
**THE SPONTANEOUS ORDER AND
PREFERENCES FOR FLEXIBILITY**

Carlo Ludovico Cordasco

A thesis submitted in partial fulfilment of the requirements for the
degree of Doctor of Philosophy

Department of Politics

University of Sheffield

March 2019

Contents

Abstract	3
Acknowledgments	4
Introduction	6
The Spontaneous Order as a Cooperative Enterprise	26
Introduction	26
Dore vs Bakewell	30
Complexity	32
Governing Complexity and the 'Organic' Account of the Spontaneous Order	35
A non-organic account of the Spontaneous Order	42
Certainty and Spontaneity	51
Hayek's Utilitarianism	55
Conclusion	58
The Spontaneous Order and Its Instrumentalist Character: A Cri- tique	61
Introduction	61
The Lockean Account	63
The Instrumentalist Character	70
A Critique	73
Derivative distributions of rights and the Instrumentalist Character .	82
The Asymmetry Objection	84
Conclusion	87
Social Norms and The Instrumentalist Character of The Sponta- neous Order	89
Introduction	89
Three Ways in which Social Norms shape our Communities	91

Identifying Social Norms	95
Social Norms and Coercion	103
Social Norms and The Spontaneous Order	108
A Second Asymmetry Objection	114
Conclusion	117
Preferences For Flexibility and The Spontaneous Order	119
Introduction	119
Social Environment Shaping Rights	120
Emergent Justifications	124
Contractarianism: Abstraction and Instrumentalism	133
Preferences For Flexibility	147
Preferences for Flexibility, Social Environment Shaping Rights and Bargaining Procedures	159
Conclusion	163
Preferences For Flexibility and The Framework For Utopia	165
Introduction	165
Agreement Through Preferences For Flexibility	166
The Persistence of Disagreement	171
The Framework For Utopia	176
The Minimal State as a Meta-Utopia	182
Is the Minimal State the best means to approximate Utopia?	187
The More-Than-Minimal State in Approximating Utopia	192
Conclusion	196
Conclusion	198
.	198
Works Cited	210

Abstract

Designing political institutions able to secure the conditions for cooperation among persons who exhibit widely diverse perspectives, underpinned by different structures of preferences and goals, constitutes, perhaps, the greatest challenge of contemporary political theorising. F. A. Hayek's theory of the spontaneous order provides one key insight on what political institutions of complex societies should look like: since only individuals know their structures of preferences, goals and their relevant circumstances, theorists and planners, by virtue of their inability to collect such dispersed knowledge, cannot design and devise and devise fine-grained systems of rules aimed at defining people's specific terms of cooperation. In order to create the conditions for cooperation in complex societies, Hayek suggests, we must design political institutions that define a large protected sphere of actions, which equips individuals with the ability to shape their specific terms of cooperation on the basis of their own local knowledge. Hayek identifies such an institutional arrangement in the protection of Lockean rights to life, liberty and property.

In this thesis, I attempt to show that an aptly modified version of Hayek's theory of the spontaneous order lends itself to a contractarian justification. In particular, I aim to demonstrate that individuals, who are uncertain about their future preferences and goals, have instrumental reasons to converge on institutional arrangements that define a large protected sphere of action, which allows them to define their specific terms of cooperation on the basis of local knowledge which will unveil to them during their life paths. In fact, individuals' uncertainty about their future structures of preferences and goals elicits the emergence of preferences for flexibility, which make them abstract from their current set of preferences and goals when facing a social contract bargaining scenario, and invite them to choose institutional arrangements which leave the door open for adaptation to changes in their future identities. Preferences for flexibility, I aim to show, can solve much political disagreement stemming from people's seemingly incompatible structures of preferences and goals.

Acknowledgments

Although I loved my doctoral research, these years have not been easy. Unjustified depression and (more than) unjustified euphoria are, I take it, constitutive parts of most Ph.D journeys, to the point that one always comes to think about dropping out. If I did not, the merit is, first and foremost, of my two supervisors: Matt Sleat and Alasdair Cochrane. They have been incredibly patient with my huge delays and always prone to encourage me in the worst moments. Matt, in particular, has followed me at each and every step, reading many incomprehensible documents I sent for revisions, and providing amazing feedback. His willingness to help has taught me much about working in academia, and if I will have the chance to continue with my academic career, I will make sure to follow his example. Both on a personal and on a professional level I could not have hoped for a better supervisor.

I also wish to thank all the academic and administrative staff of the Department of Politics at The University of Sheffield. It has been an incredible privilege to feel integral part of this friendly and intellectually vibrant academic community. Sarah Cooke, in particular, has provided incredibly valuable help, support and guidance during these years.

I am grateful to Sebastiano Bavetta, Pietro Navarra and Cristina Bicchieri for inviting me to take part in a John Templeton's sponsored project at the University of Pennsylvania, where I had an amazing time working in a great interdisciplinary research group. Sebastiano, in particular, has been a fantastic mentor and a good friend and I hope to have the possibility to continue working with him in the coming years.

I am indebted to Mario Rizzo, Richard Epstein and Nick Cowen for offering me a visiting fellowship at the New York University School of Law, where I spent invaluable time discussing the contents of this thesis with a great group of scholars who are now also very close friends.

I also owe a big thank you to Raimondo Cubeddu. I developed my interest in the theory of spontaneous order under his supervision at the University of

Pisa, and the present work is, at least in part, the result of what I have learned during those years. Raimondo is a father to me, and I sincerely hope to have other many pleasant philosophical discussions with him in the coming years.

I wish to say thank you to anyone who has provided meaningful feedback on single chapters of the thesis through endless personal or email exchanges. In particular, I must thank Adam Bartha, Jerry Gaus, Martin Van Hees, Billy Christmas and the late Carlo Argenton. Carlo's wonderful and gentle soul will always remain in my heart.

I also had the privilege of receiving great feedback from my two Ph.D. examiners: Mark Pennington and Edward Hall. Their suggestions were illuminating and brought to me to reflect on aspects I had entirely neglected. I do believe that the present thesis is a much better work as a result of their criticisms.

In the past few years, I lost two very important figures in my life: my aunt Rossana and my grandma Sisa. They happened to be the greatest supporters of everything I wrote about, though they knew very little about it. I will keep their memory alive in the hearts of all I can reach.

Last, but not least, I wish to thank my parents, Ketty and Giancarlo, and my wife Dalia. I am incredibly grateful for their unconditional love and support. I wish to dedicate the present work to them.

Introduction

An individual Jesuit is to the utmost degree of abasement the slave of his order though the order itself exists for the collective power and importance of its members.

- John Stuart Mill, *On Liberty*

John Stuart Mill's rich and multifaceted account of liberalism has largely inspired my research during the Ph.D. In fact, although only a few references to his writings will be appearing in the next chapters, the present work takes as its constitutive aim that of responding to two main questions arising from the above quoted passage. Namely, if social orders exist through and for their members, why do we often observe evolutionary patterns leading to states of affairs in which members of a certain association become mostly subjects of formal and informal rules, rather than active shapers of their social environment? And, how could we design our social orders in a way that would make them cooperative enterprises from which everyone stands to gain?

If we follow Mill in maintaining that social orders exist or should exist primarily for their members, we should also conclude that our search for the constitutive aims of our communities cannot lie outside the domain of preferences, interests and goals that we, qua members or would-be members of our communities, exhibit. In fact, these preferences, interests and goals cannot be merely regarded as features one may plausibly ignore in delineating the terms of our cooperation, as there is no other plausible standpoint from which to start our political theorising.

If we follow this line of thought, communities exist, or should exist, precisely to the extent that we, qua members of our communities, have instrumental reasons to build, join them and comply with their rules, in that most of our personal aims and enterprises depend upon cooperative schemes operating in the background. As such, the cooperative schemes one attempts to design are supposed to meaningfully respond to what we are and what we aim to achieve.

Therefore, we should be worried about states of affairs in which adaptation to these preferences, interests and goals is undermined. In particular, we should be worried about the emergence of institutional arrangements which lack, at the very outset, or lose, in due course, appeal from the standpoints of many fellow members of our communities. Such a worry, I suggest, is two-fold. First, there is a normative concern about social orders failing in attaining one of their main constitutive aims, which is to advance our chances to successfully pursue our life plans that provide meaning to our human endeavour on earth. Second, we should be concerned with how such a failure could potentially undermine the stability of our communities. In fact, members who fail to see their communities as cooperative enterprises from which they stand to gain could fail to form reasons to comply with their rules.

One striking example of the sort of phenomena Mill was concerned with is offered by Elizabeth Anderson's book *Private Government*:

Consider some facts about how employers today control their workers. Walmart prohibits employees from exchanging casual remarks while on duty, calling this "time theft." Apple inspects the personal belongings of their retail workers, who lose up to a half-hour of unpaid time every day as they wait in line to be searched. Tyson prevents its poultry workers from using the bathroom. Some have been forced to urinate on themselves, while their supervisors mock them. About half of U.S. employees have been subject to suspicionless drug screening by their employers. Millions are pressured by their employers to support particular political causes or candidates. If the U.S. government imposed such regulations on us, we would rightly protest that our constitutional rights were being violated. But American workers have no such rights against their bosses. Even speaking out against such constraints can get them fired. So most keep silent. (Anderson, 2017 p. xix).

Anderson's aim, in the book, is to explore the peculiarities of the labor mar-

ket in a free market economy, and to spell out how its main constitutive features differ from those of ordinary markets in which buyers and sellers walk away after the transaction is over. In particular, Anderson is concerned with the sort of formal and informal hierarchies, partly arising from differences in bargaining power between employers and employees, and with how such hierarchies could likely generate states of affairs in which many of our fellow citizens silently and, more or less, voluntarily become subject to schemes of rules from which they have little to gain and much to object to.

Anderson's worry is not merely highlighted by considerations of justice with respect to workers' conditions, but also driven by more mundane concerns about stability. In fact, we could hardly expect our communities to be peaceful cooperative enterprises when formal or informal distributions of rights among our fellow citizens are such 'so as to systematically undermine the interests of identifiable groups of people in serious or gratuitous ways' (Anderson, 2017, p. xxii).

Examples of this sort are also frequently found at the political level, where lack of institutional adaptation to citizens' structure of preferences and goals is somewhat systematic. For instance, consider the case of Piergiorgio Welby, Italian poet and painter, who was diagnosed with muscular dystrophy when he was only seventeen. In 1997, the progression of the disease was such that he had to resort to mechanical ventilation, artificial feeding and a speech-synthesiser to communicate. In 2006, he was able to send a video letter to the president of the Italian Republic asking for the interruption of mechanical ventilation and artificial feeding under sedation. However, in Italy, there was and still is no law regulating assisted suicide such that it is hard to tell whether aiding someone in such circumstances is against the law.

A poll made in 2006 revealed that the 64% of Italian voters were in favour of some form of euthanasia,¹ and such a tendency has substantially increased over the past few years with supporters of euthanasia reaching the 74% in 2017.² Yet, despite a widespread consensus over the possibility of assisted suicide, forms of

¹See Veltri (2006).

²Today (2017).

institutional recognition are found wanting. In this regard, in 2017, Fabiano Antoniani, a former dj, victim of a terrible car accident, which left him blind and tetraplegic, had to travel to Switzerland in order to have assisted suicide performed on him. Interestingly, this was made possible by the help of an Italian MP, who has been under investigation for the crime of instigation to suicide.³

Sometimes, lack of adaptation to people's structure of preferences and goals is driven by mechanisms widely studied in social psychology. Pluralistic ignorance, for instance, explains, in part, the persistence of a wide range of despicable and sum-negative practices going on in many communities.⁴ In this regard, Cristina Bicchieri's important and large body of theoretical and empirical work on social norms has shown that the persistence of norms of discrimination against blacks, or of norms of revenge, or of norms sustaining practices such as Female Genital Mutilation and corruption, is partly driven by mistaken beliefs about other people's normative attitudes on what we ought to do in certain circumstances.⁵ For instance, a recent empirical work analysing the persistence of corruption in Nigeria has shown that some of determinants of corruption can be traced to people's mistaken beliefs about other people's normative beliefs on the opportunity of bribing public officials within certain circumstances.⁶ Were we able to express our intimate convictions about sensible cases, without the fear of social sanctions, we would be also able to respond efficaciously to many instances of lack of either formal or informal institutional adaptation.

Some other times, lack of adaptation and the persistence of despicable practices rests on path-dependence,⁷ and collective action problems which stand in

³The stories of Piergiorgio Welby and Fabiano Antoniani have been widely covered by international media such as The New York Times. See, in particular Fisher (2006); Povoledo (2017).

⁴One of the first formulations of the concept of pluralistic ignorance can be found in Miller and McFarland (1987).

⁵See, in particular, Bicchieri (2005) which illustrates her general approach toward social norms, and Bicchieri (2016), which tackles more empirical aspects and lays out diverse methods to measure norms and detect pluralistic ignorance.

⁶Koni Hoffmann and Patel (2017).

⁷A large body of literature focuses on path-dependence as the key determinant of lack of institutional change. Among others, North (1991) and (2012) provide analytic tools to investigate the ways in which path-dependence influences the evolution of institutions.

the way of converging on better states of affairs.⁸ Often, though, lack of adaptation is driven by nothing more than mere coercion. Political elites, whether autocratic or democratically elected, are seldom either detached from or unconcerned with citizens' preferences and goals. In democratic contexts, in particular, such a detachment is often explained with reference to problems connected to incentive-compatibility,⁹ or attributed to citizens' rational ignorance,¹⁰ or to the cartellisation of political parties,¹¹ or to parties' failure in conveying political demands coming from our fellow citizens.¹²

Other times, on the other hand, we simply disagree on what social and political institutions should look like. Such disagreement, whether grounded in different conceptions of the good or merely shaped by different interests, transfers to the social and political arena and often generates states of affairs that do not really resemble compromises between different standpoints, either because our views about social institutions are non-compossible, or because preference aggregation problems tend to create winners and losers.¹³

The present thesis constitutes an attempt to theorise about the cooperative schemes of our communities by taking individual preferences and goals seriously. In particular, it starts from the key assumption I have previously highlighted: social orders can be regarded as meaningful cooperative enterprises only to the extent that we, qua would-be members, have reasons to build, join them and comply with their rules, and such reasons cannot but be grounded in what we stand to gain from participating to them. In these terms, the search for institutional arrangements that could secure the conditions for cooperation consists primarily in a search for systems of rules which members of the community,

⁸For instance, a recent article by Christopher Coyne and Rachel Coyne (2014), illustrates how the currently widespread phenomenon of Female Genital Mutilation could often be plausibly explained in terms of sub-optimal strategic equilibrium.

⁹Public choice theory, especially through the important work of Stigler (1971), Peltzman (1976) and Buchanan and Tullock (1962), emphasises how citizens struggle in controlling political elites.

¹⁰See, for instance, Caplan (2011).

¹¹See, in particular, Kitschelt (2000).

¹²Urbinati, among others, has largely shaped the research agenda on the failure of political parties as a main determinant of the crisis of democracy; see, in particular, Urbinati (2014).

¹³This is what mostly emerges from the literature on social choice theory, which mostly developed after Arrow (1950).

with their own diverse standpoints, underpinned by different sets of interests, preferences and goals, would abide by on the ground of their instrumental reasons.

Crucially, the search for such an institutional framework cannot merely concern and accommodate for widespread diversity in a static fashion; which is to say that, although diversity and disagreement constitute permanent and plausibly ineliminable features of our world, their character is subject to evolutionary processes. In fact, our structure of goals and preferences is constantly subject to change, and such a dynamic character is often driven by moral innovation, which presents us with new and sometimes appealing ways of life, and by technological innovation, which enlarges the set of opportunities and aims we may want to pursue.¹⁴

Such a dynamic character of our structure of preferences and goals, which I take as another ineliminable fact of our human endeavour on earth, should make us reflect on that agreements or convergence (whether actual or hypothetical) on social institutions by no means resembles a one-shot game in which, once and if the agreement is secured, we have finally responded to the 'first political question', to borrow Bernard Williams' words.¹⁵ In fact, dynamic structures of preferences and goals require modular changes to our terms of cooperation. As such, platforms that proved helpful in allowing us to peacefully interact with one another, with the prospect of mutual advantage, may not prove helpful when circumstances change. In this regard, our search for institutional arrangements, which are able to secure the conditions for cooperation, constitutes an endless enterprise, whose indefinite character, I suggest, partly shapes the attractiveness of political theorising.

I join many prominent authors in claiming that much political theorising has been largely unconcerned with people's actual structure of preferences and goals

¹⁴Ideally, technological innovation should drive toward the enlargement of our set of opportunities; yet, the challenges posed by climate change, for instance, seem to lead us toward the opposite direction.

¹⁵Williams claims: "I identify the "first" political question in Hobbesian terms as the securing of order, protection, safety, trust, and the conditions of cooperation. It is "first" because solving it is the condition of solving, indeed posing, any others". (Williams, 2005, p. 3).

in laying out cooperative schemes we should live by. For instance, many theorists who have mostly shaped the research agenda within the literature of political philosophy in the past fifty years - of which John Rawls' large and important body of work before *Political Liberalism*,¹⁶ constitutes only the most extended and persuasive instance - often, although not exclusively, theorise about cooperative schemes and institutional arrangements by looking at what sort of demands or normative requirements would arise under idealising assumptions concerning individuals' preferences and goals.

Such idealising assumptions are made explicit by reference to moral values that individuals allegedly internalise, such as the value of moral equality of persons, and which should importantly constrain our structures of preferences and goals, thus facilitating our converge on a common set of institutions. Ryan Muldoon, in his recent and inspiring book *Social Contract Theory for a Diverse World: Beyond Tolerance*,¹⁷ borrows from Thomas Nagel the label of 'views from nowhere' in characterising these approaches, in that authors such as John Rawls, Thomas Scanlon and, indeed, Thomas Nagel, develop accounts of justice laying out terms of cooperation we would endorse if, and only if, we could somehow abstract from our own actual circumstances, such as our conceptions of the good life, our identities, our structures of preferences and goals, and internalise values which would equip us with reasons to converge on certain cooperative arrangements.¹⁸

The widespread diffusion of such an approach, I believe, partly explains the revival of *Political Realism* and the emergence of *Practice-Dependent* theorising within the political philosophy literature.¹⁹ In fact, it is rather intuitive to interpret the emphasis on the constitutive features of politics (e.g. people's beliefs, preferences, goals, social practices, conventions, existing institutions, etc.),

¹⁶I mostly refer to Rawls' works following the publication of Rawls (1958) and later included in Rawls (1971).

¹⁷Muldoon (2016).

¹⁸I refer, in particular, to Rawls (1971), Scanlon (1998) and Nagel (1989).

¹⁹In particular, I refer to the recent and growing literature following the posthumous publication of Williams (2005). See, in particular, Galston (2010), Sangiovanni (2008), Rossi and Sleat (2014). Practice-sensitive approaches, though, may also include Gaus (2010) and Muldoon (2016).

which realists and practice-dependent theorists take as supposed to play a crucial role in defining our terms of cooperation, as a reaction to their discontent with the pervasive negligence of the actual circumstances of politics in identifying normative requirements that political orders should meet in order to qualify as just or legitimate.²⁰

The present work, in this respect, belongs to this particular line of thought, in that it reacts to views from nowhere in attempting to delineate one way of finding cooperative schemes, to which we may be willing to converge on, without idealising assumptions that demand us to deliberate about political institutions through lenses which are not intimately ours. Importantly, this is not, in any way, an attempt to discredit the conceptual relevance of views from nowhere. In fact, I can only admire the academic enterprises of these great philosophers, and should be content with achieving the smallest part of what they have accomplished in their intellectual endeavours. Nor I believe that views from nowhere lack normative relevance under the current circumstances of politics. In fact, I join John Maynard Keynes in making explicit my own convictions about the fact that:

The ideas of economists and political philosophers, both when they are right and when they are wrong are more powerful than is commonly understood. Indeed, the world is ruled by little else. (Keynes, 1936, p. 383).

Ideal or utopian theorising, on the one hand, and practice-sensitive theorising, on the other, have, I believe, quite different constitutive aims: the former attempts to inspire us, qua members of our social communities, by spelling out a number of values we should internalise in order to imagine and make intimately

²⁰It is important to point out that realist and practice-dependent theorists offer widely different interpretations of the constitutive features of politics from which to start our theorising. For instance, Sangiovanni (2008) distinguishes between two different approaches to practice-dependency, cultural conventionalism and institutionalism, which reflect a commitment to two distinct kinds of constitutive features (i.e. cultural conventionalists focus on beliefs, cultural practices, etc.; institutionalists look at formal institutional arrangements taking place in a particular community). Similarly, realists such as Sleat, would suggest that people's preferences and beliefs, far from representing constitutive features of politics, are merely contingent aspects of our political orders (see, in particular, Sleat, 2016).

ours an attractive, though demanding, account of cooperation; the latter, perhaps less nobly, aims at responding to one more pressing question, which is: how are we to design our social institutions, given what we actually are?²¹ Surely, there is much appeal in imagining perfect terms of cooperation in a world of quasi-perfect persons, and the sort of *perfectionism* that is involved in such ideal theorising should make us constantly strive for improving the kinds of persons we are.²²

At the same time, though, division of labor demands to those of us who are, perhaps, less imaginative, to think about how to make a virtue out of necessity; which is to imagine and theorise on the best cooperative schemes consistent with the actual constitutive features of our world. In this respect, I strongly believe that, while ideal theorising has much to say about the sort of persons we should aim to be but less so with respect to how institutions should look like, here and now, practice-sensitive approaches have much to tell us about how institutions should look like, but less so with regard to the kinds of persons we should aim to be.

This particular aspect, I believe, brings with it a number of implications on the sort of assumptions one is allowed to make when theorising about political institutions in the attempt to take individual preferences and goals seriously. The most important of which is that we must take our world as it is, with its scarcity of resources, its transaction costs, with our more or less pronounced self-interest and our pervasive uncertainty about the sort of persons we will be in our life paths. In fact, the present work attempts to take these elements as its inevitable starting points.

However, one thing in which it indulges is in its reliance on the concept of rationality, to which I will often resort in delineating my suggestions for thinking about institutions from such a standpoint. In this regard, I am in need to spend

²¹By practice-sensitive theorising, I mostly refer to theories that take people's beliefs and values as their starting point. Once again, I have in mind Gaus (2010) and Muldoon (2016).

²²On exemplarity, see in particular Ferrara (2008), which has profoundly shaped my ideas about the relevance of ideal theorising. Alongside Ferrara, many other authors have influenced my views on ideal theory, among those: Schmidtz (2011), Gaus (2016), Estlund (2014), Hamlin and Stemplowska (2012), Simmons (2010).

a few concise words about the sort of rationality I have in mind, as very little will be said in the following pages. In particular, I wish to emphasise that the present work should be read as mostly, though unconventionally, contributing to the literature on rational choice theory concerning institutional arrangements or social contract theory, where the concept of rationality is spelled out in terms of a person's capacity to form a certain life plan and choose the best available means to pursue it, on the basis of one's rational beliefs on how to achieve it.²³

In particular, in delineating cooperative schemes and institutional arrangements that I regard as appealing from various diverse standpoints, I will largely rely on the assumption of persons' rationality in looking at and evaluating them. Such an assumption, though, when applied to social contract theorising, is, perhaps, no less demanding than other assumptions concerning people's internalisation of moral values, as it requires, on our part, an effort in thinking about social institutions in an instrumentalist fashion, that is in judging their appropriateness on the basis of how they advance our interests and facilitate the attainment of our goals. In this regard, those of us who believe that the main constitutive aim of social institutions is to realise an account of justice, which somehow ignores the advancement of one's life plans, will find the present work of little interest. Yet, if we take individual preferences and goals seriously, the answer to the question: 'what should social institutions look like?' must emerge from our rational deliberation about which sets of institutions would best advance our aims.

Perhaps, such an enterprise requires more than mere rationality. David Gauthier, for instance, frames it in the language of morals when he asks: "What theory of morals can ever serve any useful purpose, unless it can show that all the duties it recommends are truly endorsed in each individual's reason?" (1986, p. 1). In this regard, if social contract theory is a moral enterprise, or starts from moral assumptions, it substantively differs from other approaches, in that

²³Such a definition of rationality is rather conventional and shows no significant differences from that of other authors belonging to the social contract tradition. For instance, Rawls (1999) states that "our decision is perfectly rational provided that we face up to our circumstances and do the best we can", p. 349.

it assigns normative relevance to people's actual preferences and goals, and takes as its main normative aim that of searching for institutional arrangements that would foster our ability to cooperate as a means to pursue goals embedded in our life plans.

Such a normative standpoint, I should emphasise, is by no means original. We can, indeed, enlist many approaches within modern and contemporary political thought that assign to individual preferences and goals such a strong normative weight in defining our terms of cooperation. Chief among those, Thomas Hobbes' contractarian approach in laying out the structure of the *Leviathan* and John Locke's account of the emergence of Civil Government from the state of nature (1689). I should also mention instrumentalist accounts of democracy dating back to Mill (1861), and more recently developed by theorists such as Robert Putnam (1994), David Estlund (2009) or Helene Landemore (2012); theories of justice as mutual advantage such as David Gauthier's (1986) or Russell Hardin's (1988), de-idealised accounts of public reason such as Gerald Gaus (2010), classical liberal theories which include also consequentialist justifications of political institutions such as Robert Nozick's *Framework For Utopia* (1974),²⁴ Chandran Kukathas' *Liberal Archipelago* (2003), Anthony De Jasay's contractarianism (1989), James Buchanan's constitutional liberty (1991), theories of the spontaneous order such as Friedrich Hayek's (2012) and Robert Sugden's (2004, 2018), and, especially, recent social contract theories such as Muldoon (2016) and Moehler (2018). In fact, all these authors share, at least, a partially instrumentalist view of social institutions, according to which one of the constitutive aims of our communities is to advance individuals' interests by creating the conditions for cooperation. This is not to deny that we may endorse certain institutional settings on different grounds, but merely to highlight that responsiveness and adaptation to preferences and goals do play a crucial normative role when designing the institutions of our communities.²⁵

²⁴Clearly, Nozick's main argument for the minimal state does not take a consequentialist form. Yet, the framework for utopia, which he regards as an alternative justificatory framework for the minimal state, contains consequentialist elements.

²⁵For instance, democratic theorists often integrate instrumentalist justifications of demo-

In the following pages, I shall be primarily concerned with Hayek's work and, in this respect, one main preliminary methodological remark is in order: this thesis does not attempt to reconstruct the rich and interesting historical and intellectual context surrounding Hayek's endeavour in highlighting his important contributions to political theory. In fact, the aim of this thesis is only minimally historical. In this respect, to borrow Bernard Williams' words, the present work could be read as an instance of engaging with 'history of philosophy' as opposed to 'history of ideas',²⁶ where the former identifies attempts to reflect on political and philosophical concepts in an abstract fashion, in an effort to highlight their normative or explanatory force in different contexts, and the latter refers to historical approaches to ideas with the aim of emphasising the contextual elements that contributed to their emergence.

In fact, I'll be focusing on one specific object of Hayek's inquiry, that is his theory of the *spontaneous order*, and on the epistemological premises that underlie his approach, in an effort to propose certain amendments and offer original insights on the problem of cooperation in a diverse and bounded world. This, I should stress, is by no means a new methodology in political philosophy, and many theorists have succeeded in rescuing and reformulating philosophical concepts in an effort to explore and highlight their relevance in addressing issues that are of pressing interest to the academic community.²⁷

However, one main question immediately arises in the mind of the experienced political philosopher, a question which I undoubtedly owe an answer to: why the spontaneous order? After all, one might notice, Hayek's intellectual endeavour does not, at least explicitly, belong to the social contract tradition, nor has he dealt with normative questions about the relevance of individual preferences and goals in shaping social institutions, nor can he be regarded as

cratic decision making procedures with principled justifications based on the value of political equality (e.g. Estlund 2009).

²⁶See Williams (2002).

²⁷The list is incredibly large, but I would like to mention a few works that had a relevant impact during my personal and philosophical journey during the Ph.D.: G. A. Cohen's *Karl Marx's Theory of History: A Defence*; Gerald Gaus' *The Order of Public Reason*; Ken Binmore's *Natural Justice*; John Tomasi's *Free Market Fairness*.

major figure in contemporary political philosophy debates.

In this respect, there are three main reasons I wish to offer. First, Hayek's account of the spontaneous order constitutes, although not explicitly, an instance of theorising around social contracts. In particular, I interpret the main constitutive aim of the theory of spontaneous order as that of searching for institutional arrangements which are able to create the conditions for fruitful cooperation among fellow members of a community who exhibit diverse preferences, goals and aims. Such a formulation, I suggest, lends itself to contractarian justifications, as the institutional arrangements Hayek attempts to design could, ideally, be endorsed from a wide variety of standpoints.²⁸ Moreover, the emphasis on institutional adaptation to changing circumstances, which is pervasive in Hayek's work,²⁹ aptly responds to the purpose of thinking about social institutions in a dynamic fashion, which is by imagining social institutions that can be endorsed from widely different standpoints *over time*, in spite of the dynamic character of the structures of preferences and goals exhibited by members of a community.

Second, Hayek's account of the spontaneous order is, perhaps, the most sophisticated inquiry into the search for cooperative social institutions, within complex scenarios such as pluralistic societies. Throughout his large body of works, Hayek tackles crucial issues concerning the impossibility of designing from scratch, and in full details, meaningful systems of rules aiming at regulating our interactions in a cooperative manner, as theorists or planners are unable to collect local knowledge dispersed among various members of our communities, and to act on the basis of it. In this respect, his inquiry into social orders constitutes an attempt to theorise about institutional arrangements which are able to create the conditions for cooperation without relying on central planning, and by using knowledge that is dispersed among fellow members of our communities.

²⁸A similar interpretation of Hayek's theory of the spontaneous order has been provided by Sugden (1993) and Lister (2013).

²⁹Especially in Hayek's earlier works on the distribution of knowledge in complex systems. See, in particular, Hayek (1945).

The great lesson we draw from Hayek is, I suggest, that if we wish to take individual preferences and goals seriously in designing institutional arrangements of cooperative communities, we are to imagine systems of rules that leave in the hands of individuals the ability to shape their specific terms of cooperation. This particular aspect, I hope to show, has interesting implications on social contract theorising, in that it invites us to converge or agree on a set of institutions which do not lay out, ex-ante, our specific terms of cooperation. Our agreement, indeed, should rather concern general sets of institutions which define a protected sphere within which we are able to act in the pursuit of our own goals, on the basis of our own local knowledge. Hayek envisages such an institutional arrangement in the protection of the *Lockean triad*, which is in the protection of rights to life, liberty and property. The underlying idea is that such a protected sphere would allow individuals, exhibiting different and dynamic structures of preferences and goals, to continuously negotiate their specific terms of cooperation, on the basis of their local knowledge.

Third, theories of the spontaneous order, I suggest, equip us with great tools for designing basic institutional arrangements that would foster our ability to cooperate also at the *meta-community* level. This is the case of Robert Nozick's discussion on the *Framework For Utopia*, where, in proposing us the minimal state as the institutional arrangement which best approximates Utopia, he suggests that the protected sphere instantiated by the distribution of rights underpinned by the minimal state, allows us to create the conditions for the emergence of countless communities that would aptly respond to people's diverse structures of preferences and goals.³⁰

All these elements persuaded me to start my investigation around social contract theory through the lenses and with the tools of Hayek's theory of the spontaneous order, and, in many regards, the present work could be read as an attempt to frame Hayek's work in social contract terms, and analyse the difficulties it faces and how it could be improved.

The idea of casting Hayek's theory of the spontaneous order in social contract

³⁰See Nozick (1974), ch. 10.

terms, I suggest, has two main merits. First, there is a question of authenticity of interpretation. In fact, despite numerous attempts to characterise Hayek's work as ultimately, though unconventionally, belonging to the utilitarian tradition,³¹ I regard the theory of the spontaneous order as primarily concerned with how institutions should distribute opportunities to members of a community in an effort to foster cooperation from widely different standpoints. Perhaps, the final aim of cooperation consists in maximising aggregate utility, but, if it is so, such an aim is pursued by means of designing institutions that (nearly) all members of a given community regard as satisfactory, such that they would continuously find instrumental reasons to sustain and comply with the cooperative system of rules.

Second, and most importantly, social contract theory would add another layer of justification to Hayek's political theorising. In particular, if the theory of the spontaneous order can be sensibly framed in these terms, we would succeed in making it appealing from a different and sound normative standpoint. Although I do not address meta-ethical issues in the present work, I wish to stress that I regard the enterprise of offering diverse justifications for a given account of political authority, and its institutional declination, as a central task of political theory.³² In particular, I find persuasive the idea according to which if diverse normative standpoints resist reasonable philosophical scrutiny, and their free-standing plausibility is ultimately affirmed, there is some value in assessing political theories on their ability to soundly respond to intuitions emerging from multiple normative standpoints.

Before I start with the present investigation, I wish to provide a brief summary of the present thesis in order to provide a general overview of its contents. In this respect, in the first chapter, I will lay out my interpretation of Hayek's account of the spontaneous order, in an effort to illustrate his contributions to the research on cooperative institutions within complex social orders, and in order to highlight the contractarian aim which lies in the background of his project.

³¹See, in particular, Gray (1982), Lukes (1997) and Johnston (1997).

³²Such a consideration, I must emphasise, has been in some ways inspired by Johnatan Dancy's *Ethics without Principles*.

In drawing such an interpretation, I will posit particular emphasis on one main normative element that I find emerging from the idea of the spontaneous order, which consists in the crucial relevance that institutional adaptation to people's preferences and goals should play in defining our cooperative schemes. In fact, I look at spontaneity as a property that social orders possess in their actual operations, and which consists in their ability to create the conditions for cooperation by defining a protected sphere of action which allows members of the community to shape their specific terms of cooperation in due course, according to their dynamic structures of preferences and goals.

However, despite the various merits that I attribute to Hayek's account of the spontaneous order, I will attempt to show that we have reasons to be ultimately unsatisfied with his proposals concerning institutional arrangements. In particular, I am unconvinced by the claim according to which the protection of the Lockean triad would represent, *sic et simpliciter*, the best institutional arrangement in allowing us to design cooperative schemes to which members of a community continuously have reasons to abide by. In this respect, I shall offer two main arguments which will be the object of the second and the third chapter. First, such an institutional arrangement could potentially lead to distributions of rights which may systematically prevent some members of our community from shaping their terms of cooperation, according to their dynamic structures of preferences and goals. This particular problem could lead to states of affairs in which some members cease to see their communities as cooperative enterprises from which they stand to gain, thus failing in meeting the main constitutive aim of the theory, and undermining the stability of their communities. Second, I shall attempt to show that a similar effect is brought about by informal institutions such as social norms. In particular, certain social norms, by increasing the costs for undertaking certain courses of action, may alter our preference ranking over the available strategies, thus compromising our ability to shape our specific terms of cooperation in ways that meaningfully respond to our structures of preferences and goals. When such alterations become systematic, some members may cease to see their community as a cooperative enterprise from which

they stand to gain, thus undermining its stability.

These particular problems give rise to two main desiderata that an amended version of the theory of the spontaneous order should meet. First, while preserving a protected sphere which would allow members of the community to shape their specific terms of cooperation, distributions of rights should be designed and devised so as to preserve a certain *sufficientarian* distribution of what I shall refer as *social-environment shaping rights*. This broadly defined set, includes all those rights whose exercise makes members of a community continuously able to define their specific terms of cooperation according to their dynamic structure of preferences and goals (e.g. economic rights, rights of expression). In this regard, the sufficientarian threshold that I envisage aims at equipping every member with the ability to shape their terms of cooperation, so as to preserve the instrumentalist character of their communities. In particular, such a sufficientarian threshold has as its main aim that of preventing the emergence of states of affairs in which some members cease to see their community as an enterprise from which they stand to gain. Second, policies aimed at dismantling or preventing the emergence of social norms, which would, in turn, prevent some fellow members of the community from shaping their specific terms of cooperation in ways that meaningfully respond to their dynamic structures of preferences and goals, should be devised.

These amendments, I suggest, should make us consider Hayek's proposed institutional arrangements as very appealing from a contractarian framework of justification. In fact, if we are able to design cooperative schemes which define a protected sphere that allows members of a community to continuously shape their specific terms of cooperation, according to the dynamic structures of preferences and goals, we are also able to secure compliance with a community's rules in an entirely instrumentalist fashion, thus fostering its stability.

However, one might object that there is much wishful thinking in claiming that rational individuals, with their own life plans, preferences and goals would converge on a unique set of institutions. After all, persons' specific goals and aims, within pluralistic societies, point toward widely different directions. In

light of the fact of pluralism, and consequent disagreement about institutional arrangements, the possibility to secure an agreement on a unique set of institutions would ultimately appear as chimeric. In this regard, in the fourth chapter, I will attempt to offer one main argument which could partly respond to the pervasive disagreement about political institutions, stemming from seemingly incompatible preferences and goals. The argument rests on how one's uncertainty about future aims drives the emergence of preferences for flexibility when facing one-way, long and wide-ranging choice scenarios such as decisions about institutional arrangements.³³ Particularly, I will attempt to show that people's uncertainty about future preferences and aims invites them to abstract from their current sets of goals and to come to the social contract bargaining table by considering a wider range of goals that they predict could come to be salient in the future.

The core idea behind the concept of flexibility is that we have instrumental reasons to choose institutional arrangements that would allow us to design our lives in due course, according to preferences and aims we come to possess within our entire life path. As such, institutional arrangements should not be tied to narrow sets of preferences and goals that are salient to us when the bargaining procedure takes place if, by tying our choice to such a set, we are prevented from (or we increase the costs for) pursuing goals that will come to be salient in the future. Preferences for flexibility, thus, emerge from the consideration that, in light of our own uncertainty about future goals and aims, we have reasons to keep the door for new opportunities constantly open, thus giving up on the idea of defining our specific terms of cooperation *ex ante*. Such an uncertainty, along with the desire to be able to shape our life according to goals and aims we will come to possess during our life path, invites us to choose our institutions by enlarging the set of specific goals we possess when the bargaining procedure takes place, and to converge on general sets of institutions that leave in our hands the possibility of defining our specific terms of cooperation at later stages.

³³The idea of preferences for flexibility is borrowed from Kreps (1979).

The interesting implication of preferences for flexibility, within bargaining scenarios, is that by enlarging bargainers' sets of goals, they make more room for an overlapping consensus over institutional arrangements, thus helping us in achieving coordination among agents exhibiting seemingly incompatible preferences. Flexibility, though, cannot respond to all sorts of disagreement. In fact, there may be situations in which agents' larger sets of goals, elicited by preferences for flexibility, are broadly non-compossible, such that no institutional arrangement would be satisfactory from everyone's standpoints. This persistent disagreement clarifies that any political theorising that wishes to take diversity seriously should be oriented toward a model of cooperation embodying multiple social contracts underpinning diverse sets of political institutions, rather than focusing on a unique set allegedly able to accommodate for everyone's preferences and goals.

In this regard, the fifth chapter deals exactly with models of cooperation embodying multiple communities, in an effort to find institutional arrangements, at the meta-community level, that would help in creating the conditions for the emergence of as many communities as possible, in order to respond to the challenge posed by persistent disagreement. In particular, I analyse Nozick's account of the framework for utopia, where he attempts to justify the minimal state as the meta-community institutional arrangement that best approximates the possible-worlds model. In this chapter, I put Nozick's theory under scrutiny and attempt to show that the minimal state faces two main challenges in the route to approximate the possible-worlds model: first, collective action problems stand in the way of creating new communities; second, informal sanctions, underpinned by social norms, may increase the costs for creating new cooperative enterprises or joining already existing ones. Such challenges, I hope to show, invite amendments to Nozick's theory that seems to point toward institutional arrangements, at the meta-community level, more demanding than the minimal state.

I wish to conclude this introduction by highlighting that the focus on preferences for flexibility is, in my mind, what ultimately shapes the originality of

the present thesis and sets it apart from other relevant works tackling social contract theory from the lenses of instrumental rationality. In fact, although prominent authors such as Muldoon (2016), Sugden (2018), Moehler (2018) have explored the possibility of convergence on cooperative sets of rules from a wide variety of diverse standpoints, their contributions either delineate social contract scenarios in which agents, with radically different and rather defined structure of preferences and goals, deliberate on sets of rules, in an effort to show that diversity does not compromise the possibility of an agreement on the ground of mutual advantage (Muldoon, 2016), or argue that convergence on a common set of rules is possible even if agents seek to maximise their opportunities in the light of their current and defined structures of preferences and goals (Sugden, 2018), or suggest that convergence is possible if agents exhibit a forward-looking attitude, without fully delineating the instrumental grounds for such an attitude (Moehler, 2018).

In this respect, the present work should be read as complementary to these approaches. In particular, I attempt to provide two main contributions that could potentially offer additional strength to these arguments: first, the uncertainty surrounding agents' future structures of preferences and goals fosters the possibility convergence on a common set of institutions insofar as it brings agents to consider a larger set of preferences and goals, including those they may come to possess in the future, that creates more room for an overlapping consensus; second, preferences for flexibility nicely capture, and analytically lay out, the idea of forward-looking agents, that is of agents that care about their continuous ability to shape their social environment according to preferences and goals they come to possess.

The Spontaneous Order as a Cooperative Enterprise

Introduction

In the introduction, I offered three main reasons for narrowing down our object of investigation to Hayek's account of the spontaneous order. First, I have claimed that Hayek's theory can be plausibly seen as consistent with the main constitutive aim of the present work, which is to theorise on cooperative schemes and institutional settings by taking individual preferences and goals seriously; second, I have suggested that Hayek's investigation on the spontaneous order constitutes, perhaps, the most sophisticated inquiry on cooperative institutions within complex contexts such as pluralistic societies; third, I have argued that Hayek's theorising around institutional arrangements equips us with great tools in thinking about cooperative institutions at the meta-community level.

The present chapter, in this regard, attempts to lay down a charitable description of Hayek's account of the spontaneous order, in an effort to support the first two claims. In particular, I shall attempt to show that Hayek's investigation into the spontaneous order can plausibly be read as an inquiry toward cooperation-enhancing institutional arrangements within complex social orders, in which information is widely dispersed among fellow members of our communities, and where theorists and planners are largely unable to collect the relevant knowledge in order to fruitfully respond to people's preferences and goals, both in a static and a dynamic fashion.³⁴

The key assumption surrounding Hayek's works, and one that is shared by most social scientists and political philosophers, is that securing the conditions for cooperation is the first real task of social orders. Cooperation, indeed, can be regarded as a necessary hub-enterprise, operating in the background of our

³⁴The present work only marginally deals with the reconstruction of Hayek's thought, and when it does, it confines the investigation to specific aspects that I find of philosophical interest. In this regard, bibliographical references to more general and systematic investigations into Hayek's thought, and also providing some contexts to his intellectual endeavour, are in order. Two main books, in particular, successfully embark in such an enterprise: Caldwell (2004) and Ebenstein (2003).

social interactions, which allows us to engage with most other sub-enterprises that shape our life plans and provide meaning to our endeavour on earth.

The search for cooperative institutions operating in our social orders, though, is by no means easy, as two main broad challenges stand in our way. First, social orders are not parochial associations, which members create and join in the pursuit of a shared aim. In fact, most of our ineradicable and rational disagreement about what social institutions should be like is largely driven by the fact that our life plans are diverse and often incompatible. Second, social orders, unlike parochial associations, are complex in that they are populated by many individuals exhibiting widely diverse structures of preferences and goals encompassing various aspects of their lives. Such a complexity, I will attempt to show, undermines our ability to design political institutions *as if* we could somehow collect all the relevant information required in order to design social institutions that will be found attractive from a wide variety of standpoints.

The challenge posed by disagreement, stemming from the fact of diversity in our structures of preferences and goals, is one that many philosophers and social scientists have taken seriously. Responses, I suggest, can be framed as broadly falling in two main categories. First, there is an attempt to confine the extent of our disagreement through accounts of *reasonableness*, which would guide and simplify our search for cooperative institutions. For if we rule out certain preferences and goals, and identify shared aims allegedly stemming from certain facts about human nature, our intellectual endeavour in searching for cooperative institutions is certainly made easier, as we can, in principle, start building from that in designing fully-fledged and detailed cooperative schemes and institutional arrangements.³⁵ Second, there is an attempt to theorise about cooperative social institutions by taking individuals preferences and goals as they are, and try to design simple and abstract institutional arrangements that would work as basic cooperative framework which could be allegedly endorsed by a wide variety of standpoints.³⁶

³⁵This route, I believe, characterises *contractualist* accounts such as Rawls (1971) and Scanlon (1998), but is also followed by authors such as Nagel (1986).

