73). At the third level, the “predictive science of economic behaviour” (XII, 10) makes empirically refutable predictions about the way people react to environmental stimuli. The fourth level, the “behaviouristic science” (XII, 12), incorporates insights from psychological research that point towards “motivational patterns that may be considerably more complex than the simple postulates of standard economic theory” (XII, 12). This may even lead one to completely drop the utilitarian framework most economists employ.

Buchanan does not want to restrict economics to the first level. This “general but empty” (XII, 8) logic of choice is of no use when proposing practical reforms. Most of Buchanan’s work can be labeled explanatory (second level) or predictive (third level). Among the latter is his work on market and government failures, in which he predicts to what extent the outcomes of egoistic actions within a certain institutional context will be Pareto-optimal (XII, 134-135). Surprisingly, Buchanan explicitly states that it is the fourth way that “offers the only avenue of advance for social science” (XII, 13). In this respect, he stresses that institutions influence behavior “that may be, but need not be, economically motivated” (XII, 15).

As I have shown, the problem is that Buchanan does not fill in the details of such a behavioristic science. The reason might lie in his conviction that models lose their explanatory power as soon as they start incorporating a broader array of motivations. In the end, they may no longer be able to function as a guide for solving normative issues, even though this was Buchanan’s main aim from the start (XII, 135-137). In this respect, one can roughly distinguish between two ways of thinking about economics. The first – economics as the science of freedom – mainly focuses on the freedom of individuals and therefore refuses to limit its analysis to purely egoistic considerations. The second – economics as situational determinism – analyzes how egoistic actions lead to different results as the institutions that channel these actions change. Both views have their advantages and disadvantages. While the first may be empirically more plausible, it risks turning into an empty theory, which is clearly crippled when it comes to making predictions about individual behavior and designing institutional reforms for the better. While the second view allows one to improve society by changing its rules, it is problematic to the extent that self-interest does in fact not provide the main motivation for most citizens. I want to argue that economics should steer clear both from an all too narrow model of individuals (second and third level) and an all too broad model (first level).

As a result, I agree with Buchanan who favors the behavioristic science of economics, which is modeled on the basis of empirical research (fourth level).

Whilst Buchanan does not consider the *Homo Economicus* model to be a completely accurate description of individuals, he still believes in its general empirical validity: “the *Homo economicus* construction *is* an abstraction from reality. Its purpose is that of allowing economists to impose intellectual order on the observed chaos of human interaction, without excessive distraction detail in dimensions of the analysis that are not centrally relevant” (XII, 73). In contrast, I think the *Homo Economicus* model fails to capture what it is to be a human being (I, 249). The fact that people realize that they are more than automatic pleasure-seekers is what distinguishes them from other animals. Endogenously changing preferences are characteristic for human beings, who have a sense not only of what they are but also of what they could be. People are to a certain extent free to become who they want to be. As Buchanan himself puts it, “we are, and will be, at least in part, that which we make ourselves to be” (I, 247). As I have argued, people can rise above their natural inclinations (I, 251-259). With Harry Frankfurt, I have stressed the importance of higher-order preferences that allow individuals to reflect upon their first-order preferences.

Even though science necessarily makes abstraction from the complexities of reality, it should never leave out the essence of its object. Economists explain human actions as the fulfillment of a one-dimensional set of preferences. In doing so, they leave out the very essence of man, namely his capacity to reflect upon his actions, his preferences and ultimately his life. My emphasis on the hierarchical account of human nature, in which an individual does not coincide with his bundle of first-order preferences, thus implies a criticism towards orthodox economics. Even though Buchanan from time to time seems to take a step in the right direction, he admits that his own thoughts have not “yet fully rid itself of the paradigms of neoclassical orthodoxy” (XII, 21).

## **7.2. Buchanan as a political economist and a political philosopher**

Buchanan strongly believes in the ability of citizens to change the rules and institutions of their society. In a democracy, people are not only subject to the regime they live in but they are also responsible for it. They do not only play the game of life, since they can partly determine its rules as well (I, 370). Buchanan claims that people should adopt what he calls a constitutional attitude, which consists in continuously questioning, evaluating and, if necessary, reforming the rules and institutions as they currently exist (I, 417-

418; IX, 240; XV, 16; XVI, 45-47). Trying to imagine what life would be like in a society with different rules and institutions, this attitude is really about having hopes in the future: “hope emerges for sustainable social order through the appropriate design, construction, and maintenance of *rules* that set limits on the way in which each person is allowed to order his conduct toward others” (X, xvi).

As I have already stressed, it should be taken into account that institutions do not only influence social interactions in an external way (constraining or regulating individuals and their actions) but also in an internal way (molding individuals and their preferences). Whereas Buchanan focuses almost exclusively on the first of these considerations, I have argued that both are indispensable for a proper understanding of the role of institutions in society. Joshua Cohen (2002: 380-381), who joins Rawls in his search for principles of justice, gives two reasons why institutions ought to be the main focus of political philosophy: “a first reason is that institutions play a large role in shaping economic and political outcomes, given preferences (...). The second reason is that institutions play a large role in shaping a society’s culture and the identity of members” (Cohen 2002: 381). Rawls (1993: 269) agrees that “the institutional form of society affects its members and determines in large part the kind of persons they want to be as well as the kind of persons they are”. This impact should thus be taken into account when thinking up desirable political institutions and policies. This brings me to the next point.

### 7.3. Buchanan in search of a better society

As I have shown, Buchanan combines a pessimistic belief in “the nonmalleability of basic elements in human motivation and behaviour” (XVI, 360) with an optimistic belief in the perfectibility of rules and institutions. The latter is what distinguishes him from conservatives and traditionalists who question the capacity of citizens to change the world (XVI, 177, 358). Buchanan is convinced that individuals are able to shape their own future (XVI, 415). He strongly believes that they should reform the world as it actually exists: “any proposal for change involves the status quo as the necessary starting point. “We start from here,” and not from some place else” (VII, 101). Since he wants his advice to have real-life implications, Buchanan shifts his attention from institutional design (how to create an ideal society ‘ex nihilo’) to institutional reform (how to improve current society) (XIV, 21). He argues that not every conceivable social outcome can be attained, since individual reactions to situational changes are limited. These “uniformities of human nature” (XII, 50) lead to freeriding behavior, which makes it practically



impossible to choose in a deliberate way a particular social state, even if everybody agrees that it would be desirable (XII, 34-41; XVI, 204, 208, 241). Even though the latter is physically possible, it does not lie within the domain of what Buchanan calls "behavioural feasibility" (XVI, 206) and is therefore unattainable in practice (XVI, 352). This seems to correspond perfectly with the abovementioned search for a realistic utopia in which one takes individuals as they are and institutions as they might be.

Even though his primary goal is normative, Buchanan stresses the importance of adequate theoretical models: "political economy will produce useful normative propositions only to the extent that its analytical underpinnings are correct" (II, 115). Before choosing among different sets of rules, one should first understand how these work in practice (XIV, 27; XVI, 105). Buchanan thus systematically combines the descriptive and normative aspects of his constitutional political economy (XVI, 335). This is already present in his definition of political philosophy as thinking about what politics is and what it should be (XI, 20). Since both markets and states fail in some respect, he insists that a criticism of the former should always be supplemented with an argument of why the latter will not fall into the same pitfalls. However, this also works the other way around. Before abolishing or reducing governmental activities, one should always compare it to other realistic alternatives (III, 209; XV, 50, 157).

In Buchanan's view, each individual is entitled to his own conception of the ideal society. As a consequence, it is very unlikely that all individuals will agree on a single ideal. In this respect, politics differs from science (I, 240-242). Since well-informed individuals can ultimately disagree at the political level, it is misleading to speak of truth here (XVI, 175). Because there is no external point of view from which an omniscient being can evaluate things, there is simply no truth to discover, independently from individuals and their preferences (XVI, 118-120)<sup>71</sup>. This is in line with Buchanan's contractarianism, in which the focus is not so much on the results of political decision-making – as it is in the truth judgment conception of politics – but on the process itself (XVI, 366-367).

Starting from society as it exists here and now, Buchanan wishes to improve things by analyzing which institutional reforms would be agreed upon by all (XIV, 255; XVI, 98-99, 372-374, 410-411; XVIII, 72, 86). If no proposal is unanimously agreed upon, he has to conclude that the current institutional

---

<sup>71</sup> With respect to tax reforms, for example, Buchanan argues that "there is no "best" tax system in terms of evaluation by external criteria for the simple reason that there is no single set of criteria upon which agreement exists or can exist" (XVI, 317). This is exactly why Buchanan wants to evaluate tax schemes "on the basis of the desires of the taxpayers themselves" (XVI, 317).

structure is Pareto-optimal. However, this does not mean that the status quo is the best of all possible worlds (XV, 390; XVI, 114, 451). Even though there is automatically a bias in favor of the current situation (XVI, 389), Buchanan does not consider it to be free from faults. Accordingly, he refuses to treat current institutions as sacrosanct (XVI, 49). In order to move “from what we have to what is proposed” (XV, 418), Buchanan favors “a continuing critical examination of the institutions that we observe and which should be designed to serve our objectives” (XV, 420).

In essence, this is what his constitutional approach is really all about, namely “concentration on the institutions of social order and on reforming these institutions as opposed to simply going out and looking at what would be ideal and discussing the ideal type of policy without recognition of the institutions of social order” (XVI, 102). It is therefore not so much the ideal that is crucial, but the way of getting there. If a certain ideal is unattainable, because of previously made decisions or because of the relatively fixed boundaries of human nature, it is important not to become completely immobilized. In this respect, Buchanan gives the example of a marriage gone wrong. At some moment in time, one might come to think that it would be ideal if there never were a marriage. Be this as it may, one should still start from the situation as it exists and try to think what steps would improve the condition of both partners (XVI, 449). My main criticism towards Buchanan is that he focuses exclusively on the rules that govern the actions of both partners and not on the fact that the partners themselves can change. This basic insight may not only explain the problem, it might open up possibilities to solve it as well.

#### **7.4. Summing things up**

Buchanan’s emphasis on rules, institutions and norms is based upon his insistence on preferences as the starting point of analysis. He judges it better to redesign the rules of society than to try and change its members: “good games depend on good rules more than they depend on good players” (X, 167). Buchanan suggests that Monty Python’s opening quote is misleading. After all, there is no real dilemma here. It is precisely because we are “simply spirally coils of self-replicating DNA” that we should focus on the game of life “where we make up the rules while we are searching for something to say”. Buchanan might thus argue that it belongs to the genetic constitution of humans that they will ultimately serve their self-interest. In the following chapter, however, I will show that other motivations – like altruism and public-spiritedness – have evolved through processes of natural selection as well.

In contrast, I have tried to show that in the game of life the players are at least as important as the rules. In this respect, I have argued in favor of a richer and more adequate model of the individual. To be sure, situating it in a context that shapes its identity is not necessarily at odds with methodological individualism. After all, the latter only claims that the individual is the ultimate – and not the only – explanatory variable and thus does not imply “a refusal to examine the institutional or other forces which are involved in the moulding of individual preferences and purposes” (Hodgson 1988: 54). Stressing the individual's ability to reform institutions, I believe the individual remains – and ought to remain – central when explaining what goes on in society.

Because I do not think that Buchanan's analysis of institutions (as externally channeling individual actions in mutually beneficial ways) is entirely without value, I do not want to replace it completely. While the politically economic approach to politics – changing institutions – complements that of social engineers – changing individuals – both have their blind spots. Each focusing on different aspects, they have to be combined if one wants to understand and improve society.

My main criticism is thus that Buchanan, who aims to change the institutional structure as it exists here and now, does not start from the individual as it exists here and now. I have tried to show that his reliance on the Homo Economicus model draws him into the pitfall he himself warns against, namely “to advance proposals that fail to recognize man for what he is” (XVIII, 323). This is why I have argued in favor of a richer and empirically more adequate model of the individual. In this respect, I will focus more fully on the insights gained by authors like Samuel Bowles and Herbert Gintis.



## CHAPTER 6

### THE ALTERNATIVE OF SAMUEL BOWLES AND HERBERT GINTIS

---

*“If all that really matters is that individuals should satisfy as many as possible of their desires in the order of their strength, then it matters how they live together, what kind of community they form, only to the extent that their living one way rather than another enables them to get more of what they want at the cost of less effort (...). If, however, we suppose that men want more than to succeed as much as possible in satisfying one desire after another, their own or other people’s; if we suppose that they also care what sort of persons they are (...), what kinds of life they lead, it is easy to see how they come to be attached to the community as much as to one another. How men see themselves, as they are or as they would like to be, is intimately connected with their mental images of the community; they are not mere competitors and collaborators, however benevolent, in a market for the supply of personal wants; they are members of society, and their hopes and feelings, both for themselves and others, would not be what they are apart from their group loyalties”*

(Plamenatz 1966: 175)

#### 1. Why study Samuel Bowles and Herbert Gintis?

##### 1.1. Who are Samuel Bowles and Herbert Gintis?

Samuel Bowles and Herbert Gintis are among the most important heterodox American economists of the past decades. Their academic lives largely overlap and therefore show remarkable similarities. While Bowles received his doctoral degree in Economics from Harvard University in 1965, Gintis did so at the same university four years later. Since then, they have frequently cooperated in an impressive number of projects. Currently, both are Professor Emeritus at the University of Massachusetts. In addition, they are connected to the Behavioral Sciences Program at the Santa Fe Institute in Santa Fe, New Mexico. While Bowles and Gintis are economists, they both have a considerable interest in issues situated at the borders of economics, sociology,

evolutionary psychology, political science and philosophy. Whereas Bowles has taught numerous courses on microeconomics and the role of institutions in economics, Gintis has taught primarily on game theory and collective action problems.

Their intense collaboration has resulted in co-authored monographs and articles, co-edited books and jointly organized conferences and projects. Much of their research is carried out in conjunction with other members of the Santa Fe Institute and the MacArthur Research Network on Preferences. The latter is an interdisciplinary group which includes – besides Bowles and Gintis – authors like Robert Boyd, Ernst Fehr, Simon Gächter, Joseph Henrich, Dan Kahan, Daniel Kahneman and George Loewenstein, who have jointly conducted theoretical as well as experimental research. Since its inception in 1997, this Preferences Network aims to develop models of individual decision-making that can be used for policy purposes. It focuses on the evolution of cooperation, trust and prosocial behavior in general. I hope to show why and how its insights are indispensable to gain a thorough understanding of the central themes of this dissertation.

## **1.2. Why study Samuel Bowles and Herbert Gintis?**

The research conducted jointly by Samuel Bowles and Herbert focuses on a number of topics, all of which relate to this dissertation's topic of the normative implications of rationality. For example, they show how non-egoistic motives are crucial to explain individual behavior and how institutions influence their relative importance. Bowles and Gintis focus explicitly on the mutual dependence of a society's members and institutions. Not only do they describe and explain the different causal links between the individual and the institutional level, they also look at the normative conclusions one can derive from all this: "structures change because people change them. Making good the promise of the democratic accountability of social change demands an understanding of why and how people change or fail to change the rules that govern their lives. This understanding in turn requires a more searching treatment of individual choice and collective action" (Bowles & Gintis 1986: 120). Bowles and Gintis thus aim to address issues at both the explanatory level and the normative level. The close relation between explanatory models of individuals and normative views surrounding institutional design has been the focus of their work from the very beginning.

Bowles and Gintis are convinced that the crucial challenge of improving the basic institutional structure of modern societies can only be addressed properly

through a better understanding of how people behave and how their interactions produce social outcomes. Until now, the behavioral and social sciences have failed to adequately take on this task, partly because of the highly fragmentized nature of contemporary academics. In this respect, Bowles and Gintis subscribe to the argument – which is crucial for the general line of reasoning in this dissertation – that normative issues should be based on empirically adequate models of individual behavior and motivation. In what follows, I hope to show how both their explanatory models and their normative views diverge from those of James Buchanan.

To be sure, Bowles' and Gintis' primary goal is the normative one of improving society by enhancing people's capacity to govern their own lives (Bowles & Gintis 1986: 3). It is this focus on normative issues that distinguishes them from most of the other authors of the Preferences Network, whose work is mentioned here when relevant but does not form the center of attention. Bowles' and Gintis' basic ideals are those of democracy and popular sovereignty, according to which power should be accountable to those affected by its exercise (Bowles & Gintis 1986: 209). This applies not only to the political domain (the relation between the state and its citizens) but also to the economic domain (the relation between employer and employee). In order to make these more democratic and improve the lives of people, Bowles and Gintis focus on the shortcomings of the conventional economic models that dominate recent political philosophy: "we will point to the poverty of virtually all strands of economic theory as a basis for thinking about democracy, illustrate the unfortunate consequences of the dominance of economic metaphor in our political and moral thinking, and demonstrate the need for integrating genuinely political concepts – concerning power and human development – with economic reasoning" (Bowles & Gintis 1986: 7).

Bowles and Gintis aim to show that a number of theoretical assumptions in the Homo Economicus model are grossly inadequate to explain actual human behavior in a large number of cases and settings. After elucidating their intellectual background, I want to show in what ways they criticize and amend the assumptions of the economic conception of rationality when modeling individual actions and reasons. Next, I derive some normative implications from all this. Finally, I draw a number of conclusions.

## 2. Bowles' and Gintis' intellectual background

Because they aim to develop a more accurate model of individuals than the economic one, Bowles and Gintis have close links with behavioral economics.

Here, models are explicitly constructed on the basis of empirical research that focuses on the (lack of) rationality of individuals. Since they heavily stress the role institutions play in both micro- and macroeconomics (Bowles 2004: 6-7), they can be labeled institutional economists as well. In this respect, both Bowles and Gintis clearly qualify as heterodox economists, who are critical of their mainstream colleagues. They want to revive political economy by analyzing economic and political behavior on the basis of assumptions that do not fit orthodox economics. In doing so, their Marxian roots clearly show, even though these are more straightforwardly present in their earlier works than in their more recent articles and books. Instead of analyzing the individual in an atomistic way and the economy as some vacuous equilibrium, Bowles and Gintis show the importance of power at both the microeconomic and macroeconomic level.

Bowles and Gintis state that their work brings together three strands of thought: radical democracy, liberalism and Marxism. With their “political critique of economic thinking” (Bowles & Gintis 1986: x), they show in what sense liberalism – which encompasses neoclassical economics – lacks the concepts to understand modern social life. Because of its focus on the voluntary nature of exchange relations, liberalism neglects power relations in both the private and the economic domain (Bowles & Gintis 1986: 65-71). Essentially, Bowles and Gintis argue that markets are not simply neutral mechanisms that allow free individuals to voluntarily engage in mutually beneficial arrangements on an equal footing. In this respect, they are highly critical of “the conviction among social scientists and political philosophers that as an arena of voluntary contractual exchange, the capitalist economy is devoid of political content” (Bowles & Gintis 1992: 324). This immediately suggests that their views diverge widely from those of Buchanan. Bowles and Gintis refer to the conventional economic theory of markets as the Walrasian model, according to which power and coercion are absent if competitive markets are in equilibrium. After all, as Buchanan stresses, markets are characterized by free and voluntary exchange relations between its participants.

If one thinks through the neoclassical assumption that individuals are narrowly self-interested, one should conclude that the parties to a voluntarily closed contract will tend to take advantage of each other. Here, the Walrasian model assumes “that exchanges between agents in the economy can be enforced by a third party (for example, the judicial system) at no cost to the exchanging parties” (Bowles & Gintis 1992: 328). This has led economists to interpret the market as the locus of voluntary and costless exchanges and the state as the locus of power. Bowles and Gintis (1993: 85) refer to Buchanan as the preeminent defender of this neoclassical textbook view of both basic institutions. After all, the main function of Buchanan’s protective state was



exactly to enforce a politico-legal framework that allows market exchanges to run smoothly.

In what follows, I try to show how Bowles and Gintis diverge from Buchanan's views of markets and states by focusing on their notion of contested exchange, which stresses that contracts are not always perfectly enforceable: "consider agent A who purchases a good or service from agent B. We call the exchange *contested* when B's good or service possesses an attribute that is valuable to A, is costly for B to provide, yet is not fully specified in an enforceable contract" (Bowles & Gintis 1992: 332). In such cases, the enforcement of the contract is typically endogenous, which means that the parties to the exchange monitor and sanction each other. If there is no third party to enforce the contract, conflicting interests will give rise to a struggle over the terms of the contract itself. In relations where exchange is contested, "unlike the transaction of Walrasian economics, the benefit the parties derive from the transaction depends on their own capacities to enforce competing claims" (Bowles & Gintis 1993: 85).

The relevance of such endogenous enforcement mechanisms suggests that power is relevant in market relations as well. As a consequence, the conventional strategy of equating the market with freedom and the state with power no longer holds (Bowles & Gintis 1992: 353). This shows the urgent need to extend democracy to the workplace as well. The basic reasoning is quite simple: since the employer exercises power over the employee, the former should be held accountable, exactly like politicians should be held accountable for their use of power over citizens. Essentially, Bowles and Gintis (1993: 98) question the validity of the first theorem of welfare economics, according to which any equilibrium in perfectly competitive markets is Pareto-efficient. They argue that such markets do not always clear in equilibrium. Because of enforcement costs, "contested exchange markets (...) generally fail to implement socially efficient resource use, in the sense that there exist transactions that are Pareto superior to the competitive equilibrium" (Bowles & Gintis 1993: 88).

