

**Markets, Institutions and the Polanyian challenge: A  
Theoretical Study of the New Institutional Economic History  
of Douglass C. North**

**A thesis submitted to the degree of Doctor of Philosophy**

**by**

**Matthijs Krul**

**College of Business, Arts and Social Sciences**

**Brunel University, London**

**January 2016**

## Abstract

In this study, I examine the New Institutional Economic History (NIEH) of Douglass C. North from a historiographical and philosophical perspective. As a point of departure for this purpose I take North's critical engagement with the primitivism-modernism debate in premodern economic history, as represented in his early work by the 'challenge of Karl Polanyi'. This challenge, I argue, has given shape to the development of the NIEH in its various stages of theoretical elaboration. Therefore, understanding its contextual significance is indispensable for making sense of North's oeuvre as a whole. On my reading, North interpreted the challenge of Polanyi to mean combining two methodological conceptions previously not united in one work. On the one hand, North's NIEH extends the scope of economic theory to the study of the *longue durée* of economic history; while on the other hand North seeks to theorize the importance of historical variation in sociocultural institutions for understanding why there are rarely complete or well-functioning markets in most of economic history. North considers neoclassical economics suitable for neither of these purposes. Yet his critique of Polanyi's substantivist-primitivist approach is primarily based on the absence of an integration of his project with the tools of economic theory. For this purpose, North therefore adopted the theory of transaction cost economics, also called New Institutional Economics (NIE), to this new ambitious end. More than perhaps any other author North has been responsible for extending the scope and sophistication of this economics based approach in the study of economic history.

In the present work, I discuss to what extent this approach has been successful in its own aims, internally consistent, and to what extent it is plausible as a historiographical approach from an 'external' point of view. I do this by combining a close reading and interpretation of a variety of North's writings, focusing in particular on the most contemporary version of his work - which has not been much studied - with a methodological and theoretical discussion of various major themes in or aspects of his work from the viewpoints of historiography, anthropology, and philosophy of social science. These themes include (among others) North's understanding of the functioning of markets in politics and economics, his approach to choice theory, rationality, and game theory, his use or neglect of evolutionary concepts, the meaning of embeddedness in his work, and North's contractarian anthropology.

As this work shows, North's NIEH is situated in a difficult intermediate position within larger debates in economic thought: between primitivism and modernism, between substantivism and formalism (in the anthropological sense), and most significantly, between the 'new mainstream' of economic theory and the quest for successive endogenisation of the institutional context of economic behavior. This certainly speaks for the ambition and sophistication of North's historiographical approach, something which has only increased with the further development of his theory. But in his quest to unite the best insights of choice theory with New Institutional economics as well as incorporating the 'anthropological' level of fully socialized beliefs, preferences, and how they give rise to institutional variation in history, North frequently

seeks to have his cake and eat it. The persistent methodological ambiguities in his work give rise to problems of internal consistency and external plausibility, which are present from the very inauguration of his NIEH research programme. The subsequent development of his work has not, I argue, been able to overcome this fundamental problem. For this reason, while much of North's toolset and his overarching ambitions are valuable developments in economic historical theory, he does not achieve his aim of overcoming the challenge of Karl Polanyi. Without a more decisive break with his original economic microfoundations, North's NIEH project cannot ultimately live up to its grand ambitions.

## Table of Contents

<b>Title Page</b>	1
<b>Abstract</b>	2-3
<b>Table of Contents</b>	4-6
<b>Acknowledgements</b>	7
<b>Introduction: Douglass North and the ‘challenge of Karl Polanyi’</b>	8-25
I. The challenge	8-12
II. The approach of this work	12-18
III. The reception of Douglass North’s NIEH	18-24
IV. Summary of the central problems of this work	24-25
<b>Chapter 1. Douglass North’s New Institutional Economic History</b>	26-53
Introduction: beyond the New Economic History	26-31
The birth of the NIEH: from the efficient social contract to Polanyi’s challenge	31-34
North’s NIEH in its second stage	34-41
The behavioural foundations of North’s NIEH	41-45
North’s third stage: embeddedness and cooperation in <i>UPEC</i> and <i>Violence and Social Orders</i>	45-52
Conclusion	52-53
<b>Chapter 2. North versus Polanyi on markets and embeddedness</b>	54-78
Introduction: A New theory of comparative economics	54-56
Markets, institutions and embeddedness in Polanyi	56-58
Is there embeddedness in North?	58-60
North’s theory of markets	61-66
The ‘Smithian result’ and the social contract	66-70

Heroic opportunists and as-if markets: The critical reception of Douglass North's market theory	70-77
Conclusion	77-78
<b>Chapter 3. Rationality, Social Contract, and Cooperation in North's NIEH</b>	79-108
Introduction: cooperation and its constraints	79-81
Spontaneous order, cooperation, and the classical liberal tradition	82-87
Game theory, the social contract, and the context of North's NIEH	87-92
Why rationality is bounded for North	93-95
Ideology and social learning: North's approach to beliefs and rationality	95-97
'Players of the game'	97-100
Between social contract theory and economic rationality	100-105
North's ambiguities on rational choice analysis	105-108
North on the cognitive sources of institutions	108-113
Conclusion: Why would North hold an inconsistent theory?	113-117
<b>Chapter 4: North's theory of cultural evolution</b>	118-168
Introduction: towards adaptive efficiency	118-121
Section I. Examining North's theory of cultural evolution	121-119
The evolutionary dimension of North's NIEH	121-124
North and the origin of institutions	124-127
Macro-level evolution in North's account	127-130
Section II. Evaluating North's theory of cultural evolution	130-168
North's cultural evolutionary account in theoretical context	131-133
The context of evolutionary economics	133-137
What evolves?	137-144
Interpreting North's comments on sociobiology and cooperation	144-149

Social contract or cooperative species?	149-159
Gene-culture coevolution and the role of institutions	159-164
A word on morality and emotions	164-165
Conclusion	165-168
<b>Chapter 5: Polanyi's challenge revisited: historiographical and anthropological reflections on Douglass North's NIEH</b>	169-178
Introduction: revisiting Karl Polanyi	169-170
North's pessimistic anthropology	170-173
North, Polanyi, and the primitivism-modernism debate	173-178
Revisiting Polanyi's challenge	178-184
What role for economic anthropology? A sketch of background and potentials	184-192
<b>Conclusion</b>	193-197
<b>Bibliography</b>	198-218

## Acknowledgements

In arguing for a greater understanding of intersubjectivity and other-regard in theories of economic behaviour, I can hardly do without observing that no work is the product of any individual in isolation. I dedicate therefore this work to my parents, without whose support none of this would have been possible, and indeed without whom my intellectual curiosity would have come to nought. I also wish to thank Gareth Dale and Peter Thomas for their unwavering support, their helpful comments and suggestions, and occasional pep talks. Geoffrey Hodgson and Mark Neocleous provided me with helpful comments. I thank Thomas Linehan for his enthusiasm and interest. I owe a debt to Chris Hann in particular and the Max Planck Institute for Social Anthropology, Halle (Saale), in general for their patience and support. For encouragement and insight at different stages of the process, I thank Floris Cohen, Maarten Prak, Jan Luiten van Zanden and Bas van Bavel. Furthermore, I wish to recognise my friends and comrades in London and elsewhere for their willingness to exchange social, political, and intellectual ideas and sociality with me, in person or online. While I am fortunate to have too many friends and 'fellow fellow travellers' to name them all, I especially want to credit Julian Duane as friend and intellectual sparring partner of many years, and Tom Fisher for the hugs. Finally, I must mention the other loved ones in my life: Julia Krul and Maarten Oosten. Above all I want to mention Mika Naylor, whose company, love, and cups of tea defy neoclassical economics by exhibiting no diminishing returns. All errors, of course, are mine. About them I can only cite my favourite author: "The twin peaks of Parnassus are approached through some very dim valleys. If scrambling in these, without any climbing, is sometimes rewarded with a degree, one must hope that one of the peaks at least has been glimpsed from afar."<sup>1</sup>

---

1 J.R.R. Tolkien, "Valedictory Address". *The Monsters and the Critics and Other Essays* (London 2006), p. 226.

## Introduction:

### Douglass North and the ‘challenge of Karl Polanyi’

#### Section I. The Challenge

At first sight it may seem strange, but it should come as no surprise that the eminent economic historian Douglass North inaugurated his then newly conceived research programme, his New Institutional Economic History (NIEH), with an appreciation of Karl Polanyi.<sup>2</sup> Although often overlooked, his 1977 paper on what he called ‘the challenge of Karl Polanyi’ expressed two important theoretical steps for the development of North’s ideas in this field.<sup>3</sup> Firstly, it is one of the earliest and best statements of what he saw as the tasks of the NIEH that he was to give shape to over the following decades. Secondly, it is his most explicit discussion of how his conception of these tasks, and the differences between his aims and assumptions and those of the neoclassical and substantivist traditions, originate in the ‘challenge’ of Karl Polanyi. As I will argue in this work, the former cannot be understood without the latter: without Polanyi’s challenge, no Northian NIEH.

“Anthropologists, sociologists and historians”, North argues in this paper, “have long challenged the relevance of economic theory to the analysis of past societies”. However, this theory, defined in this paper as the “wealth-maximizing behavioural postulate”, has survived among neoclassically oriented economic historians nonetheless because they have generally been unconcerned by the implications of such theoretical background choices: one cannot directly test them anyway, “the proof of the pudding is in the eating”.<sup>4</sup> Indeed, for most of his early career up to the writing of this paper, North himself had perhaps been a primary example of this unconcerned attitude. He was known mainly as one of the leading exponents of Cliometrics (or ‘New Economic History’): a school defined by the application of quantitative neoclassical methods to economic historical analysis, and not generally bothered much by the claims of other social sciences.<sup>5</sup> But whether or not North had himself been similarly unconcerned before, by this point he had come to be haunted by the presence of a ghost (as he describes it): the spirit of Karl Polanyi. Polanyi, he argues, “cannot be so lightly dismissed, and if his spirit does not haunt the new economic historians, it is because they are probably not even aware the ghost exists”.<sup>6</sup> But he does, and his challenge to such economic historians is straightforward: “Polanyi contends

---

<sup>2</sup> Unfortunately, Douglass C. North (1920-2015) passed away during the revision of this work. Because this thesis emphasizes the continued relevance and influence of North’s work and his posthumous debate with Karl Polanyi, I have decided to maintain use of the present tense.

<sup>3</sup> North 1977. One previous discussion to which this paper is central is Didry and Vincensini 2011; for other contrasts of North and Polanyi, see: Ankarloo 2002; Davis 2008.

<sup>4</sup> North 1977, p. 703-704.

<sup>5</sup> See North 1997 for a reflection on North’s part on his Cliometric period and the achievements and shortcomings of that approach. For his original, more negative assessment, see North 1977, *op. cit.*

<sup>6</sup> North 1977, p. 704.



that markets have only dominated resource allocation for a brief span of history centering on the nineteenth-century Western world. Before that time... other allocative systems have characterized economic organization and these other systems are not grounded in economizing behaviour”.<sup>7</sup>

North is clearly ambivalent about the value of this viewpoint. He explains Polanyi’s theory of ‘transactional modes’, which include – besides market exchange – reciprocity and redistribution, and how for Polanyi these latter two were the dominant or characteristic such modes of the economies of premodern societies. Markets dominate only in the most recent historical period, and even then only outside the Soviet-type societies. In this, he rightly notes, Polanyi opposes both the neoclassical and the Marxist approaches. Each of these represents an opposed social force within modern market society, a society that has only emerged out of the Industrial Revolution. For North, “it is easy to find fault with Polanyi’s analytical framework”. Indeed, “there are numerous parts of his analysis that show a failure to grasp elementary economic principles”.<sup>8</sup> Economic theory, in other words, is not on Polanyi’s side. And yet North is deeply impressed by the significance of this challenge. For “the stubborn fact of the matter is that Polanyi was correct in his major contention that the nineteenth century was a unique era in which markets played a more important role than at any other time in history... To the extent that economic theory was confined to the analysis of markets, the tools of the economist were... irrelevant to an understanding of the ancient world... Economic historians have not even begun to account for such non-market allocative systems, and until they do, they can say very little about societies in which markets had very limited allocative effects.”<sup>9</sup>

North first considers Polanyi’s own approach to dealing with this problem. Besides defining transactional modes, what North calls ‘allocative systems’, Polanyi had one other major theoretical tool: a concept that in the subsequent literature would become known as embeddedness.<sup>10</sup> North here refers to it as ‘embodied’, but is clearly speaking of the same concept. He takes as good a definition of embeddedness as one could wish for from Polanyi directly: “The key to Polanyi’s system is the view that economic organization is *embodied* in society in the sense of ‘having no separate existence apart from its controlling social integument’... Transactional dispositions of natural resources, labor, produce and services are expressions of socially defined obligation and relationships.”<sup>11</sup> The consequence of this is that “the kinship, status, hierarchy and political or religious affiliations which underlie these economic structures are not explicable in terms of economizing behaviour – one can only

---

<sup>7</sup> *Ibid.*, p. 703.

<sup>8</sup> *Ibid.*, p. 707.

<sup>9</sup> *Ibid.*, p. 706.

<sup>10</sup> An excellent discussion of Polanyi’s concept of embeddedness and its various subsequent usages is Dale 2011. See also chapter 2 for a further elaboration.

<sup>11</sup> North 1977, p. 708. Emphasis added.

understand them and therefore the functioning of the economy by ‘in depth’ studies which are social, cultural, and psychological in origin”.<sup>12</sup>

This brings us to North’s own solution. To answer the challenge of Karl Polanyi means two things: it means to shift the emphasis of research to the study of premodern economic history – that is to say, the economic history of the period before the Industrial Revolution and the dominance of the market as allocation system – and it means to give an explanation of the nature and functioning of the alternative allocation systems that Polanyi argued dominated in the premodern era. But for such an explanation to answer the challenge, it must be an explanation based in ‘economic theory’, as North put it, rather than exclusively in the “social, cultural, and psychological” explanations that Polanyi favoured.

The economic theory in question is no longer to be neoclassical, the concessions to Polanyi’s critique of such thought in this 1977 paper make clear. Rather, it is to be the New Institutional Economics (NIE), in this form also known as transaction cost economics. (As I will detail further in the first chapter, North, together with Robert Thomas, had at this point already made steps towards substituting this approach for the neoclassical one in his study of economic institutions in medieval Europe.<sup>13</sup>) Explanations in terms of transaction costs would challenge the claim of anthropologists - who were Polanyi’s main sources of theoretical and empirical inspiration and to whose discipline he arguably belonged at least as much as to economic historians – to the domain of the ‘social’: the socio-political structures and cultural norms in which these economic structures were embedded. As North puts it in the final sections of this paper: “Polanyi conceived the custom, kinship arrangements, status, etc., to be fundamentally a result of non-economic forces whereas [North and Thomas’ view is] consistent with an explanation that they evolved as ways to reduce transactions costs (*sic*)... It was Karl Polanyi’s intuitive genius that he saw the issues. A transactions-cost approach offers the promise of providing refutable explanations for these ‘transactional modes’... We can and should be able to predict the direction of institutional change [though] the precise form it will take is still beyond the scope of the state of the art.”<sup>14</sup>

The significance of this discussion for understanding the nature, development, and implications of North’s NIEH should be clarified. North’s engagement of Polanyi spells out a number of the themes that are central to his NIEH approach. Its *explananda* are, as North makes clear, those of Polanyi’s challenge: the ability to explain nonmarket allocation systems in premodern economies. The NIEH would address the question whether such allocation systems are ultimately embedded in *sui generis* social, political, and cultural (what I would call ‘anthropological’) causal factors or that they can be causally explained in terms of a general economic theory: a transhistorical and transcultural need to reduce transaction costs in economic

---

<sup>12</sup> *Ibid.*, p. 708.

<sup>13</sup> See: North and Thomas 1971; 1976.

<sup>14</sup> North 1977, p. 715-716.

behaviour. This transaction cost economics is to be seen as rivalrous with neoclassical economics and of Marxism alike, in North's perception - although the references to transaction costs "changing in the margins" show North's commitment to marginalist methodology is not diminished thereby.<sup>15</sup>

We also see in embryo some of the central concerns of North's NIEH in its later development, as I will discuss in the next chapter: the shift from the study of markets to the institutions underlying markets, one that would over time develop towards a further study of the origins of human sociality, normative and ideological structures, and political formations, i.e. the 'non-economic' institutions that co-determine the functioning, embeddedness and extent of market exchange. Indeed, ultimately Polanyi's substantivism represents a greater challenge than just the question of nonmarket allocation systems for North: "The point goes much deeper than Polanyi realized, and poses a fundamental question to the economist and economic historian. How do we account for substitutes for price-making markets of which families, firms, guilds, manors, trade unions, cooperatives, etc. are organizing institutions which allocate resources instead of markets. Most fundamental of all, how do we explain government?"<sup>16</sup>

While one can question whether North is right that Polanyi was unaware of this central difficulty for economists, it does show how great for North the stakes of Polanyi's challenge are. What is at stake is not just which approach, the substantivist or the NIE one, is best able to explain the economic institutions of premodern societies. More than that, one can observe two major interrelated methodological debates in the background that make Polanyi's approach such a 'challenge' worthy of answering with a full-fledged new research programme. Firstly, the contestation as to whether the methods of mainstream economics or the theories of the anthropologists (and perhaps by extension sociologists and historians) will dominate the study of premodern economic history.<sup>17</sup> This is implicit in the 'challenge'. Polanyi's stature among, even adoption by, economic anthropologists – a recognition which unlike his influence in other disciplines has been continuous since his death – makes him the ideal foil for North's research programme, which is here founded by explicit contrast on a bedrock of 'economic theory'.

Less immediately apparent in this text is the older interdisciplinary debate between 'primitivism' and 'modernism': i.e. the question whether premodern economic structures can only be explained by reference to qualitative historical differentiation in their economic and/or social 'logics', or whether they are fundamentally of one kind and only variable as a matter of (quantitative, organizational, institutional) degree.<sup>18</sup> But this debate was of the greatest

---

<sup>15</sup> *Ibid.*, p. 715.

<sup>16</sup> *Ibid.*, p. 709.

<sup>17</sup> This disciplinary rivalry has sometimes been described by the term 'economics imperialism': see e.g. Fine and Milonakis 2009.

<sup>18</sup> There is a considerable literature on this debate. Classic statements are Finley 1985; Bücher 1893; Meyer 1913; Polanyi, Arensberg and Pearson 1957; Rostovtzeff 1957. See for some recent discussions: Scheidel and Von Reden (eds.) 2002; Meikle 1996; Bang, Ikeguchi, and Ziche 2006; Jongman 2012; Jongman 2014.

importance to Polanyi's own work and that of his historical and anthropological collaborators. North's opposition to Polanyi mirrors that of 'modernists' against the 'primitivists' in economic history and economic anthropology alike, and, arguably, it is this question more than any other that produces two identifiable schools of thought across the disciplinary divides of economic thought in general.<sup>19</sup>

## **II. The approach of this work**

I propose to examine Douglass North's NIEH against the backdrop of his confrontation with Polanyi and the deeper methodological debates in economic thought described above, spanning the disciplines of economics, economic history, and anthropology. While there is a considerable secondary literature on North, his work tends to be discussed in the context of the development of the NIE as economic theory, rather than in its significance as a new step in longstanding debates on the interpretation of premodern economic history, and the applicability of (certain) economic theory thereto.<sup>20</sup> While in recent years various other economic thinkers, theorists and historians alike, have contributed to the NIEH, I concentrate on North's work as his is both the original and most influential version and because arguably his work makes the best case for the NIEH approach. It is radically critical of neoclassical and 'neoinstitutional' approaches, its theoretical scope and apparatus is sophisticated and spans a great interdisciplinary range of problems, and his theoretical work represents within New Institutionalism the most persistent attempt at developing a fully-fledged new 'modernist' interpretation of history. It thereby has a significance far beyond the specific problems it addresses in economic theory-driven historiography.<sup>21</sup>

However, the emergence of North's NIEH has been defined not just by its opposition to primitivist and/or substantivist theory. It is also defined by a complex relation to the economic mainstream. The rise of evolutionary economic thought, part of a more general re-appreciation of the potential of Darwinian selection to understand social scientific problems, has emphasized that economic behaviour and institutions are as much the unintended effect of various rule-like structured responses as that they are the product of design. The emphasis of this research programme is therefore on the emergence and change of institutions and behaviours out of habits, rule-following and other algorithmic behavioural patterns in varying situations. Not all evolutionary approaches to economics have, by any means, followed explicit Darwinian principles. Indeed, as discussed in the chapter on North's use of evolutionary terminology, often putatively different evolutionary principles (e.g. the Lamarckian) are invoked. Some evolutionary economists have emphasized that what this 'generalized' or 'universal Darwinism'

---

<sup>19</sup> By 'economic thought' I intend a shorthand for the common problems and intellectual traditions of economic theory, economic history, economic sociology, economic anthropology, and so forth.

<sup>20</sup> For a brief discussion of this literature, see below.

<sup>21</sup> See e.g. Galiani and Sened 2014.

applies whenever the core elements of selection, variety, and inheritance operate on open and complex systems, including those of human societies.<sup>22</sup>

A classic application of evolutionary thinking to understanding institutional evolution in the economic sphere is the work of R.R. Nelson and S.G. Winter. This work in turn has had a considerable influence on the development of the New Institutional Economics, and encouraged the convergence between New Institutional and evolutionary approaches.<sup>23</sup> This emerging synthesis between NIE and certain evolutionary approaches in economics has the potential of forming a new paradigm, one that fits very well with – indeed has partially co-determined the development of – the new mainstream’s preoccupation with institutions and bounded rationality. In North’s NIEH, this evolutionary-institutional paradigm is, however, taken one step further: with a deeper quest for the origins of institutions and political order (“how can we explain government?”). Ultimately, this culminates a sophisticated combination of game theory, New Institutionalism, and contractarian philosophy to complete a process of “successive endogenisation” of sociocultural institutions, to borrow a phrase from Thrainn Eggertsson, into a single economic historical theory.<sup>24</sup>

That is to say, on the basis of these economic and philosophical foundations North erects a comparative theory for studying the interaction between markets and other allocation systems, formal (politico-legal) and informal (socio-cultural) institutions, and the social-evolutionary adaptation of societies to change under conditions of uncertainty. This argument moves across several levels of abstraction and scope: from the bedrock of human sociality to the complex norms and beliefs that constitute the institutions of social life and therefore shape economic behaviour through time to the level of long-term historical evolution and path dependence. In its most recent form, it constitutes nothing less than a “conceptual framework for interpreting recorded human history”.<sup>25</sup>

It is therefore a highly ambitious form of New Institutionalism, extensive in domain of application and crossing traditional disciplinary boundaries of the social sciences. Yet this ambition is understandable in light of Polanyi’s own attempts at formulating such a comparative theory for premodern economic formations, one that equally straddles economic history and anthropology: it is perhaps the necessary consequence of answering Polanyi’s challenge.<sup>26</sup> In this work, I will seek to clarify the steps in North’s approach from his earliest neoinstitutionalism to his fully-fledged historical theory by considering it from the viewpoint of his 1977 paper: a ‘modernist’ rejoinder to Polanyi’s ‘primitivist’ challenge. But I will also examine the fluctuating relationship between North’s NIEH and the mainstream of economic theory. This relationship is not just something which ultimately plays a major role in defining North’s specific positioning

---

<sup>22</sup> Stoelhorst 2007, p. 233-234; Dennett 1995; e.g. Hodgson 2003, 2004; Hodgson and Knudsen 2008.

<sup>23</sup> Nelson and Winter 1982, 2002.

<sup>24</sup> Eggertsson 1990, p. xiii.

<sup>25</sup> The subtitle of North, Wallis, and Weingast 2009.

<sup>26</sup> For an early discussion of this dual disciplinary role of Polanyi’s work, see: Humphreys 1969.

within the primitivism-modernism debate, but it is also one of the main sources of North's influence in a wider sense: as I will discuss in more detail shortly, much of the reception of North's work has read him as a primary exponent of New Institutional Economics and has highlighted his contributions to bringing institutions (as rules) and the importance of beliefs back into economic theory, even to the extent of an increasing convergence with the 'Old Institutional Economics'.<sup>27</sup>

This requires some explication of what is meant by the economic mainstream, to be sure a contested category. As David Colander and colleagues argue, the mainstream of economics is not simply neoclassical theory; it can be defined more usefully as the *modelling* approach to explaining social phenomena, within which the substance of the models has shifted – at least in what they call the “edge of economics” - from general equilibrium analysis to complexity and evolutionary models.<sup>28</sup> Equally, according to them the content of economic theory has increasingly shifted from perfect to imperfect markets and their institutional contexts, and from older 'pure' rational choice conceptions of individual choice to new approaches informed by behavioural economics and psychology, such as models of bounded rationality.<sup>29</sup> They therefore distinguish *orthodoxy*, equated with neoclassical economics, from *mainstream* (for which one could adopt the synonym 'conventional'), which includes the traditional neoclassical approach but also the newer interest in institutional, evolutionary, and behavioural economics – that is, insofar as those remain within the paradigm of formal modelling and/or based on the microfoundations of choice theory.<sup>30</sup>

But to understand the context of North's NIEH within this development of economic theory, this set of definitions is not sufficient. For what of approaches that have much in common with the new mainstream's interest in institutions, evolution, and bounded rationality, but that are not primarily concerned with formal modelling? North's NIEH is one of the most prominent examples of such an approach, and this would place North's work (at least in its later stages) squarely outside even the new mainstream by Colander et al.'s definitions. And yet, within even North's most developed theory, as we shall see, there are still many elements that hearken back

---

<sup>27</sup> E.g. Hodgson 2007a; Rutherford 1994; Brownlow 2010.

<sup>28</sup> Colander, Holt and Rosser 2004.

<sup>29</sup> *Ibid.*, p. 496.

<sup>30</sup> The meaning and utility of the term 'neoclassical economics' is much debated. In the literature it appears often pejoratively, and it has as many definitions as there are polemics about it. See e.g. Weintraub 1993; Nadeau 2008; Morgan 2016; or the exchange between Mirowski and McCloskey in De Marchi 1992. It has been doubted whether it applies to the contemporary mainstream at all, as per Colander 2000. Yet Arnsperger and Varoufakis 2006 argue that modern developments have not made the term obsolete, because they are incorporated into its research programme's flexibility. For present purposes, a useful definition may be rational choice utility maximization plus equilibrium economics. I elaborate on the implications of these for North's treatment of efficiency and rationality in chapters 1 and 3, where I also discuss North's own ideas of what neoclassical economics is and what is wrong with it. In his interpretation and critique of neoclassical economics, North focuses on the ergodicity of equilibrium economics and what he sees as the shortcomings of neoclassical rational choice theory, in particular its (alleged) neglect of beliefs other than in probabilistic terms.

to the theoretical and methodological foundations that have limited the discipline of economics' move away from orthodoxy despite the development of this 'new mainstream': in particular, an unwillingness to give up certain assumptions about individual rationality and its significance for the method and conclusions of orthodox choice theory. As Geoffrey Hodgson has shown, often game theory has come to be the favoured approach of those who seek to have both the microfoundational premises of orthodoxy and yet to incorporate the new mainstream's interest in bounded rationality and institutional context.<sup>31</sup>

Such 'having your cake and eating' it approaches straddle the mainstream/non-mainstream divide, joining together two apparently contradictory methodological and theoretical paradigms. In this work, I will argue North's NIEH, in particular in its more contemporary developed form, veers between this in-between position and a more thoroughgoing departure from even the 'new mainstream's theoretical and methodological premises, in a process perhaps best described as one step forward, one step back. This is not so much a question of formal modelling in the mathematical sense as a question of how far the evolutionary-institutional convergence in North's work can be reconciled with his attachment to more *theoretically* – if not methodologically – 'orthodox' preoccupations with individual, boundedly rational choice and with the welfare implications of conventional microeconomics. I will explore this through examining the consequences of this tension in North's work on his view of markets, of rationality and choice, of evolution, behavioural economics, and cooperation.

In other words, to borrow an analogy from Mary Morgan's discussion of theory and mechanism in economic models, an examination of the 'background theory' to North's NIEH 'model' allows for a better understanding of how the interpretation of the model's results is determined by these assumptions via the mechanisms in question.<sup>32</sup> Initially, the Northian answer to Polanyi's challenge hewed closely to the more neoclassical wing of the NIE. But over time, as I will show, his approach has increasingly converged with major elements of the 'new mainstream' identified by Colander, Holt and Rosser, such as the centrality of the problem of cooperation (with reference to its formalisation in game theory) and evolutionary concepts. In this sense, North's NIEH approach has become more and more remote from neoclassical orthodoxy, but whether it can escape the limitations of even the 'new mainstream' is another question, one that this work will seek to answer.

In this work, I seek therefore to explore what North's NIEH is, and to what extent it constitutes a theoretical approach to economic history that is, by dint of its combination of evolutionary and institutional conceptualizations, able to answer Polanyi's primitivist-substantivist challenge. This means first clarifying the various theoretical assumptions, levels of abstraction, and theoretical development of North's NIEH from its beginnings to the present. As other authors have rightly observed, North's NIEH has undergone successive transformations

---

<sup>31</sup> Hodgson 2007b.

<sup>32</sup> Morgan 2012; cf. Morrison 1999, p. 51.

that one can identify as different stages of his work, and each of these stages represents a significant shift in North's complex borrowing from and rejecting of conventional economic theory.<sup>33</sup> In each new major work, North's methodological attitude changes along with his further divergence from his neoclassical origins; a process that is however never consistent or complete. To elucidate this development of North's work through a close reading will be the task of chapter 1.

In the next part of this work, I will examine several major theoretical themes in North's NIEH theory, especially in its most recent and sophisticated form, each concerned with a different level of abstraction in the whole. Judging to what extent North's rejoinder to substantivist primitivism is successful depends on comprehending the inner relations of his complex conceptual apparatus. As mentioned, North's theory is highly complex and sophisticated, and moves frequently in the course of an explanation from the most fundamental level of individual rationality and choice to the level of institutions and organizations, and from there to adaptation and efficiency at the transhistorical level. This work therefore seeks to clarify what these different levels of abstraction in North's NIEH theory are, how they relate to each other, and how consistent and plausible the steps from one to the other are.

To do this, three chapters will be devoted to exploring major themes in North's theory, and the different levels of scope and abstraction to which they apply. First, in chapter 2, I will inquire what North's concept of markets and their significance for economic behaviour is, and how he uses market exchange and market discourse in his theory to understand the relationship between individual agency and economic or social institutions. This also touches on the distinction between markets and other allocation systems central to his foundational 1977 paper, as well as the classic New Institutional opposition between markets and hierarchies. Since North posits these as solutions to transaction cost problems in the vein of NIE theory, I ask to what extent this approach is useful for understanding economic behaviour in premodern economic systems, especially given North's concession that markets played a relatively minor role in such formations. I also examine North's discussion of markets and competition with an eye to the significance of rhetoric in economics: North's frequent invocation of "economic and political markets" and the use of language derived from the marginalist theoretical tradition in economics provides an intriguing contrast with his more 'Polanyian' moments. In particular North's interpretation of politics, the realm of formal institutions, as operating in a 'market-like' way, and the degree to which for North competitive markets operate both as an analytical model and as a normative *desideratum* has not frequently been discussed in the existing literature. I also discuss the ideological components of North's rhetorical structure, in particular his reliance on (sometimes implicit) contractarian views of society and the classical liberal view of markets, competition, and the state that emerges from North's overall historical narrative.

---

<sup>33</sup> Brownlow 2010; Groenewegen, Kerstholt and Nagelkerke 1995; Ménard and Shirley 2014.



Then in chapter 3 I will study what one might call North's own microfoundations: that is to say, his theories and assumptions regarding individual rationality and choice as the basis of his understanding of how individuals and institutions interact. The desire to provide microfoundations of more general theory in some form of individual rational behaviour is arguably - next to Colander et al.'s emphasis on modelling - one of the hallmarks of mainstream economics, and how North chooses or justifies his microfoundations, i.e. his assumptions about individual rationality and agency, therefore significantly affects the structure of his argument.<sup>34</sup> Identifying these - often implicit - assumptions and assessing their function in his theory is for that reason essential to understanding how well the Northian approach succeeds in its larger aims. This chapter will concentrate in particular on North's critique of rational choice and the explicit as well as implicit forms of bounded rationality and 'socialized' choice he develops in the course of his most recent work. This examination gives me an opportunity to clarify North's complex and contradictory relationship with choice theory and the neoclassical microeconomic tradition on the one hand, and his disavowal of rational choice approaches and disembodied individual choice on the other hand. In so doing, I hope to show how the problem of rationality in North's work illustrates the position of his work in general as a powerful and ambitious, but often incomplete and internally contradictory project of 'successive endogenisation'.

Subsequently, I will turn to North's use of evolutionary themes and arguments, which are especially evident in his 2005 work *Understanding the Process of Economic Change*, but also appear elsewhere.<sup>35</sup> If chapter 3 is primarily concerned with the micro-level of North's theoretical edifice, chapter 4 will be dedicated mainly to the macro-level. As North introduces in *UPEC* and elsewhere an evolutionary framework for his historical narrative of institutions and economic performance, I use this opportunity to evaluate the plausibility of this narrative. In part, this involves critically examining North's evolutionary claims in light of theory in evolutionary economics and sociobiology. If the identification of evolutionary mechanisms such as natural selection is to succeed, the appeal to evolution must be more than a merely metaphorical use of words, and this is one aspect I examine here.<sup>36</sup> But I also discuss another dimension of evolution in North's work: his turn to cognitive theory and evolutionary anthropology to buttress his argument for the significance of beliefs and institutions in achieving human cooperation. I will discuss North's claims in this regard and assess his use - or omission - of anthropological and behavioural economic sources in his argument, and draw out the implications for North's general treatment of cooperation. Finally, I give an assessment of North's neglect of sociobiology and present an argument for the significance of sociobiological

---

<sup>34</sup> On microfoundations in NIE, see e.g. Powell and DiMaggio (eds.) 1991; Powell and Colyvas 2008. Critical discussions of the significance of microfoundations for economic theory in general can be found in e.g. Weintraub 1979; Rizvi 1994; Van den Bergh and Gowdy 2003.

<sup>35</sup> North 2005.

<sup>36</sup> See also: Rosenberg 2008, p. 196-198 for a brief discussion.

theory, in particular gene-culture coevolution, as well as behavioural and experimental economics for a discussion of the evolution of cooperation.

The concluding chapter returns to Polanyi's challenge, and draws out the anthropological inquiry that underlies the investigations described above. In this work I intend to read North's NIEH and its foundational concepts not so much from the angle of economic theory or methodology per se, but rather in order to draw out what one might call the 'hidden anthropology' of North's approach. By this I mean that in order to meaningfully answer the question to what extent Douglass North can answer Polanyi's challenge, one must consider their debate one involving competing views of the significance of economic theory for understanding premodern societies, and how these different assessments are in turn based on different approaches to human cooperation, institutions, and social structure. Insofar as such debates about the economics of institutions in historical change are always concerned with how best to understand the expression of human nature as socially and culturally mediated individuals, this discussion has an inescapable (economic) anthropological dimension.<sup>37</sup>

But North's anthropological assumptions and views are not necessarily explicit: as the analogy with modelling suggests, the background theory to a model is often not much specified, let alone explicitly justified. In fact, such an investigation also requires a constant alertness to whether North's explicit statements of his theoretical commitments and ambitions are actually reflected in the theory he presents. In the course of this work, it is such questions that will guide the analysis and the presentation of his works. In short, this means the critical content of this work is interpretative in nature. It combines 'internal' critique of North's work at various levels of abstraction, i.e. understanding North's work with an eye to his overall ambition to answer Polanyi's challenge, with 'external' considerations from the viewpoint of competing 'anthropologies', i.e. judging the plausibility of either position in light of alternative approaches and theoretical resources. For the latter purpose, I attempt in each chapter in which I discuss one of the themes in North's work to also suggest alternatives or complements to his approach, which I draw freely and merely as sketches of possibilities from (heterodox) economics, economic anthropology, or sociobiological literature, or other sources as relevant.

In the final chapter, I will sum up my findings in both regards by reassessing Polanyi and North's competing approaches. I will also there consider how a different conceptual frame, drawing together my suggestions from the above (sub)disciplines and making a more explicit break with the economic and ideological limitations of even North's institutionalism, can help achieve the aims that North set himself in accepting this 'challenge' – and thereby go beyond either Polanyi or North in the primitivism-modernism debate. To the extent that my reading of North convinces, and I am able to show the strengths and limitations of North's NIEH as a

---

<sup>37</sup> Wilk and Cliggett 2007, ch. 1.

position in the primitivism-modernism debate compared to Polanyi's strengths and weaknesses, this work can be seen as a case study in that debate.

### III. The reception of Douglass North's NIEH

This historiographical, philosophical and anthropological angle is not the usual way in which the literature hitherto has approached Douglass North's work. In order therefore to contextualize my approach, it is worth giving a summary overview of some of the reception of North's NIEH theory, both in its laudatory and critical dimensions. There is no space here to consider the whole of the literature on the NIE or even just on North's writings, so I will concentrate on some salient studies of North's NIEH insofar as they touch on the main themes of the present work.

Usually, as mentioned above, his work is perceived as primarily a contribution to the NIE. Particularly prominent in the reception of North are the points of continuity and contrast with the Old Institutionalism, namely respectively their shared critique of neoclassical economics – history and institutions matter - and simultaneously what is often seen as the NIE's reliance on (certain) rational choice approaches and its emphasis on individual agency in the creation and change of social institutions. One such comparison, repeatedly made, is between Douglass North and Gustav Schmoller, exponent of the German Historical School.<sup>38</sup> However, to understand North's work properly means to see not just its commonalities with the NIE as a whole, in particular the influential theoretical work of Oliver Williamson, but also its differences with the Williamsonian school. Indeed, North has been sufficiently recognised as distinctive because of his interest from the very beginning in history, and for his more thoroughgoing disavowal of rational choice approaches, neither of which are pronounced in Williamson's work. For this reason, North's NIE has sometimes been distinguished from the Williamsonian 'neoinstitutionalism'.<sup>39</sup> But what makes North's version of NIE theory distinctive is a subject of ongoing debate, not least because of the evolving nature of his NIEH itself, as we will see. Many NIE enthusiasts praise North and other NIE authors for their reliance on 'theory' as opposed to the alleged lack of strong theorizing that vitiated the Old Institutional tradition.<sup>40</sup> Moreover, North is widely recognized as having advanced institutionalism considerably by his redefinition of institutions as rules and his extension of policy and economic considerations about institutions beyond both the public choice and neoclassical paradigms.<sup>41</sup>

But as both Rutherford and Helge Peukert have pointed out, simply crediting North with further developing 'theory' is not always a helpful way to consider North's contribution. The older institutionalist tradition was not lacking in theory; rather, between them and North's New Institutionalism stand a different set of theoretical and methodological assumptions, even when

---

<sup>38</sup> Richter 1996; Peukert 2001.

<sup>39</sup> Eggertsson 1990, p. 8-9.

<sup>40</sup> E.g. Coase 1984, p. 230; Langlois 1986, p. 5. See also the discussion in Mäki, Gustafsson and Knudsen 1993, p. 19-23.

<sup>41</sup> See Bates 2014 for a brief summary.

both identify similar problems in economic orthodoxy. This is well illustrated by Peukert's discussion of the commonalities and differences between North and Gustav Schmoller, referred to above.<sup>42</sup> From her perspective, it becomes clear that some of North's theoretical contributions are specific to him, and some are shared with the NIE in general; but equally, that some are responses to problems posed by *any* institutionalism, as Malcolm Rutherford has also argued.<sup>43</sup> Peukert aptly summarizes Rutherford's comparative work on this last point: "The problems [of institutionalism in general] can be formulated as trade-offs between five complementary but dichotomous research strategies and perspectives: formalism vs. anti-formalism, individualism vs. holism, rationality vs. rule following, evolution vs. design, and efficiency vs. reform."<sup>44</sup> Insofar then North has deviated from the mainstream of the New Institutional tradition as well as from the OIE, he has done so within a larger set of problems shared with all institutionalist theorists, including, one might add, Karl Polanyi.

While this context must be kept in mind, North is more than an institutionalist economic theorist, although he has been this too. The first salient observation is to note that what he shares with Polanyi but not with the great majority of the institutionalist economic tradition, New or Old (some of the GHS authors excepted), is his orientation towards economic *history*. On this point, the literature is somewhat sparser. North is certainly widely recognized as an economic historian, not least by the Nobel Prize committee of the Bank of Sweden.<sup>45</sup> However, few have sought to evaluate his contributions specifically with an eye to how he has joined his specific responses to these general questions of institutionalism with justifying his New Institutional approach to economic *history* in particular. Rutherford's identification of these five axes that define the opposing solutions to the general problems of institutionalism certainly invite a comparative study between North's positions as compared to the rival approaches of authors from the GHS and OIE. Indeed, Rutherford and Peukert's works cited above can be read as contributions to that aim. But I am concerned here that such an approach neglects the historical dimension of North's work, his ambition to use New Institutional tools for the purposes of examining the long-term structure of premodern (and contemporary) economic history. There, the comparison with the sparring partner of his own choice, Polanyi, seems more relevant.

North himself has always in disciplinary terms presented himself as an economic historian, not as an economic theorist *pur sang*. Yet even early on he was concerned with justifying the special role that economic historical theorizing had to play in clarifying and helping solve conceptual questions in economic theory, as well as vice versa. Indeed, in a paper published mere months before the Polanyi paper, titled "The Place of Economic History in the Discipline of Economics", North described the (in his view) declining interest of economists in economic

---

<sup>42</sup> Peukert 2001.

<sup>43</sup> Rutherford 1994, p. ix; Peukert 2001.

<sup>44</sup> Peukert 2001, p. 93.

<sup>45</sup> [http://www.nobelprize.org/nobel\\_prizes/economic-sciences/laureates/1993/north-facts.html](http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/1993/north-facts.html).

history as a “market failure of economics”.<sup>46</sup> He stressed the importance of new, transaction cost economics-based, theory in economic history both for the improvement of economic understanding and for the survival of economic history as a (sub)discipline. So much so, in fact, that Francisco Boldizzoni, one of the few major critics who has focused on North’s role as an economic historian, has accused him of developing his theory solely in order to fill a gap in the academic career market.<sup>47</sup>

I do not share that view. Rather, I think North should be taken seriously when he pointed so early on in his career to the double bind of economic historians like himself who took primary inspiration from orthodox economic theory: they were irrelevant to economists when merely applying their theories to the recent past, and equally “anthropologists, sociologists and historians” were able to dismiss economic history on the grounds of its neoclassical theory, which could not possibly have anything to say about periods *before* this recent past. Tellingly, as primary examples of such critics with rival approaches he cited Karl Polanyi, as well as his close collaborator Moses Finley and the Marxist economic historian Immanuel Wallerstein.<sup>48</sup> As mentioned, I take as the central purpose of North’s work since to answer these rivals with a renewed economic history, explicitly oriented towards the *longue durée* of premodernity. But North has from the beginning sought to do so by a creative application of new insights from economic theory, since its continuing importance – *pace* the Polanyian and Marxist critiques of economic theory in historical analysis – is the basis of his argument against Polanyi in his 1977 paper.