³⁶This, I believe, is the route that contractarian theorists such as Gauthier (1986), Buchanan

Hayek's inquiry into the spontaneous order, I suggest, should be seen as primarily contributing to this second approach. In particular, his proposal for a basic institutional arrangement protecting a set of rights highlighted by the Lockean triad - life, liberty and property - should be read as the result of an investigation toward cooperative social institutions that members of our communities, with widely diverse structures of preferences and goals, could plausibly endorse. More specifically, I shall argue, such an institutional arrangement would allegedly possess the merit of distributing certain rights whose exercise allows us to shape our social environment in ways that make it adapt to our structure of preferences and goals.

In fact, what emerges from Hayek's investigation into the spontaneous order is that the problem of cooperation within complex social orders cannot be dealt with by attempting to collect all the relevant information required to design fine-grained institutions, in an effort to respond to people's structure of preferences and goals. Such an enterprise is ruled out by the constitutive complexity of our communities and is bound to result in largely defective sets of institutions. Instead, the main task of the theorist is to define a basic framework of rules that would create the conditions for the emergence of cooperation-enhancing platforms, where the exercise of one's rights, in the widest range of circumstances, would drive the evolution of social orders.

This is, I believe, the constitutive aim of the spontaneous order, and this is, I suggest, what the idea of *spontaneity* amounts to: a feature of social orders, ultimately depending on a basic framework of rules governing them, which allows for the emergence of states of affairs, which are the product of decentralised interactions, and which meaningfully respond to people's structures of preferences and goals.

In this chapter, I shall lay out this particular interpretation of the spontaneous order, also in an effort to dispel some possible doubts linked to alterna-

(1991), Binmore (2005), Gaus (2010), Muldoon (2016) and Mohler (2018) have followed. Moore (1996) nicely captures the distinction between contractualism and contractarianism, and frames it precisely as turning on the concept of reasonableness, which is crucial in contractualist accounts and absent in contractarian theories.

tive accounts of the spontaneous order, which allegedly deem as spontaneous any order unintentionally arising from individuals' interactions. These *organic* accounts of the spontaneous order characterise spontaneity as an ineliminable feature of our orders, in that our communities are always subject to intra-group and inter-group evolutionary forces leading to the emergence of systems of rules and states of affairs which differ from our initial positions. As such they are useless in distinguishing between social orders according to their institutional arrangements, for any system of rules is, ultimately, the product of unintentional adaptation to evolutionary forces.

In this regard, organic accounts are either entirely descriptive, in that they merely aim at spelling out how evolutionary forces affect our institutional arrangements and, as a consequence, our communities, or normatively faulty, in that their predictive claim, which would allegedly consist in suggesting that evolutionary forces would necessarily drive our institutions toward cooperation-enhancing arrangements, seems largely mistaken. This latter aspect, I believe, has been convincingly tackled by authors such as De Jasay (1989), Sugden (1993) and Gaus (2006), who have shown that institutional adaptation driven by evolutionary forces does not imply our ability to solve problems connected to cooperation, nor it necessarily implies a tendency toward cooperation enhancing arrangements.

Importantly, this is not to deny that such an alternative and puzzling account of the spontaneous order actually emerges from Hayek's account. In fact, various references to Hayek's work supporting this particular interpretation can be provided, but merely to suggest that this is neither the most charitable way to frame his important work, nor the best starting point in order to further our investigation into the theory of the spontaneous order. In this regard, the reader will notice that the aim of the present work is only minimally exegetical, as the main aim is to lay down an account of the spontaneous order that is plausible, workable and of normative interest to those who wish to theorise about cooperative institutional arrangements by taking preferences and goals seriously.

The chapter starts with a comparison between two ideal communities, Dore and Bakewell. Both communities have as their main constitutive aim that of designing institutional arrangements able to secure cooperation, given members' structures of preferences and goals. What distinguishes them is their different degree of complexity, which, I hope to show, demands different *kinds* of institutional arrangements. In particular, while Dore can be meaningfully ruled by designing fine-grained systems of rules, Bakewell cannot, as its complexity calls for an entirely different approach in designing its institutional framework. Bakewell's institutional arrangement, in other words, calls for spontaneity.

Building from such a comparison, I shall attempt to illustrate the idea of the spontaneous order I have above sketched.

Dore vs Bakewell

Imagine you live in the very small community of Dore. A few hundreds hard-working people inhabiting a beautiful and rich land. They manage to live well, produce and consume everything they need, such that the town needs no interactions whatsoever with the near and enormous Bakewell. There is not much to do in Dore. Most fellow members of the community work until 6 pm, then go to the one and only (though great) restaurant, The Old Vicarage, then home. One school, one sport club and one mall. There are no significant changes in demographic dynamics, as the population has always been stable with zero social mobility as new-born members tend to undertake the same career path of their parents and seem to enjoy it.

You are unemployed, and permanently so, a very privileged position in Dore reserved for the intellectuals. In fact, you spend your entire day thinking about ways to improve the community. At some point of your intellectual endeavour, you come to realise that, in order to have a better understanding of how to improve the status quo, you could write a book which describes how is it to live in Dore, and then subject each and every practice you have described to careful scrutiny.

Surprisingly, though, the book is taking much time to write, as you manage to describe very many aspects in full details. This is because social interactions and practices going on in Dore are quite simple and, as such, you are able to provide quite an in-depth account of them.

The book is now over and Mr Forge, the mayor, who is committed to improving the conditions for cooperation, by expanding the set of opportunities of his fellow members of the community, wants to see you in order to know whether you have gained great insights on to make Dore a better place to live. You have a few policy proposals, but one big idea comes to your mind: wouldn't it be great to go to Bakewell for a long trip, check out how life is there, write it down and see if there is anything you could plausibly implement in Dore? Surely, Dore and Bakewell are quite different communities, but something can still be learned. Therefore, you tell Mr Forge about your plan and successfully apply for a grant that will generously cover for your expenses in Bakewell for two years.

Bakewell is a beautiful, large and rich city, with many restaurants of any kind, some of which you find bad, some other incredibly good. There is a lot of technological innovation going on, and also moral innovation. People in Bakewell like to experiment, their lives are adventurous in that they engage with all sorts of enterprises. For instance, they play a wide variety of sports, most of which are unknown to you, and change their job quite frequently as new opportunities always come out from changes in technological constraints. Moreover, people in Bakewell exhibit a rather dynamic structure of preferences and goals, such that their life plans always change, and so does the range of enterprises they engage with. They move up and down in the social scale quite rapidly, and continuously update their knowledge through education, no matter how old they are.

At some point, you seem to realise that if the book about Dore took six months to write, two years may not be enough to describe everything that is going on in Bakewell. Moreover, every time you manage to write a chapter, you feel like it does not really describe how things are actually going in Bakewell,

but merely how they used to go when you observed them.

Two years have passed, your grant is now over and so is your book, and, perhaps unsurprisingly, it is quite different from what you expected. First, it is shorter than the one on Dore; second, the books are different in kinds. The one on Dore is incredibly long (and boring), as you were able to describe the lives of your fellow citizens up to an incredible level of details, such that you can tell what each is doing at particular times of the day. The book on Bakewell, on the other hand, is a book about the main constitutive features of the community: its main industries, technological innovation, religious practices, sports, etc. The books are different because there was no way to account for everything that is going on in Bakewell, in the same way you did for Dore.

Dore is a simple system: only a few hundreds fellow members, with very repetitive interactions among them, little moral or technological innovations and little experimentation. In Bakewell, on the other hand, people have with very wide interests, it is the theatre of many innovations (even dumb ones) and new ways of living are tried out. In this regard, if the books are different, it is not merely a problem of time constraints: Bakewell was too complex to be investigated that way. Its description called for a higher degree of generality compared to the book on Dore.

Complexity

A system's complexity is a function of the number of its elements, of the interactions among them and of the interactions between elements and the environment.³⁷ When elements and interactions increase, so does the overall complexity of a given system. There is no fixed amount of elements or interactions to call a system 'complex', but there are some features that complex systems exhibit and which allow us to distinguish them from simple ones, chief among those is non-linearity.³⁸

³⁷For a general introduction to complexity see Mitchell (2009). See also: Ladyman, Lambert, Wiesner (2012); Cotsaftis (2007); Kauffman (1993); Gell-Mann (1988).

³⁸Ladyman, Lambert and Wiesner (2012), in particular, highlight seven specific features characterising complex systems: nonlinearity, feedback, spontaneous order, lack of central

Nonlinear systems are those in which the *superposition* principle does not apply. This is to say that, when we introduce a certain stimulus to a given system, the system is nonlinear if changes in outputs are not proportional to the stimulus we have been introducing. More generally, we refer to nonlinearity to define systems that are more than the mere sum of their parts.

To show the peculiarity of nonlinear systems let us turn, again, to Dore and Bakewell. Particularly, imagine that we introduce a certain stimulus in both systems (e.g. a certain technological innovation), how would they react? Within simple systems, such as Dore, such a stimulus shall generate predictable outcomes. For instance, we may reasonably suppose that a certain technological innovation will facilitate the pursuit of certain ends, leaving almost intact the other constitutive elements of the community. Complex systems such as Bakewell, on the other hand, could react unpredictably, possibly affecting a wide encompassing range of practices. In fact, the high number of elements and interactions among them, shape the sensitivity of complex systems to small changes introduced within them.

For instance, consider the invention of the washing machine, which Hans Rosling, in a popular Ted Talk, compellingly claimed to be one of the major technological innovation of the twentieth century.³⁹ In simple systems, such as Dore, we may suppose that its introduction would help in the pursuit of its constitutive aim, washing the laundry, and buy some spare time to those who are up to that particular task. This time can be fruitfully spent by engaging with other sorts of enterprises going on in the community of Dore. However, if these other enterprises are limited and stable, counterfactuals about what would happen once the washing machine is introduced, would come down to a few options. On the other hand, in complex societies such as Bakewell, such a breakthrough innovation would presumably give rise to wide variety of enter-

control, emergence, hierarchical organisation and numerosity. However, some of them do not seem either necessary or sufficient to establish the conditions for the complexity of a given system; some others, on the other hand, seem implied by the concept of nonlinearity (e.g. emergence).

³⁹See, Rosling (2010). Actually, the invention of the washing machine dates back to 1767. However its diffusion and widespread adoption remains a phenomenon of the twentieth century.

prises, as the spare time gained through the use of the washing machine would allow many people to engage with various existing activities and to come up with new ones.

Complexity, thus, has crucial implications with regard to how we describe, or attempt to make predictions about the evolution of complex systems. As we have seen, describing Dore was easy, although time consuming; describing Bakewell, on the other hand, was challenging in that it required the ability to abstract from the specific elements of the system and to identify the main constitutive features of the community, with the aim to provide meaningful information to the readers. This sort of abstraction, though, more than a mere author's choice, is dictated by the impossibility of accounting for all the variables in play. It is not just that we lack time to describe what is going on in Bakewell; in fact, we also lack access to all the required information.

When we are asked to describe complex systems such as Bakewell, we may attempt to act as foxes who know many little things, but we are likely better off as the hedgehog who knows one or a few big things. In fact, if our ability to collect all the relevant information is ultimately lacking, we will likely be doing a better job by giving up on the project of laying down everything that happens in Bakewell, and jump up to a higher degree of generality. For instance, if asked to describe how people spend their vacations in Bakewell, we should probably surrender on the idea of collecting information about each and every citizen's planned destinations, and account for general trends emerging from statistical researches.

Hayek was among the earliest social scientists to provide an account of how degrees of generality of our descriptions or explanations vary (or, better, should vary) according to the degree of complexity of the objects we analyse. Quite relevant to this particular point, is his distinction between fine-grained explanations, which can be aptly provided in accounting for how a simple system works, and explanations of the principle, which aptly suit our needs when dealing with complex systems.⁴⁰

⁴⁰Much of Hayek's contribution to complexity and degrees of explanation can be found in

Essentially, the claim behind the distinction consists in that complex systems do not lend themselves to detailed descriptions, but rather call for a different level of abstraction, which allows us to say something accurate about how a given system works without providing fine-grained information about its general functioning.

Governing Complexity and the 'Organic' Account of the Spontaneous Order

Despite being incredibly boring, the book on Dore sold as well as that on Bakewell. In fact, everyone was very excited to have their name on it. Mr Forge is excited as well, and wants to see you in order to ask about the sorts of policy proposals they could import from Bakewell. He knows that citizens of Dore tend to be very conservative but is quite sure that there is something you can suggest after having carefully studied the laws of Bakewell.

However, this is not the case. In fact, your stay in Bakewell taught you another great lesson: complex and simple orders call for different kinds of rules.

in Hayek (1952); (1955) and (1964). The following quote may, perhaps, help in clarifying the present argument: "The usual kind of explanation which we give, e.g., of the functioning of a clockwork, will in our sense be merely an explanation of the principle. It will merely show how the kind of phenomena which we call clockworks are produced: the manner in which a pair of hands can be made to revolve at constant speeds, etc. In the same 'general' way most of us are familiar with the principles on which a steam engine, an atomic bomb, or certain kind of simple organisms function, without therefore necessarily being able to give a sufficiently detailed explanation of any one of these objects so that we should be able to construct it or precisely to predict its behaviour [...] Even where we are able to construct one of these objects, say a clockwork, the knowledge of the principle involved will not be sufficient to predict more than certain general aspects of its operation. We should never be able, for instance, before we have built it, to predict precisely how fast it will move or precisely where its hands will be at a particular moment of time", Hayek (1952, p. 185). Hayek also pioneered the research on degrees of explanation related to the complexity of human brain: "there is an absolute limit to what the human brain can ever accomplish by way of explanation - a limit which is determined by the nature of the instrument of explanation itself, and which is particularly relevant to any attempt to explain particular mental processes. If our account of the process of explanation is correct, it would appear that any apparatus or organism which is to perform such operations must possess certain properties of the events which is to explain. If explanation involves that kind of joint classification of many elements which we have described as 'model-building', the relation between the explaining agent and the explained object must satisfy such formal relations as must exist between any apparatus of classification and the individual objects which it classifies", Hayek (1952, p. 185). A careful analysis to Hayek's contribution in the analysis of complex systems can be found in Weimer (1982); Marsh (2010); Gifford Jr. (2007); Allen, Strathern and Baldwin (2007); Gick and Gick (2001); Agonito (1975); Dempsey (1996); D'Amico and Boettke (2010); Horwitz (2008).

Particularly, while Mr Forge can design and devise a fully-detailed set of rules, on the basis of members' structure of preferences and goals, which guides them, at each and every step, in an effort to improve cooperation, the same is not true for Bakewell.

During your time in Bakewell, you managed to interview the mayor, Miss Bragazzi. She has been very open to you about their ideas of government. Bakewell is a big city, and they cannot design and devise fully detailed sets of rules on the basis of citizens' preferences and goals; especially considering that preferences and goals tend to vary quite a lot. Fully detailed sets of rules, guiding members of the community at each and every step, call for a fine-grained knowledge of the system to which they are meant to apply, but, unfortunately, Bakewell does not really resemble a simple system. In this regard, small errors in designing the detailed set of rules could make the system dramatically deviate from our predictions, possibly undermining cooperation in such a complex environment.

The difference between Dore and Bakewell lies mainly here. Mr Forge can account for almost each and every variable and is, thus, able to design fine-grained rules that tackle particular problems or aim at improving small aspects. Miss Bragazzi, on the other hand, cannot. She is bound to higher degree of generality when thinking about how to improve the status quo. She has to identify general rules that would robustly facilitate cooperation among members of the community exhibiting a wide and dynamic range of preferences and goals, rather than trying to factor in all the possible variables in play in the attempt to lay down fine-grained rules. Dore is governed as an *organisation*, whereas Bakewell is ran as a *spontaneous order*. Although both Mr Forge and Miss Bragazzi are committed to creating the conditions for fruitful cooperation for the members of their communities, their strategy changes, as Dore and Bakewell are different in kind.⁴¹

One main question arises about what sorts of rules should Miss Bragazzi

⁴¹Hayek highlights the distinction between orders and organisations in Hayek (2012) pp. 44-52.

design and devise in complex orders such as Bakewell. In particular, how is Miss Bragazzi to design the sort of institutional framework that makes Bakewell a spontaneous order? If we follow Hayek's account tightly, such a question may not even make sense. Rules governing spontaneous orders, indeed, simply emerge from complex evolutionary mechanisms, such that there seems to be no space for human design or intervention. Gerald Gaus (2006), in particular, identifies three main mechanisms that Hayek isolates as the main drivers of the evolution of rules: group survival, group growth, and an endogenous mechanism that captures individual non-random deviations from existing rules.⁴² The first two mechanisms are based on the idea of an inter-group selection of rules:

Although the existence and preservation of the order of actions of a group can be accounted for only from the rules of conduct which individuals obey, these rules of conduct have developed because the individuals have been living in groups whose structures have gradually changed. In other words, the properties of the individuals which are significant for the existence and preservation of the group, and through this also for the existence and preservation of the individuals themselves, have been shaped by the selection of those individuals from the individuals living in groups which at each stage of evolution of the group tended to act according to such rules as made the group more efficient. (Hayek, 1967, p. 72).

And, again:

These rules of conduct have thus not developed as the recognized conditions for the achievement of a known purpose, but have evolved because the groups who practised them were more successful and displaced others. They were rules which, given the kind of environment in which man lived, secured that a greater number of the groups or individuals practising them would survive. (Hayek, 2012, p. 19).

⁴²Many other commentators have carefully analysed Hayek's account of evolutionary mechanisms shaping the evolution of institutions. See, in particular, Vanberg (1986); Hodgson (1991); Sugden (1993); Whitman (1998).

There are then [in the social life of modern man] all the remains of the traditions acquired in the successive types of social structures through which he has passed - rules which he did not deliberately choose but which have spread because some practices enhanced the prosperity of certain groups and led to their expansion, perhaps less by more rapid procreation than by the attraction of outsiders. (Hayek, 2012, p. 492).

Most of the steps in the evolution of culture were made possible by some individuals breaking some traditional rules and practising new forms of conduct not because they understood them to be better, but because the groups which acted on them prospered more than others and grew. (Hayek, 2012, p. 493).

According to these two mechanisms, a process of imitation of more successful groups shapes the evolution of rules. If, for instance, Peter and Kate's group, under the set of norms k , managed to survive or to achieve better outcomes than Alf and Betty's group did, under set z , the process of imitation will lead Alf and Betty to shift towards k . Such a shift may not be the result of Alf and Betty's explicit will or awareness but rather emerge from a partially unintended process of imitation.

The endogenous mechanism, on the other hand, works at the intra-group level and is shaped by competition between individuals:

Now there is reason to think that Hayek's account of social evolution is similar to this more complex picture. Hayek writes that a person's "thinking and acting are governed by rules which have by a process of selection been evolved in the society in which he lives." Understood thus, it looks now as if his project is to explain how each rule (not the order of actions itself) evolved within the society. That this project may rely not only to the rule's ability to produce a competitive order of actions, but the rule attractiveness to individuals, is suggested by Hayek's remark that "[t]he competition on

which the process of selection rests must be understood in the widest sense. It involves competition between organized and unorganized groups no less than competition between individuals.” This stress on individual competition and the evolution of rules suggests that, instead of a competition between social orders, Hayek has in mind a competition between individuals within a social order that leads to the selection and evolution of rules. (Gaus, 2006, p. 244).

Specifically, Gaus has in mind an intra-group mechanism that is ultimately shaped by nonrandom deviations from the existing set of rules. According to it, rules that do not entirely satisfy Alf or Betty are more likely to be violated and, thus, replaced with more efficient rules. For instance, if Betty expects to be better off by following *b* than by complying with *a*, she will likely deviate; and if such a deviation would also make Alf better off, then Alf would imitate Betty and *b* will likely replace *a*.

Hayek’s *organic* account,⁴³ though, seems to depict a ‘might is right’ scenario in which the evolution of rules toward cooperation enhancing directions is spontaneously driven by evolutionary forces, and, therefore, where we are left with very little to say about the rules governing the complex orders we inhabit. Attempts to change our sets of rules may, indeed, be easily regarded as naive if we attribute such an inherited wisdom to our current set of rules, and posit such a faith in evolutionary forces guiding their transformation. To put it bluntly, according to the organic account, any order would be spontaneous as long as

⁴³The term organic is inherited from Carl Menger, who first provided an account of how evolutionary mechanisms may drive institutions toward cooperation enhancing directions. In 1883, Menger offered a pioneering distinction between organic and pragmatic institutions. Pragmatic identifies institutions which are the result of the explicit will and design of individuals, and organic refers to institutions which arise spontaneously and unintentionally through decentralised individuals’ intentional actions. For instance, in his discussion on the origins of money, Menger (1963) [1883], Book 3, Ch. 2) points out that the decline of barter can be attributed to the difficulties in finding the ‘exact match’ within direct exchanges, and to the need of finding more marketable and divisible goods in order to carry out exchanges. According to Menger, thus, the emergence of money is the unintended result of individuals aiming to solve their problems in carrying out their exchanges rather than the result of centralised human design. Menger’s account of institutions constitutes a refinement of previous accounts developed by the Scottish moral philosophers of the eighteenth century such as Hume (1740), Mandelville (1705), Ferguson (1767) and Smith (1776). For a careful analysis of the history of the theory of the spontaneous order, see Cubeddu (1993).

we do not interfere with the evolution of its rules, as any order is subject to evolutionary forces driving the transformation of its set of rules, shaping the conducts of its members.

Both mechanisms, however, fall prey to well known collective action problems. In fact, the process of imitation, which allegedly drives us to follow sets of rules that led other groups to outperform ours, is by no means justifiable on instrumentalist grounds. Some sets of rules, indeed, may be optimal for the group and yet detrimental to us, individually. When this happens, sets of rules that would be beneficial to nearly everyone may not be adopted precisely because of the temptation to free ride on other people's efforts. On a similar fashion, non-random deviations from a certain rule may simply constitutes Alf and Betty's best response to one another possible strategies whilst being overall detrimental to both of them. In other words, evolutionary mechanisms might make us stuck within prisoner's dilemma scenarios, thus making us unable to eventually converge on sets of institutions underpinning better terms of cooperation.

In order to further illustrate how Hayek's organic account of the spontaneous order falls prey to collective action problems, consider the following example offered by Cristina Bicchieri:

Until not long ago, a Sicilian man who "dishonored" another man's daughter or sister had to make amends for the wrong by marrying the woman or pay for his rashness with his own life. The objective was to restore the family's lost honor, but the social norms dictating the ways in which this could be done were the only means available to identify honor in those circumstances. One may think that some form of monetary compensation would have worked equally well, if not better, in the case in which a marriage was impossible. It would have spared one, perhaps many, lives. But accepting a monetary compensation was not revenge, and since nobody would have ever accepted such an atonement, nobody would have ever thought of offering it. Approving of the man who exacts revenge, calling him a

”man of honor”, does not necessarily involve approval of the norm as rational or efficient. Even if one thinks a norm unjust and useless, it may be difficult not to conform, since violation involves a collective action problem: nobody wants to be the first to risk social disapproval by breaking the norm openly. (Bicchieri, 1990, pp. 838-39).

In this example, although norms of revenge are perceived as irrational or inefficient, they still persist insofar as non-compliance bears the risk of social disapproval for each member involved. Although shifting towards a different set of norms, thus, abandoning norms of revenge, would provide a better framework of cooperation, such norms still persist because complying with them constitutes each member’s best response to each member’s possible courses of actions.

In this regard, Anthony De Jasay notices, there is no reason to suppose that what drives the perpetuation of institutions, such as systems of rules, is other than their fittingness to survive:

A health service which healed some people and made most others dependent on doctors, hospitals, and drugs would certainly be fit to survive; it would create demand for itself and establish the pre-conditions of its own propagation. A prison system in which petty criminals became hardened and unreformable, or asylums that made the unhinged even madder, would likewise be self-perpetuating (De Jasay, 1989, p. 77).

Essentially, Hayek might have failed to recognize that the features that allow for the self-perpetuation of institutional settings may have little to do with their ability to provide the conditions for cooperation, which in turn would enhance group growth or group survival. In fact, in many cases, institutions can be self-perpetuating notwithstanding their ineffectiveness in responding to our needs as a group.

A non-organic account of the Spontaneous Order

The organic account just sketched, I suggest, is not really an account of the spontaneous order, but rather a description of how grand and spontaneous evolutionary forces silently operate in the background of our social orders. In this regard, both Dore and Bakewell, regardless of their different degrees of complexity and of the kind of rules in play, are subject to them. For organisations such as Dore too would be shaped by mechanisms of imitation and competition that would drive members of the community to non-compliance with certain rules, or to different interpretations of rules devised by Mr Forge, or to shifting toward different sets of rules.

The evolutionary forces that Hayek describes, we might plausibly grant, could shape our communities toward cooperation enhancing directions, in the very long run; in the same ways in which Darwinian accounts of evolution shape species' adaptations to their environment. Hayek's organic account, in this regard, provides us with a rather analytic description of the mechanisms that allegedly drive social orders' evolution, but is inexplicably silent about how to improve the conditions for cooperation.⁴⁴

Moreover, the organic account of the spontaneous order is in sharp contrast with many suggestions Hayek advances on how rules of cooperation, in complex systems such as Bakewell, could be designed and devised. In fact, Hayek consistently makes a number of claims concerning basic institutional arrangements that would allow us to peacefully and fruitfully interact with one another. Consider the following quotes:

The main features of all somewhat more advanced legal orders are sufficiently similar to appear as mere elaborations of what David Hume called the "three fundamental laws of nature, that of the stability of possessions, of transference by consent, and of the performance of promises. (Hayek, 1960, p. 158).

⁴⁴I follow Gaus (2006) in interpreting Hayek's description of evolutionary forces shaping our institutions as distinct from Hayek's account of the spontaneous order.

Property, in the wide sense in which it is used to include not only material things, but (as John Locke defined it) the 'life, liberty and estates' of every individual, is the only solution men have yet discovered to the problem of reconciling individual freedom with the absence of conflict. Life, liberty and property are an inseparable trinity. There can be no law in the sense of universal rules of conduct which does not determine boundaries of the domains of freedom by laying down rules that enable each to ascertain where he is free to act. (Hayek, 2012, p. 102).

Here, Hayek clearly states his endorsement for institutional arrangements that design a protected sphere for members of our social orders through a system of rights that protects 'property in the wide sense'. Such an endorsement, though, is not merely a matter of preferences, as institutional arrangements underpinned by the Lockean triad, are claimed to be the 'only' solution to secure peaceful cooperation. In this regard, one might attempt to reconcile the organic view and Hayek's position by claiming that reasons for endorsing institutional arrangements, underpinned by Lockean triad, stem precisely from the fact that such an institutional arrangement is the product of evolutionary forces which drive us toward cooperation-enhancing directions. So formulated, the Lockean triad would be regarded as a plausible candidate institutional arrangement because it resisted to evolutionary forces which allegedly shape our system of rules toward cooperation-enhancing directions.

Such an attempt, though, would mistakenly characterise Hayek's position, as he provides an *apriori* argument in favour of such institutional arrangements. An argument that does not rest on backward-looking explanations concerning the inherited wisdom of systems of rules, but rather looks analytically at how they could advance our ability to cooperate.

In order to introduce such an apriori argument, let us dig deeper into the Dore vs Bakewell analysis. Dore, we have assumed, is a simple system. There, the mayor is able to devise rules of conduct that enhance cooperation, by col-

lecting the relevant information about members of the community's preferences and goals. Such rules may plausibly take the form of fine-grained commands directed toward specific members of the community, such as do x under circumstances y . For instance, we may suppose that Mr Forge, after collecting all the relevant preferences of some members of the community, comes to realise about their desire to socialise during dinners, and, as such, issues a fine-grained rule which allows these members to coordinate on a fixed time to meet at the Old Vicarage. Since the mayor knows all the relevant circumstances, each relevant member is happy about the rule, as it would cut transaction costs implied by decentralised coordination.

Fine-grained rules, though, cannot meaningfully be implemented in Bakewell as Miss Bragazzi is largely unable to collect all the relevant information (preferences toward restaurants, working hours, etc). Moreover, citizens of Bakewell exhibit a wide range of incompatible circumstances, preferences and goals, such that coordination on such small detailed enterprises would perhaps be impossible. For instance, we may suppose that there is no single menu able to accommodate for the wide variety of preferences exhibited by citizens of Bakewell, or that given the flexibility of working hours, there is no way to coordinate on a time of the day that would suit everyone's needs, or that many members of Dore do not really exhibit such a pronounced preference for socialisation that would convince them to dine together.

Cooperation, in Bakewell, works at a higher degree of generality and fine-grained rules aiming at specific purposes would not to the job. In Bakewell, Hayek would plausibly suggest, we must rely on a basic framework of rules that allows for coordination in a context where knowledge is widely dispersed among members of the community, and where there cannot be any planner able to collect it and cut transaction costs by issuing fine-grained rules that would enhance our ability to cooperate in each and every enterprise we wish to engage with.

Hayek aptly frames the distinction between rules that meaningfully apply to simple and complex orders as respectively *thesis* and *nomos*. Thesis defines fine-

grained rules aiming at specific purposes, whereas *nomos* identifies general and abstract rule aiming at general purposes. Simple orders governed through fine-grained rules take the form of organisations (or *taxis*), whereas complex orders governed by general and abstract rules are spontaneous orders (or *kosmos*). Thesis and *nomos* are not interchangeable, nor they can be mixed. In fact, simple orders, such as Dore, can be meaningfully ruled as organisations, as Mr Forge is able to collect all the relevant information required in order to issue fine-grained rules aimed at improving specific instances of cooperation.⁴⁵ Complex orders such as Bakewell, on the other hand, call for *nomos* as we are unable to factor in all the relevant variables in order to devise rules that would achieve specific purposes:

We shall see that it is impossible, not only to replace the spontaneous order by organization and at the same time to utilize as much of the dispersed knowledge of all its members as possible, but also to improve or correct this order by interfering in it by direct commands. Such a combination of spontaneous order and organization it can never be rational to adopt. While it is sensible to supplement the commands determining an organization by subsidiary rules, and to use organizations as elements of a spontaneous order, it can never be advantageous to supplement the rules governing a spontaneous order by isolated and subsidiary commands concerning those activities where the actions are guided by the general rules of conduct. This is the gist of the argument against 'interventionism' or 'intervention' in the market order. The reason why such isolated commands requiring specific actions by members of the spontaneous order can never improve but must disrupt that order is that they will refer to a part of a system of interdependent actions determined by information and guided by purposes known only to the several acting

⁴⁵Simple orders can be also ruled as spontaneous orders, but transaction costs implied by decentralised coordination would suggest that cooperation is better enhanced by fine-grained rules.

persons but not to the directing authority. (Hayek, 2012, p. 49).

In this regard, Hayek notices, any attempt to issue fine-grained rules aimed at improving specific terms of cooperation, within complex orders, would likely fail in its aim and alter the overall order of actions that makes cooperation possible. Nomos, on the other hand, has as its main task that of creating the conditions for cooperation by defining a system of general rules that is able to use knowledge widely dispersed among individuals. Knowledge that a planner would not be able to collect.

Perhaps, the best example to illustrate this particular aspect is provided by the market. In Dore, for instance, resources can be aptly distributed without relying on property rights. Everyone, we might assume, owns everything and Mr Forge is entitled to make distributive decisions. In fact, he knows how much members of the community need, consume and produce, and is able to guide distribution of resources through fine-grained rules. In Bakewell, on the other hand, this is not the case. Miss Bragazzi is unable to collect all the relevant knowledge, and even if she could collect it, it would be extremely hard to design and devise rules that aptly respond to such knowledge. This particular inability gives rise to a need for a system of general rules that helps us coordinating in contexts where transaction costs cannot be eliminated in a centralised fashion.

Property rights, in particular, are up to this task. For they allow us, through the price system, to convey relevant information concerning individual preferences, reserve prices, people's needs, etc. Such information guides agents' decisions about what to produce and how much, without resorting to omniscient planners. For instance, when I buy an iPad for 800£, I communicate, through the price system, that I like the product and that the value of having an iPad exceeds, given my use case, the 800£ I have spent. Aggregate information concerning people's preferences and reserve prices, conveyed through the price system, shape Apple's decisions about how many iPads should be produced, but also provide crucial guidance to other market players about what sorts of

products would encounter people's needs and preferences.⁴⁶

What property rights do, I suggest, is to activate the *filtering device* of consent. In fact, by allowing us to consent to particular transactions, they allow us to shape agents' economic decisions on the basis of information we convey through the price system. In fact, my 800£ spent on the iPad tell other agents something about my utility function and shape their decisions about what to produce and how much, thus helping in achieving coordination. Moreover, the filtering device of consent allows us to achieve dynamic coordination, that is coordination in contexts where structures of preferences are largely subject to change. In fact, economic decisions, through the price system, continuously convey information about our structure of preferences, thus making economic agents continuously adapt to changes. Hayek nicely captures this aspect in *The Use of Knowledge in Society*:

If we can agree that the economic problem of society is mainly one of rapid adaptation to changes in the particular circumstances of time and place, it would seem to follow that the ultimate decisions must be left to the people who are familiar with these circumstances, who know directly of the relevant changes and of the resources immediately available to meet them. We cannot expect that this problem will be solved by first communicating all this knowledge to a central board which, after integrating all knowledge, issues its orders. We must solve it by some form of decentralization. But this answers only part of our problem. We need decentralization because only thus can we insure that the knowledge of the particular circumstances of time and place will be promptly used. But the "man on the spot" cannot decide solely on the basis of his limited but intimate knowledge of the facts of his immediate surroundings. There still remains the

⁴⁶See Hayek (2012): "In the market order each is made by the visible gain to himself to serve needs which to him are invisible, and in order to do so to avail himself of to him unknown particular circumstances which put him in the position to satisfy these needs at a small cost as possible in terms of other things which it is possible to produce instead", p. 276. Hayek highlights the crucial coordinating role played by the price systems in various other works, especially (1952) and (1935).

problem of communicating to him such further information as he needs to fit his decisions into the whole pattern of changes of the larger economic system. (Hayek, 1945, p. 524-525).

Fundamentally, in a system in which the knowledge of the relevant facts is dispersed among many people, prices can act to coordinate the separate actions of different people in the same way as subjective values help the individual to coordinate the parts of his plan. [...] The most significant fact about this system is the economy with which it operates, or how little the individual participants need to know in order to be able to take the right action. (Hayek, 1945, pp. 526-527).

Crucially, though, general rules able to create the conditions for the emergence of a spontaneous order such as the market, do not require complex evolutionary mechanisms in order to emerge; for our ability to provide general explanations (or explanations of the principle) of how complex systems work, also transfers to our ability to design and devise general rules that would create the conditions for spontaneous orders to emerge. Nomos, thus, does not need to arise organically but may also be the product of human design. The complexity of a system, indeed, does not prevent us from understanding its general functioning, nor to grasp and devise general rules that would improve its efficiency. What is prevented is our ability to form a detailed understanding of how the complex system works (e.g. to account for each and every element, for their specific interactions, etc.), and to design and devise fine-grained rules aimed at changing specific interactions of its elements.⁴⁷

For instance, we are able to assess the merits of property rights in enhancing our ability to cooperate, but we are unable to make predictions about specific distributions brought about by the market order, because such distributions

⁴⁷In this regard, we should be reading Hayek's critique of legislation not as aimed at dismissing any attempt to interfere with the overall order of actions organically arising from evolutionary forces, but rather as a critique toward attempts to achieve specific purposes through the design of fine-grained rules. For further investigation on Hayek's critique to legislation, see Posner (2003); Baumgarth (1978) and Kuznicki (2011).

arise as the unintended result of countless transactions carried out by agents who act on the basis of their own local knowledge. In this regard, general rules aimed at defining property rights, in ways that would create the conditions for the emergence of a market order, can be plausibly designed and devised, whereas fine-grained rules aimed at bringing about specific distributions cannot, as they go beyond our intellectual ability to factor in all the relevant variables in play.

Therefore, the attribute of spontaneous cannot rest on how *nomos* emerges, as rules defining property rights may indifferently arise from complex evolutionary mechanisms or be the result of a human design. The attribute of spontaneous, I suggest, refers to a particular feature of complex orders in their actual operations. Particularly, it concerns orders' ability to be continuously shaped by decentralised actions of their members, who act on the basis of their local knowledge. In this regard, the market is an instance of spontaneous orders precisely because, through the filtering device of consent, we are able to shape its particular elements (i. e. economic actors' decisions) in ways that make them adapt to our structure of preferences and goals, and without resorting to an ideal central planner to which we would have to convey all the relevant information and on the basis of which distributive decisions would be taken.

The idea of spontaneity I have just highlighted can be also exported to domains which do not strictly concern economic actions. In fact, Hayek's account of the spontaneous order is supposed to tell us something interesting about fruitful cooperation besides the market order.

In particular, institutional arrangements underpinned by the Lockean triad, which Hayek envisages as the basic institutional setting of the spontaneous order, have to secure the largest protected sphere for individuals to act in the pursuit of their ends, on the basis of their local knowledge. The rationale for securing such a protected sphere lies precisely in that only specific individuals know the particular circumstances, preferences and goals shaping their decisions and, as such, decentralised decision-making would equip them with the ability to shape their terms of cooperation in ways that would enhance the chances to satisfy their own ends.

Perhaps, rights of expressions constitute the best example in order to clarify this particular aspect. A right to free speech, for instance, defines a protected sphere for *performative utterances*. When I promise to my supervisor to write a chapter by the end of the week, I do so on the grounds of local knowledge which shapes my beliefs about the time required to complete the chapter. Such an utterance, though, is also performative in that it intentionally *creates* a particular situation: if my supervisors can rely on my promise, she will be able to organise her schedule accordingly. Through performative utterances, based on our local knowledge of specific circumstances, thus, we shape our terms of cooperation in ways that allegedly respond to our preferences and goals.⁴⁸ Examples concerning the right to free speech, though, may simply concern non-performative normative utterances. For instance, when my supervisor, on the basis of her own local knowledge, emails me about how bad the chapter is, her normative utterance is aimed at making me redraft it by following guidelines that would meet the standard she has in mind. Even descriptive utterances, I suggest, could play a similar role, although indirectly. For instance, when I speak to Italian colleagues about how is it to live in the UK academic environment, I convey information, based on my local knowledge, which shape their beliefs about particular features and that could, in turn, shape their future careers' decisions.

Rights of free movement, perhaps, constitute another pregnant example. In fact, choices about where to live are normally shaped by our own local knowledge of the relevant circumstances (other than economic ones). For instance, one's choice to move to the UK can be motivated by a variety of reasons (e.g. one's professional aspirations, sentimental engagement, etc.), whose relevance is made explicit only to those who actually possess them. As such, defining a protected sphere which allows individuals to filter the relevant options, would enhance their ability to shape their social environment according to their structure of preferences.

⁴⁸Austin (1962) provides the first analytic account of performative utterances clarifying the ways in speech may generate changes in our social world.

Attempts to reduce such a protected sphere or to interfere with it through fine-grained rules would, in Hayek's mind, undermine our ability to create the conditions for cooperation. In fact, by limiting the ways in which individuals are able to act on the basis of their local knowledge, we also limit their ability to shape their terms of cooperations in ways they would find convenient; and by issuing fine-grained rules aimed at producing specific distributive outcomes, we alter the overall order of actions that makes cooperation possible.

Certainty and Spontaneity

Hayek's arguments against this thesis are not exhausted by his considerations around complexity and the impossibility of achieving specific purposes through central planning, but also concern the idea of *certainty*. In this regard, we may say, without overstating its importance, that certainty is at the very core of the enterprise of the spontaneous order. In fact, it is not by accident that the definition of order Hayek provides is ultimately parasitic on that of certainty:

[Order is] A state of affairs in which a multiplicity of elements of various kinds are so related to each other that we may learn from our acquaintance with some spatial or temporal part of the whole to form correct expectations concerning the rest, or at least expectations which have a good chance of proving correct. (Hayek, 2012, p. 35)

Hayek's emphasis on the importance of forming correct expectations about what surrounds us suggests that spontaneous orders must secure a basic threshold of certainty which allows individuals to meaningfully interact with one another. For if such a threshold is not met, our ability to pursue our own goals is severely undermined. We may think, for instance, about the role that certainty plays in fulfilling daily tasks such as buying food, watching a soccer game or going out for a date. All these enterprises require not only joint efforts, but also the expectations that such efforts will be performed (e.g. I would not go out to buy food if I do not expect the grocery store to be open; I would not turn on

the TV if I do not expect Juventus to be playing, and I would not go out on a date if I do not expect my date to show up).

Certainty related to such detailed tasks, though, cannot be secured through institutional arrangements. In fact, rules aimed at improving our terms of cooperation in such specific instances would require fine-grained interventions, which are precluded in the phase of institutional design. Institutional arrangements, then, are supposed to create the conditions for the emergence of certainty, through the design of general rules that would limit our sphere of action.

Here, one may observe, lies a rather crucial trade-off between maximising certainty, through general rules, and the ability of an order to be shaped by the decentralised actions of individuals who act on the basis of their own local knowledge. In fact, general rules aimed at maximising certainty would presumably be oriented toward restricting the boundaries of one's actions in the pursuit of one's goals, for such restrictions would plausibly enhance our ability to form correct expectations on what is going on around us. For instance, property rights pose constraints on what we may legitimately do with certain objects, thus increasing our certainty about what other members of the community may legitimately do; similarly, rules imposing opening times to local shops would allow us to form clear cut expectations on the basis of which we set up our schedule. Reducing the protected sphere of individual action, indeed, would dramatically reduce counterfactuals about what we can legitimately expect, improving our acquaintance with the overall order of actions.

In this regard, if one takes certainty as a feature of our orders which has to be maximised, our orders would dramatically depart from the definition of spontaneity I have laid out. For maximising certainty requires rules reducing as much as possible individuals' range of actions, which, in turn, undermines our ability to shape our social environment in ways that make it adapt to our structure of preferences and goals. Hayek, thus, cannot be committed to the claim that the main task of spontaneous orders is primarily that of maximising certainty, for spontaneous orders and orders where certainty is maximised are very different, and perhaps opposite, things.

In this regard, I suggest, Hayek's stress on the relevance of certainty is meant to pinpoint that his proposed institutional arrangement gives rise to an overall order of actions in which individuals are able to form broadly correct expectations on what is going on around them, while still allowing for adaptation through decentralised actions of individuals who act in the pursuit of their own goals on the basis of their own knowledge. Hence, the institutional arrangements underpinned by Lockean triad must be seen as pursuing two different aims: on the one hand, it has to define the boundaries of our protected sphere of action in order to secure a basic threshold of certainty, which would allow us to form meaningfully correct expectations on what surrounds us; on the other, it has to create the conditions for spontaneity by defining the largest protected sphere for individuals to pursue their own goals on the basis of their own knowledge.

Furthermore, Hayek's emphasis on certainty should be seen as instrumental to his critique to fine-grained rules interfering with the spontaneous order of actions. For if institutional arrangements underpinned by the Lockean triad allow us to form broadly correct expectations on many instances of the order we inhabit, fine-grained rules, by altering the overall order, can lead to states of affairs in which our expectations are systematically disappointed.

The discussion on certainty allows us to respond to one main critique advanced by Joseph Raz against Hayek's reliance on the requirements of generality and abstractness of the law as a way to secure the conditions for the emergence and the preservation of a spontaneous order. In particular, Raz's critique is centered on Hayek's emphasis on the relevance of the rule of law expressed by the following quote:

Stripped of all technicalities this means that government in all its actions is bound by rules fixed and announced beforehand—rules which make it possible to foresee with fair certainty how the authority will use its coercive powers in given circumstances, and to plan one's individual affairs on the basis of this knowledge. (Hayek, 1944, p. 112)

In contrast, Raz argues that adherence to the requirements of generality and abstractness, which partly shape the concept of the rule of law, do not tell us anything substantive about how a particular order will look like. For general and abstract laws could, in principle, do many despicable things such as to institutionalise wrongful discrimination or to dramatically restrict individual freedom:

The law can violate people's dignity in many ways. Observing the rule of law by no means guarantees that such violations do not occur. But it is clear that deliberate disregard for the rule of law violates human dignity. It is the business of law to guide human action by affecting people's options. The law may, for example, institute slavery without violating the rule of law. But deliberate violation of the rule of law violates human dignity. (Raz, 1979, p. 221).

Raz's argument is certainly correct, but substantially misunderstands the extent to which Hayek relies on the adherence to the rule of law in laying out his account of the spontaneous order. In fact, Hayek's proposed requirements of generality and abstractness are merely meant to highlight two main things: first, that we are ultimately unable to design and devise fine-grained rules aimed at improving our specific terms of cooperation, in ways that would meaningfully respond to individuals' structures of preferences and goals; second, that if we attempt to design and devise fine-grained rules we may alter our ability to form correct expectations on what surrounds us.

Hence, Hayek is certainly not committed to claim that the requirements of generality and abstractness of rules are to be regarded as sufficient in order to secure the conditions for the emergence of a spontaneous order.⁴⁹

In this regard, it may be useful to pinpoint how Hayek's arguments differently apply to distinct aspects. First, the knowledge problem is supposed to highlight our inability to design fine-grained rules that would meaningfully improve our

⁴⁹A critique similar to Raz's has been developed by Gray (1981). For further investigation on Hayek's account of the rule of law, see Baumgarth (1978).

terms of cooperation by responding to our particular circumstances. Thus, any meaningful attempt to design cooperative rules must possess the requirements of abstractness and generality.