The Walrasian paradigm that dominates welfare economics is based on two assumptions, both of which Bowles and Gintis deem fundamentally wrongheaded. The first assumption is that "contractual claim enforcement is executed at zero cost and hence may be considered exogenous" (Bowles & Gintis 1993: 98). The state is supposed to enforce contracts without any cost to the exchanging parties themselves. Bowles and Gintis have shown that this is problematic by stressing that exchanges are often contested and, hence, that enforcement is often endogenous to the exchanging claimants. The second assumption is that of exogenously given preferences that stipulate the goals for

the individual to realize: “the goals are determined by individual preferences (whether selfish or altruistic), and thus action is instrumental to the satisfaction of pre-given wants” (Bowles & Gintis 1986: 20). As I have shown in my analysis of Buchanan, this leads to a view of institutions as guiding the actions of wholly formed individuals rather than forming these individuals. Accordingly, institutions “are evaluated primarily on the basis of their ability to record, aggregate, enforce, or satisfy preexisting interests. The manner in which institutions engender preferences and interests is thereby obscured” (Bowles & Gintis 1986: 22). Bowles and Gintis thus propose to move away from the instrumental conception of action aimed at fulfilling exogenously given preferences. As I will show later on, their alternative essentially constitutes what I have tried to capture in the expressive conception of rationality.

If contracts are complete and exogenously enforced, conventional economic models predict and explain individual behavior quite well (Ben-Ner & Putterman 1998: 337). However, if contracts are incomplete – and this is often the case – a lot of people act in ways that do not serve their self-interest: “most social relations are not governed by explicit contracts but by implicit informal agreements. Moreover, when explicit contracts exist they are often highly incomplete, which gives rise to strong incentives to shirk” (Henrich et. al. 2004: 74). Take the labor market, where the promise of the employee to provide some amount of work effort typically cannot be externally enforced: “work is subjectively costly for the worker to provide, valuable to the employer, and costly to measure. The manager-worker relationship thus is a contested exchange. The endogenous enforcement mechanisms of the enterprise, not the state, are thus responsible for ensuring the delivery of any particular level of labor” (Bowles & Gintis 1992: 333). If contracts are incomplete, external control mechanisms will not suffice and intrinsic motivation will start to play an important role. If employees identify with the goals of their company and participate in its governance, they will develop an intrinsic motivation to work. Consequently, surveillance costs will be lower (Bowles & Gintis 1986: 78).

Clearly, the assumption of egoism is under pressure here. Even though conventional economists think that relations on the work floor are typically governed by self-interest, research has shown that both employers and employees act on the basis of internalized norms of fairness and honesty, even when they know that opportunism would serve them better (Gintis 2003: 164; Gintis et al. 2005: 15). In addition, Bowles and Gintis (1993: 88-89) also stress that contested exchange leads one to question the assumption of exogenous preferences: “while the assumption of exogenous preferences strains credulity in the Walrasian model, it is simply incoherent in a model of contested exchange. Unlike the Walrasian model, where agents are “endowed” with preferences that they then take to market, contested exchanges shape the

character and consciousness of the exchanging parties (...). Because the exchange is durable and personal, the exchanging parties have an interest in shaping the structure of the transaction to mold the personalities, objectives, and other characteristics of the other parties to the exchange”. Dropping the central Walrasian assumptions, Bowles and Gintis (1993: 99) thus distance themselves from neoclassical economics.

As I have already mentioned, a last way of situating Bowles and Gintis is to refer to Analytical Marxism. This strand analyzes normative issues in the work of Karl Marx from the perspective of rational choice theory. Bowles and Gintis thus share with other Analytical Marxists their explicit focus on normative issues (Mayer 1994: 2). In this sense, Analytical Marxism is quite similar to Public Choice theory, which also employs economic tools to analyze the public domain<sup>72</sup>. However, Analytical Marxists typically amend some of the assumptions of conventional economists. For example, they focus more on the impact of social and structural factors on individuals, their identities and their preferences, thereby filling in the gap in conventional economics concerning preference formation. As a result, Analytical Marxists give a more complex view of the relation between individuals and the institutions that guide their lives: “individuals are formed by society, and these individuals react rationally to their environments to produce tomorrow’s environment, which in turn produces individuals who think somewhat differently from before” (Roemer 1986: 196). As I will show more fully later on, the individual and the institutional level should be understood as mutually dependent on each other. While some Analytical Marxists – like Jon Elster – continue to defend a strong version of methodological individualism (Mayer 1994: 3), I will show that only a more sophisticated view has the explanatory power needed to understand the complex interplay between rational individuals and social institutions.

### 3. Bowles’ and Gintis’ theoretical assumptions

In this section, I spell out the ways in which Bowles and Gintis amend the assumptions that dominate the Homo Economicus model. In order to do so, I focus on their research on prosocial actions and the motivations that underlie such actions. Among the latter, I distinguish between self- and other-regarding

---

<sup>72</sup> John Roemer (1986: 193) – perhaps the best-known defender of Analytical Marxism – defines it as “the programme of applying neoclassical tools to studying what might be called Marxian questions”. As I will show more fully in this chapter, however, I believe that several Analytical Marxists do not merely adopt but quite thoroughly amend the assumptions of the economic conception of rationality.

preferences (3.1). Next, I focus on strong reciprocity, which does not fit this motivational dichotomy (3.2). In order to solve the evolutionary puzzle all this poses, I first analyze the standard accounts that dominate the fields of psychology, philosophy, economics and other social sciences and their respective problems (3.3). Then, I focus on what I believe is the most plausible account of the evolutionary origins of strong reciprocity (3.4). Finally, I sum up the theoretical implications for the conventional Homo Economicus model (3.5).

### 3.1. The motivational dichotomy

In order to set out Bowles' and Gintis' criticisms of the conventional Homo Economicus model, I will focus on those actions that are prosocial and thus confer benefits to others<sup>73</sup>. More specifically, I focus on cooperation, which can be defined in two ways: "the broad definition includes all forms of mutually beneficial joint action by two or more individuals. The narrow definition is restricted to situations in which joint action poses a dilemma for at least one individual such that, at least in the short run, that individual would be better off not cooperating" (Richerson et al. 2003: 358). Cooperation refers to the level of actions, which can arise from a wide variety of individual motivations, considerations and dispositions. Most modern economists and political philosophers assume that people act in ways that serve either their private interest or the public interest. This corresponds to the distinction between self- and other-regarding preferences.

#### 3.1.1. *Self-regarding preferences*

Cooperation can arise from purely self-interested considerations. Following John Stuart Mill and Adam Smith, economists have focused on mutually beneficial trade relations when explaining cooperation. Indeed, there is no need to presume that socially desirable or morally praiseworthy actions always arise from socially desirable or morally praiseworthy motivations<sup>74</sup>. As I have shown

---

<sup>73</sup> Evolutionary theorists would frame such actions in terms of altruism: "a behavior is altruistic when it increases the fitness of others and decreases the fitness of the actor" (Sober & Wilson 1998: 17). However, I reserve the term altruism for the level of the reasons that motivate prosocial behavior.

<sup>74</sup> The thought that one does not have to be noble in order to act in noble ways is the main reason why notions like cooperation should be defined in a morally neutral way. Tolerance, for example, can be defined as allowing for some practice that one deems

extensively, economists like Buchanan defend the assumption of universal egoism, stressing that people in their public roles of citizen, politician or public servant are no different from people in their private roles of consumer or producer.

Everyday as well as experimental observation, however, suggests that not all forms of cooperation can be understood as egoistically motivated. When joint action is needed – for example to provide a public good – purely self-interested individuals will not cooperate, since freeriding will benefit them most, regardless whether others will cooperate or not. While Buchanan's work is valuable in showing that collective and prosocial action can arise from purely self-regarding considerations, the fact that humans cooperate extensively suggests that these considerations are not as omnipresent as he would think and thus that his reliance on the assumption of universal egoism is wrongheaded. A number of prosocial actions, like anonymous giving and participating in massive demonstrations, simply cannot be understood in self-interested terms. Framing the economic conception of rationality in terms of thin rationality, Michael Taylor (1988: 85) wonders: "if these actions are not thin-rational, are they rational in some other sense, or are they simply not rational at all?". As I have shown, I believe the correct answer is the first one.

### *3.1.2. Other-regarding preferences*

Cooperation is thus often motivated by other than self-interested considerations. Among such other-regarding preferences, one can distinguish between altruism and public-spiritedness. While altruists take the furtherance of another's interest as their main goal, public-spirited people concentrate on the public interest. The implicit presupposition in the bulk of modern political philosophy is that individuals, at least in their public roles of citizen, public servant or politician, are concerned with the public good, however this may be defined. Some of them conceive of it along the lines of a general will – as Jean-Jacques Rousseau does – and assume that people try to find out what it entails and how to realize it. But even those who conceive of it as some function of private interests – as social choice theorists do – implicitly trust the people working in governments to take the furtherance of these interests as their main objective. As Buchanan has shown, political philosophers often defend a bifurcated image of individuals, who are modeled as egoistic within the private

---

unacceptable while having the power to do something about it. Such a broad definition, which does not stipulate which reasons one may have for deciding to tolerate what one deems wrong, goes against the grain of recent literature in which tolerance is generally understood as inherently virtuous (Engelen & Nys 2008).

realm of the market and as public-spirited within the public realm of the state. Rousseau (1762: 26), for example, argues that “each individual can, as a man, have a private will contrary to or different from the general will that he has as a citizen. His private interest can speak to him in an entirely different manner than the common interest”.

Both altruism and public-spiritedness perfectly fit the model of neoclassical economists, who follow Lionel Robbins (1935: 24) in analyzing individuals as instrumentally motivated to achieve their ends, whatever these may be. In contrast with their classical colleagues, they refuse to stipulate what the content of preferences is or should be. In an egocentric framework, preferences can be self- or other-regarding. Whereas egoists prefer to help only themselves, altruists prefer to help others and public-spirited people prefer to achieve the public interest. These paradigmatic types form the extremes of a motivational dichotomy. In between, there lies a whole continuum of individuals with mixed motives. Julian Le Grand (2003: 27) describes both extremes of this motivational dichotomy in terms of knaves and knights: “knaves can be defined as self-interested individuals who are motivated to help others only if by so doing they will serve their private interests; whereas knights are individuals who are motivated to help others for no private reward, and indeed who may undertake such activities to the detriment of their own private interests”. He argues that both types of considerations interact in complex ways and motivate people in their public roles. While people do not always act in a purely knightly manner, one should not – as Buchanan does – exclude such motivations completely (Le Grand 2003: 38, 118).

### **3.2. Beyond the motivational dichotomy**

The problem with the motivational dichotomy is that it fails to explain certain actions. While some individuals indeed seem to act in purely egoistic or public-spirited ways (Fehr & Gächter 2000: 162; Gintis 2000: 316), a majority of individuals act in ways that cannot be understood by either of these models.

#### *3.2.1. Reciprocity*

It has been argued that people often base their actions on norms prescribing reciprocity, which “means that in response to friendly actions, people are frequently much nicer and much more cooperative than predicted by the self-interest model; conversely, in response to hostile actions they are frequently

much more nasty and even brutal” (Fehr & Gächter 2000: 159). Sometimes, the reaction is immediate, direct and perfectly equivalent to the initial action. If I buy something in a shop, I pay the exact price of the product in order to take it home with me. However, not all sorts of reciprocity are characterized by such a strict relation between action and reaction. As Adam Smith (1759: Part III, Chapter VI, 9) notices, reciprocity in interactions between kin, friends and acquaintances are typically less determinate.

People tend to answer a gift with a gift (positive reciprocity) and punish antisocial behavior (negative reciprocity) (Fehr & Gächter 2000: 160-164). Reciprocators have internalized the norm that costs and benefits in a relationship are to be balanced out roughly and hence consider it their duty to repay gifts and sanction people who do not perform their fair share of the work. Since reciprocity can involve more costs than benefits, it cannot readily be understood as self-interested and thus goes against classical economics. Here too, it has been argued that reciprocity provides the individual with a feeling of satisfaction, because it makes him feel good or because it stops his conscience from nagging. As I have already suggested, this does not provide an adequate account of the principled nature of such actions: “a person who is motivated by the warm glow that comes from having done one’s duty is not acting out of duty but engaging in narcissistic role playing” (Elster 1989a: 53).

Once more, neoclassical economists come to the rescue, arguing that reciprocity is not necessarily egoistically but always instrumentally motivated. Reciprocators aim to achieve the goals they deem important, like furthering their self-interest (in the case of trading), making another happy (in the case of giving) or upholding a social norm (in the case of imposing social sanctions). Here too, I want to argue that this also forms an inadequate account of the specific nature of reciprocal behavior. Reciprocators act in a non-instrumental way, since they are concerned neither with circumstances nor with outcomes. They reciprocate, simply because they think it is their duty. In contrast with the forward-looking nature of instrumental behavior, dutiful behavior is typically backward-looking. In Kant’s words, it is typically categorical and not hypothetical in nature: I fulfill my duty, no matter what the consequences of my actions may be (Kant 1785: 69).

### *3.2.2. Strong reciprocity: definition*

In what follows, I want to show that people often act on the basis of strong reciprocity, which can be defined as “a propensity to cooperate and share with others similarly disposed, even at personal cost, and a willingness to punish

those who violate cooperative and other social norms, even when punishing is personally costly” (Bowles & Gintis 2000: 37). This immediately suggests the following distinction: “a person is a strong reciprocator if she is willing (a) to sacrifice resources to be kind to those who are being kind (= strong positive reciprocity) and (b) to sacrifice resources to punish those who are being unkind (= strong negative reciprocity)” (McElreath et al. 2003: 136)<sup>75</sup>.

Reciprocity can be either weak (when the individual expects benefits for himself) or strong (when the individual receives no benefits whatsoever). Strong reciprocity thus differs from the abovementioned forms of reciprocity that are ultimately egoistic in nature (Bowles & Gintis 2003: 433). It motivates people to reward friendly actions and to punish hostile actions although this is costly for themselves (Henrich et. al. 2004: 56).

### 3.2.3. *Strong reciprocity: empirical evidence*

While evidence for strong reciprocity can be found in a number of real-life situations – like neighborhoods, families and work places (Fehr & Gächter 2000: 168) – I want to focus on experiments with one-shot interactions between completely anonymous strangers. Laboratory experiments allow researchers to control the conditions and to eliminate the possibility of mutual or long-term benefits to arise. As such they are extremely suited to demonstrate the presence of strong reciprocity (Fehr & Fischbacher 2005: 8; Gintis 2003: 161). Experimental evidence can thus provide insights that ethnographic and anthropological research often lacks: “anonymous one-shot experiments allow us to distinguish clearly between behaviors that are instrumental towards achieving other goals (reputations, long term reciprocity, and conformance with social rules for expediency sake) and behaviors that are valued for their own sake” (Henrich et. al. 2004: 10).

As I have shown, the behavior of only a minority of the subjects in Ultimatum games can be understood in self-interested terms. In general, such egoists (about a quarter of all subjects) are clearly outnumbered by strong reciprocators (about half of all subjects) (Fehr & Gächter 2000: 162; Gintis 2000: 316). Proposers who make offers larger than strictly required might still be egoistically motivated. They might fear that lower offers are rejected, ultimately leaving them with nothing. However, this is no longer the case in so-

---

<sup>75</sup> Other definitions all come down to the same. Gintis (2000: 313), for example, defines strong reciprocity as “a propensity to cooperate, respond to cooperative behavior by maintaining or increasing co-operation, and respond to noncooperative free-riders by retaliating against the ‘offenders’, even at a personal cost”.



called Dictator games, where proposals cannot be rejected at all. While purely self-interested proposers would clearly keep everything for themselves in such cases, only some proposers actually do (Fehr & Fischbacher 2005: 8-9; Henrich et al. 2004: 27, 72, 373). The fact that most of them offer something is best explained in terms of strong reciprocity (Gintis 2003: 168-170; Gintis et al. 2005: 12).

Similarly, responders in Ultimatum games who reject positive offers cannot be understood as trying to serve their self-interest (Bowles & Gintis 2002: 125). Accepting any proposal is always more beneficial than rejecting it. Most of the responders report that they consider low offers to be unfair and that they want to punish unfair behavior. The fact that equivalent offers by a computer are more likely to be accepted than offers by a person also suggests that low offers are rejected not because they are disadvantageous but because they are perceived as unfair (Bowles & Gintis 2000: 42). Responders typically react to the intentions of others rather than that they are instrumentally concerned about the outcome itself.

### 3.3. Standard accounts of the evolutionary origins of prosocial actions

Up to here, I have focused exclusively on the considerations and reasons that motivate individuals to act as they do. These form the so-called proximate causes of their actions. In what follows, however, I want to lay bare their ultimate causes, which refer to the evolutionary origins from which actions spring: “the task is to identify the ultimate, not proximate, causal factors, because it is entirely straightforward to identify some proximate causes: the beliefs and desires of the agent at the time of acting (...). Dissatisfaction with this type of explanation arises because it leaves unanswered the question of *where* those beliefs ultimately come from (...). A search for the ultimate causal factors is thus a search for the real reasons that people cooperate, trust, and so on” (Alexander 2005: 106).

Sober and Wilson distinguish between three accounts of such ultimate causes. First, hedonism stipulates that the only ultimate preferences are to obtain pleasure and to avoid pain (Sober & Wilson 1998: 224). Second, egoism stipulates that ultimately only self-interest matters, which includes not only psychological but also material benefits (Sober & Wilson 1998: 226). In this view, every individual’s ultimate goal is to benefit himself. It does not deny the existence of altruistic preferences but claims that these are only instrumental to fulfilling ultimately egoistic preferences (Sober & Wilson 1998: 2). Third, altruism claims that some ultimate desires are aimed at the well-being of others

for their own sake (Sober & Wilson 1998: 228). In sum, “the theories agree that people sometimes want others to do well; the debate concerns whether such desires are always instrumental or are sometimes ultimate” (Sober & Wilson 1998: 201). To decide this debate, I will now analyze more fully the evolutionary origins of prosocial and cooperative actions.

From Charles Darwin (1871) on, evolutionary thinkers have been puzzled by the fact that people who help others and thus sacrifice their chances to reproduce have been able to survive natural selection (Gintis 2003: 156; Richerson et al. 2003: 387). While classical economists focus on the broad definition of cooperation, it should be clear that I focus here on the narrow one. Self-sacrificing – rather than self-interested – cooperators bear the costs of their actions, while egoists freeride on their efforts. At first sight, one would thus expect egoists to be relatively fitter and thus to outcompete altruists. Indeed, “altruism is the very opponent of the survival of the fittest” (Sober & Wilson 1998: 19). In what follows, I will analyze the standard solutions to this apparent “puzzle of prosociality” (Gintis 2003). They try to show that genuine altruism is an illusion and that its origins can be framed in ultimately egoistic terms (Fehr & Gächter 2002: 137; Fehr & Henrich 2003: 55).

### *3.3.1. Kin selection*

Accounts of kin selection are based on the existence of a “selfish gene” (Dawkins 1989)<sup>76</sup>. According to these accounts, cooperation has evolved among people who are genetically related. It argues that the genetic constitution of humans has developed over hundreds of thousands of years in which people lived in small groups with lots of relatives. Over time, they have developed a cooperative disposition, which ensures the survival of both the individual and his relatives. Since genes have not yet adapted to current circumstances – large societies with one-shot anonymous interactions with strangers – people still tend to cooperate, even with nonkin. In this view, widespread cooperation in modern societies is largely irrational.

---

<sup>76</sup> Bowles and Gintis (2003: 441) argue that Dawkins employs the term ‘selfishness’ in a tautological way, defining selfish behavior as behavior that arises from traits that on average generate higher payoffs and that will thus be favored by natural selection. Dawkins’ selfish gene model does not justify the use of the assumption of universal egoism: “the “selfishness” of the gene and the selfishness of the person are quite different matters” (Ben-Ner & Putterman 1998: 32). As I will show more fully later on, natural selection operates at multiple levels.

However, anthropological research shows that prehistoric humans did not live in groups of relatives closed off from the rest of the world. They have always traveled a lot, thereby often encountering strangers (Fehr & Henrich 2003: 74-75; Gintis et al. 2005: 26). As I will show more fully later on, this has enabled humans to develop a capacity to detect whether they are interacting on a repeated basis or not.

### *3.3.2. Direct reciprocity*

The basic rationale of the second strand, labeled direct reciprocity, is simple: I'll scratch your back if you'll scratch mine. This account focuses on short-term benefits as the main factor in explaining cooperation. As I have already suggested, however, not all forms of reciprocity are so direct and strict. Moreover, stressing mutually beneficial exchanges, direct reciprocity fails to explain why people take on costs in order to punish others who do not cooperate. If there is nothing to gain from such punishment in the short run, direct reciprocators will leave it at that, in contrast with strong reciprocators who will invariably react to antisocial behavior.