At the same time, fewer have observed that this thereby necessarily implies that North’s work is not just a critique of neoclassical economics and its ahistoricity and asociality, but an ongoing critical engagement with the uses and disuses of the ‘new mainstream’ for history as well. As this new mainstream has emerged out of the old, if we follow the interpretation of Colander, Holt and Rosser, so too has North’s approach moved in the same direction with it. Insofar as he has then sought to defend economic theory against its Polanyian and Marxist critics, he has been able to invoke this new mainstream of imperfect information, transaction costs, bounded rationality, and (evolutionary) game theory, not the neoclassical and Cliometric approaches that characterized his earliest contributions. But this raises the question, not often discussed in the literature, to what extent this appeal is sufficient to counter the Polanyian challenge, and thereby fulfil its promise for North’s own purposes.

That is not to say that the significance of North’s complex attitude to economic theory for his ambitions in economic history nor what I call the anthropological dimensions of his work have gone entirely unobserved. In economic anthropology itself the New Institutionalism has received some cautious interest, in particular among those anthropologists for whom economics

---

<sup>46</sup> North 1976, p. 462.

<sup>47</sup> Boldizzoni 2011, p. 6.

<sup>48</sup> North 1976, p. 463.

is more a source of theoretical inspiration than purely a rival discipline.<sup>49</sup> Nonetheless, this work, too, has tended to concentrate on the ‘neoinstitutionalist’ – that is to say, more neoclassical and ahistorical – NIE as developed by Oliver Williamson, rather than North’s work. Beyond this, although these are more the exception than the rule, some authors have also emphasized the importance of Karl Polanyi’s thought for understanding North’s theory and purpose. But by my knowledge, the only other work to interpret North’s NIEH specifically as a response to Polanyi’s challenge is Claude Didry and Caroline Vincensini’s working paper on this topic. In this work, they outline how his understanding of institutions as ‘rules of the game’ differentiates his approach from what I have called the ‘neoinstitutional’, in this case the work of Oliver Williamson (economic theory) and Mark Granovetter (economic sociology).<sup>50</sup>

This view of institutions allows North, they argue, to move beyond the dichotomy of ‘markets vs hierarchies’, or in this case ‘markets vs institutions’ that characterizes the neoinstitutional approach. This is a significant insight, not least since the latter authors are important reference points for the development of NIE theory as a whole and for North’s own work. But unlike Didry and Vincensini, I am not content merely to compare these terms to neoinstitutionalist theory. We must also question whether this use of institutions, as well as North’s arguments about adaptive efficiency and about the historicity of markets, are valid “vectors of historical comparatism” (to use their phrase).<sup>51</sup> While I share their identification of these themes as central to North’s NIEH approach, as opposed to Williamsonian neoinstitutionalism, it is one of the burdens of this work to examine its success: both in terms of North’s own ambitions and vis-à-vis the rival anthropologies identified by North in his - so to speak - foundational papers of the mid-1970s.

In a more critical vein, Daniel Ankarloo has also noted the significance of the ‘Polanyian moment’ in North. In two papers, Ankarloo critiques NIE theory, concentrating on Williamson and North; and in a subsequent conference paper he has extended this critique specifically to North’s NIEH. This makes his work, although mostly unpublished, an important part of the rather small secondary literature on the NIEH itself.<sup>52</sup> Central to his critique, as also in this work, is the concern with the uses of historicity in North and the NIE in general: the assumption that, in Oliver Williamson’s words, “in the beginning there were markets” and that analysis of economic institutions should depart from the assumption of generalised market exchange.<sup>53</sup>

For Ankarloo and Palermo, the consequence of this assumption, while perhaps convenient for the neoinstitutionalist model of a trade-off between markets and hierarchies in modern economies, is that one of the central aspects of ‘capitalism’ – the presence of generalised market exchange – is assumed when the purpose of analysing economic institutions is to explain

---

<sup>49</sup> Ensminger (ed.) 2002.

<sup>50</sup> Didry and Vincensini 2011.

<sup>51</sup> *Ibid.*, p. 12.

<sup>52</sup> Ankarloo 2002; Ankarloo and Palermo 2004; Ankarloo 2006.

<sup>53</sup> Williamson 1975, p. 21; Ankarloo and Palermo 2004.

precisely such phenomena. As they put it polemically: “The problem is twofold: first, by assuming the existence of one primordial institution in capitalism, Williamson fails to explain it and then makes it universal; second, by identifying this primordial institution with the market, he inevitably falls into a sterile idealisation of this institution. Had Williamson assumed an initial starting point of pure planning or of a mixed institutional set-up, he would have been unable to explain some institutions of capitalism and the project would have been contradictory anyway. But, the problem is that the choice of the market as a natural institutional arrangement is also apologetic, since it idealises the market, transforming it from a historically defined institution into a universal category. And this process of idealisation occurs without any clear analytical justification.”<sup>54</sup>

Of course, Ankarloo recognizes that in North’s case, this Williamsonian assumption does not hold so simply, especially not in the later and more developed versions of his NIEH. “Having, at least in principle, further and further distanced himself from standard orthodox conceptions North has come the closest within NIE to acknowledging the merits in Polanyi and the demerits in orthodoxy. This has rendered him rather popular within unorthodox anti-neoclassical circles...”<sup>55</sup> Here, however, the contrast between Polanyi’s approach and that of North illustrates for Ankarloo where the problem lies.<sup>56</sup> Because this comparative critique touches on several major themes discussed here, it is worth a long quotation:

“To make a long story short, the disturbing problem of North’s economic history is all the same that the decisive institutions and human characteristics that are specific to capitalism have to be assumed to exist even when he admits that market capitalism is not at hand. Otherwise his explanations of institutions and organisations do not work. The market in North’s explanations remains the universal yardstick – the eternal reference point. But if we acknowledge that markets and the institutions of capitalism originate from somewhere, and evolve, we cannot assume them from the start. The market is not a universal yardstick for the comparison of institutions.

North knows all this, but in order for the explanation to work the economic logic of the market has to be described as present, even in the “absence of the market”. This is due to the fact that institutions evolve as rational responses to relative price changes, by, albeit “constrained”, micro-rational economic men, in his model. But such calculation cannot explain the origins of the market. Calculating what to buy only makes sense when there already are commodities on the market to buy. These calculations cannot at the same time be the cause of the market for these commodities.”<sup>57</sup>

Here, precisely North’s conception of institutions as rules, and markets and other allocation systems as alternative sets of such rules, that Didry and Vincensini saw as an advance over the

---

<sup>54</sup> Ankarloo and Palermo 2004, p. 417.

<sup>55</sup> Ankarloo 2002, p. 18.

<sup>56</sup> *Ibid.*, p 16-17.

<sup>57</sup> *Ibid.*, p. 18-19.

Williamsonian approach is criticized as assuming what needs to be demonstrated no less than Williamson's own market premise. The importance of historicity in North's NIEH is only a misleading appearance. For Ankarloo, the result is that insofar as North assumes institutions are the product of rational market action, and thereby that the operation of a market-like logic is assumed even when very 'alternative' allocation systems prevail, the historical differentiation in economic institutions and allocation systems so important to Polanyi is reduced to a set of variants of a general 'as-if market'. "History is adapted to support preconceptions of handed-down economic theory – and in this process history is not a source of knowledge, but is transformed into a 'narrative', where actual history is said to work 'as if' the neoclassical principles of 'market behaviour' are in ubiquitous operation – explaining not only 'the economy', but also close on all other aspects of social life."<sup>58</sup>

North's subsequent attempts to create a more sophisticated theory, by moving from the origins of allocation systems in their economic institutions to the origins of economic institutions in the structure of social and political life, has not found any more favour with the critics in this line. For Ankarloo, the result of the extension of the NIEH to ideology, culture, and other inputs into the formation of institutions is that the 'neoclassical' economic theory North intends to save is rendered irrelevant to the model itself: 'the economic' has disappeared, even as the theory has extended far into the domain of other social sciences.<sup>59</sup> We find a similar critique of Douglass North's work in some of Ben Fine and Dimitris Milonakis' critical writings on the history of mainstream economic thought. In discussing what they call the shift from "the principle of pricing to the pricing of principle" - i.e. the stages from Cliometrics to a developed anti-neoclassical NIEH theory I will discuss in the next chapter – Fine and Milonakis observe that "in North's hands, transaction costs has become a catchall phrase to be used as an explanatory tool in almost any historical or theoretical context... it becomes possible in principle to explain the existence of any institution in history by invoking the high transaction costs of some other theoretically possible institution."<sup>60</sup> As they conclude: "In the absence of markets, explaining the (historical) existence of another institution (labour services) through a cost/benefit analysis in relation to the market is historically meaningless."<sup>61</sup>

#### **IV. Summary of the central problems of this work**

Despite such observations, these critics also recognise the significance of North's work, not least insofar as this has entailed a systematic distancing from 'neoclassical' orthodoxy – as we will see, repudiations of neoclassical economics are a recurrent theme in North's theoretical writings. Given this fact, it may appear puzzling how Ankarloo as well as Fine and Milonakis persist in classifying North's work as neoclassical in its foundations, while recognising his attempts to

---

<sup>58</sup> Ankarloo 2006, p. 17.

<sup>59</sup> *Ibid.*, p. 12.

<sup>60</sup> Milonakis and Fine 2007, p. 49, 51.

<sup>61</sup> *Ibid.*, p. 50.



move beyond it.<sup>62</sup> This ambiguity is worth examining further, and is one of the motivations behind the present work. The critics' judgement that North has failed in his ambition to move beyond neoclassical economics rests on elaborate economic methodological considerations, involving precisely those questions about the nature of markets and of 'economic theory' that North saw as fundamental to his rebuttal of Polanyi. This latter critique is very important for assessing the success of North's enterprise, and for this reason I will have cause to refer to Ankarloo and Fine and Milonakis' observations again in this work.

But these critiques do not enter into the structure of North's most advanced theory in any great detail, nor do they assess the merits of North's arguments much beyond this methodological critique of his assumptions about markets. In this work I will seek to clarify where this ambiguity in economic and historical methodology comes from and how it manifests itself in the work as a whole. To do so requires a full-length examination of each of the levels of scope and abstraction in North's work, from the microfoundations to his ideas about historical evolution. It also means observing how these considerations in economic methodology operate as instances of the larger problem of primitivist and modernist perspectives: whether differences in economic behaviour and institutions in history require a different kind of social theory to explain them than the "economic approach" or "economic theory" *tout court* - as even the new mainstream still often sees it - can offer. The critics as well as the supporters of North's work emphasize that the central question of his approach, and the fundamental ambiguity within it, is how his view of institutions can contribute to a better historical understanding of this observed economic differentiation.

In other words, whether North's approach can make sense of markets is part of a larger question: what is at stake when the new mainstream of economic theory, as exemplified by North's developed NIEH theory, is applied to premodern history. Can it succeed – in terms of internal consistency as well as 'external' plausibility - in explaining the embeddedness of economic institutions in social, cultural, and psychological factors? What are the consequences for economic history of the anthropological assumptions about human sociality and individual rationality that underlie the 'economic approach' of the new mainstream North defends? These are questions characteristic of the primitivism-modernism debate, raised once more by North's unique answer to 'the challenge of Karl Polanyi'. This work provides some preliminary answers to these questions, and in so doing contribute to a better understanding of his NIEH project, as well as to an assessment of the success of this project.

---

<sup>62</sup> Fine and Milonakis 2009, p. 81; Milonakis and Fine 2007, p. 53-54; Ankarloo 2006, p. 2.

# Chapter 1: Douglass North's New Institutional Economic History

## Introduction: Beyond the New Economic History

In considering the origins and development of North's NIEH theory, a number of things must be kept in mind. Due to the complexity of his theoretical apparatus and the great variety of specific terms and concepts he has introduced in it, it would be unhelpful to rely wholly on a chronological overview of his thought. However, without a sense of chronology, the various stages of his NIEH research programme and even the theoretical context in which it was born will be lost to view. Of necessity, therefore, this chapter will not be a full and thorough discussion of all of North's theoretical writings, nor will it be entirely chronological and give due to every stage in the growth of the NIEH. It will concentrate on the specifically historical-theoretical works, rather than North's more purely economic-theoretical contributions or his few purely empirical case studies. In this way, the task of drawing out the 'hidden anthropology' and methodological constraints on North's own work is best accomplished. This is not to dismiss, of course, the importance of the wider intellectual-historical inputs into North's NIEH, especially given North's frequent reliance on theory from economics, politics, biology and anthropology to support his analysis. But in this chapter, I will first concentrate on North's NIEH theory itself.

Note that this also means that the more applied, case study like works of North – such as his attempts to estimate the size of transaction costs within the American economy, for example – will be given relatively short shrift. Not because these are not important, but because, as we will see, North himself has generally disobeyed his own strictures that theoretical disputes cannot be fought out at that level and are only fruitfully debatable in testable propositions.<sup>63</sup> As much philosophy of social science can attest, disentangling theoretical claims for the purposes of individual hypothesis testing is often impossible.<sup>64</sup> North's theory mostly stands or falls on its own merits, and whatever testable hypotheses have been derived from it by North or others have had little effect on the development of the theory as a whole. Therefore, I feel justified in mainly concentrating on his theoretical monographs, with a bias towards his most sophisticated and contemporary work. The latter not only provides a best case for North's NIEH approach, compared to his earlier work, but has also been discussed much less in the secondary literature – not least as much of the latter dates from before the publication of *Understanding the Process of*

---

<sup>63</sup> E.g. North 1977, p. 715.

<sup>64</sup> Most notably this problem is described in the so-called Quine-Duhem thesis. For discussions relevant to economics, see: Stanford 2013; Sawyer, Beed and Sankey 1997; Hausman 1992, p. 306-307. Philosophers of economic theory for this reason often theorize the field through a Lakatosian perspective: see e.g. Heijdra and Lowenberg 1986.

*Economic Change* (2005) and North's collaboration with John Wallis and Barry Weingast, *Violence and Social Orders* (2009).<sup>65</sup>

Although Douglass North's best-known works date from the 1980s onwards, their roots go back to his very first reconsiderations of neoclassical orthodoxy. That is to say, while North's NIEH research programme was developed out of the response to Polanyi's challenge and the need for a theory that could account for premodern economic history, these problems first appeared on the intellectual horizon when North ceased to be a Cliometrician and became an institutionalist. We first find North in the theoretical literature as one of the main exponents of Cliometrics, a view diametrically opposed to those of his most contemporary works. We must begin, therefore, with this school of economic history and North's role within it.

Cliometrics, also known as the New Economic History, is a product of the post-war period, becoming prominent for the first time in particular in the United States in the 1960s. It arose as a response to two things: the perceived need to go beyond the traditional economic historical approaches, which were often closer to social history than to economics, with a method that would be based on the best neoclassical economic theory of the time, as well as the development of new mathematical and statistical techniques in econometrics that would make the application of neoclassical quantitative models to historical cases possible.<sup>66</sup> For this reason Robert Fogel – one of its most influential practitioners – dubbed it simply “econometric history”.<sup>67</sup>

From the very beginning, therefore, Cliometrics was committed to deriving its methodology, as well as the interpretation of its models, from the ‘economics approach’ of the day in much the same way as North uses that term in his discussion of Polanyi: taking for granted that mainstream economic theory offers both a necessary and sufficient means to understand economic behaviour in any historical case studied. The implications of this attitude for the relations with the ‘traditional’ economic historians, whose narrative approaches looked to history as an academic discipline for inspiration rather than to economics, are summed up well by the New Economic Historian, Lance Davis: “Once terms are defined and the questions framed, an explanatory model must be built. Since the subject is economic history, the models frequently draw on economic theory for their form. These models provide a basis for sorting the useful from the useless, they suggest methods of possible verification, and they provide simplified 'explanations' of a far from simple world. With a model the assumptions are made explicit and it is possible to examine reality in the light of the logical system. Like mathematical economics, the 'new' economic history 'may not be able to say much, but at least the reader is aware of what has been said'. Traditional economic history also has its models, but all too frequently they are implicit, not explicit.”<sup>68</sup>

---

<sup>65</sup> North 2005; North, Wallis, and Weingast 2009.

<sup>66</sup> The classic statement is Fogel 1966; a more contemporary perspective is given in Diebolt 2012.

<sup>67</sup> As in the title of Fogel 1964.

<sup>68</sup> Davis 1966, p. 658.

Here we find for the first time an appearance by Douglass North as a major theorist in economic history – but not yet anything like the kind of theorist he would become. Rather, he appears as, *nota bene*, “chief propagandist and entrepreneur” of Cliometrics in the first meetings of the Cliometrics forum at Purdue University.<sup>69</sup> We are here not primarily concerned with his role as Cliometrician, but rather with his critique of this school of thought. It is, however, worth noting that despite North’s subsequent defection from this approach, Cliometrics is still one of the most widely practiced and influential schools or methods in economic history, with Robert Fogel being North’s co-laureate for the only Nobel Prize in Economics ever awarded for primarily historical study. In terms of theoretical prestige within economics-oriented historiography, it is perhaps only rivaled by the NIEH itself. Of course, a systematic critique of Cliometrics cannot be engaged in here; our purpose is to trace the development of the NIEH.<sup>70</sup>

Our main cause for mentioning Cliometrics is because of its subsequent role: before the ‘Polanyian moment’, North’s theoretical work was formulated for an important part in terms of a critique of the Cliometric approach he had previously adhered to. In this sense then, Cliometrics initially played the role of foil to North’s own theoretical development, in which it would later be joined more significantly by Polanyi – or rather, by the substance of his ‘challenge’. This is perhaps best illustrated by reference to his Presidential Address, held upon taking that position at the head of the Economic History Association in the United States. That North had been selected for this position is a sign of his stature already at that time. But his use of the occasion to critique Cliometrics is more interesting, especially because of the substance of this critique. His lecture, ‘Beyond the New Economic History’, expresses a fundamental ambivalence about the virtues and downsides of what this reliance on the ‘economics approach’ had achieved: “What the new economic history contributed was the systematic use of theory and quantitative methods to history. The use of a scientific methodology [*sic*] has put a distinctive stamp on this approach, which clearly delineates it from the old economic history, but it is the theory that provides a particular cast to the contribution. It is the systematic use of standard neo-classical economic theory which both has provided the incisive new insights into man's economic past and also serves to limit the range of enquiry.”<sup>71</sup>

We may identify here two characteristics of North’s thought post-Cliometrics that he would retain for the rest of his theoretical career. Firstly, the commitment to seeing “the systematic use of theory”, namely mainstream economic theory, as the *sine qua non* of scientificity: this, plus quantitative methods (less prominent subsequently), amount to a “scientific methodology” *tout court*. This also heavily implies that the traditional economic historians did not have a scientific methodology, thereby in turn suggesting strong and particular

---

<sup>69</sup> *Ibid.*, p. 659.

<sup>70</sup> But see for a historiographical critique: Boldizzoni 2011. A general discussion of methodological problems of causal inference in social science can be found in Freedman 2010. For economic-philosophical problems with causal inference in econometrics, see e.g. Reiss 2013, p. 117-141, 161-173.

<sup>71</sup> North 1974, p. 1.

standards of scientificity and justification. But if the need to rely on some form of theory that can pass muster by economic standards of scientificity is one thing – perhaps reflected in his mostly rhetorical appeals to the importance of testing - neoclassical economics being that theory is another thing. And this second characteristic, North’s dissatisfaction with using neoclassical economics to play that role, is evident here.

Continuing his critique, he writes of this theory: “The limitations [of the New Economic History] are those of the theory. Neo-classical economic theory has two major shortcomings for the economic historian. One, it was not designed to explain long-run economic change; and two, even within the context of the question it was designed to answer, it provides quite limited answers since it is immediately relevant to a world of perfect markets-that is, perfect in the sense of zero transaction costs: the costs of specifying and enforcing property rights. Yet we have come to realise that devising and enforcing a set of rules of the game is hardly ever costless and the nature of these costs is at the very roots of all economic systems' problems.”<sup>72</sup> Here the second enduring characteristic cited above is further specified. If the dismissal of traditional approaches to economic history is based on their inadequate use of economic theory, the dismissal of neoclassical economics in turn is based on two more specific considerations. Firstly, neoclassical economics is not able to deal with the *longue durée*, as the *Annales* historians called it: the problem of premodern economic history we encountered in the Polanyi paper is already presaged here. Secondly, neoclassical models are insufficiently realistic, in the sense that they make assumptions that render the results irrelevant to real world economic systems. All such systems encounter market failures and require institutional solutions to such failures, ‘rules of the game’: without them, little relevant can be said about the ‘game’ itself.

Such a consideration might lead one to expect North to develop an interest in applying the approaches of other social sciences not based in neoclassical economics, insofar as for example economic anthropology has been concerned with understanding economic institutions. But this possibility is not open to North due to his requirement that any scientific approach to such problems still be rooted in some form of mainstream economics, which is inherently universal in scope. The question at this stage is not how institutions should be methodologically understood, but simply what theory would best replace neoclassical economics in describing the universal problems of resource allocation, the subject of economics as a discipline: ““Let me emphasise that a study of the rights associated with the use and transfer of resources is as relevant in socialist societies as it is in capitalist ones. The rules of the game determine efficiency and the distribution of income in any society: classical Greece, the Roman Empire, the manorial system, or Yugoslavia today. To say that government owns the means of production or even that there are very limited markets and therefore that the study of property rights is irrelevant is simply to fail to understand that *all economic systems face a common set of problems about the use and*

---

<sup>72</sup> North 1974, p. 2.

*transfer of resources*, whether done via the market, via government, or via voluntary organisation.”<sup>73</sup>

Effectively, he has already chosen his theory: it is to be the NIE, for whom such problems appear as choices between markets and hierarchies (vertical or horizontal). The NIE is of course mainstream without being neoclassical (orthodox), in Colander, Holt and Rosser’s formulation – at least in methodological terms. (This should not mislead one into thinking it was necessarily popularly accepted among economists and economic historians at the time, or North would have had no need to defend it.) This makes it ideally suitable for North’s search for a theory that would accept the general methodological criteria and norms of scientificity of economic orthodoxy, but that would not imply its neoclassical assumptions and models. Moreover, as North would explain, such an approach has an advantage in dealing with the longer-term historical dimension as well: “Our emphasis on the last two hundred years, from the Industrial Revolution onward, is a misallocation of scholarly resources. We should spend much more time on the preceding 9800 years of man’s economic history than on the last 200. (...) In fact, the overriding issue of man’s economic history has been the relationship between population growth, diminishing returns to a relatively fixed factor, and man’s efforts to alter institutional arrangements to overcome this dilemma. Our emphasis on the present blinds us to the fact that few of man’s economic problems are new - that most have recurred endlessly in the past.”<sup>74</sup> The combination of the real-world ‘relevance’ (or realism) of institutionalism with the universality of the NIE approach is North’s justification for its superiority over rival approaches.

What this does not yet tell us is what central *concepts* for North define both the ‘economics approach’ - or ‘economic theory’ *tout court* - including its virtues he would defend against Polanyi in his subsequent paper, and the alternative to neoclassical theory he is proposing here. He does not call the latter the NIE, but rather uses its more revealing name for this purpose: transaction cost economics. In a neat summary, he defines this central conceptual question on both counts: “(1) Neo-classical theory has been a powerful tool of analysis of the new economic history and has demonstrated repeatedly that it can shed light upon our economic past. In fact, I would put it stronger: A theory of choice - the self-conscious application of opportunity cost doctrine - is essential to the framing of meaningful questions in economic history. (2) Transaction costs are the link between neoclassical theory and a broader theory of property rights. The explicit historical study of transaction costs opens up new horizons for the economic historian. Much of the productivity change in past history has been a consequence of reduced transaction costs and their study suggests a quite radically different history than we read in the standard explanation.”<sup>75</sup>

---

<sup>73</sup> *Ibid.*, p.3; emphasis added.

<sup>74</sup> *Ibid.*, p. 5.

<sup>75</sup> *Ibid.*, p. 3-4.

Let us unpack these two theses in light of this problem of conceptual definition, i.e. in how North makes this theory choice operative. It is clear that the central characteristic of the new approach is to be *transaction costs*: it is these that link the general but irrelevant models of neoclassical economics to a “broader theory of property rights”. We do not yet know what that broader theory would be, but it is clear that it would have to derive its main inspiration from an NIE-type approach, centered on transaction costs. Yet the notion of a ‘link’ with neoclassical economics once more shows the ambivalence, already so clear here, of North’s attitude to neoclassical economics. Why link to something that is not real-world relevant? Because, for North, the other central concept the alternative economic historical method must have is the orthodox “theory of choice: the self-conscious application of opportunity cost doctrine”. Choice theory, the microeconomic foundation of neoclassical economics, is to be retained. It is *essential*, i.e. a strict requirement, for not just a scientific but even simply a *meaningful* approach to economic history. This simultaneous rejection of neoclassical economics with an appeal to the indispensability of its microfoundations implies a necessary outcome of theorizing on this basis: North would have to build his subsequent macro-level theory, however different from neoclassical economics, on microfoundations that are not at all so different. In the below discussion of the birth of the NIEH proper and its subsequent development as a research programme, we will see how North reconciles these aspects, and what this entails for his theoretical apparatus.

### **The birth of the NIEH: from the efficient social contract to Polanyi’s challenge**

The Presidential Address does not, of course, come out of nowhere. Indeed, it is not the first time institutionalism appears in North’s work. While the Address is his best systematic theoretical justification of the need for transaction cost economics to replace neoclassical theory in economic historical analysis, at least in the period before the Polanyian moment, in previous years he had already made some first steps towards this approach. In 1970 North’s paper with Lance Davis on institutional change, written in a consciously neoclassical vein, had revealed the weaknesses of this approach to understanding that subject for him. Rather remarkably, in the abstract the authors describe the theory deployed in their paper as “at some points woefully weak and the explanations at times incredibly simplistic.”<sup>76</sup> With this began, perhaps, the search for an alternative. It should therefore come as no surprise that we find in his joint writings with Robert Thomas in this period an approach merging neoclassical and institutional arguments. Characteristic of this is their paper on the sources of Western economic growth. While the structure of the economy and its growth presented in this paper is a traditionally neoclassical model, in which Malthusian demographic factors, relative factor prices, and technological change interact, the novelty rests in the idea that one outcome of these mechanisms of neoclassical growth theory was the development of new and historically unique institutions. These institutional innovations and the changes in property rights they entailed allowed the West

---

<sup>76</sup> Davis and North 1970.

to shift from the historical virtually zero growth pattern to one of explosive economic expansion. Put theoretically, the neoclassical Malthusian model gives way to an institutional explanation for historical differentiation in the West in the modern era – here beginning roughly in the 18<sup>th</sup> century, though with its roots in the Middle Ages.<sup>77</sup> Most interestingly for our purposes, attached to this more or less ‘orthodox’ model is an explanation of institutional formation and change based on transaction costs, which in turn result from uncertainty, externalities, economies of scale, and the costs of information.<sup>78</sup>

An application of these ideas is found in their joint paper on the manorial system of the Middle Ages, alluded to above. This paper is interesting for two reasons: one because it reveals the growing interest in explaining longer-term economic historical developments, including changes in allocation systems, on North’s part, and secondly because the theory North and Thomas argue for in this paper uses Marxism as a foil. In this pre-Polanyian phase, the questions of historical comparativity are not yet so clear to North, but he and Thomas note the absence of a consistent ‘theoretical’ explanation (i.e. one based on economic theory, not historical explanation) of the shift from feudalism to capitalism. Since neoclassical approaches do not provide one either, the default is the Marxist, something clearly uncongenial to North and Thomas. For this reason, the ‘proof of the pudding’ is in demonstrating an approach based on “economic theory” that can do better than the Marxist explanation. This theory, once again, is the economics of transaction costs and institutional change – especially the institution of property rights.<sup>79</sup> How is this done? North and Thomas reimagine feudal serfdom not as a form of involuntary servitude or as a form of ‘direct exploitation’, but rather as a *contractual arrangement*: a contractarian analysis of feudalism where serfs accept their labor dues in return for protection.<sup>80</sup> The institutions that allowed this contractarian relationship to function were the customs of the manor, which both stated the traditional dues of serfs and protected them from arbitrary changes in the ‘contract’ in question. By analyzing these customs as property rights institutions, stating the division of property in labor between the different parties, orthodox contract theory can be applied, despite the observation that such theory was designed “in the context of a market economy”.<sup>81</sup>

However, a neoinstitutionalist approach, one complementary to this contractarian view and to orthodox contract theory, is also necessary. The explanation for this peculiar contract arrangement is the absence of a central authority able to provide protection, the high cost of contracting that requires binding political and normative structures, and the high risks of trading and of providing subsistence goods through markets. It is precisely the absence of markets that created the manorial system, in other words. For North and Thomas, this analysis is therefore a

---

<sup>77</sup> North and Thomas 1970.

<sup>78</sup> *Ibid.*, p. 6.

<sup>79</sup> North and Thomas 1971, p. 777-778.

<sup>80</sup> *Ibid.*, p. 778-779.

<sup>81</sup> *Ibid.*, p. 783.



demonstration of the strengths of their New Institutional approach: only by understanding the property rights arrangements and contractarian normative system as “fundamental institutions”, operative at a societal level, on which arise the individual contractual arrangements more familiar to neoclassical economists, can one make sense of manorial serfdom. The neoclassical case of individual contracting over property rights becomes a special case of a more normative, contractarian, and institutional approach favored by North and his colleagues.<sup>82</sup> The ‘fundamental institutional arrangements’, whose importance is reflected in the high costs of changing them, are necessarily cases of rules operating in societies or economic formations as a whole, whose agents are groups, not individuals.<sup>83</sup> This means, then, that North required a theory of comparative fundamental institutions, differentiating various economic formations according to the interaction of neoclassically understood economic behavior – private contracting in markets – with more fundamental collective or intersubjective normative structures and ‘contractarian’ political arrangements. In other words, North needed a comparative theory of the embeddedness of economic behavior in political and social institutions. Albeit, at this stage, only yet insofar those affect property rights. This sets the stage for the Polanyian encounter.

Since the challenge of Karl Polanyi for North’s approach has been discussed at some length in the introduction, I will not repeat that here, except to note how Polanyi’s challenge transforms the neoinstitutionalist interpretation of historical change North had been developing so far. This is not simply a question of replacing Marxism with substantivism as the rival theory to be beaten at its own game, although North’s dismissal of Marxist approaches in favor of more interest in (if not always approval of) Polanyi and the work of anthropologists can subsequently be noted.<sup>84</sup> More importantly, the impetus of Polanyi’s challenge has been to provide the sophisticated historical comparative theory that North wanted to meet the challenge ‘head on’. As we will see, the reappearance of the concept of ‘embeddedness’ in North’s work through its legacy in economic sociology also signifies the endurance of the Polanyian challenge in the background of North’s work, even as it is equally oriented towards problems of economic growth and institutional change in a more traditional way.

Moreover, from the Polanyian moment onwards we will see the scope and ambitions of this theory expand further and further, incorporating ever more of what we will call “successive endogenization” of Polanyi’s ‘social background’ into his institutional theory, and this will also mean an increasing distance between North’s NIEH and the neoinstitutionalist and neoclassical elements still quite dominant in the work just discussed. This expansionary evolution is the

---

<sup>82</sup> *Ibid.*, p. 785-786.

<sup>83</sup> *Ibid.*, p. 786: “When conflict has arisen about fundamental institutional arrangements, that *group* which achieves power has tended to impound its new rules in a written “constitutional” document... This incentive has proved greatest when a *group* has foreseen that its decision-making ability... may not be permanent”.

<sup>84</sup> See e.g. his brief and rather dismissive review of a collection of essays on political economy by anthropologists, where North complains about the “strait jacket imposed by simple-minded Marxian theory”, but notes the influence of Polanyi. “More than anything else”, he concludes, “this book demonstrates the naivete of anthropologists on the subject.” North 1979a, p. 1088.

necessary result, I think, of his adoption of Polanyi's challenge. This is a point to be kept in mind, which I will explore in the discussion below as well as in the historical discussion of institutionalism and Polanyi's work in chapter 2. I will now turn to the next two stages of North's work, the 'middle period' of the full-fledged development of his NIEH theory and the most recent stage in which it evolves into a grand historical narrative straddling the social sciences.

Understanding North's approach to Polanyi as one of successively endogenizing 'the social' into 'the economic', rather than the opposite approach suggested by Polanyi, depends on defining the 'social' as well as 'the economic' into which it is endogenized. Eggertsson does not go beyond mentioning endogenization of "social and political rules" and the "structure of political institutions", without further defining the meaning of 'social' or 'political' in this case.<sup>85</sup> Since the point of Polanyi's embeddedness concept is to defend endogenization in the other direction, to see the economic (whatever it may be in a given case) as part of a wider social framework, it may make sense to keep Polanyi's definitions of these terms in mind when understanding this term. For Polanyi, "the individual's *motives*, named and articulated, spring as a rule from situations set by facts of a non-economic—familial, political or religious—order", as contrasted with the sphere of the economic, where "fear of hunger and hope of gain" are the dominant motives.<sup>86</sup>

Yet Polanyi also defines the economy as the sphere of the production and distribution of goods. More useful than the focus on motivation alone (given also the discussion of rationality below), we may follow Polanyi's larger research programme in "drawing on what he calls 'the societal approach' ... conceiving of economic life as a *totality* of relations and institutions that goes beyond transactions of goods and services. Here, rather than investigating the changing place of economy in society, Polanyi offers a method for studying economic life that is in contrast with the methodological individualism of economics. As a result, embeddedness emerges as a methodological principle, and not as an analytical proposition... Thus, 'economic', in its substantive sense, refers to all interactions with nature and other human beings in the pursuit of livelihood, and not to a specific type of behaviour".<sup>87</sup> As Polanyi put it, "the substantive meaning of economic derives from man's dependence for his living upon nature and his fellows. It refers to the interchange with his natural and social environment, insofar as this results in supplying him with the means of material want-satisfaction"<sup>88</sup>. This as contrasted with 'the social': "the social process is a tissue of relationships between man as biological entity and the unique structure of symbols and techniques that results in maintaining his existence"<sup>89</sup>. This I take to include all of the institutions and symbolic spheres of interaction with other humans and

---

<sup>85</sup> Eggertsson 1990, p. xiii.

<sup>86</sup> Polanyi 1968, p. 85, 82.

<sup>87</sup> Gemici 2008, p. 15, 17.

<sup>88</sup> Polanyi, Arensberg and Pearson 1957, p. 243.

<sup>89</sup> Polanyi 1968, p. 116.

with nature insofar as *not* covered by the definition of the economic immediately previously given.

### North's NIEH in its second stage

In the Polanyi paper, North had written that “I wish to make the affirmative point that as yet we have not even tried to see how far economic analysis will take us in explaining institutional arrangements. (...) Transaction cost analysis is a promising framework to explore non-market forms of economic organisation.”<sup>90</sup> The first of the great NIEH theoretical monographs, his 1981 book *Structure and Change in Economic History*, is an expression of this idea.<sup>91</sup> Note that this represents a gradual extension in terms of the domain of his research programme. Whereas before the Polanyian moment the purpose of his neoinstitutional analysis was to understand the economic impact of ‘fundamental’ institutions insofar as these represented what I have called a contractarian solution to widespread market failures, as in the case of the feudal European manor, now this is to be extended to nonmarket allocation systems in general. Perhaps a subtle difference, but one worth noting: the effect of Polanyi’s primitivist viewpoint, in which the absence of market exchange as organizing principle of economic behavior is central, can here be felt.

While most of the institutionalist theory in *Structure and Change* clearly builds on the first NIEH statements discussed above, the purpose is now clearly put in terms of a long-term comparative analysis of institutional structures and the different economic formations that result from them, with an aim to studying the relative ‘*performance*’ of each of these structures. This performance mainly consists of the advantages security of property rights offers for the emergence of efficient markets, which in turn stimulate economic growth. This aspect of the work draws still on mainstream economic theory ideas of the benefits of market exchange and private property, i.e. the economic desideratum as equalizing private and social returns to entrepreneurial activity as much as possible. But the purpose is also more explicitly still than in the paper with Robert Thomas to provide a rival theory of long-run historical differentiation in economic systems to primitivist accounts, here mainly identified with the Marxist<sup>92</sup>; so that one reviewer at the time described it, perhaps somewhat surprisingly, as a “useful complement to recent attempts... by Perry Anderson and Immanuel Wallerstein”.<sup>93</sup> Most importantly, in this stage the contractarian nature of institutions, in other words their presentation as the results of some process of market-like bargaining over security versus property rights, moves more to the background (without wholly disappearing). North in this work more clearly perceives feudal

---

<sup>90</sup> North 1977, p. 709.

<sup>91</sup> North 1981.

<sup>92</sup> “The only building blocks the economic historian has to work with are classical, neoclassical and Marxian theory...” North 1981, p. 59. The influence of Marxism in turn on Polanyi’s own positioning in the primitivism-modernism debate (as well as on the latter’s commitment to socialism) is well attested: see Dale 2010, p. 241; Selwyn and Miyamura 2014; Block 2008.

<sup>93</sup> Millward 1983, p. 964-965.

elites, as well as those of other allocation systems in the premodern past, as essentially predatory. But the institutions sustaining their extractive powers undergo change mainly through exogenous factors familiar from standard neoclassical approaches: technological change, demographic shifts, and expansion in volume of trade, as well as the more ‘neoinstitutional’ concern with the costs of measuring goods. Effectively, then, his approach in this work is to identify the neoclassical factors normally held constant and to propose NIE theory as a means to make these constants into variables, so that long-term economic change is better understandable. This illustrates again the methodological ambivalence of North’s NIEH as an institutionalist theory that both criticizes and departs from neoclassical assumptions.<sup>94</sup>

One other important development since the Polanyi paper is the realization that if the cultural, social, and psychological inputs into normative institutions is to be taken seriously, and especially to be - as Polanyi had argued - a subject of comparative analysis, this requires a study of ideology. At this stage, this concept is perhaps not yet very well developed in North’s work, it is fair to say; at this point North’s discussion of ideology as well as the role of free riding in institutional reproduction and change is mainly concerned with pointing to gaps in the theory than in filling them.<sup>95</sup> But this book does signal for the first time in a serious way the need for a theory based on the ‘economics approach’ like North’s to deal with questions of power, altruism and ideology. It is worth observing this because these are precisely the kind of subjects that the more orthodox economic historians preferred to avoid and which Marxists, Polanyians and others thought it one of the advantages of their own framework to address.

Moreover, it had an important consequence: North’s abandonment of the contractarian efficiency view of institutions. Indeed, in his subsequent monograph, North opens by contrasting the purpose of *Structure and Change* as well as his later work with the neoinstitutionalist approach of Oliver Williamson, for whom – according to North – institutions or the results of institutional competition are always optimal solutions to substitute for markets. Including ideology, the predatory nature of elites, and free riding problems into the economic historical theory means, for North, to question such a Panglossian assumption: he would come to dismiss this neoinstitutional efficiency view as ahistorical.<sup>96</sup> Insofar as the spirit of Karl Polanyi pointed North the way to taking nonmarket allocation systems seriously, and therewith their cultural and social, i.e. ideological, foundational institutions, this abandonment of an overly Panglossian market-like interpretation of institutions is another major consequence of the Polanyian moment.

In quite a bit of the less recent secondary literature, especially the critical works, this 1981 book is taken as a representative example of North’s NIEH approach. For an important part this is because the latter’s most recent works were not yet published at the time of these reflections, of course, but it is regrettable because much of the substance of his 1981 book is restated in more

---

<sup>94</sup> A similar observation can be found in Rostow 1982, p. 299; Cf. North 1978a, p. 963.

<sup>95</sup> North 1981, p. 68.

<sup>96</sup> *Ibid.*, p. 7.

complex, ambitious, and sophisticated ways in subsequent works, beginning with his 1990 book *Institutions, Institutional Change and Economic Performance (IICEP henceforth)*.<sup>97</sup> For this reason, I will not dwell much further on *Structure and Change* and move on to the latter, which is perhaps the ‘best case’ expression of North’s NIEH in this stage. One can suffice with noting that if the previous work is mainly about defending the NIEH approach against neoinstitutional, neoclassical and Marxian rivals, the burden of *IICEP* is to clarify the conceptual apparatus: it is therefore of great significance to understanding the terms of North’s historical framework.

North proceeds in *IICEP* to “[delve] much more deeply than the earlier studies into the nature of political and economic institutions and how they change. The specification of exactly what institutions are, how they differ from organisations, and how they influence transaction and production costs is the key to much of the analysis.”<sup>98</sup> In fact, *IICEP* sets out as clearly and concisely as can be expected the fundamental concepts mentioned in the book’s title and their interaction in North’s framework, and unlike his previous works, is explicitly aimed not just at economists and economic historians but at “other social scientists” as well.<sup>99</sup> North establishes that the central *explanandum* for the NIEH is not institutions per se, but *cooperation*, and how various allocation systems and institutional frameworks help (or do not help) to bring it about. However, keep in mind that for North efficient markets are the optimal form of such social cooperation, and therefore other allocation systems resulting from different institutional structures are judged by that standard: “the central focus is on the problem of... the cooperation that permits economies to capture the gains from trade that were the key to Adam Smith’s *Wealth of Nations*. The evolution of institutions that create an hospitable environment for cooperative solutions to complex exchange provides for economic growth.”<sup>100</sup>

The question whether such ‘Smithian cooperation’, i.e. capturing the gains from trade, succeeds or not depends in a given historical economy on the institutional framework of that society in question. If the subject of investigation for North’s NIEH at this stage is then the problem of how to achieve efficient market exchange that promotes economic growth, the method to get there is the by now familiar means of NIE-inspired analysis of institutions. For North, no ambition will do short of replacing the received comparative methods, both of neoclassical economics and of the other social science theories not “integrated into economics and economic history” with the New Institutionalist approach to economic history. As he proclaims: “That institutions affect the performance of economies is hardly controversial. That the differential performance of economies over time is fundamentally influenced by the way institutions evolve is also not controversial. Yet neither current economic theory nor cliometric history shows many signs of appreciating the role of institutions in economic performance because there as yet has been no analytical framework to integrate institutional analysis into

---

<sup>97</sup> North 1990.