Second, the requirements of generality and abstractness are also important in securing a basic threshold of certainty, for fine-grained rules would unpredictably interfere with the overall order of actions and would yield the risk of disappointing our expectations.

Third, the knowledge problem demands us to design general and abstract rules which defines the largest protected sphere under which individuals can act in pursuit of their own ends, on the basis of their own knowledge, in an effort to create the conditions for the emergence of terms of cooperation that would meaningfully respond to their structure of preferences and goals.

Hayek's Utilitarianism

One question that is often debated among commentators of Hayek's work, and which is particularly relevant in assessing the normative standing of the theory of the spontaneous order, concerns its standard of justification. In particular, one may wonder whether Hayek's endorsement of the institutional arrangements underpinned by the Lockean triad is ultimately driven by the utilitarian aim of maximising the overall aggregate wealth.⁵⁰ In fact, Hayek's emphasis on the need for a protected sphere that would allegedly allow individuals to pursue their own ends on the basis of their own local knowledge seems to imply a preference-satisfaction utilitarian account.

Hayek's position, in this particular regard, is puzzling as his work is pervaded by harsh criticisms toward the utilitarian tradition:

The trouble with the whole utilitarian approach is that, as a theory

⁵⁰Some commentators suggest that Hayek is quite incoherent with respect to the main constitutive aim of the spontaneous order. In particular Gray (1982), Lukes (1997, p. 74), and Johnston (1997, p. 613), suggest that Hayek sees as the main scope of social institutions that of maximising aggregate preferences in a utilitarian fashion. Sugden (1986) and Andrew Lister (2013), on the other hand, read Hayek's approach as ultimately concerned with how institutions should distribute opportunities to members of our communities, in an effort to foster our ability to cooperate from widely different standpoints.

professing to account for a phenomenon which consists of a body of rules, it completely eliminates the factor which makes rules necessary, namely our ignorance. It has indeed always amazed me how serious and intelligent men, as the utilitarians undoubtedly were, could have failed to take seriously this crucial fact of our necessary ignorance of most of the particular facts, and could have proposed a theory which presupposes a knowledge of the particular effects of our individual actions when in fact the whole existence of the phenomenon they set out to explain, namely of a system of rules of conduct, was due to the impossibility of such knowledge. It would seem that they never grasped the significance of rules as an adaptation to this inescapable ignorance of most of the particular circumstances which determine the effects of our actions, and thus disregarded the whole rationale of the phenomenon of rule-guided action.[...] Man has developed rules of conduct not because he knows but because he does not know what all the consequences of a particular action will be. And the most characteristic feature of morals and law as we know them is therefore that they consist of rules to be obeyed irrespective of the known effects of the particular action. How we should wish men to behave who were omniscient and could foresee all the consequences of their actions is without interest to us. Indeed there would be no need for rules if men knew everything-and strict act-utilitarianism of course must lead to the rejection of all rules. (Hayek, 2012, p. 187).

Hayek's critique toward utilitarians rests, once again, on the knowledge problem. In particular, Hayek would claim, against act utilitarians, that we are ultimately unable to perform calculations about which course of actions would bring about the state of affairs in which aggregate utility is maximised, and, against rule utilitarians, that individuals are unable to perform calculations in order to devise a fine-grained system of rules aimed at bringing about the state

of affairs in which aggregate utility is maximised.

Hayek's arguments against utilitarian calculations are certainly pertinent but mistakenly assume that utilitarian standards of rightness and decision procedures, aimed at maximising utility, are the very same thing. Such a distinction can be aptly formulated in terms of questions to which these two different enterprises respond. In particular, standards of rightness are supposed to answer to 'what makes an action right?', whereas decision procedures answer to 'how am I to deliberate about which course of action I should follow?'.⁵¹ Hayek's critique, in this regard, would meaningfully apply to utilitarian theories purporting implausible calculations about which course of action would bring about states of affairs in which utility is maximised, or to rule-utilitarian theories which transfer calculations to which set of rules we should devise, but utilitarian theories are not committed to decision procedures in which such calculations are to be performed. In this regard, one may find morally compelling a utilitarian standard of rightness without necessarily being committed to implausible calculations in maximising aggregate utility.

Hayek's position, I suggest, can be aptly framed as utilitarian, in terms of standards of rightness, and as a simil-contractarian in terms of the decision procedure which leads to the choice of institutional arrangements that would maximise utility. Such an unconventional position clearly emerges from the following passage:

We may of course aim at the 'greatest happiness of the greatest number' if we do not delude ourselves that we can determine the sum of this happiness by some calculation, or that there is a known aggregate of results at any one time. What the rules, and the order they serve, can do is no more than to increase the opportunities for unknown people. If we do the best we can to increase the opportunities for any unknown person picked at random, we will achieve the most we can, but certainly not because we have any idea of the

⁵¹Distinctions between utilitarian standards of rightness and decision procedures can be found in Brink (1986), Railton (1984), Stark (1997).

sum of utility of pleasure which we have produced. (Hayek, 2012, p. 190).

Here, Hayek clearly states that the constitutive aim of the spontaneous order is that of increasing opportunities for unknown people, through the design of general rules that would allow them to pursue their own goals on the basis of their own knowledge. Such an institutional arrangement may possibly bring about a state of affairs in which aggregate utility is maximised, but if it will, it would not be on the basis of an implausible calculation that should bring us to devise certain sets of rules, as such calculation is precluded to the human mind.

This peculiar constitutive aim emerging from the decision procedure Hayek has in mind, and which leads to design institutional arrangements underpinned by the Lockean triad, suggests that Hayek's theory of the spontaneous order could plausibly lend itself to contractarian standards of justification. In fact, by ultimately leaving in the hands of individuals the opportunity to shape their specific terms of cooperation, spontaneous orders could be plausibly endorsed from a wide variety of standpoints. Moreover, by empowering individuals in shaping their terms of cooperation according to their structure of preferences and goals, we allegedly secure that they continuously have reasons to cooperate, as the spontaneous order would adapt and respond to their structure of preferences.

These two aspects, I believe, shape the attractiveness of the theory of the spontaneous order, as it becomes appealing from various standpoints. In fact, individuals exhibiting widely diverse goals would be able to shape their specific instances of cooperation, through decentralised actions, thus shaping their social environment in ways that would make their communities as cooperative enterprises from which they stand to gain.

Conclusion

Concluding, let us summarise the main points we have tackled in this chapter. First, through the Dore vs Bakwell analysis, we have laid out Hayek's knowledge

problem which planners face within complex orders such as Bakewell. In particular, we have illustrated how the complexity of orders prevents us from forming a fine-grained understanding of their specific elements and, in turn, to design and devise fine-grained rules which would meaningfully improve our terms of cooperation within specific circumstances. Complex orders, in this regard, call for a higher degree of generality in designing institutional arrangements that would enhance our opportunities to cooperate.

Second, we have shown that, in Hayek's mind, institutional arrangements underpinned by the Lockean triad meaningfully serve our purpose to improve our terms of cooperation. In fact, they are able to secure a basic threshold of certainty, while defining a protected sphere which allows individuals to act in pursuit of their ends, on the basis of their local knowledge. Such a protected sphere enhances our terms of cooperation because individuals are able to shape, through their decentralised actions, specific instances of cooperation.

Third, we have seen that the definition of spontaneity cannot meaningfully rest on how institutional settings of a given order have arisen, for arrangements creating the conditions for the emergence of a spontaneous order could indifferently arise through complex evolutionary mechanisms or as the result of human design. In this regard, the definition of spontaneity can be aptly framed as a feature which orders possess in their actual operations, and which consists in an order's ability to be continuously shaped by the decentralised actions of its members.

Fourth, we have illustrated that, although the theory of the spontaneous order could be plausibly regarded as ultimately justified in a utilitarian standard of rightness, its decision procedure points to a contractarian justificatory framework. In fact, the constitutive aim of the spontaneous order is that of increasing opportunities for unknown members of the order, through designing and devising institutional arrangements that ultimately leave in their hands the power to shape their specific terms of cooperation.

These four aspects are meant to set the stage for the critiques I shall advance in the next two chapters, where I will attempt to show that institutional

arrangements underpinned by the Lockean triad are ultimately insufficient in making sure that all members of a spontaneous order are continuously able to shape their terms of cooperation, in ways that would meaningfully respond to their structures of preferences and goals. This particular problem may lead to states of affairs in which some members of the community would cease to find their community as a cooperative enterprise from which they stand to gain, undermining the instrumentalist character of Hayek's theory.

This particular problem, I will attempt to show, may emerge as a result of two main challenges: first, large-scale consequences of small-scale transactions, allowed by the Lockean triad, may generate distributions of rights which prevent some members of the community from shaping their social environment in ways that meaningfully respond to their structure of preferences and goals; second, social norms, which increase the costs for undertaking certain courses of actions that would allow members of the community to shape our social environment, may undermine their ability to shape their terms of cooperation in ways that they would find satisfactory.

The Spontaneous Order and Its Instrumentalist Character: A Critique

Introduction

In the first chapter, I have defined spontaneity as a feature which orders possess in their actual operations, and which ultimately consists in their ability to be shaped by the decentralised actions of individuals, who act in the pursuit of their own goals, on the basis of their own local knowledge. This particular feature allows to those who take part in the overall order of actions to shape their specific terms of cooperation in ways that meaningfully respond to their particular circumstances, preferences and goals.

Hayek, we have seen, broadly identifies in the Lockean triad the institutional arrangement which creates the conditions for the emergence and preservation of a spontaneous order. In fact, the protection of property rights in the 'wide' sense is supposed to fulfill two main distinctive tasks: on the one hand, general rules aimed at defining the boundaries of our legitimate actions are supposed to secure a basic threshold of certainty, which allows members of a given order to form broadly correct expectations on what surrounds them; on the other, it secures a large protected sphere for individual to act in the pursuit of their goals, on the basis of their own knowledge. The second task, in particular, is what grants to orders their status of spontaneous. In fact, the more we enlarge the sphere of individual actions, the more we allegedly allow for orders' adaptation to the structure of preferences and goals that specific members of our community exhibit.

What emerges from Hayek's account of the spontaneous order is, I suggest, a link between the structure of rights instantiated by the institutional arrangements purported by the Lockean triad, and our ability to use knowledge dispersed among members of a given community. In fact, rights assigned by such an institutional arrangement are supposed to define a set of token acts which allegedly allow members of a community to shape their specific terms of

cooperation on the basis of their own knowledge. In fact, by exercising specific instances of our rights, through token acts permitted by such an institutional arrangements, we are able to convey information which shapes our social environment, thus defining our specific terms of cooperation.

This interpretation of theory of spontaneous order, I suggest, brings with it one main appealing aspect, which consists in its *dynamic* instrumentalist character. In particular, if individuals are able to define their specific terms of cooperation, through their decentralised actions, protected by the Lockean triad, it should follow that they possess the ability to *continuously* shape their community as an enterprise from which they stand to gain.

In this chapter, I shall present the Lockean account of the spontaneous order in greater depth and offer a critique. Such a critique aims at showing that the institutional arrangement upheld by the Lockean account defines merely an initial distribution of rights that can be unintentionally altered through fully voluntary interactions. Such alterations, in turn, may bring about derivative distributions of rights under which members of the community are prevented from shaping their specific terms of cooperation. When this is the case, they may cease to see their community as a cooperative enterprise from which they stand to gain, and form instrumental reasons to break their compliance with its rules, thus threatening its stability. As such, the spontaneous order fails in achieving its main constitutive aim, which is to secure the general conditions for cooperation.

This chapter sets out one main desideratum that plausible candidate institutional arrangements, aimed at creating the conditions for the emergence and preservation of a spontaneous order, must meet. Such a desideratum consists in defining a basic protected sphere of individual actions which cannot be altered through voluntary transactions, and which would grant to members of the community the continuous ability to shape their social environment in ways that would meaningfully respond to their structure of preferences and goals. For if such a protected sphere is not secured, some members may cease to see their community as a cooperative enterprise from which they stand to gain, thus un-

dermining its instrumentalist character and, in turn, its stability. Furthermore, this chapter paves the way for a similar critique, which will be the object of the next chapter. Such a critique will be centered on the role that *social norms* play in restricting the set of courses of actions which are available to us. In particular, by limiting our sphere of actions, social norms may prevent members of the community from shaping their specific terms of cooperation in ways that meaningfully respond to their structure of preferences and goals, thus undermining the instrumentalist character of a spontaneous order.

The chapter is organised as follows: within the first part I analyse the Lockean account, in the attempt to show, in greater depth, how the institutional arrangements Hayek envisages would serve the purpose of enhancing our decentralised ability to shape our specific terms of cooperation; within the second part I outline how such institutional arrangements might lend itself to be sustained in an instrumentalist fashion; within the third and final part I offer a critique to the Lockean account according to which the initial distribution of rights can be unintentionally altered in ways that undermine members' reasons to comply with the rules of their community.

The Lockean Account

In the first chapter, I have spelled out an organic interpretation of the spontaneous order which situates the concept of spontaneity in the origin of institutional arrangements governing our communities. In particular, we have seen that Hayek lays out a number of complex evolutionary mechanisms which allegedly guide our institutions toward cooperation-enhancing directions.

If one takes seriously the organic account, and draws its implicit conclusions, planners and theorists are left with very little to say on how to design and devise institutional arrangements that would ultimately improve our terms of cooperation. In fact, government agencies would be bound to codify and enforce systems of rules brought about by such evolutionary mechanisms.

Such an organic interpretation of Hayek's account of the spontaneous order

is, I suggested, in two main respects implausible. First, it falls prey to well known collective action problems which ultimately undermine its validity. Second, it is not consistent with Hayek's analysis of complex orders. In fact, in the previous chapter I attempted to show that the complexity of a particular system does not prevent the theorist from forming a general explanation of how it works, nor one is prevented from grasping, designing and devising systems of rules that would improve an order's functioning. What Hayek's analysis of complex systems rules out is merely our ability to form a fully detailed understanding of how a complex system works and to design and devise fine-grained rules that would improve its functioning within specific instances.

If we transfer Hayek's analysis of complex systems to social orders, we may plausibly conclude that theorists or planners are not prevented from designing and devising general rules aimed at creating the conditions for the emergence of improved terms of cooperation, by securing a protected sphere which allegedly allows for the use of knowledge dispersed among members of a given community. What is ruled out, indeed, is merely theorists or planners' ability to design and devise fine-grained rules aimed at improving our specific terms of cooperation.

Therefore, if evolutionary mechanisms select, due to collective action problems, systems of rules which are somehow inadequate to create the conditions for cooperation, theorists or planners are not prevented from designing better systems. In this regard, Hayek's endorsement of institutional arrangements underpinned by Lockean triad should be read as an attempt to theorise on systems of rules through the lenses of our ability to form a general understanding of how complex social orders work and might be improved, without necessarily relying on complex evolutionary mechanisms which allegedly grant to spontaneously emerging institutions the status of being fit to make us cooperate.

In this regard, I suggested that a plausible interpretation of Hayek's intuitions on how evolutionary mechanisms shape our institutions toward cooperation enhancing directions, would consist in confining Hayek's claim to long term evolutionary processes silently operating in the backgrounds of our social orders. As such, the fact that particular systems of rules emerged at specific times and

spaces, does not tell us anything substantive about institutions' ability to make us cooperate, as evolutionary mechanisms may, in due process, select systems of rules which make us stuck in non-cooperative scenarios.

These particular aspects, I believe, respond to two main critiques advanced against Hayek's theory of the spontaneous order, which rest on two mistaken understandings of Hayek's account. The first consists in claiming that Hayek leaves no space for institutional design, as government agencies are bound to codify institutions that have spontaneously emerged.⁵² The second, on the other hand, claims that if Hayek leaves space for institutional design, he does so by only posing the constraints of generality and abstractness of rules, but such constraints do not tell us anything substantive about the social orders we inhabit, for different general and abstract rules may give rise to very diverse orders.⁵³

To the first objection, we may aptly respond that Hayek's analysis of evolutionary mechanisms, which allegedly shape our institutions toward cooperation enhancing directions, is not meant to bind us to uncritically accept systems of rules spontaneously emerging through individual interactions and undergoing complex evolutionary mechanisms. In fact, Hayek equips us with tools to analyse complex social orders, and to design and devise systems of general and abstract rules which would create the conditions for the emergence of cooperation.⁵⁴

To the second objection, on the other hand, we may reply that Hayek's requirements of generality and abstractness of rules do not exhaust the desider-

⁵²See Hodgson (1994).

⁵³See, in particular, Raz (1976), Hamowy (1971) and Gray (1981).

⁵⁴On this particular aspect my view is similar to Gerald Gaus: "It is sometimes claimed that Hayek's thought is, at bottom, contradictory: he insists on our ignorance of social processes but, out of his analysis of ignorance, he generates prescriptions about what we should do. If we know enough to say why socialism won't work, then we must have good enough insights into the economic order to intervene to promote social goals. This, though, is wrong. As we have seen, Hayek believes that we can know quite a lot about the principles on which complex orders operate, and this theoretical knowledge allows us to say that some system states cannot be achieved and that some ways of organizing social cooperation are more efficient than others; we also know that we are unable to predict the course of, or control, the complex order itself. Thus, the analysis of complexity provides sound reasons against planning — seeking to control the emergent order. There is nothing contradictory about a mix of knowledge of principle and ignorance of what is a good plan or optimal policy". (Gaus, 2006 pp. 252-53).

ata in our search for cooperative institutions. What these requirements do, I suggest, is merely to prevent us from designing fine-grained rules aimed at improving specific instances of our cooperation, or at achieving particular distributions of rights. In fact, we have seen that such an attempt is ruled out by Hayek's considerations on complexity on two main grounds: first, fine-grained rules aimed at specific purposes are likely to fail in their aims; second, fine-grained rules would alter the overall order of actions which allows members of the community to form meaningful expectations on what surrounds them.

Generality and abstractness, thus, are needed in order to secure a basic threshold of certainty and to rule out the possibility of achieving specific purposes through fine-grained rules. However, by no means such requirements are supposed to represent sufficient conditions for a spontaneous order to emerge or to be preserved.

Many critiques advanced to Hayek's account of the spontaneous order, I argue, rest on a mistaken interpretation on the definition of spontaneity. In particular, critics tend to tie the very idea of spontaneity either to the origin of institutions, thus inferring that Hayek's account demands us to stick with spontaneously evolved institutions, or to the outcomes that such institutions would bring about, which commentators identify in the maximisation of aggregate utility, thus objecting that unspecified general and abstract rules are insufficient in securing that we succeed in our utilitarian aim.

However, I suggest, the concept of spontaneity embeds, first and foremost, a procedural desideratum in looking for institutions that would indirectly bring about on overall maximisation of aggregate utility, and which concerns the aim of designing and devising systems of rules that would allow us to overcome the knowledge problem. In particular, Hayek's search for institutional arrangements is, on the one hand, constrained by the requirement of certainty, which binds us to define institutional arrangements in a very general and abstract fashion, and, on the other hand, driven by the aim of designing systems of rules that would allow individuals to shape, through their decentralised actions in the pursuit of their own goals and on the basis of their local knowledge, their specific terms of

cooperation.

Spontaneity, thus, refers to a feature that social orders possess in their actual operations, and which consists in their ability to be continuously shaped by the decentralised actions of members of a community.

In particular, we may pinpoint the main features of the Lockean account of the spontaneous order as follows:

- (1) A spontaneous order secures a protected sphere of action, defined by the Lockean triad, which allows individuals to act in the pursuit of their own goals, on the basis of their local knowledge.
- (2) Individual actions, protected by the Lockean triad, allow individuals to shape their specific terms of cooperation in ways that meaningfully respond to their structures of preferences and goals.
- (3) The protected sphere, thus, allows for the use of knowledge dispersed among members of the community. In fact, specific terms of cooperation are not designed and devised ex-ante, but rather shaped by individual actions.
- (4) The protected sphere also secures a basic threshold of certainty which allows individuals to form broadly correct expectations on the order of actions surrounding them.

In this regard, critics such as Raz are certainly right in claiming that Hayek's emphasis on the importance of the requirements of generality and abstractness, cannot secure the conditions for the emergence of a spontaneous order, but mistaken in their assumption that these desiderata exhaust the orienting tools that Hayek provides us in searching for the institutions of the spontaneous order.

The search for such institutions, indeed, is driven by the need to define a protected sphere that would, on the one hand, allow members of a community to act in the pursuit of their own goals, on the basis of their local knowledge; and, on the other hand, to secure a basic threshold of certainty that would allow members of a community to form broadly correct expectations on what

surrounds them.

As we have seen, Hayek identifies such an institutional arrangement in the protection of the Lockean triad. However, Hayek is never particularly clear in laying out how he arrives at this conclusion, as he addresses the issue by merely claiming that distributions of rights implied by the Lockean triad would, on the one hand, 'determine boundaries of the domains of freedom by laying down rules that enable each to ascertain where he is free to act' (Hayek, 2012 p. 102), and, on the other, create the best conditions for the members of a community to pursue their own ends:

The thesis of this book is that a condition of liberty in which all are allowed to use their knowledge for their purposes, restrained only by rules of just conduct of universal application, is likely to produce for the best conditions for achieving their aims; and that such a system is likely to be achieved and maintained only if all authority, including that of the majority of the people, is limited in the exercise of coercive power by general principles to which the community has committed itself. (Hayek, 2012, p. 53).

In *The Constitution of Liberty*, Hayek briefly lays out a definition of freedom, which he spells out as 'that condition of man in which coercion of some by others is reduced as much as possible in society' (Hayek, 1960, p. 11). Such a general definition reveals two main features: first, Hayek frames the concept of freedom in purely negative terms; second, he restricts the domain of coercion to a relational concept which takes place among individuals and not between individuals and objects.

However, we would over-construe Hayek's thinking by arguing that the distribution of rights implied by the Lockean triad is arrived at by calculating which sets of rights would minimise coercion by other agents. In fact, Hayek's primary aim is not to reduce coercion *per se*, but rather to define a protected sphere of action, which would grant to individuals the best conditions for the pursuit of their own goals.

Such a protected sphere is allegedly made possible by institutional arrangements that assign to members of the community a set of claim rights to life, liberty and property. In fact, the underlying thought is that our liberty rights to act within the boundaries of the Lockean triad, alongside with the correlative duties not to interfere with our actions, would secure a large range of actions which, in turn, would create the best conditions for the pursuit of our own goals. For instance, claim rights to property should not be seen as emerging from a complex calculation which involves counterfactuals about states of affairs in which coercion by others is minimised, but rather as emerging from Hayek's considerations on how the boundaries established by property rights would allow for the emergence of cooperation-enhancing platforms such as the market, which, in turn, would grant us the possibility of shaping through our actions, our specific terms of cooperation.

In fact, I maintain that what makes these rights crucial in Hayek's account is not that their distribution minimises coercion, but rather their peculiar role in defining an unspecified set of actions through which we are able to shape our social environment in ways that meaningfully respond to our particular circumstances, preferences and goals. These include our ability to act in the pursuit of our own goals on the basis of our own knowledge, to convey information through the price system, to communicate our terms of cooperation through performative utterances, etc.

In particular, I suggest, we may aptly look at the set of rights underpinned by the Lockean triad as *social environment shaping rights*, in that they define a protected sphere for *social environment shaping token actions* which equip us with the ability to continuously renegotiate our terms of cooperation, according to our structure of preferences and goals. This peculiar way of looking at the Lockean triad, I suggest, should make us reflect on that Hayek does not merely value the fact that such rights are formally distributed and protected, but also that we have the ability to *exercise* them. For our continuous exercise of such rights is what ultimately drives the evolution of a spontaneous order and what makes us able to continuously shape our terms of cooperation in ways that

meaningfully respond to our own circumstances. In fact, if we are prevented from shaping our specific terms of cooperation through the exercise of these rights, we may ultimately cease to see our community as an enterprise from which we stand to gain and form reasons not to comply with its rules, thus undermining its stability.

The Instrumentalist Character

Hayek, thus, conceives claim rights to life, liberty and property as purely instrumental. However, we have seen that the instrumentality of rights cannot be interpreted in a strict utilitarian sense, as what drives Hayek's search for institutional arrangement is not a fine-grained complex calculation from which he supposedly arrives at a set of rules that would lead to the maximisation of aggregate wealth. Such calculations, indeed, are precluded to human planners.

Lockean rights, indeed, are selected because they 'increase opportunities for unknown people' (Hayek 2012, p. 190), by handing to members of the community the ability to shape their specific terms of cooperation, on the basis of their own particular circumstances. This particular condition that obtains through Lockean rights is, I suggest, precisely what the idea of the spontaneous order amounts to: a state of affairs in which members of the community are continuously able to define their specific terms of cooperation, through token actions protected by the set of rights underpinned by the Lockean triad.

This formulation, I suggest, has one great source of appeal: namely, by handing to members of the community the ability to shape their specific terms of cooperation, it allegedly makes community as enterprises from which all members stand to gain, lending the Lockean account of the spontaneous order to be sustained in an *instrumentalist* fashion.

In fact, by reserving us a protected sphere under which we can act in the pursuit of our own goals, and on the basis of our own local knowledge, and by allowing us to shape our specific terms of cooperation, we are allegedly able to shape our social environment in ways that meaningfully respond to our structure

of preferences and goals, thus having an incentive to continuously take part in the cooperative venture.

This source of appeal is precluded to institutional arrangements that either attempt define in fine-grained details our terms of cooperation, or which would restrict our set of social environment shaping rights. In fact, the former would fail in achieving the aim of improving our terms of cooperation in ways that meaningfully responds to our structure of preferences, as planners are unable to collect all the relevant knowledge. The latter, on the other hand, would prevent us from shaping our specific terms of cooperation on the basis of our local knowledge, thus binding us to either accept or reject a cooperative scheme.

As an instance of the former, we may imagine institutional arrangements which, for example, aim at distributing material resources on the basis of our own needs and preferences, without the need of property rights. Such an attempt would not merely fall prey to objections concerning scarcity of resources, but would also fail in its constitutive aim, due to planners' inability to correct all the relevant knowledge, even under relaxed assumptions concerning scarcity.

As an instance of the latter, on the other hand, we may imagine institutional arrangements that, for example, restrict freedom movement, or that pose severe limits to freedom of speech, or that make markets for certain goods illegal. Such restrictions, indeed, would limit members of the community in the pursuit of their own goals, on the basis of their own knowledge, and would also prevent them from shaping their social environment in ways that meaningfully respond to their structure of preferences.

Moreover, if we take at face-value Hayek's claim according to which such a distribution of rights would likely bring about states of affairs in which aggregate utility would be maximised, the appeal of the theory would double. In fact, institutional arrangements underpinned by the Lockean triad would, on the one hand, allegedly equip each and every member with reasons to comply with the community's rules, while bringing about a state of affairs in which aggregate utility is likely to be maximised, and without the need of inculcating utilitarian standards of rightness to members of the community, who may lack reasons for

acting in ways that would allegedly maximise the overall aggregate utility.

Notice that I am not alone in interpreting Hayek's account of the spontaneous order as purporting a set of institutions to which members of the community have instrumental reasons to comply with. In fact, Sugden (1993), suggests that Hayek's theory of the spontaneous order can be aptly framed as purporting a set of rules which members of a community continuously choose to follow precisely because they benefit from taking part to the cooperative venture:

We cannot know in advance that a particular spontaneous order will benefit a particular individual. This leaves us with the criterion that a social institution has value to the extent that it tends to assist each individual in the pursuit of his or her ends, whatever those ends may be. This is Hayek's criterion of general welfare.

This criterion can be expressed very naturally in contractarian terms, by imagining that the structure of rules that govern a society had to be agreed by all individuals. In considering alternative sets of rules, each individual's concern is with her ability to achieve her own ends: she will approve a set of rules to the extent that it tends to assist her in the pursuit of her ends. Thus, a set of rules which is recommended by Hayek's criterion of general welfare is also one which, when expressed as the terms of a possible social contract, everyone can agree on. (Sugden, 1993, p. 413).

Despite Sugden's claim according to which Hayek's formulation of the spontaneous order could be framed in contractarian terms, he does not really have in mind a contractarian bargaining scenario in which agents converge on Hayek's proposed set of rules. Sugden rather imagines a different sort of contractarianism which does not entail convergence on a set of rules from an external standpoint, such as the state of nature, as it only entails an agreement that takes place from within the spontaneous order:

Thus the question is not, "Is this society one that we would have chosen in some state of nature?" Rather, it is: "Given where we are,

can we agree to continue to live by these existing rules; or, if not, can we agree to any changes?" (Sugden, 1993, p. 421).

What Sugden seems to suggest is that Hayek's theory might not be justified in a full-blown contractarian fashion, as bargainers who are in the state of nature, might be driven by considerations that lead them to prefer different institutional arrangements. However, those who are already in the spontaneous order continuously form reasons to comply with its rules, as their participation to the cooperative venture equips them with the ability to shape their specific terms of cooperation.

Essentially, Sugden claims that compliance with Hayek's account of the spontaneous order is obtained in an instrumentalist fashion, in that all members, who find themselves already in the spontaneous order, have reasons to comply with its rules because playing cooperatively would enhance their chances to satisfy their own goals.

This particular feature embeds another crucial aspect which consists in the ability of a spontaneous order to secure compliance with its rules even in *dynamic* fashion. In fact, by defining a protected sphere of action which allows members of the community to continuously shape their terms of cooperation, through their decentralised actions, the spontaneous order is supposed to continuously enhance members of the community's chances to pursue their own goals. As such, compliance with the rule of a spontaneous order is supposed to secure dynamic *stability*, as everyone has a dynamic incentive to participate to the cooperative venture.

A Critique

The instrumentalist character of the theory of the spontaneous order provides much appeal to the theory. In fact, on the one hand, institutional arrangements underpinned by the Lockean triad would continuously generate incentives for members of the community to comply with the rules of a spontaneous order, thus securing stability, and, on the other, compliance with such a set of rules would

allegedly bring about states of affairs in which aggregate utility is maximised.

The instrumentalist character, we have seen, is due to two main and strictly connected features that define the very idea of spontaneity: first, the institutional arrangements underpinned by the Lockean account of the spontaneous order are allegedly able to create the best conditions for all members of the community to pursue their own goals, on the basis of their own local knowledge; second, members of the community are continuously able to shape their terms of cooperation in ways that meaningfully respond to their structures of preferences and goals. In Hayek's mind, we have seen, these particular features obtain by defining a protected sphere of actions through the Lockean triad.

In what follows, I will attempt to move a critique to the Lockean account of the spontaneous order, which impinges precisely in its alleged ability to sustain dynamic compliance with its rules in an instrumentalist fashion. In fact, I will suggest, individual transactions taking place within a spontaneous order may alter the distribution of rights in ways that systematically prevent some members of the community from pursuing their own goals on the basis of their own knowledge, and from shaping their specific terms of cooperation in ways that meaningfully respond to their structure of preferences and goals.

The argument, in a nutshell, works as follows:

- (1) The Lockean account of the spontaneous order entails a distribution of rights to life, liberty and property.
- (2) Such a distribution of right is chosen on the ground that it defines a protected sphere of actions which creates the best conditions for members of the community to pursue their own goals on the basis of their own knowledge, by allowing them to shape their terms of cooperation.
- (3) Such an institutional arrangement allegedly equips members of the community with instrumental reasons to comply with the rules of the spontaneous order, thus securing stability.
- (4) Members of the community can alter the distribution of rights

underpinned by the Lockean account through consent.

(5) Derivative distributions of rights arising from individual transactions may bring about states of affairs in which in some members of the community are prevented from pursuing their own goals on the basis of their own knowledge and from shaping their terms of cooperation in ways that meaningfully respond to their structure of preferences.

(6) Derivative distributions of rights may, in turn, undermine the instrumentalist character of the spontaneous order, as some members may cease to see their community as an enterprise from which they stand gain, thus failing to form instrumental reasons to comply with its rules, and undermining its stability.

Since we have already dealt with (1), (2) and (3), we may aptly start with (4), which, in its simplest formulation, claims that members of a community are allowed to change the distribution of rights underpinned by the Lockean account through consent.

Clearly, the ability to change the distribution of rights cannot meaningfully refer to scenarios in which members of the community voluntarily converge on a new system rules (e. g. a scenario in which property rights are left unassigned), as this would stand in contrast with (3). In fact, if members of the community are said to have instrumental reasons to comply with the set of rules of a spontaneous order, it would not make sense to seek convergence on a different set. Alterations, in this regard, are to be seen as the large-scale consequences of individual transactions in which one sells and buys portions of rights.

For instance, consider the case of the market. Previously, we have seen that the greatest virtue that Hayek ascribes to market processes consists in their ability to collect, through the price-system, information that helps coordination between sellers and buyers, and which aptly signals to entrepreneurs which products members of the community are in the market for. This particular effect obtains because members of the community, by transacting, convey crucial

information about their preferences and reserve prices. The ability to convey such information, on the one hand, allows members of the community to shape their specific terms of cooperation, and, on the other, allegedly creates the best conditions for the pursuit of their own goals on the basis of their own local knowledge. In order to illustrate this particular aspect, we may resort to James Buchanan's illustration of how market processes allow us to produce changes in the environment that would ideally respond to our demands:

Each dollar vote in the market becomes "positively effective" to the individual, not only in providing him with a unit of the chosen commodity or service, but also in generating changes in the economic environment. In either of these senses a dollar vote is never overruled; the individual is never placed in the position of being a member of a dissenting minority. (Buchanan, 1954, pp. 334-343).

Buchanan rightly suggests that 'dollar voting' is distinct from political voting systems in which votes spent on dissenting minorities are ultimately overruled, in that they fail to produce meaningful responses to our structure of preferences and goals. In fact, all dollars spent on the market enter the price-system thus generating changes in the economic environment which allegedly respond to the information we convey. However, unlike political voting, dollar voting may place members of the community in very unequal positions in transacting. And such unequal positions define, in turn, members' ability to shape their terms of cooperation in ways that meaningfully respond to their structure of preferences and goals. In fact, members of the community, who find themselves in a relative power position as a result of market distributions, can generate changes to the economic environment that are more responsive to their structure of preferences and goals.

Such differences in bargaining power may generate states of affairs in which some members of the community are prevented from bringing about any significant change in the economic environment that would meaningfully respond to their preferences (5). This particular problem could potentially undermine the

instrumentalist character of the community, as some members may cease to see their community as an enterprise from which they stand to gain (6).

Hayek answers to this particular problem both in *The Road to Serfdom* and in *Law, Legislation and Liberty* by arguing in favor of a *unified minimum* which would insure members of the community against misfortune:

There is no reason why in a society which has reached the general level of wealth which ours has attained should not be guaranteed to all without endangering general freedom. There are difficult questions about the precise standard which should thus be assured... but there can be no doubt that some minimum of food, shelter, and clothing, sufficient to preserve health and the capacity to work, can be assured to everybody. Indeed, for a considerable part of the population of England this sort of security has long been achieved. Nor is there any reason why the state should not assist [...] individuals in providing for those common hazards of life against which, because of their uncertainty, few individuals can make adequate provision. Where, as in the case of sickness and accident, neither the desire to avoid such calamities nor the efforts to overcome their consequences are as a rule weakened by the provision of assistance – where, in short, we deal with genuinely insurable risks – the case for the state’s helping to organize a comprehensive system of social insurance is very strong. [And] there is no incompatibility in principle between the state’s providing greater security in this way and the preservation of individual freedom. Wherever communal action can mitigate disasters against which the individual can neither attempt to guard himself nor make the provision for the consequences, such communal action should undoubtedly be taken. (Hayek, 1944, p. 147-148).

There is no reason why in a free society government should not assure to all, protection against severe deprivation in the form of an assured

minimum income, or a floor below which nobody need to descend. To enter into such an insurance against extreme misfortune may well be in the interest of all; or it may be felt to be a clear moral duty of all to assist, within the organised community, those who cannot help themselves. So long as such a uniform minimum income is provided outside the market to all those who, for any reason, are unable to earn in the market an adequate maintenance, this need not lead to a restriction of freedom, or conflict with the Rule of Law. (Hayek, 1976, p. 87).

The sort of proviso that Hayek sets on distributions of rights emerging from market processes, though, does not seem to rest on the desideratum of securing the instrumentalist character of our communities. In fact, he frames the idea of providing a unified minimum ultimately as a social insurance against the uncertainties of life, rather than as a mechanism devoted to make sure that our social orders remain a cooperative enterprise. This, I believe, drives him to underestimate the role that inequalities unintentionally arising from market processes could play in bringing about states of affairs in which one cease to see her social order as a cooperative enterprise from which she stands to gain.

In fact, if one takes seriously the claim according to which rights are instrumental to the preservation of the spontaneous order, redistributive actions should take place not merely as a safety net on the grounds of the uncertainties of life, but also in order to secure that members of the community are still able to shape their specific terms of cooperation in ways that meaningfully respond to their structure of preferences and goals.

Moreover, inequalities in bargaining power, unintentionally arising through market transactions, may also shape our decisions to accept terms of cooperation that would make us alienate other rights. In this regard, Elizabeth Anderson offers an example which I have already quoted in the introduction:

Consider some facts about how employers today control their workers. Walmart prohibits employees from exchanging casual remarks

while on duty, calling this "time theft." Apple inspects the personal belongings of their retail workers, who lose up to a half-hour of unpaid time every day as they wait in line to be searched. Tyson prevents its poultry workers from using the bathroom. Some have been forced to urinate on themselves, while their supervisors mock them. About half of U.S. employees have been subject to suspicionless drug screening by their employers. Millions are pressured by their employers to support particular political causes or candidates. If the U.S. government imposed such regulations on us, we would rightly protest that our constitutional rights were being violated. But American workers have no such rights against their bosses. Even speaking out against such constraints can get them fired. So most keep silent. (Anderson, 2017 p. xix).

Here, Anderson pregnantly expresses her worries about how inequalities could undermine our ability to shape our terms of cooperation in the labor market. In fact, differences in bargaining power may place employers in an incredibly privileged position, which allows them to entirely determine the conducts that employees are to follow.

The worry, here, is not merely that largely unequal positions in bargaining power shape one's decision to give up on rights, thus bringing about a state of affairs which most of us would find intuitively despicable. But rather that, by giving up on specific instances of their rights, employees increasingly lose their ability to shape their terms of cooperation in ways that meaningfully respond to their structure of preferences. And this yields the risk of bringing about states of affairs in which the stability of a community is threatened, as some members cease to see it as a cooperative enterprise from which they stand to gain.

There are other examples that strongly suggest that we should be worrying about derivative distributions of rights emerging from market transactions. Especially transaction concerning social environment shaping rights, such as rights to freedom of expression.

For instance, consider the case of contracts of silence or non-disclosure agreements. Suppose, for instance, that I had a very bad experience at Bob's restaurant. In fact, the food was bad, and the service even worse. Bob is very much concerned about my experience, as he worries that I could write a bad review on tripadvisor. As such, he proposes me a contract of silence: I sell my right to write a review in exchange for some other goods.

By selling my silence, I restrict the set of social environment shaping actions that fall within my right to freedom of expression. Clearly, selling my right to write a review about Bob's restaurant does not undermine my ability to convey signals that would allow me to shape my terms of cooperation, broadly understood. Such an ability, indeed, is a function of the set of all specific actions that possess this particular character.

This, however, far from being consolatory, is a source of worry. In fact, if I do not perceive a specific instance of selling my right to freedom of expression as problematic in terms of the overall impact on my ability to shape the social environment, I also fail to grasp that, by selling many specific instances of this particular right, I largely restrict my ability to perform social environment shaping actions, to the point that I could find myself unable to shape my specific terms of cooperation.

In this regard, large-scale unintended consequences of small scale transactions concerning one's right to freedom of expression, may bring about a state of affairs in which contracts of silence are somewhat pervasive: they concern the relationship between employers and employees, prenuptial agreements, controversies between customers and manufacturers, the relationship between patients and doctors, etc.⁵⁵

⁵⁵Garfield (1997) provides an extensive list of cases in which contract of silence have now become the norm. He also suggests to deny the enforcement of many instances of these contracts, as they allegedly threaten public access to information: "My thesis is that contracts of silence threaten public access to information and, therefore, warrant careful judicial regulation. While recognizing that parties may voluntarily enter into contracts of silence, and that parties may receive separate compensation for their commitments to silence, I nevertheless recommend that courts deny enforcement to these contracts when the public interest in access to the suppressed information outweighs any legitimate interest in contract enforcement", p. 266.

Someone may object that if I do not perceive specific instances of selling my right to free speech as problematic in terms of the overall ability to convey information and making the social environment adapt to my preference structure, this is the result of irrationality. As such, the argument I have been offering is defective insofar as it claims that only agents with cognitive limitations in the tractability of the decision problem may be affected by this sort of consequences.⁵⁶

However, our inability to shape the social environment does not merely come from our defective evaluation of the impact of specific instances of selling one's right of expression. In fact, the ability to shape the social environment in ways that make it adapt to our preference structure partly depends on other people's transactions that are fully rational. In fact, our ability to shape our terms of cooperation, through social environment shaping actions, is also a function of other people's rights. For instance, if Alf systematically buys other people's specific instances of their right to free speech, his ability to shape his terms of cooperation may annihilate mine, whether or not I decide to sell him any right (e. g. the case of paid protesters).

As such, distributions of rights preventing some members of the community from shaping their specific terms of cooperation, or from pursuing their own goals on the basis of their own knowledge, may be brought about by transactions over which they have no control.

Hence, we should conclude that the Lockean triad, on its own, is unable to secure dynamic compliance in an instrumentalist fashion from all members of the community. In fact, inequalities arising from individual transactions may bring about states of affairs in which only some members of the community maintain the ability to pursue their own goals and to shape their specific terms of cooperation. When this obtains, some members of the community may cease to see their community as a cooperative enterprise from which they stand to

⁵⁶Notice that such an objection rest on an overly demanding account of rationality. In fact, it is utterly demanding to suppose that members of a community would be able to assess the impact that small-scale transactions, in which one sells her silence, would have in undermining one's ability to shape her terms of cooperation.

gain, thus undermining its stability.

Derivative distributions of rights and the Instrumentalist Character

One thing that I wish to emphasise is that our concern with inequality is not principled. In fact, if we take seriously the claim according to which the price system is the best device we have at our hands to convey decentralised information which would bring changes to the economic environment in ways that meaningfully respond to our structure of preferences, we are bound to accept that inequalities are an inevitable outcome of market distributions.

Markets for social environment shaping rights, such as rights to freedom of expression, are by no means an exception, in that the very act of selling one's right constitutes a way to shape one's social environment. In fact, I may not care about performing certain social environment shaping actions, as I do not care about bringing about certain changes in my social environment in some particular aspects, and rather be willing to trade particular instances of a certain right in exchange for instances of some other rights, which would allow me to generate changes in other aspects of the social world.⁵⁷ Therefore, to pose constraints on one's ability to trade off rights would, in itself, bring about a state of affairs in which one's ability to shape her specific terms of cooperation, in ways that meaningfully respond to her structure of preferences, is restricted. Inequalities emerging from such transaction, thus, are a price we must willingly pay if we care about preserving the spontaneous orders.

However, the preservation of a spontaneous order calls also for limitations to our right to consent to transactions, or for redistributive actions aimed at mitigating inequalities in the distribution of rights unintentionally emerging from individual transactions, as they may lead to states of affairs in which some members of the community are prevented from shaping their specific terms of cooperation, thus undermining the instrumentalist character of our institutional

⁵⁷On how selling rights would constitute an instance of shaping one's environment, see Muldoon (2016), Sen (1970), Gibbard (1974) and Basu (1984).

arrangements.

Our concern, thus, is not toward inequalities *per se*, nor it is grounded in republican conceptions of freedom underpinned by the concept of independence or non-domination.⁵⁸ Rather, it stems from considerations that implicitly follow from Hayek's account of the spontaneous order and which consist in preserving an order's ability to continuously adapt to members of the community's structure of preferences and goals, through their ability to shape their specific terms of cooperation.

In this regard, the problem of Hayek's account of the spontaneous order lies in that the distribution of rights underpinned by the Lockean triad may, in origin, succeed in granting us the ability to equip us with the best conditions to pursue our own goals on the basis of our own knowledge, but may fail to maintain its instrumentalist character in its actual operations, as derivative distributions may compromise our ability to shape our specific terms of cooperation, thus bringing about states of affairs in which some members of the community fail to see their community as a cooperative enterprise from which they stand to gain.

This main problem might be encapsulated by Nozick's claim according to which 'liberty upsets patterns'(Nozick, 1974, p. 160-164). In fact, the system of rules Hayek envisages as the institutional arrangement of the spontaneous order, are chosen precisely on the ground that they establish distributive patterns that allegedly shape our reasons to continuously comply with the rules of our communities.