### *3.3.3. Reciprocal altruism*

According to the third strand, which Trivers (1971) labeled "reciprocal altruism", individuals help others if it is probable that they will be repaid in the future. Their altruism is thus contingent upon the altruism of others. The underlying mechanism is the same as that of direct reciprocity. Even though the benefits are more loosely connected to each other, they constitute material incentives for cooperation in long-term interactions (Fehr & Gächter 2002: 137).

The problem here is that a lot of relationships typically do without the sort of bookkeeping that is predicted by this account. Joan Silk distinguishes between exchange relationships, in which "benefits are given with the expectation that they will be reciprocated" (Silk 2003: 37) and communal relationships, in which "benefits are given according to the other's need, and receiving a benefit does not create an obligation to reciprocate" (Silk 2003: 44). Partners in communal relationships tend to cooperate without counting. Among friends, help is given because it is needed, not because reciprocity is expected. Experimental evidence shows that a calculating attitude is widely considered to be antithetic to friendship (Silk 2003: 44-46).

In the same vein, it is crucial to see that reciprocal altruism cannot explain strong reciprocity (Bowles & Gintis 2004: 26). Expecting long-term benefits from cooperation, reciprocal altruists act in a forward-looking way. In contrast, strong reciprocators are backward-looking in that they simply react to the actions of others, cooperating with cooperators and punishing freeriders. Moreover, reciprocal altruism is limited to small groups of people who interact frequently. In large groups, freeriding thrives and cooperation based on reciprocal altruism inevitably declines. In addition, people often do not know whether the interaction they engage in will recur on a repeated basis. In light of these problems, reciprocal altruism cannot explain the large scale of human cooperation in modern societies (Bowles & Gintis 2003: 432; McElreath et al. 2003: 136-140).

### *3.3.4. Indirect reciprocity*

The fourth account, labeled indirect reciprocity, stresses the importance of reputation in explaining cooperation. When interactions are not one-shot and anonymous, people can acquire a positive image by acting cooperatively. In such situations, cooperation does indeed tend to increase (Henrich et al. 2003: 448). A more specific account views cooperation as a costly signal of some hidden but valuable characteristic of the cooperator (Bowles & Gintis 2003: 437). Refusing low offers, for example, can be a way of sustaining a reputation of being tough. Precisely because this behavior is costly (in the short run), it signals to others that one has specific valuable characteristics. The fact that one is able to take on such costs shows that one's chances of survival are large: "many nonhuman organisms make costly displays to signal their qualities to others. Costly signaling theory suggests that advertising costly traits indicates phenotypic quality because only individuals of such quality can exhibit maximal expression of these traits (...). For such behaviors to be stable signals, they must be observable, costly, reliable in the sense that they must vary with the quality of the trait being communicated, and of course, they must ultimately provide some benefit to the sender" (Henrich et. al. 2004: 424). Even though cooperation is costly in the short run, it is advantageous in the long run, which allows the individual and his cooperative disposition to survive and flourish in the evolutionary struggle for life.

The problem here is that both indirect reciprocity and costly signaling are plausible only in small groups. Whereas the former assumes that parties keep track of past behaviors of others, the latter assumes the presence of a public that can interpret the signals. Both accounts thus fail to understand why people cooperate on a large-scale with anonymous strangers of whom they have no

information whatsoever (Henrich et al. 2003: 453-457). To explain this thought more fully, I will now focus on a more general flaw shared by all of the abovementioned accounts.

### *3.3.5. Strong reciprocity as a maladaptive trait*

While all of these accounts have some explanatory force, they fail to provide an exhaustive explanation of human cooperation. The basic problem of these accounts is that they are plausible only in relationships between a limited number of people who interact frequently. They are inapt to explain large-scale cooperation between anonymous strangers in one-shot interactions (Boyd & Richerson 2005: 160; Fehr & Fischbacher 2005: 7).

There is a widespread tendency to view strong reciprocity “as a by-product of one of these other ultimate accounts of human cooperation” (Fehr & Henrich 2003: 60-61). However, I agree with Fehr and Henrich (2003: 55) that the latter “do not provide satisfactory evolutionary explanations of strong reciprocity. The problem with these theories is that they can rationalize strong reciprocity only if it is viewed as maladaptive behavior, whereas the evidence suggests that it is an adaptive trait”. People are motivated by a cooperative disposition, the argument goes, which makes evolutionary sense in repeated interactions but not in one-shot interactions with strangers. Humans have long lived in small groups where interactions are typically repeated. In such circumstances, a cooperative disposition indeed enhances the fitness of both the group and the individual. Since humans have not evolved genetically in recent history, they still tend to cooperate, even though such behavior is no longer adapted to current circumstances. After all, one-shot interactions with strangers have become the rule rather than the exception.

Nevertheless, humans have developed a capacity to detect and take into account whether they are interacting repeatedly or not (Fehr & Fischbacher 2005: 14-16). Indeed, cooperation decreases if one moves from repeated interactions with people one knows to one-shot interactions with strangers, even though it remains quite pervasive among the latter. Similarly, cooperation gradually breaks down over time if the subjects have no possibility to build a positive reputation (Fehr & Henrich 2003: 65-69; McElreath et al. 2003: 139-140). In a particular series of experiments, subjects play an iterated Ultimatum game. While responders in the first condition could not build up a reputation, they could in the second condition, because proposers there knew the full history of their behavior. The results suggest that people are clearly capable of distinguishing between one-shot and repeated interactions (Henrich et. al.

2004: 79). Clearly, they are not fooled into thinking that cooperation will ultimately pay off in situations where reputation effects and the corresponding long-term benefits are in fact impossible.

Even though all of the abovementioned accounts refer to evolutionary circumstances that humans share with non-human animals like primates, there is little evidence for them among other animal species. If such self-sacrificing behavior proves beneficial in the long run, why is it not more widespread in the animal world (Bowles & Gintis 2004: 17; Henrich et al. 2003: 445; McElreath et al. 2003: 127-130; Richerson et al. 2003: 379)? While there is a lot of cooperative behavior that humans share with other mammals, the level and scope of cooperation among humans is vastly greater than that among other species. People cooperate on a scale that is unique in the animal world: “perhaps the most singular feature of *Homo sapiens* is the scale on which humans cooperate. In most other species of mammals cooperation is limited to close relatives and (maybe) small groups of reciprocators” (Boyd & Richerson 2005: 133). In addition, the intensity of cooperation is unparalleled in nonhuman animal species. It is safe to say that humans cooperate in ways, settings and degrees unequalled in the animal world (Fehr & Fischbacher 2005: 6; Henrich et al. 2004: 410-411).

In this respect, I want to argue that strong reciprocity is a uniquely human phenomenon and thus particularly apt to explain the uniqueness of human cooperation. It forms the solution to the evolutionary puzzle of human cooperation with nonkin strangers in large groups (Boyd & Richerson 2005: 241).

### **3.4. The evolutionary origins of strong reciprocity**

Experimental evidence from Ultimatum, Dictator and Public Goods games overwhelmingly shows that a majority of people act prosocially, even in one-shot and anonymous interactions (Bowles & Gintis 2002: F425; Kahan 2005: 344-345; McElreath et al. 2003: 126). Crucial in this respect is that cooperation “occurs among genetically unrelated individuals and under conditions that rule out direct reciprocity and reputation formation” (Fehr & Gächter 2002: 137). While selfish genes, direct reciprocators, reciprocal altruists, indirect reciprocators and costly signalers all receive benefits from cooperation, strong reciprocators take up costs even though repeated interaction and reputation effects are impossible: “the essential feature of strong reciprocity is a willingness to sacrifice resources in both rewarding fair and punishing unfair behavior, even if this is costly and provides neither present nor future

economic rewards for the reciprocator” (Fehr & Henrich 2003: 57). While the standard accounts of the evolutionary origins of prosocial actions can explain some forms of cooperation, they fail to account for this specific form of reciprocity and thus do not live up to their own aspiration of explaining the whole range of human cooperation (Bowles & Gintis 2003: 430).

### *3.4.1. Cultural group selection*

If one wants to explain the uniquely human phenomenon of strong reciprocity, one should refer to uniquely human characteristics. An adequate account of human cooperation should thus refer to “capacities that are unique to, or at least much more highly developed in, *Homo sapiens*” (Bowles & Gintis 2003: 429). I want to argue that culture is a good candidate, since it is based on a number of distinctively human cognitive, linguistic and psychological capacities. As such, it is unique to humans who are capable of devising, learning and enforcing a wide range of social norms (Richerson et al. 2003: 379; Sober & Wilson 1998: 158). In what follows, I want to show how this focus on culture makes plausible a specific variant of group selection in the evolutionary process of natural selection.

When internalized, social norms lead to the emergence of groups based on nonkin characteristics. They enable people to suppress differences within the group they belong to (thus limiting competition within groups) and enlarge differences with other groups (thus increasing competition between groups). All this leads to fewer and less intense conflicts within groups and more and more intense conflicts between groups (Bowles & Gintis 2004: 25-26; Henrich et al. 2003: 463; Richerson & Boyd 2004: 57). Ultimately, groups will clash and face the threat of dissolution. At such moments in evolutionary history, mechanisms inducing cooperation on the basis of long-term benefits become less relevant, since the probability of such benefits will decrease sharply over time. In the end, cooperation based on direct reciprocity, reciprocal altruism, indirect reciprocity and costly signaling will break down. In contrast, strong reciprocators will continue to cooperate: “a small number of strong reciprocators, who punish defectors without regard for the probability of future repayment, can dramatically improve the survival chances of human groups” (Bowles & Gintis 2003: 434). Circumstances of low within-group competition and high between-group competition thus allow individually costly but group-beneficial behavior to arise (Henrich et al. 2003: 464; Richerson & Boyd 2004: 202, 214).

Bringing about benefits at the level of the group, strong reciprocity has been able to develop and survive (Fehr & Henrich 2003: 77; Gintis et al. 2005: 30). The basic rationale is that groups with strong reciprocators will outcompete those without them. Because of the increased survival chances of their group, strong reciprocators are not outcompeted by egoists (Bowles & Gintis 2004). This mechanism of group selection was already defended by Darwin (1871: 166) in his attempt to explain human morality: “a tribe including many members who, from possessing in a high degree the spirit of patriotism, fidelity, obedience, courage, and sympathy, were always ready to aid one another, and to sacrifice themselves for the common good, would be victorious over most other tribes; and this would be natural selection. At all times throughout the world tribes have supplanted other tribes and as morality is one important element in their success, the standard of morality and the number of well-endowed men will thus everywhere tend to arise and increase”. As Howard Margolis (1984: 26) notices, “the groups whose members have a propensity to act in group interest will have a selection advantage over the groups deficient in that propensity”.

However, the idea of group selection is controversial. During the 1960's and 1970's, it came under vigorous attack among evolutionary theorists (Sober & Wilson 1998: 6). The argument goes that groups would have trouble distinguishing themselves. Nevertheless, empirical research clearly shows that vast differences between different societies (Henrich et. al. 2004: 18-22). Indeed, characteristics at the group level are much more important in explaining differences in behavior than characteristics at the level of individuals (Henrich et. al. 2004: 33). While people are genetically almost identical, cultural factors enlarge between-group differences and can thus make group selection a strong force even in the absence of genetic relatedness (Sober & Wilson 1998: 192).

Here, I want to stress that behavioral and motivational traits are handed down from generation to generation not only through genes but also through cultural factors such as education and socialization (Henrich et al. 2003: 459). While quite a few mechanisms in human cooperation are innate, cultural and institutional factors play a crucial role as well. People do not only inherit the genes of their ancestors, but also their language, norms, habits, institutions or, in short, their culture (Richerson et al. 2003: 361)<sup>77</sup>. It is therefore perfectly

---

<sup>77</sup> According to Peter Richerson and Robert Boyd (2004: 132), these capacities first evolved about half a million years ago, when hominids had brains that were about the same size of ours. Since culture did not exist prior to this, one can argue that some form of biological group selection played a role in human evolution until then. However, most evolutionary biologists question the plausibility of biological group selection. The main problem is that it ignores that, next to competition between groups, within-group



possible for cultural factors to play a role in natural selection: “the fact that a behavior is transmitted culturally should not be taken to mean that it is nonheritable. Cultural differences between human groups are often stable over long periods of time and are faithfully transmitted to descendant groups. They are heritable in the sense that offspring units resemble parent units, which is all that matters as far as the process of natural selection is concerned” (Sober & Wilson 1998: 114).

To further illustrate that cooperation is often based on cultural factors like norms, I want to refer to experimental evidence from Ultimatum games. The finding “that kindergarteners accept minimal offers about 70 percent of the time, third and fourth graders 40 percent of the time, and adults < 10 percent of the time, implies that any fairness-based norms or cooperative strategies are learned” (Henrich et al. 2004: 219).

For natural selection at the level of groups to be relevant, two requirements must be fulfilled. First, members of one group should behave uniformly different than members of other groups. Second, these differences should be inherited. Both requirements can be fulfilled through the working of genetic or cultural mechanisms. Groups can behave very differently without being genetically different and these behavioral regularities can be transmitted from generation to generation through social norms and institutions (Sober & Wilson 1998: 149). Natural selection thus incorporates selection not only on genetic but on cultural variation as well (Boyd & Richerson 2005: 400).

#### *3.4.2. Multilevel selection*

Selection at the level of groups forms an alternative to kin selection and selection at the level of individuals, since the latter focus exclusively on genetic evolution. While selfish gene accounts are often considered to be the fatal criticism to group selection, it actually shows the plausibility of selection at levels other than the individual one. Just like individuals are vehicles for genes to reproduce themselves, so are groups vehicles for individuals to reproduce

---

competition between individuals remains relevant. While a group of altruists may thrive better on the whole, their fitness gains will be shared by egoists who invade the group. Altruists thus tend to be exploited by egoistic freeriders, so that natural selection can be expected to eliminate altruists from the gene pool (Dawkins 1989: 7-8). As Todd Zywicki (2000: 88) rightly stresses, “the ability to control free riding is the fundamental hurdle to a group selection theory”. Here, it is crucial to see that social norms and the human abilities to enforce and transmit them can explain how opportunities for egoists to invade groups of altruists can be reduced. Once more, cultural factors thus render selection at the level of groups more plausible.

themselves. To capture the insight that natural selection occurs not only at the individual level, but also at the level of groups, Sober and Wilson (1998: 331) speak of “multilevel selection”. While group selection is a relevant factor in the survival of the fittest, it does not imply that other accounts are useless. They all offer perspectives that generate different insights (Sober & Wilson 1998: 98).

Self-sacrificing cooperative behavior is simply incomprehensible if selection operates only at the level of individuals. If selection only plays within a group, egoists will outcompete others. However, if selection between groups plays a role, then prosocial and cooperative behavior is more likely to evolve. Consequently, if selection occurs at different levels, the final outcome depends on the relative importance of both evolutionary forces: “between-group selection favors the evolution of altruism; within-group selection favors the evolution of selfishness. These two processes oppose each other. If altruism manages to evolve, this indicates that the group selection process has been strong enough to overwhelm the force pushing in the opposite direction” (Sober & Wilson 1998: 33). The conclusion is simple: prosocial behavior can evolve only if group selection mechanisms are sufficiently strong.

The basic tenet in the debate on the importance of nature and nurture is to assume that behavior is either genetically determined (nature) or learned by the individual (nurture) (Boyd & Richerson 2005: 7-9). The concept of cultural group selection, however, allows one to see that both cultural and genetic elements fit the evolutionary framework. After all, the latter is broad enough to accommodate the insight that natural selection produces adaptations that have nothing to do with genetic adaptations (Sober & Wilson 1998: 337). The fact that cultural variation, next to genetic variation, can give rise to group selection can explain why human cooperation is unique in both scale and intensity (Boyd & Richerson 2005: 204).

### *3.4.3. The role of norms and emotions*

If one allows for selection at the level of groups, cooperative behavior is more plausible, since groups in which cooperative behavior is prescribed will have an evolutionary advantage over other groups. In this respect, it is useful to analyze more fully the role of social norms in motivating prosocial behavior. Neither positive nor negative strong reciprocators are motivated by instrumental considerations. Instead, they adhere to particular social norms (Fehr & Henrich 2003: 55). Whereas positive strong reciprocity motivates people to cooperate with those similarly disposed, negative strong reciprocity motivates people to take on costs to punish norm-violators. Because punishment of defectors is

anticipated by potential defectors, it substantially increases norm-guided cooperation, thereby enabling groups to maintain levels of cooperation much higher than those in a society of egoists (Gintis 2003: 160-166; Gintis et al. 2005: 15-16; McElreath et al. 2003: 141). The presence of negative strong reciprocators who sacrifice their own interests to enforce prosocial and thus group-beneficial norms is thus crucial in explaining cooperation (Bowles & Gintis 2004: 17).

It might be raised that norms do not necessarily prescribe prosocial or cooperative behavior. Norms of revenge, for example, prescribe behavior that is aimed at decreasing rather than increasing the well-being of others. Other norms – like those surrounding etiquette – prescribe behavior that is costly for the individual itself and does not confer any benefits whatsoever. Nevertheless, I want to argue that they decrease within-group differences and increase between-group differences through the use of seemingly arbitrary symbolic traits like styles of dress, cuisine or dialect. Such norms thus play a crucial role in creating behavioral homogeneity among the members of a group: “human social groups are never genetically uniform, but they are often quite uniform behaviorally, especially when the behaviors are reinforced by social norms” (Sober & Wilson 1998: 150). They ensure that behaviors in a group remain more or less uniform, even though its members are constantly changing. As I have suggested, this behavioral homogeneity matters rather than its underlying mechanisms: “group selection “sees” only the behavioral product, not the process that creates the product” (Sober & Wilson 1998: 156). Norms thus produce the behavioral regularities needed for group selection to be plausible. In addition, the ability of humans to imitate the behaviors of successful individuals has the effect of homogenizing groups, which further supports the plausibility of group selection: “since selection operates on any form of heritable variation and imitation and teaching are forms of inheritance, natural selection will influence cultural as well as genetic evolution” (Richerson et al. 2003: 366).

While this is not always the case, social norms generally prescribe prosocial behavior. They motivate members of a group to avoid conflict with and practice benevolence and generosity toward their fellow members (Sober & Wilson 1998: 172). After all, such norms will generally improve the fitness of the group and will thus be favored by natural selection: “groups with prosocial internal norms will outcompete groups with antisocial, or socially neutral, internal norms” (Bowles & Gintis 2003: 440).

Strong reciprocity and norms mutually support each other. Motivated by internalized norms, strong reciprocity ensures that such norms are upheld and enforced. Taken together, both factors and their mutual interaction explain a

substantial part of human cooperation: “cultural group selection and moralistic punishment are both important to explaining cooperation. Cultural group selection will favor groups with high frequencies of moralistic punishment, and it helps ensure that moralistic punishment enforces functional norms. Moralistic punishment, as we have said, plays a considerable role in maintaining between-group variation on which cultural group selection acts” (Boyd & Richerson 2005: 144).

Internalized norms of fairness, for example, induce the strongly reciprocal disposition to reward fair and punish unfair behavior, which in turn helps to maintain such norms. The relatively costless mechanism of monitoring and sanctioning by peers, which allows groups to uphold cooperation, gains importance in situations where contracts are incomplete or where their enforcement is very costly. In both cases, informal norms are in general more rather than less efficient than formal enforcement mechanisms (Fehr & Gächter 2000: 168, 178). Informal norms thus often provide a cost-efficient solution to collective action problems (Kahan 2005: 347-362). The public good of establishing civil order, for example, can be achieved voluntarily on the basis of internalized norms, whereas it takes a lot of monitoring, policing and enforcing measures if such norms are absent. Later on, I will show more fully that it is important not to undermine but to support social norms.

There is a close link between norms and emotions, which is crucial to understand the behavior of those who violate norms (and experience shame), those who suffer from this (and experience anger) and third-party group members (who experience moral outrage towards violators and empathy towards victims). Emotions are crucial in motivating adherence to norms and stimulating cooperation (Bowles & Gintis 2003: 433; Henrich et al. 2003: 446; McElreath et al. 2003: 144). With respect to positive strong reciprocity, I want to refer to individuals who do not live up to the existing norms and are therefore devalued in the eyes of their fellow group members. With respect to negative strong reciprocity, I want to refer to individuals who suffer from the violation of a norm and to other group members. Both will be prepared to punish norm-violators, simply because it allows them to express their moral outrage with such behavior, which they consider to be unfair. Strong reciprocity is thus best understood as the expression of specific emotions, which are triggered by the fact that one has internalized certain norms that help constitute one's identity (Alexander 2005: 111).

Here, I want to argue that the central role of emotions does not make strongly reciprocal and norm-guided behavior irrational. The standard “view of emotions popular in the social and biological sciences is that emotions should be invoked to explain deviations from the norms of rationality” (McElreath et

al. 2003: 142). In this view, emotions are motivational forces that tend to lead a rational person astray. In contrast, I want to think of emotions as indispensable parts of human decision-making that can but do not necessarily impede its rationality (Damasio 1994: 194; Elster 1999: 156-159).