<sup>98</sup> *Ibid.*, p. vii.

<sup>99</sup> *Ibid.*, p. viii.

<sup>100</sup> *Ibid.*, p. vii.

economics and economic history. The objective of this book is to provide such an underlying framework. The implications of the analysis suggest a reexamination of much social science theorizing in general and economics in particular, and provide a new understanding of historical change.”<sup>101</sup>

For the NIEH, institutions are the rules and norms of 'the game'. These rules are constraints, created by humans, on human action. These institutions, in turn, “structure incentives in *human exchange, whether political, social, or economic.*”<sup>102</sup> This small sentence can easily hide the enormous extension of the domain of North’s NIEH approach in the direction of the general comparative theory of economic behaviour and social structure suggested by the works of Marx or Polanyi. Institutions structure incentives not just in the classically ‘economic’ sphere of market exchange, but in all other spheres as well. Instead of making market exchange (or its hierarchical alternatives) in the modern age the domain of institutionalist theory, as Williamson and others do, the domain is now ‘human exchange’ altogether.<sup>103</sup> Institutions have their effect on economic performance by their impact on the costs of transaction and production, in accordance with the transaction cost economics approach. These costs of transaction and transformation make up part of the total costs of an economy.<sup>104</sup> Each institutional arrangement, on which the 'players' of its 'game' rely, provides pervasive externalities, which can reproduce the need for these institutions, but also provide incentives for incremental change. This incremental change is introduced into the system by what North calls “entrepreneurs in political and economic organisations”, who change the system “at some margin”.<sup>105</sup>

It is worth considering the implications of this idea of institutional change, since this topic will reappear in the more contemporary works of North discussed both below and in subsequent chapters. They are firstly that the suggestion seems to be that the political sphere of institutional change operates in some way analogously to the neoclassical economic idea of entrepreneurship: firms or individual market agents translate, under conditions of limited information, existing deviations from general equilibrium (‘gaps in the market’) into opportunities for profit, depending on their degree of risk aversion. The structure of the market (number of firms, commodities supplied) is changed by entrepreneurial activity such that entrepreneurs privatize the residual social benefit to themselves as residual profit, a return for their risk-taking. In Anthony Endres and Christine Woods’ terms, “entrepreneurs will arrive at optimally imperfect decisions when considering whether or not to exploit a profit opportunity which reveals itself as a market divergence between revenue and costs in market disequilibrium states.”<sup>106</sup>

---

<sup>101</sup> *Ibid.*, p. 3.

<sup>102</sup> *Ibid.*, p. 3. Emphasis added.

<sup>103</sup> Unfortunately, no definition for the term ‘human exchange’ is provided. It seems to mean any kind of social, i.e. intersubjective, interaction.

<sup>104</sup> North 1990, p. 6.

<sup>105</sup> *Ibid.*, p. 8.

<sup>106</sup> Endres and Woods 2006, p. 191-192 (“*Modern Theories of Entrepreneurial Behavior*”); Demsetz 1983.

Secondly, the larger implication drawn from this is that the processes of contestation and institutional change in a society, which we may for simplicity call politics, are market-like. While it is clear that in North's perspective the states of disequilibrium or market failure that permit entrepreneurial activity to obtain returns are pervasive, it equally appears that the incentive structure for institutional change is analogous to the incentive structure for profit: entrepreneurs change institutions to their own benefit along some margin, and do so depending on their 'taste' for (political) risk. In this organizations play the equivalent role of firms (and indeed include firms), namely collective agents made up of individuals who join them in order to better effect the benefits of institutional change (or preservation) as suits them: "Organizations are created with purposive intent in consequence of the opportunity set resulting from the existing set of constraints (institutional ones as well as the traditional ones of economic theory) and in the course of attempts to accomplish their objectives are a major agent of institutional change."<sup>107</sup> But the incentive structure of entrepreneurs or organizations, given a particular set of institutional constraints, is in turn constrained by limited information. Here, North departs slightly from this neoclassical model of institutional entrepreneurship, insofar as – having abandoned the efficiency assumption of institutions – he attaches weight to the way in which entrepreneurs perceive and interpret these arbitrage opportunities. We have seen this already in embryonic form in his observation about the importance of ideology. As we will see, this factor will play an increasingly central role in the further development of his NIEH theory.

The result is that while transaction costs in market exchange may generate imperfect property rights arrangements, similar transaction costs in "political markets" as well as the ideologies ("imperfect subjective models") of the agents involved may sustain such imperfections over time.<sup>108</sup> This distinction between institutions as constraints and organizations and 'political entrepreneurs' as agents permits North to counter Armen Alchian's more neoinstitutional approach, in which evolutionary pressures of competition will weed out inferior institutions in favor of contextually better ones, until the point where, in neoclassical terms, there are no bills left on the sidewalk.<sup>109</sup> Since institutions are the 'environment' and the organizations and political entrepreneurs the 'organisms' in this analogy, North seems to suggest, the latter adapt to the former, creating institutional 'lock in' – what he would later refer to as *path dependence*. The costs and benefits of changing the environment itself, after all, depend on the organism in question and its abilities to affect it. They cannot be taken for granted.

North's model of political action and change, then, entails a combination of fairly conventional assumptions about markets and entrepreneurship and their application in a more evolutionary framework to processes of institutional change, even when the latter have not traditionally been studied as part of that model of entrepreneurial activity. It is much the same in

---

<sup>107</sup> North 1990, p. 5.

<sup>108</sup> *Ibid.*, p. 8.

<sup>109</sup> Alchian 1950; North 1990, p. 7-8, 81.

his model of intellectual change and its effects on institutions.<sup>110</sup> This is, I would suggest, quite in the general line of North's use of NIE theory. Having corrected neoclassical microeconomics to account for evolutionary and historical dynamics and for imperfect information, the domain of its assumptions about individual economic behaviour and rationality is extended to the study of institutions. These institutions in turn are conceived of as constraints on that same individual behaviour. In this way, the interaction between individual (or collective agent) and institutional context is analogous to that in the neoclassical model between entrepreneur and disequibrated market: the model of individuals' relations to their society is that of agents to a market-like setting. I will take up these themes further in chapter 3.

Before moving on to North's further development of the NIEH, one important further theoretical innovation must be pointed out. This has to do with the effects of this individual-institutional dynamic on North's larger comparative picture. Recall that, for North, the purpose of this institutional understanding is ultimately to provide a comparative theory of different, especially nonmarket, allocation systems and their respective performance. So far, this performance has been understood simply in terms of economic growth, along the lines of conventional growth theory, with the caveat that only institutional arrangements favouring secure private property rights can achieve the virtuous cycle between technological and entrepreneurial innovation and relative factor price shifts that such growth theory describes. The standard criterion in neoclassical economics for evaluating such arrangements is allocative efficiency: the stipulation in marginalist theory where efficiency is defined such that every good or service is produced with marginal benefit to consumers being equal to the marginal cost. Under standard (highly unrealistic) Pareto conditions, only a private, competitive market economy can provide allocative efficiency. But for North, this criterion is irrelevant to economic historians for much the same reasons as provided by his general critique of applying neoclassical economic theory to this field: it does not allow for the study of the actual historical development of economies in a realistic way.

North therefore substitutes the concept of *adaptive efficiency*.<sup>111</sup> This revisits Alchian's argument, while also drawing on the ideas of Friedrich Hayek.<sup>112</sup> Adaptive efficiency, for North, is not easy to define. But the main idea is that there exists an efficiency of institutional change itself. For a society to be adaptively efficient, it seems it must be made as easy as possible to undertake entrepreneurial activities with respect to institutions, so that processes of trial and error in societal arrangements is maximized. The result of this would be that, under conditions of pervasive uncertainty, the development of "tacit knowledge" in response to changing

---

<sup>110</sup> North describes the intellectual or ideological activities of abolitionists in the United States as a situation where "individuals could express their abhorrence of slavery at relatively little cost to themselves and at the same time exact a very high price from slave owners." North 1990, p. 85-86.

<sup>111</sup> North 1990, p. 80-82. Cf. Pelikan 1987, p. 29. A more extensive discussion of the implications of this concept can be found in chapter 4.

<sup>112</sup> E.g. Hayek 1960.



circumstances is maximally encouraged. Another element is making the elimination of inferior elements of the organizational structure easier, for example through bankruptcy laws. Generally, the upshot of this concept seems to be to suggest, through an evolutionary process, the social benefits of Schumpeterian “creative destruction” and of “decentralized decision-making processes”.<sup>113</sup> I will not discuss this at length here, as these themes are developed in more extensive ways in his more recent works, as shown below.

### **The behavioural foundations of North’s NIEH**

In this section, dealing with the works from *IICEP* to *Violence and Social Orders*, I will concentrate mostly on North’s behavioural assumptions, on his conception of embeddedness (explicit and implicit), and on the development of the evolutionary-historical dimension to his theory. I will begin with the first. Recall that the purpose of NIEH theory is to explain under what conditions institutions do or do not favour the emergence of ‘Smithian cooperation’: by which I mean North’s idea, which he attributed to Adam Smith, that cooperation in society consists of the arrangement of markets and property rights such that individuals capture the gains from trade. This somewhat narrow concept of cooperation diverges from the colloquial use. It therefore should not perhaps be too easily taken to exhaust the meaning of the term for Douglass North. But this type of cooperation, at least, being the *explanandum* of the NIEH’s institutional theory makes it the most important such form of cooperation in that theory. There is no room here to go into why North should see this notion, closely related to the traditional economic conception of market efficiency discussed above, as a form of ‘cooperation’; this and the allegedly ‘Smithian’ concept of markets North refers to will be discussed further in chapter 2. I merely here highlight the prominence of this problem of cooperation since this term is one of the guiding threads in the complex conceptual web that North has woven from *IICEP* to the present day.

Observe that posing cooperation as a central problem also implies the possibility, indeed the pervasiveness, of non-cooperation. To understand this, we need to return to North’s critique of orthodox approaches and his complex relationship to the microeconomic methodology of mainstream economics. Chapter 3 of *IICEP* is dedicated to a discussion of the behavioural assumptions North attributes to the methodology of orthodox economics, ones that he wishes to reject. North observes that while the strict behavioural assumptions common to neoclassical economics are rarely thought to be realistic in the sense of accurately reflecting human behaviour in real life, they are nonetheless routinely treated by such economists as “still the best game in town for studying politics and the other social sciences.”<sup>114</sup> However, North is not so sure that this is justified, for two reasons. Firstly, as he says, “The motivation of the actors is more complicated (and their preferences less stable) than assumed in received theory.” Secondly, “the

---

<sup>113</sup> North 1990, p. 80-81.

<sup>114</sup> *Ibid.*, p. 17.

implicit [assumption] that the actors possess cognitive systems that provide true models of the worlds about which they make choices or, at the very least, that the actors receive information that leads to convergence of divergent initial models... is patently wrong.”<sup>115</sup>

Why does this matter? Because, North argues in *IICEP*, “the behavioral assumptions that economists use” – by which he means what he views as neoclassical economists – “rest fundamentally on the assumption that competitive forces will see that those who behave in a rational manner, as described above, will survive, and those who do not will fail; and that therefore in an evolutionary, competitive situation (one that employs the basic assumption of all neoclassical economics of scarcity and competition), the behavior that will be continuously observed will be that of people who have acted according to such standards.”<sup>116</sup> We have already seen an evolutionary argument of this type in Alchian’s neoinstitutionalist model of institutional change and convergence. North continues by pointing out that these neoclassical economists’ use of methods and concepts from evolutionary theory has generally dovetailed with their use of individual utility maximization, so that the properties of natural selection are thought to be operative, at least by analogy, in human social behaviour - even at the macroeconomic level - as much as in genetics.<sup>117</sup> Importantly, he observes how even apparently altruistic behavior can and has been included in such rational choice (expected utility) evolutionary models of economic and social behaviour, as for example in Gary Becker’s work.<sup>118</sup>

So where does this go wrong, according to North? Precisely the questions of motivations and mental models or ideology. For one, there is his notion we have already observed, namely that motivation comes with a cost, that following one’s own preferences (or “values and interests”) is only likely if the cost of doing so does not outweigh the benefits. This includes altruistic behaviour and norm-following generally. “The evidence we have with respect to ideologies, altruism, and self imposed standards of conduct suggests that the trade-off between wealth and these other values is a negatively sloped function. That is, where the price to individuals of being able to express their own values and interests is low, they will loom large in the choices made; but where the price one pays for expressing one's own ideology, or norms, or preferences is extremely high, they will account much less for human behavior.”<sup>119</sup> However, the content of neither ideology nor altruism is of much concern to North here.

The other problem is that of information. Citing Herbert Simon’s discussion of the consequences of the perfect information assumption, North points out that one implication is the complete convergence between the actual environment and the individual agent’s knowledge of

---

<sup>115</sup> *Ibid.*, p. 17.

<sup>116</sup> *Ibid.*, p. 19; Winter 1986 (see ref in book).

<sup>117</sup> *Ibid.*, p. 20-21; Hirshleifer 1987.

<sup>118</sup> *Ibid.*, p. 21; cf. Becker 1981.

<sup>119</sup> *Ibid.*, p. 22.

it. In the absence of such an assumption, lack of (accurate) information and limited computational capacity on the part of the agent play a major role in explaining the importance of why ideologies arise in the first place, and why they matter for the understanding of economic behaviour.<sup>120</sup> Perfect information plus known utility functions give you predictable behavior, which - together with tractability - was the justification of these assumptions in economic theory in the first place. With limited information and limited computational capacity on the part of agents, they will require external or internal guides to complement their given 'rational' ability to make choices under conditions of uncertainty. Institutions, ideologies, mental maps: such mental constructs play, North suggests, precisely that role.<sup>121</sup> This is not simply a question of optimizing institutions, but of the functional explanatory power of ideologies and norms in their own right, whether 'optimal' or not: norms of fairness can often operate to constrain individuals to choices that avoid the 'optimal' free-riding behavior that rational choice models would predict.<sup>122</sup> But in this work, North allows that he does not yet have a theory why ideology should work this way. Nor does it seem immediately clear how this can be reconciled with the view of ideologies as complementary to, rather than substitutes of, the rational choice view in which one responds so to speak 'ideologically' depending on the relative cost of doing so.

The thesis of *IICEP* could be summed up by North's statement that "the condition of the world throughout history provides overwhelming evidence of much more than simple rational noncooperative behavior. The behavioral assumptions of economists are useful for solving certain problems. They are inadequate to deal with many issues confronting social scientists and are the fundamental stumbling block preventing an understanding of the existence, formation, and evolution of institutions."<sup>123</sup> The solution is to incorporate uncertainty and limited information into the picture. This allows for the explanatory necessity of institutions (to reduce transaction costs) and of ideologies (to reduce menu costs, i.e. the costs of choice itself).

In this way, the neoclassical model can still operate in the entrepreneurial relation of individuals to institutions and to their pursuit or neglect of ideological values in practice, but this microeconomic foundation is now complemented by institutional analysis drawn from the NIE tradition and Herbert Simon's procedural, cognitive psychology approach to individual choice. Finally, a loose evolutionary framework emerges from this set of theoretical positions. It rejects the interpretation of cultural evolution as a Panglossian optimization process. Rather, it suggests what one might call an 'evolutionary marginalism', social (cultural) as well as economic:

---

<sup>120</sup> *Ibid.*, p. 23; Simon 1986.

<sup>121</sup> *Ibid.*, p. 24.

<sup>122</sup> "The consequent institutional framework, by structuring human interaction, limits the choice set of the actors."  
*Ibid.*, p. 25.

<sup>123</sup> *Ibid.*, p. 24.

incremental rational changes in the margins of institutional frameworks add up to longer-term cultural evolution (for good or ill) and path dependence.<sup>124</sup>

The latter is a concept North borrowed from Paul David and W.B. Arthur, and in *IICEP* is described as the kind of local monopolies of technology, even when inferior in the long run, that may emerge in the case of differential rates of growth or differential returns to scale. More generally, it describes in North's work the kind of local optima or equilibria that, even when less efficient than existing or future alternatives, nonetheless persist. Following Arthur's discussion, North suggests there is a self-reinforcing mechanism that occurs when for a given innovation there are returns to scale, learning effects (which lower costs as prevalence increases), coordination effects (so costs are reduced with frequency in a given context or among a given group of agents), or because of expectations of future prevalence. The consequences are also fourfold. These are the possibility of multiple equilibria, the persistence of inefficient solutions or innovations, the problem of lock-in (the difficulty of changing technologies once they have been adopted), and finally what he defines as path dependence *sec*, which really summarizes the whole of these mechanisms and their implications: "the consequence of small events and chance circumstances can determine solutions that, once they prevail, lead one to a particular path".<sup>125</sup> While this discussion is here (and in Paul David's influential QWERTY example<sup>126</sup>) primarily concerned with adoption of technology in imperfect markets, it is easy to see how North could subsequently apply it to institutional choice.

But even so, the problem of cooperation remains: "the fact that individuals acted upon those [ideological] perceptions to overcome the free-rider problem is more difficult to explain."<sup>127</sup> So for North, at this point, we find once again the overall problem framed in terms of how to explain "the imbeddedness (*sic*) of informal constraints in societies" and the variety of ways this im- or embeddedness has operated in history: "How have societies diverged? What accounts for their widely disparate performance characteristics?"<sup>128</sup>

All this as regards the relation between the behavioural assumptions discussed by North in *IICEP* and his concern with embeddedness and cooperation. The Polanyian problem of the historical differentiation of institutions and its role in generating different 'allocation systems', i.e. economies, is to be explained first and foremost at the level of individual motivations and constrained choices. The partially neoclassical, partially New Institutional behavioural

---

<sup>124</sup> *Ibid.*, p. 44-45.

<sup>125</sup> *Ibid.*, p. 93-94.

<sup>126</sup> David 1985.

<sup>127</sup> North 1990, p. 25.

<sup>128</sup> *Ibid.*, p. 6.

assumptions serve to explain why institutions and ideologies matter.<sup>129</sup> At the macro-level, they result in historical path dependence and the persistence of imperfect equilibria, subject to adjustment at the margin. However, besides the insufficient content of ideology in this model, this does not yet explain sufficiently how ideology interacts with the remaining rational choice assumptions regarding the ‘meso-level’ of individual-institution dynamics. North still has to show how ‘Smithian cooperation’ can come about despite the constraints on choice he has described. Equally, we need to understand how rational non-cooperation can turn into cooperation (‘altruism’) to enable states and institutions to come about in the first place, given he has by this point abandoned the rational choice contractarian view of such matters he held in his works with Robert Thomas.

Finally, North’s use of the concept ‘embeddedness’ (or ‘imbeddedness’, ‘imbodied’, and other synonyms), being such a core aspect of Polanyi’s historical comparative approach, cannot be read as fortuitous. Yet Polanyi gave it a definite functional explanation. For him, as we will see in the next chapter, the separation of the functions of markets, money, and trade implied market exchange was not the dominant form of economic integration in the premodern world, and therefore anthropological explanations rather than the ‘economistic fallacy’ were justified.<sup>130</sup> What kind of embeddedness relation, modernist rather than primitivist in theory, can North substitute?

### **North’s third stage: embeddedness and cooperation in *UPEC* and *Violence and Social Orders***

We must now turn to the remaining theoretical literature on these subjects in order to see how North answers these questions. I will first consider how North in his most recent works of NIEH theory has developed his thinking on the problems of cooperation and sociocultural evolution. I will therefore discuss each of the two major post-*IICEP* books in chronological order of their appearance, explaining both the new steps in the theory and the ideas about ‘Smithian cooperation’, ideology and evolution they imply.

As we have seen, over time the emphasis of the structure of NIEH theory has changed, with the more conventional exogenous constraints on growth such as technology and demography fading more into the background, and the importance of ‘social’ phenomena such as ideology, cooperation, and other mental elements affecting fundamental cognitive uncertainty being foregrounded. Where in *Structure and Change* and even in *IICEP* ideology still lacked content, it has become a centrepiece of the theory itself in *Understanding the Process of*

---

<sup>129</sup> Recall North describes his microfoundations, his theory of individual choice in *IICEP*, as explicitly complementary to the neoclassical ones: North 1990, p. 4.

<sup>130</sup> Dale 2010, p. 113, 125.

*Economic Change* (henceforth *UPEC*).<sup>131</sup> Nonetheless, there is also considerable continuity. The themes discussed above and the NIE framework in which they are formulated are developed further in scope and sophistication in North's most recent books, but the essential ambitions formulated in the Polanyi paper, and the justifications of method of the Presidential Address remain valid even so. Indeed, North practically opens *UPEC* with the observation that its theory is "an extension – a very substantial extension – of the new institutional economics".<sup>132</sup>

This extension consists of the following. Put schematically, in the previous works the emphasis was on how institutions are causes of (differentiated) economic performance (*Structure and Change*) and subsequently how individuals interact with institutions to create a particular political economic dynamic (*IICEP*). In *UPEC*, however, North goes one step further: the central question is now how institutions themselves arise. If the interaction between individuals and existing institutional frameworks constitutes what I have called the meso-level of North's NIEH theory, *UPEC* is decidedly about the micro-level. For North, the problem is to move analytically from the improved behavioural assumptions discussed above to the formation of ideology and culture, which in turn generate institutional arrangements that are inherited by subsequent generations. The whole then constitutes an evolutionary process of learning.

Although based on evolved human psychology and sociality, this evolutionary process is for North not wholly reducible to Darwinian, because of the centrality of intentionality: the beliefs of the 'players' of the 'game' matter. "The selection mechanisms in Darwinian evolutionary theory are not informed by beliefs about the eventual consequences. In contrast, human evolution is guided by the perceptions of the players; choices—decisions—are made in the light of those perceptions with the intent of producing outcomes downstream that will reduce uncertainty of the organizations—political, economic, and social—in pursuit of their goals. Economic change, therefore, is for the most part a deliberate process shaped by the perceptions of the actors about the consequences of their actions."<sup>133</sup> The larger implications are about how such learning does or does not contribute, via institutions, to economic performance: "The focus of our attention, therefore, must be on human learning—on what is learned and how it is shared among the members of a society and on the incremental process by which the beliefs and preferences change, and on the way in which they shape the performance of economies through time."<sup>134</sup>

Of course, North realizes that this presumes knowledge of what exact combination of 'genetic' and 'cultural' traits ultimately produces individual economic behaviour as well as the

---

<sup>131</sup> North 2005.

<sup>132</sup> *Ibid*, p. 10.

<sup>133</sup> *Ibid.*, p. viii.

<sup>134</sup> *Ibid.*, p. viii.

normative frameworks that generate formal and (especially) informal institutions. But he is agnostic on this question. For him, it is sufficient that natural selection should have endowed humans with a limited cognitive capacity, limited enough to make uncertainty the basis of human action, but capacious enough to make intentionality relevant. This constitutes his ‘theory of consciousness’: one that moves from uncertainty to ideology to institutions to large-scale and long-term economic change.<sup>135</sup> Beyond this, as he states, “we may never completely untangle the complex interconnections between the genetic and cultural attributes, but combining the two enables us to make sense of the human condition over time even if some of the combinations are arbitrary assertions at this point.”<sup>136</sup>

What matters is that biological evolutionary theory is not enough, but that it must be combined with an economically sound theory of intentionality – for which of course the NIE is the source. If uncertainty and sociality are to be explained in Darwinian terms as the result of our physical limitations and needs, North has no truck with doing the same for the products of intentionality. He rejects the memetic approach in favour of the more traditional economic understanding of ‘culture’, i.e. the domain of human intentional action and its products that is not explainable in terms of immediate economic interest. As he explains: “Cultural traits do not possess attributes parallel to those of genes and indeed the growing literature of the new institutional economics makes abundantly clear that institutions must be explained in terms of the intentionality of humans. Informal norms develop that blend the moral inference of genetic origin with the intentional aims of humans, which together provide the backbone of what we should mean by the term culture.”<sup>137</sup>

*UPEC* therefore provides North’s NIEH with a theory of how institutional economics can analyse culture in this sense.<sup>138</sup> The way to do so is to understand the structure of culture and its institutions by means of a game theoretical model: individuals are players and the rules of the game are given by institutions and the costs of maintaining them as external constraints on the individual strategies of these players. “All organized activity by humans”, North proclaims, “entails a structure to define the ‘way the game is played,’ whether it is a sporting activity or the working of an economy. That structure is made up of institutions—formal rules, informal norms, and their enforcement characteristics.”<sup>139</sup> This game theoretical approach to the artifactual framework can provide us with the basis for a virtually complete theory, one that can move from individual cognition under uncertainty to transaction cost economics and historical

---

<sup>135</sup> *Ibid.*, p. 38-39.

<sup>136</sup> *Ibid.*, p. 40.

<sup>137</sup> *Ibid.*, p. 42. Indeed North elsewhere refers to the “Lamarckian aspects of culture”: p. 30.

<sup>138</sup> As North says in his introduction: “What I did not consider in earlier studies was the character of *societal* change and the way humans understand and act upon that understanding of societal change.” North 2005, p. viii. Emphasis added.

<sup>139</sup> *Ibid.*, p. 48.

differentiation in economic institutions and performance. “The structure that humans create to order their political/economic environment is the basic determinant of the performance of an economy. It provides the incentives which shape the choices humans make... The strength of informal norms and the effectiveness of enforcement play a key part in the story. Where do the rules, informal norms, and for that matter the effectiveness of enforcement, come from? They are derived from the beliefs humans have.”<sup>140</sup>

In a nutshell, this works as follows. We begin with uncertainty as the foundation of human action. For North, there are five ways uncertainty interacts with belief formation: “1. Uncertainty that can be reduced by increasing information given the existing stock of knowledge. 2. Uncertainty that can be reduced by increasing the existing stock of knowledge within the existing institutional framework. 3. Uncertainty that can be reduced only by altering the institutional framework. 4. Uncertainty in the face of novel situations that entails restructuring beliefs. 5. Residual uncertainty that provides the basis for “non-rational” beliefs.”<sup>141</sup> Beliefs and the ‘stock of knowledge’ therefore interact to form belief systems which constitute structured responses to uncertainty.

‘Belief systems’ are internalisations of the environment, limited in their accuracy by cognitively induced uncertainty. These belief systems incorporate norms about social life as well as understandings of that life, so that they unite the beliefs and desires of traditional intentional theory. People impose institutions on societies based on these systems in order to achieve their desired outcomes within those societies. Informal institutions are simply outward manifestations of the norms incorporated in belief systems and are difficult to change, whereas formal institutions are the immediately political, constitutional form of such norms and are easier to change. Because both institutions and the belief systems are inherited across generations, the trajectories of societies, insofar as they are shaped by them, are subject to path dependence. This means not just the mere fact of institutional inheritance, but also the slowness and incrementality of changes in their structure. (Recall the distinction between the slow-changing ‘fundamental’ institutions and the easier to change political ones in the works with Robert Thomas.) This incrementalism about cultural evolution is to be explained by the fact that institutions give rise to organizations, which then have a stake in existing institutional frameworks and may oppose their change. (Recall here, in turn, the discussion of similar themes in *IICEP*.) Therefore, constraints operate at several levels: individuals are constrained in their beliefs by uncertainty, individuals or

---

<sup>140</sup> *Ibid.*, p. 48.

<sup>141</sup> Note that for the purposes of North's exposition, 'knowledge' is not necessarily 'justified true belief', just “accumulation of regularities and patterns in the physical and human environment that result in organised explanations of aspects of those environments.” *Ibid.*, p. 17.



groups are constrained in their choices by their beliefs and institutions, and societies are constrained by path dependence.<sup>142</sup>

This constrained path dependence of societies in turn determines their ‘economic performance’. The artifactual frameworks described above interact with the familiar givens of technological change and demographic change to produce historical change. Economic performance can only be gotten right if the institutional structures in question change along with the rest, in all three their aspects (informal, formal, and enforcement). This requires that those with the organizational power to change institutions at a given point, or the individual entrepreneurs with the necessary ability, make those changes. If not, the inheritance of previous institutional arrangements and beliefs, in other words culture, and the inheritance of existing organisational structures, in other words politics, will constitute a case of flawed social learning.<sup>143</sup> The ultimate lesson of *UPEC* is then perhaps this: “Successful economic development will occur when the belief system that has evolved has created a “favorable” artifactual structure that can confront the novel experiences that the individual and society face and resolve positively the novel dilemmas. Failures will occur when the novel experiences are so far removed from the artifactual structure of the evolved belief system that individual and society do not have the “building blocks” of the mind and artifactual structure to resolve the novel problems. If we are going to come to grips with an understanding of the differential performance of different parts of the world both over time and cross-sectionally in the modern world it is here that we must begin... That is what is meant by adaptive efficiency; creating the necessary artifactual structure is an essential goal of economic policy.”<sup>144</sup>

One of the chapters of *UPEC* in which North discusses these determinants of performance over time is called “getting it right and getting it wrong”.<sup>145</sup> North is by no means sanguine about how easy institutional adaptation is in practice: for him, much of history is the story of people ‘getting it wrong’.<sup>146</sup> Nonetheless, we will recall that in the second stage of his NIEH, North suggested the key to successful adaptation was the development of a decentralised flexibility that allowed the necessary institutional change to happen. Now, North is ready to specify this idea further in three ways. Firstly, he stresses the importance of what he calls a shift from personal to impersonal exchange. This idea plays an analogous role to the Polanyian shift from other dominant transactional modes to market exchange playing that role: it is what distinguishes the modern economy from previous allocation systems. This shift is essential to achieve cooperation, which as we have seen before has been defined as capturing the gains from trade, in other words

---

<sup>142</sup> *Ibid.*, p. 49-52. Cf. Vanberg 1994.

<sup>143</sup> *Ibid.*, p. 116-117.

<sup>144</sup> *Ibid.*, p. 69-70.

<sup>145</sup> *Ibid.*, ch. 9.

<sup>146</sup> *Ibid.*, p. 116.

private benefits in market exchange: “Personal exchange relies on reciprocity, repeat dealings, and the kind of informal norms that tend to evolve from strong reciprocity relationships. Impersonal exchange requires the development of economic and political institutions that alter the pay-offs in exchange to reward cooperative behavior.”<sup>147</sup>

This then brings us to the second specification, namely the importance of political institutions. We have already observed how in previous formulations the protection of private property rights was the *sine qua non* for economic performance in North’s theory. More generally, political institutions need to be such that impersonal exchange can be performed with a minimum of transaction costs: a staple of NIE theory. This requires order and state enforcement of private agreements and of an economically desirable institutional order alike. But this contractarianism is more Lockean than Hobbesian, insofar for North the risk of state expropriation is as real as that of insufficient enforcement. Order therefore necessarily entails “a strong but limited polity”; perhaps not coincidentally the traditional demand of liberal political theory.<sup>148</sup> The third specification is the need for a property arrangement so that the benefits of new knowledge or ideas is distributed effectively in light of the other dimensions of the institutional framework: knowledge is to be dispersed at low costs of transacting.<sup>149</sup> This last point is essentially simply an application of the transaction cost approach, and I will therefore not discuss it in detail here; except to note that this transaction cost approach dovetails well with the more conventional measure of market performance in economic history, namely by measuring levels of price volatility. On this argument, market performance consists of the ability of markets to withstand external shocks without great volatility or great differences between one region or market and another. With low transaction costs, arbitrage becomes easier and so relative prices respond to compensate for the volatility induced by the external shock.<sup>150</sup> While I do not intend to discuss the economic historical literature on the concept of economic performance here, this affinity of concepts is worth noting because for North it is important to the application of his NIEH ideas to contemporary development economics, a point I will return to below.

The final NIEH theory monograph I will discuss here is North’s most recent, his collaboration with John Wallis and Barry Weingast titled *Violence and Social Orders*.<sup>151</sup> Insofar as *UPEC* was concerned with the emergence of belief systems out of uncertainty and the impact of such cultural frameworks, when joined with bounded rationality, on institutional arrangements, *Violence and Social Orders* puts the state, politics, and formal institutions central. Nonetheless, the argument in this book as a whole is equally concerned with the origins of

---

<sup>147</sup> *Ibid.*, p. 117-118.

<sup>148</sup> *Ibid.*, ch. 8; p. 117-119.

<sup>149</sup> *Ibid.*, p. 121.

<sup>150</sup> See e.g.: Van der Spek, Van Leeuwen and Van Zanden (eds.) 2015, p. 3-5.

<sup>151</sup> North, Wallis, and Weingast 2009.

‘cooperation’ – as always, in the sense North attributed to Adam Smith – and of human sociality itself. In this approach embeddedness, as a term to describe the relations between economic behaviour and institutions and those of the ‘artifactual framework’ of a given society at large, plays an important role as well. As indicated above, I will discuss these aspects of North’s theory as it stands today, in that order: cooperation, evolution, and finally embeddedness. I will also raise some larger questions from this overview of North’s NIEH theory. But first, we have to see what this book has added to the Northian approach.

In *UPEC*, North argued for the importance of a shift from personal exchange to impersonal exchange, and for the emergence of an impersonal and limited state power to enforce the formal institutions. *Violence and Social Orders* is in part a working out of this suggestion. In order to create an origin story for the emergence of human sociality and the 'artifactual framework', *Violence and Social Orders* divides history into a progressive evolution, if not necessarily a linear one, of different types of state. While this contains various categories and sub-categories, on my reading the most important of these is the distinction between what they call 'natural state orders' and 'open access orders'. The authors describe the former as follows: “The logic of the natural state follows from how it solves the problem of violence. Elites – members of the dominant coalition – agree to respect each other’s privileges, including property rights and access to resources and activities. By limiting access to these privileges to members of the dominant coalition, elites create credible incentives to cooperate rather than fight among themselves. Because elites know that violence will reduce their own rents, they have incentives not to fight. Furthermore, each elite understands that other elites face similar incentives. In this way, the political system of a natural state manipulates the economic system to produce rents that then secure political order.”<sup>152</sup>

The open access order, by contrast, is simply the opposite of this: an order where the state has a monopoly on violence but allows free entry and exit for citizens into political organisations and the economy, and where the state takes on a wholly impersonal character, which ensures equality. The open access order is without further ado identified with the state regimes of the Western world today, whereas the natural state orders characterise all non-Western countries today as well as essentially all state orders of the past. Proceeding from the self-regard of individuals, game theoretical models are used to support the idea that repeated cooperation will be difficult to sustain given the benefits to free riding and cheating, requiring an external enforcer of the 'rules of the game' of some sort. This enforcer then is whatever elite is concocted out of temporary alliances of self-seeking individuals and turns out to be strong enough to transform into a state. The transition from this stage to the paradise of the open access order is of course the real problem in the book. The answer to this is a situation in which it becomes in the

---

<sup>152</sup> *Ibid.*, p. 18.

rational interest of the elites to treat each other impersonally, rather than violently, and subsequently to extend access to the impersonal institutions so created to a wider and wider section of the population.<sup>153</sup>

## Conclusion

But what is the foundation for this dynamic at the level of the individual and their interaction with the ‘economic system’, whether subject to predatory elites or otherwise? Even if we allow this as a plausible explanation of the political economic transition from premodern to modern allocation systems, it must be directly based on the behavioural assumptions and micro-level concepts already identified. In other words, it must have its microfoundations within the theory. North, Wallis, and Weingast therefore have recourse to economic anthropology. But they do so in a way intended to counter the ideas of the anthropologists they rely on.

North, Wallis, and Weingast go back to the Neolithic revolution and the rise of elites, which for them arise naturally out of the increasing size of human communities. Based on the work of Allen Johnson and Timothy Earle on the evolution of human societies, the authors theorise the origins of states as stemming from the ability of elite figures, when organized together, to limit a Hobbesian war of all against all. They do so in explicit contradistinction to the prevailing approaches of economic anthropology. “State is a term of art with a specific meaning in anthropology, but less so in political science and economics. States are distinguished from chiefdoms by size and structure and include formal administration of government. For anthropologists, states do not appear until populations rise into the hundreds of thousands. In contrast, what we define as the natural state arises as societies reach populations of one thousand or more, and new forms of integrated political and economic organisation develop to limit violence. As Earle recognises, “The fundamental dynamics of chiefdoms are essentially the same as those of states, and . . . the origins of states is to be understood in the emergence and development of chiefdoms”. We add the logic of natural state to the approach of Johnson and Earle: the key link that constrains military power is embedding the individuals who direct military power in a network of privileges. By manipulating privilege, interests are created that limit violence.”<sup>154</sup>

In other words, the key assumptions are the pervasiveness of violence in the ‘state of nature’ (pre-state societies), the role of the ‘natural state’ in stemming this violence, and the centrality of the creation and distribution of rents to doing so. Indeed, the whole explanation of premodern economic history (and also contemporary non-Western societies), being the history of ‘natural states’, is summarized by the authors as follows: “The dynamics of natural states are the

---

<sup>153</sup> *Ibid.*, p. 26.

<sup>154</sup> *Ibid.*, p. 53; Earle 1997; cf. Johnson and Earle 2000. Internal references omitted.

dynamics of the dominant coalition, frequently renegotiating and shifting in response to changing conditions. If adjustments lead to more power and rents based on personal identity, institutions become simpler and organizations less sophisticated, and the society moves toward the fragile end of the progression of natural states. If adjustments lead to more power based on durable agreements, institutions become more complex and organizations become more sophisticated, and societies move toward the mature end of the progression. No compelling logic moves states in either direction.”<sup>155</sup> With some rhetorical exaggeration, one can jest that for North and colleagues the history of all hitherto existing society is the history of rents; it is rents that restrain violence, while simultaneously being extracted by the constrained violence exercised by predatory elites.<sup>156</sup> The one exception are the open access orders achieved by the West, which have taken the logic of reducing transaction costs and increasing institutional adaptability to a maximum, thereby maximally achieving the gains from trade.

In the next chapter, we will now turn to North’s ‘theory of markets’. By this I do not mean a theory ever explicitly formulated in those terms, but rather a closer examination of the way North conceives of how markets function within larger social structures, how he justifies his appeal to the ‘Smithian result’, and the functionalist philosophical structure of North’s market concept. By using the recently repopularised term ‘embeddedness’, I hope to show how the contrast with Polanyi implied in the way North has taken up his ‘challenge’ helps clarify the particulars of North’s approach to market exchange in economic history. I will also discuss some critiques of North’s interpretation of social and political noncooperative coordination as market-like, and discuss the differences between North’s approach and the neoinstitutionalist view of markets of the Williamsonian tradition.

---

<sup>155</sup> *Ibid.*, p. 73.

<sup>156</sup> For a comparable perspective from public choice theory, see: Olson 1993.

## Chapter 2: North versus Polanyi on markets and embeddedness

### Introduction: A New theory of comparative economics

In the previous chapter I have set out North's NIEH theory and its historical development, identifying a number of concepts and themes that on my reading constitute the core of the theory and its defining *differentia specifica*. In particular, I have tried to show how North's approach has shifted over time in domain as well as conception: from an efficiency view of institutions to one of multiple equilibria, from an explanation of premodern societies in terms of reducing transaction costs alone to framing this by study of the origins and inheritance of institutions, from market failures as the object of inquiry to ideology, cooperation, and embeddedness. Finally I have tried to show the role of the Polanyian moment in giving North the impetus to remake New Institutionalism into a rival comparative historical theory of great scope and ambition.

In this chapter, I want to explore two aspects of North's NIEH works in more detail. Firstly, I want to deepen our understanding of the relationship between Polanyi and North in terms of the specific problems for historical and economic theory posed by Polanyi's primitivist approach. This specific conception of the relationship between economic behaviour and sociocultural institutions Polanyi proposed is encapsulated in his (late) concept of 'embeddedness'. It is interesting that this same concept not only appears in a modified form in North's 1977 paper, but subsequently re-emerges in the work of Mark Granovetter, whose approach to economic sociology has much in common with the ideas of the NIE (or at least its Northian branch).<sup>157</sup> In order therefore to clarify what precisely Polanyi's challenge might consist of, a brief exploration of the interconnectedness between Polanyi's theoretical framework for studying economic history and the purpose of the NIE as a special kind of 'substantivism' in its own right may be useful. In particular, I want to focus on how Polanyi's challenge, as North took it up, had been shaped by his idiosyncratic view of price-making markets and the role of economic theory (supply and demand) in understanding their functioning.

This leads me subsequently to discuss North's own, rival approach to understanding markets. Firstly, this means examining more closely the complex analytical relationship between individual self-interest, transaction costs, and market efficiency in his theory. Secondly, we must clarify what North means by what he calls the 'Smithian result'. This term, and variant references to Smith in the context of perfect market competition, plays an important normative role in North's theory. We must therefore examine this (dubious) apparent attribution to Adam Smith of the insight that once an institutional order permits well-functioning markets, the logic of neoclassical economics does obtain, and economic growth as well as adaptive efficiency are furthered thereby.

---

<sup>157</sup> E.g. Granovetter 1985; 2005.

Necessarily, such an examination also requires looking at the ideological context of this idea in North's work, where I want to point to the importance of social contract theory in his presentation of the 'choice' between markets and nonmarket allocative systems, the question raised in his 1977 paper. The political-ideological background to this, in classical or ordoliberal theories of the market inspired by Friedrich Hayek, must also be observed, although I will say more about this in chapters 3 and 4. Finally, I will discuss some of the previous critiques of North's theory on precisely this point, the understanding of the role of the market and its institutional structure when compared with institutional orders not primarily based in market exchange: a point important for both Polanyi and North. Since North's response to Polanyi begins with this discussion about markets and other allocation systems, and since much of North's theorizing of the historical function of market exchange rests on more implicit than explicit assumptions, this examination will lay the foundation for subsequent discussions of North's rationality concept, his use of evolutionary theory, and his relationship to economic anthropology.

As Sally Humphreys has pointed out, the implications of Polanyi's project are that if orthodox economic theory could not simply be applied to premodern economic formations, one would need a "new theory of comparative economics".<sup>158</sup> North himself clearly interpreted Polanyi's challenge as constituting one such theory. My argument is that his own NIEH approach should be seen in this light as another, namely one intended to demonstrate Polanyi's primitivist premises are not required for such a theory to succeed. The present chapter therefore stands on two legs. One is getting a better grip on North's attempt to answer Polanyi's challenge from this "economics approach" by examining what Polanyi's approach to markets and their embeddedness in sociocultural and political institutional structures actually entailed. The other is to clarify North's subsequent development of a different kind of substantivism about markets, first and foremost based on transaction cost economics but entailing a much wider range of theoretical positions. Some of these positions have striking areas of overlap with Polanyi's, but in other respects North's approach to markets is much more indebted, I will argue, to a contractarian liberal tradition about the virtues of markets as engines of growth and coordination.