Therefore, if derivative distribution of rights fail in shaping our reasons to comply, it follows that Hayek's theory calls for amendments to the system of rules he envisages. Such amendments may be aptly framed either as limitations to our right to consent to individual transactions, or as redistributive actions aimed at restoring members' of the community's ability to shape their specific terms of cooperation.

⁵⁸See, in particular, Pettit (2012) and List and Valentini (2016).

The Asymmetry Objection

A plausible objection to the amendment I have just proposed may insist on that by suggesting that derivative distributions of rights, emerging from the Lockean triad, may compromise the instrumentalist character of a certain community and that, as such, agents have instrumental reasons to incorporate the amendment, I do not show that political institutions could be designed and devised so as to prevent these evolutionary patterns. In fact, one may plausibly point out that Hayek's epistemological premises, as illustrated in the first chapter, imply the impossibility of designing fine-grained institutional arrangements in order to achieve certain desired outcomes.

Such an objection, I believe, has some important merits in that it highlights a potential *asymmetry* in my discussion about political institutions from an Hayekian standpoint. In fact, it may seem that if, on the one hand, I argue in favour of the institutions of the spontaneous order precisely on the ground of their ability in allowing cooperation through decentralised decision-making, without the need for fine-grained institutions that would require an omniscient planner, on the other, I suggest that such fine-grained rules are ultimately required in order to make the theory to work. In particular, it might seem that I suggest that, although a centralised political authority is ultimately unable to create the conditions for cooperation through fine-grained rules, I do imply this ability when introducing the amendment.⁵⁹

In this respect, I wish to offer one main counter-argument. In particular, I wish to suggest that the proposed amendment does not require fine-grained institutional design in order to prevent evolutionary patterns that would compromise the instrumentalist character of a given community because it does not aim at achieving *specific* outcomes. In fact, complex problems can be addressed in both complicated and simple ways, and while complicated solutions require fine-grained knowledge, simple ones do not.⁶⁰

In order to illustrate this particular point, let me propose the following

⁵⁹I wish to thank Mark Pennington and Edward Hall for pointing out this particular aspect.

⁶⁰On this particular aspect see Epstein (1997).

example. Suppose that Betty manages a tennis club and wishes to develop a system of rules that would allow members to make efficient use of the two available tennis pitches. Assume also that Betty is unable to collect the relevant and dispersed knowledge about members' preferences and needs, and that, as such, she cannot come up with a detailed set of rules that would optimally allocate each time slot to specific members. Hence, she decides to set up a booking system that functions on a first-come and first-served basis. Essentially, members can log on to the booking system and reserve the time slots that match their needs.

However, Betty soon realizes that some members tend to book time slots during rush hours for a long time span. In particular, these members reserve well in advance both tennis pitches from 9 am to 8 pm every weekend. As a result, many other members who, due to other commitments, can only play during those hours never manage to book a tennis pitch. Hence, they threaten to end their subscription and join a different tennis club. Assume also, for the sake of the argument, that if these members quit, Betty would be unable to fund the maintenance of the tennis' pitches due to the lack of revenues coming from subscriptions.

In this scenario, Betty is faced with the necessity of designing and devising a new system of rules that would keep these members into the club by allowing them to reserve a sufficient number of time slots per month. However, there are various ways of approaching this problem. For instance, Betty could attempt to collect preferences and needs of neglected members and reserve them a certain set of *specific* time slots on the basis of these preferences and needs. This attempt would be aimed at generating specific outcomes on the basis of fully detailed knowledge that she aims to gather from individual members. However, if obtaining such knowledge is ultimately impossible, perhaps because neglected members cannot tell in advance what their schedule will look like in the coming week, Betty's efforts in assigning specific slots on the basis of neglected members' preferences and needs are likely to fail.

A second approach, though, could prove feasible. For instance, Betty could

prevent the 'abusing' members from booking the tennis pitches too in advance, or reserve both tennis pitches to neglected members one weekend per month. In fact, Betty knows that these members are willing to play on weekends, but is ultimately unable to assign specific time slots as she cannot collect the required fine-grained information. Hence, by reserving the tennis pitches one weekend per month to neglected members, she allows them to determine which specific slots would suit their needs without the need for collecting fine-grained and dispersed knowledge.

These two approaches are different in the degree of knowledge required by Betty. In fact, while the first assumes Betty's ability to collect neglected members' preferences and needs in a fully detailed fashion, the second requires only a general knowledge of members' needs. However, such a difference in degree also shapes a difference in kind when analysed from a diverse perspective. In particular, while in the first scenario Betty aims at laying out in full details the members' terms of cooperation, by assigning each slot in a fully detailed fashion, in the second scenario she designs a general framework of rules that partly leaves in the members' hands the possibility to shape the allocation of time slots.

In this respect, the amendment that proposes to prevent transactions or to redistribute rights, in an effort to prevent derivative distributions that would compromise the instrumentalist character of a given cooperative enterprise, should be read in the latter fashion. Essentially, I do not aim to propose that centralised political authorities are able to design and devise fine-grained institutions able to achieve specific outcomes on the basis local and dispersed knowledge. Rather, I suggest that a general set of rules that leaves in the hands of fellow members of a community the possibility to shape their terms of cooperation, and maintain the community's instrumentalist character, can be designed and devised.

In this respect, I wish to stress, there is no significant difference, from an Hayekian standpoint, in claiming that we can plausibly design and devise a system of rules (such as the bundle of rights embedded in the Lockean triad) that

allows members of a community to shape their terms of cooperation according to their structure of preferences and goals, and suggesting that derivative distributions of rights that would compromise the instrumentalist character of a community can be prevented by redistributing rights or preventing certain kinds of transactions. In fact, both efforts do not require fine-grained knowledge of members of the community's preferences and goals, nor they aim at achieving specific outcomes.

The idea of a redistribution of rights, indeed, is different from designing fine-grained institutions aimed at shaping terms of cooperation in a fully detailed fashion, on the basis of knowledge that a planner cannot plausibly collect. And such a difference lies precisely in that the aim of such a redistributive policy is not specific, in that it does not aim at generating specific outcomes. Rather, it seeks to allocate rights in a way that leaves in the hands of the fellow members of a community the possibility to shape their social environment according to their structures of preferences.

Conclusion

Concluding, let us summarise some of the aspects I have been tackling within this chapter. First, I have attempted to illustrate one peculiar feature of the spontaneous order, which consists in its instrumentalist character. Particularly, I have been trying to support Hayek's claim according to which the Lockean triad, by virtue of allowing for adaptations to members of the community's structures of preferences and goals, equips members of the community with instrumental reasons to be part of the cooperative venture.

Second, I have offered an objection to the Hayekian formulation of the spontaneous order that tackles its alleged ability to keep its instrumentalist character over time. Particularly, I have been trying to show that the Lockean triad merely defines an initial distribution of rights which allegedly possesses an instrumentalist character, but may be unable to secure compliance in an instrumentalist fashion once the initial distribution is altered through individual transactions.

In fact, derivative distributions of rights may prevent some members of the community from shaping their specific terms of cooperation, such that they would have no interest in continuing to comply with its rules.

The next chapter, on the other hand, shall be devoted to show that our worry about the difficulties in sustaining compliance in an instrumentalist fashion should not be limited to formal institutional settings. In fact, informal institutions such as social norms and conventions may undermine the instrumentalist character of our communities even when formal institutions are able to equip individuals with the ability to continuously shape their terms of cooperation.

Social Norms and The Instrumentalist Character of The Spontaneous Order

Introduction

In the second chapter, I offered a critique of the Lockean account of the spontaneous order which tackles its instrumentalist character. Particularly, I have attempted to show that derivative distributions of rights, unintentionally emerging from individual transactions may prevent some members of the community from shaping their specific terms of cooperation. When this obtains, members may fail to see their community as an enterprise from which they stand to gain, thus undermining the community's instrumentalist character and its stability. This particular problem relevantly affects the appeal of the theory, as it undermines its constitutive aim, which is that of creating the best conditions for members of the community to pursue their own goals, on the basis of their own knowledge, and to shape their specific terms of cooperation.

If the critique is successful, I have claimed, it posits a desideratum on Hayek's account of the spontaneous order. In fact, it demands to limit members' right to trade off goods and rights, or it requires to redistribute rights when derivative distributions prevent some members of the community from shaping their specific terms of cooperation in ways that meaningfully respond to their structure of preferences and goals.

In this chapter, I shall offer a similar critique. This time, though, I will not be concerned with distribution of rights arising from individual voluntary transactions involving goods and rights, but rather with how social norms affect members of the community's exercise of rights that equip them with the ability to shape their social environment. In fact - I shall attempt to show - social norms may prevent members of the community from exercising their social environment shaping rights even under formal distributions of rights that allegedly grant them the ability to do so. This particular problem may lead to states of affairs in which some members of the community are prevented from making their

social environment adapt to their changing structures of preferences and goals, thus undermining the instrumentalist character of our communities.

Social norms are a particular kind of informal institutions unintentionally emerging from countless individual interactions (Hayek 1973, Schelling 1978, Brennan et al 2013) and sustained through informal sanctions (Bicchieri 2006, Brennan et al 2013). These two key features distinguish them from formal rules, such as those enforced by constitutions or laws. In fact, unlike formal laws, social norms are not the result of the will or design of a legislator, as they normally emerge as an equilibrium of strategic interactions (Bicchieri 2006, Ullmann-Margalit 1977, Lewis 1969, Sugden 1986, Young 1993). In addition, and unlike formal rules, social norms are not enforced through the threat of force. In fact, compliance is sustained through informal and decentralised sanctions which, by assumption, exclude the threat or the use of force toward those who fail to comply. Their weak coerciveness, I suggest, may be aptly captured by the fact that, unlike formal rules, social norms do not prevent members from undertaking certain courses of actions but merely increase the costs for doing so.

Such a distinction, however, should not lead us to underestimate their power in shaping our social environment. Their role, indeed, is even more pervasive than that of formal rules. In fact, social norms do not merely regulate non-public behaviours, by which I identify conducts that are unsuited to be regulated by the law, but also affect our set of opportunities in ways that define the substantive contents of rights and may shape our compliance with formal rules.

Analogously to the second chapter, my main aim is to show that norms can alter the exercise of some of our rights up to the point that some members are prevented from shaping their specific terms of cooperation in ways that meaningfully respond to their structure of preferences. When this is the case, once again, communities lose their instrumentalist character and are thus bound to instability.

If successful, such a critique posits a further desideratum to any candidate amended formulation of the theory of the spontaneous order, which is that of preventing the emergence of social norms that would constitute a threat to the

exercise of social environment shaping rights, or to dismantle already existing norms that affect the exercise of these rights.

The chapter is organised as follows: within the first section, I outline how norms shape our communities; within the second section, I deal with the identification of social norms in the attempt to distinguish them from descriptive norms; within the third section, I analyse social norms' status in terms in terms of coerciveness, in order to distinguish them from formal rules; within the fourth section, I offer a social norms based critique to Hayek's account of the spontaneous order.

Three Ways in which Social Norms shape our Communities

Consider the case of Carla, a promising Ph.D. in orthodontics with a brilliant academic career ahead of her. She lives in Caltanissetta (Sicily), where she has also started a private dental practice. Her social life is permeated with social events, and fully enriched by her friends whose company she enjoys very much.

Carla has recently started dating Mario - an entrepreneur - and they seem to get along quite well. Mario, though, does not belong to Carla's social circle and can hardly aspire to be welcomed among Carla's friends. In fact, entrepreneurs, in Caltanissetta, do not enjoy a good reputation and are often labeled as 'money-driven merchants'. Carla's friends, on the other hand, are all academics. Their lives are devoted to unveiling the mysteries of science, or to unravel the deep meanings of life, in order to make the world a better place. Sure, they take some spare time from their intellectual endeavours, and they mostly devote it to appreciate the marvels of poetry.

Carla has a crush on Mario, but is ultimately unwilling to introduce him to her social circle. In fact, she is afraid of informal sanctions which would follow. Mario is a hard working guy, who minds his business, and Carla's friends would ultimately find him as unsuited to sustain profound conversations on the meaning of life. After all, he is a money-driven merchant. How could he possibly understand what they are up to?

Carla is torn by the situation, but ultimately realises that Mario is not the right guy. She would have too much to lose in carrying on with the relationship and, perhaps, marrying him. Her friends would not understand, and she would end up by being excluded from the social circle that has fully shaped her joyful life so far. Perhaps, she will consider dating Alberto. In fact, her friends would be delighted to have him on board for their weekly reading group on Derrida. Alberto, indeed, is a highly regarded and widely published Professor of Philosophy at the prestigious University of Caltanissetta, and would perfectly qualify as the sort of partner Carla should aim.

In Carla's story, we are presented with a bundle of social norms which affect non-public behaviours. In particular, they affect a private sphere that is intuitively unsuited to be governed by the law: either because formal regulation would lack effectiveness, or because it would require invading one's privacy up to an unacceptable extent, or because it would violate one's individual autonomy. In fact, within a liberal framework, marriage choices, alongside a wide range of other important choices that provide meaning to our lives (e. g. career's paths, education, parental, etc.), usually fall within a protected sphere which is supposedly impenetrable by formal rules.

Yet, social norms importantly shape these choices, or, at least, alter the payoffs of our choice scenarios. In Carla's social circle, for instance, we may suppose that a norm prescribing not to date money-driven merchants is in play. Although Carla may exhibit a preference for continuing her relationship with Mario, her desire to fit in, alongside the fear of sanctions yielded by deviations from the norm, shape her preference ranking over the available courses of actions, and, in turn, shape her decision to break up. Were the norm not in play, perhaps Carla would have acted differently.

The impact of social norms, though, is not restricted to the domain of private behaviours. In fact, they also shape the substantive contents of specific rights underpinned by formal rules, such as the right to freedom of expression. In order illustrate this, let me offer two examples.

In 2013, Guido Barilla, chairman of a popular brand of pasta, said, during

a radio interview, that he would never approve a script for one of Barilla's commercials which pictures a gay family as the main protagonist. Barilla poorly justified his statement by resorting to the idea of the traditional family as the bedrock of the society he wishes to live in.

The immediate reaction on the media was of obvious disdain with many celebrities proposing boycotts. Barilla's competitors also brilliantly took advantage of his gaffe in the attempt to gain the market share of those who used to be loyal Barilla customers and who understandably felt profoundly offended by Barilla's words. Garofalo, for instance, introduced the following catchphrase on their pasta packaging: "It doesn't really matter with whom you cook it, it just needs to be al dente". Thus, it comes by no surprise that, after a couple of months, Barilla ended up producing their first gay-friendly commercial.⁶¹

A similar story had as main protagonists the worldwide famous fashion designers Domenico Dolce and Stefano Gabbana. In an interview for the Italian magazine *Panorama* from January 2015, Mr Dolce told the interviewer: "You are born to a mother and a father, or at least that's how it should be. I call children of chemistry, synthetic children. Rented uterus, semen chosen from a catalog". The interview prompted Elton John to launch the hashtag #boycottDolceGabbana and the successful boycotting campaign induced the fashion designers to release an apologetic statement.⁶²

Barilla and Dolce and Gabbana's stories provide an example of how social norms interact with specific rights, underpinned by formal institutional arrangements. In particular, they illustrate how social norms may shape our preference ranking over courses of actions which are formally protected by individual rights, and which should ideally be insulated from external constraints. In fact, under liberal institutional arrangements, the right to free speech is not merely asserted but also protected through formal provisions (e.g. the first amendment in the US or object of article 19 of the United Nations Universal Declaration of Human Rights, according to which 'Everyone has the right to freedom of opinion and

⁶¹On Barilla's story see Davies (2013).

⁶²For an exhaustive summary of Dolce and Gabbana's story, see Ember (2015).

expression;⁶³ this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers'). Yet, norms do affect our freedom of speech to a large extent. In fact, informal sanctions, such as discrimination by customers who felt offended by Barilla and Dolce and Gabbana's words, or the expectation of such sanctions, increase the costs for expressing one's sincere opinions about delicate matters.⁶⁴

A third instance of how social norms affect our communities concerns their role in shaping our compliance with formal rules which forbid or require certain specific conducts. For instance, consider the case of laws requiring to wear helmets on motorcycles. In many Naples' neighborhoods is not uncommon to get informally sanctioned if one complies. Plausible explanations might resort to 'bully' attitudes, although some recent documentaries point to the fact that helmets often identify members of criminal organisations, such as the Camorra, who use them in order to protect their identity while committing criminal activities. What is sure, though, is that informal sanctions strongly undermine citizens of Naples' compliance with the rule.

On the other hand, there are many cases in which norms sustain our compliance with formal laws. For instance, Cialdini et al (1990, 1991 and 1998) provide interesting insights on how social norms enhance our compliance with rules aimed at reducing littering, and Posner (2000) shows how social norms profoundly affect our choices with respect to tax compliance.

On a more general level, we may think about broader attitudes generating macro social norms concerning law-abidingness. In the south of Italy, for instance, a popular joke aptly pictures traffic signs as no more than a prudential advice. In fact, it is not uncommon to be mocked when stopping at the red light, as those who do are deemed as either narrow-minded, stupid or cowards. In other communities, on the other hand, non-compliance with formal rules is perceived as outrageous, even though compliance may appear as bringing about

⁶³General Assembly of the United Nations (1948) pp. 4–5.

⁶⁴Basu (1998) and Fiss (1985) precisely focus their attention on how norms affect rights to freedom of expression.

overall sub-optimal states of affairs.

Identifying Social Norms

Brennan et al (2013) classify theories of social norms into two main categories. On the one hand, we have reductive accounts, which identify norms as well-established social practices; on the other, we have accounts introducing normative elements.

The reductive approach can be illustrated by the following example. Suppose Betty, Mary and Alf enjoy meeting every Friday after work at the university pub and to update each other about their ongoing research. There are no scheduled meetings but merely an empirical expectation that everyone will show up, on that particular time of the week, at that particular place. Empirical expectations merely represent expectations about the likelihood that a certain event will occur, and are shaped by the observation of regularities in their respective behaviours. In fact, Betty, Mary and Alf's expectations do not entail any ought on their reciprocal behaviours, nor there are sanctions associated with defections. However, defections, by changing their empirical expectations, may undermine the norm. In fact, if Betty, Mary and Alf lack independent reasons to go to the university pub on Fridays, repeated defections may lead to abandoning the practice.

Norms of this sort are aptly captured by David Lewis' account of conventions (1969), which defines the conditions for the identification of norms in the following form:

A regularity R in the behavior of members of a population P when they are agents in a recurrent situation S is a convention if and only if it is true that, and it is common knowledge in P that, in almost any instance of S among members of P,

- (1) almost everyone conforms to R;
- (2) almost everyone expects almost everyone else to conform to R;

- (3) almost everyone has approximately the same preferences regarding all possible combinations of actions;
 - (4) almost everyone prefers that any one more conform to R, on condition that almost everyone conform to R;
 - (5) almost everyone would prefer that any one more conform to R', on condition that almost everyone conform to R',
- where R' is some possible regularity in the behavior of members of P in S, such that almost no one in almost any instance of S among members of P could conform both to R' and to R. (Lewis, 1969, p. 78).

In fact, Betty, Mary and Alf (1) normally go to the University pub on Fridays after work, (2) exhibit empirical expectations about the fact that (almost) everyone will show up, (3) are happy to meet at the University pub, as it is close to their workplace (4) are willing to go to the pub provided that (almost) everyone will, (5) prefer that any one shows up, provided that (almost) everyone does.

Lewis' account, thus, explains norms in terms of salient solutions to pure coordination games, where players' preferences are perfectly aligned. In fact, Betty, Mary and Alf already exhibit a preference for meeting and catching up about their research, and the observation of regularities in their respective behaviours indicates that going to the university pub after work on Fridays would satisfy their preference. Therefore, 'go to the university pub on Friday nights after work' constitutes a rule that provides a salient solution to their shared preference for coordination.

Therefore, Bicchieri notices, Lewis' conditions capture scenarios in which our conformity with a particular norm is always dictated by straightforward self-interest:

We conform because such norms make life easier for us, because we want to 'fit in' or do the right thing – as when we adopt a new

fashion – or simply because they provide evidence of what is likely to be effective, adaptive behavior, as when we bought Internet stocks because many people we know were buying them and were doing well. Often there are good prudential or informational reasons to “do as the Romans do.” Conformity to a descriptive norm may be motivated by a desire to imitate others’ behavior in uncertain or ambiguous situations (Bicchieri, 2005, p. 34).

Often, though, scenarios to which norms apply do not resemble situations in which agents’ preferences are perfectly aligned, but rather where mixed motives come into play. In Carla’s story, for instance, she is torn between two non-compossible aims: on the one hand, she exhibits a desire to fit in and to avoid incurring in informal sanctions; on the other, she has a preference for dating Mario.

In Carla’s scenario Lewis’ (3) does not obtain, as Carla exhibits preferences which are unaligned to those shared by members of her social circle. In this regard, the norm prescribing not to date the money-driven merchant shapes an impure coordination game, as Carla’s compliance comes at the cost of giving up on her aim to continue dating Mario.

Lewis’ conditions, we may observe, are unable to account for Carla’s scenario, as they cannot not explain how compliance with the norm taking place in Carla’s scenario is sustained, given the clash of preferences between Carla and members of her social circle. In this regard, Lewis’ conditions merely account for *descriptive* norms, that is for norms that highlight a solution that perfectly accommodates for all players’ preferences over the available courses of actions. These norms do not entail any sacrifice in complying with them, as they merely highlight to players a way to solve their struggle in coordinating.

Mixed motives scenarios, indeed, require normative elements in order to explain compliance with norms. In particular they require us to lay out what drives agents’ compliance with norms in choice scenarios where agents exhibit unaligned aims.

One plausible solution is offered by rational choice models, according to which our behaviours are ultimately shaped by the sanctions yielded by non compliance with norms (Rommetveit 1955, Axelrod 1986, Coleman 1990). The normative element is precisely captured by sanctions, whose expectation invites members of a community to comply with certain norms even when compliance comes at the cost of sacrificing one's other aims. Such an approach explains our behaviours as the result of a payoff-maximising strategy in which agents deliberate on which course of actions they should undertake by factoring in the expected costs yielded by sanctions. Rational choice models, thus, would explain Carla's choice as emerging from a deliberation procedure that weighs the benefits of dating Mario and the costs of expected sanctions coming from her social circle.

However, Bicchieri (1990, 1997, 2006) points out that even rational choice models offer us a rather incomplete approach to norms' compliance. In fact, these models are unable to explain the persistence of norms that are not upheld by sanctions. In order to illustrate Bicchieri's argument we may aptly frame Carla's scenario as a situation in which non-compliance with the norm prescribing not to marry the money-driven merchant does not yield sanctions. Under such circumstances, Bicchieri would claim, compliance can still obtain but rational choice models are ultimately unable to explain why.

In this regard, Bicchieri suggests that our compliance with norms, in mixed motives scenarios, is partly shaped by a preference for conformity developed by members of a community who learn to act cooperatively. In particular, such a preference would give rise to a meta-norm inviting compliance with specific norms, in specific circumstances. Let us analyse her formulation:

Let R be a behavioral rule for situations of type S , where S can be represented as a mixed-motive game. We say that R is a social norm in a population P if there exists a sufficiently large subset $P_{cf} \subset P$ such that, for each individual $i \in P_{cf}$:

Contingency: i knows that a rule R exists and applies to situations

of type S;

Conditional preference: *i* prefers to conform to R in situations of type S on the condition that:

(a) Empirical expectations: *i* believes that a sufficiently large subset of P conforms to R in situations of type S;

and either

(b) Normative expectations: *i* believes that a sufficiently large subset of P expects *i* to conform to R in situations of type S; or

(b') Normative expectations with sanctions: *i* believes that a sufficiently large subset of P expects *i* to conform to R in situations of type S, prefers *i* to conform, and may sanction behavior.

A social norm R is followed by population P if there exists a sufficiently large subset $P_f \subset P_c f$ such that, for each individual $i \in P_f$, conditions 2(a) and either 2(b) or 2(b') are met for *i* and, as a result, *i* prefers to conform to R in situations of type S. (Bicchieri 2006, p.11)

The meta-norm is captured by one's conditional preference to conform, provided certain conditions obtain. In fact, if members of the community internalise the preference for conformity, they also develop a default position which explains their compliance with norms even under the absence of sanctions, and in the presence of mixed motives. In this regard, Carla's compliance with the norm may obtain even under the absence of sanctions, as her preference for conformity could, in principle, account for her decision.

Bicchieri's requirements for the emergence of a conditional preference for conformity add an important element to Lewis' account, in the form of normative expectations or normative expectations yielding sanctions. Normative expectations, indeed, capture the oughts that were missing in Lewis' account and which allow us to explain compliance in mixed motive scenarios.

In order to illustrate this particular aspect, let us analyse Carla's story under

Bicchieri's conditions. Carla, we assumed, has a preference for dating Mario. However, she also exhibits a preference for conformity which is driven by empirical and normative expectations: empirical expectations constitute beliefs about the behaviours of her friends, whereas normative expectations entail second order beliefs about how one ought to behave. In particular, in our scenario, Carla believes that everyone complies with the norm and believes that her friends believe that she ought to conform with the norm, and may sanction deviations.

Empirical expectations alone are unable to explain Carla's decision to break up with Mario. In fact, the mere fact of believing that others undertake a particular course of actions, *per se*, does not provide a reason for conformity, as it does not entail normative attitudes on one's behaviour. However, if empirical expectations are accompanied by second order beliefs about how one should behave, they can explain conformity even in the absence of sanctions. In fact, when normative expectations obtain, Carla's reasons to conform to the norm is shaped by the desire not to disappoint other people's normative attitudes, even in the absence of sanctions.

In this regard, one may wonder to what extent Bicchieri's conditional preference for conformity brings any explanatory value to the identification of norms that is not already embedded in the concept of normative expectations. In particular, one may suggest that normative expectations, even in the absence of informal sanctions, already yield some sort of psychological penalties in the form of disappointing other people's normative attitudes, which alone would explain one's compliance with a specific norm. In fact, in Carla's scenario, her compliance with the norm might simply be explained in terms of her desire not to disappoint her friends' normative attitudes, without resorting to conditional preferences for conformity.

Such an inclusive definition of sanctions would allow us to get rid of Bicchieri's conditional preference for conformity, while still allowing us to explain compliance with norms in mixed motives contexts where non-psychological sanctions do not follow our deviations from the norm. This is, indeed, the route followed by Brennan et al (2013) in laying out their requirements for the iden-

tification of norms. According to their account, indeed, a particular normative principle P is a norm within group G if and only if:

- (i) A significant proportion of the members of G have P-corresponding normative attitudes; and
- (ii) A significant proportion of the members of G know that a significant proportion of the members of G have P-corresponding normative attitudes. (Brennan et al., 2013, p. 29).

Such a definition excludes from the relevant conditions the unwarranted requirement of agents' conditional preference for conformity and merely points to common knowledge about the existence of normative attitudes toward certain behaviours. Moreover, differently from Bicchieri, Brennan et al also exclude empirical expectations from their list of necessary conditions. In fact, they suggest, norms may still exist even when we are unable to observe widespread compliance:

The Moldovans: Virtually all Moldovans judge that one mustn't urinate in public swimming pools, are disposed to disapprove of anyone who does so, and are disposed to regard such attitudes as perfectly justified. Moreover, that Moldovans generally feel this way about urinating in public swimming pools is a matter of common knowledge among Moldovans. Despite this, and unbeknownst to the Moldovan public, urinating in swimming pools in Moldova is in fact widespread. Everyone does it regularly. To be sure, no one fesses up to the fact. Indeed, they do everything they can to hide it. On the rare occasions that they discover that someone else has done it, they are outraged. When they do it themselves, they feel guilty (though perhaps also a certain frisson of guilty pleasure). They have no idea, and would be appalled to discover, that everyone is behaving in exactly the same way. (Brennan et al., 2013, p. 20).

The example is meant to suggest that, despite widespread non-compliance

with the norm 'do not urinate in public swimming pool', a norm still exists insofar as Moldovans do everything to hide their non-compliance on the ground of their fear of sanctions yielded by corresponding normative attitudes. However, one may notice, the example does not really rule out Bicchieri's condition of empirical expectations, but merely suggests that a norm can exist despite agents' empirical expectations being mistaken.⁶⁵

In fact, Moldovans would hardly form second order beliefs about other people's normative attitudes against the act of urinating in public swimming pools, in the absence of empirical expectations. Moreover, under circumstances in which non-compliance can be easily unveiled, empirical expectations, in order to emerge, must be consistent with actual practices. In fact, we could hardly form empirical expectations which stand in sharp contrast with other people's observable conducts. Therefore, we must conclude that empirical expectations are a necessary condition for the identification of a norm. In fact, they are required in order to form widespread second order beliefs about other people's normative attitudes toward a particular behaviours, and, in turn, to shape the credibility of the threat posed by sanctions associated with deviations from the norm.

In this regard, I wish to propose a set of conditions for the identification of norms that draws both from Bicchieri (2006) and Brennan et al (2013). According to the proposed formulation:

A normative principle R is a social norm in situation type S, in a population P, when:

(1) a significant part of P has empirical expectations: a large part of P believes that a large part of P will conform to R in situation type S.

(2) A significant proportion of P knows that a significant proportion of P have R-corresponding normative attitudes and are willing to

⁶⁵This particular situation obtains only under the condition that gaining access to other people's conducts is impossible or particularly costly, thus resembling a particular case of pluralistic ignorance. See Miller, Dale, and McFarland (1987).

sanction deviations in situation type S (where sanction may identify both psychological and other kinds of informal punishment).

Such a formulation, I suggest, aptly responds to the intuitions I have outlined above. First, it gets rid of Bicchieri's unwarranted condition of conditional preference for conformity, as norms' compliance can be simply explained in terms of sanctions. Second, we maintain Bicchieri's condition concerning empirical expectations, which are needed in order to form second order beliefs about other people's normative attitudes. Third, we merge Bicchieri's (b) and (b') and Brennan et al (i) and (ii) into (2).

Social Norms and Coercion

With (1) and (2) outlined, we can move on to investigating how social norms affect our behaviours. In order to do so, let us turn back to the examples that I offered within the first section. In Carla's story, the existence of a norm N prescribing not to date a money-driven merchant requires that:

- (1) A large proportion of the members of Carla's social circle have empirical expectations that others will conform to N and,
- (2) A large proportion of the members of Carla's social circle knows that a large proportion of the members of Carla's social circle has N-corresponding normative attitudes and is willing to sanction deviations.

How does the existence of N affect Carla's behaviour? Within the first section, I proposed that the best way to capture the effects of norms is to consider how the costs of sanctions alter our ranking over the available courses of actions. In fact, in Carla's example, we may explain her behaviour as ultimately shaped by a deliberation procedure which weighs her preference for dating Mario and the costs of sanctions (including psychological ones) associated with deviation from N. Normative expectations, indeed, by virtue of increasing the costs for continuing to date Mario, alter her preference ranking over the available courses

of actions and shape her decision to break up. In fact, were the norm not in play, she would have had reasons to follow a different path.

In Barilla and Dolce and Gabbana's examples, there are further interesting elements to consider. In fact, their controversial speeches indicate that they either lacked normative expectations or underestimated the intensity of sanctions associated with deviations. The boycotting campaign following their interviews, indeed, makes them reconsider their previous actions and converge on a course of action that is not subject to sanctions.⁶⁶

Analogously, social norms shaping our compliance with formal rules alter our preference ranking over the available courses of actions by increasing the costs for undertaking certain courses of actions. In fact, turning back to the example concerning laws requiring to wear helmets on motorcycles: agents have a preference for complying that is shaped by the existence of a formal rule, and that is out-weighted by the expectations of sanctions associated with deviation from the social norm proscribing to wear helmets.

In this regard, it may be useful to assess social norms in terms of their coerciveness, in an effort to distinguish them from formal rules. Such an assessment crucially turns on the definition of coercion that we employ. Hence, let us consider two classical definitions offered by Hans Kelsen and J. R. Lucas:

As a coercive order, the law is distinguished from other social orders. The decisive criterion is the element of force — that means that the act prescribed by the order as a consequence of socially detrimental facts ought to be executed even against the will of the individual and, if he resists, by physical force. (Kelsen, 1967 [1934], p. 34).

[W]e are concerned with the enforcement of decisions: we are considering the conditions under which decisions will be carried out regard-

⁶⁶Clearly, one may suggest that the boycotting campaign makes them reconsider and update their beliefs, such that Barilla's decision to produce a gay friendly commercial, and Dolce and Gabbana's apologising statement may be the result of a cognitive process which involves the internalisation of different moral values. This is, indeed, perfectly plausible. However, my goal here is to highlight how norms alter our preference rankings on the available courses of actions, without making people internalise the normative attitudes implicit in the social norm.

less of the recalcitrance of the bloody minded. We therefore define force in terms of bloody-mindedness, of what happens irrespective of how recalcitrant a man is, of what happens to him willy-nilly. Force, then, we say, is being used against a man, if in his private experience or in his environment either something is being done which he does not want to be done but which he is unable to prevent in spite of all his efforts, or he is being prevented, in spite of all his efforts, from doing something which he wants to do, and which he otherwise could have done by himself alone. A man is being coerced when either force is being used against him or his behaviour is being determined by the threat of force. (Lucas, 1966, p. 57).

In the above quoted passages, Kelsen and Lucas propose us two formulations that identify as the main distinctive element of coercion the threat or use of force toward the coerced. According to these definitions, social norms cannot be said to be coercive as they lack such a distinctive element. In fact, informal sanctions, by assumption, do not entail the use of physical force in order to secure agents' compliance with norms.

However, although such a definition is able to capture the distinction between formal and informal rules, it fails in accommodating our intuitions about the distinction between descriptive and social norms. In fact, if sanctions were not in any way coercive, we would hardly explain why people would conform to conducts they would otherwise consider as sub-optimal. In this regard, more recent accounts of coercion have focused on conditional threats that may not include the use of force or violence to compel one to perform a certain action, thus allowing us to explain compliance with social norms. Nozick's account, in particular, has been largely influential, and sets out the minimal conditions for coercion as follows:

1. P aims to keep Q from choosing to perform action A;
2. P communicates a claim to Q;

3. P's claim indicates that if Q performs A, then P will bring about some consequence that would make Q's A-ing less desirable to Q than Q's not A-ing;
4. P's claim is credible to Q;
5. Q does not do A;
6. Part of Q's reason for not doing A is to lessen the likelihood that P will bring about the consequence announced in (3). (Nozick, 1969, pp. 441-445).

If we reconsider our examples under the light of Nozick's account of coercion, social norms seem to fall under the same category of formal rules. For instance, in Carla's story:

- (1) Members of Carla's social circle have normative attitudes toward N;
- (2) Carla knows that a large proportion of the members of her social circle shares these attitudes;
- (3) Members of Carla's social circle are willing to sanction deviations from N, increasing Carla's costs for undertaking her preferred course of actions;
- (4) Carla believes that sanctions will take place if she does not conform;
- (5) Carla decides to break up with Mario;
- (6) Part of Carla's reasons for not continuing to date Mario is to avoid sanctions from members of her social circle.

Nozick, indeed, by failing to distinguish between different kinds of consequences one may bring about in securing agents' compliance, also fails in accommodating our intuitions about the distinction between formal and informal rules, which precisely rests on the threat of the use of physical force.

Moreover, condition (5) by tying coercion to cases of successful coercion, also ties coercion to one's subjective assessment of the relevance of the threats posited by coercers, bringing us to implausible conclusions. In order to illustrate this, let us consider two scenarios:

Carla's friends would only trivially sanction Carla's decision to continue dating Mario (e. g. Carla would not be invited to the next meeting of the reading group on Derrida). Yet, Carla perceives the costs of the sanction as too high and breaks up with Mario.

Carla's friend would strongly sanction Carla's decision to continue dating Mario (e. g. they would destroy her reputation). Yet, Carla is in love with Mario and decides to continue dating him.

Oddly enough, Nozick would conclude that coercion takes place in the first, but not in the second scenario. Further developments of Nozick's account of coercion, indeed, distinguish between coercion and coerciveness.⁶⁷ According to such accounts, unsuccessful attempts to coerce do not rule out the coerciveness of a particular scenario, which rests on the existence of conditional threats that coercers intend to bring about if the coercee performs a certain action.

However, these accounts of coercion are still unable to accommodate for the distinction between formal and informal rules, as they do not distinguish between threats involving the use of force, and those merely involving informal sanctions. In this regard, and restricted to the purpose of the present work, we may offer a distinction between strongly and weakly coercive rules, where 'strongly' identifies rules enforced through the threat or use of force, and 'weakly' identifies rules sustained by the threat of informal sanctions, thus excluding the use of force. At the same time, we may keep the distinction between coercion and coerciveness in order to avoid tying the definition of coercion merely to cases of successful coercion. As a result, we obtain the following taxonomy:

⁶⁷See, in particular, Carr (1988) and Lamond (2000).

	Strongly Coercive	Weakly Coercive
Coercion	Successful Coercion through the threat of force	Successful Coercion through informal sanctions
Coerciveness	Unsuccessful Coercion through the threat of force	Unsuccessful coercion through informal sanctions

Such an account, indeed, although unable to distinguish between social norms yielding sanctions of different intensity, has two crucial advantages. On the one hand, it accommodates our intuitions concerning the distinction between formal rules, such as laws and constitutions, and informal rules such as social norms. On the other, it allows us to explain compliance with weakly coercive rules, such as social norms, thus accommodating for the distinction between descriptive norms, which we comply with merely on the basis of self-interest, and social norms, which require the threat of sanctions in order to secure agents' compliance.

Social Norms and The Spontaneous Order

Social norms, by virtue of being weakly coercive, shape our behaviour without penetrating the protected sphere that liberal institutions design for us. In fact, it would be weird to claim that, in Carla's story, informal sanctions associated with deviation from the norm prescribing not to marry a money-driven merchant, would violate Carla's rights, such that her friends should be formally coerced into refraining from sanctioning her. Analogously, it would be too demanding to suggest that Barilla's and Dolce and Gabbana's customers should be coerced into refraining from boycotting, on the ground that the boycott would undermine

Barilla's and Dolce and Gabbana's right to politely express their own opinions on delicate matters.

In fact, coercion toward sanctioners brings with it three main drawbacks: first, there is a problem of efficacy, in that behaviours of this kind are hard to identify, and compliance with laws forbidding them is even harder to secure; second, there is a concern about people's privacy as the identification of this sort of behaviours may require interference with one's private life; third, there is a concern about individual autonomy, as one may point out that, for how undesirable these attitudes and sanctions are, citizens of a free society should be allowed to hold on to them.⁶⁸

Yet, social norms do regulate our behaviours and shape many decisions which profoundly affect our lives. This particular fact, I suggest, has important implications with respect to Hayek's account of the spontaneous order. This is because Hayek considers the protected sphere defined by institutional arrangements underpinned by the Lockean triad as instrumental in order to create the best conditions for all members of the community to pursue their own goals, on the basis of their own knowledge. In particular, he considers members of the community's ability to act within the boundaries of the Lockean triad as conducive to their ability to continuously shape their specific terms of cooperation, through their exercise of social environment shaping rights.

However, I suggest, social norms may undermine Hayek's claim, leading to states of affairs in which the instrumentalist character of our communities, and, in turn, their stability is threatened. The argument may be summarised as follows:

- (1) The Lockean account of the spontaneous order entails a distribution of rights to life, liberty and property.
- (2) Such a distribution of right is chosen on the ground that it defines a protected sphere of actions which creates the best conditions for

⁶⁸Bartlett and Gulati (2016) specifically tackles informal discrimination by customers and offers a rather interesting discussion concerning how governments may reasonable deal with informal sanctions stemming from normative attitudes.

members of the community to pursue their own goals on the basis of their own knowledge, and allows them to shape their specific terms of cooperation.

(3) Such an institutional arrangement allegedly equips members of the community with instrumental reasons to comply with the rules of the spontaneous order, thus securing stability.

(4) Social norms may prevent some members of the community from shaping their specific terms of cooperation in ways that meaningfully responds to their structure of preferences and goals.

(5) Social norms may, in turn, bring about states of affairs in which some members of the community cease to see their community as an enterprise from which they stand to gain, thus undermining its instrumental character and, in turn, its stability.

Since we have dealt with (1), (2) and (3) in the previous chapters, we may aptly start with (4). Social norms, we have seen, increase the costs for pursuing certain courses of actions, which fall in our protected sphere, and may, in turn, alter our preference ranking over the available courses of actions. In Carla's scenario, for instance, informal sanctions associated with deviation from the norm 'do not date the money-driven merchant', make her refrain from continuing dating Mario, which she would have done in the absence of the norm. Analogously, in Barilla's and Dolce and Gabbana's stories, informal sanctions associated with their deviation from a norm proscribing controversial speech on delicate matters, make them ex-post converge on conducts which are sanction-free.

This peculiar ability of norms brings with it crucial implications when sanctions affect members of the community's social environment shaping rights. In fact, these rights allegedly serve the purpose of defining a protected set of actions, through which members are able to convey information about their specific terms of cooperation, in order to shape their social environment in ways that meaningfully respond to their structure of preferences and goals. One's failure to perform social environment shaping actions, that reflect one's structure of

preferences and goals, results in one's failure to shape her specific terms of cooperation in ways that meaningfully respond to her structure of preferences and goals. Systematic failure to perform sound social environment shaping actions may generate states of affairs in which one ceases to see her community as an enterprise from which she stands to gain, thus undermining its stability.

In order to further illustrate this problem, let us consider the following scenarios:

Betty asks Bob to co-author a paper. Betty would indifferently use a TeX editor or a Word processing software, and asks Bob about his preferred solution. Bob hates TeX, and has never really learned how to use it. However, Bob feels pressured by an ongoing social norm in his department. In fact, everyone uses TeX, and those who do not are subject to mockery. Therefore, he chooses TeX.

Some members of the community of Broomhill like Michael Buble. However, such a preference must be hidden, as a large proportion of members of Broomhill would regard it as shameful. In fact, only foolish people could possibly enjoy listening to such a disengaged music genre. Michael Buble organises a concert in Broomhill, but those who appreciate him feel pressured not to go, as their preference could be easily unmasked. Perhaps, they could go to the next concert, which is scheduled in July, when Broomhills' intellectuals will be on vacation. However, nobody shows up to Buble's concert and he never comes back to Broomhill.

In these scenarios, we are presented with two instances of one's failure to convey information that reflects one's structure of preferences and goals, and which, in turn, results in one's failure to shape her specific terms of cooperation in ways that meaningfully respond to one's structure of preferences and goals.

In the first case, Bob chooses to write the paper on a TeX editor, despite the possibility of a better option at hand. His performative utterance in answering

Betty generates a specific term of cooperation which Bob does not really appreciate. In the second case, members of Broomhill fail to convey signals, through the price system, that would bring Michael Buble back for a concert in July.

Clearly, specific instances in which one complies with a norm, despite idiosyncratic preferences, do not undermine one's ability to shape her social environment in ways that meaningfully respond to her preferences and goals. Yet, many instances, encompassing various aspects of one's life, may bring about states of affairs in which one systematically fails to define specific terms of cooperation in ways that one finds acceptable. This, I suggest, may result in one's failure to see her community as a cooperative enterprise from which she stands to gain, thus undermining its stability (5).

One may plausibly object that social norms, by virtue of being only weakly coercive, do not prevent members of the community from exercising their social environment shaping rights in ways that meaningfully respond to their structure of preferences and goals, but merely increase the costs for doing so. Therefore, if, for instance, we choose to refrain from exercising our right to free speech on the ground of informal sanctions, we do so insofar as we fail to grasp the relevance that the exercise of our right to free speech has in shaping our social environment. As such, my argument is defective as it only applies to people that are not fully rational.

This argument has some merits: in fact, social norms do not bring about distributions of rights in which I am contractually bound and, thus, unable to exercise my right to free speech at later stages. Moreover, such an objection does not rest upon an overly demanding account of rationality. In fact, I may come to realise, through a process of trial and error, that failure to exercise my right to free speech has crucial consequences on my ability convey information that would allow me to shape the social environment by making it adapt to my preferences and goals. Once I acknowledge it, I can always exercise my right to free speech at later stages, as failure to exercise it in the past does not bind me not to exercise it in the future.

However, such an argument overlooks how pervasive the role of norms is and

how hard is to convey information that could free us from informal sanctions. In this regard, two counter objections can be offered: first, our compliance with specific norms is often shaped by pressing needs; second, when norms are well-established, our sound exercise of social environment shaping rights does not produce meaningful changes.

For instance, in Barilla and Dolce and Gabbana's stories, their compliance with the norm is shaped by the pressing need of protecting their businesses and is also driven by the fact that their controversial speeches would not either produce any substantive change in their social environment, nor bring about states of affairs in which their right to hold controversial positions will not be affected by informal sanctions.