### 3.5. Complementing Homo Economicus with Homo Reciprocans

Strong reciprocity cannot be understood within the egoistic and instrumental framework of economics. The strongly reciprocal willingness to cooperate and punish those who do not cooperate arises from an adherence to internalized social norms (Kahan 2005: 343). Norm-guided behavior should not be understood as means towards the realization of an individual's goals but as based on principles and commitments he deems important. It goes against the instrumental character of the economic conception of rationality, according to which people care only about the consequences of economic interactions and not about the processes through which these arise (Gintis 2000: 316; Gintis et al. 2005: 18-20). Against the general tenor of conventional economics, which is to define rationality in terms of self-interest and outcome-orientation, I continue to defend that actions are rational if they arise from what the individual himself considers to be good reasons. Consequently, the fact that strongly reciprocal behavior is norm-guided does not necessarily make it irrational.

Strong reciprocity thus forms an anomaly for the instrumental Homo Economicus model. While I clearly grant that both pure egoism and pure altruism can occur, strong reciprocity cannot be reduced to either of these extremes or to some hybrid composite of both. Rather, it is situated beyond this motivational dichotomy. In this respect, it has been argued that a majority of people behave more like Homo Reciprocans than like Homo Economicus (Bowles & Gintis 2003). In contrast with the latter, "*H. reciprocans* is a conditional cooperator whose penchant for reciprocity can be elicited under circumstances in which personal self-interest would dictate otherwise" (Gintis 2000: 316). His backward-looking disposition to react friendly to friendliness and hostile to hostility lies beyond the motivational dichotomy that is situated completely within an instrumental framework: "people are not generally the self-interested actors of neoclassical economics, since they value treating others fairly and will incur personal costs to do so. Nor are people the unconditional altruists of utopian thought, since they want to hurt free-riders and other norm-violators" (Gintis 2000: 316). In contrast with the former who freeride no matter what, strong reciprocators are prepared to take on costs even if this

yields no personal gain whatsoever. In contrast with the latter who cooperate no matter what, they cooperate only if others do their fair share as well.

In sum, the available experimental evidence in Ultimatum and other games shows that a majority of people act like Homo Reciprocans. In contrast with Homo Economicus, “Homo reciprocans cares about the well-being of others and about the processes determining outcomes – whether they are fair, for example, or violate a social norm. He differs in this from the self-regarding and outcome oriented Homo economicus” (Bowles & Gintis 2000: 37). Whereas the classical Homo Economicus is a full-blooded egoist, Homo Reciprocans is willing to benefit others, even when this yields no present or future rewards whatsoever. Whereas the neoclassical Homo Economicus always acts instrumentally to achieve certain goals in the future, Homo Reciprocans reacts to the past behavior of others, even if this inhibits the realization of his personal aims.

To further indicate the empirical adequacy of this Homo Reciprocans model one can refer to iterated Public Goods games. These are typically characterized by initially high levels of cooperation that decay towards the end. According to most economists, subjects only understand the nature of this game and the superiority of the egoistic freeriding option after playing it a number of times. However, this does not fit the empirical finding that cooperation decays less when the number of repeated interactions is increased to the knowledge of the players. In this light, strong reciprocity provides a more plausible reason why cooperation decays: “public spirited contributors want to retaliate against free-riders and the only way available to them in the game is by not contributing themselves” (Gintis 2000: 317). There are two possible evolutions if strong reciprocators are confronted with freeriders. First, cooperation can start to unravel, since increasingly more people refuse to contribute. If others no longer cooperate, why should I? Second, cooperation can be upheld when freeriders are punished by strong reciprocators who feel that they are being treated unfairly. The latter evolution is likely if people have the means to sanction freeriders (Bowles & Gintis 2000: 43; Fehr & Fischbacher 2005: 13-14; Gintis 2003: 165-166).

In this respect, I want to stress that both the Homo Economicus and the Homo Reciprocans are best understood as ideal types. They are abstractions that highlight certain characteristics of individual motivation that are often closely intertwined in reality. Real-life people who are motivated purely by egoism, altruism, public-spiritedness or strong reciprocity are rare. Instead, most of them are motivated by a whole gamut of considerations that gain different relative weight in different contexts. Some of these differences can be traced back to the fact that they belong to different groups, while others are

purely individual. Both forms of motivational heterogeneity suggest the usefulness of different models complementing rather than replacing each other. I thus do not want to substitute one all too simplistic model for another: “it is tempting to react to the widespread experimental evidence of non-selfish behaviors by replacing the *selfishness axiom* with some equally simple and universal assumption about human behavior. If *Homo economicus* has failed the experimental test, maybe *Homo altruisticus*, *Homo reciprocans*, or some other simplified version of a panhuman nature will do better. The diversity of behaviors we have observed leads us to doubt the wisdom of this approach” (Henrich et. al. 2004: 50).

#### 4. Normative implications of Bowles’ and Gintis’ theoretical assumptions

In this section, I want to think through the argument that different explanatory models of individuals have different normative implications in that they justify different institutions and policies. Like I have done in my analysis of Buchanan’s views, I will now address the normative issues surrounding institutional design. As I have argued extensively, these should be based on empirically adequate models of individuals and their motivations. It should by now be clear that the latter are not as homogenous as economists suggest: “policy-making, no less than the grand projects of constitutional design, risk irrelevance if they ignore the irreducible heterogeneity of human motivations” (Bowles & Gintis 1999: 3). In my view, the pluralism at the individual level is reflected at the institutional level, where a combination of various institutional mechanisms is to be favored.

##### 4.1. The institutional dichotomy

Here, I want to show in what sense the abovementioned motivational dichotomy between self- and other regarding preferences has led to a dichotomy in normative thinking about institutional design. This institutional dichotomy, which dominates conventional debates in modern political philosophy, concerns the proper task division between markets and states (Bowles & Gintis 2002: F419). In short, it consists of two contending camps “advocating *laissez faire* on the one hand or comprehensive state intervention on the other as *the* ideal form of governance” (Bowles & Gintis 2005: 379).

#### *4.1.1. Markets*

As I have shown in my analysis of Buchanan, theorists who assume universal egoism tend to praise markets for their capacity to coordinate self-interested actions into socially desirable outcomes: “for them, the safest assumption to make when constructing government institutions and formulating government policy is Hume’s maxim that everyone is a knave. Inevitably, therefore, they endorse the competitive market as the principal means for organising economic and social production; for, as we know from the works of David Hume’s near contemporary, Adam Smith, the market is the method by which self-interest can be harnessed to serve the common good” (Le Grand 2003: 12). Their focus on the invisible hand mechanism typically leads to a predilection for laissez-faire politics. They tend to be wary of governments, because public servants can be expected to serve their private interests rather than the public interest. The state should be constitutionally constrained to a minimum in order to avoid the excesses of unbridled rent-seeking. Its only task is to allow markets to function properly and to correct for their failures by providing additional external incentives, like rewards for compliant citizens and sanctions for deviant ones (Kahan 2005: 341-342).

#### *4.1.2. States*

In contrast, theorists who model individuals in their public roles as altruistic or public-spirited tend to favor a quite pervasive state. Public-spirited government officials can be trusted to do what is needed to serve the common good. In addition, the fact that such citizens are willing to serve the interests of others suggests that they can be expected to pay the taxes to realize the public interest they agree on (Le Grand 2003: 13).

Recently, however, consensus has grown that neither the market nor the state is perfect. Markets fail in achieving unanimously agreed on goals because of freeriding problems and negative externalities. This is where governments come into play. Because of rent-seeking behavior, however, these fail as well. As a result, twentieth century theorists have been debating the extent to which markets and states should complement each other. There is an implicit consensus among these theorists that the main function of both institutions is to bring into correspondence the private and the public interest through the provision of material incentives.



As a disclaimer, I concede that I leave out a number of theoretical possibilities here. First, there is Thomas Hobbes (1651: chapter XVII) who argues that egoistic individuals need a state with absolute and unlimited authority to prevent a war of all against all. They mutually consent to obey to such a Leviathan, since they realize that it is a necessary evil to make social life possible at all. However, the question remains whether self-interested people will actually consent to such a dictatorial regime. After all, one can expect it to serve nothing but its own self-interest. Second, there are anarchists who argue that public-spirited individuals do not need a coercive state, since they will achieve a harmonious society on a voluntary basis. However, a society where order is achieved without formal rules and constraints is a romantic fiction (Buchanan 1999: VII, 5-9). Since disagreement can persist among public-spirited citizens as well – for example on what the public interest exactly consists of – the necessity of some enforcement mechanism is evident. This mechanism can be a formal one like the state or a more informal scheme of social norms. In what follows, I want to make clear what I mean by the latter.

## 4.2. Beyond the institutional dichotomy

The crucial role of strong reciprocity and the norms on which it is based shows the inadequacy of the conventional debate on markets and states. More specifically, I want to show that it gives rise to a third basic institution that is neglected all too often, namely that of communities<sup>78</sup>. This forms the institutional counterpart of strong reciprocity and the still unconventional Homo Reciprocans model.

Bowles and Gintis (1998: 3) define a community as “a structure of social interaction characterized by high entry and exit costs and nonanonymous relationships among members”. The crucial difference between markets and states on the one hand and communities on the other hand lies in the impersonal, anonymous and occasional character of social interactions within the former: “market interactions are characterized by ephemerality of contact and anonymity among interacting agents while idealized state bureaucracies are characterized by long-term anonymous relationships” (Bowles & Gintis 1998:

---

<sup>78</sup> This by no means implies that there are only three institutions in modern societies (Abercrombie, Hill & Turner 1994: 216; Bowles & Gintis 1998: 22). Partly for reasons of space, I will not go into the role of institutions like families, churches, classes, clubs and firms. While some can be fitted in the category of communities quite well, a number of differences remain. For example, whereas churches and firms typically have a formal and hierarchical decision-making structure, communities are based on informal norms and sanctions.

6). In contrast, communities can be defined as groups of “people who interact directly, frequently, and in multi-faceted ways” (Bowles & Gintis 2005: 381). As such, they are based on fundamental aspects of the identities of their members. Traditional definitions of communities by sociologists and social psychologists systematically stress the sense of belonging to a group of people who feel related to each other (McMillan & Chavis 1986: 4).

#### *4.2.1. Egoistic preferences and their amendments*

In such informal networks, interactions are primarily based on the strongly reciprocal willingness of members to cooperate with others so disposed and to punish those who violate shared norms. As I have shown, this clearly suggests that the assumption of universal egoism has its limits. In my view, communities should not be perceived as some leftover from premodern societies, doomed to fade away in modern societies (Bowles & Gintis 1998: 23). Their usefulness lies in their ability to solve the governance problems that markets and states fail to address: “why do communities persist despite their inability to exploit the efficiency-enhancing properties of markets and the advantages of universal enforcement of rules provided by states? One reason is the capacity of communities to foster cooperative behavior” (Bowles & Gintis 1998: 3). Like markets and states, communities provide solutions to coordination problems that arise within social interactions in modern societies. Since they are able to partly solve the failures of markets and states, they provide a desirable complement to both. Even though they do not allow individuals to achieve direct mutual benefits – like markets do – and they do not have a monopoly over coercive enforcement – like states do – communities have the capacity “to avert or attenuate costly coordination problems of the prisoner’s dilemma type” (Bowles & Gintis 1998: 22).

One could question whether communities are a desirable form of social organization: “communities work because they are good at enforcing norms, and whether this is a good thing depends on what the norms are” (Bowles & Gintis 2002: F428). As I have shown, however, norms in general prescribe prosocial behavior towards fellow community members. Because communities are typically characterized by high entry and exit costs, interactions occur on a repeated and enduring basis and, as a result, community members tend to treat each other decently: “communities as we have defined them structure social interaction in ways that foster: (a) frequent interaction among the same agents; (b) partly as a result, low-cost access to information about other community members; (c) a tendency to favor interactions with members of one’s own community over outsiders; and (d) restricted migration to and from other

communities. These structural characteristics, we will show, contribute to the ability of communities to promote pro-social behavior” (Bowles & Gintis 1998: 6).

The capacity of communities to deal with freeriding problems without having to resort to costly enforcement schemes allows societies to achieve high levels of cooperation on a voluntary basis. Generally accepted informal norms make more formalized enforcement mechanisms and incentive schemes largely superfluous. They reduce the need for coercion by a large bureaucratic apparatus. Where social norms are present, governmental intervention is needed to a lesser extent. As such, societies can achieve cooperation without heavily relying on the coercive power of the state (Etzioni 2000: 177).

As I have shown, informal norms and more formal institutions mutually support each other. Where emotions like shame are common, punishment of norm-violators will be particularly effective and, above all, cheap. All one needs to induce shame is a simple frown (Fehr & Henrich 2003: 77). As a result, high levels of group cooperation are achieved and sustained at limited cost (Bowles & Gintis 2003: 439). Informal and formal compliance mechanisms thus complement – rather than substitute – each other in inducing prosocial behavior.

All this suggests a number of ways in which communities can solve the coordination problems that characterize market and state failures. First, since members interact repeatedly, the costs of gathering information of others decrease and its benefits increase. Because they interact frequently and directly, community members have information about each other that is not available in anonymous relations within markets and states. Second, since community members know that they will continue to interact in the future, they have an incentive to act favorably towards their partners in order to avoid future repercussions. This is enhanced by the limited mobility between groups, which arises from their high entry and exit costs.

However, this suggests another problem communities face. They rely heavily on the distinction between insiders and outsiders: “membership has *boundaries*; this means that there are people who belong and people who do not” (McMillan & Chavis 1986: 4). Cultural group selection and the communities to which it gives rise tend to cause malign behavior towards member of other groups: “group selection favors within-group *niceness* and between-group *nastiness*” (Sober & Wilson 1998: 9). If based on race, religion or sex, this inherently exclusive rationale of a shared identity can lead to morally repugnant situations.

However, while communities are based on a divisive rationale of ‘us’ versus them’, they do not automatically give rise to a completely fragmented society.

As I have already suggested, I do not favor an anarchistic ideal in which communities are the only institution to govern social interaction. Some form of government is needed to enforce a minimal framework of rules and make sure that basic human rights are not violated. In addition, there are wider communities, like those at the level of the nation-state, which prevent such morally troublesome situations to arise. They allow people to think of themselves as members of a society with a shared concern for the public good: “participants must share a common identity as citizens that is stronger than their separate identities as members of ethnic or other sectional groups” (Miller 1990: 284). If a community exists at the level of the nation-state, its members will conceive of themselves as parts of that collectivity. Although identification with smaller-scaled communities tends to be more intense, nationality often constitutes an important part of people’s identity<sup>79</sup>. It allows one to empower communities, while avoiding their possibly perverse logic of ‘us’ versus ‘them’ (Bowles & Gintis 2002: F431-432).

#### *4.2.2. Exogenous preferences and their amendments*

The importance of communities in modern societies shows the inadequacy of the assumption of exogenously given preferences as well. Bowles and Gintis argue in this respect that liberalism – including its variant defended by neoclassical economists – neglects the role of communities in governing social life. It focuses on the moments when individuals know what they want, but neglects the moments when their preferences and identities are still being formed: “liberalism claims that the marketplace and the ballot box allow people to get what they want. But liberalism is silent on how people might get to be what they want to be, and how they might get to want what they want to want” (Bowles & Gintis 1986: 125). Liberalism thus has a huge blind spot. While it shows that people make choices, it ignores the fact that choices also make people. Ignoring the impact of people’s actions on their own preferences and identities, liberalism remains “indifferent or hostile to the formation of those loyalties and social bonds upon which a vibrant democracy must depend. This is nowhere more clear than in its devaluation of decentralized autonomous communities” (Bowles & Gintis 1986: 139).

---

<sup>79</sup> David Miller (1990: 239) aptly describes the importance of a community at the national level, in which people share a common national identity: “each member recognizes a loyalty to the community, expressed in a willingness to sacrifice personal goals to advance its interests”.

The institutional structure of a society should not be narrowed down to markets and states but include communities as well: “a democratic society must foster the proliferation of vital and autonomous self-governing communities standing between the individual and the state” (Bowles & Gintis 1986: 139-140). Focusing exclusively on the state as the realm where power is exercised, liberalism neglects the opportunities for people to organize themselves in a vast number of ways at this intermediate level. As I have shown, communities allow individuals to actively participate in public life, thereby reducing the need for an expansive state, which ultimately – as Buchanan’s worst case scenario of the Leviathan shows – poses an inevitable threat to individual freedom and democratic accountability.

### **4.3. Normative implications at the level of the state**

At the policy level, theorists employing the Homo Economicus model agree with governmental measures only if they offer the prospect of mutual benefits in the long run (Bowles & Gintis 1999: 4). In contrast, the Homo Reciprocans model shows the importance of cooperation that is contingent not upon the presence of long-term benefits but upon a perceived balance between rights and obligations (Bowles & Gintis 2000: 44). Most people feel that the benefits and burdens of public arrangements should be distributed more or less evenly.

Crucial in this respect is the notion of desert. Most people regard redistributive and compensatory government measures as fair and legitimate when they favor people who are not responsible for their own suffering. They are generally considered legitimate if aimed at the poor who cannot do much about their situation and thus do not deserve to be poor. The other side of the same coin is that most people do not support such measures if they favor people who make no effort whatsoever to engage in socially valued work. Hence, there is some truth in the cliché that hardworking taxpayers do not like being duped by people who could work but lazily choose not to. The perceived legitimacy of welfare programs thus crucially depends on “one’s views of why the poor are poor, and specifically one’s beliefs about the relative importance of effort rather than luck” (Bowles & Gintis 2000: 47). If people oppose welfare programs and egalitarian policies, it is not because they do not want to pay for them, but because they offend their strongly reciprocal adherence to norms of fairness: “policies that reward people independent of whether and how much they contribute to society are considered unfair and are not supported” (Bowles & Gintis 1999: 16).

According to Bowles and Gintis (1991: 1), public policies should thus tap the public's sentiments: "the welfare state is in trouble not because selfishness is rampant (it is not), but because many egalitarian programs no longer evoke, and sometimes now offend, deeply held notions of fairness, encompassing both reciprocity and generosity". In order to devise desirable institutions, such motivations should not be neglected. In this respect, Philip Pettit (2002: 290-307) argues that there are two strategies to design and implement institutions in a world of rational individuals, namely the managing and the motivating strategy. The managing strategy aims to devise public policies by taking into account the fact that people often act in altruistic, public-spirited and strongly reciprocal ways. In contrast, the motivating strategy stresses that self-interest motivates people not to comply and concludes that institutions should be organized in such a way that compliance becomes more attractive in self-interested terms. However, since individuals differ in the extent to which they voluntarily comply and since it is hard to rig the pay-offs in a personalized way, the argument goes, one "should look to the most self-interested person we can imagine – the knave, in traditional parlance – and make sure that that agent will face penalties that are harsh enough, or rewards that are high enough, to elicit compliance with the institution in question" (Pettit 2002: 276). This corresponds to Buchanan's defense of the assumption of universal egoism for normative purposes. Partly due to Buchanan and other Public Choice theorists, this egoistically motivated *Homo Economicus* has become the standard model of individuals for theorists who engage in institutional design (Le Grand 2003: 9). Not only has it led to a general predilection for markets over states, it also cleared the ground for the conventional policy strategy of promoting the public interest by materially rewarding prosocial behavior and punishing antisocial behavior.

Pettit (2002: 276-277) mentions three problems this line of reasoning encounters: "first, many agents will comply with the demands of a public institution on the basis of a spontaneous, non-egocentric pattern of deliberation. Second, the introduction of sanctions apt for knaves is likely to switch these people out of such non-egocentric deliberation into more self-interested reflection (...). And third, this being so, the introduction of knavish sanctions is likely to reduce the level of compliance overall, not to increase it. These propositions are empirical in character but are borne out in much research". Whereas the first problem refers to the fact that not all preferences are egoistic, the others refer to the fact that preferences can change endogenously under the influence of institutions and policies. I will go deeper into both issues in what follows.

#### *4.3.1. Egoistic preferences and their amendments*

The conventional strategy among politicians and their advisers is the motivating one: “the belief that people are relatively self-regarding (...) may be behind the assumption of economic policy-makers and institutional designers that no proposal is plausible unless it can be shown to be ‘incentive-compatible’: that is, unless it can be shown that people will have self-regarding reasons for going along with what the proposal requires” (Pettit 2002: 227). However, this does not always produce optimal outcomes if people are motivated by other than self-regarding considerations. That is why the managing strategy tries to detect, support and reinforce the motivations that guide the behavior of most people most of the time. It aims to achieve social order and cooperation on the basis of the wide array of considerations that motivate most people (Pettit 2002: 277). It is thus to be employed if one wants to take people as they are and not as the knaves they might be: “the managing strategy would give us a world fit for ordinary, more or less virtuous people, the motivating strategy would give us a world fit for knaves” (Pettit 2002: 305). Only after getting the most out of the citizens who comply voluntarily, will it try to motivate the more self-interested citizens through sanctions and rewards.