Finally, I will also make some critical comments on North's approach to markets and discuss some alternative approaches to their role in (premodern) economic history – in particular with an eye to Polanyi's own use of anthropological theory to study this subject. If I therefore violate disciplinary strictures against concentrating on philosophical and theoretical problems over practical empirical studies or new models, so be it. If this work suffers from what Clifford Geertz called "epistemological hypochondria", I justify it as a prerequisite for improving explanation in economic history and by the impact of the NIEH on wider debates in the social sciences.<sup>159</sup> Even so, the literature on markets and market theory is far too vast to allow even an approximation of an exhaustive study of these topics; the present discussion will have to serve

---

<sup>158</sup> Humphreys 1969, p. 166.

<sup>159</sup> Geertz 1988, p. 71.

merely to highlight a number of important aspects of North's approach to markets, and to consider their strengths and weaknesses compared to Polanyi's.

### **Markets, institutions and embeddedness in Polanyi**

With this in mind, let me now return to the themes to be discussed below. In this chapter, I will discuss the afterlife of embeddedness, in the Polanyian sense, in the work of North and his engagement with the reappearance of this concept in New Institutionalism-inspired economic sociology. This also allows us to situate North's work more clearly in the context of his relation to Polanyi. Among Polanyi's theses on the embeddedness of economic behavior in social and cultural norms in premodern societies was the contention that markets, money, and trade were separate social functions that did not operate along the same lines in such societies as they do when generalized market exchange is the dominant 'transactional mode'.<sup>160</sup> His claims regarding ancient markets, that they were fundamentally not determined by 'supply and demand' but embedded in social and political structures supporting other dominant transactional modes, is one of the most controversial aspects of his primitivist economic comparative theory. For North, as we have seen, the universality of the 'economics approach' is central, and for this reason he cannot consistently accept such a primitivist account. Yet we have also seen that North has been readily willing to concede the limited nature of market exchange in premodern societies.

Because of this, after the brief discussion of the afterlife of Polanyian embeddedness in North I will consider how North perceives the nature of markets and market exchange in such societies. I will also discuss his theories of 'political markets' and their role in extending the analysis precisely to such institutional embeddedness as Polanyi was concerned with. This point, North's description of markets or indeed of 'the economic' as equivalent to market-like phenomena, has been most often the target of his detractors' critiques. But embeddedness is a notoriously vague term, one which has been used in different ways by different authors and even within the work of one, for example that of Polanyi, can give rise to many ambiguities of meaning.

For at least one eminent Polanyi specialist, there is no doubt that "the logical starting point for explaining Polanyi's thinking is his concept of embeddedness. Perhaps his most famous contribution to social thought, this concept has also been a source of enormous confusion."<sup>161</sup> This confusion is at least in part the fault of Polanyi's own inconsistent use of the term. As Kurtulus Gemici has pointed out, for Polanyi the term 'embeddedness' came to symbolize two separate theoretical viewpoints. On the one hand, it entails Polanyi's argument for the impossibility of separating a sphere of economic logic from the larger social-institutional context in which this form of 'the economic' (however defined) operates. On the other hand, Polanyi uses it to indicate the difference between the functioning of markets in premodern societies

---

<sup>160</sup> See e.g. Gudeman 2008, p. 95.

<sup>161</sup> Block 2001, p. xxiii.



(embedded) and in modern societies (disembedded). There is therefore a contradiction in his use: in some sense all economies are embedded in institutions, whereas in another sense the transactional mode in which market exchange dominates is a case of disembedding from institutions.<sup>162</sup>

As Gemici suggests, this is partially to be explained by Polanyi's shifting interests. In *The Great Transformation* (TGT), Polanyi uses it only twice, but in his later works (which are more relevant for the present study) the term gains in significance.<sup>163</sup> This can be explained by Polanyi's increasing concern with differentiating in what historical economies the independent logic of the market, represented for him by the thinking of mainstream economics, operates from those historical economies in which this logic does not operate and therefore such thinking is inapplicable.<sup>164</sup> In *TGT*, the main thrust of his argument<sup>165</sup> is to explain the emergence of modern market society and its negative effects, indeed its unsustainability in every sense; in his later economic historical/anthropological works, his argument is at least as much about the applicability of economic theories as about the history of markets and trade.

Standard economic theory is fine for the period covered by *TGT*, even if Polanyi disapproves of that type of society. But the division between modern and premodern is established by the inapplicability of such theory to the societies of the past that had different dominant transactional modes. As Polanyi put it in a 1947 essay: "The market mechanism, moreover, created the delusion of economic determinism as a general law for all human society. Under a market-economy, of course, *this law holds good*. Indeed, the working of the economic system here not only "influences" the rest of society, but determines it... In terms of the present article, instead of the economic system being embedded in social relationships, these relationships were now embedded in the economic system."<sup>165</sup> In premodern economies the embeddedness relationship was the reverse. That is to say, economic behavior was not a question of determination by the operation of the market, but by different *motives*: "The individual's motives, named and articulated, spring as a rule from situations set by facts of a non-economic—familial, political or religious—order".<sup>166</sup> As Hannes Lacher puts it: "For political economy and neo-classical economics, the profit motive was a given, deduced from human nature or the structure of competition. For Polanyi, by contrast, the question is how such a specifically economic interest came into existence."<sup>167</sup>

Polanyi's idea of embeddedness, now taken as historical differentiation in the sense described above, i.e. where the motives that guide the economic actions of individuals are dependent on the institutional context in which they operate, rests on two foundations. The first

---

<sup>162</sup> Gemici 2008, p. 3-4.

<sup>163</sup> Polanyi 2001.

<sup>164</sup> *Ibid.*, p. 12-14.

<sup>165</sup> Polanyi 1968, p. 70; Cited in Gemici 2008, p. 12-13. Emphasis added.

<sup>166</sup> Polanyi 1968, p. 85.

<sup>167</sup> Lacher 1999, p. 316.

is the motivational one. As we have seen, embeddedness meant for Polanyi rejecting the idea that certain motives, or logics of choice, operated throughout human history and defined a separate sphere of the economic. The ‘substantive’ view of the economy, for Polanyi, is after all simply defined as the sphere of the production and distribution of the necessities of life, of the ‘livelihood of man’ – also the title of a posthumous collection of his essays.<sup>168</sup> This distinguishes itself consciously from the motivational universalism of the standard economic approach, at least in its choice theory, which is considered to apply whenever a choice is made between scarce means for competing ends – in practice usually mediated by relative prices in some form.<sup>169</sup>

Here, I want to concentrate on the second point: the institutional context. Insofar as embeddedness has served, as Gareth Dale says, “as a doorway for explorations of the relationship between economic behavior and the social integument”, it is clear that if Polanyi and North differ strongly on how to understand the former, they share an interest in the latter, in the form of the institutional framework.<sup>170</sup> To sum up, Polanyi’s embeddedness thesis appears as a challenge for the ‘economic approach’ insofar as it rejects its universality; but insofar as concerned with historicizing the institutional context of ‘markets and other allocation systems’, it appears as rather complementary to North’s ambitions. Both approaches do have something in common: taking individuals as the primary agents of analysis. Postponing the methodological discussion of rationality inherent in this aspect to the next chapter, I will now discuss how North’s has taken up the historical differentiation aspect of the embeddedness thesis and to what extent it has succeeded in improving on Polanyi on this point.

### **Is there embeddedness in North?**

In Polanyi we find that the term ‘embeddedness’ is introduced early on (in 1934)<sup>171</sup>, more or less *en passant*, then plays only a minor role in the theoretical analysis (in *TGT*), only to subsequently blossom into a much more significant concept for the whole. Curiously, something similar can be said for the use of this term, or virtually synonymous terms, in Douglass North’s writings. I observed previously that in discussing Polanyi’s challenge, North notes that “the key to Polanyi’s system is the view that economic organization is *embodied* in society in the sense of ‘having no separate existence apart from its controlling social integument’. Transactional dispositions... are expressions of socially defined obligation and relationships.”<sup>172</sup> Although North here writes ‘embodied’, we may take this as his definition of Polanyi’s embeddedness concept, as it is difficult to imagine what else it might refer to. But in subsequent works, this term or synonyms do not appear much at all, just as with Polanyi’s first monograph: in *IICEP* for example there is only one relevant use of the word ‘embedded’, namely where North states that “the incentives embedded in the institutional framework direct the process of learning by doing” – not even a

---

<sup>168</sup> Polanyi 1977. See also North’s critical review of this work: North 1978b.

<sup>169</sup> The classic formulation of Robbins 1932.

<sup>170</sup> Dale 2010, p. 195.

<sup>171</sup> Dale 2011.

<sup>172</sup> North 1977, p. 708. Emphasis added.

context very close to Polanyi's concerns.<sup>173</sup> In *UPEC*, embedded and embeddedness appear three times, and in contexts closer to Polanyi's use; and in *Violence and Social Orders*, it suddenly gains a great prominence, being mentioned (by my count) no fewer than eighteen times in Polanyian formulations.<sup>174</sup> Polanyi's own name does not reappear in these monographs (only his brother does), so the increasing emphasis on embeddedness is all the more interesting.

Polanyi may be present in North's concept indirectly as well. One important 'afterlife' of embeddedness has been in economic sociology, where Mark Granovetter rediscovered the concept in a 1985 paper.<sup>175</sup> In this paper, Granovetter wanted to distance himself from what he called the "undersocialized image of man" in neoclassical economic theory as well as the "oversocialized image of man" in Polanyi's work.<sup>176</sup> Associating the latter with 'embeddedness' *tout court*, he suggested his own approach, an NIE-inspired concept of 'network embeddedness', as a middle ground between the two. As Granovetter argued, "much of the utilitarian tradition, including classical and neoclassical economics, assumes rational, self-interested behaviour affected minimally by social relations... At the other extreme lies what I call the argument of 'embeddedness': the argument that the behaviour and institutions to be analysed are so constrained by ongoing social relations that to construe them as independent is a grievous misunderstanding."<sup>177</sup>

Granovetter has since repudiated the association of embeddedness with Polanyi, claiming he did not have Polanyi in mind and did not mean to reintroduce it.<sup>178</sup> But there is at least an analogy between Granovetter's view of network embeddedness and North's ambiguous position between neoinstitutionalism and Polanyi's challenge.<sup>179</sup> This makes it suggestive that precisely this term, so associated with Polanyi (even if not original to him), should have appeared inescapable for describing the effort to reconcile the 'economic approach' to economic behaviour with how it was 'constrained by ongoing social relations'. So much so, that in this paper sometimes Granovetter slips into using embeddedness for his own position. It is therefore imaginable that for all Granovetter's subsequent protestations, North found this Polanyian concept amenable for his own middle ground approach, perhaps even drawing on Granovetter's initial formulation. But of course, none of this constitutes more than circumstantial evidence pointing at the afterlife of Polanyi's embeddedness problematic in North's work.

At the very least we can say this: there is a late return to the embeddedness concept in North as there was in Polanyi's own work, a return – I would suggest - prompted by *Violence*

---

<sup>173</sup> North 1990, p. 81.

<sup>174</sup> I have not counted other uses of 'embedded', such as to simply indicate 'part of' or 'found in', which North frequently uses when referring to legal provisions as part of larger formal constitutional systems.

<sup>175</sup> Granovetter 1985.

<sup>176</sup> *Ibid.*, p. 481-93.

<sup>177</sup> *Ibid.*, p. 481-482.

<sup>178</sup> Granovetter in Krippner et al. 2004, p. 113.

<sup>179</sup> For a critique of Granovetter's ambiguities as analogous to those of New Institutionalism, see: Beckert 2009.

*and Social Orders* as the most complete formulation of a general comparative economic history comparable to Polanyi's own. For this, North clearly needed some concept to describe the relation of the individual, and his economic action, to the social whole and its normative structure: the *faits sociaux totaux* so important to Polanyi as well.<sup>180</sup> Characteristic are formulations like "the set of social interactions, organizations, and networks in which individuals are embedded" and "rules and norms, by themselves, are not self-sustaining; they must be embedded in a larger structure of organizations and beliefs".<sup>181</sup> North's use of the term is clearly somewhat broader than Polanyi's, but gets at the same problem: how to relate individual economic behaviour - including within and towards institutions, organisations, and political life - to the institutional structure as a whole and its constraining effect on such behaviour. Which is just to say: Polanyi and North are both institutionalists in their own way, and the reappearance of embeddedness in *Violence and Social Orders* to describe the nature of this institutionalism is one piece of evidence for the convergence of their institutionalisms at the level of explanatory theory, if not on the salience of economic microfoundations.<sup>182</sup>

The point of this discussion is not primarily philological, but rather to allow me to use the concept of embeddedness to discuss the points of overlap and contrast between North's and Polanyi's views of what the economy is, and in particular what markets do within them; and to relate these to the problem of the absence of generalized market exchange in premodern societies and how the individual relates to the social-institutional context in those societies. From his 'Polanyian moment' to *Violence and Social Orders*, one cannot escape the sense that he was consciously or implicitly setting up his own view of this complex problematic as a Polanyian one, even if his answers to this problem converged with Polanyi's view only in part. We have seen they converge in a very general sense, namely insofar as both accepted normative institutional structures as the fundamental constraints on economic behaviour in the premodern world (if in Polanyi's case not in the modern). We have also seen that in the 1977 paper North conceded Polanyi's assertion that the premodern world was largely dominated by transactional modes other than market exchange. So what of markets then? Do they at least agree on what those are, and how they function? If not, can North consistently formulate a theory of embeddedness as a response to Polanyi's challenge that nonetheless has a very different view of markets than Polanyi had?

After all, for Polanyi the different function and determinants of markets and the separation between markets, money, and trade was at least as important as economic methodology for his primitivist case. For some critics, North's interpretation of markets and their historical function is fundamentally indebted to economic perspectives irreconcilable with a Polanyian embeddedness project. Insofar as this is true, Polanyi's inconsistent embeddedness theory would

---

<sup>180</sup> Humphreys 1969, p. 166.

<sup>181</sup> North, Wallis, and Weingast 2009, p. 28, 260.

<sup>182</sup> A similar argument is made by Didry and Vincensini 2011.

be matched by an equally inconsistent embeddedness theory on North's part, at least as far as the theory of the market is concerned.

### North's theory of markets

In order to get a sense of what North's theory of markets and their historical function is, in light of the above discussion, I begin once more at the beginning. Inevitably this means first another close reading of the 1977 paper on Polanyi, but now with a different focus than before. Directly following on his discussion of Polanyi's embeddedness viewpoint in the 1977 paper, North points out (correctly) that one implication of Polanyi's argument is that exchange and trade must not necessarily imply "economic motivation" – by which he appears to mean a profit motive. He points to Polanyi and George Dalton's famous arguments regarding administered trade in ancient Babylonia and the *kula* trade of the Trobriand Islanders – a set piece of economic anthropology taken from Bronislaw Malinowski – as illustrations of this argument.<sup>183</sup> As we have seen, far from reacting as many of the strict 'modernists' might, or as one might expect from an orthodox economic historian, and simply denying the truth of this Polanyian claim, North goes out of his way to accept this premise. In fact, for him this is even true in the post-Industrial Revolution economies of the modern world, so that he falls clearly on the 'always embedded' side of interpreting that Polanyian concept. In a sense familiar from Coase, Williamson, and the wider NIE tradition, North points out how firms and governments also undertake economic activity in modern market societies, so that even there price-making markets do not fully dominate all transactions.<sup>184</sup>

For North, the challenge is to explain the fact that "substitutes for markets have dominated exchange in past societies" (and to a lesser extent to explain these today as well), and to do so *without* accepting Polanyi's assertion regarding economic motives. As we have seen, for North in the 1977 paper we have to see how far "economic analysis" will take us in explaining these nonmarket transactional modes, and economic analysis implies rational behaviour "in the economic sense of the word".<sup>185</sup> This much is familiar by now: this is his argument for the advantages of transaction cost economics as a theoretical framework for doing so.

What we have not yet investigated, however, is what understanding of the market underlies this argument regarding market and nonmarket transactions. Insofar as North has accepted Polanyi's premise, and based the NIEH on this challenge, he has to accept a particular definition of markets that is consistent with Polanyi's argument about the separation of markets, money, and trade at both institutional and motivational levels. This point did not escape North, and for this reason his Polanyi paper also includes a discussion of how to define markets for historical comparative purposes. For Polanyi this was the central heuristic, but in the orthodox economic literature this is not a familiar starting point of inquiry. As North observes: "it is a peculiar fact

---

<sup>183</sup> North 1977, p. 708.

<sup>184</sup> *Ibid.*, p. 709.

<sup>185</sup> *Ibid.*, p. 709, 714.

that the literature of economics and economic history contains so little discussion of the central institution that underlies neoclassical economics: the market.”<sup>186</sup>

It is clear to North that Polanyi’s definition of markets is a very specific one, namely the *price-making market*. The absence of this type of market exchange, with its independent endogenous logic of supply and demand, in premodern economies is the cornerstone of the latter’s analysis of economic historical differentiation.<sup>187</sup> Polanyi importantly distinguished markets *per se* from what he called ‘price-making markets’, one in which supply and demand act as independent forces to determine the relative prices of goods and which is integrated, or tends to integrate, into a larger system of market exchange.<sup>188</sup> As North points out, this seems to apply to markets with “a large number of buyers and sellers, a variety of goods, an agreed-upon medium of exchange and an enforced set of property rights”.<sup>189</sup> For North, as one might expect, it is the latter that is decisive: the property rights arrangements (and their enforcement) determine transaction costs (information costs, measuring costs, and so forth) and thereby the viability of such markets compared to other forms of exchange.<sup>190</sup> The salient point is that “the costs of defining and enforcing property rights... lead to non-price allocation of many goods and services today... transactions costs in the ancient past would have been an insuperable barrier to price-making markets throughout most of history.”<sup>191</sup>

As mentioned, for the neoinstitutionalist perspective on ‘markets versus hierarchies’, such as that of Oliver Williamson, the choice between market contracting and hierarchical non-price structures is between the twin poles in an efficiency tradeoff. North here appears to extend this concept of markets and other ‘allocation systems’ (in North’s jargon) as a choice within a tradeoff structure to Polanyi’s distinction between price-making markets and other ‘transactional modes’ (in Polanyi’s). For North, the NIE version of the economics approach can be extended to viewing this type of ‘choice’ for one or another of these transactional modes as a response to transaction costs. So what determines the ‘choice’ of institutional arrangement? Here North draws directly on the neoinstitutional literature. While pointing out that the literature differs on the precise nature of the firm (a hierarchical structure, a legal fiction, a joint production arrangement), North agrees with all that it is a “wealth-maximizing institution which substitutes for price-making markets”. And so, North says, “*it is reasonable to assume that the forces that*

---

<sup>186</sup> *Ibid.*, p. 710.

<sup>187</sup> Polanyi’s argument has the caveat that the period of Hellenistic Greece to the heyday of the Roman Empire were, for him, characterized by such markets. Polanyi therefore, quite consistently, considered those societies as of one piece with modern ‘capitalism’. Polanyi 1977, p. 273. See also the discussion in chapter 5.

<sup>188</sup> Polanyi 1957, p. 255; Stanfield 1999, p. 214-215.

<sup>189</sup> North 1977, p. 710.

<sup>190</sup> *Ibid.*, p. 709-710.

<sup>191</sup> *Ibid.*, p. 710.

*lead to the substitution of firms for markets today may also help us to explain the variety of forms of economic organization in past societies.*"<sup>192</sup>

In other words, the neoinstitutionalist choice theoretical approach can, by extension, explain the past 'choice' for nonmarket forms of allocation. If firms substitute for markets in order to maximize wealth, given the constraints necessitating a tradeoff between them, North seems to argue, why should not other allocation systems be just like firms and also choices within a wealth-maximizing tradeoff? It is reasonable to assume the forces of competition, motivation, or whatever may be involved are no different in past societies as in the present, insofar this type of institutional choice is concerned. Firms or other nonmarket allocation systems act in this framework as substitutes for price-making markets (leading to the 'nonprice allocation' referred to), which therefore serve as a benchmark. Even insofar as North concedes Polanyi's empirical claim regarding price-making markets in premodern history, i.e. even in their absence, a wealth-maximizing motive is still assumed *in the 'choice' between allocation systems*. This view of the market-like functioning of allocation systems and their institutional embeddedness has serious consequences for North's subsequent NIEH research programme.

It is by this perspective that in the 1977 paper the *kula* trade becomes a "least-cost trading solution where no system of enforcing the terms of exchange between trading units exists", i.e. in the absence of the contractarian state. Similarly, the 'ports of trade' Polanyi described as fundamental to administered trade in the classical and medieval worlds are re-described as risk-reduction solutions underpinned by norms (informal institutions). Since prices did change from time to time (as Polanyi admitted), price making markets might even originate in such phenomena.<sup>193</sup> In a similar way, ideology is measured in a later paper as the costs people are willing to incur rather than free ride off public goods (such as informal institutions), and the second category specifies "exchange in which kinship ties, friendship, and personal loyalty all play a part in constraining the behaviour of participants" in order to reduce transaction costs.<sup>194</sup> Precisely those aspects which for Polanyi's anthropologically derived approach are the fundamental differences in motivation and meaning attached to economic behaviour in general here appear as special cases of a more general rule concerning exchange throughout history, without differentiation as to what kind of 'markets' are involved in this exchange. Although rejecting neoclassical Walrasian models as well as neoinstitutional efficiency approaches, North's view is – perhaps *malgré lui* – at this more fundamental level not so far removed from the standard economic understanding of what 'exchange' is and how it is reducible to a gains-oriented conception of individual action. (Indeed, North here also shares with standard economic theory the view that production is a special case of exchange, quite contrary to for example the

---

<sup>192</sup> *Ibid.*, p. 711. Emphasis added. Note that by North's later definition, a firm would not be an 'institution' in the strict sense.

<sup>193</sup> *Ibid.*, p. 713-714.

<sup>194</sup> North 1984, p. 257-258.

traditions of Marx, much economic anthropology, and indeed the approaches of classical and Old Institutional Economics.)

A similar case can be made for North's conception of 'political markets'. This is a term he frequently favours to describe the structural characteristics of political institutions, organisations and processes independent of their specific historical content. Here, too, North is quite clear that he rejects modelling such processes as if the results must be some form of efficiency, whether defined in Pareto terms or in terms of favourability to economic growth. Indeed, North doubts rather understandably that we would know what a Pareto efficient governmental structure would mean<sup>195</sup>; and as regards economic growth as a measure of efficiency, he questions empirically whether in economic history political structures have often promoted it.

The latter point is worth elaborating on, because in the process of commenting on this, North on certain occasions also clarifies what he thinks that type of efficiency in the 'political market' would look like (if we did have it): "Just as the efficiency of an economic market can be measured by the degree to which the competitive structure, through arbitrage and efficient information feedback, mimics or approximates the conditions of a zero-transaction-cost framework, so an efficient political market would be one in which constituents accurately evaluate the policies pursued by competing candidates in terms of the net effect on their well-being; in which only legislation (or regulation) that maximized the aggregate income of the affected parties to the exchange would be enacted; and in which compensation to those adversely affected would ensure that no party was injured by the action. To achieve such results, constituents and legislators would need to possess true models that allowed them to accurately evaluate the gains and losses of alternative policies; legislators would vote their constituents' interests-that is, the vote of each legislator would be weighted by the net gains or losses of the constituents, and losers would be compensated so as to make the exchange worthwhile to them - all at a transaction cost that still resulted in the highest net aggregate gain."<sup>196</sup>

North of course rejects any implication that this is actually what happens in such 'political markets' in most countries today, let alone in premodern history. This is his argument against the neoclassical approach as well as the public choice school: they need to take the pervasiveness of transaction costs seriously, and allow for ideology in the sense that people may have different (more or less true) individual models of how the world and the political process operates. Both act as constraints on their action within and towards a political structure. Also, as an economic historian North never fails to point to the path dependent nature of such an institutional "matrix".<sup>197</sup> But as with the notion of efficient market exchange in the economic sphere, compared to substituting institutions, here too the notion of the efficient political market described above functions as a benchmark. In fact, occasionally it is even more explicit: for as

---

<sup>195</sup> E.g. *Ibid.*, p. 256.

<sup>196</sup> North 1993a, p. 66-67. Cf. North 2005, p. 53-54.

<sup>197</sup> *Ibid.*, p. 64.



North states on one occasion, “the task as I see it is to structure the institutional framework so as to approximate this ideal model.”<sup>198</sup>

We can find ample proof of an equivalent conception in all the stages of North’s NIEH project, even up to the most recent theory. *UPEC* virtually opens with the observation that “various kinds of markets (political as well as economic) have different margins at which competition can be played out, the consequence of the structure we impose will be to determine whether the competitive structure induces increasing economic efficiency or stagnation”.<sup>199</sup> This particular formulation also recalls North’s argument for the incremental nature of institutional change as pursued by individual ‘political entrepreneurs’ acting in their own interests: they operate ‘on some margin’, implying shifts in the relative (shadow) ‘prices’ in a political or institutional ‘market’. “The outcome is a mixture of both economic and political decisions that in the aggregate affect the performance in individual political and economic markets as well as determine the direction of the economy as a whole.”<sup>200</sup>

Note that this market benchmark, so to speak, determining the way North speaks about political and economic behaviour and processes operates for him independently of motive: “It is something else again exactly what the intentions of the players are. The overall direction of economic change will reflect the aggregate of choices made by political and economic entrepreneurs with widely diverse objectives, most of them not concerned with the consequences for overall performance.”<sup>201</sup> These ‘entrepreneurs’ need not have economic performance at the macroeconomic level in mind. Rather more likely, North seems to suggest, their aim is “lining the pockets of the players”.<sup>202</sup> But insofar as these actions – constrained by the usual mental bounds - mirror the supply and demand (price and quantity) mechanisms of the market, they will have efficient consequences: “the Smithian result ensues”.<sup>203</sup> If not, the implication goes, the institutional structure and/or property rights arrangement is inefficient and should be (or have been) altered.

Political structures may often be inefficient and path dependent, individuals acting within them ideologically and institutionally constrained – even so, this does not make politics any different than any other market with high levels of transaction costs, and does not make the ideal efficient market, geared towards economic growth, any less relevant as a model for understanding *and* evaluating political formations. This point is essential to North’s market-like conception of institutional and political ‘matrices’ or ‘artifactual frameworks’: the same criteria

---

<sup>198</sup> North 1992, p. 16. The ‘task’ in this case is one for developmental policy makers, as this particular paper was written for the occasional papers series of the International Center for Economic Growth.

<sup>199</sup> North 2005, p. 1.

<sup>200</sup> *Ibid.*, p. 80.

<sup>201</sup> *Ibid.*, p. 79.

<sup>202</sup> *Ibid.*, p. 79.

<sup>203</sup> *Ibid.*, p. 80.

to evaluate them in terms of efficiency are the criteria used for understanding their function and historical ‘meaning’ for the purposes of explanation.

### **The ‘Smithian result’ and the social contract**

As we have seen, in the more developed stages of his NIEH theory, North recurrently makes use of the concept of the ‘Smithian result’ (or equivalent phrases). This term serves him as shorthand, it seems, to describe the efficiency and coordination results one can expect from well-functioning – that is to say, institutionally well-supported – modern markets, of the kind whose internal functioning (‘laws of supply and demand’) is according to both North and Polanyi best described by standard neoclassical models. This concept requires further examination if we are to understand North’s market theory. What is this result, and what does Adam Smith have to do with it?

In the conclusion of section 1 of *UPEC*, we are told that “when economic markets are so structured that the players compete via price and quality rather than at non-productive margins then the Smithian result ensues.”<sup>204</sup> This seems to capture a rather more neoinstitutional insight, one that has affinities with the public choice literature: namely the trade-off not so much between markets and hierarchies, as between rent-seeking through non-market means and the pursuit of marginal advantages within the market. In the most classical models, such advantages come about through technological innovation improving productivity and thereby giving the innovator (temporarily) a competitive edge, permitting the obtaining of super-normal returns. But in the same work we are told more about the ‘Smithian result’ than this rather conventional economic observation. For Smith’s result, North makes clear, is a special case implied by neoclassical theory. For this reason, in the more general case – described by his own NIE-based approach – the result does not obtain, although at all times, as mentioned above, for North it would be ideal if it *did* obtain.

North makes this explicit when a little later in *UPEC* he points to a number of problems with the neoclassical model that reduces it, and the Smithian result with it, to this special case status. There are four such problems: the “movement from personal to impersonal exchange” (a redescription of Polanyian embeddedness and disembedding, one is tempted to think) is necessary to obtain the Smithian result. This requires “fundamental rethinking at odds with our genetic heritage”, namely by shifting from small group cooperation to large group noncooperative coordination (more on this in the following chapters). What one needs is therefore for markets to become general and large in scale, so that the necessary ‘impersonality’ is the result. Secondly, “Adam Smith’s specialization and division of labor— the necessary condition for achieving such markets—is really specialization of knowledge”. This, North argues, cannot be done by a “price system” alone, and requires institutional support to coordinate and achieve public goods. Thirdly, and perhaps most importantly, there is a direct relationship

---

<sup>204</sup> North 2005, p. 79-80.

between market competition and the ‘Smithian result’. And this is not easy: for “all well-functioning factor and product markets must be structured to provide incentives for the players to compete at those margins, and those margins alone, that induce growing productivity. Only then do we realize Smith’s beneficent result. Moreover, in a dynamic world with changing technology, information costs, and politics there is nothing automatic about the structure changing in response to these changing parameters to continue to produce efficient markets.” Fourthly, achieving all this requires strong political support, perhaps even initiative, and this in turn demands political institutions strong enough to generate and sustain public goods, but weak enough not to interfere with market efficiency or individual property rights.<sup>205</sup>

What we are told about the ‘Smithian’ view of markets is then this: that with his name is associated a “beneficent result”, the nature of which is not stated, if and only if markets are structured in a competitive way such that they foster productivity increases. Such productivity increases are meant here presumably as opposed to rent-seeking, as indicated in the previous mention of Smith; the latter forms one of the main subjects of his subsequent book with John Wallis and Barry Weingast, *Violence and Social Orders*. But this beneficent result is not just a matter of channelling competition through markets, but also of having as general a market structure as possible: only by politically and institutionally providing for scale and scope in market exchange does one achieve “the potential envisioned by Adam Smith when he viewed the wealth of nations as being a function of the size of markets”.<sup>206</sup> Finally, towards the end of the work we are told that one additional problem for achieving Smith’s result is the Hayekian problem of ignorance: people do not sufficiently realize the correctness of this Smithian analysis of markets, and this prevents a structuring of political and economic institutions favourable to it. As North suggests, “It is more than two hundred years since Adam Smith explained the underlying sources of the wealth of nations but the extent to which such views are embedded in the decision-making process of those shaping political/economic change is problematic... Adam Smith’s wealth of nations was a function of specialization and division of labor. But the logic of specialization and division of labor implies a world in which individuals know a great deal about their specialty but in consequence know less about the rest of their world. Hayek emphasized the crucial additional point that in consequence individuals can have only a very imperfect understanding of the overall character of the political/economic system.”<sup>207</sup>

Although on other occasions North does refer more explicitly to Adam Smith’s actual writings on the 18<sup>th</sup> century commercial economy<sup>208</sup>, the strong impression here generated is that Smith is not really invoked as an economic thinker, but rather as a shorthand for a particular model of the capitalist market. Smith stands for the belief that competitive and well-structured markets are ultimately both beneficial in economic terms and have the capacity to coordinate

---

<sup>205</sup> All citations in this paragraph from North 2005, p. 84-85.

<sup>206</sup> *Ibid.*, p. 84.

<sup>207</sup> *Ibid.*, p. 156, 162.

<sup>208</sup> E.g. North, Wallis, and Weingast 2009, p. 205.

individual actions and desires in a society large enough that most interactions are impersonal and noncooperative. This latter aspect, the need for coordination under noncooperative conditions is also, on my reading, the context of North's many references to game theory and the 'players' of the 'game', something I will discuss in more detail in chapter 3. Significant is here to note that North simply uses Smith to defend an assumption about the overall desirability of competitive markets, but to also point to the difficulties of achieving and sustaining such markets *at the level of institutional and political support*. Using Smith to make the neoclassical perfect competition argument a special case of a more general – and considerably less optimistic – framework allows North to take the desirability and efficiency of classical competitive markets for granted, and to shift the subject of contention to a deeper level 'beneath' the market, namely the 'artifactual structure'.

One may wonder how much this view really has to do with Adam Smith's own work; indeed North seems singularly unconcerned with this problem, despite the extensive literature on the complexities of Smith's writings and the degree to which the view of Smith as the founder of 'free market economics' is a misrepresentation.<sup>209</sup> Why would North do this? One may perhaps only be able to give a speculative answer to this, and the why may be less important than the role this 'Smithian result' plays in his overall argument. Even so, I will introduce a to my mind plausible explanation here, one which I will explore further in discussing other major themes in North's work as well. My suspicion is that this use of Smith serves a triple purpose:

- 1) in that 'Smith' stands for the way markets can turn coordination problems into mutually beneficial solutions, an attribution that not coincidentally has a pedigree in game theoretical approaches in economic and public choice theory<sup>210</sup>;
- 2) that 'Smith' stands for the ability of markets to promote the division of labour, distribute scarce knowledge, increase productivity, and other economic behaviour that is beneficial for economic growth and thereby for general well-being (this is clear from North's references to the 'wealth of nations');
- 3) that Smith's 'two faces', i.e. his emphasis on the way markets enable cooperative solutions to strategic games as well as his interest in their background in personal sympathy and moral values, represent each of the two aspects of North's own approach to markets, namely the special case described by neoclassical theory and the general case for which the emphasis must be on social institutions (norms and values as rules of the game). In both cases, Smith can be said to stand for the suggestion that society works best when both the market itself is relatively free, competitive, and impersonal, and when the normative institutional order is oriented towards market exchange.<sup>211</sup>

---

<sup>209</sup> E.g. Lamb 1974; Brown 1997; Kennedy 2009.

<sup>210</sup> A classic statement is Tullock 1980.

<sup>211</sup> See Smith 1998 for an argument along these lines.

Not all of this is explicit in North's own writings in this way. The references above are to authors in the public choice and economic theory tradition close to his own, but that does not mean we can simply conflate them; rather, they may serve to provide us with an idea of how to clarify this aspect of North's work. This is therefore emphatically an interpretation of why North should seek to use this notion of the 'Smithian result' in this way, without much clarification or justification in his works. But as mentioned, the equation of Smith to the wealth generating and cooperation-enabling abilities of markets certainly is explicit. North uses for this the claim that institutions do, or should, serve 'capturing the gains from trade'. For example in *IICEP*, we are straightforwardly told that "the central focus is on the problem of human cooperation - specifically the cooperation that permits economies to capture the gains from trade that were the key to Adam Smith's *Wealth of Nations*. The evolution of institutions that create an hospitable environment for cooperative solutions to complex exchange provides for economic growth."<sup>212</sup>

That trade has such gains, and that they can be captured by institutional arrangements that allow an approximation of competitive, open and non-rent-seeking markets is taken as given. In the background of such an assumption, I think, is North's commitment to what one might call a classical or ordoliberal perspective on society, something which comes to the fore in his contractarian approach to the formation of institutions. I will elaborate on both in the next chapter; but suffice here to say that North has made little secret of his fundamental political sympathy with the classical liberal tradition.<sup>213</sup> Partially this is exemplified by his effusive praise (though not without criticism) of Friedrich Hayek, partially it becomes clear through his own description of his trajectory in economic thought: moving from Marxism through neoclassical economics to an interest in what makes efficient markets possible. Since he abandoned Marxism, his work has throughout been characterized by the search for the right way of providing an effective but limited government, competitive markets, strong individual rights, and a tendency to explain institutional orders as emerging out of organically formed social contracts.<sup>214</sup> One may also point, as some have done, to "his status as Senior Fellow of the Hoover Institution at Stanford University [and] his participation in activities with the Cato Institute".<sup>215</sup>

North himself makes his classical liberal perspective on freedom in the market more clear when, in an essay in *National Review* on the problems with ever-expanding government tasks, he concludes: "From my perspective, individual liberty is inextricably entwined with the options—the alternatives—available to individuals in a society. By this definition individual liberty has been seriously eroded. The choices of occupations, the decision to hire, fire, or promote employees, the exploitation of natural resources, the establishment of new enterprises, the

---

<sup>212</sup> North 1990, p. vii.

<sup>213</sup> A brief discussion with biographical information and autobiographical reflections can be found in Klein and Daza 2013.

<sup>214</sup> The latter is already explicitly evident in his 'efficiency' explanations of the medieval manor as a social contract, see: North and Thomas 1971. But it reappears in his later works, as discussed in chapter 3. See also Bates 1988 for an early critique.

<sup>215</sup> Klein and Daza 2013, p. 527.

determination of quality standards for products, the disposition of one's earnings all are increasingly more circumscribed than in the past... What have we obtained in return—greater economic security for individuals, a more desirable distribution of income, the reduction of environmental deterioration? Aside from the difficulty of agreeing about what is a desirable distribution of income, the assessment of these benefits (particularly as compared to hypothetical alternatives) would differ widely. I am somewhat sceptical about our successes in these directions. Yet obviously, we have experienced some degree of success. But the cost in terms of my definition of individual liberty has been substantial.”<sup>216</sup>

To point to this context is to give a plausible political explanation of why for North the emphasis should be shifted from the given *desideratum* of competitive markets and small governments to the institutional context that makes it possible. On my argument, it is this role that is played at the various levels outlined above by the references to the ‘Smithian result’. To say that North’s politics have classical liberal influences is not to reduce his theory to his political views, or to say that it stands or falls with it; nor is it to simple-mindedly accuse North of being a *laissez faire* enthusiast in the style of those pseudo-‘Smithians’ associated with the Adam Smith Institute and similar bodies. For North, as for the actual Adam Smith, the institutional-normative context of markets is at least as important for what markets do as the degree of *laissez faire*.

Even so, the ‘Smithian’ framework – whether actually to do with Smith or not - provides for a remarkably different idea about markets than Polanyi’s careful distinctions between markets, trade, and money, and between price-making and non-price-making markets, despite North’s willingness (in 1977) to concede the merits of Polanyi’s approach for understanding the premodern economy. Smith therefore stands, one suspects, not just for an analytical assumption, but also for a normative criterion: the *desideratum* of modern, impersonal, competitive market exchange, along the lines of the models of the neoclassical colleagues North otherwise excoriates. If he disagrees with his colleagues at the analytical level, he certainly disagrees with them less than with Polanyi when it comes to the desirability of competitive market exchange as the primary allocation system – provided one gets the institutions right.

### **Heroic opportunists and as-if markets: The critical reception of Douglass North’s market theory**

For some, this approach of North’s has been one of the sources of the strength and viability of North’s NIEH research programme. As I have mentioned, the secondary literature on North frequently places him in the context of NIE theory as a whole, or even in the wider context of institutionalist economics as such. North’s interest in the interaction between individual and institution-as-rule, and his appreciation of the significance of normative and social conceptions and cohesions for that interaction, form a contrast to the approach of Coase, Williamson, and

---

<sup>216</sup> North 1983, p. 812.

fellow neoinstitutionalists for which North's work is often valued. The partial convergence, but also remaining differences, between North's approach and the concerns of the OIE (even as far back as the German Historical School) is also a frequent subject of discussion in the literature.<sup>217</sup> Not all of this literature properly separates North's later work - focused as it is on bounded rationality, ideology, cultural evolution, and problems of cooperation - from his former efficiency-based model of institutions. Some of it predates his latest, most sophisticated and ambitious publications.<sup>218</sup> But among the literature that engages with North's developed NIEH theory comparatively in this way, only a few works stand out in their discussion of North's theory of historical markets in the context of this work's anthropological and philosophical concerns. I will here discuss and evaluate a few of these contributions, before giving an assessment of the problems with North's market-centric perspective on history. In closing, I will comment briefly on the suggestive role of market rhetoric in North's work.

In separate publications, Claude Didry and Caroline Vincensini as well as Don Kanel have observed the importance of North's 'Polanyian moment' for his market theory.<sup>219</sup> Indeed, the former suggest to some extent that North's Polanyian engagement differentiates him from the other New or neo-institutionalists on precisely this point: the nature and historical context of market exchange.<sup>220</sup> What is striking is that for both these commentators, among the very few to have noted the importance of Polanyi's substantivist (anthropological) institutionalism for North, the latter's work is seen in some sense as a continuation or extension of the former's approach. The reason for this is that they identify Polanyi's challenge as the significance of the market/non-market distinction, one that North has clearly taken up.<sup>221</sup> In this, North's approach is rightly identified by these authors as opposed to the markets/hierarchies distinction of Williamsonian neoinstitutionalism and the neglect of institutions in neoclassical economic theory. As Didry and Vincensini put it in very Polanyian terms (though referring to what I have called neoinstitutionalism): "'new institutionalism' ignores this challenge by sticking to the coexistence of two spheres, market and non-market, in economic activities, the latter referring to 'social structures' in which the economy is embedded."<sup>222</sup>

These authors are clear that for them, North's market theory is a clear improvement over those of his neoinstitutionalist and neoclassical rivals, especially insofar as North has accepted Polanyi's challenge, and much of Polanyi's framework for thinking about markets besides. Kanel defines North's market theory as follows: "transaction costs... prevent availability of [public goods] through market transactions, and a non-market way of making [them] available is a better allocation of resources, but only if social pressure and an internalized sense of honorable

---

<sup>217</sup> E.g. Peukert 2001; Rutherford 1994; Groenewegen, Kerstholt and Nagelkerke 1995; Harriss 2002.

<sup>218</sup> E.g. Field 1981.

<sup>219</sup> Didry and Vincensini 2011; Kanel 1985.

<sup>220</sup> Didry and Vincensini 2011, p. 3.

<sup>221</sup> *Ibid.*, p. 3-4; Kanel 1985, p. 815-816.

<sup>222</sup> *Ibid.*, p. 3.

behavior keeps free riders in line”.<sup>223</sup> He subsequently states his agreement with the usefulness (suitably adapted) of this theory for historical explanation. For Didry and Vincensini, North’s cue from Polanyi allows him to make “institutions as explanatory factors of historical dynamics by going beyond the opposition between market and non market dimensions of the economy”, clearly a *desideratum*.<sup>224</sup>

I use these examples not because these authors have been particularly influential in the secondary literature on North, but because they clearly illustrate how even those who have observed the importance of Polanyi’s challenge for North, and the way North has taken up Polanyian ideas about markets and historical differentiation, still judge the former’s NIEH approach primarily by comparison with the main alternatives for whom institutions are not such a central variable. Although these works stand out for their focus on the Polanyian challenge, this means of evaluating North’s theory is fairly representative for the secondary literature on the whole. It tends to emphasize, for good reason, the advantages of North’s approach over the Williamsonian or the neoclassical one for understanding embeddedness and ‘vectors of historical comparativity’, as Didry and Vincensini put it.