One may also object that social norms, by virtue of being the result of countless individual interactions, embed, in themselves, knowledge that is dispersed among individuals. As such, they already shape specific terms of cooperation in ways that meaningfully respond to our structures of preferences and goals. However, such an objection fails to distinguish descriptive norms, taking place in scenarios where agents' preferences are perfectly aligned, from norms guiding our behaviours in mixed motives scenarios. In fact, in pure coordination games, solutions perfectly embed knowledge that is dispersed among agents; in mixed motives games, on the other hand, norms compliance comes at the cost of sacrificing one's other aims. These aims are not conveyed to other agents and, thus, we fail in shaping our specific terms of cooperation according to our structure of preferences and goals.

Social norms, thus, posit a threat to an order's ability to be continuously shaped by its members. And this particular problem brings with it the risk of undermining a community's instrumentalist character, as some members of the community may fail to see their communities as a cooperative enterprise from which they stand to gain.

As such, social norms call for a further desideratum that sound institutional arrangements, underpinned by a theory of spontaneous order, should aim to satisfy, which is that of preventing or limiting the emergence of norms that

impede adaptation, or to dismantle those which are already in play.

Although the present work does not deal with how institutional arrangements could aptly respond to weakly coercive rules, sustained by social sanctions, it may point to a flourishing literature on soft paternalistic policies, such as nudging. In particular, it may point to the works of theorists such as Cass Sunstein and Richard Thaler, whose aim is to design institutional provisions that help us in overcoming undesirable patterns without resorting to the threat or use of force.⁶⁹

A Second Asymmetry Objection

In the previous chapter, I have highlighted a plausible objection to the possibility of introducing my amendment to Hayek's account of the spontaneous order. In particular, it might seem that, although I argument in favour of the Hayekian account of the spontaneous order on the ground of its ability in creating the conditions for cooperation without requiring fine-grained institutions that would need local and dispersed knowledge to be designed, I imply the planner's ability to collect such knowledge in designing institutions that can prevent derivative distributions of rights that would compromise a community's instrumentalist character.

In response to such an objection, I have argued that the proposed amendment does not require fully detailed knowledge of the fellow members of the community's preferences and goals because it does not aim at achieving *specific* outcomes. In particular, following Hayek's distinction between general and fine-grained explanations, I have been trying to show that detailed knowledge of members of the community's preferences and goals is required when one attempts to design institutions aimed at specific outcomes, whereas general aims can be pursued by means of general policies (e.g. redistribution of rights, preventing kinds of transactions, etc.).

In this chapter, though, I have introduced a second amendment that could

⁶⁹In particular, see Thaler and Sunstein (2008), and Sunstein (2014).

posit further difficulties. In fact, I have suggested that centralised political authorities should design and devise institutions in order to dismantle or prevent the emergence of norms that would compromise the instrumentalist character of cooperative enterprises. This seems far more demanding than imagining a set of rules that would redistribute rights in a way that continuously allows individuals to shape their social environment according to their structure of preferences and goals. In fact, informal institutions such as social norms emerge from a complex interactions between people's beliefs, preferences, behaviours, environment and formal institutions. Hence, one may plausibly object that preventing their emergence requires the fine-grained knowledge that I deemed as unnecessary in the previous chapter.

Once again, I wish to start responding to this objection by suggesting that complex problems do not always require complicated solutions. In fact, the more our solutions are complicated, the more they are sensitive to our mistakes in diagnosing, in a fine-grained way, what the problem is. However, one may suggest that the further layer of complexity embedded in social norms precludes the possibility of simple solutions, and that tackling norms unequivocally requires fully detailed knowledge of all factors involved.

In this respect, let me offer the following example: suppose Betty owns a very large firm, with thousands of employees. Over the past few years she has developed an IT solution that allows them to work remotely in order to improve their work-life balance. In fact, many employees have complained about the long working hours in the office and how this aspect dramatically affects their families. However, she soon realises that, despite her efforts, a large amount of employees still maintain the same working patterns whilst complaining about it.

Betty's employees do not suffer from bipolar personality disorder. In fact, they are simply subject to a social norm that pervades their workplace: they observe other people staying long hours in the office, and they think that other people think they should do it as well. The persistence of such a norm could be driven by all sorts of factors: for instance, some older employees may informally

sanction colleagues who work remotely because they used to spend long hours in the office in their whole career and wish to perpetrate the same ordeal to younger employees, or, perhaps, some employees might think that, by following older colleagues working patterns, they will more likely progress in their career paths. However, Betty is faced with the problem of improving work-life balance in her firm, and such a problem cannot be merely resolved by means of creating the opportunity to work remotely. In fact, creating such an opportunity has changed very little in terms of working patterns.

Here, once again, Betty can tackle the problem in two very different ways: she could either try to collect fine-grained information from her employees (e.g. by asking why they continue to spend such long hours, who sanctions them, whether they think their career path is affected by working remotely, etc.) and tailor her solution to such information (e.g. by approaching sanctioners, or by reassuring employees about their career paths), or try to find a solution that would tackle the norm in a very general way: for instance, she could set as a default position for employees to work remotely, and suggest that their presence in the office is required only for important meetings.

Changing the default position has a number of advantages. In fact, it allows Betty to change both empirical and normative expectations at the very same time. More importantly, though, it does not require fine-grained knowledge of all the factors involved in the perpetuation of the norm. In fact, whether the norm is sustained by peer pressure, or by beliefs surrounding career paths, the updated default position is able to dismantle it regardless of the factors involved.

This example, I suggest, clarifies that although norms are, indeed, complex matter, and that diagnosing their causes is often a prohibitive enterprise, changing them does not necessarily require fine-grained knowledge, unless one wishes to achieve specific outcomes.⁷⁰

However, there is a further difficulty with norms that must be highlighted. Namely, while derivative distributions of rights, that would compromise the in-

⁷⁰Other examples of norms' diagnosis, measurement and change are nicely offered by Bicchieri (2016).

strumentalist character of a cooperative enterprise, can be tackled by reference to people's general preference for keeping their ability to shape their social environment according to their structure of preferences and goals, social norms cannot be addressed in such a generic fashion because they apply to subgroups of communities. As such, even though they may not require fine-grained knowledge of people's structures of preferences, they do require a certain in-depth investigation in order to diagnose them.

Although such a difficulty makes changing or dismantling norms much more challenging than tackling formal institutions, there is no principled epistemological argument, from an Hayekian standpoint, that would warrant the impossibility of addressing them. In fact, Hayek's epistemological standpoint addresses the impossibility to provide fully detailed explanations for complex problems, and, as a consequence, to tackle them in fine-grained ways.

Conclusion

Concluding let us summarise the main points that I have addressed within the present chapter. First, I outlined three main ways in which social norms shape our social orders: norms regulating non-public behaviours, norms defining the substantive contents of rights underpinned by formal laws, and norms shaping our compliance with formal institutions.

Second, I illustrated how norms peculiarly act on our choice scenarios. In particular, we have seen that informal sanctions increase the costs for pursuing certain courses of actions, possibly altering our preference ranking over the available options.

Third, we have illustrated how social norms may significantly affect our ability to shape the communities we inhabit, preventing us from conveying information that would allow us to define specific terms of cooperation that meaningfully respond to our structures of preferences and goals. This particular problem, we have seen, calls for a further desideratum on theories of the spontaneous order, which is that of preventing the emergence of norms that would dramatically

limit our ability to shape our social environment.

Therefore, we have set two main desiderata that a sound theory of the spontaneous order should aim to satisfy:

- (1) To prevent the emergence of distributions of rights under which some members of the community are prevented from exercising their social environment shaping rights.
- (2) To prevent the emergence of social norms that would prevent some members of the community from exercising their social environment shaping rights.

In the next chapter, I shall attempt to show that a theory of the spontaneous order that meets these two desiderata could plausibly be justified in a contractarian fashion. In fact, it requires a rather minimal agreement on institutional arrangements that leave in the hand of members of the community the ability to shape their specific terms of cooperation.

Preferences For Flexibility and The Spontaneous Order

Introduction

In the previous chapters, I offered one main critique of Hayek's account of the spontaneous order. Particularly, I have attempted to show that derivative distributions of rights, emerging from individual transactions, and social norms, arising from individual interactions, may generate states of affairs in which some members of the community lose their ability to shape their social environment by making it adapt to their structure of preferences and goals, thus undermining the instrumentalist character of the community.

These problems seem to call for amendments to Hayek's proposed institutional arrangements which would allow for preserving social orders' ability to be continuously shaped by members of the community, in ways that meaningfully respond to their structures of preferences and goals, in an effort to maintain our communities as cooperative enterprises from which everyone stands to gain. In particular, they seem to call for two main desiderata:

- (1) To prevent the emergence of distributions of rights under which some members of the community are prevented from exercising their social environment shaping rights.
- (2) To prevent the emergence of social norms that would prevent some members of the community from exercising their social environment shaping rights.

However, such desiderata seem to call for further justifications, as transactions carried out within our communities are fully voluntary. In fact, one may object that derivative distributions of rights bringing about states of affairs in which some members are prevented from making their social environment adapt to their structure of preferences and goals are emergently justified, in that they are the expression of members' consent.

In the first part of the chapter, I deal with such an objection in the attempt to show that emergent justification cannot play any plausible justificatory role with regard to unintended derivative distributions of rights brought about by fully voluntary interactions. In fact, I shall argue, the justificatory power of consent is limited to particular transactions and their foreseeable results.

In the second part, I attempt to demonstrate that the two desiderata can be justified in a contractarian fashion. Particularly, I shall attempt to show that members of the community, who are uncertain about their future structures of preferences and goals, could form preferences for flexibility, which are preferences for institutional arrangements that do not lay out, *ex-ante*, their specific terms of cooperation. Rather, they would have reasons to endorse institutional arrangements that leave in their hands the ability to continuously renegotiate their terms of cooperation according to changes in their structure of preferences and goals.

Preferences for flexibility are construed entirely in an instrumentalist fashion. In fact, I shall attempt to show, members of the community have instrumental reasons to make the sort of abstractions required in order to exhibit preferences for flexibility. My argument, in a nutshell, is that the dynamic character of our structure of preferences and goals calls for abstracting from our current set of specific goals when making choices that have long and wide-ranging effects on our lives.

Social Environment Shaping Rights

In the previous chapters, I have framed Hayek's Lockean account of the spontaneous order as embedding the constitutive aim of creating the best conditions for all members of the community to pursue their own goals, on the basis of their own knowledge. In this regard, Hayek's proposed institutional arrangements are supposed to sustain compliance with a community's rules in an entirely instrumentalist fashion. In fact, members of the community are allegedly made able to shape their specific terms of cooperation, through the exercise of their

social environment shaping rights protected by the set of rights underpinned by the Lockean triad, and thus able to shape their community as cooperative enterprises from which they stand to gain.

The instrumentalist character is not supposed to play any justificatory role. In fact, it cannot either.⁷¹ Its relevance stems from the fact that consequentialist theories, not merely aiming at defining a standard of rightness, leave us wanting of an outline of basic institutional arrangements and of convincing explanations for how compliance with such arrangements is to be sustained. Particularly, explanations of this sort should focus on what kinds of reasons members of a community possess to comply with a certain system of rules.

In this regard, the two critiques that I have proposed within chapters 2 and 3 tackle precisely the ability of the Lockean account of the spontaneous order to sustain compliance with its rules in an instrumentalist fashion.

These two critiques, we have seen, focus on specific kinds of rights, to which I referred as social environment shaping rights. The peculiarity of these rights rests on the role they play in equipping us with the ability to convey certain information to our fellow members of our community. Particularly, by exercising these rights we are able to let others know on what terms we are willing to cooperate. Consider the example concerning economic transactions which I have previously offered. In an institutional context in which economic rights are widely protected, prices emerge out of a large amount of voluntary interactions among economic actors. Whenever we are in the market for an iPad worth £800, by conveying information to other market players, through the price system, we shape market players' strategic choices, whose aim is to respond to our preferences and goals.

To provide a more intuitive example that captures the social environment shaping power of actions protected by economic rights, let us think about crowdfunding platforms such as Kickstarter or Indiegogo. There, prospective manufacturers present projects that potential customers are asked to back up, through pledges. Specifically, on the one hand of the platform, prospective manufactur-

⁷¹See, in particular, Piper (1986).

ers show demo versions of their products, illustrating their capabilities; on the other, customers are able to evaluate them and, eventually, pledge to buy the item when available. Such a platform is cooperation enhancing in many ways. In fact, it allows manufacturers to know whether their products meet customers' needs and desires, and whether the proposed price point generates enough demand in order to sustain mass production of a particular item. More relevantly to our discussion, platforms of this kind provide customers with a certain ability to shape the market, as their pledges determine a product's entry on the market, its pricing point and the quantity produced.

The social environment shaping feature of certain rights, I have claimed, plays a crucial role in sustaining a community's instrumentalist character. This is due, I suggest, to the dynamism of our structure of preferences and goals. In order to illustrate this particular aspect, let us turn back to the Dore vs Bakewell analogy, which I proposed within the first chapter. Recall that Dore is a relatively simple social order, where members of the community have relatively stable preferences and goals such that eventual changes may be perfectly predicted by the planner. Such predictability makes the social environment shaping features of rights ultimately useless, as members of the community do not need to convey information about their terms of cooperation. In fact, the planner is already able to collect preferences and goals and to design institutional arrangements which reflect people's specific terms for cooperating. In Bakewell, on the other hand, such a possibility is precluded. In fact, we have seen that the dynamic structure of preferences and goals of the members of the community makes the planner unable to design and devise systems of rules that reflect people's terms for cooperating. In such a scenario, cooperation depends on the ability of members of the community to convey information to the other members about their specific terms for cooperation, and such an ability rests on the exercise of social environment shaping rights.

Unfortunately, there is no meaningful way to identify the set of social environment shaping rights in more detail. This is because, I suggest, most rights allegedly protect actions that could potentially serve as a means to convey infor-

mation about one's terms for cooperation, but also protect actions which would not really serve the purpose of shaping our social environment. For instance, aside from examples concerning economic rights, we have seen that our right to free speech may be used to express and convey information which are needed to find common grounds for cooperating; at the same time, our discourses may be already happening in a context of common grounds, and thus lacking the social environment shaping aspect.

Often, when we act within the boundaries of the protected sphere defined by Hayek's proposed institutional arrangements, we perform social environment shaping actions, without really aiming to. In fact, when we are in the market for an iPad, we are not enticed by our ability to affect market players' strategic choices, nor we care about it. In this regard, most social environment shaping actions are undertaken in a rather unintentional way.

Such an unintentional character of the exercise of social environment shaping rights helps us in clarifying the rationale of the two critiques I have previously offered. In fact, members of the community exercising these rights may not perceive the role that such actions play in conveying information about their terms of cooperation, and this explains why members of a community may be keen on selling rights or portions of these rights in exchange for some other goods or rights that do not grant the same social environment shaping feature. It may also explain why they may waive the exercise of these rights on the ground of informal sanctions that they perceive as too burdensome. Moreover, since the boundaries of our inability to convey information about our terms of cooperation are ultimately blurry, members of the community may be unable to discern when waiving or selling a particular right would bring about one's inability to shape their specific terms of cooperation in ways that meaningfully respond to our structure of preferences and goals.

Emergent Justifications

The role that social environment shaping rights play or should play in theories of the spontaneous order is of profound relevance. This is due to the particular emphasis that such theories posit on adaptations to rapid change in the environment.⁷² The importance of adaptation is also implicit in the search for institutional arrangements that possess an instrumentalist character. In fact, we have seen that when institutional arrangements are unable to respond to changes in our structure of preferences and goals, we may not find profitable to continue playing our cooperative strategy. Particularly, if the chances of satisfying our own goals are not enhanced, or are reduced, by playing the cooperative strategy, a community loses its instrumentalist character, generating grounds for defection and resulting instability.

Within the previous chapters, we have seen that states of affairs in which some members of the community are prevented from shaping their social environment, thus undermining a community's instrumentalist character, can be brought about unintentionally. Particularly, they may emerge as large-scale consequences of individual transactions and interactions. In this regard, I have proposed to incorporate limitations to our transactions involving social environment shaping rights in order to prevent such states of affairs from emerging.

One, however, may object that distributions of rights arising from voluntary transactions require stronger justificatory grounds to be either prevented or amended. In fact, they are emergently justified, by virtue of arising through actual consent. Particularly, according to this objection, distributions of rights, brought about by individual transactions to which members of the community consent to, are, in themselves, expression of actual consent.

I take such an objection to carry two different implications: on the one hand,

⁷²In this regard, it is worth re-quoting Hayek (1945): "If we can agree that the economic problem of society is mainly one of rapid adaptation to changes in the particular circumstances of time and place, it would seem to follow that the ultimate decisions must be left to the people who are familiar with these circumstances, who know directly of the relevant changes and of the resources immediately available to meet them. We cannot expect that this problem will be solved by first communicating all this knowledge to a central board which, after integrating all knowledge, issues its orders", p. 524.

actual consent within individual transactions yields a strong justificatory power which transfers to distributions of rights arising from such transactions; on the other, derivative distributions of rights spontaneously emerge from individual transactions that members of the community want to bring about; as such, we should not worry about the lack of instrumentalist character within emerging states of affairs, as they are the result of transactions that members of the community wish to bring about, and it is unlikely that they would intentionally commit to distributions of rights that would prevent them from shaping their specific terms of cooperation.

In this regard, let me start by considering what emergent justifications of institutional arrangements may look like. Particularly, let me start by illustrating a popular version of the concept of emergent justification that we can extract from Nozick (1974):

It is not clear how those holding alternative conceptions of distributive justice can reject the entitlement conception of justice in holdings. For suppose a distribution favoured by one of these non-entitlement conceptions is realized. Let us suppose it is your favorite one and let us call this distribution D1; perhaps everyone has an equal share, perhaps shares vary in accordance with some dimension you treasure. Now suppose that Wilt Chamberlain is greatly in demand by basketball teams, being a great gate attraction. (Also suppose contracts run for a year, with players being free agents.) He signs the following sort of contract with a team: In each home game, twenty-five cents from the price of each ticket of admission goes to him. (We ignore the question of whether he is "gouging" the owners, letting them look for themselves.) The season starts, and people cheerfully attend his team's games; they buy their tickets, each time dropping a separate twenty-five cents of their admission price into a special box with Chamberlain's name on it. They are excited about seeing him play; it is worth the total admission price to

them. Let us suppose that in one season one million persons attend his home games, and Wilt Chamberlain winds up with \$250,000, a much larger sum than the average income and larger even than anyone else has. Is he entitled to this income? Is this new distribution D2 unjust? If so, why? [...] If D1 was a just distribution, and people voluntarily moved from it to D2, transferring parts of their shares they were given under D1 (what was it for if not to do something with?), isn't D2 also just? (Nozick, 1974, pp. 160-161).

Essentially, Nozick's argument can be read as claiming that, given an initial distribution of rights that members of a community consider as just, derivative distributions emerging from fully just transactions will also be just. In fact, derivative distributions of rights are said to be emergently justified, in that they emerge from an initial just distribution, through transactions possessing a certain property P (i. e. being fully just). As a consequence, any attempt to keep or restore the initial distribution would be either preventing just transactions or requiring further transactions which may not be fully just.

In Chamberlain's story, just transactions are fully voluntary transactions. This is to say that derivative distributions of rights arising from transactions are emergently justified as long as there is actual consent from the parties involved in such transactions. In this regard, to keep the initial distribution would require preventing members of a community from consenting to particular transactions, and restoring the initial distribution would command further transactions to which members of a community may not be willing to consent. David Schmidtz nicely captures the strong justificatory force yielded by consent:

Actual consent has justificatory force over and beyond the teleological force of the reasons people have for consenting. Freely given consent is intrinsically a kind of authorization; by consenting, one gives others a right to expect from oneself that which one has consented to do, to give, or whatever.' Nor is actual consent particularly rare or difficult to secure within a range of typical human endeavors.

To give an example not directly relevant to the creation of governments, we observe consent on a small scale whenever we observe an ordinary exchange of goods between two people. What do we ordinarily think of as justifying such exchanges? There are two answers. We could argue that the exchange's results further the participants' goals (better than their alternatives). For epistemic reasons if nothing else, however, we usually are more inclined to focus on whether the process of negotiation and exchange is unforced, not fraudulent, and so on. In other words, when the process accords with these and any other constraints applying to it, it fully realizes the justificatory force latent in actual consent. The first approach is teleological, looking to the exchange's outcome. The second is emergent, looking for compliance with constraints on the process by which the outcome arises. (Schmidtz, 1990, pp. 97-98).

According to Schmidtz, actual consent seems to yield two main justificatory features. First, there is a general presumption that fully voluntary transactions are mutually advantageous, in that rational persons who engage in a transaction do so in order to enhance the chances to satisfy their goals; second, outcomes of these transactions are obtained by complying with the constraints expressed by a certain property P, which is the property of being fully voluntary.

However, Schmidtz notices, there is one main question arising from emergent justifications of distributions of rights emerging from actual consent. Namely, to what extent the strong justificatory force that actual consent carries within individual transactions is transferable to large-scale distributions of rights emerging from such specific transactions.⁷³ In fact, large-scale distribution of rights, emerge from what Schmidtz refers to as an invisible hand processes, where individual consent is limited to particular transactions, and cannot possibly entail

⁷³See Schmidtz (1990): "The problem is that what people consent to are individual transactions, rather than to the order that spontaneously emerges from them. In other words, that an outcome arose by consent does not entail that people consented to it. (Analogously, people are willingly doing what produces the greenhouse effect, but that does not mean they are consenting to its production)" p. 100.

consent to large-scale consequences of such transactions.

Particularly, following Cohen (1977), there are two main aspects to illustrate. The first is concerned with the extent to which participants to a just transaction are able to foresee the large-scale consequences on the distribution of rights that such transactions would yield. For instance, in Chamberlain's story, those who buy the tickets are driven by the desire to watch Wilt Chamberlain play, and such an outcome is brought about by the particular transaction. However, it is not clear that, by buying a ticket, we consent to something more than watching Wilt Chamberlain play. In fact, it is not obvious that derivative distributions of property rights emerging from that particular transactions are something which we are able to foresee and willing to consent to.

Second, even if participants to a just transaction are able to foresee the particular consequences that such a transaction would yield over the larger distribution of rights, such distribution is also brought about by other people's transactions over which they have no control. In fact, in Chamberlain's story, Wilt's privileged position, in terms of the distribution of property rights, arises from countless transactions over which singular participants have no bearing. In this regard, even under the assumption that each participant engages with specific transactions with the ability to predict the effects that each transaction has over the larger distribution of rights and, thus consents only to those that would produce sound outcomes, these specific instances of consent do not express consent over the distribution of rights arising from all transactions.

In fact, suppose Alf, Betty and Wilt start with an initial endowment of £1. Alf and Betty are extremely eager to watch Wilt play and willing to pay fifty cents each. Such transactions carry a strong justificatory force as Alf and Betty engage in them with the prospect of bettering their lives; moreover, both are able to foresee the particular effect on the larger distribution of rights emerging from their singular transactions and agree to it. Particularly, Alf consents to making Wilt richer of fifty cents, and the same applies to Betty. However, it is not clear that singular transactions, taking place under the assumption of Alf and Betty's ability to foresee the large-scale consequences on the distribution of

rights of each transaction, express also their consent to the distributive outcome arising from both transactions. Alf may be consenting to make Wilt richer of fifty cents but his consent is limited to the distributive outcome which that particular transaction yields. In fact, he may find a distributive outcome in which Wilt's endowment amounts to £2 unacceptable, and the same applies to Betty.

In such a scenario, the distributive outcome would be emergently justified as it arises from fully voluntary transactions. However, actual consent to singular transactions and to the outcome that singular transactions generate, does not carry any justificatory power when dealing with large scale consequences of all transactions. Clearly, Alf and Betty may be content with the overall distributive outcome of both transactions, but, if they are, they are so by accident, and not through a process of deliberation that sees them as multilaterally consenting to the outcome.

Considerations emerging from Wilt Chamberlain's example, help us in elucidating the critiques that I have offered within the previous chapters. In fact, it should be now clear that consent to particular transactions cannot imply a consent to derivative distributions of rights. Consent to small-scale transactions does not merely fail in justifying unforeseeable outcomes generated by singular transactions, but also fail in justifying outcomes generated by all sorts of transactions, over which we have no control. This, once again, helps us in explaining why members of a community who are motivated to keep their ability to convey information about their terms of cooperation may still consent to particular transactions involving social environment shaping rights, and also why states of affairs in which they lose such an ability may be brought about without their consent. We may also notice that similar implications apply to the critique offered in the third chapter. In fact, our choice not to exercise social environment shaping rights, on the ground of informal sanctions, does not, in itself, constitute an instance of consent to states of affairs generated by such a choice. Moreover, one may argue, our choice not to exercise social environment shaping rights, may be driven by informal sanctions which would make our decision to

waive the exercise of a particular right as less than fully voluntary.

This brings us to the conclusion that actual consent to small-scale transactions plays no justificatory role at all with respect to derivative distributions of rights. In fact, on the one hand, the justificatory power of actual consent cannot be transferred to derivative distributions of rights; on the other, there is no guarantee that consent to individual transactions shall generate large-scale distributive patterns that would preserve the instrumentalist character of our communities. In fact, even if members of the community share an interest in preserving it, large-scale distributive patterns cannot be said to be the result of their actual consent.

If the sort of invisible hand emergent justification Nozick seeks to provide in Chamberlain's example fails in delivering any justification to large-scale emerging distributive patterns, Schmidtz invites us to consider another alternative. Particularly, he asks to consider the possibility of contractarian versions of emergent justifications:

Consider contractarianism as a theory about how emergent justification might work. In a contractarian bargaining process, members of a large group seek a collective agreement. Consent to the agreement is taken as a sign that the agreement is mutually advantageous. It is by no means a guarantee, however. (At least, it does not guarantee ex post advantage, which is presumably what bargainers really care about.) People enter the agreement without the benefit of hindsight. Nor does actual consent presuppose rationality in the idealized way that hypothetical consent does. But actual consent carries emergent force regardless, so long as, for example, failures of foresight are not due to fraud. [...] Contractarian accounts of the state's emergence are distinguishable, at least in a rough sense, from invisible hand accounts. In contractarian models, intentional collective actions leads to an intended and mutually agreeable result. In invisible hand models, bargaining occurs among shifting and relatively

small subsets of the collective. The larger scheme of stable society evolves through a series of relatively small-scale exchanges and is an unintended result of such exchanges. There are various agreements between individuals, but there is nothing resembling an agreement to create the emerging social order. The social order emerges spontaneously. (Schmidtz, 1990, pp. 98-99).

Contractarian bargaining procedures, I shall argue, offer us a valuable framework to cast our concerns about emerging distributions of rights that may undermine the instrumentalist character of our communities. However, before I get to this point, I wish to contend, contrary to Schmidtz, that contractarian accounts too may prove ineffective in justifying larger schemes spontaneously emerging from individuals' transactions, once the set of rules bargainers agree upon is made to work. In fact, bargainers' collective action, alone, solves only one part of the problem connected to invisible hand emergent justifications. Particularly, it only responds to our worry about large-scale consequences being brought about by individual transactions which are not under everyone's control. However, it leaves open the problem of justifying unforeseeable larger schemes arising from individual interactions, allowed by the set of rules on which bargainers have converged.

In order to show this, let us consider two different bargaining solutions. In the first, Alf, Betty and Wilt converge on a certain distribution of rights; in the second, they further stipulate that certain patterns, eventually arising out from their interactions, allowed by the initial distribution of rights, will be ruled out, either by means of preventing them from engaging in certain transactions or by restoring a certain distribution of rights after such transactions occur. These two agreements differ in one main crucial respect; namely, the first solution is exclusively concerned with setting constraints on what counts as a legitimate transaction, whereas the second engages also with patterns eventually arising from such transactions. Although such a difference seems to rest merely upon the extension of the object of the agreement, it is also a difference in kind. In

fact, by looking at patterns, bargainers transform the emergent justification of the procedure into a sort of teleological justification. Particularly, they define a set of states of affairs that a certain distribution of rights may justifiably bring about, and rule out a set of illegitimate patterns.

This brings with it interesting differences with respect to the extension of the justification provided by actual consent within the bargaining procedure. Specifically, while the first solution can hardly be said to justify unintended patterns arising from Alf, Betty and Wilt's transactions, allowed by the distribution of rights they have agreed to, the second, extends its justificatory scope to states of affairs arising from such transactions. In this regard, the more the agreement of the bargaining procedure leaves the door open for derivative and unintended distributive patterns over which members of the community have no meaningful control, the less actual consent can transfer its justificatory power to states of affairs eventually arising from individual transactions. In fact, consent consistently applies to the object of the agreement, but its justificatory power does not equally apply to derivative distributive patterns over which Alf, Betty and Wilt have no control.

Contractarian accounts, I suggest, are rather peculiar justificatory devices, as they, de facto, hand to bargainers to the task of outlining the structure of justification of social orders, whether emergent or teleological. In fact, they may be looked at as two stages justificatory devices: a first stage, which defines the properties that bargaining procedures should possess (e.g. how consent is to be exercised, how bargainers come to an agreement, etc.), and which eventually determines the emergent character of the agreement, and a second stage which hands to bargainers the decision about the structure of their agreement, which could take either an emergent or a teleological form.

The problem lies in that if the structure of the agreement takes an emergent form, like in Alf, Betty and Wilt's first solution, unless patterns arising from their interactions are entirely foreseeable, the justificatory power of the agreement cannot be meaningfully transferred to such patterns. On the other hand, if patterns are entirely foreseeable, the structure of the agreement comes to take

a teleological form, for Alf, Betty and Wilt, by agreeing to a set of rules that can only bring about a certain set of patterns, also agree on which patterns can be brought about and which have to be ruled out.

In this regard, one may wonder to what extent Alf, Betty and Wilt's second solution is an instance of emergent justification. Surely, it emerges from bargainers' actual consent and, most importantly, is the result of bargainers' collective action, which answers to parts of our concerns about invisible hand accounts of emergent justification (i.e. large-scale consequences are brought about by a collective action involving all agents simultaneously, rather than emerging from individual transactions). However, what provides extended justificatory scope to their agreement is the fact that it assumes a teleological form; for if it did not, states of affairs unintentionally brought about by their interactions, allowed by the scheme of rules they have converged upon, would hardly be justified by their agreement. This, I think, carries interesting implications with respect to the reach of emergent justifications, for it shows that if we seek to justify larger schemes emerging from individual transactions, contractarian accounts, in principle, are unable to respond to our worries, unless the bargainers' agreement takes a teleological form. I find such an aspect of particular relevance for the purpose of the present work, as it shows that if we are to provide a contractarian justification of a theory of the spontaneous order, the mere agreement on a Lockean distribution of rights is unable to justify derivative distributions of rights in which some members of the community are prevented from exercising social environment shaping rights that allow them to convey information about their terms of cooperation. In fact, such a state of affairs might be unintentionally brought about through an invisible hand process, which fails to provide any substantive justification to derivative distributions of rights.

Contractarianism: Abstraction and Instrumentalism

In the previous chapters, I have highlighted the instrumentalist character that Hayek ascribes to the Lockean distribution of rights. In particular, we have seen

that the protected sphere of action these rights grant to us, allegedly allows members of a community to shape their specific terms of cooperation, and, in turn, equip them with reason to comply with the community's rules, as members see their community as an enterprise from which they stand to gain.

These aspects provide the Lockean account of the spontaneous order with a contractarian nuance. In fact, if institutional arrangements purported by the theory equip members of the community with instrumental reasons to comply with them, then such an arrangement could be said to represent a plausible solution to a bargaining procedure where agents come to define terms of cooperation. On a more general level, we could notice that all instrumentalist theories could be backed up by a contractarian justification, as they represent, at least, plausible solutions to a bargaining procedure.

Moreover, Hayek's arguments highlighting the uniquely suited position of the Lockean distribution of rights in enhancing members of the community's chances to satisfy their own goals on the basis of their local knowledge, may suggest that Hayek envisages his proposed institutional arrangements as the best solution to a contractarian bargaining procedure among agents exhibiting widely different standpoints. However, Sugden notices, contractarian theories need not be that ambitious. Particularly, he argues:

A more modest form of contractarianism is possible, which does not seek to design the institutions of a whole society according to wholly external standards of rationality. Rather, the object is to evaluate possible changes in the institutions of an existing society, using a criterion of agreement that is defined relative to the knowledge and the conventions that prevail in that society. Thus the question is not, "Is this society one that we would have chosen in some state of nature?" Rather; it is: "Given where we are, can we agree to continue to live by these existing rules; or if not, can we agree to any changes?" (Sugden, 1993, p. 421).

Sugden further claims that this less ambitious form of contractarianism also

captures Hayek's theory of the spontaneous order which he finds to represent an "agreement among individuals within an ongoing social order, rather than among rational agents who stand outside any particular society". (Sugden, 1993, p. 393).

One way to articulate Sugden's claim is to argue that Hayek's theory of the spontaneous order is not captured by a one-shot bargaining procedure where members of a community get to agree on institutional arrangements but rather by hypothetical and infinitely repeated bargaining procedures taking place within an initial position that resembles members of the community's actual position and from which they continuously negotiate better terms of cooperation. According to this view, hypothetical agreements should be seen as sequential coordination games in which members of the community identify changes to their terms of cooperation that would represent Pareto-improvements and decide to converge on them. The problem with such an interpretation is that changes to members of the community's terms of cooperation, within the spontaneous order, do not really take place through bargaining procedures, but rather through an invisible hand process shaped by the individual exercise of social environment shaping rights (e.g. right to transact, rights of expression). States of affairs arising from invisible hand processes, then, can hardly be seen as solutions to coordination games as members of the community do not really get to agree on them.

In this regard, another way to cast Sugden's point is to argue that contractarian agreements on institutional arrangements need not to be backed up by external standards of rationality, and that, as such, a contractarian justification of Hayek's theory of the spontaneous order could be provided without resorting to idealised bargainers that stand outside the ongoing social order. According to such a view, actual members of a community, from their actual standpoints, could potentially converge on the institutional arrangements purported by Hayek's theory of the spontaneous order.

The approach I have just highlighted is the one that I wish to follow in arguing in favour of an amended version of the Lockean account of the spon-

taneous order, which includes the two desiderata that I have outlined within the previous chapters. Before I get to this point, however, I need to clarify a few aspects. Particularly, I need to articulate more on the distinction between approaches to social contract embodying external standards of rationality and on why, alongside with Sugden, I take external standards of rationality as ultimately unable to serve the purpose of justifying institutional arrangements that possess an instrumentalist character.

In this regard, let me start by briefly outlining the idea of employing external standards of rationality within social contract theories. In order to do this, I suggest, we may start by looking at the ways in which theorists model agents within bargaining procedures, and, in particular, at the ways in which they model agents' moral motivations. This aspect, indeed, constitutes the turning point that allows us to distinguish between Kantian approaches, such as those undertaken by Thomas Scanlon and John Rawls, and Hobbesian accounts, developed by theorists such as David Gauthier and Jan Narveson.⁷⁴

The Kantian tradition, which is usually referred to as contractualism, frames social contract agreements as taking place among agents who find themselves already in a moral domain. Particularly, contractualist theorists follow two main routes: the first consists in imagining bargaining scenarios in which agents have already internalised a certain moral principle (e. g. the principle of moral equality), the second consists in designing bargaining scenarios in which agents are required to abstract from a range of information specific to their own identities, such as to assume a standpoint that is ultimately different from the one they would exhibit in light of their own particular circumstances.

The first route, within the Kantian tradition, is the one developed by Scanlon, whose contractualist account is based on the concept of reasonableness, rather than mere instrumental rationality, where reasonableness is meant to capture our respect for others as morally free and equal persons. Particularly, Scanlon outlines a framework of justification according to which rightness and

⁷⁴In particular, I have in mind Scanlon (1998) and Rawls (1971) on the one hand, and Gauthier (1986) and Narveson (1988) and (2001), on the other.

wrongness of principles are grounded on the concept of reasonable rejectability:

In the contractualist analysis of right and wrong, what is presupposed first and foremost is the aim of finding principles that others who share this aim could not reasonably reject. This aim then brings other reasons in its train. Given this aim, for example, it would be unreasonable to give the interests of others no weight in deciding which principles to accept. (Scanlon, 1998, p. 192).

Scanlon's principle of reasonable rejectability, as a tool to adjudicate the soundness of other principles, follows one's internalisation of the principle of moral equality. In particular, Scanlon suggests that we all, qua member of humanity, should recognise that other agents possess the same moral worth and that, as such, our deliberation procedures in assessing rightness and wrongness of principles, must capture the fact of our equal moral worth. The device of reasonable rejectability, thus emerges from our search for justificatory devices that best capture the fact of equal moral worth.

Such a contractualist framework of justification, thus presupposes that agents, who come to a bargaining procedure already care about formulating principles that other could not reasonably reject. And agents care precisely because they have internalised the principle of equal moral worth, which, in turn, would equip them with instrumental reasons to choose the device of reasonable rejectability in assessing rightness and wrongness of other principles.⁷⁵

The obvious limit with Scanlon's account consists in that the internalisation of the principle of moral worth, which should provide bargainers with instrumental reasons to care about other people's standpoints, does not obtain through instrumental reasons. In fact, agents' ability to internalise the principle of moral worth is contingent on three main aspects: first, on our ability to grasp it; second, on our ability to identify the reason-giving force of moral judgments; third,

⁷⁵It is important to notice that Scanlon is not really concerned with social contract bargaining procedures, nor his form of contractualism is directed at generating decisions on institutional arrangements. In fact, the theory is mostly concerned with moral principles rather than political decisions. However, the ways in which he frames the contractualist framework is helpful in illustrating the concept of external standards of rationality.

on our ability to see the priority of the reason-giving force of moral demands over other considerations (e.g. one's reason to promote his well-being).

Therefore, if agents fail in one of those enterprises, they enter the bargaining scenario without instrumental reasons for endorsing the principle of reasonable rejectability as the main adjudicatory device in assessing the rightness or wrongness of other principles. If they do, the outcome of the bargaining procedure is ultimately indeterminate, as agents would disagree about the optimal adjudicatory device and, in turn, on which principles would be ultimately right and wrong.

If, on the other hand, bargainers are demanded to assess the rightness and wrongness of principles through the device of reasonable rejectability, despite not having internalised the principle of moral worth, their choices will lack normative force once the bargaining procedure is over. In fact, agents will lack instrumental reasons to comply with principles chosen through the device of reasonable rejectability precisely because they fail to internalise the principle of moral worth.

In this regard, if one is to assess Scanlon's theory from an instrumentalist standpoint, one is faced with a dilemma: in fact, either the outcome of the bargaining procedure is indeterminate, in that bargainers strongly disagree on the appropriateness of the device of reasonable rejectability in assessing the rightness and wrongness of principles, insofar as they disagree on the normative relevance of the principle of moral worth, or principles chosen through the device of reasonable rejectability will lack normative force once the procedure is over, as agents would fail to form instrumental reasons to comply with them.

The second path, within the contractualist tradition, is the one traced by Rawls, whose account of the original position surely represents the most influential design of a social contract bargaining procedure. Rawls asks us to imagine a hypothetical situation in which agents come to agree on a set of rules or principles defining their terms of cooperation.⁷⁶ The peculiarity of the original

⁷⁶Particularly, Rawls writes: "It is clear, that the original position is a purely hypothetical situation. Nothing resembling it need ever take place, although we can by deliberately

position lies in that agents are required to abstract from a range of information specific to their own identity and their position within their existing society. More specifically:

No one knows his place in society, his class position or social status; nor does he know his fortune in the distribution of natural assets and abilities, his intelligence and strength, and the like. Nor, again, does anyone know his conception of the good, the particulars of his rational plan of life, or even the special features of his psychology such as his aversion to risk or liability to optimism or pessimism. More than this, I assume that the parties do not know the particular circumstance of their own society. That is, they do not know its economic or political situation, or the level of civilization and culture it has been able to achieve. The persons in the original position have no information as to which generation they belong. (Rawls, 1999, p. 118).

Agents within the original position find themselves behind a veil of ignorance that shapes their instrumental reasons for converging on the two principles of justice Rawls' highlights.⁷⁷ Particularly, although Rawls models agents' psychology as mutually disinterested, the exclusion of a range of information concerning agents' own identities and positions within existing societies from the choice scenario, deliver us a bargaining solution that captures an account of justice as fairness. In fact, Rawls clarifies:

The original position is defined in such a way that it is a status quo in which any agreements reached are fair. It is a state of affairs in which the parties are equally represented as moral persons and the outcome is not conditioned by arbitrary contingencies or the relative balance of social forces. Thus justice as fairness is able to use the

following the constraints it expresses simulate the reflection of the parties" Rawls (1999, p. 104).

⁷⁷Rawls (1999), p. 52.

idea of pure procedural justice from the beginning. (Rawls, 1999, p. 104).

The main benefit coming from Rawls' original position, compared to Scanlon's account, consist in that it aptly responds to problems connected to moral and political disagreement within the bargaining procedure. In fact, while Scanlon's account leaves the door open for disagreement insofar as bargainers may fail to internalise the principle of moral equality, which in turn equips them with instrumental reasons to choose the principle of reasonable rejectability as the main justificatory device, Rawls, by making bargainers abstract from arbitrary contingencies, is able to frame a bargaining scenario in which agents exhibit similar standpoints.

All agents, regardless of their moral motivations, stemming from arbitrary contingencies, will be able to converge on the two principles of justice once the veil of ignorance is made to work. In fact, through abstraction, all bargainers come to possess the very same structure of preferences and goals and their agreement is, thus, secured.

Therefore, differently from Scanlon, Rawls seems not to presuppose that agents have internalised the principle of moral equality once the veil of ignorance is made to work. In fact, parties to the bargaining procedure are supposed to 'discover' the principles of justice through the bargaining procedure. Convergence on the two principles of justice he envisages, indeed, does not allegedly emerge from a moral commitment to the idea of moral equality, but rather constitutes a Nash equilibrium reached by parties who are mutually disinterested.

However, abstraction comes at a cost. In particular, there are two main worries linked to employing abstractions in modeling agents within a bargaining scenario: first, there is a concern with respect to the doubly hypothetical character of the outcome; second, there is a worry about the meaningfulness of choices made by abstracting from morally arbitrary features. The first worry, introduced by Dworkin (1975), suggests that the result of a hypothetical bargaining procedure, among hypothetical agents, could hardly be said to be bind-

ing. In fact, there is no reason to suppose that we would assign any normative force to a contract hypothetically signed by other agents, on our behalf, without our authorisation.⁷⁸ The second worry, on the other hand, questions agents' ability to make meaningful choices behind thick veils of ignorance such as the one Rawls introduces in framing his bargaining scenario. This latter objection is twofold: on the one hand, it is claimed, persons are unable to entirely abstract from some of their peculiar features and, thus, unable to come to the bargaining procedure entirely untied from their conception of the good; on the other, even if agents could succeed in abstracting from morally arbitrary features, the choices they make would lack meaningfulness, in that they would leave aside considerations that are incredibly relevant to their choice about social institutions. Particularly, Sandel argues:

The original position achieves too much detachment from human circumstances, that the initial situation it describes is too abstract to yield the principles Rawls says it would, or for that matter, any determinate principles at all. Such an objection would most likely take issue with the veil of ignorance on the grounds that it excludes morally relevant information, information necessary to generate any meaningful results. It would argue that the notion of the person embedded in the original position is too formal and abstract, too

⁷⁸Particularly, Dworkin claims that an "hypothetical contract is not simply a pale form of an actual contract; it is no contract at all" (Dworkin, 1975, p. 18). However, it is important to notice that such a critique partially misses the target with respect to Rawls' intent with the original position. In fact, as Freeman argues, the original position is not intended to secure an agreement that binds us, but rather to make us reflect on fair terms of cooperation among free and equal persons:"The original position (like social contract doctrine generally) is not intended to bind us to any promise we might make. Rather its purpose is to clarify what we now think and are committed to believe, assuming that we conceive of ourselves as free and equal citizens and have a willingness to cooperate with others as equals on grounds of mutual respect. The conception of justice the parties would adopt identifies, Rawls believes, the conception of justice that we regard - here and now - as fair and supported by the best reasons (PL, 26). Rawls presumes that as members of a democratic society we in fact think of ourselves as free and equal citizens, and that we are drawn to affirm the idea of a well-ordered society as an ideal of fair social cooperation among free and equal people like ourselves. These are "ideas of practical reason" implicit in the political consciousness of members of a democratic society (see PL, 107, 110). The point of a social contract in the original position is to clarify the implications of these ideas and democratic ideals we endorse". (Freeman, 2007, p. 344.

detached from contingency to account for the requisite motivations. Where the first objection complains that the thin theory of the good too thick to be fair, the second contends that the veil of ignorance is too opaque to yield a determinate solution. (Sandel, 1982, pp. 27-28).