Since people are motivated by a mix of considerations of which the relative weight varies from individual to individual and from context to context, it is important to rely on a combination of both managing and motivating strategies. However, given the omnipresence of motivations that induce people to spontaneously cooperate, the managing strategy is generally superior to the motivating one. The fact that social norms have an impact on individuals and their preferences thus clearly has normative repercussions: “the more one is blind to the importance of internalization, the more one would be inclined to increase fines and jail sentences to curtail crime. But if one understands internalization and the ways it can be enhanced, one would rely more on character education, shaming, and peer groups to change people’s behavior” (Etzioni 2000: 165). To be sure, while exclusively relying on monetary incentives is not always ideal, a complete absence of sanctions is likely to lead to corruption as well (Pettit 2002: 299-300). In this respect, it is useful to introduce a system where second-time and third-time offenders of rules receive progressively harsher penalties for not complying. This sends a signal to all citizens that freeriding behavior does not pay off.

Embracing the insight of motivational heterogeneity opens up a number of options that are unthinkable under the assumption of universal egoism: “institutional arrangements can affect the pattern of social outcomes by selecting among agents of different types as well as by altering incentives for

agents” (Brennan 1996: 272). The fact that not all people are purely egoistic suggests the desirability of screening procedures that aim to select people who tend to cooperate without a costly sanctioning scheme (Brennan 1996: 258-262; Pettit 1996: 81; Pettit 2002: 297-298). If some people voluntarily act in socially desirable ways, one might want to search the right man for the right job. It is, for example, useful to attract more public-spirited individuals to hold public offices. In this respect, it may be sensible not to pay politicians and public servants too generously, since this would attract people who are only in it for the money. If, however, the wages are too low, incompetent candidates might be attracted, which increases the risk of corruption and bribery (Brennan 1996: 273-274). The issue here is that of finding the middle road between two extremes. Putting aside the issue where exactly this lies – this should be decided on an empirical basis – I want to stress that such considerations simply cannot pop up if one assumes universal egoism, as Buchanan does in his theory of constitutional choice.

#### *4.3.2. Exogenous preferences and their amendments*

Bowles and Gintis (1986: 10) stress that “politics produces people”. In contrast with Public Choice theorists like Buchanan, they deem it wise to drop the assumption of exogenously given preferences when analyzing the political domain: “we argue that an adequate conception of action must be based upon the notion that people produce themselves and others through their actions (...). Individuals and groups, accordingly, act not merely to *get* but to *become*. The politics of becoming, we believe, provides a central corrective to both the normative and the explanatory dimensions of traditional political theory” (Bowles & Gintis 1986: 22). In order to show more fully what the implications of all this are, I return to the more specific issue of the role of prosocial motivations in a political context.

If other- and process-regarding preferences exist and induce socially desirable behavior, one should take precautionary measures to support and guarantee their continued existence. As James Madison writes in ‘Federalist Paper’ no. 57, “the aim of every political constitution is or ought to be first to obtain for rulers men who possess most wisdom to discern, and most virtue to pursue the common good of the society; and in the next place, to take the most effectual precautions for keeping them virtuous” (Pettit 1996: 81). This interpretation of the Founding Fathers as trying to devise a constitution by detecting and supporting non-egoistic motivations in politicians could not diverge more from Buchanan’s focus on the need to keep egoistic politicians in check by means of strict constitutional limitations.



#### 4.3.2.1. Crowding effects: definition

The insight that the presence of prosocial motivations is not to be taken for granted is especially relevant when thinking about institutional issues. Crucial in this respect is the notion of crowding-out, which means that non-egoistic dispositions and preferences disappear under the influence of institutions and policies that reward egoistic behavior. The opposite effect, which is labeled crowding-in, occurs when institutions and policies support and strengthen non-egoistic dispositions. Bruno Frey, an author well-known for his work on crowding effects, distinguishes between extrinsic and intrinsic motivations and argues that introducing the former can lead the latter to decline: “external interventions undermine intrinsic motivation under identifiable conditions. This effect is theoretically and empirically well grounded” (Ben-Ner & Putterman 1998: 439).

#### 4.3.2.2. Crowding effects: evidence

Frey and Jegen (2001: 589) analyze the evidence surrounding the existence and relevance of crowding effects: “there exist a large number of studies, offering *empirical evidence* in support of the existence of crowding-out and crowding-in. The study is based on circumstantial evidence, laboratory studies by both psychologists and economists, as well as field research by econometric studies”. As for circumstantial evidence, they refer to the example of parents paying their children to mow the lawn. Once this practice is established, children tend to perform household tasks only if there is some material benefit attached to it. With respect to laboratory evidence, studies by psychologists like Edward Deci and economists like Ernst Fehr and Simon Gächter clearly suggest the relevance of crowding-out effects (Frey & Jegen 2001: 597-600). As for field evidence, one can refer to a school’s daycare center that decided to sanction parents who came late to pick up their children. The number of late-coming parents actually increased, because of the changed perception of the service. After a monetary fine was introduced, the intrinsic motivation of parents to show up in time was reduced, because they felt that the teachers were being paid for the trouble of having to stay longer (Frey & Jegen 2001: 602).

## 4.3.2.3. Crowding effects: psychological mechanisms

In order to understand crowding effects, it is useful to analyze the underlying psychological mechanisms. External interventions can reduce the individual's self-determination if they are interpreted as a form of control and thus as neglecting his intrinsic motivation (Frey & Jegen 2001: 594). The fact that he is offered money to perform some task may lead him to feel that his authentic involvement is not really appreciated (Ben-Ner & Putterman 1998: 481). In general, more uniform interventions will tend to crowd out the intrinsic motivations at play, since these are not properly recognized. Paying everyone to perform an action that some would have done out of an inner sense of duty, may actually push the latter to do so simply for the reward.

The main problem is thus that external interventions can shift the so-called "locus of control" (Deci 1971: 105). Whereas people perceive of themselves as being in charge when performing an action on a voluntary basis, they no longer do so when they start to receive money for it. If one wants to avoid crowding-out effects, one should make sure that external interventions are perceived as supporting intrinsic motivations: "external interventions *crowd-out* intrinsic motivations if they are perceived to be *controlling* and they *crowd-in* intrinsic motivation if they are perceived to be *acknowledging*" (Ben-Ner & Putterman 1998: 444). People who receive positive verbal support do not feel bought out and are more likely to continue to act on the basis of their intrinsic motivation. Experimental evidence supports the view that it is the nature of the external interventions that largely determines their impact on intrinsic motivation: "(a) when money was used as an external reward, intrinsic motivation tended to decrease, whereas (b) when verbal reinforcement and positive feedback were used, intrinsic motivation tended to increase" (Deci 1971: 105).

According to Pettit (1996: 73-77), crowding-out effects occur when institutional contexts induce virtual self-interest to come into play and – as a manner of speech – to shift from the back of their head to right before their eyes. Material rewards and sanctions might induce individuals to shift from their usual, norm-guided routines to a more calculating attitude. The introduction of penalties and rewards can thus trigger people into thinking in more egoistic and outcome-oriented ways (Pettit 2002: 293). Aiming explicitly at the individual's self-interest, such incentives cause the salience of egoistic considerations to arise or to increase.

#### 4.3.2.4. Crowding effects: normative implications

The main point is that crowding-out effects can cause the conventional motivating strategy to produce counterproductive results: “the introduction of deviant-centered sanctions would tend to switch agents from a nonegocentric to an egocentric mode of deliberation. Conclusion: The introduction of deviant-centered sanctions is likely to do more harm than good” (Pettit 1996: 73). It forms the main argument in favor of the managing strategy, which aims to design policies that support those socially desirable actions that are performed voluntarily. This complier-centered strategy explicitly aims to strengthen rather than undermine such spontaneous compliance (Pettit 1996: 81-87).

In what follows, I try to show in what ways crowding-out effects affect the relation between citizens and politicians. Crowding out can occur if policies and laws are based on the principle that citizens are not to be trusted: “a constitution implying a fundamental distrust of its citizens, and seeking to discipline them, tends to crowd out civic virtue and undermines the support which citizens are prepared to give towards the basic laws” (Frey & Jegen 2001: 604). Citizens who perceive the state and its employees as trustworthy and fair will comply more readily to its rules, laws and decisions than citizens who have a negative perception of the governmental apparatus. While there is a general willingness to pay taxes among the former, the latter generally consider taxes to be unjustly high and thus try to evade them. In Frey’s words, “the intrinsic motivation to pay one’s taxes – or tax morale – depends strongly on the extent of trust the citizens have in the political system. When individuals are alienated from government and do not think that they are treated fairly by the political process, they are more inclined to pursue their self interest, i.e., to evade taxes, taking into account only the expected probability of being punished” (Ben-Ner & Putterman 1998: 452).

Policy-makers should therefore try to support and promote the trust citizens have in each other and in public institutions (Gintis et al. 2005: 30, 33). If people trust each other, they tend to cooperate more extensively, making costly government intervention largely superfluous. If this is needed after all, citizens tend to believe it is justified. As a result, there is considerable room for large-scale collective action by means of the state (Richerson et al. 2003: 377). If government officials can ensure that a vast majority of citizens contribute their fair share to the public good, their policies will generally be perceived as reasonably fair and legitimate, which harbors the trust citizens have in politics. If, for example, almost everybody pays his taxes, these are perceived as justified, which helps uphold high levels of tax compliance. If, however, people

believe that an increasing number of citizens no longer contribute, their own inclination to do so will gradually erode. Here too, it is strong reciprocity that motivates people rather than self-interest, which would spur them to evade their tax duties altogether, regardless of what others do.

With regard to politicians, it is crucial to see that treating them as untrustworthy knaves might turn into a self-fulfilling prophecy (Goodin 1996: 41; Pettit 2002: 295). If implemented in the design of institutions and policies, the assumption of universal egoism might actually induce them to look for opportunistic occasions, making them less and less virtuous. In Frey's concise words, "a constitution designed for knaves tends to drive out civic virtues" (Frey 1997: 44). This constitutes perhaps the most important problem of Buchanan's theory of constitutional choice, since it undermines his argument in favor of constitutional constraints that limit rent-seeking in politics. The vicious cycle all this gives rise to can be broken by focusing on non-egoistic motivations: "the strong reciprocity model thus underscores the anxiety that too readily accepting the public choice picture can make it the reality of our political life" (Kahan 2005: 364).

There are indeed a number of informal norms that discourage the tendency of politicians to serve only their own interests (Kahan 2005: 364). In fact, it is the institutional constellation of democratic politics that requires them to publicly justify their actions and opinions in other than purely self-interested terms. Even though politicians referring to the public interest might be hypocritical at first – they do not really care about it – the psychological mechanism of cognitive dissonance reduction leads them to adjust their opinions to their utterances (Elster 1998: 104). Jon Elster (1998: 12) has labeled this the "civilizing force of hypocrisy". The largely public nature of political decision-making processes thus ultimately induces them to act honestly, sincerely and impartially.

#### 4.3.2.5. Crowding effects: theoretical relevance

As I have shown in my analysis of Buchanan's views, "the assumption of universal self-interest is an analytic device designed to test out the incentive properties of alternative institutional arrangements" (Brennan 1996: 257). This exercise is subverted if one assumes altruism or public-spiritedness on the part of the people. However, Brennan (1996: 258) stresses that the conventional motivating strategy that is based on the assumption of universal egoism is appropriate only if two requirements are fulfilled: "first, that agents do not invariably pursue the public interest (properly identified) for its own sake (...);

and second, that the institutional devices so designed do not undermine any public-interested motivations that may be in place". As I have shown, both of these conditions are problematic. First, at least a number of people are intrinsically motivated to serve the public interest. While this does not justify an assumption of universal public-spiritedness, it does suggest that the assumption of universal egoism is inapt to say the least. Second, institutional measures that rely on material incentives to motivate self-interested people to act prosocially can be counterproductive. Crowding-out effects thus clearly undermine the assumption of egoism when thinking about desirable institutions. If different institutional schemes induce different motivations, the whole rationale of modeling each and every individual as egoistic in order to evaluate these schemes is rendered superfluous.

Crowding-out effects provide a serious anomaly for the economic conception of rationality, since they go against the so-called relative price effect. By making a particular behavior more rewarding, economists assume that this option becomes more attractive and hence that more individuals will perform it. However, if crowding-out occurs, an increase in rewards leads to a decrease in the behavior at hand (Frey & Jegen 2001: 590). A clear example is that of a community in Switzerland that had to decide whether or not to allow a nuclear waste repository being sited in their town. While half of the respondents agreed to accept it without compensation, only a quarter agreed to accept it with compensation (Frey & Jegen 2001: 603-604).

Since both effects can occur simultaneously, the final outcome can vary: "in general, both the relative price effect and the crowding-out effect are active, so that external intervention has two opposite effects on the agent's performance. Whether intervening is beneficial from the principal's point of view depends on the relative size of the two countervailing effects" (Frey & Jegen 2001: 593). It thus has to be analyzed empirically what the net result will be in different situations (Ben-Ner & Putterman 1998: 455). According to Julian Le Grand (2003: 40, 168), empirical evidence shows that the provision of material incentives does not automatically reinforce egoistic considerations. Instead of turning all knights into knaves, such incentives can crowd in more altruism and public-spiritedness if they are perceived as a symbolic recognition of valued actions. As I have suggested, all this largely depends on the nature of the incentives provided. In most cases, a mix of different kinds of measures will prove optimal.

#### 4.4. Normative implications at the level of the market

##### 4.4.1. *Egoistic preferences and their amendments*

The fact that not all people act like knaves holds not only in the public but also in the private realm. If the assumption of universal egoism were to apply in real life, one would expect crime, theft and cheating to prosper everywhere. In fact, for markets to function properly there is need not only for a set of coercive governmental rules but also for a minimal morality among its participants. Even though it cannot be framed in exclusively egoistic terms, a minimal amount of trust is needed for social order to be possible at all (Ben-Ner & Putterman 1998: 417; Brennan 1996: 259-260). Smith's invisible hand does not rely exclusively upon knavish motives but on a proper mix of egoistic and moral motivations (Ben-Ner & Putterman 1998: 9). As I have already suggested, it is thanks to trust and social norms (communities) that markets can function as well as they do: "large-scale, market-based societies could not function without well-coordinated norms for dealing with anonymous, one-shot, monetary interactions" (Henrich et. al. 2004: 163).

##### 4.4.2. *Exogenous preferences and their amendments*

Providing material incentives to induce people to cooperate might diminish mutual trust, which in the end may undercut the positive effect of such incentives altogether. Consequently, this conventional strategy may turn out to be a self-defeating one: "field and laboratory research suggests that incentives, far from solving collective action problems, can sometimes actually magnify them by dissipating trust" (Kahan 2005: 343). After all, they tend to mask the motivations of one's fellow citizens, which may lead one to question whether or not they will continue to cooperate voluntarily. Why else would such material incentives be provided? Material incentives (which characterize both governmental policies as well as market relations) can thus crowd out strongly reciprocal and norm-guided actions (which typically characterize interactions in communities) (Gintis et al. 2005: 20).

In this respect, I want to refer to the phenomenon of crowding out in the labor market. Punctual workers can become clock-watchers after harsh penalties are imposed for arriving late at the office. This can lead to a reduction in work morale and effort. Moreover, material benefits tend to attract money-seekers to positions that were previously taken by people who performed their job right

because they felt it had intrinsic value or because this was required by their work ethics. Take the example of caring labor. If one increases the pay, this does not automatically result in more and better care: “a higher price may even have a crowding out effect by eroding the kind of values which underlie the motivation for caring labor” (Ben-Ner & Putterman 1998: 181). Once more, the managing strategy is clearly superior to the motivating one.

Max Weber (1914: 40-41) distinguishes between associative relationships, which are based on a congruence of one's interests, and communal relationships, which are based on the feeling that one belongs together. Market interactions belong to the former, because they primarily allow parties to improve their situation. In this respect, Weber (1914: 636) notices the potentially disastrous effects of anonymous market interactions on socially desirable motivations: “the market community as such is the most impersonal relationship of practical life into which humans can enter with one another (...). Where the market is allowed to follow its own autonomous tendencies, its participants do not look toward the persons of each other but only toward the commodity; there are no obligations of brotherliness or reverence, and none of those spontaneous human relations that are sustained by personal unions. They all would just obstruct the free development of the bare market relationship, and its specific interests serve, in their turn, to weaken the sentiments on which these obstructions rest”. While not explicitly, Weber questions the assumptions of egoistic and exogenously given preferences. He clearly argues that the egoistic reasoning that dominates market behavior is antithetic to the nobler motivations that dominate personal relationships.

Conventional economists praise the market for its ability to optimally allocate goods and services without assuming benevolence or altruism. Bowles and Gintis (1986: 130) cast doubt on this conventional view of the role of the market as exclusively concerned with the production and distribution of goods: “a constitutive theory of the economy must address not only the question of who gets what and why, it must also ask who gets to become what and why. A theory that focuses exclusively on production runs the risk of seeing economic activity as simply a process of *getting* rather than also a process of *becoming*”. They thus stress that markets can have an impact on its participants as well. Markets reduce not only the need for non-egoistic considerations, but also these considerations themselves (Bowles & Gintis 1986: 129).

According to Jean Ensminger, there are two distinct and opposite effects that markets can have on its participants (Henrich et. al. 2004: 357). First, markets can induce more self-interested behavior and erode nobler motivations. Since egoistic behavior tends to be rewarded in market settings, the Homo

Economicus model works as a self-fulfilling prophecy here<sup>80</sup>. Second, it has been argued that markets induce considerations like honesty, good manners and a concern for reputation. People who engage more frequently in market relations are more likely to follow norms of cooperation, fairness and reciprocity than others, since they are familiar with anonymous, one-shot interactions and the mutual benefits these can produce (Henrich et. al. 2004: 356). People who never engage in such interactions, the argument goes, tend to distrust people who they do not know personally and are thus less willing to cooperate. Indeed, in Western, highly developed societies, where markets play a prominent role in everyday life, cooperation is abundant.

Ensminger is rather optimistic about the overall effect of markets, stressing that they can reinforce socially desirable motivations. She concludes that people's concern for reputation may become so entrenched that it emerges in anonymous, one-shot interactions as well (Henrich et. al. 2004: 380). While some would argue that this implies that everything can ultimately be reduced to self-interest, I think it is warranted to conclude that norms of cooperation, fairness and reciprocity have an autonomous impact on behavior that can override egoistic considerations. As I have suggested earlier on, it is thanks to such norms that market exchanges can generate mutual benefits (Henrich et. al. 2004: 429).

Bowles and Gintis clearly do not share Ensminger's optimism in this respect. In their view, the market can be valued both positively and negatively: "it is exactly the anonymity of the market that renders it so attractive when considered from the standpoint of instrumental action and so ominous when considered as a formative influence on human development" (Bowles & Gintis 1986: 129). They argue that the anonymity in market exchanges renders less likely the evolution of social norms that are crucial in solving coordination problems (Bowles & Gintis 1993: 95-96). Behavior based on norms of cooperation becomes less viable and stable if markets spread<sup>81</sup>. This thought has been succinctly formulated by Bowles and Gintis (1993: 97): "it has been suggested that *homo economicus* produced capitalism, meaning roughly that human nature being what it is, the evolution of the capitalist rules of the game is both likely and desirable. But this may be just backwards, or at least one-sided; one could equally argue that capitalism produced *homo economicus*". While

---

<sup>80</sup> This phenomenon also comes to the fore in experiments that show that students tend to act more egoistically after taking economics classes (Frank et al. 1993).

<sup>81</sup> An additional consequence might be that people become less concerned with politics if markets become increasingly more important in social life: "markets might undermine democratic political participation through limiting the stakes and reducing the opportunity costs of not participating" (Bowles & Gintis 1986: 135).



the first sentence summarizes Buchanan's position, the second largely coincides with my criticism of it.

## 5. Conclusion

### 5.1. Summing things up at the individual level

Experimental evidence shows not only that strong reciprocity motivates people to reward prosocial and punish antisocial behavior in anonymous one-shot interactions, but also that they do so to a greater extent in repeated interactions. In addition, less people act in strongly reciprocal ways if the costs of doing so increase (Fehr & Fischbacher 2005: 19). This immediately suggests that the Homo Economicus model is not without value (Alexander 2005: 108). The fact that non-egoistic preferences are widespread in social interactions throughout the world by no means implies that egoism plays no role at all. Experiments provide substantive evidence that a minority of people systematically serve their own interests. In addition, accounts of kin selection, direct reciprocity, reciprocal altruism, indirect reciprocity and costly signaling do explain some forms of human cooperation. They show that egoism can be behind reciprocity when interactions are repeated or when reputation building is possible (Fehr & Fischbacher 2005: 16). Stressing the importance of strong reciprocity, however, I have aimed to criticize the general tendency to explain all forms of cooperation as based on purely egoistic and instrumental considerations.

Most economists and evolutionary psychologists implicitly argue that strong reciprocity is irrational. As I have already shown, however, there is no need to claim that rationality presupposes self-interest or outcome-orientation (Gintis 2000: 320; Gintis et al. 2005: 5). The minimal conception of rationality only stipulates that rational actions are based on what the individual considers to be good reasons. As such, it is perfectly possible to think of the dutiful and norm-guided aspects of strong reciprocity as rational.

The evolutionary perspective is valuable in that it analyzes where such reasons come from and how they are formed. It shows that they should be modeled in ways that do not fit the conventional Homo Economicus model. Insights gained from evolutionary theory thus lead Bowles and Gintis to criticize and amend the all too strict requirements of economic rationality. Pointing towards the importance of both genetic and cultural factors affecting individual preferences and identities, evolutionary insights are crucial in understanding

not only individual behavior but also the social and institutional structure people live in.