The question to what extent North’s market theory in the NIEH takes Polanyi’s arguments seriously and has a good answer to the latter’s critique of the ‘economistic fallacy’, including his substantivist perspective on markets and market exchange, threatens to be overshadowed by this excessive focus on less institutional economists. Similarly, this line of argument tends not to ask to what extent North’s virtues are also his defects: that is, whether in situating himself between Polanyian ideas and neoinstitutionalist and neoclassical ideas does not lead to an incoherent approach to the embeddedness of institutions. Granovetter found a solution in sociological network analysis, but his theory did not engage with Polanyi and North’s primarily historical concern. Because the nature of markets in history is one of the central issues raised by Polanyi’s challenge, this makes the question whether North really can have it both ways – to have his Polanyi and his economic approach too - all the more urgent.

North has also been criticized by a number of authors on his view of the market/non-market relationship and his privileging of perfect competition in markets as a benchmark. Some of these are quite wide ranging and discuss also his historiography and views on individual rationality and motivation, subjects I will discuss in the following chapters. Here, I will concentrate on the problems identified in the literature with this idealized markets as benchmark approach. Particularly significant are the ambiguities in distinguishing competitive markets as interpretative heuristic and as normative ideal in North’s theory, and the problem of seeing institutions as substitutes or complements for markets even where markets (by North’s own

---

<sup>223</sup> Kanel 1985, p.818.

<sup>224</sup> Didry and Vincensini, p. 4.



admission) have in premodern history often not been dominant or even significant allocation systems.

One such critic is Daniel Ankarloo, who has written several critiques of the theory of markets in New Institutionalism, using both Williamson and North as cases of this approach. Ankarloo objects to this conception of markets at several levels. Firstly, he observes that there is an apparent inconsistency between the use of transaction costs as ‘shadow prices’, that is as the “underlying costs of exchange”, and individuals acting rationally, in however bounded a form of rationality, to reduce transaction costs and maximize gains.<sup>225</sup> After all, if transaction costs are not already reflected in market prices, how are agents – boundedly rational or otherwise – to know what the transaction costs are and act according to standard choice theory in response to these relative costs? If this is allowed, then it is allowed that either individual rationality in market settings is not just a response to relative costs, or that if transaction costs do not appear in a market setting as prices, there is no way for agents to obtain information about such transaction costs. In either case, transaction costs cannot explain individual choices on their own. Only the institutions generated by transaction costs can, and therefore institutions rather than transaction costs should explain choice behaviour.<sup>226</sup> In other words, in explaining institutions by transaction costs, North puts the cart before the horse, as far as Ankarloo is concerned.

Secondly, Ankarloo has extensively criticized what he calls the ‘as-if economic history’ produced by the market theory of New Institutionalism. Although this critique has been aimed at a wide range of NIE and associated authors, including Ronald Coase, Oliver Williamson, and even Morris Silver, it is also applied to North’s NIEH approach.<sup>227</sup> With ‘as-if economic history’ Ankarloo seems to describe the effects of this market benchmark on the explanation of non-market phenomena in North’s NIEH. If the market is the explanation for the existence and nature of the non-market, there is potentially a problem of assuming what needs to be proved. Ankarloo hones in on just this point. As he writes, “the market in North’s explanations remains the universal yardstick – the eternal reference point. But if we acknowledge that markets and the institutions of capitalism originate from somewhere, and evolve, we cannot assume them from the start. The market is not a universal yardstick for the comparison of institutions.”<sup>228</sup> Yet he notes how North in his discussion of Polanyi’s challenge – a point of departure for Ankarloo’s argument as well – acknowledged the insignificance or absence of market exchange as allocation system in premodern history.

The point of North’s approach is to show that the economics approach can explain such allocation systems, their institutional structure and the direction of change regardless. But, Ankarloo suggests, given the above problem North can only do this by assuming the very market

---

<sup>225</sup> North 1989, p. 661; Ankarloo 2006, p. 8-10.

<sup>226</sup> Ankarloo 2006, p. 9-10.

<sup>227</sup> Ankarloo 2002; Ankarloo and Palermo 2004.

<sup>228</sup> Ankarloo 2002, p. 18-19.

relations he has just denied: “This is due to the fact that institutions evolve as rational responses to relative price changes, by, albeit “constrained”, micro-rational economic men, in his model. But such calculation cannot explain the origins of the market. Calculating what to buy only makes sense when there already are commodities on the market to buy... North fails to acknowledge the fact that the “free” individual – the micro-calculating consumer – is the result of capitalist markets, rather than the other way around. In North’s theory this individual has no history, and that is a critical limitation on his program of “history matters”. This problem of neoclassical theory is not solved at all. It is inherited.”<sup>229</sup>

In several works, Ankarloo has argued that this incoherence of the NIEH originates in the NIE’s ambiguous position between what he terms ‘neoclassical orthodoxy’ and the desire for a more realistic theory of historical and present economics than that orthodoxy allows. In words applied to Oliver Williamson’s approach, but clearly intended to describe the NIE as a whole, he argues: “Williamson tries to solve these dilemmas by escaping from them. Markets are everywhere. Capitalist micro-rationality is a universal trait of human beings (even in the absence of prices and markets)... But granted the fact that capitalism — including capitalist markets and firms—is a result of history, the market too should be considered as a consequence, not the cause of historical development.”<sup>230</sup> Hence, the economic history produced by the NIE approach is ‘as if’ economic history: it is history ‘as if’ the economy and politics were structured by markets throughout.<sup>231</sup>

These two objections Ankarloo has formulated are, I believe, closely related. In both cases, the problem results from the market as benchmark, as means of explaining institutional structures *even in the acknowledged absence of significant price-making markets*. Polanyi’s point about premodern economic history had been precisely to point out that the absence of such price-making markets required an institutional explanation of economic behaviour that could *not* be rooted in responses to relative prices. Therefore, institutions could only be explained in terms of the social and cultural characteristics of the societies in question, and behaviour in the economic sphere, i.e. the transactional modes, only explained in terms of those institutions. It is clear that the NIE approach must reject such an assumption, for its *raison d’être* is the ability of an economics approach to explain institutions and the choice between markets and nonmarket economic formations. Much of the neoinstitutionalist tradition has undertaken this in a somewhat Panglossian manner.<sup>232</sup> For them, institutions must be efficient responses to transaction costs, or else private contracting would have supplanted them.<sup>233</sup> This can be supported only by a

---

<sup>229</sup> *Ibid.*, p. 19.

<sup>230</sup> Ankarloo and Palermo 2004, p. 420.

<sup>231</sup> *Ibid.*, p. 419.

<sup>232</sup> This is the basis for Granovetter’s critique of their approach as ‘undersocialized’: Granovetter 1985. For a similar critique, see Hodgson 1991.

<sup>233</sup> Williamson 1975, p. 8.

relatively strong idea of rationality and by Williamson's explicit ahistoricity: the assumption, strongly assailed by Ankarloo among others, that "in the beginning there were markets".<sup>234</sup>

We must consider to what extent Ankarloo's critique applies to North, at least in the later stages of North's work. After all, North does not say that in the beginning there were markets, and he explicitly abandoned and rejected the efficiency model of institutions as unrealistic. (I postpone discussion of how compatible this is with North's rationality assumptions to the next chapter.) Indeed, not all of Ankarloo's critiques seem to me to hit the mark. In his paper on North, Ankarloo alleges that for him, too, "explanation of these phenomena [institutions and transaction costs] starts from orthodox equilibrium: a 'zero transaction cost world', where institutions do not exist." Since it is not obvious how one would get to a world of institutions from a world of zero transaction costs, Ankarloo reasons one must then concede the existence of institutions independent of transaction costs, which defeats the purpose of the transaction cost argument.<sup>235</sup> But this is to underestimate the historicity of North's approach: for him, the benchmark of competitive markets is not an *empirical* starting point. Rather, it is a way of analytically moving from the neoinstitutional framework to a still more realistic and historical discussion: why nonmarket allocation systems often prevail and why institutions matter, indeed persist despite their inefficiency.

We must allow that North does not claim that markets have always existed, nor that institutions need be efficient. Also, as the discussion of embeddedness and North's response to Polanyi has shown, North's NIEH approach does not concentrate on the markets vs hierarchies tradeoff, but rather on why markets often do not appear; or when they do, why they do not function according to the efficiency models of the neoinstitutionalists and neoclassical economists alike. In that sense, North is more of a critic of the Williamsonian view - at least implicitly and on occasion explicitly - than Ankarloo and other critics allow when they incorporate North in their sweeping critiques of the NIE in general. As I have argued before, it is not unproblematic to see North's work as simply an application of existing NIE theory, especially as regards the later stages of the NIEH.

However, Ankarloo's general point about the potential incoherence of North's still market-centric approach is well-taken, in my view. It is perhaps possible to read his work as an inquiry into the origins of markets as well as nonmarket allocation systems. But in taking up Polanyi's challenge he committed himself to providing an explanation of these that could rest on choice theoretical microfoundations and on explaining variation in institutions by differing transaction costs. North has certainly made serious attempts at estimating transaction costs empirically; but always in those Western economies already taken as benchmark to begin with, and there is no consensus on how they are to be measured.<sup>236</sup> While some inspired by New Institutionalism have

---

<sup>234</sup> *Ibid.*, p. 21.

<sup>235</sup> Ankarloo 2002, p. 10-11.

<sup>236</sup> Wallis and North 1986; Wang 2003.

sought to measure “non-market transaction costs”, this has generally come in the shape of studies measuring the costs of obtaining government licenses or of finding trading partners and the like.<sup>237</sup> But these are relevant examples for the Williamsonian paradigm in studying the relative costs of competitive market exchange versus government action as substitute. The same applies for studies that have focused on the quality of institutions in contemporary societies for achieving growth.<sup>238</sup> As Ning Wang points out, generally “what is measured or indexed in these studies is not transaction costs per se, but the cost of institutional inefficiency or poor governance.”<sup>239</sup>

These studies, whatever their value, do not address the problem of the meaning or measurement of such ‘non-market transaction costs’ in societies where private contracting in response to relative prices are not the dominant allocation system, which is North’s ‘challenge’. This problem is conceptual: it is not obvious how one could say what the transaction costs in a given historical economy are except insofar as individuals are attempting to contract through markets and prices are able to adjust to reflect the relevant costs. Such phenomena certainly have existed in economic history, and the theory of transaction costs may explain relevant aspects of such markets. North for example rightly points to the importance of standardized weights and measures for reducing costs of transacting in formalized market-places.<sup>240</sup> Equally, it may be possible to interpret the *effect* of informal institutions (norms, bonds of trust, and their in-group enforcement) in trading settings as reducing transaction costs and thereby making e.g. caravan trade or other long-distance trading operations in premodern settings easier.<sup>241</sup>

It is quite another thing, however, to explain *nonmarket* behavior by these means, especially in societies where, by North’s own admission, most economic transactions do not take place by the price-sensitive exchange of privately owned commodities. It is telling that the wider NIE-inspired discussion of transaction costs in empirical studies has virtually exclusively focused on cases of private trade for gain. This suggests Ankarloo’s critique points to a serious weakness in the analysis. Even insofar as it is possible to identify the function of institutions in such a market setting as reducing transaction costs in some form, or as expressing the interests of particular powerful agents in maintaining higher transaction costs to extract rents – this as regards both economic and ‘political markets’ - this functionalist explanation is not generalizable. Where such individual contracting is not already operative in a way sufficiently analogous to the ‘benchmark’ case of perfect markets, no functional explanation in terms of transaction costs is possible. The *kula* trade cannot, *pace* North, be explained as a least cost trading solution because it first must be demonstrated that it is, in fact, a ‘trading solution’, and secondly, there must be a measure of the transaction costs operative in the situation that can explain the *kula* trade and

---

<sup>237</sup> E.g. De Soto 1989; Benham and Benham 1998.

<sup>238</sup> Rodrik 2000; Rodrik, Subramanian and Trebbi 2002.

<sup>239</sup> Wang 2003, p. 8.

<sup>240</sup> North 1991, p. 100.

<sup>241</sup> *Ibid.*, p. 100; Greif 1993; Greif 2012.

explain why precisely this ‘solution’ was found rather than another. It is not obvious that North’s approach is capable of either.

This is where North’s own ambiguous position between methodologies causes difficulties for his theory. North’s dismissal of Williamsonian efficiency explanations is effectively a dismissal of a relatively powerful functionalist explanatory method that rational choice efficiency assumptions provide. Williamson can assume that any given institution must be efficient for the market in some way, and therefore try to investigate what aspects of market failure would make this given institution ‘necessary’ from that viewpoint. One can doubt how plausible such explanations are, but the functionalist mode allows the model to be at least fully determined: there is always a definitive explanation for why an institution exists. In North’s case, this efficiency assumption is rejected. Moreover, North admits that ideology affects the formation of institutions, that path dependence can lock in institutional patterns across time, and that markets are often minimal or absent.

## **Conclusion**

Given these concessions, North’s view of markets and institutions is certainly more realistic and historical than Williamson’s approach. But it also means that he has very little left by which he can give any determinate explanation of a given institutional framework, *unless he assumes that the situation is functionally analogous to self-regarding rationality in a setting of market exchange*. This explains perhaps his insistence on concepts like ‘political markets’. But North cannot demonstrate that the *kula* trade, or the medieval Catholic church, or the Kwakiutl potlatch are functionally equivalent in this way, because there are no prices that could reflect transaction costs and he may not assume that these institutions are efficient in market terms. He can therefore only *assume* this functional equivalence; and this despite having denied himself the assumptions of institutional efficiency and the appropriate model of individual rationality he needs to make this assumption work. Therefore while North allows that markets and other allocation systems are embedded in institutional frameworks, he cannot fully carry through this idea. If he did so, he would have to abandon the notion that transaction costs can explain institutions, and to do so would be to give up on the NIE framework of his theory. He must either concede too much to Polanyi and other primitivist (and anthropological) institutionalisms and give up on the idea that transaction cost economics can explain the ‘choice’ between market and nonmarket institutions, political or economic. Or he must concede too much to the “economics approach”, and give up the attempt to provide a theory of institutions and historical change that is able to explain premodern economic formations in a realistic fashion. This is the dilemma North’s NIEH faces, at least as far as its theory of markets and other allocation systems is concerned.

To wrap up, let me return to the problem I have identified in North’s inconsistent functionalism about markets and transaction costs. That North’s theory needs some type of

functional explanation also has consequences for the coherence of his assumptions at other levels. Firstly, it needs to be seen how the explanation of individual behavior in terms of intentionality, i.e. beliefs and desires, can be reconciled with such a model absent certain assumptions about rationality and efficiency. North is keen to emphasize the boundedness of rationality as well as the importance of ideology and other mental content for the behavior of individuals towards and within institutions. Moreover, his view of rationality underwrites the problem of cooperation that the 'choice' between markets and other allocation systems is supposed to solve. In the next chapter I will discuss to what extent North's theory is coherent at this level.

## Chapter 3: Rationality, Social Contract, and Cooperation in North's NIEH

### Introduction: cooperation and its constraints

In this chapter, I will examine the interaction between the behavioural (choice) foundations of North's NIEH theory and the role of ideology, cognition, social learning, and other foundations for informal institutions within it. Since in this domain in particular North's distancing from neoclassical economics is most rhetorically marked, and yet at the same time most incomplete, this requires in part a review of the well-established debate around rationality and rational choice approaches in social theory. Considerations of space and coherence do not allow me to engage here in a full discussion of all aspects of the rationality assumptions common in mainstream economic theory and their problems, nor to fully explore the vast literature on free riding and the prisoner's dilemma. Rather, I will try to focus on a few salient elements of the mental foundations of behaviour as depicted in North's approach: (bounded) rationality, ideology, and cooperation.

In the discussion of North's approach to markets, I have only considered some aspects of these insofar as they relate to market theory: namely the contrast between Polanyi's motivational approach and North's 'as if' approach to market orientation. What makes North's response problematic, I argued, was that while North wishes explicitly to acknowledge the importance of ideological and moral considerations in economic behaviour as much as 'institutionalised' behaviour in general, he does not draw Polanyi's conclusion that if these matter, they should also matter for our interpretation of the economic phenomena themselves. Insofar as monetary exchange, markets, and trade are constituted by the behaviour of individuals, their ideological and mental motivations as well as representations involved in these economic activities should, Polanyi argued, affect both meaning and function of the economy in our model of a given society.<sup>242</sup>

North, too, acknowledges the importance of ideology and gives moral, religious, and other considerations a central role in his theory in the appearance of informal rules; yet for him, ideology and moral norms are often claimed to be merely *constraints* on individual behaviour. Even where containment of violence, and questions of power and rent-seeking are involved, the fact remains that North assumes that "the forces that lead to the substitution of firms for markets today may also help us to explain the variety of forms of economic organization in past societies" and (implicitly at least) that therefore individual motivation is *in this respect* merely constrained, but not fundamentally changed, by the operation of morality and ideology. Indeed, here North appears to have recourse to a familiar economic argument from the rational choice literature: namely that although people may not historically consciously (motivationally) pursue the reduction of transaction costs, we can understand institutional formation and change *as if*

---

<sup>242</sup> See e.g. Polanyi 1947, p. 98-101.

they do. As he acknowledges, “it is something else again exactly what the intentions of the players are. The overall direction of economic change will reflect the aggregate of choices made by political and economic entrepreneurs with widely diverse objectives, most of them not concerned with the consequences for overall performance”. But even with this *caveat*, his assumption remains that individuals act on institutional frameworks as ‘players’ of a given ‘game’, in which they seek to “line their pockets”.<sup>243</sup> Since these statements (appearing on the same page of *UPEC*) are not clearly reconciled, I must assume in line with North’s general ‘as if’ approach that this is intended as an instrumentalist approach to modelling such interactions, rather than a descriptive one. Even so, it is difficult to remove the impression of a fundamental contradiction.

An important role is here played by this rhetoric of game theory, which requires consideration both in terms of the methodological implications of North’s use of it and in terms of its rhetorical significance. As I have indicated in the previous chapter, to my mind these methods are not neutral in their implications. Depending on how they are used and how they are justified - two aspects I will investigate in this chapter - particular concepts of rationality, including those modeled by game theoretical methods, have consequences for the explanatory role of individuals and of institutions alike. They affect these at two levels: at the level of the relationship of agents to institutions, and at the level of individual agency itself. In North’s treatment of these problems, it appears all the more that they cannot just be seen as modelling assumptions for the sake of tractability and simplification, but at least as much as part and parcel of what one might call the ‘pessimistic anthropology’ of much social contract theory.<sup>244</sup>

In other words, to explore North’s shifting conceptions of rationality is also to explore his contractarian view of institutions and their role in achieving human cooperation, a central concept to much of his more recent work. For North, certain assumptions about gain-oriented motivation, the pervasiveness of violence, and the prevalence of prisoner’s dilemma type situations play a major role in his accounts of how institutions function: whether they do or do not overcome free riding, constrain violence, and (in his words) ‘achieve the Smithian result’. This aspect of his work I have highlighted already in the previous chapter, but it will prove important again in the present discussion. If the existence or nonexistence of this ‘Smithian result’ is the criterion for how well institutions enable cooperation in a particular historical society, one may expect this view to also affect North’s ideas about how individual rationality works and how it brings about particular institutional orders rather than others. We move therefore here to the most ‘micro-level’ analysis in North’s theory. His approach, or rather approaches, to rationality and ideology engender the interaction between individuals and institutions in his NIEH, and these in turn sustain the larger edifice in which institutions and their function for cooperation are part of a contractarian, market-oriented vision, such as described in

---

<sup>243</sup> North 2005, p. 79.

<sup>244</sup> E.g. Bruni and Zamagni 2013, p. 2. A term originally drawn from Christian theology. See: Pannenberg 2004, p.172; Roberts 2010. Cf. Cliteur 1990 on a similar worldview in Hayek.



the previous chapter. Moreover, this meso-level in turn supports the macro-level evolutionary dynamic of human culture (artifactual frameworks) North identifies, which - although departing from our most basic cognitive capacities and limitations - provides a grand narrative explanation of recorded human history as the unfolding of greater or lesser levels of ‘adaptive efficiency’.

The purpose of this chapter is not to engage with these methodological concepts for their own sake. Indeed, the literature on rationality, the social contract, self-interest or even game theory (which is only implicitly referred to in North’s work) could fill many bookcases. Rather, in this chapter I want to engage more deeply with some of the presuppositions of North’s thinking about the classic economic and social problem of how individual agency can sustain large scale cooperation, and what role markets and institutions play in this process. I have said much about the latter, and will say more yet, but the question of cooperation and the formation of spontaneous order, and the individual rationality or lack thereof that determines the difficulties of approximating an optimal outcome in aggregating individual agency remains. North’s references to the ‘players’ of the ‘game’ or to the difficulties and desirability of achieving the ‘Smithian result’ are merely examples of how he conceives the relationship between individual agency, the mediating role of markets and institutions, and the *desideratum* of low-cost exchange. To make sense of this, we must have a – necessarily extremely brief - look at the context of the contractarian tradition in North.

I will focus here first on the ‘spontaneous order’ tradition of Hayek and his idiosyncratic mixture of epistemology inspired by evolutionary theory and classical liberal politics, which I think has had considerable influence on North’s thinking on the above concepts. This may also serve as a framing device for more practically discussing this otherwise rather broad, even unbounded, range of problems in social theory as they appear throughout North’s work. In chapter 4 I will then return to this tradition, but concentrating on the evolutionary dimension in this Hayekian tradition. Subsequently, I will discuss how this tradition has expressed itself in economics in game theory and the debates about cooperation, rationality, and spontaneous order as they are interpreted through this form of rational choice analysis.

In light of these discussions, I want to clarify what it means for North to refer to the ‘players of the game’ who are self-interested (‘lining their own pockets’), and why this creates a point of tension with his critique of neoclassical economics as based on rational choice theory. As discussed in chapter 1, North’s critique of neoclassical economics emphasizes in particular the importance of a substantive, non-probabilistic meaning of beliefs (or ‘ideology’) for a proper concept of rationality, something rational choice theory does not in his view comfortably accommodate. Yet his game theoretical language and his references to rational choice of institutions as a self-interested process would suggest otherwise. Since in standard choice theory as expressed in game theory rationality has a specific set of criteria, derived from the axiom of utility maximisation, North’s attempt to simultaneously use this rationality concept and critique it becomes a puzzle. I will then explore the difficulties of North’s use of ‘bounded rationality’ and – later in his oeuvre – cognitive science as solutions to this puzzle.

## Spontaneous order, cooperation, and the classical liberal tradition

In the classical liberal economic and sociological tradition, the concept of spontaneous order has long been important. Deriving from the thinkers of the Scottish Enlightenment in particular, such as Adam Smith (as invoked by North), but also David Hume and Adam Ferguson, the concept of spontaneous order suggests the aggregation of the actions of very many individuals in such a way that society does not become – as Hobbes suggested – a *bellum omnium contra omnes*, but that rather, for example in the generation of public goods, a cooperative order ensues in which the whole is more than merely the conflict between its parts. In the classical liberal interpretation of this approach, in particular in the tradition of Herbert Spencer and Friedrich Hayek, this idea is further specified. In this tradition, it is economic coordination through the market that enables cooperation, and to be precise, cooperation *through the action of self-interested individuals*. Already in the work of Bernard Mandeville and in Spencer the concept of ‘private vices, public benefits’ plays a foundational role in the analysis of spontaneous order.<sup>245</sup> For Hayek, it was specifically the problem of limited information, implying the ubiquity of tacit and local knowledge, that made cooperation only possible through market coordination. As Peter Boettke notes on Hayek’s interpretation of the Smithian legacy: “The voluntary action of thousands of individuals, each pursuing their own interests, generates and utilizes economic information that is not available to any one individual or group of individuals in its totality... Smith’s case for economic liberty amounts to an argument and demonstration that individuals pursuing their self-interest can, and will, produce a social order that is economically beneficial... It is this emphasis on the use of contextual knowledge that underlies the critical defense of the liberal order from Smith to Hayek”.<sup>246</sup>

One operative word for this idea of mutual benefit, mediated by the market, is cooperation. This word is not always clearly defined in the relevant economic literature, and in any case can have a range of meanings. In game theory, it means the opposite of defection, in other words to choose that strategy which would be mutually welfare enhancing if jointly followed (say in a 2-player noncooperative game): the prisoner’s dilemma game revolves around the possibility that such a mutual cooperation may not be possible by following self-interested rationality.<sup>247</sup> This does express the more general notion that ‘cooperation’ implies some form of mutual (economic) interaction such that it enhances welfare for both parties in traditional welfare economic terms. In this political economic tradition, cooperation then comes to have a very particular significance (and signification), one that leans on the contractarian pre-history of the Scottish Enlightenment and its liberal progeny itself. In the social contract political philosophy of the early modern age, the main problem was to consider how many different individuals could coordinate their

---

<sup>245</sup> For an anthropological interpretation of this literature, see e.g. Cronk 1988.

<sup>246</sup> Boettke 1990, p. 64-65.

<sup>247</sup> This meaning of cooperation as a welfare-enhancing strategy of mutual benefit is not to be confused with the concept of a ‘cooperative game’, which depends on the possibility of enforceable agreements prior to moves in the game itself. I will here be concerned with non-cooperative games, in which such enforceable agreements (or even any communication) are not assumed possible prior to the (first) game.

behaviour and interests in such a way as to generate a legitimate and (preferably) mutually beneficial form of government. While the answers varied considerably from one theorist to the other, the problem as such remained fairly stable: one could say as a question of how to make possible sufficient coordination (collective action) that a society, thus formed, could achieve cooperation, i.e. mutual welfare benefit.

In this liberal contractarian tradition one can trace through Spencer and Hayek, then, we find two vehicles of such coordination: the free market, as seen above, and the sphere of social institutions. The latter coordinates the psychological or cultural propensities or norms of individuals in much the same way that the market, through the price mechanism, coordinates their local information (beliefs) and preferences. Institutions constitute the ‘order’ implied in spontaneous order, a social arrangement that can be more or less favourable to the operations of decentralized market exchange. As Boettke summarizes the Hayekian view: “The key question for the social theorist is how the various and diverse images of reality that individual minds develop could ever be coordinated to one another. The social institutions that arise through the voluntary association of thousands of individuals serve to guide individuals in the process of mutual accommodation.”<sup>248</sup> Through the spontaneous formation of institutions, and through the process of market exchange, a spontaneous order – a form of collective cooperation – arises. Not coincidentally, this emphasis on *order* as the (desired) outcome of collective action, either as the form taken by cooperation or as the preconditions for it, is found also throughout *Violence and Social Orders*. Order as in safety from violence is a precondition; order in the sense of a (more or less) stable set of institutions with welfare-enhancing properties (relative to the Hobbesian state of nature) is the result.

The difficulty with this perspective is the question of how desirable results should come about by means of coordination, and indeed whether this leads to desirable results at all. In the Spencer-Mandeville tradition, the outcome of the process of spontaneous order is often depicted as optimizing, so that ours is the ‘best of all possible worlds’, or would be if we let the process do its work. In the Hayekian tradition, such optimization and efficiency is explicitly disavowed: for him, there was nothing about either coordination through the market nor through social institutions that necessitated any kind of optimality about the result, only the impossibility of improving on it by any form of *pre-planned* coordinating action, especially by the state.<sup>249</sup> For David Ramsay Steele, this claim – particularly strong in Hayek’s work of the 1980s, such as *Knowledge, Evolution, and Society* entails a change in Hayek’s own relationship to the classical liberal tradition of spontaneous order: from a liberal claim that the resulting order would be one characterized by desirable traits of cooperation to the conservative claim that however undesirable the order, it simply is as it is and cannot be consciously improved upon.<sup>250</sup>

---

<sup>248</sup> Boettke 1990, p. 69.

<sup>249</sup> *Ibid.*, p. 75-76.

<sup>250</sup> Steele 1987; cf. Hayek 1983, p. 36.

Although whether or not this entails a shift within Hayek's own perspective is disputed, it is perhaps illustrative for Hayek's influence on North that the latter has followed a somewhat similar trajectory: from the efficiency contractarian approach, where the medieval manor was as it was because it was optimal that way, to an explicit disclaiming of such optimality, preferring to see – as Hayek did – the process of institutional change as an often decidedly *inefficient* process. Like Hayek, he reorients the discussion away from the efficiency characteristics of a particular economic order to the evolution of institutional orders in general, where any optimality is at best local and temporal, but where the evolutionary characteristics of institutions and market exchange join together to form *some* spontaneous order, even if it is from the viewpoint of neoclassical economics a bad one. However, as Viktor Vanberg has noted for Hayek's case, this leaves open the question of why it is that the evolution of institutions should be such that it favours cooperative outcomes, as the spontaneous order represented by the free market suggests.<sup>251</sup> Hayek may have thought this outcome not 'optimal' in some strict sense, but certainly at least better than any other institutional order, for it is manifestly the case that many institutional orders have come into being that were not based (integrated, in Polanyi's terms) on free market exchange. I will engage in chapter 4 with the question of how and whether an evolutionary account can save the Hayekian argument.

North himself was quite explicit about the influence Hayek and his intellectual genealogy had on his own work. On one occasion, he declared Hayek to have been the “the greatest economist of the twentieth century, and by a long way” and even “if you look for people who really want to try and understand the world, Hayek came closer to that ideal than anybody who has ever lived.”<sup>252</sup> On one occasion, North discussed at greater length the contribution of Hayek to his own understanding of the process of economic change.<sup>253</sup> His discussion of Hayek focuses in particular on how Hayek has helped ‘us’ understand how intentionality and institutions contribute to generating economic change as well as constraining such change within the boundaries of inherited belief and normative systems. Moreover, in this paper North himself suggests a union of Hayek's insights with the ‘Smithian’ concept of expanding market exchange as driver of increases in productivity. This forms for him the basis of his approach to the study of economic change.<sup>254</sup>

The main difference between North and Hayek is the former's rejection of the idea of spontaneity in spontaneous order, at least insofar as this involves an *a priori* rejection of the possibility of social engineering of any kind. For North, we are rather condemned to social engineering because of our intentional engagement with our environment and each other, but we do – and here he has common ground with Hayek – very often fail to achieve what we aim for

---

<sup>251</sup> Vanberg 1986, p. 83.

<sup>252</sup> Cited in Klein and Daza 2013, p. 531.

<sup>253</sup> North 1999.

<sup>254</sup> *Ibid.* 1990, p. 80-81.

when doing so.<sup>255</sup> But besides this one mostly definitional objection, Hayek is clearly a major contributor to the shape of North's most mature work. Indeed, Hayek's ideas in *The Sensory Order* and elsewhere are also the foundation of North's shift towards cognitive science and the centrality of how beliefs are formed, which frame the discussion in *UPEC*, arguably his most developed theoretical monograph. The layer of New Institutional economic analysis in terms of institutions and path dependence is, for North, "complementary to and in the spirit of his framework".<sup>256</sup> Ultimately, for North "our economic history can be rendered intelligible in a Hayekian framework", joining an evolutionary approach to cognition and trial-and-error to the New Institutional conceptual toolset and the argument, shared by North and Hayek alike, that getting the institutions right is a question of fulfilling the promise of "Smith's happy conclusion".<sup>257</sup>

That said, it is worth noting that for North, this has been an impetus to study precisely the *problems* of coordination, with an emphasis on the frequent failure of the 'right' combination of institutions and market exchange to come about. It is this that is the subject of the present chapter, postponing (as mentioned) the evolutionary-cognitive dimension to the next. That North's theoretical inspiration, and indeed often (seemingly) normative standpoint regarding what form of institutionalised cooperation is most desirable from an economic point of view derives much from Hayek and this larger classical liberal contractarian tradition is beyond doubt.<sup>258</sup>

But North's comparative economic historical research gives it a decidedly more pessimistic cast. Partially this may be because much of economic history revolves around nonmarket forms of allocation and often very seemingly dysfunctional political institutional orders, as his discussion of Polanyi points out. But I suspect that his approach to economic rationality also has much to do with his pessimistic interpretation of the liberal contractarian tradition. As we have already seen, North's rejection of the 'efficiency approach' to interpreting institutional change goes hand in hand with his ever more emphatic rejection of the rationality criteria he attributes to neoclassical economics, preferring instead the bounded rationality approach of Herbert Simon. In a Hayekian vein, North excoriates the unrealistic nature of assuming optimizing forms of rationality and complete information. But the more he does so, the more restrained individual reason, the more difficult it becomes to see how a spontaneous order emerging from such reason could have the desirable properties associated with cooperation that the Smithian tradition (as interpreted by North) suggests. This applies in particular when such reason is assumed to be self-interested.

---

<sup>255</sup> Here he defers to the work of Vanberg mentioned above. See North 1990, p. 96n22.

<sup>256</sup> North 1990, p. 84.

<sup>257</sup> *Ibid.*, p. 95, 86.

<sup>258</sup> In this vein, Francis Fukuyama has also credited to Hayek "the systematic study of how order, and thus social capital, can emerge in spontaneous and decentralized fashion" as "one of the most important intellectual developments of the late twentieth century." Fukuyama 2000, p. 103.

Here the prisoner's dilemma and its implications become significant, especially in light of North's invocations of self-interested players of a game seeking to line their pockets as an explanation of institutional change. In a nutshell, the rational pursuit of self-interest by different 'players' of the 'game', the prisoner's dilemma implies, can lead to outcomes that are suboptimal from the point of view of all players involved.<sup>259</sup> The spontaneous pursuit of self-interested action therefore can have decidedly undesirable outcomes, which is a major challenge to the more 'Panglossian' version of the spontaneous order tradition. As Vanberg notes, the problem here is that this tradition needs to distinguish between coordination problems, which are self-enforcing once individuals can be brought together to pursue collective action, and those kinds of problems, exemplified by the prisoner's dilemma, where "there is no 'natural' harmony between constitutional and action interests, even if there is perfect agreement on the former between all members of the community".<sup>260</sup>

That is to say, the pursuit of self-interest may, when sufficiently coordinated, lead to positive cooperation in the design of social institutions (as rules or norms) and be sustainable because of mutual self-interest; but this does not guarantee that the pursuit of self-interest *within* such an order necessarily has this property. A classic example is the economic observation that while cooperation in the form of exchange may be mutually profitable to both parties, it is to greater individual advantage of each party to steal from and at the expense of the other. The result of rational self-interest in this case would be mutual theft, mutual non-cooperation, leading to the least desirable outcome in welfare terms.<sup>261</sup> It would seem then that insofar prisoner's dilemma type situations may be ubiquitous, rather than coordination problems, this poses a problem for the classical liberal emphasis on the mutual benefits of rational self-interest. The spontaneous emergence of institutions or conventions to solve this problem cannot be assumed, as in the coordination problem case, because the mutually rational strategy in a prisoner's dilemma is mutual defection, and no coordination can take place (*ex hypothesi*).

The question is then under what conditions there may be compensating factors that can realign incentives, provide additional incentives, or make cooperation otherwise more likely. This problem, relevant for the prisoner's dilemma, speaks to the evolution of reciprocity and its role in the formation of institutions, and to the evolutionary argument about spontaneous order more generally, as discussed in chapter 4. But another question is under what conditions coordination can take place in the first (i.e. coordination) type of problem, for which 'free' market exchange

---

<sup>259</sup> Technically, prisoner's dilemma games are games in which for any noncooperative equilibrium there exists at least one payoff vector associated with some non-equilibrium pure strategy  $n$ -tuple that is Pareto superior to it. Schotter 1981, p. 24.

<sup>260</sup> Vanberg 1994, p. 65.

<sup>261</sup> A point made by Mueller 2003, p. 9-10, and which has formed the basis of much study of the role of the state and rule enforcement in realigning incentives, as discussed below. In the form of the public choice tradition, this approach to state enforcement has also had a considerable influence on North, not least in bibliographic terms (e.g. North's regular publications in public choice oriented journals). See also Miller, Benjamin and North 2013.

is, in the classical liberal political economy of Hayek, North, and indeed Vanberg himself<sup>262</sup>, an exemplary case. For North, moreover, one cannot ultimately expect the decentralized free market to do the work of coordinating so as to make the ‘Smithian’ cooperative result possible if one does not have the proper institutional structure. This proper structure therefore needs to come about, not only despite coordination problems, but also despite prisoner’s dilemma type situations.

We can therefore read North as putting on the research agenda the integration of the two into a single account of how institutions and the state (or ‘elites’) mediate between both sets of problems. To do this, he must sustain a flexible and in a sense ‘pessimistic’ account of the limits to individual rationality and the ‘suboptimality’ or ‘inefficiency’ of many real-life historical orders, while simultaneously defending an account of market exchange in such a context that legitimates the assumption that, given the ‘right’ institutional order, in this kind of situation the problem will be mainly of the coordination type, and not of the prisoner’s dilemma type. Cooperation in the normative sense depends then on the evolution and ‘spontaneous’ emergence of the right norms and rules (as well as psychological states such as trust and the right beliefs), such that they become self-enforcing; where ‘right’ is defined as being favourable to the ‘Smithian’ process of mutual advantage through market exchange. The payoffs in the prisoner’s dilemma game will be changed so that the results are a coordination problem, one for which the Hayekian tradition of markets and spontaneous traditions has already established a solution. In short, in this tradition one can say the cooperative outcome first and foremost *is* the ‘Smithian result’.<sup>263</sup>

### **Game theory, the social contract, and the context of North’s NIEH**

Before moving on to the specifics of North’s arguments on these themes, such as his discussions of ideology and cognition, it is necessary to develop the argument a little further. This may take us slightly away from North’s work, but it helps if we can understand how the two traditions leading up to that oeuvre – the classical liberal tradition of Smith and the (often more pessimistic) social contract tradition of Hobbes, Locke, and Hume, both already united in some fashion in his inspiration, Hayek – are joined specifically in the postwar economic theories and ideas that North’s arguments respond to. I have mentioned previously public choice theory and the more neoinstitutionalist types of New Institutional Economics as the immediate context of North’s writing, and authors in these schools are the most common reference points in North’s own discussions, especially those more oriented towards economic theory rather than historical analysis. So to understand North better, and to grasp the meaning of his emphasis on ideology’s independent role or his reliance on terms like ‘players of the game’ when theorising the

---

<sup>262</sup> Vanberg 1994, p. 73.

<sup>263</sup> In his Nobel Prize lecture, North equates ‘cooperation’ with “anonymous, impersonal exchange” “capturing the gains from trade”. North 1993c, section V. Of course, I do not mean to suggest North wholly neglects the issue of public goods.

relationship between individuals and institutional orders, I think a bit more context is helpful for my subsequent argument.

Why games? Why should coordination and prisoner's dilemma problems matter so much? It is undeniable that North uses virtually no formalized game theory. And yet the language of game theory, and the emphasis on how difficult it is to achieve cooperation through rational behaviour, indeed comments on how difficult rationality as a concept is, are omnipresent in his work. As Lee Cronk explains, the kind of economic problems dealt with in the tradition of Hayek are the problems of spontaneous order, and this means the "largely unintended results of the actions of individuals".<sup>264</sup> Or as Carl Menger, not only a founder of marginalist economics but also a major figure in the study of institutions and spontaneous order in his own right, put it: "how can it be that institutions which serve the common welfare... come into being without a common will directed towards establishing them? (...) The unintended result of innumerable efforts of economic subjects pursuing individual interests".<sup>265</sup> Of course, the state, or some central planner, or some other formal authority can establish rules, but this would not answer the question how such authorities come into being themselves, nor how they can succeed in creating welfare enhancing institutions.

Underlying this question is the 'pessimistic anthropology' I have referred to before: an anthropology because, as Cronk has argued, the problem of relating individual values, beliefs, and reasons to the emergence of symbolic, economic, or institutional patterns is one that has immediate bearing on the domain of anthropology (even if anthropologists have not always recognized it in those terms); pessimistic, because North's contractarian approach departs from a pessimistic idea about human (bounded) rationality in a 'state of nature', and because of his interpretation of most historical societies as generally suboptimal, even dysfunctional, path dependent local equilibria, emerging from the interaction of such blinkered individuals. North is much more inclined than Menger or Smith to ask why cooperation does not (optimally) come about rather than why it does, and is pessimistic both in the sense of expecting improvements to be difficult and in the sense of relying on a pessimistic idea of human nature. The latter constitutes a mixture of rational self-interest constrained by ignorance and ideology and norm following behaviour in an unstable synthesis, as discussed further below.

So this is one of the main questions pursued in North's NIEH, often in a more pessimistic form than Menger's presentation: how it can be that such institutions often do *not* come about, or only suboptimally. And it is a question that has been extensively discussed in the theoretical schools or traditions in the immediate backdrop of North's work, such as public choice theory and various NIE theories. Since the question is ultimately concerned with the welfare effects, or payoffs, of the strategic (self-interested) action of individual agents in particular institutional settings, game theory has been one of the main tools used in exploring such problems. That the

---

<sup>264</sup> Cronk 1988, p. 282.

<sup>265</sup> Menger 1985, p. 146, 158.



problems of social contract theory are very similar – how can order come about from the state of nature? – and also representable as strategic action in particular institutional settings, including a Hobbesian setting where no prior institutions or conventions exist, makes attempts at unification of these problems via game theory tempting. Combined with some strong economic theory, such as the New Institutionalism, it permits exploration of all sorts of variants of ‘invisible hand’ explanation, including of the state itself.<sup>266</sup> I think North’s use of the terminology, and his explorations of rationality and beliefs should be seen in this light.

To understand how these questions then came to be formulated in the NIE paradigm, a comparison with the work of Andrew Schotter may be helpful. Next to North and Williamson, Schotter is arguably one of the most influential NIE theorists – be it in a decidedly neoinstitutional vein – and he has extensively relied on game theory to develop the NIE argument, particularly in his 1981 book *The Economic Theory of Social Institutions*.<sup>267</sup> Schotter’s approach is somewhat different from North’s: his definition of institutions revolves around social conventions as the *outcome* of rule-based action, not the rules themselves, and he sees institutions primarily as information conveying mechanisms in the style of Hayek, a claim North does not defend explicitly.<sup>268</sup> Moreover, for Schotter the primary assumption is a rationality assumption that North precisely abandoned at around the same time: the notion that institutions will come into being as solutions to market failures in a ‘functionalist’ way such that they can be fully explained as – possibly path dependent – outcomes of strategic self-interested action in such a situation of market failure. In Schotter’s approach, the individual is a “selfish, maximizing agent who is capable of coordinated social action”.<sup>269</sup> This leaves out some of the problems of the content of ideology and ‘worldviews’ that North later became concerned with, as discussed below.