Sandel's claim is of particular interest as it seems to suggest that either there is no place for rational decision making behind thick veils of ignorance, or that rational decision making would lead to results that lack meaning, as bargainers lack orientation from the sort of things that mostly matter to their lives: namely, their conception of the good, alongside with the goals that it makes salient to them. In order to capture Sandel's critique, we may imagine a choice scenario in which are to pick between two alternative routes but we lack an ultimate purpose to reach either destinations. Surely, there may be objective grounds to evaluate the two options. For instance, one route may be shorter than the other, or presenting less obstacles; however, such objective assessments would still lack meaningful guidance with respect to our destination.

Rawls has an answer to Sandel's point, which takes the form of a list of goods, the primary goods, that rational persons are supposed to want whatever their goals are.⁷⁹ The list is meant to provide guidance to members of a community in choosing the basic structure of their cooperative enterprise.⁸⁰

⁷⁹Particularly, Rawls specifies the list as follows: "First, the basic liberties as given by a list, for example: freedom of thought and liberty of conscience; freedom of association; and the freedom defined by the liberty and integrity of the person, as well as by the rule of law; and finally the political liberties. Second, freedom of movement and choice of occupation against a background of diverse opportunities; Third, powers and prerogatives of offices and position of responsibility, particularly those in the main political and economic institutions; Fourth, income and wealth; and Finally, the social basis of self respect". (Rawls, 1982, p. 362).

⁸⁰In particular, he suggests: "Therefore how can they reach a sensible decision [on the principles of justice]? One might reply that the rationality of a person's choice does not depend upon how much he knows, but only upon how well he reasons from whatever information he has, however incomplete. Our decision is perfectly rational provided that we face up to our circumstances and do the best we can. Thus the parties can in fact make a rational decision, and surely some of the alternative conceptions of justice are better than others. Nevertheless, the thin theory of the good which the parties are assumed to accept shows that they should try to secure their liberty and self-respect, and that, in order to advance their aims, whatever these are, they normally require more rather than less of the other primary goods. In entering into the original agreement, then, the parties suppose that their conceptions of the good have a certain structure, and this is sufficient to enable them to choose principles on a rational basis". (Rawls, 1999, p. 349).

However, one might suggest that Sandel's objection is somehow robust to Rawls' response. In fact, whether or not Rawls' list of primary goods equips bargainers with sufficient information to converge on a set of principles of justice, once the bargaining solution is achieved, agents start to evaluate their choice under the light of their own actual circumstances. In particular, they evaluate principles of justice in the light of their own actual conception of the good, of their own talents, of their own relative position in society, and may question the meaningfulness of their choices taken behind the veil of ignorance.

For instance, suppose bargainers are to deliberate about how to distribute property rights on a currently unowned piece of land. When bargainers abstract away from arbitrary contingencies, their choice is guided by Rawls' list of primary goods. Particularly, they are to rationally deliberate on a distribution of property rights of the land that maximises their fair share of primary goods. Once the decision is taken, though, Betty realises that the distributive principle she picked behind the veil of ignorance does not reflect her actual instrumental reasons. Suppose, indeed, she has personal motives for claiming the property of a particular piece of that land, which has been assigned to another member of the community. Particularly, suppose she has sentimental attachment to it or that she has a personal project that can only take place in that particular part of the land. These personal motives are exactly the sort of arbitrary contingencies that the veil of ignorance requires us to strip away. Yet, they are an important source of instrumental reasons for deliberation. Betty, indeed, may perceive that instrumental reasons that apply to her behind the veil of ignorance, differ from instrumental reasons that apply to her under the light of arbitrary contingencies. In fact, were she able to take into account her sentimental attachments or her personal projects, she would have picked a different distributive principle.

Rawls attempts to respond to this problem by resorting to idea of strains of commitment to the principles chosen behind the veil of ignorance. Particularly, he argues that the requirement of strict compliance is already embedded within the bargaining scenario. In fact, he claims:

Agents are rational in that they will not enter into agreements they know they cannot keep, or can do so only with great difficulty. Along with other considerations, they count the strains of commitment. Thus, in assessing conceptions of justice the persons are to assume that the one they adopt will be strictly complied with. (Rawls, 1999, p. 129).

Parties to the bargaining procedure are thus supposed to work out the basic structure of their terms of cooperation so as to assure that they converge on principles that shall generate universal support, such that the agreement is not reached in vain.

However, one might wonder how are agents to evaluate their strains of commitment given that they do not have access to the arbitrary contingencies that give rise to their instrumental reasons. In this regard, Rawls argues that strains of commitment do not come from evaluating the two principles under the light of arbitrary contingencies, for arbitrary contingencies are irrelevant from the standpoint of justice. Particularly, he claims that the aim of the veil of ignorance is 'to rule out those principles that it would be rational to propose for acceptance, however little the chance of success, only if one knew certain things that are irrelevant from the standpoint of justice' (Rawls, 1999, p. 17).

This answer, though, poses obvious limits if one is to assess Rawls' theory from an instrumentalist standpoint. In fact, if agents are to evaluate their commitment to the principles of justice, from the standpoint of justice, the veil of ignorance becomes morally loaded, for one needs to internalise the demands of justice in order to enter the bargaining procedure and to place herself behind the veil of ignorance. In fact, one may internalise different demands and object that the level of abstraction Rawls demands to bargainers is ultimately unwarranted.

In this regard, the advantages of Rawls' original position over Scanlon's contractualism are only apparent. For the sort of moral and political disagreement that the original position is able to accommodate is reintroduced in the form of disagreement about the demands of justice one is to internalise in order to

place herself behind the veil of ignorance. More specifically, people who internalise different demands of justice, shaped by one's own conception of the good, will propose different bargaining scenarios, incorporating different levels of abstraction, or perhaps no abstraction at all.

This poses a sort of trilemma: in fact, if bargainers are forced to abstract, they may find hard to pick principles they will strictly comply with, under the light of instrumental reasons provided by arbitrary contingencies (as they do not have access to them); if bargainers are to evaluate their strains of commitment under the light of instrumental reasons provided by arbitrary contingencies, they may find it hard to agree on a common set of principles; if abstraction is undertaken on a voluntary basis, bargainers, who fail to internalise Rawls' intuitions about justice as fairness, shall disagree on the level of abstraction required by the bargaining procedure about the basic structure of their terms of cooperation.

There is, indeed, an important trade-off between the thickness of the veil of ignorance we posit on bargainers, in modelling social contract bargaining scenarios, and the normative relevance of the outcome of the bargaining procedure to actual persons. Particularly, the more we demand actual members of the community to abstract from their arbitrary contingencies, the more we design impersonal reasons that guide their choice in deliberating about the basic structure of our society, either in the form of primary goods or by assuming their internalisation of the principle of moral equality, the less the outcome of the bargaining procedure will carry normative relevance to them, from their own actual standpoint.

It is this concern toward thick veils of ignorance that drives theorists in the Hobbesian tradition, such as Gauthier and Narveson, to develop contractarian accounts that attempt to ground the possibility of an agreement on instrumental rationality. Particularly, Gauthier attempts to offer:

A contractarian rationale for distinguishing what one may and may not do. Moral principles are introduced as the object of fully volun-

tary ex ante agreement among rational persons. Such an agreement is hypothetical, in supposing a pre-moral context for the adoption of moral rules and practices. But the parties to agreement are real, determinate individuals, distinguished by their capacities, situations, and concerns. In so far as they would agree to constrain on their choices, restraining the pursuit of their interests, they acknowledge a distinction between what they may and may not do. As rational persons understanding the structure of their interactions, they recognize a place for mutual constraint, and so for a moral dimension. (Gauthier, 1986, p. 9).

Theorists in the Hobbesian tradition, thus, seek to construe contractarian justificatory frameworks that do not require abstraction from one's arbitrary contingencies. In fact, their attempt is to find unique solutions to bargaining scenarios in which bargainers are not required to internalise moral principles, whose normative relevance they fail to grasp. Agents, indeed, come to the bargaining procedure with their own instrumental reasons, shaped by their own structures of preferences and goals, guiding their choice over their terms of cooperation.

The underlying assumption of theorists such as Gauthier and Narveson is that agents are instrumentally motivated to find an agreement over cooperation-enhancing platforms, as without such an agreement they would be worse-off from their own personal standpoint.⁸¹ Therefore, moral and political disagreement do not undermine the possibility of a contractarian agreement, provided that the terms of cooperation give each and every member instrumental reasons to comply with them. Which is to say, under the condition that members of the community see their cooperative enterprise as an enterprise from which they stand to gain.

This approach is the one I intend to follow in order to provide a contractarian

⁸¹See Narveson (1988), when he claims that our instrumental reasons to find a cooperation enhancing agreement are due to the fact that 'we are vulnerable to the depredation of others, and second because we can all benefit from cooperation', p. 148.

justification of the amended version of the Lockean account of the spontaneous order. Unlike Gauthier and Narveson, though, I attempt to model agents as engaging in some sort of abstraction from their structure of preferences and goals when dealing with contractarian bargaining scenarios. However, differently from Rawls, I attempt to show that agents have instrumental reasons to abstract. Particularly, I suggest that agents see the sort of abstraction I outline as ultimately serving their interests in maintaining their ability to shape their social environment in ways that adapt to the dynamic character of their preferences and goals over their lifetime.

Preferences For Flexibility

What emerges from the previous section is that in framing our bargaining scenario we must be careful in considering trade-offs connected to abstraction. Particularly, we must make sure that the veil of ignorance we posit on our bargainers, in order to secure the possibility of an agreement, does not undermine the normative relevance (to actual persons) of solutions to the bargaining procedure. In this section, I wish to show that there is some plausible level of abstraction that would meet such a desideratum. Moreover, I argue, such a proposal has one crucial advantage: namely, it can be justified in an instrumentalist fashion, that is to say that rational persons have instrumental reasons for employing this level of abstractions when making certain kinds of choices, including choices about basic institutional settings of our communities. In fact, I wish to suggest, we employ this level of abstraction when facing many important choices within our lives.

In order to illustrate this point, let me start by considering a choice scenario concerning educational choices for our children. When we are to pick among different schools for our children often we lack specific goals. We may have a vision of them twenty years from now, but such a vision is ultimately blurry; our goal is to see them fully realised and happy but we are unable to tell what is going to take for them to flourish. For instance, they may flourish as astronauts,

academic philosophers or chefs, but we lack solid grounds to predict it and, thus, are unable to tie our educational choice to a precise pattern. Surely, we may have a job in mind for them, but if we attempt to think about their educational choices under the light of their inclinations, our decision would probably be aiming at maximising their opportunities, leaving to them the chance to shape their lives in due course.

The same applies to our own educational choices. Particularly, when we approach college, we may still be uncertain about what sorts of career path we wish for ourselves. We are mature enough to have a general life plan, or to have a set of preferences and goals that rules out certain options and make others salient, but, ultimately, we may come to this choice with the awareness that, whatever specific inclination we have here and now, our structure of preferences and goals is dynamic and that, as such, making educational choices that tie us to present inclinations could drive our life pattern to very undesirable directions. To show this particular aspect, imagine Betty is passionate about acting and that she is offered a position at a great acting academy in Bakewell. At the same time, she has successfully applied for a prestigious Law School in Dore. The choice scenario that Betty faces is quite challenging as she knows that her structure of preferences and goals, here and now, would make her choose the acting academy over the law school but, at the very same time, she recognises that such a choice would dramatically restrict her life path. Were she ever going to change her mind about being an actress, the skills gained at the acting academy would hardly be transferable to other careers and her position on other job markets would be disadvantageous compared to the prospects of a law degree. Moreover, although she has a preference for the acting academy, she is able to grasp the relevance of an education in law, in that she sees the good that such an educational pattern would bring to her. In Agnes Callard's words, Betty also exhibits an 'aspiration' toward being a law school student as she can grasp the relevance of the values associated with that path.⁸² She may not

⁸²See Callard (2018), in which she develop a truly fascinating account of what an aspiration is. In particular, she captures the idea of aspiration as a cognitive state in which we grasp the

entirely like what she knows about being a law student and a lawyer afterwards, and that is part of the reasons why she still prefers the acting academy, but she also knows that by cultivating the values of which she grasps the relevance, she may come to appreciate her life as a law student, and associated career paths, more fully.

The first interesting aspect of Betty's scenario is that the choice she faces cannot be made by merely taking into consideration the specific goals she has here and now, as she knows that educational choices are long-ranging, in that they are to accommodate for changes in our structure of preferences and goals in the distant future. In fact, choosing our education is not like choosing courses at the restaurant: there, we have a precise layout of the available options and the eventual uncertainty linked to dishes we have never tasted is of minor importance, as our choice is not supposed to accommodate for changes in our tastes in the future. In fact, choosing courses at the restaurant merely restricts our opportunity set for that particular meal. There is no need to be forward-looking as that particular choice will not affect our future choices. It is this particular need to accommodate for future and yet partially unknown goals that makes Betty's choice so problematic. In fact, on the one hand she possesses a clear preference for the acting academy, on the other she is aware of the dynamic character of her structure of preferences and goals. Since the educational choice is supposed to have long ranging effects, in that it shall affect her future opportunities, she has to find a way to account for them.

What makes choices concerning educational paths peculiar, though, is not merely their long-ranging nature but also their wide-ranging implications. In fact, we may imagine scenarios in which choices with long-ranging effects only marginally affect our set of opportunities. Suppose, indeed, Betty is asked to choose a brand of jeans she will be committed to wear for the rest of her life. Surely, her preferences may change in due course, such that she may regret having chosen one brand over the other, but her decision would only negligibly affect her life. This is due not merely to the triviality of the object of Betty's choice, relevance of some particular enterprise even though we are unable to fully assess its merits.

but also to the fact that such a decision concerns a hermetically sealed compartment of her life, which would hardly generate repercussions on meaningful facets. Educational paths, on the other hand, do not merely define our opportunity set with respect to a crucial compartment of our life, but rather affect other meaningful aspects (e.g. they provide us with tools to fulfill many different tasks, not strictly related to the career path we have chosen for ourselves; they largely define our social circle, etc.). Long and wide-ranging choices, I suggest, should be seen as multi-purpose tools, in that they are not merely supposed to accommodate for changes in our structure of preferences within one particular aspect, but rather within several and *prima facie* unrelated meaningful facets of our lives.

In order to further illustrate the concept of multi-purpose tool, we may think about choosing computers. Particularly, imagine Betty is passionate about design and that, as such, she is naturally inclined toward buying an iPad Pro. In fact, the Apple Pencil, the high pixel-density and 120Hz refresh rate screen would provide an ideal set up for her drawings. The other option meeting her budget constraint is an uglier and thicker ThinkPad, with lower pixel density and less design capabilities. On the pros side, though, the ThinkPad is largely upgradable, has a much more powerful CPU, and an operating system, alongside with its software ecosystem, which allows for a wide variety of tasks. What should Betty choose? The answer to such a question turns on whether the device Betty is in the market for, is supposed to be her main device; which is to say, whether that computer is supposed to fulfill more than design-related tasks. For instance, were Betty to use her computer also for general educational purposes, she may find the iPad Pro as mostly unsuited to tasks such as word-processing or programming. The problem lies in that, if Betty is partially unaware about the sort of tasks she will be required to fulfill, her choice cannot be merely tied to her passion for design, as choosing an iPad would prevent her from getting some work done with respect to tasks that are not design-related.

Unlike computers, educational paths possess a third peculiar feature which consists in their being largely, though not entirely, one-way routes. In fact,

computers can be easily tried, returned, swapped or replaced; educational paths, on the other hand, cannot. Surely, we may imagine Betty giving a shot to the acting academy and realising after a few months that she will be better off by taking some other route, or deciding to go back to education after obtaining her degree at the acting academy. However, costs associated with such decisions are by no means negligible and, often, our first choice is one we are likely to stick with for our entire life. This particular aspect invites deeper reflections on our choice scenario, as we may be happy to commit ourselves entirely to a particular activity under the condition that its effects, although wide and long-ranging, can be reverted as we please, or that we will still be able to commit to completely different kinds of activities afterwards, and quite often we make one-way decisions under the false assumption that they are reversible.

The one-way nature, alongside with the long and wide-ranging character, of Betty's choice scenario define its complexity, and make us wonder how is it possible to choose educational paths on the basis of information we can only partially grasp. In this regard, I suggest that a plausible answer is offered by the idea of flexibility.

Kreps (1979), in particular, captures our problem as follows:

In many problems of individual choice, the choice is made in more than one stage. At early stages, the individual makes decisions which will constrain the choices that are feasible later. In effect, these early choices amount to choice of a subset of items from which subsequent choice will be made. This paper concerns choice among such opportunity sets, where the individual has a "desire for flexibility" which is "irrational" if the individual knows what his subsequent preferences will be.

A simple example is that of making reservations at a restaurant. Imagine that the only way that restaurants vary is in the menu of meals which they will serve. The individual is assumed to know the menus at all restaurants that he might select. Eventually, the

individual will choose a meal, but his initial choice is of a restaurant/menu from which he will later choose his meal. Let Z be the set of possible meals, with generic element z . Let X be the set of all conceivable menus, with generic element x . That is, X is the set of nonempty subsets of Z . Initial choice is the selection of one x from some subset of X (the set of available menus), and subsequent choice is the selection of one z from the x chosen. For simplicity, assume that Z is finite [...] Suppose that the individual prefers a menu containing only steak to one containing only chicken. But he strictly prefers a menu with both steak and chicken to either of the first two, because it gives him greater flexibility [...] Why should anyone desire such flexibility? The individual choosing a menu would likely make some statement about being uncertain about something or other. For some reason, he is unsure about what will be his mood on the evening in question. Perhaps he is unsure about what he will have had for dinner on the previous evening. Perhaps he cannot explain why he is uncertain about his future mood, but he claims to be unsure of it nonetheless. This rationalization is naturally termed "uncertainty about future tastes. (Kreps, 1979, p. 7).⁸³

Essentially, in Kreps, the need to accommodate for the uncertainty about future tastes equates to Betty's need to accommodate for changes in her structure of preferences and goals in the distant future. Preferences for flexibility, within Kreps' choice scenario, result in an agent's choice that includes both available options as she is uncertain about her future tastes. In particular, I suggest, we may look at Kreps' agent as attempting to abstract from her current tastes in an effort to choose the menu that would likely accommodate for changes in her structure of tastes in the future. Similarly, Betty may attempt to abstract from her current structure of preferences and goals in an effort to choose an educational path that would also accommodate for changes in her structure

⁸³For similar accounts concerning preferences for flexibility, see Koopmans (1962), Stigler and Becker (1977).

of preferences and goals. In particular, she may find the prospect of the Law School as more flexible than that of the Acting Academy, as it would open up much more career paths which are likely to enlarge her set of opportunities in ways that would accommodate for changes in her future structure of preferences and goals, without entirely preventing her from cultivating her current passion for acting.

One peculiarity of Betty's scenario lies in that, differently from Kreps', there are associated trade-offs. Particularly, while Kreps' agent is able to choose menus that contain both steak and chicken, Betty's alternative paths shall constrain her opportunities in one direction or the other. In fact, the Acting Academy is likely to maximise her chances to pursue an acting career path, at the costs of narrowing down her set of opportunities with respect to other career paths; whereas the Law School would enlarge the set of opportunities paving the way for countless career paths while reducing her chances to become a professional actress. Another peculiarity consists in that while Kreps' agent is inflexible with respect to alternative opportunity sets (i.e. steak, chicken, steak and chicken), Betty's uncertainty about career paths is different in degrees. This is, in part, because the relationship between Betty's educational choice and career paths is less straightforward than the one between menus and food, and, in part, due to the fact that Kreps' agent has perfect knowledge of how her tastes may change in the future, whereas the same does not apply to Betty, whose salient opportunity sets are the result of sequential process of discovery, which may occur at later stages within her educational path.

This latter peculiarity, I argue, provides Betty with even more reasons to abstract from her current structure of preferences and goals and choose the Law School, as she is largely unable to assess the ways in which such a structure may evolve in the future. In fact, the acting academy seems to require a full commitment over a single career path at the very outset, whereas the Law School would make Betty able to procrastinate her decision about career paths to further stages, where she will have a fuller understanding of her structure of preferences and goals and of their evolutionary patterns. This particular

aspect clarifies that preferences for flexibility are primarily a function of one's uncertainty about future goals, which is to say that more uncertainty about one's future tastes calls for more abstraction from one's current set of goals and for more flexible choices, especially if paths to achieve the current set of goals are to narrow down one's opportunity set to one's current preferences.

At this point, one may wonder what exactly the concept of flexibility captures. Particularly, one may ask whether it is merely supposed to express one's risk aversion toward non-ideal circumstances, or if there is more on the table other than mere risk assessment. In this regard, I wish to highlight that preferences for flexibility are not the result of a minimax strategy, in that they are not meant to make us choose strategies that will maximise our payoffs under the worst case scenario. Specifically, Betty's motivation for choosing the Law School over the Acting Academy is not aimed at, say, securing the best salary, under conditions in which she fails to be among the top 10% students at either the Acting Academy or the Law School. In fact, preferences for flexibility are meant to highlight our preference for keeping our ability to shape our life plans in due course, by making it adapt to the changing structure of our preferences and goals. Particularly, preferences for flexibility are the expression of our rational desire to continuously shape our lives according to values, preferences and goals that we come to have within our entire existence, rather than being constrained by past choices that cannot adequately account for the sort of persons we are and we will be in the distant future. In this regard, I wish to emphasise that preferences for flexibility are not merely preferences for patterns that open us as many opportunities to satisfy goals compatible with our broader life plans. In fact, the primary aim is to preserve our ability to adapt our lives to changing preferences and goals.⁸⁴

There are two main worries arising from the idea of flexibility. First, it

⁸⁴There is, however, a rather unconventional way to frame preferences for flexibility as risk minimising strategies. Particularly, we may plausibly interpret Betty's choice as ultimately driven by her preference to minimise the risk that opportunities emerging from her educational path will fail to match Betty's evolving structure of preferences and goals. As such, the Law School, by virtue of opening much more career paths, without preventing her from cultivating her passion for acting, would be the minimax strategy.

may seem that we come to exhibit preferences for flexibility only insofar as we irrationally privilege the satisfaction of future goals over present ones. Second, it may seem that flexibility demands us to forgo on meaningful projects in order to preserve our ability to satisfy future goals that have little normative significance to us, when the choice scenario takes place. These two worries, albeit similar, highlight two different aspects: on the one hand, there is a concern related to irrational rates of temporal discounting, as one may plausibly object that Betty's choice to go to the Law School is the result of a process of deliberation that posits too much weight on the satisfaction of future goals; on the other, rates of the temporal discounting that privilege the satisfaction of future, and partially unknown, goals pose problems with respect to the meaningfulness of flexible choices; in fact, one may point out that Betty's reasons to go to the Acting Academy are highlighted by a meaningful goal, that of becoming an actress, that carries stronger normative significance to her compared to future, and partially unknown, goals that demand her to go to the Law School, when the choice scenario takes place. Therefore, one may join Sandel in claiming that the sort of abstraction required by flexibility demands too much detachment in that it asks us to forgo on projects we find of value.

In order to answer these worries, let me first point out that preferences for flexibility are not the result of a process of deliberation that assumes rates of temporal discounting, but rather the other way around. In fact, what drives Betty's choice to go to the Law School is not the satisfaction of future, and partially unknown, goals but rather her present desire to preserve her ability to shape her life according to preferences and needs she comes to possess over her entire existence. Although such a preference for flexibility implicitly embodies the satisfaction of future goals, it is not a future preference. However, someone may object that the mere fact that rates of temporal discounting emerge from preferences for flexibility partially begs the question. For if preferences for flexibility end up by demanding us to posit too much weight on future goals, it may not be rational to develop them. A successful answer, thus, has to further elaborate on how preferences for flexibility emerge and what kind of demands

they posit on us.

In this regard, let me answer by illustrating that preferences for flexibility do not demand the same level of abstraction on all persons. In fact, we have seen that they are a function of one's uncertainty which, I suggest, is also a function of one's complexity, where complexity can be highlighted by one's aspirations. For instance, in Betty's choice scenario, we assumed that she is able to grasp the relevance of values associated with being a law student and related career's paths and is also able to see that these values may come to possess a more central role as she goes on with her life path. In this regard, preferences for flexibility demand different levels of abstraction according to the strength, extension and variability range of one's aspirations.

To show this, imagine a scenario in which Betty lacks the sort of imaginative power required to develop aspirations, aside from her passion for acting. Particularly, assume that Betty's goal of becoming an actress is the only reason for undertaking an educational path. Here, clearly, Betty fails in exhibiting preferences for flexibility with respect to her educational choice. She may still exhibit them with respect to other aspects of her life (e. g. where to live, marriage choices, parental choices, etc.), but flexibility plays no role in choosing her educational path as she is unable to grasp the relevance of values associated with other career paths.

Essentially, Betty sees herself as able to shape her life according to her goals and needs only from within that particular framework. In this regard, one may argue that preferences for flexibility, far from being too demanding, have a too limited reach in making us abstract. For if we fail to exhibit aspirations we also lack incentives to abstract. However, such a conclusion rests on implausible claims about persons' complexity. Particularly, it rests on assumptions highlighting our inability to grasp values associated with opportunities and patterns that are not strictly related to our present preferences and goals.

There is another interesting objection to the idea of flexibility that does not involve demeaning assumptions concerning persons' complexity. Particularly, one may point out that flexibility is still too demanding if it requires us to forgo

on our main aspirations on the ground of less fully developed aspirations. For instance, in the first scenario, Betty's preferences for flexibility demand her to choose the Law School over the acting academy, despite the fact that the acting career is her most developed aspiration. In this regard, flexibility would demand us to forgo on project whose meaningfulness is much more salient to us when the choice scenario takes place.

However, such an objection rests on the assumption that our main aspirations require full commitment on our part. Particularly, it assumes that we cannot cultivate values associated with less fully developed aspirations without sacrificing our ability to cultivate values and reach goals associated to our main aspirations. In the first choice scenario, though, going to the Law School does not necessarily prevent Betty from cultivating her aspiration toward an acting career, as she could still manage to devote some of her time to that particular aspiration. On a similar note, we may highlight recent tendencies, among young tennis players, to undertake an educational paths not necessarily related to the tennis profession, alongside with their career path as professional tennis players. Although these tendencies seem to capture more risk-averse attitudes with respect to the possibility of failure rather than our preference for flexibility, they clarify that even demanding career paths do not prevent us from cultivating values that lie outside our main aspirations. In fact, cultivating less fully developed aspirations could also add something or reinforce the strength of our commitment to our main aspirations.

However, the objection still stands for there may be cases in which there is an important trade-off between cultivating less fully developed aspirations and cultivating our main aspirations. Particularly, there may be cases in which cultivating less developed aspirations would significantly reduce our chances to satisfy goals connected to our main aspirations. The question, then, is what flexibility would demand. Predictably, the answer is: it depends. It depends on the strength of our main aspirations, on the strength of less fully developed aspirations, on the terms of the trade-offs, and on how restricting associated patterns are.

For instance, in Betty's choice scenario, she may be willing to further investigate the strength of her main aspiration and compare it with less fully developed aspirations; she may also analyse to what extent the Law School would reduce her chances to pursue an acting career, and evaluate how restrictive the two educational paths are in relation to other possible careers she may find of value. In this regard, I should like to stress that, although Betty may be willing to evaluate these aspects from a risk assessment perspective, this is not what preferences for flexibility are meant to highlight. Particularly, our preferences for flexibility are not meant to capture the rationality of choosing strategies that accommodate for non-ideal scenarios. In fact, preferences for flexibility highlight our present and meaningful desire to be able to shape our life according to goals and needs we come to possess over the course of our existence. In this regard, they invite us to carefully consider what sorts of persons we are likely to be in the distant future, in order to make sure not to undertake paths that would prevent us from shaping our life in due course.

Clearly, preferences for flexibility should not be taken as granted. In fact, we may fail both in grasping the nature of the choice scenario we face, by neglecting its one way character or its long and wide-ranging effects, and in assessing how volatile our current set of preferences and goals is. Particularly, we may fail in evaluating the paths we choose for ourselves under the light of a larger domain, which is the domain of our entire existence. Such a level of abstraction is by no means easy to achieve as, to paraphrase Augustine, our hands are full. In fact, we are often entirely absorbed by specific projects and goals to the point that we lose the ability to look at our choices by taking into account the kinds of persons we will be in the future. Nonetheless, to show preferences for flexibility is to show a full understanding of the dynamic character of our values, preferences and goals, a character over which we have only marginal control. In fact, I argue, rational persons have instrumental reasons to reach this level of abstraction in making choices that profoundly affect their life plans. As by abstracting from our current set of goals we converge on options that would equip us with the ability to continuously renegotiate what we do, in the attempt of reconciling it

with who we are.

Preferences for Flexibility, Social Environment Shaping Rights and Bargaining Procedures

The sort of abstraction implied by preferences for flexibility, thus, does not undermine the possibility of making rational and meaningful choices. There are, indeed, various cases in which we have instrumental reasons to reach this level of abstraction. In fact, choices that have profound effects on our lives demand us to put our current set of goals and preferences under deeper scrutiny and that we consider the sort of persons we may be in the distant future.

The case for flexibility, I aim to show, is even more intuitive within contractarian bargaining scenarios, where we come to define the basic set of institutions that shall govern the interactions among members of a community. It is more evident because, by defining the terms of our cooperation, basic institutional arrangements also shape the kind of opportunities that members of the community shall have in many different aspects of their lives. Bargaining scenarios, indeed, possess the very same features of Betty's scenario: they are one-way decisions and their effects are both long and wide-ranging. Moreover, since their effects are pervasive within many different aspects our lives, they call for abstraction from one's specific goals in these various facets. In fact, rational agents, approaching the bargaining table, recognise that their terms of cooperation cannot be shaped by current sets of preferences and goals, if, by doing so, they prevent institutional adaptability to the structure of goals they may come to possess at future stages of their lives. Particularly, they have reasons to converge on cooperation-enhancing platforms that leave constantly open the possibility for further renegotiation. That is to say that they have reasons to endorse a basic institutional arrangement that continuously allows them to shape their social environment by making it adapt to their dynamic structure of preferences and goals. More specifically, they have reasons to pick institutional arrangements that would allow them to satisfy preferences and goals that fall

within the set of aspirations carrying values they can grasp the relevance of.

Clearly, the level of abstraction implied by preferences for flexibility may not solve all the problems connected to disagreement among bargainers. In fact, the veil of ignorance agents posit on themselves is not thick enough to make them possess the very same set of preferences and goals. In this regard, it is still possible to imagine situations in which bargainers' sets of aspirations are radically different, to the point that there is no set of institutions able to accommodate them. This particular fact, clarifies that preferences for flexibility are inconclusive with respect to the possibility of a universal agreement on a single set of social institutions.

In particular, there are two difficulties I wish to highlight. The first is common to all orthodox rational choice contractarian accounts and concerns people's moral beliefs over just distributive principles. Particularly, contractarian accounts tend to construe distributive principle as arising from bargaining scenarios as a function of people's structure of preferences and goals. This means that rational agents are supposed to choose distributive principles that maximise their chance of satisfying their individual preferences and goals. In this regard, they cannot accommodate for widely diverse moral beliefs over just distributive principles that are not formed instrumentally. This means that members of the community who, for instance, share deontic beliefs over equal distribution of resources, regardless of how such a distribution is going to affect one's chances to satisfy goals and preferences, cannot possibly be content with a contractarian bargaining procedure. This is because these members are willing to sacrifice the benefits of cooperation on the altar of their moral beliefs over just distributions.

The second difficulty, on the other hand, is peculiar to the theory I have offered and concerns the fact that preferences for flexibility are also a function of members of the community's age. In this regard, one may plausibly object that older societies would be less prone to converge on institutional arrangements that leave the door open for continuous adaptation. The answer to such an objection is twofold: on the one hand, I suggest, our close ties with future generations equip us with reasons to design institutional arrangements that

would allow them to shape their social environment, in the same way in which we attempt to make educational choices for our children that would allow them to flourish in their own peculiar way; on the other hand, I wish to suggest that we should not underestimate our interest in keeping our ability to shape our social environment even at a later stage of our life. Surely, members of the community who approach the end of their existence are less inclined to abstract from their specific goals as they can predict that their structure of preferences and goals will be less subject to dynamic changes; yet, changes in our structure of preferences and goals do not pre-alert us with deadlines. We may encounter them at different stages of our lives without any ability to predict when they would take place and what would trigger them. In this regard, preferences for flexibility, although less pronounced as the time passes, are always rationally formed.

However, despite these difficulties, such a thin veil of ignorance modelled by preferences for flexibility brings with it great benefits. Particularly, we do not have reasons to worry that the result of the bargaining procedure shall lack normative relevance to actual agents. In fact, agents choose to abstract from their specific goals on the ground of instrumental reasons. This means that there is no rational ground for evaluating the result of the bargaining procedure in light of their specific goals, as they are ultimately irrelevant to their choice. The question, then, shifts to the Lockean account of the spontaneous order. More specifically, we are now to evaluate whether institutional arrangements purported by the Lockean distribution of rights would encounter the endorsement of agents exhibiting preferences for flexibility. In this regard, the answer rests on the plausibility of Hayek's claim with respect to the unique ability of the Lockean distribution of rights in bringing about cooperation-enhancing platforms where members can continuously update their terms of cooperation. Particularly, if we are persuaded by Hayek's argument, a Lockean distribution of rights would certainly be a plausible bargaining solution. In fact, such an institutional arrangement would capture the need for adaptability that agents exhibiting preferences for flexibility would share. However, what preferences for

flexibility clarify, is that members of the community have reasons to maintain their ability to shape their social environment overtime, as this allows them to renegotiate their terms of cooperation in the light of changes in their structure of preferences and goals.

In this regard, they have reasons to incorporate the two desiderata I have been offering. Particularly:

- (1) they have reasons to endorse limitations to their right to transact in order to prevent states of affairs in which they lack the ability to shape their social environment.
- (2) they have reasons to prevent the emergence of (or to dismantle) social norms that would generate states of affairs in which they lack the ability to shape their social environment.

In fact, these two amendments would resolve problems connected to unintended derivative distributions of rights arising from individual interactions allowed by the Lockean triad. Particularly, members of the community, who exhibit preferences for flexibility, would have instrumental reasons to accept limitations to their ability to transact, if such limitations are posited in order to prevent the emergence of states of affairs in which some members would lack the possibility of conveying information that would allow them to continuously shape their specific terms of cooperation in ways that would meaningfully respond to their structure of preferences and goals.

In this regard, I suggest, agents exhibiting preferences for flexibility would attribute particular relevance to the peculiar role that the distribution of social environment shaping rights plays within our communities. In fact, they would have reasons to ensure that such a distribution makes all (or nearly all) members of the community able to continuously convey information about their terms of cooperation, as this would grant to the basic institutional arrangement an instrumentalist character. In fact, members of the community have:

- (3) an interest in their community remaining a cooperative enterprise from which they stand to gain.

(4) an interest in that all (or nearly all) members find the community as a cooperative enterprise from which they stand to gain.

While (3) is intuitive, (4), as we have seen within the previous chapters, is attributed to the fact that when members of the community fail to see their cooperative endeavour as an enterprise from which they stand to gain, they lack instrumental reasons to comply with the basic structure, thus, undermining the stability of the community.

This particular aspect implicitly calls for a sufficientarian interpretation of the distribution of social environment shaping rights, where the basic threshold is to be identified according to the amount of rights that equips members with the ability of conveying information about their terms of cooperation. For if the basic threshold is not met, (4) is not met either.

In this regard, one may wonder why bargainers would call for a sufficientarian distribution of social environment shaping rights, rather than for a more demanding distribution (e. g. an equal distribution). The answer to such a question lies in the trade-off between our ability to shape our social environment and limitations to our right to transact (including redistributions of rights). In fact, our right to transact is, in itself, a means to shape our social environment in ways that make it adapt to our structure of preferences and goals. Therefore, limitations to our right to transact are also limitations to our ability to shape our social environments. A sufficientarian account, in this regard, captures the fact that bargainers have reasons to pose limitations to their right to transact only insofar these limitations are needed to preserve the stability of the community.

Conclusion

Concluding, let me summarise the two main achievements of this chapter. First, I have attempted to show that emergent justifications lack justificatory force in dealing with derivative distributions of rights. Particularly, I have shown that the justificatory power of actual consent within individual transactions cannot be easily transferred to large-scale consequences generated by such transactions.

This, we have seen, has crucial implication for our own purposes, as it does not rule out the possibility of positing amendments to the Lockean distribution of rights that would limit our right to transact in order to prevent the emergence of states of affairs in which some members of the community are unable to shape their social environment by making it adapt to their structure of preferences and goals.

Second, I have shown that such limitations can be justified in a contractarian fashion. Particularly, I have highlighted a bargaining scenario in which agents come to define the basic institutional arrangements of their community, exhibiting preferences for flexibility. These preferences emerge from the fact that bargainers have reasons to abstract from their current set of preferences and goals, as basic institutional arrangements are supposed to regulate their interactions overtime, thus, including stages of their lives in which their structure of preferences and goal will be subject to change. Preferences for flexibility, in this regard, capture need for institutional arrangements that allows for continuous renegotiations of members of the community's terms of cooperation. In fact, the dynamic nature of their structure of preferences and goals calls for cooperation-enhancing platforms in which members of the community are able to shape their social environment.

In the next chapter, I shall deal with persistent disagreement. That is with the sort of disagreement agent would still exhibit after preferences for flexibility are elicited. Persistent disagreement, I shall suggest, requires that we think about cooperation not merely within the community, but also at a meta-community level. In fact, if one cares about designing institutional arrangements that are able to accommodate for a wide variety of standpoints, our political theorising should deal with designing institutional arrangements, at the meta-community level, that would create the conditions for the emergence of communities that aptly suit people's structures of preferences and goals.

Preferences For Flexibility and The Framework For Utopia

Introduction

In the previous chapter I suggested that the amended theory of the spontaneous order allows us to respond to some pressing challenges related the fact of diversity. In particular, I have claimed that preferences for flexibility provide an interesting tool to secure convergence on institutional arrangements even when people's preferences and goals are, at a first glance, incompatible.

This particular effect obtains when individuals are uncertain about their future aims. In fact, uncertainty invites them to abstract from their current set of preferences and goals when coming to the bargaining table. In particular, individuals' uncertainty elicits Kreps' preferences for flexibility in designing their preferred institutional arrangements, and, thus facilitates coordination within social contract bargaining scenarios.

In this chapter, I attempt to deal with persistent disagreement, that is with the sort of disagreement, stemming from non-compossibility of preferences, goals and conceptions of the good, that cannot be accommodated by the emergence of preferences for flexibility.

If we take diversity seriously, and if we care about accommodating disagreement stemming from non-compossibility of goals, our political theorising cannot be restricted to the domain of designing institutional arrangements at the community level, but should rather look at how institutions operate at the meta-community level. The underlying idea is that, since there is no single community able to accommodate for all sorts of disagreement, we should be looking for institutional settings that would allow us to create as many communities as possible in order to successfully respond to our preferences, goals and conceptions of the good.

This is exactly Nozick's aim in delineating the framework for utopia, where he attempts to justify the minimal state as the institutional arrangement that

best approximates the possible-worlds model. In this chapter, I put Nozick's theory under scrutiny and attempt to show that the minimal state faces two main challenges in the route to approximate the possible-worlds model: first, collective action problems stand in the way of creating new communities; second, informal sanctions, underpinned by social norms, may increase the costs for creating new cooperative enterprises or joining already existing ones. Such challenges, I attempt to show, invite amendments to Nozick's theory that seems to point toward institutional arrangements, at the meta-community level, more demanding than the minimal state.

The chapter is organised as follows: within the first section, I illustrate how preferences for flexibility help us in accommodating some sort of disagreement; within the second section, I show how certain kinds of disagreement, stemming from non-compossibility of preferences and goals, will still persist within pluralistic societies; within the third section, I propose Nozick's discussion of utopia as an appealing framework in order to respond to the fact of diversity; within the fourth section, I put under scrutiny Nozick's claim according to which the minimal state represents the institutional arrangement that best approximates the idea of utopia; within the fifth section, I propose certain amendments to Nozick's theory.

Agreement Through Preferences For Flexibility

Within the previous chapter, I have highlighted the idea of preferences for flexibility and illustrated how flexibility provides instrumental reasons for abstracting from specific preferences and goals when facing one-way, long and wide-ranging choice scenarios, such as decisions about the basic institutional arrangements of our social orders. Particularly, I have attempted to show that bargainers' uncertainty about future preferences and aims invites them to abstract from their current sets of goals and to come to the bargaining table by considering a wider range of goals that they predict could come to be salient in the future.

The core idea behind the concept of flexibility is that we have reasons to choose institutional arrangements that would allow us to design our lives in due course, according to preferences and aims we come to possess within our entire life path. As such, institutional arrangements should not be tied to narrow sets of preferences and goals that are salient to us when the bargaining procedure takes place if, by tying our choice to such a set, we are prevented from (or we increase the costs for) pursuing goals that will come to be salient in the future.

Let us briefly turn to Kreps (1979) in order to summarise the main argument. Suppose Betty is asked to choose among three different menus for tomorrow's dinner: i contains only steak, ii contains only chicken, and iii contains both steak and chicken. Betty strictly prefers steak over chicken, and, as such, she strictly prefers i and iii over ii , but is indifferent between i and iii as both contain steak. If we introduce additional costs attached to iii (e.g. iii is pricier or, say, the quality of the steak is inferior compared to i), Betty would strictly prefer i over iii . However, if Betty is uncertain about her mood tomorrow, and her mood may affect her preference toward either steak or chicken, her preference ordering may substantially change as she develops preferences for flexibility. In fact, let m be a variable capturing her mood tomorrow, such that, if m_s , she would strictly prefer to eat steak and, if m_c , she would rather have chicken. Assume further that Betty is able to assign probabilities about her mood tomorrow, such that m_s shall occur with probability p and m_c with $1 - p$. If there are no further costs attached to iii , Betty would strictly prefer iii over i , as iii will satisfy her preference for either steak or chicken with probability 1. If, on the other hand, we attach further costs to iii , Betty's preference ordering between i and iii shall ultimately rest on the amount of costs attached to iii and to the distribution of probabilities about her future mood. Particularly, the more she is uncertain, the more such an uncertainty will offset the eventual costs attached to iii , as it calls for deferring her choice between steak and chicken to tomorrow, when she will be clear about her mood. The sort of abstraction elicited by preferences for flexibility, thus, rests on our uncertainty about future preferences. In fact, the more we are uncertain about our future preferences,

the more we wish to abstract from goals that are salient to us when the choice scenario takes place, and rather consider larger sets which include goals that may be salient in the future.

Preferences for flexibility, I suggest, identically apply to choices concerning institutional arrangements (IA), under the assumption that agents choose IAs on the ground of the opportunities they provide in order to achieve goals and aims that are of value to agents. In fact, suppose Betty is asked to express her preference among IA_1 , IA_2 and IA_3 . Assume further that each institutional arrangement provides different sorts of opportunities to satisfy different goals. For instance, imagine that IA_1 provides opportunities to satisfy x , IA_2 provides opportunities to satisfy y , and IA_3 fosters her chances to satisfy x and y . If Betty shows interest only in x , she would rank IAs as follows: $IA_1 \sim IA_3 > IA_2$. If we attach some costs to IA_3 (e.g. it provides opportunities to satisfy x and y , but less opportunities to satisfy x , compared to IA_1), then, Betty would strictly prefer IA_1 over IA_3 . However, if Betty is uncertain about the relevance of y , in that it may become a salient goal in the future, such an uncertainty may offset the costs associated with IA_3 and shape her preference ranking as follows: $IA_3 > IA_1 > IA_2$. Betty's uncertainty, once again, invites her to abstract from the salience of x and to choose institutional arrangements by considering goals that are less salient when the choice scenario takes place, but that may be of relevance at later stages. Furthermore, the sort of flexibility-driven abstraction that is at stake, changes her preference ranking, making IA_3 her preferred institutional arrangement.

One interesting implication of preferences for flexibility is that they help us in achieving coordination among agents exhibiting seemingly incompatible preferences, when such preferences are compossible. In order to show this, we may turn back to our example concerning menus, and transform it into a mixed motives game played by Alf and Betty, where Betty strictly prefers steak over chicken and Alf strictly prefers chicken over steak. In such a game, even assuming players' preference for coordination, the incompatibility of their tastes may prevent them from coordinating. Such a problem, though, is solved if Alf

and Betty have reasons to abstract from their current preferences on the ground of their uncertainty about future tastes.