## 5.2. Summing things up at the institutional level

The central argument in this dissertation is that decisions concerning institutional design and reform should take people as they are. Because of their all too narrow theoretical assumptions, conventional economists fail to address some of the most crucial normative issues in modern societies. Bowles, Gintis and their collaborators remedy this problem by providing a richer view of the multitude of individual motivations. In this respect, they try to live up to the requirement that normative views on institutions should take into account the wide motivational array of individuals: “within most social actors, self-seeking impulses exist alongside principled and even altruistic motives. In designing our institutions we ought at least to take account of that fact, which might (as a further principle) be describe as *sensitivity to motivational complexity*” (Goodin 1996: 41). Pointing towards the value of communities, Bowles and Gintis show that the conventional debate on the tasks of markets and states is to be rephrased if it is to be relevant for institutional issues that concern actual individuals.

Communities have both an instrumental and an intrinsic value. First, they are able to partly address the freerider problems that markets and states inevitably face: “communities solve problems that might otherwise appear as classic market failures or state failures” (Bowles & Gintis 2002: F422). Through the enforcement of prosocial norms, they help to achieve and sustain levels of cooperation higher than would be expected on the basis of the assumption of universal egoism (Bowles & Gintis 1998: 4). The fact that community members have internalized shared norms allows them to engage in collective action without having to resort to costly and coercive state intervention (Bowles & Gintis 2000: 46). In short, communities can render coercive and costly governmental intervention superfluous. Governments can thus save themselves a lot of trouble simply by welcoming and supporting the informal norms that ensure high levels of cooperation in communities.

Second, communities also have intrinsic value. People are essentially social creatures who define themselves as members of a group. Seeing their own origins and destinies as inextricably bound up with those of that group, they feel isolated without it: “community is not merely a matter of sentimental attachment to other persons, but enters deeply into identity in such a way that, cut off from the relevant community, a person’s life would lose an important

part of its meaning” (Miller 1990: 234). Crucial to speak of communities is the emotional connection that accompanies the group member’s identification with a shared history, tradition, language or symbol (Margolis 1984: 128-129; McMillan & Chavis 1986: 8). Because of their purely formal and anonymous nature, interactions in markets and states do not allow people to act on fundamental aspects of their identities. Communities provide a suitable antidote to such alienating contexts, since they enable people with shared identities to express who they are. In contrast with both markets and states, which are based on the liberal goals of mutually beneficial cooperation and moral neutrality, communities thus allow individuals to act on their comprehensive doctrines.

As people engage less and less in repeated and personal interactions, they “may develop fewer loyalties and affections and may thus feel less regard for others” (Ben-Ner & Putterman 1998: 49). The very fact that an individual finds himself in circumstances where self-regarding preferences prevail may lead to a gradual erosion of his other- and process-regarding preferences. However, there is no reason to despair, since these effects can be countered and even reversed: “interactions within families and small groups may permit desirable other- and process-regarding preferences to be rewarded (...). Habituation to other-regarding or virtuous behavior in the small-group setting may (...) predispose individuals to behave similarly in situations involving outsiders” (Ben-Ner & Putterman 1998: 43). The personal relations within communities can thus serve as a counterweight against the impersonal ones within states and markets (Ben-Ner & Putterman 1998: 324).

It has been debated to what extent one can speak of an increasing process of individualization in modern societies. In order to show the need for reacting to the dissolving bonds between the members of a society, I want to refer to the example of care for the elderly. Whereas this used to be done on an informal and voluntary basis by neighbors and family members, one can argue that people nowadays rely more heavily on more formalized public arrangements. However, these have a number of downsides, like the increased costs for society in general and, more importantly, the deteriorating well-being of the elderly themselves. After all, placement in large homes often leads to increased isolation and ensuing depressions. Once more, one can see both the purely instrumental and the more elemental value of communities coming to the fore.

As this chapter’s opening quote by John Plamenatz suggests, the notion of communities is indispensable as soon as one sees that the economic conception of rationality is too narrow to capture the essence of what it is to be human. Because of the impersonal character of interactions within markets and states, these institutions hardly leave room for the aspects that characterize

expressive rationality. In contrast, communities allow people to express who they are and to form themselves in ways they deem desirable. After all, an important part of their identity is still defined by the groups they belong to: “people see themselves not merely as individuals with essentially private interests and concerns, but as tied to social units in such a way that in answer to the question ‘Who are you?’ they will say, among other things, ‘I belong to ...’” (Miller 1990: 234).

What matters is that one finds a suitable combination in which each institution performs its own valuable tasks in mutually supporting ways. Take, for example, David Miller’s defense of market socialism. While he gives a central role to markets, which efficiently provide the goods and services people want, he realizes that these need to be complemented by a state that sets the basic rules, rectifies income inequalities and supplies the public goods that markets fail to provide (Miller 1990: 295-298). Yet, Miller (1990: 319) stresses that such a state does not necessarily turn into an ever increasing bureaucratic apparatus: “it should not be the benevolent colossus of socialist myth; but nor, for the same reason, need it be the malevolent leviathan of libertarian nightmare”. In addition, he stresses the value of communities, especially at the national level: “it is because they share a common identity that citizens are able to reach a genuine consensus on matters of policy; without this underlying identity the other components of the system would fall apart” (Miller 1990: 18-19). This shows that it is perfectly possible to combine (the advantages of) a free market, a quite extensive state and a comprehensive community.

### **5.3. Bringing together the individual and the institutional level**

The view of individuals as serving the public interest, which used to be dominant amongst political philosophers, has recently been replaced by the less naïve view of individuals as serving their private interests. I have argued that both extremes of this motivational dichotomy are inappropriate for normative purposes. As Le Grand (2003: 2) puts it: “policies designed on the assumption that those who work in the public sector are basically knights are likely to have disastrous consequences if in fact most of those individuals are predominantly knaves. But the same may be true for policies fashioned on the basis of a belief that people are knaves if the consequence is to suppress their natural altruistic impulses and hence destroy part of their motivation to provide a quality public service”. Theorists should take into account a wider array of motivations when aiming to improve institutions and policies (Fehr & Gächter 2000: 161, 178; Kahan 2005: 366). While empirically adequate models of individuals need to be

developed more fully, I believe the Homo Reciprocans model provides an appropriate basis in this respect.

This shift in the way individuals are modeled is likely to have significant normative repercussions: “good policies and constitutions are those that support socially valued outcomes not only by harnessing selfish motives to socially valued ends, but also by evoking, cultivating, and empowering public spirited motives” (Bowles & Gintis 2002: F430). Even though I have stressed the superiority of the managing strategy that takes such motivations into account, the motivating strategy has its value as well. Even though policies providing material incentives to induce citizens to cooperate can be counterproductive, this is surely not always the case. They can be useful in preventing a completely lawless situation to arise. If they are explicitly targeted at freeriding knaves, they can strengthen the trust that more knightly citizens have in the institutions that govern their lives. As such, they can bolster rather than undermine the widespread predisposition to cooperate (Kahan 2005: 346-351).

Whereas liberal theorists tend to view other-regarding dispositions and motivations like altruism and public-spiritedness as scarce resources that should not be relied upon too heavily, Bowles and Gintis show that a lot of people act in a number of ways that cannot be understood in self-interested terms. In addition, they show that institutions should cultivate prosocial actions and motivations, because they might disappear if not exercised on a regular basis. The analogy with scarce resources that are used up if employed too often is misleading in that prosocial motives can actually fade away if they are not put into practice.

At the individual level, I have focused on strong reciprocity as falling beyond the motivational dichotomy. At the institutional level, I have focused on communities as falling beyond the institutional dichotomy. Failing to incorporate the empirical evidence supporting the widespread presence of strong reciprocity, the conventional debate on the task division between markets and states is outdated. In this respect, I have stressed that welfare economists and normative political thinkers should focus more on communities.

People generally act according to the roles attached to the institutional context within which they act. In the market, egoism is the central motivation. In the state, people tend to act in more public-spirited ways. In communities, motivations like strong reciprocity, generosity, loyalty, solidarity, trust and shame are crucial. If one assumes universal egoism, it is impossible to understand the importance of families, states and communities in modern societies. If one wants to explain and appreciate the central role these

institutions continue to fulfill, one should allow for more motivational heterogeneity than conventional economists tend to do (Ben-Ner & Putterman 1998: 54-57).

I acknowledge that the individual and the institutional level do not correspond as neatly as I may have suggested. All of the abovementioned notions are best understood as ideal types that do not actually exist in their pure form. Just like individuals are typically motivated by a multitude of considerations, societies always consist of a complex mixture of elements that characterize these institutional ideal types. Moreover, such an amalgam of institutions mutually complementing each other is highly desirable, since they simply cannot function properly without each other (Miller 1990: 249-251). A governmental framework is indispensable for smooth market relations, just as much as norms (communities) help govern the interactions between consumers and producers, employees and employers (markets) and citizens and their political representatives (states). Markets and states would simply break down if it were not for the widespread tendency of individuals to adhere to the informal norms that form the backbone of their communities. The main challenge thus lies in “developing an institutional structure such that states, markets and communities are mutually enhancing” (Bowles & Gintis 2002: F431).

## CONCLUSION OF PART III

---

*“Social actors are transformed by their very acts (...). Social choice transforms not only the rules of the game, but the subjects of social life themselves”*

(Bowles & Gintis 1986: 118)

### 1. Conclusions at the explanatory level

As should be clear by now, Samuel Bowles and Herbert Gintis diverge from James Buchanan in their analyses of individual behavior, motivation and rationality. This goes for all the levels I have distinguished in the first part of this dissertation. Whereas Buchanan focuses on the instrumental nature of actions and the egoistic and exogenous nature of preferences, Bowles and Gintis amend all of these assumptions. As their empirical research on strong reciprocity shows, people often act on the basis of internalized norms of cooperation, reciprocity and fairness, regardless whether this benefits them or not. This clearly demonstrates that they do not aim to achieve some goal but simply want to express their adherence to such norms and their attachment to the community they live in. It also shows that people often are not egoistic, not even in some indirect way. In addition, the fact that such cooperative, reciprocal and fair actions can be crowded out under the influence of an inappropriate use of monetary incentives suggests that people's preferences should not be taken for granted (which is what one does if one assumes these to be exogenously given).

In recent years, there has been an increasing focus on institutions in a broad range of scientific disciplines. More and more economists, sociologists, political theorists and philosophers embrace the insight that social action is always set within an institutional context. These institutions, however, not only form external constraints on individual choices, but also have a more internal impact on people: “the same contextual factors that constrain individual and group actions also shape the desires, preferences, and motives of those individual and group agents” (Goodin 1996: 20). Perhaps the main conclusion of the third part of this dissertation at the explanatory level is thus that institutions alter both people's choice options (by constraining them) as well as their preferences (by shaping them).

Individuals and institutions mutually influence each other. First, the institutional structure of a society is formed through the decisions of its members. Individual behavior merges into social norms and institutions. Second, individuals are in turn affected by the institutions that constitute the environment within which they are socialized. According to Douglas North, the social sciences should focus explicitly on “the underlying interrelationships between the rules of the game that humans devise to structure human interaction and the way those rules evolve in the interaction between humans and their environment, an environment which changes as a result not only of external natural forces, but also of changes induced by the players themselves” (Ben-Ner & Putterman 1998: 491). In order to get an adequate understanding of the complex interplay between both individuals and institutions, one thus needs an adequate model of both.

This perfectly ties in with my points of criticism towards the all too narrow economic conception of rationality: “an understanding of this two-way relationship requires a more comprehensive characterization of human motivation than that which lies at the core of standard economic models” (Ben-Ner & Putterman: xvii). While the first part of this dissertation hints at ways in which the economic models can and should be amended, this third part shows that this should always be done in a way that incorporates the impact of institutions. I have tried to show that institutions, rules and norms have some sort of robustness that makes them a relevant factor in any model within the social sciences.

It is interesting to see whether all this has repercussions for the assumption of methodological individualism. I believe the insights of Bowles and Gintis show the need to amend Buchanan’s all too individualist approach to institutions. Avner Ben-Ner and Louis Putterman (1998: 45) pose this issue as follows: “a full understanding of the effects of preferences on institutions, and of institutions on preferences, requires a view of their relationship as operating in two directions at once. But researching such a two-way relationship rigorously presents serious methodological challenges. If both institutions and values are endogenous, what can be the starting point for one’s analysis?”. In this respect, one could say that orthodox economists typically cling on to methodological individualism, whereas sociologists typically stress that the basic explanatory force is primarily situated at the institutional and thus structural level. In this respect, I want to follow the middle road, which is described by Goodin (1996: 17): “even the staunchest advocate of rational choice models as explanatory tools must concede that people’s preferences (which are the driving force in that model) (...) come, ultimately, from structures of past experiences, prior socialization or social location. And even the staunchest advocate of structural explanations cannot escape the fact that there have to be agents – albeit



“socially constructed” ones – to serve as the carriers and enforcers of those structural imperatives, which those agents inevitably reshape in the process of reapplying and reinforcing them”.

In the first part of this dissertation, I unambiguously adopted the principle of methodological individualism, according to which social phenomena should ultimately be explained by referring to individuals and their actions. After studying the works of Analytical Marxists like Bowles and Gintis, however, a more sophisticated view seems to be in order. The fact that individuals are shaped by social factors such as institutions leads one to question whether the former indeed form the ultimate basis for explanations within the social sciences. After all, what do preferences, beliefs and intentions – or, in short, reasons – matter, if they result from social, institutional and structural factors in society? Should we not replace the methodological individualist stance with a structuralist one?

Michael Taylor (1988: 94) convincingly argues that “a pure individualist explanation would have the field to itself *only if* the causes of the attitudes and beliefs which cause action are themselves nothing but actions and properties of individuals”. In this case, the individual is indeed the ultimate explanatory variable. This view assumes that social structures are nothing but individuals interacting with each other. This reflects Buchanan’s view of institutions as resulting from voluntary exchange relations between rational individuals. In his view, institutions exist only because they serve the purposes of these individuals and will collapse if they no longer do so.

While Taylor (1988: 95) grants that structures do not exist completely independently from individuals, he believes that such a purely individualist view is too radical: “certainly, a structure typically emerges as a result of, and is maintained or transformed by, the actions of individuals. But it is not *the same thing* as these actions”. However, this immediately suggests that a purely structuralist account is inapt as well. While structural factors like institutions play an important role in society, they are partly formed through intentional actions of individuals: “these structures and situations are themselves in significant part the products of intentional actions” (Taylor 1988: 97). That is why I have repeatedly stressed the value of Buchanan’s optimism regarding people’s capacity to change the structural aspects of society for the better.

The basic conclusion is that neither individuals nor institutions can provide the ultimate rock-bottom foundation for explanations within the social sciences. One should try to find the middle road between an exclusive focus on free and intentional individuals on the one hand and an exclusive focus on structural factors on the other hand. After all, the fact that individuals, their preferences and identities are partly formed by structural factors does not make them

irrelevant. Similarly, the fact that these structural factors are in part result from individual actions does not make them irrelevant. Consequently, one needs “both individualist explanation of structures (and other macrophenomena) and structuralist explanation (amongst other kinds of explanation) of individual attitudes and beliefs. To deny either side of this supposition is to deny *any* causal force either to structures or to individuals, to attach *all* the explanatory power to one or the other” (Taylor 1988: 94). Since the individual and structural levels determine each other, both kinds of explanations are necessary (Mayer 1994: 211).

This leads me to nuance my views regarding individualism. At the ontological level, Philip Pettit (1993: 123-155) convincingly argues in favor of what he labels holistic individualism. While he does not deny the existence of social and structural regularities (holism), he remains an individualist in claiming that these do not eliminate or dominate intentional regularities (individualism) (Cuypers 2002: 93-97). Pettit stresses that an individual’s actions are not completely determined by structural factors (individualism), while acknowledging that his beliefs and preferences essentially depend on the people surrounding him (holism) (Van Liedekerke 2002: 116). At the methodological level, I endorse Pettit’s defense of “explanatory ecumenism, according to which intentional, structural, historicist, and rational choice styles of explanation are complementary enterprises” (Pettit 1993: 229). As I have shown, both intentional and structural explanations are valid and can mutually complement each other. There is no reason why the validity of either sort of explanation would undercut that of the other (Pettit 1993: 253). The notion of holistic individualism allows one to bring together both the individual (part I) and the institutional level (part III).

## 2. Conclusions at the normative level

The works of James Buchanan and Analytical Marxists like Bowles and Gintis provide a number of useful insights in normative issues surrounding institutional design. As Buchanan himself stresses, “there are important normative implications to be derived from the public choice perspective on politics, implications that, in their turn, carry with them an approach to institutional reform” (XIII, 19). In the end, his aim is not merely to analyze what individuals and institutions look like, but to suggest how things can be changed for the better (III, 306).

As I have shown, Buchanan ends up defending a libertarian social order that heavily relies the market’s invisible hand mechanism. The only task the minimal

state has to fulfill is to make sure that the market functions well. Governmental intervention is thought of as a necessary evil only to be employed if its alternative of free market interaction leads to even worse outcomes. In contrast with Buchanan who explicitly bases his theory of constitutional choice on the Homo Economicus model, Bowles and Gintis (1999: 2) focus on motivations that lie beyond this model in order to create more room for legitimate government intervention. A more sophisticated view of rational individuals shows that the conventional debate between defenders of either the market or the state is grossly outdated and incomplete. Stressing the importance of actions based on social norms prescribing cooperation, reciprocity and fairness, it enables one to argue in favor of less market and state and more community.

As I have shown, James Buchanan applies a catallactic perspective to the constitutional choice of a basic institutional structure. Bowles and Gintis (1986: 117) show that the assumptions that are implicit in doing so have a blind spot: “among the more compelling metaphors for society is the market (...). This view admirably captures the intentional aspect of human activities reflected in the act of choice, but it ignores structural determination; the systematic way in which the rules of the game produce social outcomes independently of the wills of the actors themselves”. Buchanan fails to adequately understand the impact of institutions and norms on society and its members, which is exactly the main focus of most Marxist and structuralist theorists. However, the latter, in turn, fail to capture the fact that people can act freely within the rules and even intentionally change these rules.

In order to take both aspects into account, Bowles and Gintis (1986: 118) propose to understand society as “an ensemble of games. The game analogy is immediately attractive as it evokes both action and structure (...). We integrate choice, structural determination, and history by conceiving of realms of social action as games in which both the rules and the players are continually transformed (...). The rules of the various games define the meaning and effectiveness of action on the part of the players, but these rules are in turn altered by the players themselves”. As this conclusion’s opening quote suggests, Bowles and Gintis make use of this analogy in order to stress the mutual dependence of institutions of a society and its members. They argue that neoclassical economists – primarily because of their assumption of exogenous preferences – have focused exclusively on the rules of the game: “political theorists and constitutional thinkers since the late eighteenth century have taken *Homo economicus* as a starting point (...). Good rules of the game thus came to displace good citizens as the *sine qua non* of good government” (Bowles & Gintis 2005: 379). Since people are assumed to have exogenously given reasons, the only way to improve society is to change its structural

characteristics. With Bowles and Gintis, I want to revive attention to the fact that well-ordered societies consist of good citizens as well as good institutions.

While Bowles and Gintis make use of the same analogy as Buchanan, their normative conclusions diverge completely. This results from their acknowledgement of the fact that individuals can change when playing the game of life. Their empirically more adequate views of both individuals and institutions show that Monty Python's lyrics are too simplistic in presenting the whole issue as a dilemma. As their focus on cultural factors in natural selection shows, people are not "simply spirally coils of self-replicating DNA". However, neither "is life just a game where we make up the rules while we are searching for something to say". It is simply absurd to believe that individuals can freely and intentionally choose the rules that guide their lives. Nevertheless, even though people are partly determined by their genetic, cultural, social and institutional backgrounds, there is room for constructivist attempts to improve society. I will go into this thought more fully in the general conclusion.

## GENERAL CONCLUSION

---

*“For myself, I am an optimist – it does not seem to be much use being anything else”*

(Churchill 1954: 8603)

This dissertation is based on the conviction that the notion of rationality is crucial if one wants to explain human behavior. I hope it furthers the search for consistent and philosophically pertinent alternative conceptions of rationality that criticize and complement the conventional – but alas still dominant – economic one. Each of these alternatives captures aspects that are essential to understand human beings as rational creatures. Whereas expressive rationality stresses that people have good reasons to act upon the things they care about, the value of instrumental rationality is based on the simple fact that rational people often care about the consequences of their actions. While it would be grossly irrational for people not to honor the basic commitments that constitute their identities, the meaning of their lives does not hinge entirely upon the accomplishments they achieve. In what follows, I try to sum up some general conclusions and show the relevance and value of this dissertation’s research project.

### 1. Avoiding a pessimistic determinism

As should be clear by now, I strongly believe in the capacity of people to design and reform the institutions that govern their lives. This optimistic attitude seems to be some kind of necessity if one wants to engage in normative political philosophy. As this conclusion’s opening quote by Winston Churchill aptly suggests, it does not seem to be much use being anything else than an optimist. If one does not believe that individuals can improve (the basic institutional structure of) their society, one ends up in passive resignation.