But more than differences, their approaches have remarkable and for our purposes useful commonalities. Schotter views institutions as results of an evolutionary ‘survival of the fittest’ process that enables social order<sup>270</sup>; they are subject to path dependence<sup>271</sup>; they should be understood as a social contract emerging from strategic action in an (imagined) state of nature<sup>272</sup>; they are the product of spontaneous order or of political bargaining (though Schotter’s book concerns only the former)<sup>273</sup>; and finally, they are a rich and varied, indeed unpredictable although reverse engineerable, substitute for market exchange in cases where the price system is absent or fails<sup>274</sup>. Moreover, Schotter, like North, defends Simonian bounded rationality and its

---

<sup>266</sup> See Ullman-Margalit 1978. Cf. Ingraio and Israel 1990.

<sup>267</sup> Schotter 1981. Curiously, he does not cite Douglass North in this work, despite the strong parallels between their approaches and North’s already well-established reputation by that point.

<sup>268</sup> *Ibid.*, p. 11, 18.

<sup>269</sup> *Ibid.*, p. 5.

<sup>270</sup> *Ibid.*, p. 1, 2.

<sup>271</sup> *Ibid.*, p. 14, 34.

<sup>272</sup> *Ibid.*, p. 20-21.

<sup>273</sup> *Ibid.*, p. 28.

<sup>274</sup> *Ibid.*, p. 118.

implications of satisficing<sup>275</sup>, and similarly criticizes neoclassical economics as hopelessly inadequate because of its neglect of nonmarket institutions, myopically focusing on ‘perfect’ market equilibria alone.<sup>276</sup>

Because of these commonalities, despite North’s more historical focus and subtly different definition of institutions, I think Schotter’s work helps us understand more about the relationship between rationality, strategic ‘games’, and institutional orders in North’s work. Since North often seems to assume a theoretical background of NIE theory here that he does not develop at any great length, such as in his references to the ‘players of the game’ or how bounded rationality relates to self-interest and “the motivations of the players”, their interpretation is a matter of contextualizing. For this, a comparison with Schotter’s approach can give us a clue as to why North’s discussion of rationality, further described below, takes the shape it does, and what games have to do with it. This also helps in developing a more explicit formulation of the ‘pessimistic anthropology’ of the state of nature that underlies North’s idea of social orders.<sup>277</sup>

What Schotter contributes is the understanding that insofar institutions emerge as substitutes for the price system in cases of market failure, they need to be explicable in terms of the interests of the individuals that help them come about. The origins of institutions are therefore a question of strategic interaction, as described by game theory. For example, cooperative games, in which communication and bargaining are central, have been at the foundation of public choice theory and the analysis of what North would call ‘formal institutions’, which for this tradition (as for North) involves bargaining between elites and between elites and constituents.<sup>278</sup> North’s different definition of institutions is also important seen in game theoretical context: drawing on the work of Leonard Hurwicz and others, Schotter remarks that on this definition of institutions “each social institution can be considered a set of rules that specify or constrain the behaviour of agents in various social and economic situations. If these situations can be specified as games of strategy whose equilibria can be calculated, then... the proper institutional structure (rules of the game) can be determined that yields these outcomes as equilibria”. In other words, a game theoretical backdrop exists to using institutions as rules of the game, because such an approach permits a more *normative* comparative approach, as is indeed part of North’s work.<sup>279</sup>

However, Schotter’s view of institutions as outcomes of the strategic interaction process, as regularities so established, helps us understand how one can *noncooperatively* go from the ‘state of nature’ to the emergence of such rules in the first place: it addresses the origin problem that concerns North a great deal as well (see below and chapter 4). Game theory defines under what conditions social institutions could arise as the result of individual self-interested action in situations resembling coordination games or prisoner’s dilemmas (as well as other types of

---

<sup>275</sup> *Ibid.*, p. 148-149.

<sup>276</sup> *Ibid.*, p. 150-154.

<sup>277</sup> For other influential approaches to game theory and the social contract, cf. Binmore 1994, 1998.

<sup>278</sup> See Buchanan and Tullock 1962.

<sup>279</sup> Schotter 1981, p. 29.

games).<sup>280</sup> In short, the interaction between individuals to create institutions *as well as* the interactions between individuals and institutions once they exist (and either operate as parametric constraints or affect the payoff structure) can be seen as strategic problems for ‘players’ of a ‘game’, and this forms a basis for a New Institutionalist theoretical approach to comparative study of such institution formation and change. The emergence of spontaneous order is a game theoretical problem fundamental to New Institutional Economics and the Hayekian tradition, as Schotter’s work helps us see. It therefore makes sense to read North’s references to games in this context. A classic example is the institution of property rights (here in North’s terms), something North’s work has been much concerned with: all NIE theorists agree that their clear definition and enforcement is essential for achieving what North calls the ‘Smithian result’. But how they emerge (and with what characteristics) in the first place as the result of self-interested action is not explained in the Smithian framework, and this requires a Hobbesian ‘state of nature’ analysis that can be dealt with via game theory.<sup>281</sup> Spontaneous order is a problem of strategic nature, once self-interested individuals are assumed, as North at times does, and one that in North’s pessimistic interpretation frequently has the character of a prisoner’s dilemma type situation.

However, as discussed more in detail below, North does not always assume self-interest. Indeed, if he did, the prisoner’s dilemma could not be solved, because no *cooperative* solution to it exists except in the case of indefinitely repeated games. Although institutions have been interpreted as such in the literature, not least because of the enduring nature of many institutions despite the coming and going of the agents that, as it were, partake of them<sup>282</sup>, this does not necessarily solve the problem of the rational strategy agents would or should pursue in such situations. Indeed, in indefinitely repeated prisoner’s dilemma games, the rational strategy is indeterminate: although tit-for-tat cooperation is one possible equilibrium solution, there exist also much less cooperative ones.<sup>283</sup> This weakens the ability to explain the emergence of cooperation enhancing institutions in prisoner’s dilemma type situations, which are widely taken to be ubiquitous in collective action problems, as in the case of public goods.<sup>284</sup>

For this reason, and because of the experimental evidence that cooperation in experimental games takes place well above the levels predicted by self-interested strategic action, other solutions have been suggested. Put briefly, these tend to come in two types: either by defending a different (normative or explanatory) theory of rationality, or by suggesting ways in which norms, conventions, trust, and so forth can change the strategic situation. Both approaches are defensible from an explanatory viewpoint, although it can be argued that they involve abandoning the original prisoner’s dilemma game, because on some readings of rational choice theory they

---

<sup>280</sup> *Ibid.*, p. 30.

<sup>281</sup> Buchanan 1975, p. 9; Schotter 1981, p. 44.

<sup>282</sup> Something like this intuition may underlay North’s emphasis on the importance of perpetually lived organizations: North 2007; North, Wallis, and Weingast 2009.

<sup>283</sup> Ostrom 2005; Binmore 2015, p. 30.

<sup>284</sup> See, however, Northcott and Alexandrova 2015.

simply involve changing the payoff structure.<sup>285</sup> North, I believe, does all these to some extent. He looks for institutions that would enable cooperation where instrumentally self-interested action would not rationally permit it; he looks for the origins of institutions in ‘strong beliefs’, which are not probabilistic judgements but rather sources of such judgements and therefore of the explanation of choice behaviour; in other ways looks for the sources of the payoff distribution, sources that operate *prior* to the strategic situation itself<sup>286</sup>; and finally, by emphasizing the actual content of beliefs, attempts in a groping way to move away from the functionalist explanatory framework that underlies rational choice interpretation of institutions, as exemplified by Schotter’s work.<sup>287</sup> He also attempts to defend (in my view) various not obviously compatible interpretations of rationality, as the case demands. These I will now discuss in more detail, with the game theoretical and Hayekian context of the problems of spontaneous order, rationality, and cooperation in mind.

First, I will come to grips with North’s definition of rationality and how he uses the concept of bounded rationality to distinguish his approach from both neoclassical approaches and from those not partaking of the ‘economics approach’. Since rationality can be said to be bounded in various ways, it is important to appreciate both which bounds North considers relevant and how he justifies his use of the concept. Secondly, I will examine North’s approach to beliefs and ideology and how he relates these to the emergence and persistence of a particular institutional framework. In his more recent discussions of ideology, North has shifted from seeing it as part of the institutional framework to a source for it, and with this shift has come an increasing emphasis on particular models of human cognition and social learning. I will briefly outline what purpose these models have for North’s conception of ideology. Nonetheless, I will also point to the ongoing significance of conventional economic ideas of incentives and self-interest in North’s discussion, which creates a tension with the implications North claims for his own bounded rationality concept and its relation to economic theory.

Having established these points to define North’s initial New Institutional framework, I will then return to the themes of cooperation and the social contract as mediated through game theory. I will discuss North’s use of game theoretical concepts and language and discuss why and how he uses, or rather implies, insights derived from game theory. Given the particular modes of rationality conventional noncooperative game theory is designed to represent, this frequent rhetorical device – for little formalization of game theory appears in his work – sits uneasily with his other assertions about the nature of rationality and ideology and their role in individual agency.

---

<sup>285</sup> Such an argument (together with a rejection of the enormous significance ascribed to the prisoner’s dilemma) is defended in Binmore 2015 and Binmore 1994. Note that this does not mean that changing the game is *wrong*: just that it is a different problem.

<sup>286</sup> For an argument defending the necessity of explaining sources of preferences even in a (boundedly) rational choice framework, including norms that must exist prior to the situation analysis, see Hollis 1994, ch. 6,7, 9.

<sup>287</sup> *Ibid.*, p. 214.

## Why rationality is bounded for North

Next to the (methodologically related) rejection of the efficiency assumption, the emphasis on bounded rationality is for North arguably the main behavioural premise separating his approach from the neoclassical. The ‘rationality assumption’, by which North seems to mean the more traditional rational choice approach of orthodox economics, is for him not just an unnecessary premise but in fact a major methodological hindrance to advancing economic thought along the lines he has indicated. It prevents a proper consideration of ideology, of institutions, and of long-term evolutionary patterns, concepts that make up the bulk of North’s own historical theory. As he emphatically states in *UPEC*: “The rationality assumption has served economists (and other social scientists) well for a limited range of issues in micro theory but is a shortcoming in dealing with the issues central to this study. Indeed the uncritical acceptance of the rationality assumption is devastating for most of the major issues confronting social scientists and is a major stumbling block in the path of future progress. The rationality assumption is not wrong, but such an acceptance forecloses a deeper understanding of the decision-making process in confronting the uncertainties of the complex world we have created.”<sup>288</sup>

For North, there are two ways economic historians (and economists) should deal with this problem. Firstly, they need to take the cognitive processes of social learning into account, since these affect how humans form ideas about the world and respond to pervasive uncertainty. Secondly, the proper rationality concept is a bounded one, in which people do not live up to rational choice expectations of “substantive rationality”, but rather face imperfect information and imperfect feedback from the environment in interpreting the effects of actions. Because of this, the content of intentionality matters since understanding subjectivity becomes essential to understanding behaviour.<sup>289</sup> The boundedness of rationality appears for North both in limitations on information and in our limited cognitive capabilities: we face uncertainty *and* frequently act ‘irrationally’ in terms of the expectations of traditional economics models.<sup>290</sup>

The implications of the former are that informal and formal institutions are necessary to help us cope with our changing environment, and that such institutions may suffer from ‘lag’ relative to such changes and therefore have unintended negative effects down the line. The implications of the latter are that subjective models may often be wrong and that ideology, including political and religious beliefs, matter for understanding individual behaviour in institutional contexts. Indeed, North even suggests such beliefs, when hardened into ideologies, can be seen as imposing transaction costs on institutional change. Typically, for North one example of such subjective irrationality is the inability of many to accept ‘Smithian’ arguments for how to improve economic performance: “In a Coasian world the players would always choose that policy that maximized aggregate well-being with compensation for any losers; but

---

<sup>288</sup> North 2005, p. 5.

<sup>289</sup> *Ibid.*, p. 23-24, 64.

<sup>290</sup> *Ibid.*, p. 74, 156.

the real transaction costs are frequently prohibitive reflecting deep-seated beliefs and prejudices that translate into such prohibitive transaction costs. It is more than two hundred years since Adam Smith explained the underlying sources of the wealth of nations but the extent to which such views are embedded in the decision-making process of those shaping political/economic change is problematic.”<sup>291</sup>

In *UPEC* and *Violence and Social Orders*, North’s emphasis is understandably on explaining what theories of each he adopts for the purposes of his work, rather than elaborating on his critique of ‘substantive’ rational choice approaches. This therefore does not tell us that much about how he relates the implications of his bounded rationality concept to the existing economic literature on this subject, which would be helpful to clarify his methodological justifications. Fortunately, in a 1993 paper North undertook the latter in more detail. A number of important points stand out. Firstly, his primary justification for abandoning the standard rational choice approach is that such an approach is neither realistic nor robust: “A dissection of the rationality assumption is essential in order to incorporate much more ‘realistic’ assumptions to be derived from the diverse mental models that guide human decision-making... Taken over from economics was not only the scarcity and hence competition assumption - a robust assumption, but also the assumption of a frictionless a-institutional world and the assumption of expected utility theory, incorporated in the rationality postulate. Neither of these latter two assumptions is robust.”<sup>292</sup>

Why are they not robust? Because they are unrealistic in the sense that they cannot explain the existence and function of institutions. Since the function of institutions is to help people cope with uncertainty, assuming away such uncertainty leads to a failure to understand the dynamics of real historical economies. “In the real world the actors have in complete information and limited mental capacity by which to process that information. In consequence they develop regularized rules and norms to structure exchange. There is no implication that institutions are efficient in the sense of providing low cost transacting. Ideas, ideologies, myths, dogmas, and prejudices matter because they play a key role in making choices and transaction costs result in very imperfect markets or no markets at all.”<sup>293</sup>

We will not get there, North avers, by incorporating limited information itself into a rational choice framework, as in models of ‘rational ignorance’ and principal-agent problems. Such models cannot explain the fact that subjective models compete with each other in the form of ideologies, and since there is imperfect feedback from the environment on choices made, “there is a continuum of theories that agents can hold and act upon without ever encountering events which lead them to change their theories”.<sup>294</sup> This leads North to the second point: in

---

<sup>291</sup> *Ibid.*, p. 156. See also his adoption of Yujiro Hayami’s argument that a combination of Smithian and Confucian values lowered transaction costs in Japan from the Meiji period onwards: *Ibid.*, p. 57-58; Hayami and Aoki 1998.

<sup>292</sup> North 1993b, p. 159.

<sup>293</sup> *Ibid.*, p. 160.

<sup>294</sup> Frank Hahn cited in North 1993b, p. 159-160.

rational choice approaches, we already know everything worth knowing. Such approaches, North argues, are *right* for markets characterized by low transaction costs, but since that only describes very few societies (perhaps open access orders?), they are unlikely to be of much further use.<sup>295</sup> Institutional analysis can do the job instead, combined with theories from cognitive science, evolutionary approaches, and perhaps other social sciences too. “Simple problems, complete information, repetitive situations, and high motivation will produce conditions that lend themselves to rational choice models. As we move away from these conditions we must explore not only the immediate consequences in terms of choices but particularly the kinds of institutions that will evolve in such contexts to structure human interaction... How do we learn? Specifically, what set of circumstances lead us to change the mental models we possess and modify or alter the choices we make? Why do ideologies such as religions or political doctrines exist? They entail faith rather than reason and persist in the face of overwhelming contrary evidence. What makes some persist and others disappear? These are old questions, but cognitive science offers the promise of shedding new light upon them and in the course of doing so opening up new frontiers in the social sciences.”<sup>296</sup>

### **Ideology and social learning: North’s approach to beliefs and rationality**

Having established North’s justifications for rejecting the rational choice approach associated (by him) with neoclassical economics and public choice theory, let us now take a closer look at what answers he gives in his more recent work to those questions posed in 1993. It is clear that the scope of his theory has become vastly more interdisciplinary and requires more ‘external’ theory from social, cognitive, and biological sciences since he first set out to justify a historical and institutional approach to economic orthodoxy, and that this is in considerable part due to his changed approach to individual rationality and social context.<sup>297</sup> It is therefore no more than reasonable that North relies on the claims of particular specialists in fields like cognitive science, evolutionary anthropology, and the like to provide some microfoundations for his larger historical theory. Yet even so, he puts his particular stamp on this aspect of his work as well, both by his choice of ‘external’ theoretical premises and by the characteristic rule-based functionalism which guides his interpretation of the evidence.

North and colleagues repeatedly emphasize how difficult it is to get a grip on the workings of beliefs and ideology. Statements abound about how in their judgement the social and cognitive sciences have not yet come very far in understanding where beliefs come from and how they influence behaviour, and North therefore shows an appropriate recognition of the imprecise and contentious nature of his analysis.<sup>298</sup> In chapter 1, I explained how for North

---

<sup>295</sup> Cf. North 2005, p. 23.

<sup>296</sup> North 1993b, p. 161-162.

<sup>297</sup> In a piece written soon after the Polanyi paper, North could still simply assume that as far as rationality is concerned, all one needs to know is that “individual economic choices are determined by (1) opportunities and (2) tastes or preferences.” North 1978a, p. 963.

<sup>298</sup> E.g. North 2005, p. 16, 25, 28.

uncertainty gives rise to subjective beliefs as well as institutions (especially informal ones), where both function as heuristic devices in order to cope with the physical and human environment. For North, the trick to getting to the right institutional order from the viewpoint of economic performance is then to generate the right feedback loop between beliefs, institutions, and the ‘human environment’ in question so that the Smithian result obtains. As I explained there, North distinguishes various types of environment and various types of uncertainty; here I will not go into these further, but concentrate on the content and role assigned to ideology and learning in his larger functionalist approach.

Helpfully, North himself provides a list of methods for interpretation of behaviour that a cognitive science based model of rationality should provide. This list, already expressed in the language of the New Institutionalism, contains the following elements: “cognitive foundations focused on the dynamics of categories and mental models; heuristics as quite general processes for decision and learning; context-dependence, and, relatedly, social embeddedness of both interpretative models and decision rules; endogeneity of (possibly inconsistent) goals and preferences; organizations as behavioral entities in their own right; processes of learning, adaptation and discovery apt to (imperfectly) guide representations and behaviors also (or primarily?) in ever changing environments.”<sup>299</sup> We can see how far this takes North from the standard rationality approach of neoclassical economics, and how he emphasizes rationality operates as rules, as guides for the mind to least-cost ways of coping with complex and changing environments. But this does not tell us by what mechanism the content of these heuristics is determined.

Additionally, North suggests that learning takes place in a conformist way, so that “a common institutional/educational structure will result in shared beliefs and perceptions. A common cultural heritage, therefore, provides a means of reducing the divergent mental models that people in a society possess and constitutes the means for the intergenerational transfer of unifying perceptions.”<sup>300</sup> Ideology, indeed beliefs in general appear in this approach as a means to reduce uncertainty. For North, institutions, both formal and informal, can be seen as embodiments of such beliefs in the face of uncertainty, and while formal institutions can be changed more readily following political processes, informal institutions are slow and difficult to change consciously (this appears to be part of their respective definitions).<sup>301</sup> Moreover, since there are inequalities of power in many societies, there are differential abilities of agents (individuals, organisations) to undertake such institutional change, so that particular ideologies/beliefs will be more embodied in a lasting institutional form than others based on previous political constellations: “Thus the structure of an economic market reflects the beliefs of those in a position to make the rules of the game... (always with the caveat that their beliefs may be incorrect and produce unanticipated consequences). When conflicting beliefs exist, the

---

<sup>299</sup> *Ibid.*, p. 26.

<sup>300</sup> *Ibid.*, p. 27; ch. 5.

<sup>301</sup> *Ibid.*, p. 48-49.



institutions will reflect the beliefs of those (past as well as present) in a position to effect their choices”.<sup>302</sup>

This phenomenon, combined with the operation of social learning, creates an iterative process in which the ‘artifactual structure’ of a society (its accumulated institutionalized beliefs) contains descent with modification: modification that is, however, intentional as well as constrained by the power relations of past and present, as agents will prevent the change of such institutions as favour them. The result is the operation of path dependence: the shaping of institutions by the beliefs and power of agents in previous generations generates firstly the organisations of the present that benefit from the artifactual structure and secondly the individuals that have obtained the same belief systems through social learning. Both will act as a considerable brake on the change of existing artifactual structures, North suggests.<sup>303</sup> Path dependence plus the content of the artifactual structure then determine the divergent economic courses (and presumably also other social dynamics) of various societies in history.

It is clear then that ideology, as institutionalized in an artifactual framework, does a lot of work in North’s theory. It operates as a constraint on individual rationality in a given societal setting; it generates institutions, on which both levels of (relative) transaction costs and adaptive efficiency depends; and it creates path dependence, so that present opportunities are constrained by past decisions. Insofar as the ideologically constrained operation of individual rationality is the basis for institutions and organisations, and institutions and organisations in turn give rise to transaction costs, incentives, power relations, and all such supervening mechanisms, this trinity, the ‘scaffold’ of society as North calls it, in fact reduces to the first element. From the belief systems of individuals and the way they constrain how individuals cooperate rationally (or irrationally), all else follows. In this sense, North’s theory is ‘microfounded’. Indeed, North’s own summary of his discussion of the ‘scaffolds’ makes clear just how much depends on it: “The scaffolds humans erect not only define the economic and political game but also determine who will have access to the decision-making process. They further define the formal structure of incentives and disincentives that are a first approximation to the choice set. But the scaffold is much more. It is equally the informal structure of norms, conventions, and codes of conduct. And still beyond that it is the way the institutional structure acts upon and reacts to other factors that affect both the demographic characteristics of a society and changes in the stock of knowledge.”<sup>304</sup>

### **‘Players of the game’**

The general concept of the artifactual framework, or scaffold, clearly carries a lot of different analytical burdens. Postponing for the moment the question whether it can do so consistent with North’s different statements about rationality and belief, I now turn to North’s own observations

---

<sup>302</sup> *Ibid.*, p. 49-50.

<sup>303</sup> *Ibid.*, p. 50-52.

<sup>304</sup> *Ibid.*, p. 52.

about ‘players of the game’, the implicit references to game theory and how they express the problem of social cooperation as described in the beginning of this chapter. It is striking how the first functional role of the scaffold, even before the conventional choice theoretical claim about incentives or the normative-ideological role, is described as to “define the economic and political game”. That a game theoretical framework is present in the background of much of North’s thinking about the emergence of institutions, the emergence and persistence of state power, and indeed in describing the interaction of agents with institutions and with each other is not difficult to establish. On the very first page of *UPEC*, we are told that “the structure of constraints we impose to order that competition shapes the way the game is played”, and later that “all organized activity by humans entails a structure to define the way the game is played”.<sup>305</sup>

In fact, North is quite explicit that the whole mode of human interaction in a given social context is game-like in nature: “The game is played within a set of formal rules, informal norms... and the use of referees and umpires to enforce the rules and norms. How the game is actually played depends not only on the formal rules defining the incentive structure for the players and the strength of the informal norms but also on the effectiveness of enforcement of the rules.”<sup>306</sup> Elsewhere we find a definition of institutions as “the rules of the game”, and in *Violence and Social Orders* much of the discussion focuses on the distinction between zero-sum games (natural states) and positive sum games (open access orders).<sup>307</sup> On occasion, the game theoretical implications are made even more explicit, such as when North observes, in discussing institutions and their contribution to cooperative outcomes, that “effective institutions raise the benefits of cooperative solutions or the costs of defection, to use game theoretic terms”.<sup>308</sup>

For North and colleagues, such enforcement mechanisms build then on existing normative informal framework, which equally arises spontaneously out of the repeated interactions of individuals and their need for cooperation: “The spontaneous emergence of informal institutions is a process of innovation and imitation that takes place in a social group that is learning collectively. Individuals respecting conventions, following moral rules, and adopting social norms cause (as an unintended outcome of their action) the emergence of social order. In close-knit groups, informal institutions largely suffice to stabilize expectations and provide discipline, because the members of the group engage in personal relationships. In primitive societies, informal institutions alone can establish social order; and often there is no need for additional institutions with an explicit third-party enforcement mechanism.”<sup>309</sup>

It is therefore safe, I think, to say that for North institutions are analogous to the norms and conventions that various game theoretical approaches suggest as solutions to the cooperation problems mentioned in the introduction to this chapter. Equally, North’s argument for states and

---

<sup>305</sup> *Ibid.*, p. 1; 48.

<sup>306</sup> *Ibid.*, p. 48.

<sup>307</sup> North 1990, p.3-4; North, Wallis, and Weingast 2009, p. 133, 143.

<sup>308</sup> North 1990, p. 89.

<sup>309</sup> Mantzavinos, North and Shariq 2004, p. 77.

other norm enforcers as agents to change the payoff structure of such games so that rather than mutual defection, mutual cooperation becomes the (desirable) Nash equilibrium is a standard response to the implications of the prisoner's dilemma. Not coincidentally too, I think, is that this is also the conventional explanation or justification for the state in certain traditions of social contract theory, whether Hobbesian, Lockean, or even Humean.<sup>310</sup> In this regard, North's approach to the connection between these problems of rationality and cooperation in social and political life is conventional within certain traditions of economic and political theory, those belonging to what I have earlier described as the 'pessimistic anthropology' of his social contract theory. Indeed, given the statements about players of the game cited above, and the presentation of institutions as coordinators and of states/elites as enforcers in spontaneous order, plus finally North's ongoing concern with normatively interpreting these in terms of their enabling or prevention of achieving the spontaneous order of the 'Smithian' type, makes it difficult to interpret his work in any other way. North is not a game theorist, but he relies on concepts that have been specified in conventional game theory, and he does not hesitate to rhetorically invoke them. If conventional game theory is a Humean contractarian enterprise, as Shaun Hargreaves-Heap and Yanis Varoufakis allege, North is such a Humean *par exemple*.<sup>311</sup>

North and colleagues argue that reasoning in long-term economic historical change should move from the cognitive level, with its implications for rationality, to the institutional level and from there to the economic. The mechanism of interaction between these, giving rise to an artifactual framework, and its persistence through time creates path dependence. It is important to note that it does so by creating a particular payoff structure (often a low or negative one) in the 'game': "institutional path dependence may structure the economic game in a standardized way through time and lead societies to play a game that results in undesirable consequences."<sup>312</sup> As they write: "cognitive and institutional path dependence will ultimately lead to economic path dependence. The intuitively formulated proposition that "history matters" designates the importance of the phenomenon of path dependence, starting at the cognitive level, going through the institutional level, and culminating at the economic level."<sup>313</sup> Whatever then the constraints at the cognitive level, about which I will say more further below, they must be consistent with a game theoretical presentation of individual rationality in an institutional setting.

I will now examine what kinds of rationality must be assumed for conventional noncooperative games to lead to meaningful results, and then examine North's own varied statements about the types of rationality and their function in his explanatory framework. In the process, I will question to what extent this joining of concepts and implications derived from game theory and his contractarian approach to institutions is internally consistent with his

---

<sup>310</sup> Hargreaves-Heap and Varoufakis 1995, p. 148.

<sup>311</sup> *Ibid.*, p. 33ff. See also my comments on North's borrowing from the Scottish Enlightenment and the Spencerian-Hayekian classical liberal traditions above.

<sup>312</sup> Mantzavinos, North and Shariq 2004, p. 81.

<sup>313</sup> *Ibid.*, p. 81.

recognition of the role of ideology and his critical statements (cited earlier) about rational choice theory. Insofar as they are not consistent, we must ask then whether it would help North's overall ambitions to abandon this 'Hobbesian-Humean contractarianism' about formal and especially informal institutions.

### **Between social contract theory and economic rationality**

We have seen how coordination problems and prisoner's dilemma type problems, as analysed through game theory, suggest the need for institutional arrangements so as to overcome the implied difficulty of achieving welfare maximizing outcomes in collective action problems given self-interested rational choice. Indeed, such collective action problems have often been held to be (and have been) quite general in societies and to be a major problem central to the research agenda in the study of economic and political arrangements.<sup>314</sup> Indeed, Hume's approach to social contract theory already recognized free riding as a serious problem of collective action in situations requiring the cooperation of large numbers of agents. In essence, such formulations recognize, as Russell Hardin has pointed out, the risk of a fallacy of composition in a similar way as the prisoner's dilemma does: the interests of individuals may not line up with the interests of the group to which they belong, what he has called a "fallacious move between individual and group motivations and interests".<sup>315</sup> Indeed, this fallacy results from the assumption of self-interest and its dire consequences, in turn associated with Hobbes' social contract theory: "The modern view of the fallacy of composition in social choice is a product of the understanding of politics as self-interested... To some extent, therefore, one could credit Hobbes with the invention of social science and of explanatory, as opposed to hortatory, political theory."<sup>316</sup> But of course public goods often *do* come about, and many people do not free ride when they could. Therefore solutions to the analytical problem must exist.

What such solutions could be found? Put in a simplistic way, there are two ways of answering this problem. The first is what the conclusion of the Hobbesian approach, where state bodies or other external enforcers have the ability to change the payoff structure and turn prisoner's dilemma problems into coordination problems, either through e.g. taxes and subsidies or through the enforcement of formal institutions such as to raise the payoff of particular outcomes and lower those of others. The state itself can be taken as an instance of such a game, where rational actors will decide to allow a third party to undertake enforcement by distributing part of their payoff to it, such that this cost for each agent is lower than the marginal benefits obtained by the cooperative outcome to the dilemma.<sup>317</sup> As we have seen, North is inclined to explain the emergence of violence-reducing states or elite groups in a like manner, at least as a partway explanation. Since governments and such bodies reduce transaction costs by

---

<sup>314</sup> Olson 1965; Hardin 1982. See also Varoufakis 1991 and Hardin 2003 for the relationship of collective action problems, such as the free riding problem, to game theory and the prisoner's dilemma in particular.

<sup>315</sup> Hardin 2003.

<sup>316</sup> *Ibid.*

<sup>317</sup> Hargreaves-Heap and Varoufakis 1995, p. 148.

enforcement of formal institutions (rules, laws), they can increase the payoffs of cooperation (contracting) relative to defection (not contracting).<sup>318</sup> Indeed, the further back one goes in North's work, the more important this solution appears: North's earlier work, the state appears explicitly as a rationally chosen enforcement mechanism as a response to imperfect information and transaction costs.<sup>319</sup> In North and Thomas' model of manorial feudalism, for example, the importance of this approach for the 'efficiency view' can still be seen: in their book, the authors write that the feudal lord and his manor have "the functions of providing justice and protection" in high uncertainty environments, and that "the contractual arrangement of the classic manor can now be seen as an efficient arrangement for its day".<sup>320</sup>

North's de-emphasising of this solution is probably the joint result of his abandonment of the efficient contracting interpretation of institutions and of the realization that such a model does not provide much room for the role of informal institutions. Traditionally, the weakness of such solutions to the dilemma is precisely in explaining how the state itself then comes to exist as a stable equilibrium outcome of a strategic interaction. In North (and colleagues') later work much more attention is paid to this problem, but the issue remains that states or elites, and their transaction-cost reducing powers, are explained as the result of a bargaining equilibrium (i.e. a cooperative equilibrium) between the resistance of the 'citizens' and the predation of elites. Ultimately, North et al. can only solve this problem with reference to the problem of beliefs and values: only with shared values and legitimacy can such third party enforcement produce a credible mutually shared interest, i.e. a Nash equilibrium outcome.<sup>321</sup> The state is then not an efficient Hobbesian result, but rather a Lockean-Humean one: one where the consent of the governed and their cooperative agreement not to resist when the benefits outweigh the costs (including in less physical terms than money or safety) is necessary for the state to operate as a solution to the public goods problem.<sup>322</sup>

This points to the second solution: what one, following Hargreaves-Heap and Varoufakis, might call the Humean approach.<sup>323</sup> That is to say, the building up of 'custom and habit', as well as the influence of non-self interested rationality guided by moral considerations and sympathy, that guides choices, changing the game's characteristics or payoff structure. For example, one version of this approach might be where the prisoner's dilemma is seen to model the 'state of nature', but the emergence of normative institutions and conventions permit the building up of trust over repeated games and thereby change the dominant strategy, somewhat akin to 'tit for tat'. Indeed, valuations and beliefs, insofar as they affect choices, determine the payoffs given in

---

<sup>318</sup> North 1984.

<sup>319</sup> North 1979b.

<sup>320</sup> North and Thomas 1973, p. 32, 11.

<sup>321</sup> Contrast North 2005 with e.g. North and Weingast 1989, where incentives do all the work and beliefs play virtually no role in the argument.

<sup>322</sup> For North on the influence of Locke, see: Letter to David Nichols, May 23, 1991. For further discussion, see: Hough and Grier 2015, p. 50.

<sup>323</sup> Hargreaves-Heap and Varoufakis 1995, p. 102.

the game itself. This leads to the Hayekian view of conventions, alluded to above. Hayek's famous arguments about the 'road to serfdom' could be seen as having argued that due to pervasive uncertainty and imperfect information, the Hobbesian approach to solving the public goods (or free riding) problem will not succeed, but that the role of conventions and tradition allows the Humean approach to succeed instead.<sup>324</sup>

Much of North's more recent work can be seen as an expression of this idea, as can be taken from North's many examples and assertions of the importance of beliefs and of institutions in shaping people's behaviour. As we know, much of his criticism of neoclassical rational choice rests on his perception that by emphasizing efficiency models, they do not acknowledge the importance of beliefs or norms and conventions in regulating social life in a mutually beneficial way (albeit with the risk of path dependent suboptimal trajectories). The blending of self-interest and of convention, norm and belief in agents then generates the structure of strategic interaction between them. As we shall see, this view is not an inaccurate representation of most of North's observations on agents' individual rationality, but it poses serious philosophical problems. One is the problem of indeterminacy, which although seemingly technical, actually reveals the way appeals to 'players' of the 'game', with payoffs set by institutions, is a conceptual structure that cannot fully carry the weight North places on it. This in turn points to the conflicted and inconsistent way North has carried through his critique of self-regarding rational choice as the starting point of analysing individual agency. I will as briefly as possible expound these below, with this Northian context in mind.

Firstly, it would need to be explained where the conventions and norms themselves come from. It is difficult to argue these emerge by an analogous process of given norms and conventions that determine the payoff structure of a game without falling into an infinite regress. Indeed, this problem is somewhat analogous to the problem where in repeated prisoner's dilemma games with a known ending, backward induction of rationality means that the chain of strategies is equivalent to the choice in a one shot game; and in such a game the rational choice is to defect. One common answer to this is to resort to evolutionary explanations to provide an origin story for human conventions and cooperation, which we will discuss in chapter 4.

Related to this is the second and most important problem, which is the ambiguous status of rationality and its relation to values and beliefs in conventional game theory. This is because on standard interpretations of rational choice, it is necessary that the payoffs of the games already reflect one's preferences in a consistent way, so that the payoffs represent hypothetical rational choices. But this causes problems when in practice 'irrational' moves are made. One must either then conclude that in fact a different game is being played, because hypothetical rationality (in the sense of consistent utility maximization) is to be assumed by definition for any behaviour<sup>325</sup>; or one must concede the possibility of irrational action within a game setting. This indeed often

---

<sup>324</sup> Hayek 1944; Hayek 1960.

<sup>325</sup> Binmore 2007; Gul and Pesendorfer 2008.

happens in practice, insofar as error terms are a common representation of the possibility of accidentally selecting a suboptimal strategy in refinements of the Nash equilibrium concept.<sup>326</sup> But in such a case, the game can no longer be said to represent *the* hypothetical choices of an agent given in advance by the utility function of the individual.<sup>327</sup>

Moreover, many such games suffer from systematic indeterminacy with regard to the best rational strategy, which limits their applicability for the purposes of explanation and prediction alike. Only with the strong assumptions of the common knowledge of rationality and the common alignment of beliefs, which is to say that each agent has the same beliefs about the world and each agent knows all other agents to be rational and that they know that agent to be rational, can many of these games have a single equilibrium solution at all. The reason for this is the interrelationship between beliefs and desires, which for the purposes of instrumental rationality must be kept separate. One must have beliefs regarding the state of the world, including the payoffs of the actions of other agents, whether a known state or a probability distribution. Instrumental reason then acts to maximize your expected utility given these beliefs and the preferences that determine your payoffs. However, when beliefs actually influence preferences directly, and preferences in turn interact with beliefs, games become indeterminate. One classic example is the game where your payoffs and those of other agents depend on their beliefs whether you will be brave or cowardly, which in turn determines whether your rational action is actually to be brave or cowardly. If ‘you’ do not know what the belief of the other agents is, the game is indeterminate: one can have a probability distribution of hypothetical rational responses, but *no way of deciding which will actually emerge*.<sup>328</sup>

There exist refinements of the Nash equilibrium concept so as to account for risk, signalling, (Bayesian) learning, and uncertainty which to some extent mitigate this problem, but they do not solve it. There is simply no guarantee absent exogenously given specifications that a Nash equilibrium exists, and it is even not clear whether it is in a given case ‘rational’ to follow the strategy that obtains that equilibrium as long as there is no common alignment of beliefs. It is beliefs about others’ beliefs *and* preferences that generates indeterminacy, and therefore the separation of beliefs from payoff structures poses a major problem for any (instrumental) rational choice approach to spontaneous order that so easily invokes game theoretical concepts as North’s. The possible responses then are either to expand the concept of rationality beyond utility maximization; or to rely on evolutionary arguments as a justification of both rational choice itself and of the conventions it would give rise to; or to attempt somehow to reconcile beliefs and their effect on preferences with an instrumental interpretation of the function of institutions. North chooses the latter two. This creates, I believe, a tremendous tension in his approach to the rationality of agents, and even more so when combined with his treatment of beliefs and their role in his contractarian view of institutions. I will argue that his turn to

---

<sup>326</sup> E.g. Selten 1975.

<sup>327</sup> Reiss 2013, p. 63.

<sup>328</sup> E.g. in Geneakoplos, Pearce and Stacchetti 1989.

evolutionary explanations and his ‘cognitive turn’ are responses to this problem with his contractarian approach.

Before doing so, however, a last word on the status of conventional (or classical) game theory in the context of liberal contractarian thought may be useful. One of the reasons for the influence of conventional game theory, despite its limitations and indeterminacies, is that it further extends the technical tools for modelling individual behaviour by the twin criteria of utility maximization and the strict separation of structure and action.<sup>329</sup> These twin criteria impose on the effects of any structures on individual behaviour the nature of *constraints*: externally given limits on a choice set, given by rules, the actions of others, or even your belief structure (usually depicted probabilistically). But within those constraints, the agents are seen as maximisers of utility. Even if in principle such utility need not be monetary or even self-interested in nature, for such analysis to be practically useful their interests need at least be *self-regarding*, i.e. the pursuit of one’s *own* passions, whatever they may be. This, in turn, has often been how the contractarian position has interpreted the role of institutions. The Nash equilibrium concept is important for this tradition because it “can be seen as the only sustainable outcome of rational negotiations in the absence of externally enforceable agreements” (which would themselves need explanation).<sup>330</sup> Rational choice analysis, in the form of conventional game theory, therefore has an elective affinity with the contractarian quest to explain institutions out of rational action. It does so in two ways. Firstly, by explaining the origins of institutions out of the rational pursuit of self-regard (and often self-interest) by individuals with given beliefs and preferences; and secondly, by interpreting the persistence of institutions as constraints on individual choice, without affecting fundamentally the nature of self-regarding rationality within them. As Hargreaves-Heap and Varoufakis put it: “You want to win and the game simply constrains how you go about it.”<sup>331</sup>

So let us return to North once more. We have seen that North does not use formalized game theory much at all, but he does refer to concepts whose working out has been made explicit in the form of game theory. To the extent that, however, neither the origin nor persistence of institutions and the way ‘history matters’ for them can be explained in a realistic way in these terms, as the indeterminacies of instrumental rationality in game theory show, one must add more content to the model. That is to say, one requires specifications of the utility function, or a separate theory of the emergence of beliefs or preferences, or indeed a broader theory of rationality than instrumental self-regard. Yet each of these choices will affect the approach as a whole, and risks being *ad hoc* if the framework is otherwise left unaltered. This problem, inherent in rational choice analysis, can be seen directly in North’s ambiguous observations about the rationality of agents. On the one hand, North sees institutions frequently as ‘constraints’, claims agents pursue self-interest within institutional settings, and talks about

---

<sup>329</sup> Hargreaves-Heap and Varoufakis 1995, p. 30; cf. Giddens 1979 on ‘structuration’.

<sup>330</sup> Hargreaves-Heap et al. 1992, p. 102.

<sup>331</sup> Hargreaves-Heap and Varoufakis 1995, p. 30-33.



‘players’ of the ‘game’. On the other hand, he strongly criticizes rational choice analysis and sees it as one of the main flaws of neoclassical economics. And what role does experimental economics, with its tests of the implications of utility maximization in game theoretical settings, play in his work? Returning now to North, we will see in exploring these ambiguities that there is a permanent tension, not to say internal inconsistency, in his approach to rationality. This tension, I think, is the result of his desire to maintain the contractarian view of markets and as-if market institutions, while he simultaneously seeks to develop a much deeper view of human nature than rational choice allows.

### **North’s ambiguities on rational choice analysis**

On the one hand, we are told that ideology has the function of enabling social identification and thereby group formation<sup>332</sup>; that economic policy in the Soviet Union was path-dependently (co-)determined by Marxist ideological principles<sup>333</sup>; that ideology functions to reduce the costs of enforcement of informal institutions by encouraging conformism<sup>334</sup>; and we are repeatedly told with emphasis that strong belief systems matter because of the way they shape choices and lead to expectations about the world, such as the medieval world-view, and in a way not immediately amenable to confirmation or disconfirmation so that it can persist regardless of the economic or political choices made.<sup>335</sup> These claims suggest that the nature of rationality and irrationality itself is at issue, and that the incorporation of belief systems – including in their ‘incarnation’ as informal and formal institutions – should fundamentally alter the behavioural premises about the economic and political behaviour of agents.<sup>336</sup>

On the other hand we are told that institutions, which are expressions of beliefs or ideologies, “impose severe *constraints on the choice set* of entrepreneurs when they seek to innovate or modify institutions *in order to improve their economic or political positions.*”<sup>337</sup> Path dependence induced by institutional stability plays the same role of constraining the choice set, with “the nature of the limits to change that it imposes” left undefined.<sup>338</sup> These statements suggest rather something different, that self-interested (or at the least self-regarding) behaviour conforming to conventional norms of rationality as defined in economics remains the premise of understanding the actions of agents, but that their choice set is modified by ideological and normative concerns: an important specification of self-regarding utility maximization but not a

---

<sup>332</sup> North, Wallis, and Weingast 2009, p. 53.