In order to show this, imagine Alf and Betty are to choose among three different restaurants for tomorrow's dinner. Restaurants' menus, we have seen, may be aptly characterised as sets, such that $i = (\textit{steak})$, $ii = (\textit{chicken})$ and $iii = (\textit{steak}, \textit{chicken})$. If there are no costs attached to iii , Alf's preference ranking would look as follows: $ii \sim iii > i$, whereas Betty's is: $i \sim iii > ii$. If they show a preference for coordination, they will likely converge on iii , without the need of flexibility-driven abstraction; however, if we attach costs to iii (e.g. iii is pricier or the quality of steak and chicken is inferior compared respectively to i and ii), their preference for coordination may not be enough in securing actual coordination. In fact, let us assume that Alf obtains 3 utils from ii , 2 util from iii and zero from i . Betty, on the other hand, obtains 3 utils from i , 2 from iii and 0 from ii . Assume further that their preference for coordination is valued 1 util, when satisfied, and zero, if not satisfied.

		Alf		
		i	ii	iii
Betty	i	(4, 1)	(3, 3)	(3, 2)
	ii	(0, 0)	(1, 4)	(0, 2)
	iii	(2, 0)	(2, 3)	(3, 3)

In such a game, coordination cannot be achieved. In fact, Betty's dominant strategy is to choose i , whereas Alf's dominant strategy is to choose ii . As such, (i, ii) is a strict Nash equilibrium.⁸⁵

However, uncertainty about future tastes may change the structure of the payoffs and make Alf and Betty converge on iii . Such a possibility rests on the degree of uncertainty Alf and Betty possess, that is on how Alf and Betty distribute probabilities about their future mood. In order to show this, let m be the variable identifying Alf and Betty's mood tomorrow, where m_s identifies the mood for steak and m_c the mood for chicken. Assume, for the sake of simplicity, that if Alf's m_s is realised, his structure of the payoffs is identical to that of

⁸⁵The first entry in each combination represents the payoff to Betty and the second entry gives the payoff to Alf.

Betty above highlighted, whereas if Betty's m_s is realised, her payoffs' structure is equivalent to that of Alf above highlighted. Hence, we have four possible combinations depending on Alf and Betty's mood.

		Alf m_c		
		<i>i</i>	<i>ii</i>	<i>iii</i>
Betty m_s	<i>i</i>	(4, 1)	(3, 3)	(3, 2)
	<i>ii</i>	(0, 0)	(1, 4)	(0, 2)
	<i>iii</i>	(2, 0)	(2, 3)	(3, 3)

		Alf m_s		
		<i>i</i>	<i>ii</i>	<i>iii</i>
Betty m_c	<i>i</i>	(1, 4)	(0, 0)	(0, 2)
	<i>ii</i>	(3, 3)	(4, 1)	(3, 2)
	<i>iii</i>	(2, 3)	(2, 0)	(3, 3)

		Alf m_c		
		<i>i</i>	<i>ii</i>	<i>iii</i>
Betty m_c	<i>i</i>	(1, 1)	(0, 3)	(0, 2)
	<i>ii</i>	(3, 0)	(4, 4)	(3, 2)
	<i>iii</i>	(2, 0)	(2, 3)	(3, 3)

		Alf m_s		
		<i>i</i>	<i>ii</i>	<i>iii</i>
Betty m_s	<i>i</i>	(4, 4)	(3, 0)	(3, 2)
	<i>ii</i>	(0, 3)	(1, 1)	(0, 2)
	<i>iii</i>	(2, 3)	(2, 0)	(3, 3)

Suppose now that Alf and Betty are able to attribute probabilities on whether m_s and m_c may occur, such that $p(m_s) + p(m_c) = 1$. For instance, assume that Betty believes that m_c shall occur with probability 0.6 and m_s with 0.4, whereas Alf believes that m_c will occur with probability 0.4 and m_s with 0.6. Here's the distribution of expected payoffs:

Essentially, uncertainty about future tastes has transformed the original non-cooperative game into a coordination game with three Nash equilibria (i, i ; ii, ii ; iii, iii), and in which (iii, iii) offers the Pareto-optimal outcome.

		Alf		
		<i>i</i>	<i>ii</i>	<i>iii</i>
Betty	<i>i</i>	(4.4; 5.6)	(2.4; 2.4)	(2.4; 4)
	<i>ii</i>	(3, 6; 3.6)	(5.6; 4.4)	(3.6; 4)
	<i>iii</i>	(4; 3.6)	(4; 2.4)	(6; 6)

Similarly, preferences for flexibility can accommodate for some sort of disagreement within social contract bargaining scenarios. In fact, although agents may exhibit quite diverse sets of goals, underpinning different institutional arrangements, convergence on a common institutional setting can be secured if agents' uncertainty about future aims invites them to abstract from their current preferences and goals and to consider a larger set when coming to the bargaining table. Larger sets of goals, indeed, would foster the chances of an overlapping consensus among bargainers, thus accommodating for some sort of disagreement among different options.

The Persistence of Disagreement

Preferences for flexibility, though, cannot eliminate all sorts of disagreement. Two types of scenarios, in particular, seem to be of help in explaining the persistence of disagreement. The first concerns situations in which agents are rather certain about their future preferences (or believe to be so), such that larger sets of goals, eventually elicited by preferences for flexibility, fail to emerge.⁸⁶ The second concerns situations in which agents' larger sets of goals, elicited by preferences for flexibility, are broadly non-compossible, such that no institutional arrangement would be satisfactory from everyone's standpoints.⁸⁷

The theory of the spontaneous order, with the amendments I have presented throughout the present work, does not say anything illuminating about the pos-

⁸⁶For instance, imagine Alf and Betty are quite certain about their future tastes for tomorrow night: Alf will like to eat steak and Betty will be leaning toward chicken. If, as postulated above, there are costs associated with the larger menu, coordination is not an equilibrium as Alf and Betty would be better off by unilaterally deviating.

⁸⁷Particularly, imagine a situation in which Alf and Betty are uncertain about their future tastes tomorrow; however, their uncertainty does not generate room for an overlapping consensus over a single menu, as there simply is no menu able to accommodate for their sets of preferences.

sibility of accommodating for non-compossible goals, as its main focus is to lay out how institutions should be designed in order to allow for continuous renegotiation in a context where members of the community's preferences and goals are dynamic but compossible. In fact, we have seen that the scope of the theory is to foster adaptation to changes in members of the community's structure of preferences and goals in order to preserve the instrumentalist character of our communities. However, such an aim is of little significance in a context characterised by widespread non-compossibility of goals, as we face an antecedent challenge, which is to find an institutional arrangement that is able to make our community a cooperative enterprise in the first place, that is an enterprise from which everybody stands to gain.

The problem of disagreement stemming from non-compossibility of diverse conceptions of the good is captured by Robert Nozick in *Anarchy, State and Utopia*:

Wittgenstein, Elizabeth Taylor, Bertrand Russell, Thomas Merton, Yogi Berra, Allen Ginsburg, Harry Wolfson, Thoreau, Casey Stengel, The Lubavitcher Rebbe, Picasso, Moses, Einstein, Hugh Hefner, Socrates, Henry Ford, Lenny Bruce, Baba Ram Dass, Gandhi, Sir Edmund Hillary, Raymond Lubitz, Buddha, Frank Sinatra, Columbus, Freud, Norman Mailer, Ayn Rand, Baron Rothschild, Ted Williams, Thomas Edison, H. L. Mencken, Thomas Jefferson, Ralph Ellison, Bobby Fischer, Emma Goldman, Peter Kropotkin, you, and your parents. Is there really one kind of life which is best for each of these people? Imagine all of them living in any utopia you've ever seen described in detail. Try to describe the society which would be best for all of these persons to live in. Would it be agricultural or urban? Of great material luxury or of austerity with basic needs satisfied? What would relations between the sexes be like? Would there be any institution similar to marriage? Would it be monogamous? Would children be raised by their parents? Would

there be private property? Would there be a serene secure life or one with adventures, challenges, dangers, and opportunities for heroism? Would there be one, many, any religion? How important would it be in people's lives? Would people view their life as importantly centered about private concerns or about public action and issues of public policy? Would they be single-mindedly devoted to particular kinds of accomplishments and work or jacks-of-all-trades and pleasures or would they concentrate on full and satisfying leisure activities? Would children be raised permissively, strictly? What would their education concentrate upon? Will sports be important in people's lives (as spectators, participants)? Will art? Will sensual pleasures or intellectual activities predominate? Or what? Will there be fashions in clothing? Will great pains be taken to beautify appearance? What will the attitude toward death be? Would technology and gadgets play an important role in the society? And so on. (Nozick, 1974, pp. 310-311).

Nozick frames the problem of disagreement ultimately as a concern about sub-optimal institutional arrangements. In fact, his worry is that a unique set of institutions would require too many compromises given the fact of diversity. Particularly, Nozick's problem is that a unique set of political institutions would fail in representing the best institutional setting from the standpoints of many individuals exhibiting widely different conceptions of the good and resulting utility functions. However, the challenges posed by ineradicable disagreement go beyond the problem of sub-optimality. In fact, individuals exhibiting sets of goals which are largely non-compossible may not be merely worried about living in a community whose social and political institutions fail to best advance their conceptions of the good, but rather struggle to see their community as an enterprise from which they stand to gain. As such, they may fail to see reasons to comply with the community's cooperative scheme, thus undermining its instrumentalist character. In this regard, the challenges posited by diversity

and disagreement do not merely tackle sub-optimality but may rather undermine the possibility of securing a basic agreement under a unique set of political institutions.

This particular aspect clarifies that any political theorising that wishes to take diversity seriously should be oriented toward a model of cooperation embodying multiple social contracts underpinning diverse sets of political institutions, rather than focusing on a unique set allegedly able to accommodate for everyone's conception of the good. However, simply arguing in favour of models of cooperation embodying multiple communities underpinned by diverse social contracts cannot possibly count as an adequate response to the problem of diversity. An adequate account, indeed, has to illustrate *how* communities would be created and specify the ways in which they would be able to respond to challenges posed by widespread diversity.

Interestingly, the sort of disagreement we have laid out so far seem to deal exclusively with static preferences. In fact, following Kreps' example on uncertainty about future tastes, we have observed that, given people's widespread diversity in terms of preferences and goals, and given the non-compossibility of certain sets of goals, our model of cooperation has to embody multiple communities. However, such a static account of disagreement is not the only source of concerns about models of cooperation envisaging a unique set of institutions. In fact, disagreement may not be merely exogenous, in that it stems from incompatible preferences that institutional arrangements are supposed to accommodate once and for all, but can also result from changes in the dynamic structure of people's preferences and goals. This is to say that members of a community, who are able to reach an agreement about the set of institutions that is supposed to regulate their interactions overtime, may come to be unsatisfied about the terms of their agreement when their structure of preferences and goals changes.

The amended version of the spontaneous order attempts to respond to this particular problem by inviting convergence on distributions of rights that allow for continuous renegotiation of the terms of the agreement, by preserving

a certain distribution of social-environment shaping rights. Yet, members of the community's goals and preferences may come to be widely diverse after the agreement is reached, such that institutions cannot be modified so as to accommodate for such a diversity. Therefore, the need for a model of cooperation embodying multiple social contracts emerges not merely in order to respond to a static account of disagreement, but also in order to respond to disagreement stemming from changes in the dynamic structures of people's preferences and goals.

In fact, a community's stability cannot be taken for granted once the agreement on a set of institutions is reached, even under the condition that its members maintain the power to exercise their social-environment shaping rights in order to continuously renegotiate their terms of cooperation, according to the dynamic structure of their preferences and goals. Instability, indeed, may spontaneously emerge at later stages if members of the community come to possess sets of preferences and goals that are broadly non-compossible. As such, the attempt to shape our social environment through a continuous process of renegotiation may always leave some members of the community outside the cooperative scheme, in that they fail to see the community as an enterprise from which they stand to gain.

In order to illustrate this particular problem, let us turn back to Alf and Betty's coordination game. Let us assume that their uncertainty about future tastes, as spelled out in the example above offered, made them converge on the third menu. However, suppose that after the agreement is reached, Betty realises she will not be willing to eat chicken tomorrow. Her uncertainty about future tastes is now between steak and hummus. Similarly, Alf comes to know that he will not be willing to eat steaks, as his uncertainty is now between chicken and sushi. If there is no menu able to accommodate for their tastes, their attempt to find a plausible coordination point by communicating to each other their updated preferences is ultimately useless. In fact, their preferences about what to eat tomorrow are simply non-compossible, such that there are no coordination points that would be mutually advantageous.

Here, Alf and Betty's exercise of social environment shaping rights in order to renegotiate their terms of coordination is ultimately helpless as any proposed point of coordination would be unsatisfactory. In fact, either Alf or Betty would have little to gain from coordinating, and their agreement, thus, comes to lose its instrumentalist character. This particular aspect clarifies that the amended theory of the spontaneous order is supposed to work within communities where the dynamic structure of preferences and goals of its members is broadly compossible. When such a condition does not obtain, preferences for flexibility are unable respond meaningfully to the problem of disagreement. Crucially, non-compossibility may emerge after the agreement is reached as there is no guarantee that members of the community's sets of goals and preferences shall evolve toward compossible directions.

In this regard, any attempt to theorise about a model of cooperation embodying multiple communities, underpinning diverse sets of political institutions, has to deal also with disagreement endogenously emerging from the dynamic structure of people's preferences and goals. This is to say that a static model envisaging multiple social contracts laid out on the basis of revealed and existing preferences will not do the job, as preferences and goals are subject to a process of continuous scrutiny and change. Therefore, the model should account for the sort of dynamic disagreement spontaneously emerging from the evolving structure of people's preferences and goals.

The Framework For Utopia

The project of a political theory taking diversity seriously is one that many authors have undertaken.⁸⁸ Two rather similar approaches are particularly relevant to the present work. One has been developed by Chandran Kukathas in *The Liberal Archipelago* and the other is Nozick's framework for utopia developed in *Anarchy, State and Utopia*, in the search for an alternative justificatory framework for the minimal state. The relevance of these approaches for the

⁸⁸Among the others, see Kukathas (2003), Nozick (1974), Rawls (1971, 1993), Kymlicka (1995), Walzer (1995).

present work consists in that both start from quite similar premises. First, in both accounts diversity is taken as an ineradicable fact and not as a value worth pursuing. This is not to deny that diversity could, in principle, possess some instrumental value,⁸⁹ but to assert merely that there is nothing inherently valuable in the fact of diversity, such that social institutions should be in any way required to promote it. Second, and strictly linked to this first point, diversity is taken as a constraint in designing institutional arrangements that would allow for peaceful cooperation or coexistence among persons.⁹⁰ Third, both approaches aim to design institutional arrangements, at the meta-community level, by attempting to maximise the fittingness of communities with people's conceptions of the good. This is to say that both Nozick and Kukathas attempt to secure a state of affairs in which there are enough communities that would constitute the best possible match with people's preferences and goals. Fourth, both Nozick and Kukathas believe that communities cannot be designed in an apriori fashion, as theorists and planners are not in the position to collect persons' preferences, goals and conception of the good; as such, any institutional arrangement, at the meta-community level, has to rely on filtering devices, such as individual consent, in order to allow for the emergence of communities that would constitute the best possible matches with people's preferences.

Nozick's framework for utopia, in particular, lays out a rather analytic account of the criteria that a set of institutions should aim to maximise at the meta-community level. Although the model is built with the aim of exploring an alternative justificatory framework for the minimal state,⁹¹ its criteria, I aim

⁸⁹For instance, many authors have showed that diverse perspectives help us in resolving complex problems (Page 2010, Muldoon 2016), or have convincingly shown that experiments in living may lead to richer ways of life yielding more human flourishing (Muldoon 2016).

⁹⁰Kukathas (2003), for instance, takes diversity as a source of a problem liberalism attempts to resolve: "Liberalism is not, as Galston has argued, about the protection of diversity. If the description of liberal society offered here is defensible, diversity is not a value of any particular importance. Diversity is, in fact, not the value liberalism pursues but the source of the problem to which it offers a solution. That problem is the problem of how individuals can be free in circumstances of conflict over the question of how people should live?", p. 29.

⁹¹See Nozick (1974): "But doesn't the idea, or ideal, of the minimal state lack luster? Can it thrill the heart or inspire people to struggle or sacrifice? Would anyone man barricades under its banner? It seems pale and feeble in comparison with, to pick the polar extreme, the hopes and dreams of utopian theorists. Whatever its virtues, it appears clear that the minimal state is no utopia. We would expect then that an investigation into utopian theory

to show later in the chapter, could point toward different directions.

Let us start our investigation by introducing Nozick's model with a definition of utopia. A utopian world is the best possible world such that no better world can be imagined. Hence, improvements, within a utopian world, cannot take place: if we live in a utopian world, we could not possibly imagine a change in one of its features that would make us better off. The bestness of a world is not to be judged from any particular conception of the good,⁹² but rather from each individual standpoint. Different people would rank worlds differently and, as such, diversity gives rise to a utopian model that is constituted by sub-utopias, that is by various communities where members exhibiting similar criteria for judging worlds would come to cooperate and flourish.

In order to capture the idea of utopia, Nozick proposes us an analogy with the possible-worlds model. In the model, infinite worlds and individuals can be created and each individual can, in turn, create her own utopian world and its members by simply imagining them. Each individual, with her own imaginative power, can thus generate political institutions that would best fit with their own preferences, goals and conception of the good, such that no improvements could be imagined.⁹³ Crucially, for these worlds to qualify as utopian, they have to be stable, where stability can be defined as a world's property such that "none of their inhabitants can imagine any world they consider to be stable and of which they would rather be an inhabitant" (Bader 2011, p. 257). Hence, worlds in which some members can imagine other worlds that would better advance their own conceptions of the good, by virtue of being unstable, cannot possibly qualify as utopian.

The possible-worlds model is meta-utopian if and only if all worlds it contains

should more than serve to highlight the defects and shortcomings of the minimal state as the end of political philosophy. Such an investigation also promises to be intrinsically interesting. Let us then pursue the theory of utopia to where it leads", p. 297.

⁹²see Nozick (1974), p. 309 and p. 333.

⁹³In order to make the model meaningful, one main condition applies: "The world cannot be imagined so that it logically follows that (1) its inhabitants (or one of them) most (or nth most) want to live in it or (2) its inhabitants (or one of them) most (or nth most) want to live in a world with certain (kind of) person, and will do whatever he says, and so on". (Nozick, 1974, p. 303).

are utopian. Since individuals are allowed to create infinite worlds and join the ones that best fit with their preferences, the possible-worlds model qualifies as meta-utopian.

The possible-world model is supposed orient our investigation toward the best possible institutional arrangement, at the meta-community level, within the real world. Particularly, given that the fact of diversity prevents us from designing a unique utopian community that would suit everyone's preferences, we are to imagine an institutional arrangement, at the meta-community level, that would best approximate the model. Such an institutional arrangement would constitute the framework for utopia, as it would bring about a state of affairs in which "people are at liberty to join together voluntarily to pursue and attempt to realize their own vision of the good life in the ideal community but where no one can impose his own utopian vision upon others" (Nozick, 1974, p. 312).

Clearly, the possible-worlds model and the real world show important dis-analogies. Bader (2017), in particular, identifies two: transaction costs and coercion. The challenge posed by transaction costs is explained with reference to the fact that we cannot create communities and their inhabitants, within the real world, in the same way in which we do within the model. Creating a community and gathering other people in order to make them join is costly and the required information is often not at our disposal. Similarly, it is costly to find already existing communities that would constitute a better fit with our own conception of the good. Hence, within the real world, individuals shall converge on communities that would not correspond to their ideal ones as those are either not existent and too costly to create, or because the costs of finding a better match, among already existing communities, would overweight the expected benefits. The challenge posed by coercion, on the other hand, consists in that, unlike the possible-worlds model, in the real world some communities are stable not because their members are unwilling to search for more fitting communities or cannot find a better fit, but rather because they are coerced into staying. As such, the real world exhibits a large amount of falsely stable

communities, in that stability is driven by coercion.

However, according to Nozick, such discrepancies do not undermine the normative relevance of the model. In fact, even under the consideration of these disanalogies, realising the institutional framework that best approximates the model would constitute the best possible scenario:

Our consideration of alternative arguments for the framework, and discussion of objections to it, will make a case for (but not establish) the proposition that it would be better to realize the framework than to realize alternatives even more divergent from the possible-worlds model than it. We should note here that some of the ways the framework diverges from the possible-worlds model, though making the framework less desirable than the possible-worlds model, leave it more desirable than any other actually realizable situation. For example, in the actual operation of the framework there will be only a limited number of communities, so that for many people, no one community will exactly match their values and the weighting they give them. Under the framework, each individual chooses to live in the actual community which (putting it roughly) comes closest to realizing what is most important to him. But the problem about no community exactly fitting someone's values arises only because people disagree about their values and their weighting. (If there were no disagreement, there would be enough other people to populate the exactly desired community.) So there will be no way to satisfy all of the values of more than one person, if only one set of values can be satisfied. Other persons will have their values more or less closely satisfied. But if there is a diverse range of communities, then (putting it roughly) more persons will be able to come closer to how they wish to live, than if there is only one kind of community. (Nozick, 1974, p. 308-309).

One main question arises about how are we to evaluate and compare insti-

tutional arrangements, at the meta-community level, in order to establish their betterness in approximating the model. In this regard, two elements seem to play a major role in assessing institutional arrangements: fittingness and the number of communities that would be created under different sets of institutions. I refer to fittingness as a thick concept capturing the coincidence between the sets of goals/conceptions of the good individuals possess and the ones that the institutional structure of a certain community would facilitate in attaining. The significance of the number of communities, on the other hand, is derivative as it rests on the fact that there is no single set of institutions that would constitute a plausible match with everyone's conceptions of the good or that best satisfies everyone's preferences/goals. In other words, in a world characterised by incompatible conceptions of the good, the more communities, the better.

Clearly, for the number of communities to play a significant role, communities have to be fitting and differently so from one another. This is crucial insofar as communities with a low degree of fittingness would be unstable in the same way imagined worlds would be, in that their inhabitants would prefer to join another existing association or to create another one. Identical communities, on the other hand, would not count as relevant options. In fact, given that moving from one community to another is not costless, members of a community would prefer remaining in their own community rather than moving to another one offering the same degree of fittingness.

In this regard, one might object that the relevance of the number of communities is somewhat parasitic to fittingness and difference. In fact, a state of affairs containing n communities with low degree of fittingness would hardly be defined as a better approximation of the possible-worlds model compared to a scenario in which there is a single highly fitting community. However, counterexamples of this sort make sense only in a context characterised by low heterogeneity of goals. When individual goals reach a certain degree of incompatibility, indeed, the idea of a single highly fitting community becomes implausible as it is hard to find overlapping sets of goals among deeply diverse perspectives. This means that to approximate the possible-worlds model in a context of high heterogeneity

of goals, requires increasingly high numbers of fitting communities.

The relevance of the number of communities is also explicitly made clear by Nozick:

For example, in the actual operation of the framework there will be only a limited number of communities, so that for many people, no one community will exactly match their values and the weighting they give them. Under the framework each individual chooses to live in the actual community which (putting it roughly) comes closest to realizing what is most important to him. But the problem about no community exactly fitting someone's values arises only because people disagree about their values and their weighting. (If there were no disagreement, there would be enough other people to populate the exactly desired community.) So there will be no way to satisfy all the values of more than one person, if only one set of values can be satisfied. Other persons will have their values more or less closely satisfied. But if there is a diverse range of communities, then (putting it roughly) more persons will be able to come closer to how they wish to live, than if there is only one kind of community. (Nozick, 1974, p. 309).

In this regard, it seems that plausible candidates as institutional arrangements that would count as realising the framework for utopia would have to be assessed by looking at their ability in maximising these two operationalised features. Hence, the institutional arrangement that would realise the framework for utopia is the one that is able to allow for the creation of as many fitting communities as possible.

The Minimal State as a Meta-Utopia

Nozick's aim is to show that the minimal state could be justified as the framework for utopia, that is as the basic set of institutions that would maximise

our chances to create as many fitting communities as possible. Therefore, the minimal state, more than an institutional arrangement that is meant to work within a community, would represent a basic set of institutions that is meant to apply at the meta-community level. Such a set, thus, is meant to define an initial position from which persons could voluntarily depart by creating new communities and joining already existing ones that would constitute a better fit with their sets of preferences and goals. The normative value of the minimal state, under the justificatory idea of utopia, rests on its ability in creating the conditions for the emergence of countless communities that would aptly respond to the challenges posed by people's diversity.

As Bader (2017) notices, there are three main claims emerging from Nozick's discussion of the minimal state under the justificatory idea of utopia.

1. the minimal state is inspiring because it is a meta-utopia that counts as realisation of Nozick's possible-worlds model of utopia.
2. the minimal state is the common ground of all possible utopian conceptions and as such can be universally endorsed.
3. the minimal state is the best means for approximating utopia.

Claim (1) and (3) are evidently linked, in that (1) asserts the normative standing of the institutional setting that best approximates the possible-worlds model, and (3) affirms that such an institutional arrangement is to be found in Nozick's account of minimal state. Claim (2), on the other hand, seems to spell out another justificatory feature of the minimal state. According to it, the minimal state seems to possess a certain normative standing not merely by virtue of best approximating the possible-worlds model, within the real world, but also by virtue of the fact that it allegedly constitutes a common ground of all utopian conceptions, in that it defines the maximal institutional structure from which each and every individual could depart in order to form their own utopian world.⁹⁴

⁹⁴In fact, Bader (2017) argues: "Any institutional structure that is more extensive than the minimal state will fail to be neutral and will privilege certain views of utopia. Any

In this regard, it is unclear whether such a feature is meant to provide an alternative justificatory framework for the minimal state, or whether it is supposed to pose a constraint on plausible candidate institutional arrangements that satisfy (3). In particular, it is unclear whether an eventual institutional setting that satisfies (3) but fails in satisfying (2) could be considered justified. However, before we get to this particular problem, I suggest that we start by analysing (3), as our aim is to look for an institutional arrangement, at the meta-community level, that would be able to accommodate for as many preferences and goals as possible, through the creation of various communities, in the attempt to secure the conditions for cooperation given the fact of diversity.⁹⁵ As such, Nozick's account of utopia seems promising, in that the institutional setting that best approximates the possible-worlds model, seems exactly the sort of arrangement we are looking for.

As we have seen, claim (3) purports that the minimal state ultimately represents such an institutional arrangement. In fact, the underlying thought is that the distribution of rights operating within the minimal state allows for the creation of a fertile soil where the opportunities of creating new fitting communities and experimenting various ways of living would be maximised. Nozick's argument in support of (3) can be framed as follows:

(i) The possible-worlds model cannot be meaningfully approximated by designing communities in an apriori fashion, as this would require a certain amount of knowledge that theorists and planners cannot plausibly possess.

(ii) In order to approximate the possible-worlds model we need filtering devices that are able to use knowledge dispersed among individuals.

(iii) Individual consent constitutes such a filtering device.

restrictions on what associations can be like will ensure that there will be some people who do not consider their association to be the best possible association, but would rather be members of an association that would not be subject to those restrictions. This means that any more extensive state cannot in principle be the best for each of us but can only be the best for those adhering to the utopias that are thus privileged. More-than-minimal states will thus only be accidentally compatible with utopia, namely only in those situations in which no one should wish to be a member of an association that is subject to the restrictions of a more-than-minimal state", pp. 267-268.

⁹⁵I shall return on (2) at the end of the chapter.

(iii) The minimal state protects the bundle of rights connected to individual consent and is, thus, the best means to approximate the possible-worlds model.

The reader will be familiar with (i), as it is borrowed from Hayek's reflection on the knowledge problem, which we have previously highlighted. In fact, the underlying idea is that any attempt to design complex social orders, on the basis of people's preferences, goals and conceptions of the good, requires knowledge that is dispersed among individuals and that cannot be plausibly collected by theorists or planners:

Given the enormous complexity of man, his many desires, aspirations, impulses, talents, mistakes, loves, sillinesses, given the thickness of his intertwined and interrelated levels, facets, relationships (compare the thinness of the social scientists' description of man to that of novelist), and given the complexity of interpersonal institutions and relationships, and the complexity of coordination of the actions of many people, it is enormous unlikely that, even if there were one ideal pattern for society, it could be arrived at in this a priori (relative to current knowledge) fashion. And even supposing that some great genius did come along with the blueprint, who could have confidence that it would work out well? (Nozick, 1974, p. 313).

Hence, any community devised by human planners would likely turn out as defective.

In this regard, Nozick argues that, under conditions of limited knowledge, any plausible attempt to design institutional arrangement aiming to maximise fittingness, at the meta-community level, has to rely on filtering devices that are able to use the required local knowledge (ii). Filtering devices have two tasks to perform: on the one hand, they must rule out unfitting options from a large set of alternatives;⁹⁶ on the other, they should propose new candidate institutional

⁹⁶See, Nozick (1974): "Filter devices involve a process which eliminates (filters out) many from a large set of alternatives. The two key determinants of the end result(s) are the particular nature of the filtering out process (and what qualities it selects against) and the particular

arrangements that could be tried out, and eventually adopted or dismissed.⁹⁷ Nozick's idea is that filtering devices, in order to arrive at a desired end product of which we do not know exactly the nature (e. g. fitting communities), are easier to design than the actual community. In fact, the amount of knowledge required to design a filtering device is much less compared to that required in designing a community able to accommodate for various preferences, goals and conceptions of the good.

Only specific persons know in full details their present specific and more general aims and, as such, we are in need to devise a filtering mechanism that is able to use such dispersed knowledge in ruling out certain alternatives and proposing new plausible candidates.

Nozick identifies such a filtering mechanism in individual consent (iii). In fact, the idea is that only individuals can meaningfully filter out unfitting options and propose plausible candidate institutional arrangements in an apriori fashion.⁹⁸ However, the filtering processes operating through individual consent should not merely concern the 'design' phase, but should rather extend to actual experiments in living. In fact, his claim is that apriori designs, although devised at the individual level, through the use of local knowledge, cannot determine at the very outset which proposals are successful, as institutional arrangements must be also tried out.⁹⁹ Therefore, we need an institutional arrangement, at

nature of the set of alternatives it operates upon (and how this set is generated). Filtering processes are especially appropriate for designers having limited knowledge who do not know precisely the nature of a desired end product", p. 314.

⁹⁷See, Nozick (1974):"Furthermore, if the filtering process is of the type that involves a variable method of generating new candidates, so that their qualities improves as the quality of the members remaining after previous filtering operations improves, and it also involves a variable filter that becomes more selective as the quality of the candidates sent into it improves (that is, it rejects some candidates which previously had passed successfully through the filter), then one legitimately may expect that the merits of what will remain after long and continued operation of the process will be very high indeed", p. 314.

⁹⁸See Nozick (1974):"Any group of people may devise a pattern and attempt to persuade others to participate in the adventure of a community in that pattern. Visionaries and crackpots, maniacs and saints, monks and libertines, capitalists and communists and participatory democrats, proponents of phalanxes (Fourier), palaces of labor (Flora Tristan), villages of unity and cooperation (Owen), mutualist communities (Proudhon), time stores (Josiah Warren), Bruderhof, kibbutzim, kundalini yoga ashrams, and so forth, may all have their try at building their vision and setting an alluring example" p. 316.

⁹⁹See Nozick (1974):"But one cannot determine in advance which people will come up with the best ideas, and all ideas must tried out (and not merely simulated on a computer) to see

the meta-community level, that allows us to try out proposed alternatives. Such an institutional setting, Nozick suggests, is the minimal state (iii), where:

People try out living in various communities, and they leave or slightly modify the ones they don't like (find defective). Some communities will be abandoned, others will struggle along, others will split, others will flourish, gain members, and be duplicated elsewhere. Each community must win and hold the voluntary adherence of its members. No pattern is imposed on everyone, and the result will be one pattern if and only if everyone voluntarily chooses to live in accordance with that pattern of community. (Nozick, 1974, p. 316).

The claim is that the distribution of rights operating within the minimal state, *activates* the filtering processes involved by one's exercise of consent. As such, the end result(s) obtained through the filtering processes is a state of affairs at which individuals arrive by means of their decisions, which are based on their own local knowledge. Each person considers a wide range of possible alternative settings and filters some out; individuals choose to experiment certain communities, some of which will be abandoned or progressively improved. The end result is, in Nozick's mind, an archipelago of communities whose fittingness is secured by such filtering processes.

Is the Minimal State the best means to approximate Utopia?

The question is whether the minimal state would actually constitute the institutional arrangement, at the meta-community level, that best approximates the possible-worlds model. In order to support such a claim, one has to show that the distribution of rights operating within the minimal state would outperform other distributions in maximising the chances for the creation of various fitting

how they will work. And some ideas will come only as we are (*post facto*) trying to describe what patterns have evolved from the spontaneous coordination of the actions of many people", pp. 315-316.

communities. This is to say that the filtering processes activated by the distribution of rights purported by the minimal state would be better than other filtering devices that may be purported by different institutional arrangements.

At a first glance, Nozick's story seems perfectly plausible: individuals start from an institutional arrangement constituted by the minimal state; they are free to propose and try out various institutional arrangements that adopt different distributions of rights; after trying them out they could either reconsider their choices and try new adventures, or continue living within them. Although distribution of rights, within certain communities, may be such that individual consent no longer operates as a filtering process in order to improve the community from within, individuals always possess a right to exit, such that if they come to be unsatisfied with the institutional arrangement operating within their communities, they can always leave.

This would ideally answer the concerns that I have presented within the previous chapters about the need for a sufficientarian distribution of social-environment shaping rights and with regard to norms that impede adaptation to members of the community's changing structure of preferences and goals. In fact, the right of exit equips everyone with the possibility of abandoning the community if they come to realise that they have nothing to gain from complying with its cooperative scheme.¹⁰⁰

However, I suggest, one main obstacle should make us carefully consider Nozick's theory. It concerns a crucial disanalogy between the possible-worlds model and the real world, that is the existence of transaction costs. As we have noted earlier, one cannot create a community with the same ease of the possible-worlds model, where individuals can create a world by simply imagining it. Nor one can leave a community and join another existing one in the same way in which one joins an already existing world. Creating a community carries costs:

¹⁰⁰Surely, communities in which consent still operates as a filtering device and where social-environment shaping rights are distributed such that adaptation to changing preferences and goals is fostered, would be more stable as members would be able to continuously renegotiate their terms of cooperation; but this may not be worrisome as members who do not stand to gain from participating to the cooperative enterprise may seek to try other communities.

in fact, the design process requires efforts, gathering and persuading people to embark in a new cooperative enterprise does not come for free either, and trying out new cooperative arrangements is always risky, for one knows that no matter how appealing a certain idea is, it must always be tried out in order to evaluate its potential. Similarly, abandoning a community and joining an already existing one cannot be regarded as a cost-free process: fitting communities have to be identified and the searching process requires knowledge that may be hard to obtain.

Such worries do not necessarily undermine the validity of the theory, as Nozick's minimal state could still represent the best approximation of the possible-worlds model. However, the real challenge is not merely constituted by the existence of transaction costs, but rather by the emergence of patterns, caused by transaction costs, which severely limit the possibility of creating new communities.

This particular problem is highlighted by the fact that the creation of a new community is perfectly captured by a public good game. In fact, as for the production of public goods, the creation of a community requires joint efforts, both in the design process, and in the phase of building its structure (e.g. in the identification and acquisition of the territory, in building its infrastructures, and so forth); moreover, as for public goods, the communities imagined by Nozick are supposed to produce benefits that are not excludable to those who did not contribute in the first place, as communities could be joined at any time.

In public good games, the provision of a certain good requires joint efforts from its participants. The costs yielded by individual efforts would be outweighed by the benefits generated once the good is created; however, the structure of the game is such that each player has an incentive to forego on the initial contribution, and, as a result, the good is not created. In fact, imagine $n=6$ players are equipped with an initial endowment of 10£. Each player may either invest part or the whole endowment in a common pool that would triple the returns (the multiplier must be greater than one but less than n), or keep it. The investment's returns will be equally distributed among the players, irrespective

of their contributions.

If all players invest their total endowment, the total returns would be of 180£, which would grant to each players an amount of 30£. However, rational players have an incentive to free-ride on other players' contributions, as each invested pound would yield a return of only 50 cents.¹⁰¹ As such, the dominant strategy of each rational player is to avoid contribution, and the good is not generated.

Rational choice theory, thus, indicates that the minimal state's route to approximate the possible-worlds model is not merely affected by transaction costs yielded by the creation of new cooperative enterprises, but rather subject to patterns that would go against the creation of new communities. In fact, rational individuals would lack any incentive to contribute to the creation of fitting communities, as their dominant choice is always to free ride. Clearly, this particular problem could be overcome by making communities excludable, such that those who did not contribute in the first place would be prevented from joining the community at a later stage. However, such a solution would undermine the appeal of the theory from a different angle, as individuals would be required to create a new community every time they are unsatisfied with their current institutional setting, instead of searching for better alternatives.

The challenge posed by patterns against the creation of new communities is rather fatal, as it shows that the status quo (i. e. the minimal state) would be privileged not merely on the ground of transaction costs or risks connected to experimenting new cooperative enterprises, but also by the fact that rationality demands not to invest efforts in the creation of a new community even when individuals are certain about its returns. This is to say that individuals would eventually privilege remaining in the minimal state even when a more fitting option is at hand.¹⁰²

¹⁰¹The payoffs of free-riders and contributors are obtained as follows: $P_{fr} = (rn_c c)/N$, $P_c = P_d - c$, where r denotes the multiplier, n_c the number of contributors, and c the amount of the contributions.

¹⁰²The public good game objection, I suggest, already embeds other similar remarks that could be advanced against Nozick's theory: for instance, one may argue that, normally, those who are unsatisfied with their current institutional arrangements are the ones that are left

Therefore, it seems implausible that the minimal state would represent the best means toward approximating the possible-worlds model, as spontaneously arising patterns tend to drive individuals toward staying under the initial institutional setting, regardless of the possibility of creating more fitting communities.

Another, although less dramatic, challenge to the minimal state's route toward approximating the possible-worlds model is represented by social norms increasing costs for leaving communities or creating new ones. Particularly, the worry I wish to highlight is that transaction costs are not merely denoted by the risks associated with trying out new cooperative enterprises or by the efforts required in order to create a new community or in searching for an already existing and more fitting option, as costs may also be hiding behind informal sanctions.

The underlying idea is that a right to exit a community or a right to gather and persuade other people to create a new cooperative enterprise may also be affected by informal sanctions directed toward those who undertake such strategies. As such, even if individuals are not formally coerced into staying or into abstaining from the attempt to create a new communities, they could be informally sanctioned when trying to do so. Essentially, even assuming that patterns affecting the creation of new communities, illustrated by the public good game, can be overcome, informal sanctions may stand in the way of our projects, making increasingly costly, and eventually not appealing, the strategy of embarking into new cooperative enterprises, as costs yielded by informal sanctions may outweigh the benefits expected by creating or joining more fitting communities.

The institutional arrangements operating within the minimal state (and within other eventually existing communities) do not contain any provision aimed at dismantling (or preventing the emergence of) norms yielding such

poor by the distribution of rights operating within their communities, and, as such, lack the resources required to build a new community or join an already existing ones. In fact, the public good objection shows that no matter how resources are distributed, the status quo would always be privileged as individuals have incentives to forego on efforts required to build a new community.

sanctions, although sanctions could severely undermine the possibility of embarking into new cooperative enterprises. Such a problem is less fatal to Nozick's theory insofar as it is impossible to identify, in an apriori fashion, a pattern that would drive the emergence of norms going systematically against these strategies, as the evolution of norms, by virtue of being tightly-coupled to initial conditions and individual attitudes, is rather unpredictable. However, the emergence of such norms is still a possibility and would pervasively affect institutional arrangements attempting to approximate the possible-worlds model if no provisions are devised.

The More-Than-Minimal State in Approximating Utopia

In the previous section, I argued that the minimal state is a rather unappealing means toward approximating the possible-worlds model and, thus, the idea of Utopia. This, we have seen, is not merely due to the existence of transaction costs, shaping important disanalogies between the model and real world, but rather to the fact that such costs drive the emergence of patterns that would strongly privilege the status quo (i. e. the minimal state), even when better options are at hand. Moreover, the absence of provisions against norms that would increase the costs for the creation of new communities or for the abandonment of unfitting ones, poses another obstacle in imagining the minimal state as a fertile soil where new cooperative arrangements would be continuously generated and where individuals are able to freely move from one community to another, in the attempt to find a better match.

These particular problems, I suggest, do not undermine the normative standing of the framework for utopia. In fact, Nozick's aim of finding the institutional arrangement, operating at the meta-community level, that best approximates the possible-worlds model remains appealing. The appeal, I argue, comes from the fact that our theorising around institutional arrangements at the community level cannot possibly accommodate for the fact of diversity, as pluralistic societies will always face problems connected to non-compossibility of people's pref-

erences, goals and conceptions of the good. In this regard, political theories that wish to take diversity seriously and aim to accommodate for non-compossible conceptions of the good, would look at Nozick's possible-worlds model as the ideal scenario to approximate.

The amended version of the theory of the spontaneous order I have proposed within the previous chapters is by no means an exception, as we have seen that preferences for flexibility cannot accommodate for persistent and ineradicable disagreement stemming from non-compossible preferences and goals. Hence, the search for a framework for utopia, that is for an institutional arrangements, at the meta-community level, that best approximates the possible-worlds model, concerns us inasmuch as it concerns all theories aiming to create the conditions for cooperation, in a world characterised by ineradicable diversity and disagreement. In this regard, the highlighted challenges that the minimal state faces in the route of approximating the possible-world model are not meant to rule out the enterprise of finding the institutional arrangement that best approximates the model, but rather to offer a few improvements to Nozick's proposal.

In fact, Nozick's design of the minimal state as a means toward approximating the possible-worlds model contains crucial insights that must kept. Two key aspects, in particular, deserve our attention. First, Nozick offers an apt critique to theories aiming to leave in the hands of theorists and planners the task of designing, in full details, fitting communities. In fact, he rightly observes that theorists and planners lack the local knowledge required to design, in an apriori fashion, cooperative enterprises that would fit with people's preferences, goals and conceptions of the good, and that, as such, any community so devised would be largely defective. Second, Nozick's discussion on the necessity of filtering devices in generating fitting communities is, in many ways, illuminating. In fact, it sheds lights on that devising a filtering device able to drive toward the creation of fitting communities, through the use of knowledge dispersed among individuals, is a feasible enterprise that requires much less knowledge than the process of designing communities from scratch. Moreover, he convincingly argues that individual consent still represents the best filtering device in bringing

about an end result (i. e. fitting communities) of which we do not know the precise nature. Equally important, the idea of employing individual consent as the main filtering device allows us to respond to disagreement endogenously emerging from changes in members of the community's structure of preferences and goals, and not merely to static disagreement. In fact, the right to exit existing communities and embarking in new cooperative adventures allows individuals to continuously find better matches, in light of the dynamic character of their structure of preferences.

The challenges I have highlighted, though, point toward amendments of the institutional settings Nozick envisages as the best means toward approximating utopia. Once again, two desiderata seem to emerge. First, the difficulties posed by collective action problems, illustrated by the public good game, invite a distribution of rights able to overcome the incentive to defect in the process of creating of new communities. Second, the challenge posed by informal sanctions, eventually increasing the costs for generating new cooperative enterprises or joining others, invites us to design provisions aiming at dismantling (or preventing the emergence of) norms underpinning these sanctions.

Such desiderata seem to depict an institutional arrangement, at the meta-community level, that goes beyond Nozick's formulation of the minimal state. In fact, they seem to call for a more-than-minimal state that, on the one hand, preserves individual consent as the main filtering device toward the creation of new communities, and that, on the other, distributes rights in a way that would overcome the highlighted challenges. In particular, they seem to call for distributions of rights able to foster our ability to generate new communities, through the design of incentives that invite us to embark in new and more fitting cooperative enterprises, when they are at our hand; and for an effort toward dismantling or preventing the emergence of norms that would affect our project to continuously look for better institutional arrangements.

The proposal of a more-than-minimal state as the institutional arrangement, operating at the meta-community level, that attempts to approximate the possible-worlds model, though, faces a plausible objection we have previ-

ously highlighted. Particularly, Nozick seems to argue that the minimal state is justified not merely on the ground of its alleged ability in best approximating utopia, but also on the ground that it constitutes the maximal institutional structure compatible with the idea of utopia. The underlying idea is that the minimal state represents the only institutional setting that does not rule out any possible utopian conception, and that, as such, is the only one that could be universally endorsed at the meta-community level. More-than-minimal states, indeed, while allowing for experimentation, would pose restrictions on what communities should be like, and, as such, cannot be universally endorsed.

In particular, Bader argues that:

Any institutional structure that is more extensive than the minimal state will fail to be neutral and will privilege certain views of utopia. Any restrictions on what associations can be like will ensure that there will be some people who do not consider their association to be the best possible association, but would rather be members of an association that would not be subject to those restrictions. This means that any more extensive state cannot in principle be the best for each of us but can only be the best for those adhering to the utopias that are thus privileged. More-than-minimal states will thus only be accidentally compatible with utopia, namely only in those situations in which no one should wish to be a member of an association that is subject to the restrictions of a more-than-minimal state. (Bader, 2017, pp. 267-268).