To explain this thought more fully, I want to bring into remembrance the quote by Monty Python cited in this dissertation’s general introduction: “is life just a game where we make up the rules, while we are searching for something to say? Or are we just simply spirally coils of self-replicating DNA?”. As I have suggested, the question it raises is a genuine and interesting one. Are human

beings able to deliberately change the rules and institutions that guide their lives or will such attempts inevitably encounter the strict boundaries of human nature? However, I want to argue that its formulation is misleading in that it presents this issue in a dichotomous way. The lyrics suggest that people are either completely free to think up whatever rules they deem desirable or that they are completely determined by their genetic constitution. Instead, I want to argue that both sides of the story are relevant. When thinking about desirable rules and institutions for modern societies, one should thus take into account human nature as it has evolved. While it would be absurd to deny that people are – to some extent – determined by their genes, this does not imply that they have no control whatsoever over the ways in which they organize their lives.

In a highly insightful article, Thomas Christiano argues that rational choice approaches to the public domain inevitably fall victim to the trap of what he labels “basic structure determinism”. Referring to institutions like markets, states and communities, Christiano (2004: 124) defines this as “the thesis that the development, maintenance, and decline of basic structural institutions in society is determined by forces that are beyond the capacity of human beings to guide and design. Political institutions do not develop the way they do because human beings think that this is the best way for them to develop (...). Political and social institutions are the product of the cumulative effects of many people acting with a great variety of different purposes, and the development of the institutions overall cannot be said to be determined by any kind of design, choice, or plan”.

The crux of his argument lies in claiming “that individuals rarely have much impact on the social and political institutions in which they live. So individuals do not have incentives to try to change the constitutional order under which they live. The expected value of their actions is so low in this area that the self-interest of individuals inclines them to concern themselves with other issues” (Christiano 2004: 128). According to Christiano (2004: 132), it makes sense from a rational choice perspective to find out which institutions are desirable, but not to expect that individuals will do what is necessary to bring about these institutions: “the combination of the assumption of *homo economicus* and the exhortation to bring about the best basic structural political institutions seems to be a self-defeating approach to politics if basic structure determinism is true. And the reason for this is that it is simply not within the power of human beings voluntarily and by design to bring about desirable basic structural change”.

Accepting this kind of determinism thus generally leads to a pessimistic stance in issues surrounding institutional design. After all, it is hard to see what the use would be of articulating ideal institutional schemes if no rational individual

will try to achieve them. Basic institutions, the argument goes, are typically formed and reformed as the unintended result of the interactions of a lot of people. No single individual has control over this process. This lays bare a strange kind of tension in the views of liberal theorists like James Buchanan. Whereas they systematically stress the ability of individuals to freely and intentionally change the rules and institutions of society, they have to admit that people, whom they model as economically rational, have no incentive whatsoever to employ this ability and to engage in political or institutional matters.

There are two ways out of this pessimistic trap. First, one can stress that basic structure determinism does not deny the possibility of individuals to act intentionally and freely to achieve their goals (Christiano 2004: 125). After all, it only argues that actions intended to change institutions in deliberate ways are doomed to fail. While the basic institutional framework is not subject to the deliberate planning of individuals, the latter can still make marginal changes. This line of reasoning implies that the room for institutional design and reform is limited. Nevertheless, it allows one to remain optimistic and to put one's hopes in small but incremental changes to the institutional structure. After all, such deliberate and conscious efforts to make the world a better place are precisely what politics is all about.

Second, one can argue that individuals are not rational in the sense implied by Christiano. Since people do not act like *Homines Economici*, they may well be motivated to do what is needed to implement the institutional scheme they deem desirable. Most people tend to act upon other than self-regarding preferences whenever the circumstances do not allow them to serve their self-interest. This is what characterizes both large-scaled elections (part II) and the context of institutional design (part III). Here, motivations like altruism, public-spiritedness, fairness and reciprocity are more likely to be decisive, since the opportunity cost of forgone satisfaction of egoistic preferences is typically low (Brennan & Hamlin 2000: 19).

In this respect, it should be clear that people can have good reasons to act collectively and cooperatively. In addition, such actions typically have socially desirable effects – they confer benefits to the group in which one lives (and thus also to one's fellow members) – and thus make perfect evolutionary sense. Both voting and acting in strongly reciprocal ways typically have small costs for the individual at hand but potentially large benefits for society. Christiano is perfectly right in pointing out that such a situation leads to motivational problems among *Homines Economici*, which will be tempted to freeride on the contributions of others. Nevertheless, people who identify to some extent

with their community and have internalized its norms will have good reasons to contribute to its public good.

As I have stressed repeatedly, cooperative and prosocial dispositions are especially relevant in contexts where people have little or no incentive to base their decisions on self-interested considerations. Since individuals cannot directly further their own interests when voting or deciding on basic institutional issues, they tend to take on such occasions to express what they care about. This entails a different picture than that of basic structure determinism, in which political outcomes arise as the unintended outcome of social interactions: “if expressive voting cannot operate as an *invisible* hand it could, at least in principle, operate as a *visible* hand. That is, voters might systematically vote their views of the public interest” (Brennan & Hamlin 2000: 176). The fact that a lot of people base their public choices on expressive considerations leads to me doubt the pessimistic conclusion that they necessarily lack the motivation to reform institutions.

Christiano (2004: 138-139) is critical of this optimistic story. First, he argues that the expressive account of voting does not explain why individuals decide to vote. However, I have shown that it is, in fact, better able to do so than the instrumental account, according to which no economically rational citizen will vote. Second, he argues that empirical evidence suggests that people vote strategically rather than expressively. However, my emphasis on expressive considerations does not exclude that instrumental considerations have a marginal impact on voting decisions and electoral outcomes. As I have shown, an account that relies exclusively on instrumental considerations – which Christiano seems to defend – is downright false.

According to Brennan and Hamlin (2000: 177-180), a democratic system tends to select morally motivated politicians. After all, citizens are expected to express in elections their support for the candidate whose principles and character they identify with. This, in turn, motivates politicians to speak and act in ways that appeal to the electorate. Again, Christiano expresses his doubts. First, he argues that voters may well decide to express more flimsy or even malign feelings in the voting booth. Second, he stresses that voters have no incentive whatsoever to inform themselves about political issues, parties or candidates. Third, this leads to politicians being “chosen on the basis of superficial displays of virtue and crude political appeals at best (...). The voters will simply not have the necessary information to evaluate the real characters or platforms of politicians” (Christiano 2004: 140).

Concerning the first criticism, I have argued that the expressive considerations that can motivate voters should not be understood as a tendency to vent superficial feelings and tastes. After all, if one wants to explain why people take



the trouble of going out to vote, one should refer to the fundamental things that underlie their preferences and identities. The second criticism is beside the point in that it continues to refer to economically rational individuals and thus neglects the ways in which expressive considerations can motivate people to become involved in politics. Since the point is exactly that voters are not primarily self-interested, it does not follow that they inevitably remain politically apathetic and ignorant. If they care about the fate of their community, they will have a good reason to inform themselves on all things political. Christiano seems to ignore this in his third criticism as well. In addition, he does not take into account the possibly civilizing force of hypocrisy. If politicians are elected (at least partially) on the basis of non-egoistic considerations, they will have an incentive to speak in impartial terms. As I have shown, this may ultimately lead (initially hypocritical) politicians to actually become less self-regarding or help (honest) politicians getting elected. Moving away from the economic conception of rationality thus enables me to avoid the pessimistic stance it entails with regard to institutional design.

## 2. The value and limitations of this research project

As always, this dissertation, which is the result of four years of research, has its strengths as well as its shortcomings. While I do believe that it contains a number of valuable insights, it would be pretentious to deny that it has its limitations as well. In what follows, I mention only a few of them in order to open up a number of prospects for further research.

First, this dissertation offers no specific suggestions with respect to measures or reforms that would improve the basic institutional structure. While I do hope that it points to some general directions on the route towards a realistic utopia, I believe it is up to policy makers and their advisers to try and implement the insights of political economists and philosophers at the pragmatic level of real-world politics.

Second, while I have stressed the importance of good citizens – next to good institutions – in such a realistic utopia, it is still up for discussion what exactly this entails. I have not specified in any detailed manner which motivations and actions can or should be labeled moral, noble or virtuous. Making use of the general phrase of socially desirable motivations and actions, I hope to have circumvented discussions on such matters. After all, under some circumstances, altruistic and public-spirited preferences can lead to unintended side effects in the sense that they do not make others better off. Well-intended actions can thus result in socially undesirable outcomes. While one should

analyze more fully the mechanisms that may thwart or pervert good intentions and take them into account when designing institutions and policies, I refuse to believe that they cannot be remedied.

Third, while my emphasis on both good citizens and good institutions as parts of a realistic utopia perfectly fits Paul Ricoeur's attention to both individual and social ethics, one should remain cautious that it does not give rise to social engineering in which the formation of virtuous citizens becomes a policy goal in itself. To avoid deliberate attempts to form citizens on the basis of some ideal, I have stressed the need to uphold and support those norms prescribing prosocial action that are still present in modern societies.

Throughout this dissertation, I have relied heavily on different theoretical models to analyze both the individual and the institutional level. In this respect, I believe that these perspectives – which are all incomplete when considered on their own – should ultimately be combined when one wants to explain human behavior in all of its aspects and think about ways in which societies should be structured. I hope this dissertation opens up a number of possibilities for theorists in the 'human sciences' to join forces.

In addition, I have tried to show the need for alternative conceptions of rationality to complement the all too narrow economic one, which fails to incorporate the basic fact that people's complex motivational arrays can change endogenously under the influence of institutional reforms. It is up to social and political scientists, experimental economists, evolutionary theorists and psychologists to fill in this broad theoretical framework on the basis of empirical studies. While economists indeed tend to delegate issues of preference formation and change to other disciplines, I think they should give up some of their central assumptions if they want to benefit from such research. If this dissertation is to convince the reader of anything, I hope he will be persuaded by the need for more interdisciplinary research within the broad domain currently constituted by psychology, evolutionary theory, social sciences, psychology, political theory, moral and political philosophy.

## BIBLIOGRAPHY

---

- ABERCROMBIE, N., HILL, S. & TURNER, B.S. (1994), *The Penguin Dictionary of Sociology*, 3<sup>rd</sup> ed., London, Penguin Books.
- ALDRICH, J.H. (1993), *Rational Choice and Turnout*, in: *American Journal of Political Science*, 37 (1): 246-278.
- ALDRICH, J.H. (1997), *When Is it Rational to Vote?*, in: D.C. Mueller (ed.) (1997), *Perspectives on Public Choice: A Handbook*, Cambridge, Cambridge University Press: 373-390.
- ALEXANDER, J.M. (2005), *The Evolutionary Foundations of Strong Reciprocity*, in: *Analyse & Kritik*, 27 (1): 106-112.
- BARRY, B. (1970), *Sociologists, Economists and Democracy*, London, Collier-Macmillan.
- BECK, N. (1975), *The Paradox of Minimal Regret*, in: *American Political Science Review*, 69 (3): 908-913.
- BEITZ, C.R. (2000), *Rawls's Law of Peoples*, in: *Ethics*, 110 (4): 669-696.
- BENN, S.I. (1979), *The Problematic Rationality of Political Participation*, in: P. Laslett & J. Fishkin (eds.) (1979), *Philosophy, Politics and Society*, Oxford, Blackwell: 291-312.
- BEN-NER, A. & PUTTERMAN, L. (eds.) (1998), *Economics, Values, and Organization*, Cambridge, Cambridge University Press.
- BLAIS, A. (2000), *To Vote or Not to Vote? The Merits and Limits of Rational Choice Theory*, Pittsburgh, University of Pittsburgh Press.
- BLAIS, A. & YOUNG, R. (1999), *Why Do People Vote? An Experiment in Rationality*, in: *Public Choice*, 99 (1): 39-55.
- BOUDON, R. (1993), *Toward a Synthetic Theory of Rationality*, in: *International Studies in the Philosophy of Science*, 7 (1): 5-19.
- BOUDON, R. (1997), *Le "paradoxe du vote" et la théorie de la rationalité*, in: *Revue française de sociologie*, 38 (2): 217-227.
- BOWLES, S. (2004), *Microeconomics: Behavior, Institutions, and Evolution*, Princeton, Princeton University Press.
- BOWLES, S. & GINTIS, H. (1986), *Democracy and Capitalism: Property, Community, and the Contradictions of Modern Social Thought*, New York, Basic Books.

- BOWLES, S. & GINTIS, H. (1992), *Power and Wealth in a Competitive Capitalist Economy*, in: *Philosophy and Public Affairs*, 21 (4): 324-353.
- BOWLES, S. & GINTIS, H. (1993), *The Revenge of Homo Economicus: Contested Exchange and the Revival of Political Economy*, in: *Journal of Economic Perspectives*, 7 (1): 83-102.
- BOWLES, S. & GINTIS H. (1998), *The Moral Economy of Communities: Structured Populations and the Evolution of Pro-Social Norms*, in: *Evolution and Human Behavior*, 19 (1): 3-25.
- BOWLES, S. & GINTIS H. (1999), *Is Equality Passé? Homo Reciprocans and the Future of Egalitarian Politics*, in: *Boston Review*, 23 (6): 1-22.
- BOWLES, S. & GINTIS, H. (2000), *Reciprocity, Self-Interest and the Welfare State*, in: *Nordic Journal of Political Economy*, 26: 33-53.
- BOWLES, S. & GINTIS, H. (2002a), *Homo Reciprocans*, in: *Nature*, 415, January 10: 125-128.
- BOWLES, S. & GINTIS, H. (2002b), *Social Capital and Community Governance*, in: *Economic Journal*, 112 (483): F419-F436.
- BOWLES, S. & GINTIS, H. (2003), *Origins of Human Cooperation*, in: P. Hammerstein (ed.) (2003), *Genetic and Cultural Evolution of Cooperation*, Cambridge (Massachusetts), The MIT Press: 429-443.
- BOWLES, S. & GINTIS, H. (2004), *The Evolution of Strong Reciprocity: Cooperation In Heterogeneous Populations*, in: *Theoretical Population Biology*, 65: 17-28.
- BOWLES, S. & GINTIS, H. (2005), *Social Capital, Moral Sentiments, and Community Governance*, in: H. Gintis, S. Bowles, R. Boyd & E. Fehr (eds.) (2005), *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*, Cambridge (Massachusetts), The MIT Press: 379-398.
- BOYD, R. & RICHESON, P.J. (2005), *The Origin and Evolution of Cultures*, New York, Oxford University Press.
- BRENNAN, G. (1996), *Selection and the Currency of Reward*, in: R.E. Goodin (ed.) (1996), *The Theory of Institutional Design*, Cambridge, Cambridge University Press: 256-275.
- BRENNAN, G. & HAMLIN, A. (1998), *Expressive Voting and Electoral Equilibrium*, in: *Public Choice*, 95 (1): 149-175.
- BRENNAN, G. & HAMLIN, A. (2000), *Democratic Devices and Desires*, Cambridge, Cambridge University Press.
- BRENNAN, G. & LOMASKY, L.E. (1989), *Large Numbers, Small Costs: The Uneasy Foundation of Democratic Rule*, in: G. Brennan & L.E. Lomasky (eds.) (1989),

- Politics and Process: New Essays in Democratic Thought*, Cambridge, Cambridge University Press: 42-59.
- BRENNAN, G. & LOMASKY, L.E. (1993), *Democracy and Decision: The Pure Theory of Electoral Preference*, Cambridge, Cambridge University Press.
- BROOME, J. (1995), *Weighing Goods: Equality, Uncertainty and Time*, Oxford, Blackwell.
- BUCHANAN, A. (1995), *Ethics, Efficiency, and the Market*, Oxford, Clarendon Press.
- BUCHANAN, J.M. (1986), *Liberty, Market and State. Political Economy in the 1980s*, Oxford, Oxford University Press.
- BUCHANAN, J.M. (1999), *The Collected Works of James M. Buchanan*, Indianapolis, Liberty Fund.
- CALDWELL, B.J. (1982), *Beyond Positivism: Economic Methodology in the Twentieth Century*, London, Allen & Unwin.
- CAMPBELL, A., CONVERSE, P.E., MILLER, W.E. & STOKES, D. (1960), *The American Voter*, New York, John Wiley & Sons.
- CARLING, A. (1998), *The Paradox of Voting and the Theory of Social Evolution*, in: K. Dowding, & D. King (eds.) (1998), *Preferences, Institutions, and Rational Choice*, Oxford, Clarendon Press: 20-42.
- CHRISTIANO, T. (2004), *Is Normative Rational Choice Theory Self-Defeating?*, in: *Ethics*, 115 (1): 122-141.
- CHURCHILL, W.S. (1954), *Speech at the Lord Mayor's Banquet in London on November 9, 1954*, in: R.R. James (ed.) (1974), *Winston S. Churchill: His Complete Speeches, 1897-1963*, New York, Chelsea House Publishers: 8603-8604.
- COHEN, J. (2002), *Taking People as They Are?*, in: *Philosophy and Public Affairs*, 30 (4): 363-386.
- COLEMAN, J. (1990), *Foundations of Social Theory*, Cambridge (Massachusetts), The Belknap Press of Harvard University Press.
- CUDD, A. (2003), *Contractarianism*, in: E.N. Zalta (ed.) (2003), *The Stanford Encyclopedia of Philosophy*, available online at: <http://plato.stanford.edu/archives/spr2003/entries/contractarianism>.
- CUYPERS, S.E. (2002), *Individualistische ontologie en oecumenische methodologie in de sociale wetenschappen*, in: X. Vanmechelen (ed.) (2002), *Afhankelijkheid zonder dominantie. Over de sociale en politieke filosofie van Philip Pettit*, Leuven, Acco, 2002: 77-102

- DAMASIO, A.R. (1994), *Descartes' Error: Emotion, Reason, and the Human Brain*, New York, Putnam Publishing.
- DARWIN, C. (1871), *The Descent of Man and Selection in Relation to Sex*, London, Murray.
- DAVIDSON, D. (1980), *Essays on Actions and Events*, Oxford, Clarendon Press.
- DAVIDSON, D. (1982), *Paradoxes of Irrationality*, in: R. Wollheim & J. Hopkins (eds.) (1982), *Philosophical Essays on Freud*, Cambridge, Cambridge University Press: 289-305.
- DAWKINS, R. (1989), *The Selfish Gene*, 2<sup>nd</sup> ed., New York, Oxford University Press.
- DECI, E.L. (1971), *Effects of Externally Mediated Rewards on Intrinsic Motivation*, in: *Journal of Personality and Social Psychology*, 18 (1): 105-115.
- DOWDING, K. (2005), *Is it Rational to Vote? Five Types of Answer and a Suggestion*, in: *British Journal of Politics and International Relations*, 7 (3): 442-459.
- DOWDING, K. & KING, D. (1998), *Introduction*, in: K. Dowding & D. King (eds.) (1998), *Preferences, Institutions, and Rational Choice*, Oxford, Clarendon Press: 1-19.
- DOWNES, A. (1957), *An Economic Theory of Democracy*, New York, Harper & Row.
- DWORKIN, R. (2002), *Sovereign Virtue: The Theory and Practice of Equality*, Cambridge (Massachusetts), Harvard University Press.
- EDGEWORTH, F.Y. (1881), *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*, New York, Kelley Publications, reprinted in 1967.
- ELSTER, J. (1979), *Ulysses and the Sirens: Studies in Rationality and Irrationality*, Cambridge, Cambridge University Press.
- ELSTER, J. (1982), *Rationality*, in: G. Floistad (ed.) (1982), *Contemporary Philosophy: A New Survey*, Den Haag, Nijhoff: 111-131.
- ELSTER, J. (1983), *Sour Grapes: Studies in the Subversion of Rationality*, Cambridge, Cambridge University Press.
- ELSTER, J. (ed.) (1986), *Rational Choice*, Oxford, Blackwell.
- ELSTER, J. (1989a), *Nuts and Bolts for the Social Sciences*, Cambridge, Cambridge University Press.
- ELSTER, J. (1989b), *Solomonic Judgements: Studies in the Limitations of Rationality*, Cambridge, Cambridge University Press.