<sup>333</sup> North 2005, p. 3.

<sup>334</sup> *Ibid.*, p. 42.

<sup>335</sup> *Ibid.*, p. 5, 16, 128.

<sup>336</sup> The argument that belief systems reduce transaction costs, already alluded to above, seems to me an ambiguous case. North does not suggest they do so intentionally, of course, but if they are adopted for that purpose it would require a second order rationality (which rational choice models do not normally permit), and if not, the appeal is to a functionalist explanation rather than one based on motivated choice.

<sup>337</sup> North 2005, p. 2.

<sup>338</sup> *Ibid.*, p. 52. A classic examination of such path dependence where the limits are based on transaction costs is: David 1985.

deviation from it. On yet another occasion in the same work North says that “the complex blend of ‘rational self-interested behavior’ (the foundation of economic models) with ideological beliefs stemming from the self-awareness of humans poses a major challenge to political scientists”.<sup>339</sup> This statement in turn suggests that some unspecified *combination* of (qualified or specified) rational choice and some other form of intentionality is used.

One may agree with North that rational self-interested behaviour remains the foundation of economic models, indeed defines what an economic model is as opposed to those of other sciences; but if so, it merely illustrates North’s difficulties in deciding when traditional economic models are relevant for him and when they are not. Clearly this “complex blend” North identifies poses a major challenge to New Institutional economic historians as well. The vagaries of rationality assumptions in his work are some of the clearest examples of how his NIEH approach threatens to fall into a methodological gap: insufficiently orthodox to use the methods - and therefore the justifications of those methods - of economics, yet insufficiently distant from the practice of economists to let go of the implications and theorems that economists have traditionally taken as the strong results of their discipline.

One example of this may be given in North’s discussion of results from economic experimentation, often taken to have sufficiently significantly affected the standard behavioural assumptions of rationality that it has given birth to a separate (and possibly rival) subdiscipline of behavioural economics. Noting the divergence from classic rational choice accounts, North summarizes the results from experimental economics as follows, in a statement significant enough to quote in its entirety: “People invoke reward/punishment strategies in a wide variety of small group interactive contexts. These strategies are generally inconsistent with, but more profitable than, the noncooperative strategies predicted by game theory. There is, however, consistency with the game theoretic folk theorem which asserts that repetition favors cooperation, although we observe a substantial use of reward/punishment strategies and some achievement of cooperative outcomes in single play games. *Non cooperative outcomes are favored, however, where it is very costly to coordinate outcomes, in large groups, and even in smaller groups under private information. In large groups interacting through markets using property rights and a medium of exchange, and with disperse private information, non-cooperative interaction supports the achievement of socially desirable outcomes. Experimental studies have long supported this fundamental theorem of markets.* This theorem does not generally fail, however, in small group interactions because people modify their strict self-interest behavior, using reward/punishment strategies that enable some approximation of surplus maximizing outcomes. Seen in the light of evolutionary psychology, such behavior is not a puzzle, but a natural product of our mental evolution and social adaptation.”<sup>340</sup>

---

<sup>339</sup> *Ibid.*, p. 56.

<sup>340</sup> *Ibid.*, p. 29. Emphasis added.

Two things are going on here. Firstly there is the appeal to game theoretical models of agent interaction put to the experimental test. Although North notes that the observed behaviour does not quite conform to what rational behaviour in such settings would lead one to expect, he suggests this is a consequence of small group interaction and low transaction costs. I will return to this point in my discussion of evolutionary themes in North. Here I want to draw attention to the second point, emphasized in the quotation: here he explicitly connects noncooperative rational behaviour with the achievement of ‘socially desirable outcomes’ as long as such is mediated by market exchange. This sounds very similar to the claim that the ‘Smithian result’ of (relatively) harmonious cooperation will result from the creation of free markets low in transaction costs and barriers to entry, discussed previously. On this occasion, the claim takes the form of a ‘fundamental theorem’ of economics resulting from the operation of individual rationality in strategic settings. Noncooperative rationality, that is to say self-regarding forms of strategic rationality, are now in larger settings sufficient to achieve such aims - as long as one gets the institutions right. North’s forays into realms of intentionality far removed from conventional economic premises have here clearly come to an end: we are back at the market contractarian view North (dubiously) associates with Smith, where it is simply a question of private vices, public benefits; except now with institutions to mediate the two. This serves as another illustration of the ‘two steps forward, one step back’ character of North’s attitude to conventional economic behavioural microfoundations.

The problem in my view is not straightforwardly that North has the wrong unit of agency (individuals instead of classes, say) but rather how North explains the behaviour of these agents – bounded rationality. I have now said enough, I trust, to describe the inconsistency in North’s use and rejection of this behavioural assumption. Far from the problem being solved by reference to cognitive science, as North himself seems to suggest, it is only magnified. No amount of cognitive science is in the foreseeable future likely to present us with a theory of sufficiently predictive precision to know where and how human choices (economic or otherwise) can be derived from initial conditions. It is precisely this fact that forces social scientists into nonpredictive forms of explanation, such as functionalist and interpretative explanation; rational choice theory, in the sense of the utility maximization postulate, is one example of the latter. Even a general prediction such as knowing “the direction of change” of institutions, predictively or retrodictively, will not be derivable from cognitive theory itself, and this is one of North’s ambitions for his NIEH.<sup>341</sup> This is clear from his own discussion, for if understanding cognitive processes of acquiring and storing information were enough, why would we need an analysis of beliefs and preferences, i.e. of intentionality, as he insists? The purpose of cognitive science for North is to learn how ideas emerge and why they do, but this cannot be related to historical observables except through institutions, beliefs, and preferences.

---

<sup>341</sup> E.g. North 2005, p. 99.

As North himself acknowledges, rationality postulates cannot be avoided.<sup>342</sup> Any theory of individual behaviour requires behavioural foundations. This raises therefore the question of what *justifications* he provides for his particular choices: understanding this may shed more light on the inconsistency identified above. However, in my view these justifications are just as inconsistent as the behavioural premises they are supposed to justify, and this strengthens my impression of the reason why: North wishes to reject neoclassical economic methods for economic history, but he wants to preserve the centrality of the ‘economics approach’ when it suits him. His critique of the former does not seem to affect his commitment to the latter. Put crudely: he wants to have his cake and eat it too.

North’s own (few) discussions of justifications are revealing of this ambivalence. His justification for his repeated reliance on conventional choice theory (rational choice) is its greater testability, rather than its likelihood of accurate description.<sup>343</sup> Yet simultaneously, he dedicates a great deal of his latest works to systematically rejecting rational choice assumptions, which he associates with neoclassical economics – on the grounds that it will lead to more accurate description. This indicates that at least some of the internal inconsistency in rationality concepts for North derives from his inconsistent use of justifications for those concepts. North wants his theory to be founded on conventional choice theory because of its (alleged) greater ability to produce testable propositions, but he also rejects strong rationality assumptions because they are unrealistic and unsuitable for historical analysis.<sup>344</sup> Equally, he accepts results from experimental economics insofar as they point to the apparent refutation of utility maximization, at least insofar as the expected utility is specified in monetary terms; but he also insists on the assumption that institutions are changed because of, and individuals and organisations act on, the pursuit of the agents’ expected advantage. Although cognitively constrained, the latter is still a form of utility maximization.

### **North on the cognitive sources of institutions**

This helps us understand the odd role that cognitive science plays in North’s account. We may here use Herbert Simon’s own distinction between procedural and substantive rationality. Traditionally neoclassical economists are uninterested in procedural rationality, i.e. the actual process through which choices are made, as long as the choices in question adhere to substantive rationality, i.e. the logical criteria of rational utility maximization.<sup>345</sup> (These criteria include transitivity, completeness, and monotonicity.) As Alexander Rosenberg points out, the usual justification in neoclassical economics for ignoring procedural rationality is that the substantive form of rational choice has sufficient explanatory power that further knowledge is irrelevant.<sup>346</sup>

---

<sup>342</sup> North 1990, p. 5.

<sup>343</sup> *Ibid.*, p. 5; North 1981, p. x, 4. See also: Mäki 1993, p. 6 for a critical note on the NIE notion of ‘testing’.

<sup>344</sup> It is worth noting that this is analogous to Polanyi’s view, who also accepted marginalist value theory as applicable to the modern ‘capitalist’ world, but *not* to premodern historical societies. See: Dale 2010, p. 103.

<sup>345</sup> Simon 1976.

<sup>346</sup> Rosenberg 1992, p. 116-117.

Indeed, in some interpretations the *point* of rational choice is to remove all consideration of what happens inside the mind: rational choice provides a set of axioms which are necessary for this kind of behaviourist analysis of human behaviour.<sup>347</sup>

In the case of the former justification, the obvious problem is that the rationality criteria have often been shown to be violated in practice, and moreover that whether explanatory value of rational choice in economics is sufficient to overcome such problems is, to say the least, controversial. It is against this justification that North appears to aim his criticisms. Yet on the latter justification, utility maximization is irrefutable because *true by definition*; any behaviour can always be redescribed as maximizing some property. The only relevant criterion, then, is whether this assumption leads to testable predictions and so to improvements in knowledge, which is again controversial.<sup>348</sup> From this perspective, North's emphasis on investigating the sources of procedural rationality (and its possible failure) only makes sense if he had already rejected the assumption that humans can be usefully modelled as if they conform to substantive rationality *and* the claim that they actually conform to it, at least sufficiently for theoretical purposes. North often seems to indicate that this is in fact his view. But if that is so, it is not clear how he can insist that testable predictions should be derivable from his theory and that the traditional justifications of microeconomic choice theory, or the language of game theoretical modelling, provide him with this.<sup>349</sup> The processes of learning and cognition are to give content to human intentionality, but the more content it has, the less any choice theoretical framework recognizable from conventional microeconomics will work, and therefore the less its implications or justifications are available to North. This is precisely why most orthodox economists reject any venture into cognitive science, neuroscience, or developmental biology.

Of course, it also fulfils another function for North: namely to account for individual choice under uncertainty. In standard economics, this is the job of expected utility theory. With actually observed behaviour (choices) and beliefs about probabilities held constant, expected utility theory can work backwards to obtain a preference structure that would explain that behaviour. Nonetheless, this is only a more general version of utility maximization and no more allows for refutation of the expected utility maximization assumption than does its more specific variant with perfect information.<sup>350</sup> Again, North seems to seek explanations in cognitive theory to account for the pervasiveness of uncertainty, attributing great importance to the historical shift that the greatest danger to survival is now attributable to uncertainty about other individuals rather than to uncertainty about the (physical) environment.<sup>351</sup>

---

<sup>347</sup> Gul and Pesendorfer, *op. cit*; Rosenberg 1992, p. 119-120.

<sup>348</sup> For some discussions, see: Hogarth and Reder 1986, 1987; Elster 1986; Rosenberg 1992; Hausman 1992.

<sup>349</sup> For a critique aimed at rational choice approaches to economic history that are nonetheless close to the NIEH (e.g. Avner Greif), see Elster 2000.

<sup>350</sup> Rosenberg 1992, p. 122.

<sup>351</sup> North 2005, p. 7.

But this use of cognitive science can at most explain how imperfect information came about; it does not (yet) provide us with sufficient knowledge to know, in given environmental conditions, what choices individuals will make. For this we are left once again with some theory of rationality. This, in turn, leads us back to the strong assumptions and their interpretation, according to which utility maximization is either wrong or by definition true, and in both cases of questionable use. Hence North's recourse to 'bounded rationality', a solution that, as we have seen, isn't one. This explains North's hesitation when introducing the chapters on cognitive science: "The rationality assumption is *not wrong*, but such an acceptance forecloses a deeper understanding of the decision-making process in confronting the uncertainties of the complex world we have created."<sup>352</sup> We are not told what kind of 'deeper understanding' this is, and whether it should affect the rationality assumption or not, the same one he has called "devastating for most of the major issues confronting social scientists and is a major stumbling block in the path of future progress".<sup>353</sup> He wants us to recognize the importance of uncertainty and its implication that we must study beliefs and institutions. But while cognitive science is tasked with giving us beliefs, nothing connects beliefs and preferences other than a final resort to bounded rationality. Bounded rationality in turn consists of an application of rational choice under strong constraints, which is why North calls institutions and beliefs constraints. But rational choice must take, holding tastes constant and for a given cognitive capacity, beliefs and preferences as given.

That these justifications are not compatible with each other does not appear to occur or matter to North. In each case, he accepts one or the other of these claims because they provide justifications that are convenient relative to other theories. The anti-rational choice perspective is justifiable for North when used against neoclassical and neoinstitutionalist 'efficiency' models, and the rational choice perspective is justifiable when it confirms the validity of the conclusions of conventional microeconomics. These conclusions are in particular the explanatory power of assuming utility maximization of some sort and its contribution to testability and prediction; as well as the normative desirability of 'free and efficient' markets as modes of social cooperation, and that such markets are a necessary but - given the problem of getting the institutions right - not sufficient condition for achieving lasting economic growth and/or 'adaptive efficiency'.

Another reason why North might put so much stress on human cognitive processes is because of the role of beliefs in his most mature NIEH approach. Precisely in contradistinction to the 'unrealistic' rational choice views he attributes to the neoclassical economists, for North beliefs must play an independent causal role in history. This has several implications for him. Firstly, the independent role of intentionality precludes a simplistic evolutionary account, by which he mainly refers to memetics: cultural evolution is decidedly different from 'Darwinian evolutionary theory' because of this.<sup>354</sup> I postpone discussion of this point to the next chapter.

---

<sup>352</sup> *Ibid.*, p. 5.

<sup>353</sup> *Ibid.*, p. 5.

<sup>354</sup> *Ibid.*, p. viii.

The second role of the independent causality of beliefs is more important: beliefs are shapers of the preferences of individuals *and* constitute the stock of knowledge, i.e. the technologies for institutional change available to a society. In this way, they are both constitutive of the constraints on individual action and of the aims of that action, on North's account. Since beliefs depend on learning processes, cognitive theory (North seems to hope) will help us elucidate the process of the formation and inheritance of culture: "Culture not only determines societal performance at a moment of time but, through the way in which its scaffolding constrains the players, contributes to the process of change through time. The focus of our attention, therefore, must be on human learning—on what is learned and how it is shared among the members of a society and on the incremental process by which the beliefs and preferences change, and on the way in which they shape the performance of economies through time."<sup>355</sup>

This dual role of beliefs, however, puts North in a difficult position. As we have seen, standard approaches to strategic interaction between rational self-regarding individuals have difficulty dealing with beliefs in this double role, due to the strict difference between structure and action. And yet, as North (surely rightly) puts it, "beliefs... are both a positive model of the way the system works and a normative model of how it should work".<sup>356</sup> Where North initially had beliefs be part of the institutional structure, they are now functionally equivalent to institutions (as additional constraints) *and* sources of institutions (as an important part of the normative structure of society) *and* sources of the preferences agents pursue within the institutional framework. North is not to be accused of lack of profundity here: the difficulties he grapples with are surely the right ones, and the relationship between ideology, rationality, and social structure is at the heart of debates in social and historical sciences for more than a century. The problem for North is rather that he attributes these complex roles to beliefs while simultaneously, at other times, assuming or suggesting a much more simple (and more tractable) agency based on self-interest and/or instrumental rationality. He does not, in other words, correspondingly broaden his rationality concept in a consistent manner, and especially not to go beyond the self-regard (though not self-interest) of the individual.

His bounded rationality is, insofar as he maintains this approach, a form of rational choice, but one adjusted for greater realism of assumptions. The restriction on rational choice is then, following Herbert Simon's pioneering work on 'satisficing', a question of cognitive limitation, but this does not imply a rejection of the fundamental instrumentalism of rational choice or its foundation in a self-regarding agent.<sup>357</sup> Yet we have equally established that this is not always true for North. Philip Pettit points out the problem for rational choice theory, well-established in behavioural economic theory at this point, that "the mind postulated in rational-choice theory is that of a relatively self-regarding creature. But the mind that people display towards one another

---

<sup>355</sup> *Ibid.*, p. viii.

<sup>356</sup> *Ibid.*, p. 2.

<sup>357</sup> North 1990, p. 23; Simon 1959, 1972; See also Lilly 1994 for a cognitive interpretation of Simon's satisficing principle.

in most social settings... is saturated with concerns that dramatically transcend the boundaries of the self.”<sup>358</sup> Indeed, institutions and shared normative frameworks, ideologies and beliefs, whole cultures, practices, and habits would appear to have this characteristic, and here North is keen to include these as essential modifications of rationality for understanding actual historical events.

This ambiguity has been noted by other critics as well. In two articles on North’s rationality concept and its status in North’s work, Ben Fine and Dimitris Milonakis have criticized his approach on just this point, since the increasing importance of beliefs is something that sets him apart even within the NIE tradition. As they observe: “The new and newer economic histories can in principle explain anything on the basis of rationality by pushing out the boundaries of the endogenous/rational, one by extending the scope of the as-if-perfect market, the other by appeal to market imperfections. By contrast, North has retained a continuing commitment to the role of ideology which occupies an uncomfortable position in straddling the rational/irrational divide.”<sup>359</sup>

They rightly observe that North’s theoretical trajectory has seen him shift away increasingly from neoclassical approaches (in North’s own terms) towards a theory of norms and institutions interacting with the objective dynamic of transaction costs. Yet, building on Daniel Ankarloo’s critique, they argue that in his trajectory from Cliometrics to the transaction cost approach he has moved from (as the paper title suggests) ‘the principle of pricing to the pricing of principle’. As they argue: “there is one essential mechanism that straddles all of the analytical levels— *individual pursuit of self-interest with an as-if market mechanism prevailing throughout society*. This is already explicit in the gains from behaviors such as cheating, but for institutional impediments. But it penetrates everywhere, even to the determinants of ideology, with every principle having its price”.<sup>360</sup> The result is, for them, that “*ideology* becomes a truly marvelous and powerful explanatory factor. For it encompasses whatever is not explained by rationality (which has itself been extended by appeal to market imperfections). This is a point of departure from the newer approach, certainly as promulgated by economists, for whom only rationality exists, and appeal to any other form of behavior is arbitrary and unacceptable... There must not only be incentive but conception”.<sup>361</sup>

This is an interesting observation that dovetails with my objections to North’s inconsistent or unclear use of bounded rationality and its lack of consideration for the meaning and content of intentionality, *despite* North’s frequent appeals to the importance of such considerations. Fine and Milonakis put it as follows: “Otherwise, with frequent reference to the work of Herbert Simon, and others working on (ir)rationality from the perspective of the individual’s cognition and system of beliefs, North is most readily interpreted as always reconstructing social theories of ideology through the prism of methodological individualism, of agents making sense of their

---

<sup>358</sup> Pettit 2002, p. 238.

<sup>359</sup> Fine and Milonakis 2003, p. 553. The present discussion is in part an extension of the argument of that paper.

<sup>360</sup> *Ibid.*, p. 558.

<sup>361</sup> *Ibid.*, p. 559-560.



external environment by interacting with, and internally contemplating it... As it were, ideology, etcetera, is complicated; we do not have a complete theory; so the focus will be on one effect alone.”<sup>362</sup> Although not based on a close reading such as I have attempted here, Fine and Milonakis rightly observe that on certain occasions, North is liable to see individual gain or self-interest as the determinant of ideological ‘choices’, however much he on other occasions emphasizes the value orientation inherent in belief systems. The connection between the two is best exemplified by a conclusion to *IICEP*, which they cite: “The agent of change is the individual entrepreneur responding to the incentives embodied in the institutional framework. The sources of change are the changing relative prices or preferences. The process of change is overwhelmingly an incremental one.”<sup>363</sup>

Fine and Milonakis’ critiques have certain limitations, from my perspective. Like much of the secondary literature on North, the timing of their works means they concentrate their discussion on his works of the 1980s and 1990s. Perforce, they did not have access to *UPEC* or *Violence and Social Orders*, which show developments such critiques should address: North, Wallis, and Weingast have now supplied North’s NIEH with a model of power and conflict as well as with a story of the origins of ideology and sociality, while *UPEC* demonstrates a deepening of the ‘cognitive turn’ and a frequently more subtle and sophisticated restatement of the NIEH framework proposed in what I have for that reason called the ‘second stage’.<sup>364</sup> It is not a coincidence that North begins the discussion in *UPEC* by asking “why do rules, norms, conventions, and ways of doing things exist? What induces the mind to structure human interaction in this way?” and that for him the NIE and its theory only “takes us partway”.<sup>365</sup> In this sense then I do not think so much that North wants to concentrate on one effect alone - at least no longer - as that, faced with the contradictions involved in his shifting conceptions of rationality, he tends to fall back on received wisdom from more conventional economic theories, neoinstitutional or neoclassical. As with his approach to markets, the logic appears to be that ‘when in doubt, assume rational choice.’ His *ambition* is clearly the contrary, as most pronounced in *UPEC*, where the emphasis is on a broad conception of rationality, the incorporation of the ‘irrational’, and the need for interdisciplinarity in understanding these problems. It is rather that North’s practice lags behind his theoretical aims.

### **Conclusion: Why would North hold an inconsistent theory?**

James Buchanan, one of the main theorists of public choice, once usefully distinguished ‘orthodox economics’ and ‘constitutional economics’. For him, orthodox economics is about explaining in a Walrasian or Marshallian framework the rational interactions of agents. Constitutional economics, on the other hand, “attempts to explain the working properties of

---

<sup>362</sup> *Ibid.*, p. 562-563.

<sup>363</sup> North 1990, p. 83.

<sup>364</sup> E.g., Milonakis and Fine 2007 is largely concerned with critiquing *Structure and Change*, a work by then mostly rejected in North’s own presentation.

<sup>365</sup> North 2005, p. 11.

alternative sets of legal-institutional-political rules that constrain the choices and activities of economic agents, the rules that define the framework within which the ordinary choices of economic and political agents are made... it examines *the choice of constraints* as opposed to the *choice within constraints*.<sup>366</sup> North's intentions in the later stages of his NIEH are clearly in the former category, a species of constitutional economics in this sense. More specifically, I have suggested that the ambiguities found in North make most sense if they are seen as a conflict between on the one hand his liberal contractarian views of social evolution and his attachment to the concepts of Humean instrumental rationality that underlie them, fitting the constitutional economics tradition of the NIE; and on the other hand his ambitions to achieve a more complete and realistic economic history in which the world outside private contracting, including its 'irrational' components like belief structures and its imperfect markets and information, play the large role that they deserve.

Before concluding, let me expand a little on this. As many critics of mainstream economic thought have observed, in its historical origins as well as its methodological practices there is an intimate connection between utility maximizing general equilibrium models and the political philosophy of the market as social contract. North's 'Smithian result', as discussed in the previous chapter, is not just a series of claims about the harmonizing role of the market, but in some respects at least the neoclassical tradition can also be seen as a formalization of this political idea.<sup>367</sup> For Alexander Rosenberg, general equilibrium theory (which is in turn based on utility maximization) serves the 'Smithian result' paradigm in that it "is the formalized approach to the systematic study of this claim about how the unintended consequences of uncoordinated selfishness result in the most efficient exploitation of scarce resources in the satisfaction of wants."<sup>368</sup>

The tradition of constitutional economics then becomes the search for 'incentive compatible' institutional arrangements that best fit this allegedly 'Smithian' perspective. Orthodox microeconomics is then *ipso facto* justified by serving this aim. As Rosenberg puts it, "at a minimum the contractual arrangements justified must be compatible with the possibility of the operation of rational self-interest and the theory from which the arrangements are derived should enable us to predict what the generic consequences of self-interest will be... Adam Smith's view was that the market provided an incentive-compatible mechanism."<sup>369</sup> One might add that in this context game theory might appear a useful method to appeal to precisely because in strategic rationality the parameters of this minimum condition are explored, and moreover, the famous existence proof of general equilibrium under highly restrictive conditions that underpins neoclassical theory is also a Nash equilibrium.<sup>370</sup>

---

<sup>366</sup> Buchanan 1989, p. 64. I draw for this discussion further on Rosenberg 1992, 1994.

<sup>367</sup> Indeed it has been defended as such: Arrow 1974. Cf. Hodgson 1994a for a critical perspective.

<sup>368</sup> Rosenberg 1992, p. 219.

<sup>369</sup> *Ibid.*, p. 225.

<sup>370</sup> For a critical evaluation of this line of reasoning, see: Mirowski 2002, p. 370ff.

Ultimately, therefore, in this chapter - as in this work as a whole - I try to show that where North is closest to mainstream economic concepts of rationality, albeit in bounded form, such as where he relies most on incentives, self-interest, or game theory, his interest in the origins of ideas and the evolution of cognition fall by the wayside. On such occasions, they are casually replaced by more market-centric, normative and contractarian ideas. His constitutional economics is, seen from this perspective, a branch off the main tree of economic thought. North clings to this branch despite his explicit ambitions to leave the tree and explore the wider forest.

If my argument is correct, North's rationality concept is inconsistent both in its definition and in the justifications offered for it. On the one hand, for North bounded rationality promises to liberate his NIEH from the unrealistic strictures of neoclassical (fully) rational choice models. On the other hand, his rationale for doing so extends much further, suggesting that postulates of utility maximization itself may not be useful and that substantive rationality is irrelevant for the study of economic behaviour in premodern cases. If so, there is no obvious way to reconcile this with the deployment of an unspecified concept of bounded rationality, of which we only know that beliefs are further specified as constraints on the choice set than is commonly done in mainstream economics. North emphasizes the importance of cognitive theory for giving us a better model of individual rationality and therewith explanation of behaviour, and even justifies the NIEH as a whole by its predictive capacities and testability. Yet cognitive theory does not in his work replace the operation of the usual extensions of self-regarding individuals acting under some formalization of folk psychology, as when they are 'players' of the 'game' acting to 'line their pockets' or to change institutional arrangements in their own interests. Incentive compatibility remains the main problem, rather than a substantively different view of human individual rationality that would give flesh to the centrality of ideas and ideology to his vision. In other words, North simply does not seem able to choose, and wants to have both the critique of (mainstream versions of) economic rationality where it suits him and simultaneously fall back on just those types of rational action where necessary.

His Janus-like response on the first occasion in which he discusses experimental results in economics, of which I have observed that he saw this as supporting simultaneously his rejection of neoclassical economics and confirming his belief in the 'fundamental market theorem', I therefore interpret as follows: *it reveals more strongly North's commitment to the idea that 'the Smithian result' follows from the 'correct' political economy, i.e. the right combination of institutions, markets, and property rights, rather than to any particular rationality postulate.* To paraphrase Deng Xiaoping: seemingly for North it does not matter whether a cat is rational or irrational, as long as it catches mice.<sup>371</sup> The 'mice' are then the normative implications. To achieve the open access order, institutions must act as ways to reduce transaction costs and in

---

<sup>371</sup> From the viewpoint of rhetorical analysis, it is interesting to note the frequency of economic metaphors in North's citation of Antonio Damasio's work on the neurobiology of the self, which the former enthusiastically treats as illustrating his point about how cognitive limitations imply bounded rationality: North 2005, p. 37. ("The price of knowing risk, danger, and pain... The price of knowing what pleasure is... The coin in a Faustian bargain".)

this way economic and political life must be made as much like a neoclassical model of a competitive market as possible.

North's inconsistency derives to my mind therefore, at least partially, from the motivated cognition of his own research. By having set out in advance that his theory must confirm to certain favoured conclusions and theorems of conventional economics, he repeatedly introduces justifications that are inconsistent with the ambitions and methods he himself defends for reaching those conclusions. It is as if he says to his neoclassical colleagues: I will arrive at the same findings as you, and I will justify them in the same terms as you do, but I will reach them by means of methods and concepts diametrically opposed to yours. Such an approach is bound to lead to inconsistencies between the justifications for the theory and the theory actually deployed.

This is illustrated by the persistent issue in North's bounded rationality model: that we are not told in what way it is bounded, beyond the observation that ideology and mental models matter to understand intentionality. However true this may be, and however valuable a critical perspective on the neoclassical rational choice approach, it is not a sufficient departure from rational choice. There is an important distinction, conceptually and methodologically, between rational choice bounded by beliefs and norms, and an extension of the model of instrumental choice itself. Here we are confronted once more with the recurring problem in North – already encountered in the discussion of whether premodern economies were meaningfully 'non-market' or not – that what North says he wants to do with his theory at the level of methodology and philosophy of social science is one thing, and what he actually proposes at the more concrete level is another.

However, this leaves one aspect unexplored. For North is not so Panglossian as all this may make him sound. He rightly insists that cooperation is difficult to achieve in actual societies, even if cooperation is conceived as the 'Smithian result'. In *Violence and Social Orders* and elsewhere, he stresses the pervasiveness of violence and disorder. For him, institutional improvements that last are hugely difficult to achieve given the embeddedness of institutions in existing incentive structures and vice versa. Indeed, the (vast) game theoretical literature on the prisoner's dilemma would suggest this, and when North refers to 'players' of the 'game' and the need for states and institutions to solve coordination problems, it is clear that the implications of this literature are also present in his thought. It has long been known that under conditions of *bounded* rationality and with certain minimum assumptions of self-regard, the provision of public goods will not be established to an optimal level by a fully competitive market in the absence of some extraneous force.

Williamson's 'markets versus hierarchies' perspective is a recognition of this, as is indeed much of the New Institutional literature. It should not be surprising then that for North cooperation remains a central problem within his market contractarian framework. Often, economists have argued that state coercion should play the role of enforcing provision of public goods; at other times, the ability of individuals to contract given the right informational

incentives has played a key role. North's question where the institutional underpinnings of all this come from, how the normative structures and indeed the human orientation towards sociality and political commitment arise in the first place, should in my view be seen as a major step in research beyond these rather simple solutions. The contractarian perspective, the reliance on explanations based on strategic rationality, and the emphasis on institutionally supported market solutions do not disappear with this question. But insofar as the prisoner's dilemma makes their easy achievement through individual self-interest unlikely, North has to answer it by introducing a different dynamic to link all the levels of his theory together: cultural evolution. To this I will turn in the next chapter.

## Chapter 4: North's theory of cultural evolution

### Introduction: towards adaptive efficiency

In the previous chapter I have argued that North's work, not least in the most recent phase of his theoretical development, is characterized by a fundamental ambiguity or contradiction in terms of the rationality of agents. I have examined this through the lens of his game theoretical framing of economic historical agents as 'players' of a 'game', as well as through the lens of the relation between individual rationality and collective cooperation, especially in the form of market exchange. The latter, I have argued, remains for North the primary *desideratum*, and his often 'orthodox' views on the intrinsic merits of free competition in economics and politics thereby exist in some degree of tension with his critiques of rational choice and other methodological niceties that have hitherto (often) served as foundation for justifying those same views in mainstream economics.

This is not to reduce North's work to a simplistic *laissez faire* analysis. In North's work the developmental economics implications revolve around the *difficulties* of achieving the 'correct' institutional framework, including frameworks of belief, that would allow an 'open access order' to function as it should. But it does mean that there exists for him some combination of the 'correct' underlying institutional order with (more conventionally) a regime of strong property rights and free markets, whose necessity and desirability is given as the 'Smithian result' (discussed in chapter 2). I have argued in the previous chapter that North lacks a convincing account of the interaction between agents and institutions, of the origins of institutions themselves, and of how these result in determinate forms of behaviour. Without such an account, North must rely on the contractarian tradition and its examination of when individual self-regard does and does not allow for the collectively harmonious results the 'Smithian result' implies.

In contemporary economics, conventional game theory, especially the many explorations of the prisoner's dilemma, offer the most promising route for examining these problems. Unfortunately, as I have argued, the methodological limits of game theory often lead to indeterminacy of its results. To some extent this is due to the restrictive rationality assumptions required to achieve a (subgame) Nash equilibrium in many conventional games. Moreover, many explorations of the prisoners' dilemma and its variants demonstrate that lasting cooperation is only likely to be achievable under those very restrictive conditions (for example via tit-for-tat strategies). An obvious approach would be an extension of the rationality concept of individual agents to a more realistic and broad conception of rationality, one that does not rest solely on 'common alignment of beliefs' and perfectly informed self-regard. But as Shaun Hargreaves-Heap and Yanis Varoufakis show, such extensions of rationality only make conventional game theory less determinate and applicable.

Indeed North himself in his various critiques of neoclassical economists seems to insist on such broadening of rationality, on the importance of beliefs and ideologies beyond mere

Bayesian probability, and on the limitations of the perfect information and perfect competition tradition. Yet at other times, he is equally keen to emphasize self-interest as the guiding motive of agents in interaction with institutions ‘at the margin’ and his game theoretical rhetoric serves to underpin his contractarian presentation of the function of institutions. This ambiguity cannot be easily solved, because it rests on his desire to preserve what I suspect are mutually incompatible things. On the one hand North wants to preserve the ‘Smithian result’ as the idealized outcome of free societies, implied by the tradition of the ‘economics approach’ and its contractarian heritage, and on the other hand he insists on the necessity of accepting Polanyi’s challenge, with its implications of taking the different institutional and belief frameworks of different historical societies as fundamental to the analysis of their economies. These may appear at first glance to be compatible, but they founder on the rocks of inconsistent rationality concepts and inconsistent approaches to the market, as the last two chapters have argued. For this reason, I suggest he must choose between critiquing rational choice and conventional ideas of self-regarding agency and maintaining the idea that the ‘economics approach’ will, via New Institutional Economics, safeguard the ‘Smithian result’ and its contractarian political economic heritage.

However, in having said this North’s theoretical framework is not yet fully explored. For I have postponed a discussion of evolutionary game theory, experimental economics, and their implications to the present chapter. This is because North has yet another mechanism up his sleeve to make ‘history and institutions matter’: namely his theory of cultural evolution, by which means he moves from the rationality of individual agents to the macro-historical dynamics of institutional change he ultimately wants to describe. This mechanism, then, joins together the micro-level described in chapter 3 with the meso-level markets-and-hierarchies dynamic described in chapter 2, and adds to this a transhistorical macro-level dynamic. This last dynamic is sociocultural evolution and its effect on ‘economic performance’, an effect North describes with the term *adaptive efficiency*.<sup>372</sup>

Fitting what one might call the ‘turn to biology’ (rather than physics) as source of inspiration for the social sciences, and with North’s stated admiration for Hayek’s theory of cultural evolution and its importance for economics, North seeks to do with evolutionary arguments what his indeterminate choice theoretical approach cannot: explain the origins *and* function of institutions and beliefs in terms of their effect on efficiency of economic cooperation.<sup>373</sup> As North describes it, “put simply the richer the artifactual structure, the more likely are we to confront novel problems successfully. That is what is meant by adaptive efficiency; creating the necessary artifactual structure is an essential goal of economic policy.”<sup>374</sup> The natural mechanism guaranteeing adaptive efficiency, it turns out, is essentially a classically liberal society: a

---

<sup>372</sup> North 2005, p. 4.

<sup>373</sup> *Ibid.*, p. 51; North 1999.

<sup>374</sup> North 2005, p. 70.

legitimate government, with limited powers, strong property and individual rights, and credible commitments against expropriation by government.<sup>375</sup>

All these in turn must be underpinned by strong social norms that enable this structure to function, and this causes the difficulties policymakers face in simply proposing (as in e.g. the ‘Washington consensus’) that poor or developing countries transform their political and economic structure into a liberal one. Oftentimes, North argues, the ideological commitment is missing due to long periods of ‘disorder’, and then the results will disappoint: “Because this cultural conditioning of a society usually takes place over generations it is fundamentally difficult to establish stable consensual order in societies that have experienced persistent disorder. In such cases authoritarian order may very well be preferred by the members of that society.”<sup>376</sup>

In *Violence and Social Orders*, the link between adaptive efficiency, competition, and cultural evolution is clarified. Open access orders, the ideal types (so to speak) of the classical liberal polity defined above, have the greatest degree of adaptive efficiency because they have the greatest degree of *market competition* in them, in both the economic and political markets (as one will recall from chapter 1). This, in turn, has two beneficial effects for such adaptivity. Firstly, competition enables more ways of solving problems to be proposed and tested, which strengthens a society’s flexibility and therefore survivability in the face of a changing environment. Secondly, political competition means governments can offer more credible commitments in negotiation, which in turn allows for more conflict resolution without internal violence or ‘disorder’. In addition to these core benefits, North and colleagues also mention the beneficial economic effects of market competition, by allowing more market goods as well as more public goods to arise; the latter (perhaps a surprising claim) is seen as enabled by the impersonal nature of liberal government.<sup>377</sup> All this creates, via market competition in the liberal and economic spheres and via liberal beliefs underpinning their legitimacy, a “virtuous circle of open access and competition”.<sup>378</sup>

I will not at this stage go further into the political normative claims involved in this picture of idealized liberal society. I have alluded previously in this work to the particular classical liberal content of North (and colleagues’) normative framework, and I will have occasion to contextualize North’s work as a whole within this context in the final chapter. Here, I am more concerned with the evolutionary dynamic imputed to competition-as-adaptive-efficiency. I want to clarify more precisely what for North the relation is between the evolution of human cognition and sociality, the origin of institutions, and his narrative of economic history as the search for competitive institutions backed by commitment and belief. In the process, this will also allow me

---

<sup>375</sup> *Ibid.*, p.107-108.

<sup>376</sup> *Ibid.*, p. 108.

<sup>377</sup> North, Wallis, and Weingast 2009, p. 133-136.

<sup>378</sup> *Ibid.*, p. 140.



to ask how North's ambiguous approach to rationality discussed previously asserts itself in dealing with the cognitive and evolutionary sources of institutions, and the question of what room exists for human other-regard and altruism in his work. Finally, I will make an evaluation of his claims regarding cultural evolution, both for the origin story of human social institutions and the transhistorical classical liberal narrative of competition and efficiency.

## **Section I. Examining North's theory of cultural evolution**

### **The evolutionary dimension of North's NIEH**

First off, we have to observe the peculiarities of the term 'adaptive efficiency'. We know that North abandoned the 'efficiency view', the approach which functionally explains institutions as efficient responses in a given environment, long ago. North's understanding of what makes for efficiency in the case of a whole artifactual framework would then need some explanation, which the claims about competitiveness and commitment would provide. More importantly, however, is the term 'adaptive'. This term is unmistakably a reference to such efficiency coming about as the result of an evolutionary process: adaptations and adaptive are terms of art in evolutionary biology to describe how the unit of selection adjusts to its environment such as to increase its 'fitness'.<sup>379</sup> If the efficiency of institutional arrangements as a whole is part of an evolutionary process, as the term suggests, this would imply that societies as a whole are, for North, units of selection in cultural evolution.

There is evidence that North indeed sees things in this way. Although North begins *UPEC* by explicitly disavowing an immediate analogy between "Darwinian evolutionary theory" and the process of institutional change he sets out to describe, he attributes this difference not to the absence of selection (say), but to the intentionality of agents in bringing about institutional arrangements.<sup>380</sup> This intentionality makes the process of change one of unintended consequences and path dependences at macro-level, produced by the peculiar "blending" of beliefs and preferences (as discussed in the previous chapter) North sees as the bedrock of choice at the micro-level.<sup>381</sup> It is not obvious, however, why this intentionality should *ipso facto* prevent the operation of a process of cultural evolution. North repeatedly asserts that his difference with a 'Darwinian' approach and with evolutionary economics rests on the importance of intentionality (likely especially beliefs, given their important role in North's theory) of

---

<sup>379</sup> The concept of a 'unit of selection' is contested, and has given rise to considerable debate: see Sober and Wilson 1994; Lloyd 2012; cf. Hodgson and Knudsen 2010, p. 228. It can mean the unit that is the beneficiary of selection, as per Dawkins 1982; but either way it must distinguish units of replication from units of environmental interaction, as pointed out by Hull 2001, p. 61. Sober and Wilson 1994 favor defining it as the unit of interaction, which, *per* Lloyd 2012, is the most common solution in the literature (cf. Gould 2002, p. 624). Even so, the point here is not to delve into these debates in great detail, but to identify whether *any* unit of replication *or* interaction is defined in North's evolutionary approach.

<sup>380</sup> North 2005, p. viii.

<sup>381</sup> *Ibid.*, p. viii.

individuals.<sup>382</sup> But he provides no argument why or how this should qualify his own evolutionary account elsewhere.

We must therefore consider to what extent North's appeal to evolutionary reasoning is to a Darwinian-type process, and to that extent, how well it succeeds as such. But this requires also considering his objection to that identification: namely the importance of intentionality. By no means all evolutionary approaches to economics use an explicitly Darwinian type reasoning, as for example Richard Nelson and Sidney Winter did not do (see below for a further discussion of their work in the context of North).<sup>383</sup> Indeed the use of Darwinian evolutionary thought is by some associated with the Old Institutional tradition in direct contrast to the New Institutionalists' insistence that the purposeful action of agents makes a difference to the evolutionary mechanism.<sup>384</sup>

But North gives no explicit argument why the centrality of intentionality to his account, as he sees it, should mean that his evolutionary argument cannot be compatible with a Darwinian account. Indeed, as a commonly held notion, that the intentionality of human beings falls outside the purview of Darwinian reasoning and invalidates the application of such reasoning to the production and evolution of culture, it is simply false. Indeed, the purpose of a Darwinian theory of cultural evolution is to *explain* this intentionality, both in its evolutionary origins and in the evolutionary effects of its operation. To assume human intentionality may be valid, but to consider it a point of disjunction with Darwinian evolution is not. As Hodgson and Knudsen note: "Darwinism does not take intentionality as given. It holds that intentionality and other human mental capacities must have evolved from similar but less developed attributes among our prehuman ancestors. It insists that intentionality must be explained rather than simply taken for granted."<sup>385</sup>

This is all the more striking because North's subsequent argument does not actually imply any such disjunction. Far from seeing intentionality and its product, culture, as some form of irreducibly 'thick' social process that cannot be meaningfully explained in evolutionary terms, as some social scientists do, North himself undertakes his discussion in *UPEC* by rather suggesting the contrary. Firstly, he observes that human culture, including knowledge, beliefs, and learned habits, are a cumulative residue of past experience of members of a society, one that is inherited. This heritage not only "constrains the players" (here we are back at instrumental rationality), but in fact straightforwardly "determines societal performance at a moment of time", elsewhere called the "performance characteristics of societies".<sup>386</sup> But there is more. North subsequently rightly points out that such cultural-institutional frameworks as make up societies are themselves

---

<sup>382</sup> *Ibid.*, p. viii, 66.