As noted earlier in the chapter, this particular aspect poses an interesting question about whether we should rule out institutional arrangements that, although better suited in approximating the possible-worlds model, by virtue of better enhancing experimentation, cannot be universally endorsed due to posing restrictions on what communities should be like. In this regard, I suggest, ruling out more-than-minimal states on such a ground, seems to betray Nozick's aim of looking for the institutional arrangements that best approximate utopia, as

it would restrict the set of options to the minimal state at the very outset.

However, there is another compelling response we could offer in defence of more-than-minimal states. Particularly, we might object that Nozick's design of the minimal state is not immune to the criticism of privileging certain conceptions of utopia over others, as collective action problems seem to indicate that individuals would rationally be inclined toward the preservation of the status quo, given the incentive to defect in building new cooperative enterprises. Although such a tendency does not, in principle, rule out any conception of utopia, it severely undermines the emergence of communities that would depart from the minimal state.

Therefore, ruling out institutional arrangements on the basis of their lack of neutrality, with respect to other conceptions of utopia, could lead us to rule out also the minimal state from the set of plausible options, leaving us with nothing more than anarchy at the meta-community level.

Conclusion

Concluding, let me summarise the two main achievements of the present chapter.

First, I have attempted to show that, since preferences for flexibility are likely to fail in accommodating for all sorts of preferences and goals, our theorising cannot be merely focused on what institutional arrangements should look like within particular communities.

The fact of diversity, indeed, invites us to consider how institutional settings should look like at the meta-community level, in the attempt to facilitate the creation of cooperative enterprises that would be able to respond to the challenges posed by diverse and non-compossible preferences and goals.

In this regard, Nozick's account of utopia provides a framework in order to cast the terms of our discussion about diversity and disagreement. In fact, it sets the main desideratum for institutional arrangements operating at the meta-community level, that is to create the conditions of cooperation, through maximising the chances for the emergence of various fitting communities.

Second, I have attempted to show that Nozick's proposed institutional settings, at the meta-community level, fail in approximating the account of utopia he has in mind. In fact, although his proposed justificatory framework aptly allows us to respond to the challenge posed by persistent disagreement, the minimal state faces crucial challenges in approximating the possible-worlds model. In particular, it faces collective action problems that severely constrain people's ability to create new communities.

In this regard, I have proposed two main amendments that seem to call for a more-than-minimal state as the institutional arrangement, at the meta-community level, that best approximates the possible-worlds model. First, distributions of rights should generate incentives for the creation of new communities, in order to overcome collective action problems; second, provisions in order to dismantle norms underpinning sanctions for the creation of new communities and for the abandonment of unfitting ones should be devised.

Such two desiderata are not spelled out in full details. However, I believe that they would constitute a plausible starting point for policy-makers aiming at designing institutional arrangements, at the meta-community level, that would create the conditions for the creation of as many fitting communities as possible, in order to aptly respond to the challenges posed by diversity and resulting disagreement.

Conclusion

Perhaps, one good way to start the conclusion of the present work is to reiterate John Stuart Mill's quote that accompanied its developments during these years:

An individual Jesuit is to the utmost degree of abasement the slave of his order though the order itself exists for the collective power and importance of its members. (Mill, 2002 [1859], p. 95)

In the introduction, I have suggested that this quote nicely embodies the main assumptions and the main aims of this thesis. In particular, it captures the idea of starting our theorising about political institutions from the preferences and goals of our fellow members of our communities, whose participation to our cooperative enterprises is what ultimately provides meaning to our social orders, and it captures the aim of theorising about political institutions in an effort to design terms of cooperation that we all, qua main protagonists of our communities, would find advantageous.

More than that, I suggest, it captures our astonishment in realising that social orders, despite existing for and because of their members, continuously fail in responding to members' preferences and goals. In fact, they spectacularly fail in creating the conditions for peaceful cooperation among members who legitimately uphold different conceptions of the good, who exhibit widely different aims, who would embark in widely different enterprises, without necessarily posing a threat to each other.

Perhaps, our disagreement about what political institutions should look like is ineradicable, and, as such, social orders' failure in accommodating for our diversity is bound to persist. Perhaps, indeed, our beliefs, our conceptions of the good life, our preferences and goals, are ultimately non-compossible and underpin widely different sets of institutional arrangements, such that a convergence on a unique system of rules is implausible. This is, we have seen, Nozick's standpoint in delineating the framework for Utopia. In fact, Nozick forcefully claims

that persons, who exhibit a wide variety of preferences, goals and conceptions of the good cannot plausibly converge on a shared set of institutions that they regard as utopian. Individuals, he suggests, would describe the fine-grained features of their ideal communities in widely different fashions, and theorists are ultimately unable to design political institutions that would accommodate for everyone's standpoint:

Is there really one kind of life which is best for each of these people? Imagine all of them living in any utopia you've ever seen described in detail. Try to describe the society which would be best for all of these persons to live in. Would it be agricultural or urban? Of great material luxury or of austerity with basic needs satisfied? What would relations between the sexes be like? Would there be any institution similar to marriage? Would it be monogamous? (Nozick, 1974, p. 310).

If we take Nozick's claim at face-value, there is no way to respond to the problem of disagreement stemming from our diversity. In fact, we can hardly think about Frank Sinatra, Buddha, Socrates and many others living under the same set of political institutions and being satisfied by the terms of cooperation that such institutional arrangements design for them, as they would plausibly answer to Nozick's fine-grained questions by defining specific features of their ideal communities which are ultimately non-compossible.

Perhaps, though, most failures of our social orders, in creating the conditions for peaceful cooperation among widely different persons, rest precisely on that we theorise about political institutions *as if* we could design specific terms of cooperation in ways that would respond to Nozick's questions. Questions to which, I suggest, we would be largely unable to answer on our own, in designing our ideal community. This thought experiment is particularly revealing: what would our ideal community look like? Would it be adventurous? Would it be monogamous? Which educational paths would people engage with? Which sports would people play? My straightforward answer to these questions is: I

do not know, as I cannot specify in such a fully detailed fashion the constitutive features of my preferred terms of cooperation. And, perhaps, such rock-solid uncertainties are shared by most fellow members of our cooperative enterprises. Furthermore, one might ask, what would our ideal community look like ten years from now? Would it coincide with the one we attempt to design right now?

Our uncertainty is not limited to what our ideal community would look like, as it also extends to what our own fine-grained structure of preferences and goals would be like ten years from now. Would we like to be married? Would we like our lives to be adventurous? Would we like to play sports? And which sports? These questions are not redundant, in that our ideal community and our own structure of preferences may plausibly not coincide. In fact, we can hold on to our conceptions of the good by leading a life that is consistent with a certain structure of preferences and, yet, prefer to be surrounded by persons who are different from us: who would never engage into long lasting relationships, who would like to play curling and who would like to study astrology.

Unsurprisingly enough, many of us would be unable to predict their own structure of preferences and goals ten years from now. And if we look backwards, one can realise that our inability to grasp future identities is well grounded, as ten years ago we would hardly have predicted the fine-grained structure of preferences we exhibit here and now. Sure, we have long term 'ground projects', to borrow Bernard Williams' words,¹⁰³ and our integrity is shaped by acting consistently with what those projects demand. Yet, our projects do not exist outside our structure of preferences and goals, and if we hold on to them, in spite of the dynamic structures of our preferences and goals, it is because they are general enough to be robust to changes in our own identities, or because we are able to shape specific instances of these projects in light of the changes to our own circumstances.

Let me offer two examples in order to illustrate these particular aspects. When Betty starts her Ph.D. in philosophy she considers it a fundamental project. In fact, Betty sees it as the obvious route to her academic career.

¹⁰³Williams and Smart (1973).

A career to which she is very much committed, as she loves intellectually stimulating environments and can hardly imagine her life without engaging in philosophical discussions with her colleagues. Yet, at some point, Betty becomes frustrated by what academic life looks like. For instance, she becomes unsatisfied by the increasing administrative burdens lying on academics, and fails to grasp the meaningfulness of criteria shaping one's success within the academic environment.

An academic career, though, is by no means the only career path that a Ph.D. opens up. As such, she finds new reasons to comply with the demands of her project; reasons which are shaped by her new structure of preferences and goals, and to which the course of actions required to completing the Ph.D. still represents an apt response.

Consider now a second example. When Betty starts her Ph.D. she has a specific project in mind: she wants to lay out how widespread diversity undermines the possibility of cooperation in our societies. Yet, while writing the thesis, she realises that her arguments are unsound, as there might be a meaningful way to define the terms of cooperation that could be endorsed from widely different standpoints. Hence, she changes specific instances of her project and attempts to provide different arguments shaped by her updated beliefs.

In both examples, Betty sticks with the Ph.D. program for two different kinds of reasons. In the former, although her beliefs, preferences and goals, change with respect to the opportunity of an academic career, she finds the Ph.D. as a robust path toward pursuing different aims, shaped by her updated identity. In the latter, she sticks with the Ph.D. because she is able to define particular instances of her project in due course, thus preserving her ability to shape the contents of her thesis according to beliefs she comes to possess over the years.

These two examples reveal two different sorts of uncertainties which require two different predictive powers. In fact, I suggest, we can partially grasp which projects would be robust to changes in our structure of preferences, but we are likely to be largely unable to tell how changes in our own specific circumstances

would shape particular instances of our projects, such that we can hardly commit to define such specific instances *ex-ante*. In fact, when we do, perhaps, we fail to realise how such a commitment is likely to undermine our integrity, in that we will be unable to shape specific parts of a project in light of our updated identity.

These considerations, I suggest, are nicely encapsulated by Soren Kierkegaard's claim according to which "life can only be understood backwards, but it must be lived forwards".¹⁰⁴ In fact, we can only look at our fundamental projects through the lenses of our actual structure of preferences, which shape them, and perhaps grasp which enterprises would be robust to changes in our own circumstances. However, our ability to account, *ex ante*, for how we would like to shape specific instances of our projects is ultimately precluded to us. In this regard, a fuller understanding of the fine-grained ideal features of our own life, and of our ideal community, is one we come to possess only by looking backwards.

If we look at Nozick's exercise of assessing the incompatibility of our ideals of social order, in the light of these considerations, we may fail to grasp its relevance. In fact, Nozick is able to tell that there is no plausible set of institutions that would accommodate for Buddha and Frank Sinatra's structures of preferences precisely because he looks backwards to their lives, to what they have achieved, to the sort of enterprises they have engaged with, and to their underlying beliefs that shaped their reasons for undertaking certain courses of actions. Yet, our exercise to theorise about institutions that would accommodate for our various structures of preferences, is not an exercise we undertake by looking backwards. In fact, it is rather something we do for the sake of defining terms of cooperation which are meant to regulate our interactions, in ways that we find appropriate, right now and in the future. Our intention, *qua* members of our community, is not to shape our specific terms of cooperation in light of what we liked, on the sort of enterprises we engaged with, on our structure of preferences we used to exhibit, as cooperation is ultimately a forward looking

¹⁰⁴Kierkegaard (1980) [1844].

enterprise.

One way to verify the plausibility of this argument is precisely to ask whether Buddha and Frank Sinatra, were they living in the same time span, would have arrived at the same conclusion Nozick draws. How would we know? It is already hard to assess, ex-post, the incompatibility of the specific features of their ideal communities, that one should almost certainly give up on the enterprise of assessing whether they would have found common grounds for cooperation. Notice that our difficulties in assessing the compatibility of Buddha and Frank Sinatra's views of the ideal community, does not merely rest on our inability to know what it is like to be either Buddha or Frank Sinatra,¹⁰⁵ but, rather, on that Buddha and Sinatra, like us, had to live their lives forward, although they could only fully understand them by looking backwards. As such, their eventual terms of cooperation would not have been shaped by a full understanding of what they desired, or of the enterprise they would have engaged with, or in the light of the dynamic structures of preferences that accompanied them during their life paths. Their terms of cooperation would have rather been shaped by their partial ability to grasp which projects would have been robust to changes in their structures of preferences, and by their large inability to tell how their dynamic structures of preferences would have guided them to design specific instances of their projects. Our difficulties double if we consider that our structures of preferences are something which we partly shape through our social interactions, such that Buddha and Frank Sinatra eventually incompatible views should be assessed also by taking into account that their preferences are not hermetically sealed from their interactions.

However, what I wish to emphasise is that cooperation is an inherently forward looking enterprise, and that, since our lives can only be (perhaps) fully understood backwards, our attempt to define, at the very outset, our fine-grained terms of cooperation is a rather silly enterprise. In fact, we are plausibly able to define general terms of cooperation that will be consistent with changes in our structure of preferences and goals, as we can partially grasp which ideals

¹⁰⁵Such an argument would rest on Thomas Nagel's considerations laid down in Nagel (1974).

of the good life could be meaningfully relevant to us in the future, but we are largely prevented from laying down in full details, ex-ante, how our dynamic structure of preferences and goals would shape our preferred specific terms of cooperation, such as to define a detailed set of rules that would guide us at each and every step.

These considerations, I suggest, are more or less embedded in Hayek's account of complexity, which we have previously outlined. In fact, these two different kinds of uncertainties are captured by Hayek's claims concerning our ability to grasp the general functioning of how a complex system works, and, perhaps, to isolate which general constitutive features bring about specific outcomes within the system, thus making us able to predict the emergence of general patterns; and our inability to provide a fully detailed account of how a complex systems works, thus preventing us from outlining meaningful predictions on the emergence of specific outcomes.

In this regard, we may aptly see Hayek's theory of the spontaneous order as a response to the high degree of complexity of our social orders, which gives rise to his claim on the impossibility of central planning, and to the need for a set of institutions able to use knowledge that is dispersed among widely diverse individuals. Yet, unlike Hayek, I wish to stress that the complexity of our social orders is not merely determined by our inability to collect all people's specific preferences and goals, but also by people's uncertainty about how their structures of preferences and goals will evolve in the future. In fact, our inability to account for the fine-grained evolution of our structures of preferences and goals adds up to the complexity of our social orders.

Such an uncertainty, I suggest, grounds the rationality of forming preferences for flexibility when facing a social contract bargaining scenario and, in turn, grounds the rationality of the two desiderata I have added to the amended version of Hayek's theory of the spontaneous order. In this regard, if we do not take members of the community's preferences as fixed, our choice about institutional arrangements cannot be merely driven by considerations on how are we to design and devise institutional arrangements able to use knowledge

that is dispersed among various individuals in our society. In fact, we should also account for the fact that individual preferences and aims are subject to a process of continuous change, and that distributions of rights emerging from their decentralised interactions, may bring about states of affairs in which some members of the community are ultimately unable to shape their specific terms of cooperation according to their dynamic structure of preferences.

The idea of preserving the instrumentalist character of our social orders, by correcting distributions of rights that bring about such states of affairs, stems precisely from the consideration that institutional arrangements are not merely supposed to bring about distributive outcomes on the basis of a snapshot which captures our preferences, here and now, but rather to create the best conditions for us to lead a life according to preferences and needs which will continuously unveil during our life paths, and which we are prevented from predicting *ex ante*. In fact, distributive outcomes of all our interactions do not merely fail in carrying the same justificatory power of individual interactions, as they are also a function of interactions over which we have no control; but they also bring about states of affairs capturing our structures of preferences and goals at a given time in the past, which may not be satisfactory from our own updated standpoint.

In a similar fashion, the idea of converging on policies that could plausibly prevent the emergence of social norms, which increase the costs for pursuing our preferred and legitimate courses of actions, stems from the fact that informal sanctions may stand in the way of our ability to continuously shape our social environment by undertaking social environment shaping actions. Social norms, in this regard, are precisely like derivative distributions of rights: they emerge from a wide variety of interactions over which we have no control, and even if they could somehow capture and synthesise our structure of preferences and goals, they would do it with references to given times in the past. As such, I suggest, we have instrumental reasons to dismantle them, or to prevent their emergence, insofar as our cooperative enterprises are ultimately supposed to be forward-looking.

The most relevant message that this thesis wishes to convey consists in that our inability to predict the fine-grained evolution of our structures of preferences and goals invites us to think about institutional arrangements by abstracting from our current set of preferences and goals, and to come to the social contract bargaining table by considering a wider range of aims, whose relevance may unveil to us during our life path. This is precisely the idea behind the concept of preferences for flexibility: since we are bound to a certain degree of uncertainty concerning the sort of enterprises we may be willing to engage with in the future, we have instrumental reasons not to tie our choice about political institutions to fine-grained goals and preferences which are subject to a process of continuous evolution. Preferences for flexibility, thus, are elicited by our rational desire to be able to shape our specific terms of cooperation in due course, in light of our uncertainty about our future identities.

Perhaps, if we take seriously our uncertainty about how our structures of preferences and goals will unveil, and if we consistently imagine bargaining scenarios in which we all assume such a standpoint, most, though not all, of our disagreement will suddenly disappear. In fact, preferences for flexibility would allow for a wider overlapping consensus on a minimal agreement. An agreement which will mostly concern institutional arrangements that would equip us with the ability to shape our social environment in due course, without unwarranted constraints shaped by previous fine-grained preferences and aims.

If we are persuaded by the fact that forming preferences for flexibility is ultimately rational, the amended version of the spontaneous order that I have proposed in this thesis will appear as a plausible solution to a social contract bargaining scenario. In fact, if bargainers are uncertain about their future aims, they have reasons to converge on institutional arrangements which define a large protected sphere of actions that would allow them to shape their specific terms of cooperation in due course. In this regard, the protection of Lockean rights, which Hayek envisages as the best cooperative institutional arrangement, would equip individuals with the possibility of exercising social environment shaping rights (e. g. economic rights, rights of expression, etc.), which would allow them

to define their specific terms of cooperation according to the evolution of their structure of preferences and goals.

However, inasmuch as bargainers, who form preferences for flexibility, would plausibly endorse institutional arrangements defining a large protected sphere of actions, they would also endorse policies that would prevent or correct the emergence of states of affairs in which some members of a community are unable to shape their specific terms of cooperation in due course, according to their structures of preferences and goals. In particular, bargainers would have reasons to endorse a sufficientarian distribution of social environment shaping rights, in order to make sure that nearly all members of the community continue to see their cooperative venture as an enterprise from which they continuously stand to gain.

The considerations on preferences for flexibility, I suggest, define an intermediate position between Derek Parfit, Christine Korsgaard and David Brink's arguments for altruism, and rational egoism.¹⁰⁶ The position is intermediate in that the sort of uncertainty about our future-selves, which I describe, is thick enough to make us care about people's interests as long as such interests could become ours, but not so pervasive to form reasons to weigh all possible interests equally. In fact, I suggest, we may partially grasp which enterprises would be robust to evolution in our structure of preferences, such that we might, ex-ante, rule out certain interests.

In order to illustrate such a difference, let us quote Brink:

If [...] the separateness or diversity of persons is not so fundamental [...] Insofar as distinct individuals are psychologically connected and continuous, each can and should view the other as one who extends her own interests in the same sort of way that her own future self extends her own interests. [...] This means that each should regard the good of those to whom she stands in such relationships as

¹⁰⁶In particular, my understanding of the altruistic position is largely drawn from Parfit (1984), Korsgaard (1989), Brink (1997), whereas I draw from Shaver (1999) for the rational egoist position.

a constituent part of her overall good, just as she should regard the good of her own future self as a constituent part of her overall good. [...] Interpersonal connections and continuity can be found among intimates who interact on a regular basis and help shape each other's mental life; in such relationships, the experiences, beliefs, desires, ideals, and actions of each depend in significant part upon those of the others. (Brink, 1997, pp. 141-143).

Following Parfit and Korsgaard, Brink suggests that if it is an empirical fact that our psychological connections with our future-selves are so weak to the point that the idea of separateness of persons fails to be significant, we have reasons to care about other persons inasmuch as we care about us. This, in turn, would equip us with reasons for weighting all interests equally. However, I find such a claim implausible. The implausibility does not rest on empirical considerations about psychological connections, over which I am largely ignorant, but rather on our inability to think about our lives as ultimately led by different persons at different time spans. Perhaps, Parfit, Korsgaard and Brink empirical assumptions are broadly correct; perhaps, indeed, our future-selves really are, as a matter of fact, strangers to us, but one main problem stands in our way: namely, we are largely unable to assess what is it to think as if our relation to our future-selves is identical to that we have to other persons. In fact, we think about our future-selves through the lenses we possess, here and now, and, as such, we are unable to know what is it to be someone else (or a bat).

Preferences for flexibility, in this regard, are not meant to highlight our complete uncertainty about our future-selves. Perhaps we have no empirical grounds to believe that we have some special connections to our future-selves, and, strictly linked to this aspect, we may have no empirical grounds to believe that the lenses through which we look at our world right now, would make sense to us in the future. However, we cannot get rid of our lenses in the same way in which we get rid of specific goals in eliciting preferences for flexibility.

Therefore, our flexible standpoints in an eventual bargaining scenario would not be perspective-less, and we would always have reasons to exclude some interests and some specific goals from the set of the available options. Perhaps, our inability to get rid of our lenses shall drive us to make mistakes in our general prediction of what shall make sense to us in the future, but there is no other meaningful standpoint from which to assess our terms of cooperation.

This particular fact is what drives the discussion on institutional arrangements at the meta-community level, which closes the thesis. In fact, if preferences for flexibility cannot make us abstract in the way in which Parfit, Korsgaard and Brink think we have reasons to, some disagreement about political institutions is bound to persist. In this regard, cooperation may not be an enterprise we can engage with all together, and under the same set of political institutions, as some would filter out certain features of our community that others may find of value.

Let me conclude this thesis by suggesting that preferences for flexibility, though not exhausting the room for moral and political disagreement, may open up an interesting research path in contractarian theorising. A path that allows us to achieve some level of abstraction, which rational individuals have interests in reaching, while maintaining our authentic lenses in imagining plausible cooperative schemes. In this regard, if we are persuaded by the fact that cooperation is an inherently forward looking enterprise, and if we come to realise that our structures of preferences and goals is dynamic, preferences for flexibility could come to respond to many problems connected to moral and political disagreement. They are not the perfect answer, but a plausible and workable starting point.

Works Cited

- Agonito, R. (1975). Hayek revisited: mind as the process of classification. *Behaviorism*, 3(2), 162-171.
- Allen, P. M., Strathern, M. and Baldwin, J. S. (2007). Complexity and the limits to learning. *Journal of Evolutionary Economics*, 17(4), 401-431.
- Anderson, E. (2017). *Private Government: How Employers Rule Our Lives (and why We Don't Talk about It)*. Princeton: Princeton University Press.
- Arrow, K. J. (1950). A difficulty in the concept of social welfare. *Journal of political economy*, 58(4), 328-346.
- Austin, J. L. (1962). *How to do things with words*. Cambridge: Harvard University Press.
- Axelrod, R. (1986). An evolutionary approach to norms. *American political science review*, 80(4), 1095-1111.
- Bader, R. M. (2011). The framework for Utopia. In: Bader, R. M.; Meadowcroft, J. (Eds.) *The Cambridge companion to Nozick's Anarchy, state, and utopia*, 255-284. Cambridge: Cambridge University Press.
- Bader, R. (2017). Counterfactual Justifications of the State. In D. Sobel, P. Valentyne, and S. Wall (Eds.), *Oxford Studies in Political Philosophy*, 3: 101-131. Oxford: Oxford University Press.
- Bartlett, K. T. and Gulati, M. (2016). Discrimination by Customers. *Iowa Law Review*, 102, 223.
- Basu, K. (1984). The right to give up rights. *Economica*, 51(204), 413-422.
- Basu, K. (1998). Social Norms and the Law. In: P. Newman (Ed.) *The New Palgrave Dictionary of Economics and the Law*. London: Macmillan.
- Baumgarth, W. P. (1978). Hayek and political order: The rule of law. *Journal of libertarian studies*, 2(1), 11-28.
- Bicchieri, C. (1990). Norms of cooperation. *Ethics*, 100(4), 838-861.
- Bicchieri, C., Jeffrey, R., Jeffrey, R. C., Skyrms, B., Adams, E. W., Diaconis, P., Suppes, P. (Eds.). (1997). *The dynamics of norms*. Cambridge: Cambridge University Press.

- Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge: Cambridge University Press.
- Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford: Oxford University Press.
- Binmore, K. (2005). *Natural justice*. Oxford: Oxford University Press.
- Brennan, G., Eriksson, L., Goodin, R. E., and Southwood, N. (2013). *Explaining norms*. Oxford University Press.
- Brink, D. (1986). Utilitarian morality and the personal point of view. *The Journal of Philosophy*, 83(8), 417-438.
- Brink, D. (1997). Self-Love and Altruism. *Social Philosophy and Policy*, 14(1), 122-157.
- Buchanan, J. M. (1954). Social choice, democracy, and free markets. *Journal of Political Economy*, 62(2), 114-123.
- Buchanan, J. M., and Tullock, G. (1962). *The calculus of consent*. Ann Arbor: University of Michigan Press.
- Buchanan, J. M. (1991). *The economics and the ethics of constitutional order*. Ann Arbor: University of Michigan Press.
- Caldwell, B. J. (2004). *Hayek's challenge: An intellectual biography of F.A. Hayek*. Chicago: University of Chicago Press.
- Callard, A. (2018). *Aspiration: The Agency of Becoming*. Oxford: Oxford University Press.
- Caplan, B. (2011). *The myth of the rational voter: Why democracies choose bad policies*. Princeton: Princeton University Press.
- Carr, C. L. (1988). Coercion and freedom. *American Philosophical Quarterly*, 25(1), 59-67.
- Cialdini, R. B., Reno, R. R. and Kallgren, C. A. (1990). A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology*, 58(6), 1015.
- Cialdini, R. B., Kallgren, C. A. and Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In *Advances in experimental social psychology*, 24,

- 201-234. Cambridge: Academic Press.
- Cialdini, R. B., Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske, G. Lindzey (Eds.), *The handbook of social psychology*, 151-192. New York: McGraw-Hill.
- Cohen, G. A. (1978). *Karl Marx's Theory of History: A Defence*. Princeton University Press.
- Cohen, G. A. (1977). Robert Nozick and Wilt Chamberlain: How Patterns Preserve Liberty. *Erkenntnis*, 11(1), 5-23.
- Coleman, J. S. (1990). The emergence of norms. In Hechter, M, Opp, K-D, Wippler, R (eds.) *Social Institutions: Their Emergence, Maintenance and Effects Social institutions: Their emergence, maintenance, and effects*. Berlin: de Gruyter, 35-39.
- Cotsaftis, Michel (2009). What Makes a System Complex? An Approach to Self Organization and Emergence. In M. Aziz-Alaoui, C. Bertelle (eds.), *From System Complexity to Emergent Properties*. New York: Springer, 49-99.
- Coyne, C. J. and Coyne, R. L. (2014). The identity economics of female genital mutilation. *The Journal of Developing Areas*, 48(2), 137-152.
- Cubeddu, R. (1993). *The philosophy of the Austrian school*. London: Routledge.
- Dancy, J. (2004). *Ethics without Principles*. Oxford University Press.
- D'Amico, D. J. and Boettke, P. J. (2010). Making sense out of the sensory order. In: W. Butos (Ed.) *The Social Science of Hayek's 'The Sensory Order': Advances in Austrian Economics*, 13, 357-98. Bingley: Emerald Group Publishing Limited.
- Davies L. (2013, September 26). Pasta firm Barilla boycotted over 'classic family' remarks. *The Guardian*. Retrieved from <https://www.theguardian.com/world/2013/sep/26/pasta-firm-barilla-boycott-gay>
- De Jasay, A. (1989). *Social contract, free ride: a study of the public goods problem*. Oxford: Oxford University Press.
- Dworkin, R. (1975). The Original Position. In: N. Daniels (Ed.) *Reading Rawls*, Oxford: Blackwell, pp. 16-53.

- Dempsey, G. T. (1996). Hayek's Terra Incognita of the Mind. *Southern Journal of Philosophy*, 34 (1):13-41.
- Ebenstein, A. O. (2003). *Friedrich Hayek: A biography*. Chicago: University of Chicago Press.
- Ember, S. (2015, March 16). Remarks on Family by Dolce and Gabbana Bring Swell of Criticism. *The New York Times*. Retrieved from <https://www.nytimes.com/2015/03/17/business/media/remarks-on-family-by-dolce-and-gabbana-bring-swell-of-criticism.html>
- Epstein, R. (1997). *Simple Rules for a Complex World*. Harvard University Press.
- Estlund, D. M. (2009). *Democratic authority: A philosophical framework*. Princeton: Princeton University Press.
- Estlund, D. (2014). Utopophobia. *Philosophy and Public Affairs*, 42(2), 113-134.
- Ferguson, A. (2000) [1767]. *An Essay on Civil Society*. Hildesheim: Georg Olms Velag.
- Ferrara, A. (2008). *The force of the example: Explorations in the paradigm of judgment*. New York: Columbia University Press.
- Fisher, I. (2006, December 20). A Poet Crusades for the Right to Die His Way. *The New York Times*. Retrieved from <https://www.nytimes.com/2006/12/20/world/europe/20welby.html>
- Fiss, O. M. (1985). Free speech and social structure. *Iowa Law Review*, 71, 1405.
- Galston, W. A. (2010). Realism in political theory. *European journal of political theory*, 9(4), 385-411.
- Freeman, S. (2007). *Rawls*. London: Routledge.
- Garfield, A. E. (1997). Promises of Silence: Contract Law and Freedom of Speech. *Cornell Law Review*, 83, 261.
- Gaus, G. F. (2006). Hayek on the evolution of society and mind. In: E. Feser (Ed.), *The Cambridge Companion to Hayek*, 232-258. Cambridge: Cambridge University Press.

- Gaus, G. (2010). *The order of public reason: A theory of freedom and morality in a diverse and bounded world*. Cambridge: Cambridge University Press.
- Gaus, G. (2016). *The tyranny of the ideal: Justice in a diverse society*. Princeton: Princeton University Press.
- Gauthier, D. (1986). *Morals by agreement*. Oxford: Oxford University Press.
- Gell-Mann, M. (1988). Simplicity and complexity in the description of nature. *Engineering and Science*, 51(3), 2-9.
- General Assembly of the United Nations, (1948). *Universal Declaration of Human Rights*. Retrieved from <https://www.refworld.org/docid/3ae6b3712c.html>.
- Gibbard, A. (1974). A Pareto-consistent libertarian claim. *Journal of Economic Theory*, 7(4), 388-410.
- Gick, E., and Gick, W. (2001). F.A. Hayek's theory of mind and theory of cultural evolution revisited: Toward and integrated perspective. *Mind and Society*, 2(1), 149-162.
- Gifford, A. (2007). The knowledge problem, determinism, and The Sensory Order. *The Review of Austrian Economics*, 20(4), 269-291.
- Gray, J. (1981). Hayek on liberty, rights, and justice. *Ethics*, 92(1), 73-84.
- Gray, J. (1982). Hayek and the rebirth of classical liberalism. *Literature of Liberty*, Winter, vol. 5, No. 4.
- Hamlin, A., and Stemplowska, Z. (2012). Theory, ideal theory and the theory of ideals. *Political Studies Review*, 10(1), 48-62.
- Hamowy, R. (1971). Freedom and the Rule of Law in FA Hayek. *Il politico*, 349-377.
- Hardin, R. (1988). Bargaining for justice. *Social Philosophy and Policy*, 5(2), 65-74.
- Hayek, F. A. (1935). The present state of the debate. In: Hayek, F. A., Pierson, N. G., Von, M. L., Halm, G. N., and Barone, E. (Eds.) *Collectivist economic planning*, 201-243. London: Routledge.
- Hayek, F. A. (1944). *The Road to Serfdom*. Sidney: Dymock's Book Arcade.

- Hayek, F. A. (1945). The use of knowledge in society. *The American economic review*, 35(4), 519-530.
- Hayek, F. A. (1952). *The sensory order: An inquiry into the foundations of theoretical psychology*. Chicago: University of Chicago Press.
- Hayek, F. A. (1955). Degrees of explanation. *The British Journal for the Philosophy of Science*, 6(23), 209-225.
- Hayek, F. A. (1960). *The constitution of liberty*. Chicago: University of Chicago Press.
- Hayek, F. A. (1964). The theory of complex phenomena. In M. Bunge (ed.) *The critical approach to science and philosophy*, 332-349. London: MacMillan.
- Hayek, F.A. (1967). Notes on the Evolution of Systems of Rules of Conduct. In: F. A. Hayek (Ed.) *F.A. Hayek, Studies in Philosophy, Politics, and Economics*, 66-81. Chicago: University of Chicago Press.
- Hayek, F. A. (2012). *Law, legislation and liberty: a new statement of the liberal principles of justice and political economy*. London: Routledge.
- Hobbes, T. (2016) [1651]. *Leviathan*. London: Routledge.
- Hodgson, G. M. (1991). Hayek's theory of cultural evolution: an evaluation in the light of Vanberg's critique. *Economics and Philosophy*, 7(1), 67-82.
- Hodgson, G. M. (1994). Hayek, evolution, and spontaneous order. In P. Mirowski (ed.) *Natural images in economic thought*, 408-47. Cambridge: Cambridge University Press.
- Horwitz, S. (2008). Analogous models of complexity: The Austrian theory of capital and Hayek's theory of cognition as adaptive classifying systems. In: R. Koppl (Ed.) *Explorations in Austrian Economics*, 11, 143-166. Bingley: Emerald Group Publishing Limited.
- Hume, D. (1938) [1740]. *An Abstract of a Treatise of Human Nature*. Cambridge: Cambridge University Press Archive.
- Hume, D. (2006) [1751]. *An enquiry concerning the principles of morals*. Oxford: Oxford University Press.
- Kauffman, S. A. (1993). *The origins of order: Self-organization and selection in evolution*. Oxford: Oxford University Press.

- Kelsen, H. (1967) [1934]. *Pure theory of law*. Oakland: University of California Press.
- Keynes, J. M. (1936). *The general theory of employment, interest and money*. London: Macmillan.
- Kierkegaard, S. (1980) [1844]. *The Concept of Anxiety*, trans. Reidar Thomte in collaboration with Albert B. Anderson. Princeton: Princeton University Press.
- Kitschelt, H. (2000). Linkages between citizens and politicians in democratic polities. *Comparative political studies*, 33(6-7), 845-879.
- Koopmans, T. (1962). On Flexibility of Future Preference. No 150, *Cowles Foundation Discussion Papers*, Cowles Foundation for Research in Economics, Yale University.
- Koni Hoffman, L. and Patel, R. N. (2017). *Collective Action on Corruption in Nigeria A Social Norms Approach to Connecting Society and Institutions*. Chatham House Report.
- Korsgaard, C. M. (1989). Personal identity and the unity of agency: A Kantian response to Parfit. *Philosophy and Public Affairs*, 18(2), 101-132.
- Kreps, D. M. (1979). A representation theorem for "preference for flexibility". *Econometrica: Journal of the Econometric Society*, 47(3), 565-577.
- Kukathas, C. (2003). *The liberal archipelago: A theory of diversity and freedom*. Oxford: Oxford University Press.
- Kuznicki, J. (2011). Law, legislation, and local minima: Solving a problem in Hayek's theory of common law judging, with historical examples. *The Review of Austrian Economics* 24(3), 293-309.
- Kymlicka, W. (1995). *Multicultural citizenship: A liberal theory of minority rights*. Oxford: Clarendon Press.
- Ladyman, J., Lambert, J., and Wiesner, K. (2013). What is a complex system?. *European Journal for Philosophy of Science*, 3(1), 33-67.
- Landemore, H. (2012). *Democratic reason: Politics, collective intelligence, and the rule of the many*. Princeton: Princeton University Press.
- Lamond, G. (2000). The coerciveness of law. *Oxford Journal of Legal Studies*, 20(1), 39-62.

- Lewis, D. (1969). *Convention*. Cambridge: Harvard University Press.
- List, C., and Valentini, L. (2016). The methodology of political theory. In: H. Cappelen, T. S. Gendler, and J. Hawthorne (Eds.) *The Oxford handbook of philosophical methodology*, 525-550. Oxford: Oxford University Press.
- Lister, A. (2013). The "mirage" of social justice: Hayek against (and for) Rawls. *Critical Review*, 25(3-4), 409-444.
- Locke, J. (1966) [1689]. *Two treatises of civil government*. Montreal: Delta.
- Lucas, J. R. (1966). *The Principles of Politics*. Oxford: Oxford University Press.
- Lukes, S. (1997). Social justice: the Hayekian challenge. *Critical Review*, 11(1), 65-80.
- Mandeville, B. (1989) [1705]. *The Fable of The Bees: or, Private Vices, Public Benefits*. London: Penguin.
- Marsh, L. (2010). Hayek: Cognitive scientist avant la lettre. In: W. Butos (ed.) *The Social Science of Hayek's 'The Sensory Order': Advances in austrian economics*, 13: 115-155.
- Menger, C. (1963). *Problems of economics and sociology*. Urbana: University of Illinois Press.
- Mill, J. S. (1966) [1859]. On liberty. In: J. M. Robson (ed.) *John Stuart Mill: A selection of his works*. New York: Odyssey Press.
- Miller, D.T. and McFarland, C., (1987). Pluralistic ignorance: When similarity is interpreted as dissimilarity. *Journal of Personality and social Psychology*, 53(2), p. 298.
- Mitchell, M. (2009). *Complexity: A guided tour*. Oxford: Oxford University Press.
- Moehler, M. (2018). *Minimal Morality: A Multilevel Social Contract Theory*. Oxford: Oxford University Press.
- Moore, M. (1996). On Reasonableness. *Journal of Applied Philosophy*, 13: 167-178.
- Muldoon, R. (2016). *Social contract theory for a diverse world: Beyond tolerance*. London: Routledge.

- Nagel, T. (1974). What is it like to be a bat?. *The philosophical review*, 83(4), 435-450.
- Nagel, T. (1989). *The view from nowhere*. Oxford: Oxford University Press.
- Narveson, J. (1988). Justifying a morality. In: D. Odegard (ed.), *Ethics and Justification*. Edmonton: Academic Printing and Publishing, 267-69.
- Narveson, J. (2001). *The libertarian idea*. Peterborough. Broadview Press.
- North, D. C. (1991). *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.
- North, D. C. (2012). *Understanding the process of economic change. Worlds of Capitalism*. London: Routledge.
- Nozick, R. (1969). Coercion. In W. Morgenbesser (ed.), *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*. New York: St Martin's Press, 440-472.
- Nozick, R. (1974). *Anarchy, state, and utopia*. New York: Basic Books.
- Page, S. E. (2010). *Diversity and complexity*. Princeton: Princeton University Press.
- Parfit, D. (1984). *Reasons and persons*. Oxford: Oxford University Press.
- Peltzman, S. (1976). Toward a more general theory of regulation. *The Journal of Law and Economics*, 19(2), 211-240.
- Pettit, P. (2012). *On the people's terms: a republican theory and model of democracy*. Cambridge: Cambridge University Press.
- Piper, A. M. (1986). Instrumentalism, Objectivity, and Moral Justification. *American Philosophical Quarterly*, 23(4), 373-381.
- Posner, E. A. (2000). Law and social norms: The case of tax compliance. *Virginia Law Review*, 86, 1781-1819.
- Posner, R. A. (2003). *Law, Pragmatism, and Democracy*. Cambridge: Harvard University Press.
- Povoledo, E. (2017, December 14). Italy to Allow Living Wills and the Refusal of End-of-Life Care. *The New York Times*. Retrieved from <https://www.nytimes.com/2017/12/14/world/europe/italy-living-will-end-of-life-right-to-die-assisted-suicide.html>

- Putnam, R. D., Leonardi, R., and Nanetti, R. Y. (1994). *Making democracy work: Civic traditions in modern Italy*. Princeton: Princeton university press.
- Railton, P. (1984). Alienation, consequentialism, and the demands of morality. *Philosophy and Public Affairs*, 134-171.
- Rawls, J. (1958). Justice as fairness. *The philosophical review*, 67(2), 164-194.
- Rawls, J. (1971). *A theory of justice*. Cambridge: Harvard University Press.
- Rawls, J. (1982). The basic liberties and their priority. In: S. McMurrin (ed.) *The Tanner lectures on human values*, 3, 1-87. Salt Lake City: University of Utah Press.
- Rawls, J. (1993). The law of peoples. *Critical Inquiry*, 20(1), 36-68.
- Rawls, J. (1999). *A Theory of Justice*, revised edn. Cambridge: Belknap.
- Raz, Joseph (1979). *The Authority of Law: Essays on Law and Morality*. Oxford: Oxford University Press.
- Rommetveit, R. (1955). *Social norms and roles; Explorations in the psychology of enduring social pressures*. Minneapolis: University of Minnesota Press.
- Rosling, H. (2010). *The Magic Washing Machine*. Ted Talks. Retrieved from https://www.ted.com/talks/hans_rosling_and_the_magic_washing_machine?language=en.
- Rossi, E., Sleat, M. (2014). Realism in normative political theory. *Philosophy Compass*, 9(10), 689-701.
- Sandel, M. (1982). *The Limits of Justice*. Cambridge: Cambridge University Press.
- Sangiovanni, A. (2008). Justice and the Priority of Politics to Morality. *Journal of political philosophy*, 16(2), 137-164.
- Scanlon, T. (1998). *What we owe to each other*. Cambridge: Harvard University Press.
- Schelling, T. C. (1978). Altruism, meanness, and other potentially strategic behaviors. *American Economic Review*, 68(2), 229-230.
- Schmidtz, D. (1990). Justifying the state. *Ethics*, 101(1), 89-102.
- Schmidtz, D. (2011). Nonideal theory: What it is and what it needs to be. *Ethics*, 121(4), 772-796.

- Sen, A. (1970). The impossibility of a Paretian liberal. *Journal of political economy*, 78(1), 152-157.
- Shaver, R. (1999). *Rational egoism: A selective and critical history*. Cambridge: Cambridge University Press.
- Simmons, A. J. (2010). Ideal and nonideal theory. *Philosophy and Public Affairs*, 38(1), 5-36.
- Smith, A. (1776). *An inquiry into the nature and causes of the wealth of nations*. London: George Routledge and Sons.
- Stigler, G. J. (1971). The theory of economic regulation. *The Bell journal of economics and management science*, 2(1), 3-21.
- Stigler, G. J., Becker, G. S. (1977). De gustibus non est disputandum. *The american economic review*, 67(2), 76-90.
- Stark, C. A. (1997). Decision procedures, standards of rightness and impartiality. *Nous*, 31(4), 478-495.
- Sugden, R. (1993). Normative judgments and spontaneous order: The contractarian element in Hayek's thought. *Constitutional political economy*, 4(3), 393-424.
- Sugden, R. (2004). *The economics of rights, co-operation and welfare*. Basingstoke: Palgrave Macmillan.
- Sugden, R. (2018). *The Community of Advantage A Behavioural Economist's Defence of the Market*. Oxford University Press.
- Sunstein, C. R. (2014). *Why nudge? The politics of libertarian paternalism*. New Haven: Yale University Press.
- Thaler, R., Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth and Happiness*. London: Penguin.
- Today, (2017, March 18). Eutanasia, favorevoli tre italiani su quattro: plebiscito di "si" tra i giovanissimi. *Today*. Retrived from <http://www.today.it/politica/sondaggi-testamento-biologico.html>
- Tomasi, J. (2011). *Free Market Fairness*. Princeton University Press.
- Ullmann-Margalit, E. (1977). *The emergence of norms*. Oxford: Clarendon.

- Urbinati, N. (2014). *Democracy Disfigured. Opinion, Truth, and the People*. Cambridge: Harvard University Press.
- Vanberg, V. (1986). Spontaneous market order and social rules: a critical examination of FA Hayek's theory of cultural evolution. *Economics and Philosophy*, 2(1), 75-100.
- Veltri, C. (2006, November 29). Eutanasia, anche i cattolici favorevoli Il 50% dei praticanti dice sì a Welby. *Repubblica*. Retrieved from <http://www.repubblica.it/2006/11/sezioni/cronaca/welby2/sondaggio-ipr-eutanasia/sondaggio-ipr-eutanasia.html>
- Walzer, M. (Ed.). (1995). *Toward a global civil society*. Providence: Berghahn Books.
- Weimer, W. B. (1982). Hayek's approach to the problems of complex phenomena: An introduction to the theoretical psychology of The Sensory Order. In: W. B. Weimer, D. S. Palermo *Cognition and the symbolic processes*, Volume 2, 241-285. Hillsdale: Erlbaum.
- Whitman, D. G. (1998). Hayek contra Pangloss on evolutionary systems. *Constitutional Political Economy*, 9(1), 45-66.
- Williams, B.; Smart, J. J. C. (1973). *Utilitarianism: For and Against*, Cambridge: Cambridge University Press.
- Williams, B., 2002. Why Philosophy Needs History. *London Review of Books*, vol. 24 no. 20 pp. 7-9
- Williams, B. A. O. (2005). *In the beginning was the deed: Realism and moralism in political argument*. Princeton: Princeton University Press.
- Young, P. H. (1993). The Evolution of Conventions. *Econometrica*, vol. 61(1), pages 57-84, January.