- ELSTER, J. (1991), *Rationality and Social Norms*, in: *European Journal of Sociology*, 32 (1): 109-129.
- ELSTER, J. (1993), *Some Unresolved Problems in the Theory of Rational Behaviour*, in: *Acta Sociologica*, 36 (3): 179-189.
- ELSTER, J. (ed.) (1998), *Deliberative Democracy*, Cambridge, Cambridge University Press.
- ELSTER, J. (1999), *Strong Feelings: Emotion, Addiction, and Human Behavior*, Cambridge (Massachusetts), The MIT Press.
- ENGELEN, B. (2005), *The Politics-Economics-Ethics Continuum Revisited*, in: *Romanian Journal of Bioethics*, 3 (3): 60-69.
- ENGELEN, B. (2006), *Solving The Paradox. The Expressive Rationality of the Decision to Vote*, in: *Rationality and Society*, 18 (4): 419-441.
- ENGELEN, B. (2007a), *Alcoholica: When Sweet Amber Becomes the Master of Puppets*, in: W. Irwin (ed.) (2007), *Metallica and Philosophy: A Crash Course in Brain Surgery*, Oxford, Blackwell: 29-40.
- ENGELEN, B. (2007b), *Rationality, Norms and Institutions: In Search of a Realistic Utopia*, in: *Human Affairs*, 17 (1): 33-41.
- ENGELEN, B. (2007c), *Thinking Things Through: The Value and Limitations of James Buchanan's Public Choice Theory*, in: *Review of Political Economy*, 19 (2): 165-179.
- ENGELEN, B. (2007d), *Why Compulsory Voting Can Enhance Democracy*, in: *Acta Politica*, 42 (1): 23-39.
- ENGELEN, B. (2008a), *Book Review of 'Beyond Conventional Economics', edited by Giuseppe Eusepi and Alan Hamlin*, forthcoming in: *Review of Social Economy*.
- ENGELEN, B. (2008b), *The Sources of Cooperation: On Strong Reciprocity and its Theoretical Implications*, forthcoming in: *Theory and Psychology*.
- ENGELEN, B. & NYS, T. (2008), *Tolerance: A Virtue? Towards a Broad and Descriptive Definition of Tolerance*, forthcoming in: *Philosophy in the Contemporary World*.
- ETZIONI, A. (2000), *Social Norms: Internalization, Persuasion, and History*, in: *Law and Society Review*, 34 (1): 157-178.
- EUSEPI, G. & HAMLIN, A. (2006), *Beyond Conventional Economics: The Limits of Rational Behaviour in Political Decision Making*, Cheltenham, Edward Elgar.
- FEHR, E. & FISCHBACHER, U. (2005), *Human Altruism – Proximate Patterns and Evolutionary Origins*, in: *Analyse & Kritik*, 27 (1): 6-47.
- FEHR, E. & GÄCHTER, S. (2000), *Fairness and Retaliation: The Economics of Reciprocity*, in: *Journal of Economic Perspectives*, 14 (3): 159-181.

- FEHR, E. & GÄCHTER, S. (2002), *Altruistic Punishment in Humans*, in: *Nature*, 415, January 10: 137-140.
- FEHR, E. & HENRICH, J. (2003), *Is Strong Reciprocity a Maladaptation? On the Evolutionary Foundations of Human Altruism*, in: P. Hammerstein (ed.) (2003), *Genetic and Cultural Evolution of Cooperation*, Cambridge (Massachusetts), The MIT Press: 55-82.
- FEREJOHN, J.A. & FIORINA, M.P. (1974), *The Paradox of Not Voting: A Decision Theoretic Analysis*, in: *American Political Science Review*, 68 (2): 525-536.
- FEREJOHN, J. & SATZ, D. (1994), *Rational Choice and Social Theory*, in: *The Journal of Philosophy*, 91 (2): 71-87.
- FEREJOHN, J. & SATZ, D. (1996), *Unification, Universalism, and Rational Choice Theory*, in: J. Friedman (ed.) (1996), *The Rational Choice Controversy: Economic Models of Politics Reconsidered*, New Haven, Yale University Press: 71-84.
- FIORINA, M.P. (1976), *The Voting Decision: Instrumental and Expressive Aspects*, in: *Journal of Politics*, 38 (2): 390-415.
- FIORINA, M.P. (1997), *Voting Behavior*, in: D.C. Mueller (ed.) (1997), *Perspectives on Public Choice: A Handbook*, Cambridge, Cambridge University Press: 391-414.
- FRANK, R.H., GILOVICH, T. & REGAN, D.T. (1993), *Does Studying Economics Inhibit Cooperation?*, in: *Journal of Economic Perspectives*, 7 (2): 159-172.
- FRANKFURT, H.G. (1988), *The Importance of What We Care About. Philosophical Essays*, Cambridge, Cambridge University Press.
- FRANKFURT, H.G. (1999), *Necessity, Volition, and Love*, Cambridge, Cambridge University Press.
- FREY, B.S. & JEGEN, R. (2001), *Motivation Crowding Theory*, in: *Journal of Economic Surveys*, 15 (5): 589-611.
- FRIEDMAN, J. (1996), *Introduction: Economic Approaches to Politics*, in: J. Friedman (ed.) (1996), *The Rational Choice Controversy: Economic Models of Politics Reconsidered*, New Haven, Yale University Press: 1-24.
- FRIEDMAN, M. (1953), *The Methodology of Positive Economics*, in: M. Friedman (1953), *Essays in Positive Economics*, Chicago, The University of Chicago Press: 3-43.
- GAUTHIER, D. (1986), *Morals by Agreement*, Oxford, Clarendon Press.
- GINTIS, H. (2000), *Beyond Homo Economicus: Evidence from Experimental Economics*, in: *Ecological Economics*, 35 (3): 311-322.



- GINTIS, H. (2003), *Solving the Puzzle of Prosociality*, in: *Rationality and Society*, 15 (2): 155-187.
- GINTIS, H., BOWLES, S., BOYD, R. & FEHR, E. (2005), *Moral Sentiments and Material Interests: Origins, Evidence, and Consequences*, in: H. Gintis, S. Bowles, R. Boyd & E. Fehr (eds.) (2005), *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*, Cambridge (Massachusetts), The MIT Press: 3-39.
- GOODIN, R.E. (1996), *Institutions and Their Design*, in: R.E. Goodin (ed.) (1996), *The Theory of Institutional Design*, Cambridge, Cambridge University Press: 1-53.
- GOSSERIES, A. (2005), *Publicity*, in: E.N. Zalta (ed.) (2005), *The Stanford Encyclopedia of Philosophy*, available online at: <http://plato.stanford.edu/archives/spr2005/entries/publicity>.
- GRAVELLE, H. & REES, R. (2004), *Microeconomics*, 3<sup>rd</sup> edition, Harlow, Prentice Hall.
- GROFMAN, B. (1983), *Models of Voter Turnout: A Brief Idiosyncratic Review*, in: *Public Choice*, 41 (1): 55-61.
- HARGREAVES HEAP, S. (1989), *Rationality in Economics*, Oxford, Blackwell.
- HAUSMAN, D.M. (1998), *Rationality and Knavery*, in: W. Leinfellner & E. Köhler (eds.) (1998), *Game Theory, Experience, Rationality: Foundations of Social Sciences, Economics and Ethics in Honor of John C. Harsanyi*, Dordrecht, Kluwer: 67-79.
- HAUSMAN, D.M. & MCPHERSON, M.S. (1995), *Economics, Rationality and Ethics*, in: D.M. Hausman (ed.) (1995), *The Philosophy of Economics: An Anthology*, 2<sup>nd</sup> ed., Cambridge, Cambridge University Press: 252-277.
- HENRICH, J., BOWLES, S., BOYD, R.T., HOPFENSITZ, A., RICHERSON, P.J., SIGMUND, K., SMITH, E.A., WEISSING, F.J. & YOUNG, H.P. (2003), *Group Report: The Cultural and Genetic Evolution of Human Cooperation*, in: P. Hammerstein (ed.) (2003), *Genetic and Cultural Evolution of Cooperation*, Cambridge (Massachusetts), The MIT Press: 445-468.
- HENRICH, J., BOYD, R., BOWLES, G., CAMERER, C., FEHR, E. & GINTIS, H. (eds.) (2004), *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, Oxford, Oxford University Press.
- HOBBS, T. (1651), *Leviathan*, edited by C.B. MacPherson, New York, Penguin Books, 1968.

- HODGSON, G.M. (1988), *Economics and Institutions: A Manifesto for a Modern Institutional Economics*, Cambridge, Polity Press.
- HUME, D. (1741), *Of the Independency of Parliament*, in: D. Hume (1963), *Essays, Moral, Political and Literary*, edited by T.H. Green and T.H. Grose, Oxford, Oxford University Press: 40-47.
- INGRAM, J.K. (1888), *A History of Political Economy*, New York, Macmillan.
- KAHAN, D.M. (2005), *The Logic of Reciprocity: Trust, Collective Action, and Law*, in: H. Gintis, S. Bowles, R. Boyd & E. Fehr (eds.) (2005), *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*, Cambridge (Massachusetts), The MIT Press: 341-378.
- KANT, I. (1785), *Groundwork of the Metaphysics of Morals*, in: I. Kant (1996), *Practical Philosophy*, translated and edited by M.J. Gregor, Cambridge, Cambridge University Press: 43-93.
- KOSLOWSKI, P. (ed.) (1985), *Economics and Philosophy*, Tübingen, J.C.B. Mohr.
- LE GRAND, J. (2003), *Motivation, Agency, and Public Policy: Of Knights and Knaves, Pawns and Queens*, Oxford, Oxford University Press.
- LOMASKY, L. (1992), *The Booth and Consequences: Do Voters Get What they Want?*, in: *Reason*, 24 (6): 30-34 .
- MANSBRIDGE, J.J. (ed.) (1990), *Beyond Self-Interest*, Chicago (Illinois), University of Chicago Press.
- MARGOLIS, H. (1984), *Selfishness, Altruism, and Rationality: A Theory of Social Choice*, Chicago (Illinois), University of Chicago Press.
- MARQUETTE, J.F. & Hinckley, K.A. (1988), *Voter Turnout and Candidate Choice: A Merged Theory*, in: *Political Behavior*, 10 (1): 52-76.
- MARX, K. (1845), *Theses on Feuerbach*, in: K. MARX & F. ENGELS (1975), *Collected Works: Volume 1*, London, Lawrence and Wishart: 13-15.
- MAYER, T. (1994), *Analytical Marxism*, London, Sage Publications.
- MCELREATH, R., CLUTTON-BROCK, T.H., FEHR, E., FESSLER, D.M.T., HAGEN, E.H., HAMMERSTEIN, P., KOSFELD, M., MILINSKI, M., SILK, J.B., TOOBY, J. & WILSON, M.I. (2003), *Group Report: The Role of Cognition and Emotion in Cooperation*, in: P. Hammerstein (ed.), *Genetic and Cultural Evolution of Cooperation*, Cambridge (Massachusetts), The MIT Press: 125-152.
- MCMILLAN, D.W. & CHAVIS, D.M. (1986), *Sense of Community: A Definition and Theory*, in: *Journal of Community Psychology*, 14 (1): 1-23.

- MERRILL, S. & GROFMAN, B. (1999), *A Unified theory of Voting: Directional and Proximity Spatial Models*, Cambridge, Cambridge University Press.
- MILL, J.S. (1836), *On the Definition of Political Economy; and on the Method of Investigation Proper to it*, in: J.S. Mill (1967), *Essays on Economics and Society*, edited by L. Robbins and J.M. Robson, London, Routledge and Kegan Paul: 309-339.
- MILLER, D. (1990), *Market, State, and Community: Theoretical Foundations of Market Socialism*, Oxford, Clarendon Press.
- MOSER, P.K. (ed.) (1990), *Rationality in Action: Contemporary Approaches*, Cambridge, Cambridge University Press.
- MUELLER, D.C. (1993), *The Public Choice Approach to Politics*, Aldershot, Elgar.
- MUELLER, D.C. (2003), *Public Choice III*, Cambridge, Cambridge University Press.
- NOZICK, R. (1974), *Anarchy, State, and Utopia*, Oxford, Blackwell.
- NOZICK, R. (1993), *The Nature of Rationality*, Princeton, Princeton University Press.
- NISBETT, R. & ROSS, L. (1980), *Human Inference: Strategies and Shortcomings of Social Judgment*, Englewood Cliffs, Prentice-Hall.
- ORR, G. (2002), *Ballot Order: Donkey Voting in Australia*, in: *Election Law Journal*, 1 (4): 573-578.
- OVERBYE, E. (1995), *Making a Case for the Rational, Self-Regarding, 'Ethical' Voter ... and Solving the "Paradox of not Voting" in the Process*, in: *European Journal of Political Research*, 27 (3): 369-396.
- PEREA, E.A. (2002), *Individual Characteristics, Institutional Incentives and Electoral Abstention in Western Europe*, in: *European Journal of Political Research*, 41 (5): 643-673.
- PETTTT, P. (1991), *Consequentialism*, in: P. Singer (ed.) (1991), *Companion to Ethics*, Oxford, Blackwell: 230-240.
- PETTTT, P. (1993), *The Common Mind: An Essay on Psychology, Society, and Politics*, New York, Oxford University Press.
- PETTTT, P. (1996), *Institutional Design and Rational Choice*, in: Goodin, R.E. (ed.) (1996), *The Theory of Institutional Design*, Cambridge, Cambridge University Press: 54-89.
- PETTTT, P. (2002), *Rules, Reasons, and Norms*, Oxford, Clarendon Press.
- PLAMENATZ, J. (1966), *The English Utilitarians*, 2<sup>nd</sup> revised ed., Oxford, Blackwell.

- PRESSMAN, S. (2006), *Clap Happy: Applause and the Voting Paradox*, in: *Journal of Economic Methodology*, 13 (2): 241-256.
- RAWLS, J. (1993), *Political Liberalism*, New York, Columbia University Press.
- RAWLS, J. (1999), *The Law of Peoples*, Cambridge (Massachusetts), Harvard University Press.
- RICHERSON, P.J. & BOYD, R.T. (2004), *Not by Genes Alone: How Culture Transformed Human Evolution*, Chicago, The University of Chicago Press.
- RICHERSON, P.J., BOYD, R.T. & HENRICH, J. (2003), *Cultural Evolution of Human Cooperation*, in: P. Hammerstein (ed.) (2003), *Genetic and Cultural Evolution of Cooperation*, Cambridge (Massachusetts), The MIT Press: 357-388.
- RICOEUR, P. (1990), *Soi-même comme un autre*, Seuil, Paris.
- RIKER, W.H. & ORDESHOOK, P.C. (1968), *A Theory of the Calculus of Voting*, in: *American Political Science Review*, 62 (1): 25-43.
- ROBBINS, L. (1935), *An Essay on the Nature and Significance of Economic Science*, 2<sup>nd</sup> revised ed., London, Macmillan.
- ROEMER, J. (ed.) (1986), *Analytical Marxism*, Cambridge, Cambridge University Press.
- ROUSSEAU, J.-J. (1762), *On the Social Contract; Discourse on the Origin of Inequality; Discourse on Political Economy*, translated by D.A. Cress and introduced by P. Gay, Indianapolis, Hackett Publishing Company, 1983.
- RUSSELL, B. (1950), *Unpopular Essays*, New York, Simon and Schuster.
- SCANLON, T.M. (1998), *What We Owe to Each Other*, Cambridge (Massachusetts), The Belknap Press of Harvard University Press.
- SCHMID, H.B. (2005), *Beyond Self-Goal Choice: Amartya Sen's Analysis of the Structure of Commitment and the Role of Shared Desires*, in: *Economics and Philosophy*, 21 (1): 51-63.
- SCHMIDTZ D. (1993), *Reasons for Altruism*, in: *Social Philosophy and Policy*, 10 (1): 52-68.
- SCHMIDTZ, D. (1994), *Choosing Ends*, in: *Ethics*, 104 (2): 226-251.
- SCHUESSLER, A.A. (2000a), *A Logic of Expressive Choice*, Princeton, Princeton University Press.
- SCHUESSLER, A.A. (2000b), *Expressive Voting*, in: *Rationality and Society*, 12 (1): 87-119.

- SEN, A. (1977), *Rational Fools: A Critique of the Behavioral Foundations of Economic Theory*, in: *Philosophy and Public Affairs*, 6 (4): 317-344.
- SEN, A. (1987), *On Ethics and Economics*, Oxford, Blackwell.
- SEN, A. (2005), *Why Exactly is Commitment Important for Rationality?*, in: *Economics and Philosophy*, 21 (1): 5-14.
- SILK, J.B. (2003), *Cooperation without Counting: The Puzzle of Friendship*, in: P. Hammerstein (ed.) (2003), *Genetic and Cultural Evolution of Cooperation*, Cambridge (Massachusetts), The MIT Press: 37-54.
- SIMON, H.A. (1957), *Models of Man: Social and Rational. Mathematical Essays on Rational Human Behavior in a Social Setting*, New York, John Wiley & Sons, Inc.
- SIMON, H.A. (1990), *Alternative Visions of Rationality*, in: P.K. Moser (ed.) (1990), *Rationality in Action: Contemporary Approaches*, Cambridge, Cambridge University Press: 189-204.
- SLOTE, M. (1989), *Beyond Optimizing: A Study of Rational Choice*, Cambridge (Massachusetts), Harvard University Press.
- SMITH, A. (1759), *The Theory of Moral Sentiments*, edited by K. Haakonssen, Cambridge, Cambridge University Press, 2002.
- SMITH, A. (1776), *An Inquiry into the Nature and Causes of the Wealth of Nations*, edited by R.H. Campbell, Oxford, Clarendon Press, 1976.
- SOBER, E. & WILSON, D.S. (1998), *Unto Others: The Evolution and Psychology of Unselfish Behavior*, Cambridge (Massachusetts), Harvard University Press.
- SOLOW, R.M. (1995), *Science and Ideology in Economics*, in: D.M. Hausman (ed.) (1995), *The Philosophy of Economics: An Anthology*, 2<sup>nd</sup> ed., Cambridge, Cambridge University Press: 239-251.
- STYRON, W. (1979), *Sophie's Choice*, New York, Random House.
- TAYLOR, M. (1988), *Rationality and Revolutionary Collective Action*, in: M. Taylor (ed.) (1988), *Rationality and Revolution*, Cambridge, Cambridge University Press: 63-97.
- TESTART, A. (1998), *Uncertainties of the 'Obligation to Reciprocate': A Critique of Mauss*, in: W. James & N.J. Allen (eds.) (1998), *Marcel Mauss: A Centenary Tribute*, New York, Berghahn Books: 97-110.
- THALER, R. (1994), *The Winner's Curse: Paradoxes and Anomalies of Economic Life*, Princeton, Princeton University Press.
- TRIVERS, R.L. (1971), *The Evolution of Reciprocal Altruism*, in: *Quarterly Review of Biology*, 46 (4): 35-57.

- TULLOCK, G. (1967), *Toward a Mathematics of Politics*, Ann Arbor, University of Michigan Press.
- TULLOCK, G. (1972), *Economic Imperialism*, in: J.M. Buchanan & R.D. Tollison (eds.) (1972), *Theory of Public Choice. Political Applications of Economics*, Ann Arbor, University of Michigan Press: 317-329.
- TULLOCK, G. (1976), *The Vote Motive: An Essay in the Economics of Politics, With Applications to the British Economy*, London, The Institute of Economic Affairs.
- TVERSKY, A. & KAHNEMAN, D. (1981), *The Framing of Decisions and the Psychology of Choice*, in: *Science*, 211 (4481), January 30: 453-458.
- VANDEVELDE, A. (1993), *Rationality and Reasonableness*, in: *International Studies in the Philosophy of Science*, 7 (1): 85-86.
- VANDEVELDE, A. (1994), *Economische en menselijke rationaliteit*, in: F. Verbeure & L. Apostel (eds.) (1994), *Verwijdering of ontmoeting? Verslagboek van de lezingenreeks Wetenschap en filosofie*, Kapellen, Pelckmans: 88-105.
- VAN LIEDEKERKE, L. (2002), *Hoe virtueel is Homo Economicus?*, in: X. Van Mechelen (ed.) (2002), *Afhankelijkheid zonder dominantie: over de sociale en politieke filosofie van Philip Pettit*, Leuven, Acco: 103-125.
- VAN PARIJS, P. (1990), *Le modèle économique et ses rivaux. Introduction à la pratique de l'épistémologie des sciences sociales*, Genève, Droz.
- VEBLEN, T. (1909), *The Limitations of Marginal Utility*, in: *Journal of Political Economy*, 17 (9): 620-636.
- VON HAYEK, F.A. (1944), *The Road to Serfdom*, London, Routledge.
- VON HAYEK, F.A. (1945), *The Use of Knowledge in Society*, in: *American Economic Review*, 34 (4): 519-53.
- VON MISES, L. (1920), *Economic Calculation in the Socialist Commonwealth*, translated by S. Alder, in: F.A. von Hayek (ed.) (1935), *Collectivist Economic Planning: Critical Studies on the Possibilities of Socialism*, London, Routledge: 87-130.
- VROMEN, J. (1998), *Economen en hun homo economicus: Striptease van een persona non grata*, in: *Krisis: Tijdschrift voor filosofie*, 70 (1): 20-37.
- WALSH, V. (1996), *Rationality, Allocation, and Reproduction*, Oxford, Clarendon Press.
- WEBER, M. (1914), *Economy and Society: An Outline of Interpretive Sociology*, edited by G. Roth and C. Wittich, Berkeley, University of California Press, 1968.

WICKSTEED, P.H. (1910), *The Common Sense of Political Economy*, edited and introduced by L. Robbins, London, Routledge and Kegan Paul, 1933.

ZYWICKI, T.J. (2000), *Was Hayek Right About Group Selection After All? Review Essay of Unto Others: The Evolution and Psychology of Unselfish Behavior by Elliott Sober and David Sloan Wilson*, in: *Review of Austrian Economics*, 13 (1): 81-95.