<sup>383</sup> For Nelson and Winter's implicit use of Darwinian principles, see Hodgson and Knudsen 2010, p. ix.

<sup>384</sup> Spread 2016, p. 71.

<sup>385</sup> Hodgson and Knudsen 2010, p. 48.

<sup>386</sup> North 2005, p. viii, ix.

possible due to human evolution: partially the “genetic architecture of humans” that enables the cognitive and social behaviour so as to have culture, and partially “cultural evolution such as the development of institutions to favor larger group cooperation”.<sup>387</sup> The origin of institutions (or culture) themselves is clearly evolutionary in nature. Finally, the process of path dependence, and other factors that tend to create continuity between generations of a culture despite ongoing change, operate analogously to heritability even in North’s own description. He describes long periods of incremental changes interspersed with brief periods of rapid change as a case of “punctuated equilibrium change”, and the temporal factor of economic history is explicitly mentioned as a reason to take inspiration (be it with the above reservation) from evolutionary theory.<sup>388</sup>

It is clear that North is at times hesitant to commit too strongly to one or another position in the fraught debates on cultural evolution and evolutionary economics, but his account of the origins of institutions and of ‘adaptive efficiency’ are equally clearly meant as building on those ideas – not least on Hayek’s view of knowledge and traditions as heritable traits in a process of spontaneous cultural evolution.<sup>389</sup> Similarly, North builds on the view of democracy as an evolutionary method developed in political science.<sup>390</sup> Finally, the first six chapters of *UPEC*, roughly half the book, are dedicated to explaining adaptive efficiency, and of that, at least three chapters (3, 4 and 6), and arguably chapter 7 too, are specifically about evolutionary processes in the most literal sense.

Of course, there have often been suggestions that evolutionary theories do apply to the process of culture and institutional change, but not a Darwinian one; for example in the appeals to the theory of Jean-Baptiste Lamarck. There is something of a tradition of such appeals in the history of evolutionary approaches to institutions, including authors that North took inspiration from, as in the work of Friedrich Hayek, of Nelson and Winter (mentioned above), and even of Herbert Spencer.<sup>391</sup> But as has been understood in evolutionary biology for a considerable time, no Lamarckian inheritance mechanism can exist independently of the Darwinian: if there is a Lamarckian form of inheritance, “the Lamarckian theory can explain adaptive improvement in evolution only by, as it were, riding on the back of the Darwinian theory.”<sup>392</sup> And North is happy to invoke such a Darwinian product as the ‘genetic architecture of humans’, without which the intentionality so important to his approach could not have come to exist, and could not persist, in the first place. There can be no doubt then that whatever North’s reservations about what he calls Darwinism and intentionality or his hesitations to wade into this minefield, a full discussion of

---

<sup>387</sup> *Ibid.*, p. ix.

<sup>388</sup> *Ibid.*, p. 2-3, 65-66.

<sup>389</sup> *Ibid.*, p. 51.

<sup>390</sup> *Ibid.*, p. 56.

<sup>391</sup> See e.g. Hayek’s statement that “cultural evolution simulates Lamarckism”: Hayek 1988, p. 25; Cf. Marciano 2009. For Nelson and Winter, see: Nelson and Winter 1982; cf. Vromen 1995, p. 125, on Nelson and Winter’s misnomer of their own ‘dual’ evolutionary approach. For Herbert Spencer’s ‘Lamarckism’, see Bowler 2014.

<sup>392</sup> Dawkins 1986, p. 300. See also Hodgson and Knudsen 2010, ch. 3.

North's work requires investigating his account as a theory of cultural evolution that at least should meet criteria comparable to the Darwinian framework. Of course, this leaves open the question whether it is a consistent and complete evolutionary framework by such criteria, which is what I will now discuss. But it is worth noting that for a theory of cultural evolution to conform to Darwinian criteria, it need not be a *complete* Darwinian theory. In fact, no Darwinian evolutionary approach is complete on its own: it provides mechanisms, but these always need application in specific empirical contexts and using auxiliary theories for the specific case. It does however, at least in its modern form, define these mechanisms and the criteria for their operation with some degree of precision.<sup>393</sup>

In order to proceed with understanding this account, we have to divide the discussion according to the two aspects of cultural evolution I have mentioned: the evolution of institutions as a problem of their *origins*, and the evolution of institutions as a problem of their dynamics over time and economic effects, i.e. of their *persistence*. Put differently, the first question is how and why an 'artifactual framework' comes into being in the first place, and the second question is how it works and changes once it is there, and what this means for its adaptive efficiency over time. North's approach gives us sufficient means to discuss both aspects as part of an integrated process of cultural evolution that, if it is to succeed at all, must succeed in Darwinian terms, despite his own reservations. Although there are many different usages of the term 'evolution', it is generally agreed that for a Darwinian type process of evolution by means of natural (or artificial) selection to take place, all that is required is (high accuracy) heritability or reproduction, variability, and differential fitness (or competition, which is the same thing). North's approach gives us these: we have cultural inheritance and path dependence (heritability), we have differential fitness (adaptive efficiency), and we have variation ("the immense variation in the performance characteristics of societies" North observes has its source in different institutional 'traits'<sup>394</sup>).

Using North's own examples and ideas, I will explore his view of each of these in turn. Note that even given the possibility of an evolutionary account of these processes, as the above allows, it does not necessarily mean that such an account is the best or most plausible evolutionary account, or that the three criteria are rightly identified. In this chapter the task is therefore to evaluate how well North's account works as a description of a process of cultural evolution, whether North has taken his inspiration for this account in the right places, and having identified problems with the account, to suggest a better way of conceptualizing cultural evolution and its relation to economic history.

### **North and the origin of institutions**

---

<sup>393</sup> Hodgson and Knudsen 2010, p. 5.

<sup>394</sup> North 2005, p. ix.

While North's elaborate attention to theories of human cognition and its evolution may seem at first glance rather curious in a book on economic history, the evolutionary dimension of North's work makes the importance of this subject for his approach more understandable. We have seen that beliefs are an important complement to more explicit institutional arrangements in North's account, and the correct beliefs are essential in his view for an open access order – the ideal in adaptive efficiency - to come into being. Since beliefs emerge from the interaction of cognitive processes in the human brain with an uncertain and ergodic environment, as North ever emphasizes, that alone would suffice to pay some attention to what happens at this interface. But more than this, beliefs are also the source of institutions in the first place, despite their ambiguous instrumental role in North's theory, as we have seen in the previous chapter. Since institutional frameworks as a whole are his preferred unit of selection, it stands to reason that he would be concerned with how beliefs arise and how they affect the translation of many individual rationalities (bounded as they are) into shared institutional orders and belief structures. Or as North puts it: "the beliefs that humans hold determine the choices they make that, in turn, structure the changes in the human landscape... We begin with exploring the mind of the individual as a necessary condition to understanding societal beliefs."<sup>395</sup>

North presents his account of the evolution of cognition as a necessary complement to standard accounts because of the inadequacy of rational choice equilibrium reasoning, i.e. neoclassical economics. Here we are on familiar ground, with North presenting his favoured approach as a necessary consequence of the inadequacy of neoclassical economics' rationality concept (despite the internal inconsistencies discussed in chapter 3). Because beliefs matter, as North rightly insists, the acquisition of beliefs from the environment matters too. He surveys briefly theories of pattern recognition and the limited cognitive reasoning abilities of human minds. Then he observes how this contextualizes rational choice in a larger institutional framework, without doing away with it. This larger institutional framework is itself the process of evolutionary competition as a whole, and emerges from imperfect feedback on imperfect information and limited cognitive abilities. As he puts it: "the implication... for social science theorizing is that much of what passes for rational choice is not so much individual cogitation as the embeddedness of the thought process in the larger social and institutional context."<sup>396</sup> We again see here that the evolutionary dimension contributes to the continuous rethinking of what rationality means in North's work, without an obvious solution to problems of indeterminacy presenting itself. In this sense, the evolutionary focus on learning and cognition almost acts as a 'backup' reasoning for the problems North encounters in theorizing the interface between individuals and institutions in other works.

But of course our own cognition is the result of biological evolution, and this raises the sociobiological problem of what this means for human sociality and culture. It is I think a

---

<sup>395</sup> *Ibid.*, p. 23.

<sup>396</sup> *Ibid.*, p. 24.

striking fact that North treats this in only a summary way, and mainly with the aim of emphasizing the importance of producing culture for human behaviour. The main antagonist here is the sociobiological school of ‘evolutionary psychology’, represented by John Tooby, Lana Cosmides and others, who have long asserted that the adaptive landscape of the Pleistocene has shaped a strongly modular human brain in such a way as to render much of our behaviour only explicable as adaptations to a prehistoric context long gone.<sup>397</sup> North accepts that such ‘universals’ may operate still, as maladaptations from a bygone era in the period of human civilization, but questions how far this can take social science given the vast variation in cultural traits displayed by human societies in history. Culture matters in an evolutionary picture, he rightly emphasizes, and it is not obvious how this school of evolutionary psychology can account for that in terms of institutional variation. As he insists, “the immense variation, however, in the performance characteristics of political/economic units over time makes clear that the Lamarckian characteristics of culture must also be central to the understanding of the process.”<sup>398</sup> Note again the use of evolutionary terms to describe cultural development and change (“Lamarckian”): North is not arguing against cultural evolution as such, merely the account of this branch of sociobiology.<sup>399</sup>

Returning then to how culture emerges from “the link between the mind and the environment”, Hayek again looms large for North. Most important is Hayek’s observation from *The Sensory Order* that “the apparatus by means of which we learn about the external world is itself the product of a kind of experience. It is shaped by the conditions prevailing in the environment in which we live, and it represents a kind of generic reproduction of the relations between the elements of this environment which we have experienced in the past.”<sup>400</sup> This is another cue for North to insist on the holistic nature of learning, its dependence on the experience of the world as mediated by the human environment (the society) as much as by the natural. Rather strikingly, he cites Edwin Hutchins to the effect that “we cannot adequately understand cognition without accounting for the fact that culture, context, and history . . . are fundamental aspects of human cognition and cannot be comfortably integrated into a perspective that privileges abstract properties of isolated individual minds.”<sup>401</sup> (Again, one wonders how this squares with the account of agents in institutional change and the assumption of individual, instrumental rationality of such ‘players’ of the ‘game’ seeking to ‘line their pockets’ that we saw in the previous chapter.)

The upshot of this is to see, with Hutchins and Hayek, learning as “adaptive reorganization in a complex system”, and institutions as cognitive instruments forming adaptations to present

---

<sup>397</sup> E.g. Tooby and Cosmides 1992, cited in North 2005, p. 28.

<sup>398</sup> North 2005, p.30.

<sup>399</sup> This is also the one occasion – to my knowledge – that North discusses experimental economics and its implications. I will consider his use of these results in my evaluation of his theory further below.

<sup>400</sup> North 2005, p. 32-33; Hayek 1952, p. 165.

<sup>401</sup> North 2005, p. 33-34. Internal quotation marks omitted.

(or past, in the case of path dependence) environments.<sup>402</sup> These institutional frameworks and their greater or lesser ability to generate adaptive efficiency are then the subjects of selection. In other words, learning operates at group level as much as at individual level, because of the importance of social context to individual learning processes. Because of widespread uncertainty and continuous change, human learning and retention of knowledge is an essential evolutionary development towards culture. Institutions allow individual learning to proceed from an inheritance of culturally stored knowledge. These learning processes cause this larger whole, the institutionalized society or group, to act as the unit of selection of cultural evolution. These societies as a whole act so as to maximize their fitness (i.e. they adapt intentionally). He cites approvingly the work of Andy Clark, who writes that “when the external scaffolding of policies, infrastructure and customs is strong and (importantly) is a result of competitive selection, the individual members are, in effect, interchangeable cogs in a larger machine. The larger machine extends way outside the individual, incorporating large-scale social, physical, and even geopolitical structures. And it is the diffused reasoning and behavior of this larger machine that traditional economic theory often succeeds in modeling.”<sup>403</sup> In other words, whether or not individuals act according to conventional models of rational choice, due to selective pressure, societies (institutional frameworks) as a whole do. But these processes are imperfect and it is difficult to achieve adaptive efficiency in such institutional arrangements, something only maximizing individual free competition under conditions of legitimizing beliefs and believable commitments to property rights can do.

North makes this connection between the evolutionary origin of institutions in individual and collective learning and the macro-level evolutionary process explicit in a paragraph worth quoting in full: “When we move from Hutchins’s dynamic social group ... to the larger implications for the structure, functioning, and process of change for a whole society we can see that the cultural heritage provides the artifactual structure—beliefs, institutions, tools, instruments, technology—which not only plays an essential role in shaping the immediate choices of players in a society but also provides us with clues to the dynamic success or failure of societies through time. In essence, the richer the artifactual structure, the greater the reduction of uncertainty in making choices at a moment of time. Over time, the richer the cultural context in terms of providing multiple experimentation and creative competition, the more likely the successful survival of the society. These generalizations require careful elaboration and qualification, but they are a foundation of this study.”<sup>404</sup>

### **Macro-level evolution in North’s account**

At the macro-level, North moves from this cognitive framework to the unfolding of competitive, impersonal exchange as a mechanism making large-scale cooperation possible. This historical

---

<sup>402</sup> *Ibid.*, p. 35-36.

<sup>403</sup> Cited in: *Ibid.*, p. 24-25.

<sup>404</sup> *Ibid.*, p. 36.

process is not predetermined to succeed. Indeed, on North's account, none other than the Western world have achieved it. The means of achieving such cooperation is a change in the payoffs in the game theoretical matrix such that cooperation becomes a dominant strategy. But because in North's theory the choice set of institutional arrangements depends on beliefs and previous institutional arrangements, in other words because path dependence matters, this is not something that can be arbitrarily willed into being: "the shift from personal to impersonal exchange has produced just such a stumbling block both historically and in the contemporary world. Personal exchange relies on reciprocity, repeat dealings, and the kind of informal norms that tend to evolve from strong reciprocity relationships. Impersonal exchange requires the development of economic and political institutions that alter the pay-offs in exchange to reward cooperative behavior. The creation of the necessary institutions requires a fundamental alteration in the structure of the economy and the polity which frequently is not in the feasible set given the historically derived beliefs and institutions of the players."<sup>405</sup> Nonetheless, North is explicit that impersonal trade, the competitive market benchmark discussed in chapter 2, is the *sine qua non* of evolutionary success: "successful evolution has entailed radical alteration in economic institutions in order to make such long distance and impersonal trade viable."<sup>406</sup>

What needs clarification then is the connection between evolutionary processes, the by now familiar Northian notion of cooperation as market exchange, and adaptive efficiency. As one might expect, the missing link is the theory of transaction costs. Disclaiming any simplistic *laissez faire* views, North insists that because the evolutionary environment is constantly changing, the institutions necessary to maintain adaptive efficiency must do so too. There is therefore no simple, one-size-fits-all recipe. Rather, the lower the transaction costs in a given market situation, the more efficient market exchange and therefore the more cooperation. The constant adjustment of institutions required to maintain a low level of transaction costs is the process of evolutionary adaptation.<sup>407</sup> Because of this, a simple Washington consensus development policy will not suffice, because there is simply insufficient knowledge among all the agents concerned to achieve the aim of efficient markets (one North otherwise shares).<sup>408</sup>

North's macro-level historical account, with its characteristic combination of viewing beliefs and institutions as constraints on instrumental reason and its evolutionary concern with path dependence and adaptivity, is then perhaps best summarized in his own words. "In a Coasian world the players would always choose that policy that maximized aggregate well-being with compensation for any losers; but the real transaction costs are frequently prohibitive reflecting deep-seated beliefs and prejudices that translate into such prohibitive transaction costs. It is more than two hundred years since Adam Smith explained the underlying sources of the wealth of nations but the extent to which such views are embedded in the decision-making

---

<sup>405</sup> *Ibid.*, p. 117-118.

<sup>406</sup> *Ibid.*, p. 118.

<sup>407</sup> *Ibid.*, p. 122-124.

<sup>408</sup> *Ibid.*, p. 124-126.



process of those shaping political/economic change is problematic.”<sup>409</sup> In other words, we have known since Adam Smith that we need competitive markets to achieve wealth, and we need institutions protecting property and liberty to make such markets work, as well as a set of beliefs compatible with their maintenance. Moreover, such a competitive free market society also has the great virtue of being more adaptive than any other, and thereby able to produce a virtuous circle of self-perpetuation, the sign of its evolutionary fitness. The problem is that the ideological belief in or commitment to such competitive markets is often missing, in addition to the correct market incentives and secure property rights.

The developmental recipe North proposes then has ‘evolutionary’ flexibility, in that it concentrates on ability to adapt rather than simply ‘getting the prices right’. But the recipe itself is much the same as that of classical liberal economics, albeit with a renewed emphasis on cultural and ideological factors.<sup>410</sup> This is illustrated by the case study North uses to support his idea of adaptive efficiency in historical practice, namely the rise and fall of the Soviet Union. On North’s account, the Soviet system worked well for a while, but when confronted with a novel environment failed to adapt, lacking the institutional means to effect such change. As a result, its structure became mismatched with the present conditions, and the system collapsed – not unlike the extinction of organisms strongly adapted to a particular niche who fail to adapt to a changing environment.<sup>411</sup> This dovetails with his endorsement of Hayek’s classic Austrian school argument against socialism, namely the way the price system carries dispersed and tacit information that the survival of an economic system cannot do without. From this viewpoint, market exchange, based on prices responsive to their environment, is the ‘Smithian’ solution to the adaptivity problem.

But for North prices are not the only such carriers of intentional knowledge; institutions are too.<sup>412</sup> The story of economic history is then the story of achieving (or failing to achieve) open access orders, which is to say, the right combinations of beliefs, institutions, and competitive markets in politics and economics to achieve adaptive efficiency. In *Violence and Social Orders* this account is complemented by an account of violence and the state. After all, North has not yet shown how social life is possible in the first place, i.e., how institutions arise in a durable fashion. For North and colleagues, the root of human cooperation is the limitation of violence.<sup>413</sup> This approach deepens cooperation beyond the ‘Smithian’ functions of market exchange. So in addition to the functions exercised for achieving cooperation by the market price system and by institutional norms, cooperation now also consists of the maintenance of durable social ties in larger groups itself. These are then a prerequisite, chronologically and analytically, for the

---

<sup>409</sup> *Ibid.*, p. 156.

<sup>410</sup> *Ibid.*, p. 157-161. For another statement of the close relationship between classical or ‘ordoliberal’ *Ordnungspolitik* and North’s adaptive efficiency approach, see: Richter 2015, p. 178-179.

<sup>411</sup> *Ibid.*, ch. 11.

<sup>412</sup> *Ibid.*, p. 162.

<sup>413</sup> North, Wallis, and Weingast 2009, p. 258.

‘Smithian result’ and indeed any other institutional arrangement. (As we shall see, this emphasis makes sense, since the origins of sociality is one of the main points of debate in sociobiology and its economic implications.)

The control of violence requires rent creation in addition to beliefs and institutions that maintain order, posing an additional obstacle to the evolutionary transformation of such ossified structures to a competitive and therefore more adaptive one. “Understanding social change in actual historical events requires separating institutions, organizations, and beliefs, as well as violence, in order to track their interrelated development over time... the transition [to open access orders] provides citizens with new tools, fewer restrictions, and greater scope for impersonal relations, all of which dramatically increase the gains from specialization and exchange while reducing the risk of expropriation.”<sup>414</sup> Rule of law and democracy then appear as the correct institutions for competitive political markets as in *UPEC* strong private property rights and rules favouring efficient impersonal exchange are for competitive economic markets.<sup>415</sup> This argument is in my view best seen as complementary to that of *UPEC* and the evolutionary process it describes, not as a substitute. Arguably, *Violence and Social Orders* does for a cultural evolutionary account of ‘political markets’ what *UPEC* does for a cultural evolutionary account of economic markets. Both are therefore complementary in the larger economic historical narrative that constitutes North’s NIEH.

## **Section II. Evaluating North’s theory of cultural evolution**

Before moving on to an evaluation of this set of arguments, allow me for convenience to sum up my reading of North’s evolutionary account. Evolved intentionality – mediated by limited cognitive abilities - gives rise to beliefs and institutions of social life as heuristics in conditions of uncertainty. These beliefs and institutions shape and change the ‘human environment’, so that there is a constant changing interaction between these two domains. But the durable maintenance of social groups (on a larger scale) requires also the reduction of pervasive violence, as per a Hobbesian account, so that the institutions of the ‘natural order’ give rise to elites who can create and distribute rents in order to reduce this violence. This institutional whole gives rise to a particular incentive structure within which agents (individuals and organisations) pursue their aims in a boundedly rational fashion.

This pursuit of preferences is presented as an ambiguous and indeterminate mixture of rational pursuit of self-interest given incentives and constraints imposed by slowly changing beliefs. Norms and values, self- or other-regarding, are part of the institutional structure insofar they are translated into rules enforced in society. However, although the enforcing body need perhaps not be the state, states or other elite controllers of violence and their subsidiary bureaucracies are usually presented as the main enforcers of institutions. I.e., political

---

<sup>414</sup> *Ibid.*, p. 259, 261.

<sup>415</sup> *Ibid.*, p. 262-267.

enforcement is the main concern.<sup>416</sup> Since institutions are *constraints* on action, these norms cannot affect the nature of rational action as such, although different institutions may produce different payoff matrices for cooperative or noncooperative behaviour. Conventional forms of utility maximization appear to be still the operating assumption, albeit in an indeterminately bounded manner.

Agents seek to preserve or (incrementally) change institutions in pursuit of their own interests, a process which, being both imperfectly predictable and path dependent, further develops the non-ergodic nature of the human environment. Institutions are expressions of cultural learning and are inherited as social wholes, so that piecemeal adjustment is not easily achieved and so that cultures as a whole act as evolutionary replicators. Insofar as rent seeking and inefficient institutions and/or beliefs prevail, transaction costs will be high, thereby hindering the achievement of a competitive society geared for economic growth. As a result, many societies become locked into particular combinations of suboptimal institutions, political orders, and/or beliefs, which over the longer run become maladaptive in light of changing environments. Such societies then fail, as the Soviet example illustrates. Only insofar as institutions, beliefs, and political orders align to reduce transaction costs and to make economics and politics alike as impersonally competitive as possible can the (alleged) prediction of Adam Smith regarding the efficiency of market exchange be realized. This is what appears to be meant by economic or even ‘societal’ performance. Moreover, only through competition will a society adapt to changing circumstances and therefore survive the selection process of cultural evolution. This is what appears to be meant by adaptive efficiency.

### **North’s cultural evolutionary account in theoretical context**

One of the most striking aspects of North’s cultural evolutionary approach is how little it engages with existing relevant literature in either economics or sociobiology. Besides the brief dismissal of ‘Darwinian evolution’ as lacking in intentional explanation and an equally brief dismissal of the school of evolutionary psychology, North seems to have little interest in integrating his approach with existing evolutionary theories of culture or economic behaviour, or at least in explaining how this is to be done.<sup>417</sup> Of course, Hayek is an exception to this; but Hayek’s work mostly predates the post-war revival of evolutionary economics as well as the rise of sociobiology, and itself does not much engage with other works along these lines, so that this rather underlines the problem. The question is, based on the analysis I have offered, whether this evolutionary approach in North is founded on more than the two somewhat opportunistic functions I have identified so far: namely, to give an account of how the set of institutions and beliefs associated with classical liberalism can be better than other sets of institutions, an account not based on the approach of neoclassical economics, and to circumvent his difficulties with the rationality question by moving from evolved cognition to beliefs and from beliefs to institutions

---

<sup>416</sup> E.g. *ibid.*, ch. 2.4.

<sup>417</sup> North 2005, p. viii; 6; 28-29.

and their macroeconomic effects. In any case, for North's account, the emphasis on beliefs and the analytical steps from bounded cognition to beliefs to institutions is clearly central.<sup>418</sup>

These are absolutely legitimate and plausible areas of interest, and North is surely to be commended for his awareness of the problem where institutions come from and how they relate to beliefs, a problem that plagues the analytical separation between beliefs, preferences, and behaviour needed for rational choice approaches to work. (We have seen this problem already in the context of the discussion on game theory in the previous chapter.) But it is only a small part of the relevant problems in evolutionary economics, and an even smaller part of what is needed for a plausible and sufficiently complete sociobiological account of cultural evolution. Moreover, North's political economic assumptions (most notably the virtues of the competitive society) produce a certain 'motivated cognition' in his evolutionary account that causes him to skate over some weakly developed points: his choice of units of selection, the nature of competition between societies that would cause differential survival, and the integration of his account with the evolution of intersubjective rationality, i.e. the interaction and cooperation between individuals that makes a society and its institutions possible in the first place, to name but a few. The first of these are important for the coherence of his account of an evolutionary mechanism, whereas the latter is perhaps *the* classic problem of integrating his contractarian view of cooperation and rationality, the topic of chapter 3, with what is known (or theorized) in sociobiology.

Illustrative on this point is his one engagement with behavioural economics, namely one of the various game theoretical experiments testing rationality assumptions and their results undertaken in recent years. As we have seen before North cites Vernon Smith, Elizabeth Hoffmann and Kevin McCabe in an important summary of their work in this domain: "People invoke reward/punishment strategies in a wide variety of small group interactive contexts. These strategies are generally inconsistent with, but more profitable than, the noncooperative strategies predicted by game theory. There is, however, consistency with the game theoretic folk theorem which asserts that repetition favors cooperation, although we observe a substantial use of reward/punishment strategies and some achievement of cooperative outcomes in single play games. Non cooperative outcomes are favored, however, where it is very costly to coordinate outcomes, in large groups, and even in smaller groups under private information. In large groups interacting through markets using property rights and a medium of exchange, and with disperse private information, non-cooperative interaction supports the achievement of socially desirable outcomes. Experimental studies have long supported this fundamental theorem of markets. This theorem does not generally fail, however, in small group interactions because people modify their strict self-interest behavior, using reward/punishment strategies that enable some approximation of surplus maximizing outcomes. Seen in the light of evolutionary psychology,

---

<sup>418</sup> North 2005, ch 2-4; see also Knight and North 1997.

such behavior is not a puzzle, but a natural product of our mental evolution and social adaptation.”<sup>419</sup>

These observations are extremely significant for the integration of theories of ‘social adaptation’, the sources of cooperation, and the deviations of observed rational behaviour from the expectations of conventional game theory, which I consider among the central problems faced by a Northian approach. And yet North seems little concerned or interested by this; he has no difficulty, as we have seen, pursuing conventional game theory approaches to agents’ rationality while simultaneously appearing to endorse these conclusions. The connection with the evolution of cooperation as the central problem in accounting for the emergence of institutionalized society is not made; North simply observes that sociobiology has discussed this before, drops the subject, and moves on to the importance of the cognitive sources of ‘nonrational’ beliefs.<sup>420</sup>

In my view, North’s neglect of essential inputs into any cultural evolutionary theory is damaging for his account; but more importantly, it is a significant missed opportunity for discussing precisely this core area of overlap between problems in cultural evolution, in theories of rationality, and in evolutionary-institutional economics. I shall therefore in this section try to develop some of these points, both to show the limitations of North’s account in *UPEC* and elsewhere and to suggest directions in which, in my view, it could be fruitfully developed. This entails placing North’s cultural evolutionary theory in the context of wider debates in evolutionary economics and sociobiology. To limit this task to manageable proportions, I will not here review the whole of literatures in either of these fields, but restrict myself to a consideration of (respectively) some core problems with the nature of evolutionary mechanisms in economic structures and the sociobiological issue of the origins of cooperation and its implications for evolutionary accounts of institutions. Each of these I will discuss in a limited way, mainly seeking to draw out implications for North’s joining of evolutionary and New Institutional Economics and to show the importance of these debates for developing such an approach further. As I will show, as with North’s partial rejection of conventional microeconomic assumptions, his exploration of evolutionary dimensions of institutions is not flawed because such an approach is unnecessary, but rather for the opposite reason: North does not take his own logic to its theoretical conclusion. As before, I believe the most plausible interpretation of this phenomenon is North’s attachment to a contractarian view of cultural and economic institutions and to an idealized competitive market society as the ideal institutional arrangement.

## **The context of evolutionary economics**

---

<sup>419</sup> Hoffman, McCabe and Smith 1998, p. 350; cited in North 2005, p. 29.

<sup>420</sup> North simply concludes that “*the most important contribution of the evolutionary psychologist is explicating the underlying inference structure of the mind that appears to account for the predisposition of the mind to entertain and construct “non-rational” beliefs...*” North 2005, p. 29-30, emphasis added.

The application of evolutionary arguments to the analysis of the modern economy, in particular that of the capitalist firm, goes back to at least the work of Thorstein Veblen.<sup>421</sup> In more recent times, the work of Richard Nelson and Sidney Winter did much to put the identification of evolutionary dynamics in the economics of the firm, and by extension microeconomics in general, back on the agenda.<sup>422</sup> In their work, Nelson and Winter applied the theory of natural selection to the interaction between firms in an economy over a longer period of time, emphasizing the effects of competition on the differential economic survival of such firms. While the idea of economic competition as akin to the ‘survival of the fittest’ was nothing new (already a favourite metaphor of Spencerian liberalism in the Victorian age), more significant was Nelson and Winter’s identification of the units of replication and selection. Emphasizing that firms were not necessarily fully profit maximizing, nor the market necessarily in a state of equilibrium, they concentrated on the internal organization of the firm as the heritable equivalent to the genotype. The firm itself being the phenotypical expression of routines and habits of organisational behaviour within it, such routines were, for Nelson and Winter, replicators that had both the necessary stability and heritability and the necessary potential for random or non-random variation required for a process of selection. These are then the units of selection.<sup>423</sup>

In this way, Nelson and Winter could identify within the domain of industrial organisation and modern market competition an evolutionary process by observing how it fulfilled the necessary criteria: the existence of sufficiently stable, but mutable units, of their heritability, and of a process of differential survival. For the moment, the point is not whether Nelson and Winter were right to identify routines as units of selection in this manner, but simply to realise the significance of this argument. In Nelson and Winter’s account, the interaction between the organisation of firms and the competition between firms is identified as an *actual* process of natural selection, not merely a politically convenient *metaphor* or analogy as in the Spencerian tradition. Indeed, as Marcus Becker notes, the analysis of routines became very influential in reviving the promise of an evolutionary economics precisely because “the great promise that the concept of routines holds for evolutionary economics is that it might enable the application of an evolutionary explanation in economics”, that is to say, it would be “a cornerstone of an evolutionary theory of economic change”. With the identification of a gene equivalent, suddenly the prospect of actual Darwinian selection, not just metaphorical Darwinism, became possible in the economic sphere.<sup>424</sup>

The affinity of such approaches with institutional economics is easy enough to see. Evolutionary explanations serve to explain change, and rather than the static equilibrium approaches of perfect competition neoclassical economics, they are easily combined with imperfect markets and imperfect information assumptions. The behaviour of economic agents

---

<sup>421</sup> Veblen 1898.

<sup>422</sup> The main work is Nelson and Winter 1982.

<sup>423</sup> Becker 2004, p. 2.

<sup>424</sup> *Ibid.*, p. 2, 29; Nelson and Winter 1982, p. 134-136, 160-161; Hodgson and Knudsen 2010, p. ix, 179.

under uncertainty is an important area of study of institutional economics, as we know, and therefore the two go together very easily. One early example is Armen Alchian's evolutionary analysis of competition under uncertainty, where the main theoretical innovation was to substitute survival (fitness) for profit maximization per se in the model of firm interaction and market change.<sup>425</sup> (This paper was an inspiration for Nelson and Winter's work as well.) Moreover, the element of heritability and the constraints on incremental change as a result of random or constrained mutability of the unit of replication both imply the importance of historical trajectories of change, i.e. path dependence. In North's account, we have already encountered the importance of Alchian's emphasis on uncertainty and the role of path dependence; indeed North's NIEH is in part so interesting precisely as an exemplar of the joining of certain strands in (new) institutional economics and in evolutionary economics.

The foundation of this merger are two insights, already incipient (but not explicitly stated in these terms) in Nelson and Winter's work: that Darwinian selection is a generalizable theory not limited to the sphere of nonhuman biology, but directly applicable in reality (rather than metaphorically) to certain sociocultural human systems; and that understanding an economy means understanding economic *change* in an 'embedded' fashion, that is to say understanding the role of the institutional structure of an economy and its social context, and that such institutions must exist to complement markets in any economic structure not possessed of perfect competition and perfect information. The generalization of Darwinian theory simply follows from the nature of Darwinism itself: it is a process that applies if and only if the criteria of variation, selection (through competition), and heredity apply, although its application to the social sphere requires identification of evolutionary units beyond the genetic, 'cultural' units such as habits or institutions. The viewpoint of generalized Darwinism is defended by Howard Aldrich, Geoffrey Hodgson and colleagues when they observe that although "cultural selection processes are different from biological selection processes", it is nonetheless the case that where these criteria can be identified, "the expression of the underlying core Darwinian principles of variation, inheritance and selection differ in important ways, yet the overarching general principles remain."<sup>426</sup> Indeed, Darwinian selection is a *general* theory, describing the nature and operation in abstract terms of a particular algorithmic process. For this reason, it is not limited to the description of the operation of particular biological mechanisms alone, but applies generally wherever the algorithmic process operates on relevant entities.<sup>427</sup>

The point of contestation should not be, therefore, whether Darwinian selection can in principle be generalised, but rather where in human sociocultural life it can fruitfully be applied and to what units of selection. This also entails, however, recognising the warning of these economists that Darwinian theory, precisely because of its generality, is incomplete. It is all too easy to apply the language of selection to any given sociocultural process; it is quite another

---

<sup>425</sup> Alchian 1950.

<sup>426</sup> Aldrich et al. 2008, p. 580.

<sup>427</sup> See e.g. Dennett 1995 for an animated elaboration of this point.

thing to demonstrate its real (rather than metaphorical) applicability. As Aldrich et al. rightly write: “Darwin’s principles of evolution do not themselves provide a complete explanation. Darwinism does not provide a complete theory of everything, from cells to human society. Instead, these principles are a kind of ‘meta-theory’, or an over-arching theoretical framework wherein theorists place particular explanations. Crucially, explanations additional to natural selection are always required to explain any evolved phenomenon... Selection is a general principle, but it operates in different ways in different contexts. Likewise, the general Darwinian principle of variation applies, but it does not itself explain how variation occurs.”<sup>428</sup> North is then certainly defensible when he proposes a theory of cultural evolution, but the use of evolutionary language is not sufficient on its own to constitute an appropriate explanation. We must keep this in mind when evaluating North’s particular theory of cultural evolution. Moreover, the precise operation of natural selection in a sociobiological context is also controversial, even if the *applicability in principle* is not, as we shall see further below.

The interesting properties of Darwinian selection, well explored in evolutionary biology since the post-war period, lend themselves to the ‘elective affinity’ with institutional economics discussed above. A discussion of this affinity must be undertaken with some care. As we have seen, the NIE is divided into more neoinstitutional and more ‘heterodox’ strands, and they have different definitions of institutions. Moreover, as Jack Vromen has pointed out, there is an important difference between those evolutionary approaches concentrating on the evolution of market competition in capitalist economies, as exemplified by Nelson and Winter’s work (but arguably also that of Veblen and Commons), and those approaches interested in the more sociobiological question of cultural evolution as a whole, within which economic evolution is a (central) special case, as exemplified by Hayek and North.<sup>429</sup> This division has much in common with the distinction between neoinstitutionalism and NIE(H) proposed by Eggertsson. In both cases, the former tends to take the given ‘allocation system’ for granted and focuses on competition and markets in modern economies, whereas the latter is interested in a more macro-level perspective. In the Hayekian tradition of North, this includes larger historical dynamics, potentially spanning various allocation systems and cultural-institutional frameworks. However, the analogy is not complete. Much of the neoinstitutional tradition itself, as exemplified by the work of Williamson, Coase, and others, still hews too closely to the neoclassical model of the agent as rational contractor. It allows only for change and learning in a static, probabilistic way. Although markets are imperfect and transaction costs persist in such models, they do not allow for the role of uncertainty, nor for understanding learning and innovation as ongoing processes of finding local efficiencies. This, in turn, means that they lack a natural affinity with the blindly searching algorithmic selection process that Darwinian natural selection represents.<sup>430</sup>

---

<sup>428</sup> Aldrich et al. 2008, p. 585.

<sup>429</sup> Vromen 1995.

<sup>430</sup> Hodgson 1998.



Keeping this in mind, the affinities between the generalised Darwinian approach and at least some institutionalisms are worth restating. Darwinian selection is blind, thereby not inherently requiring any particularly sophisticated rationality nor informational acquisition on the part of its ‘phenotypes’ and none at all on the part of its ‘genotypes’. Even where no obvious equivalent to genotypes and phenotypes exists, it can still operate just fine without requiring any intentionality on the part of the replicators, *pace* North, as long as the criteria of generalised Darwinism apply. This lends itself well to its use in contexts where less than perfect rationality is assumed, and where uncertainty prevails and information is scarce. Moreover, Darwinian selection searches only for *local* ‘efficiency’, and defines such efficiency purely in terms of differentiated survival – contrary to popular belief, no actual implications of improvement, progress, or ‘optimality’ are implied by natural selection. This makes for a natural fit with approaches to economics that emphasize the contingency and limits of efficiency in markets (or other economic modes of organisation), such as institutional economics. Finally, the core of modern evolutionary theory is the interaction between a population subject to selection pressures and its environment, and in an economic analysis this readily evokes the relation between economic agents and a social-institutional context.

North’s most recent works illustrate, as we have seen, all of these elective affinities. For this reason the evaluation of his work must have implications for assessing the utility of approaches in the general theory cluster one might call ‘institutional-evolutionary economics’ (involving both the Veblen-Alchian-Nelson/Winter school and the Hayek-North school), of which one could argue his NIEH is an important part.<sup>431</sup> This perspective can help us understand better North’s emphasis on the cognitive operation of learning processes beyond simple Bayesian probabilistic models and why this emphasis is combined in the same work with the introduction of evolutionary (if, according to North, ‘Lamarckian’) mechanisms into his theory.

### **What evolves?**

It is clear then that North’s approach belongs in a more ambitious and historical form of cultural evolutionary theory than the evolutionary study of firms and market competition. The latter school has, as we have seen, to some extent converged on identifying routines and/or habits as replicators and units of selection.<sup>432</sup> That is to say, this approach to Darwinian sociobiology uses David Hull’s influential division between ‘replicators’ (those entities that are stable and preserved in the hereditary process) and ‘interactors’ (the relevant units of the population subject to selection pressure) – the generalised theory of which the genotype-phenotype distinction is the biological special case.<sup>433</sup> In the case of the evolutionary theory of the firm, the routines act as

---

<sup>431</sup> The range of works potentially within this category is vast, also in light of its potential partial overlap with Austrian school economics. Without intending any thorough literature review, see (besides works already cited) e.g.: Hodgson 2006; Hodgson and Knudsen 2010; Hodgson 2007a; Witt (ed.) 1993; Witt 2003; Loasby 1999; Metcalfe 1998; Dopfer (ed.) 2005; Schubert and Von Wangenheim (eds.) 2006.

<sup>432</sup> Becker 2006, *op. cit.*

<sup>433</sup> Hull 1980; cf. Dawkins 1982.

replicators and the firms themselves constitute the population, as mentioned. The advantage of such an approach is that it contributes to clarifying one of the central problems in the application of generalised Darwinian theory to a specific sociocultural case. This is an ongoing debate within philosophy of biology, and involves a number of different problems in a given case: what the unit of selection is, what it is that is doing the adapting, what it is that is subject to variable survival (i.e. what the replicator is), and what is the ultimate agent that ‘benefits’ from ‘success’ in the process of natural selection.<sup>434</sup> I will not review all these here, although in returning to the potential contribution of sociobiology to historical theories of cultural evolution we will encounter them again.

For my purposes, what matters is that in the specific case of the evolution of the firm, the advantage of the identification of routines and the subsequent literature theorizing its evolutionary mechanisms is that it can to some extent successfully answer these questions in a precise way. Moreover, the answers can be linked directly to empirically observable phenomena in the economic entities and behaviours in question. While this does not guarantee the theory is the most fruitful description of competition and change between and within firms in a modern market economy, it does give the evolutionary approach sufficient grounding to be more plausible than a mere ‘just so’ story. But unlike the tradition of Nelson and Winter, Alchian, and others, North is not primarily interested in an evolutionary explanation of market processes and organisational change themselves. In fact, if my argument in the previous chapters is right, he is so to speak consistently inconsistent in whether to take a more neoinstitutional approach to this domain – focusing on rational pursuit of maximum payoffs when confronted with transaction costs – or to adjust, in some indefinite manner, the model to account for a different form of rationality. North’s evolutionary mechanism is not in fact located at the meso-level of the short to medium term operation of markets, hierarchies, and organisational and institutional structures within a given historical society’s economic formation. It is located at the micro-level, in the importance of cognition and other human genetic heritage as products of evolution; and more importantly, it is located at the macro-level, in the evolution and adaptation of whole institutional orders, indeed whole societies and cultures, over the long run. As he himself has indicated, it is the tradition of Hayek, not of Veblen or Nelson and Winter, that is the most important for North’s account.<sup>435</sup> This presents two daunting challenges. If, for the sake of argument, we accept the replicator-interactor distinction conventional in evolutionary economics as useful, at each level the relevance and plausibility of the evolutionary account then has to be justified in terms of the four questions posed above. Since these are more closely connected, despite their different scopes and levels of abstraction, they require an integrated evaluation. I will begin this by concentrating on the macro-level role of evolution in his account, since it is arguably the most

---

<sup>434</sup> Lloyd 2012; Sober and Wilson 1994. See the note on page 121 above.

<sup>435</sup> Hough and Grier 2015, p. 50-51; North 2005, p. 5, 31, 52; cf. North 1999.

controversial, but this soon brings us to problems in North's micro-level evolutionary account as well.<sup>436</sup>

A helpful start is to show the points of contrast between North's approach and that of Hayek. There are two major differences between Hayek's evolutionary account of spontaneous order and North's account of adaptive efficiency. The first is that for Hayek, the rationality and intentionality of individuals was given, and not in need of explanation. Culture in Hayek's view was strictly reducible to the cumulative outcome of the choices of individuals, and the nature of these choices as such was not in question.<sup>437</sup> In North, however, the problem of the formation of beliefs and their interactive dialectic with the human (institutional) environment is a central area of concern. Unlike the Austrian tradition of Hayek, North has no difficulties per se with admitting psychology into the model of individual behaviour. This leads to the second major difference, which is that for Hayek, the "selective evolution" of rules "by imitative learning" he identified was undertaken according to the requirement of the fitness of a particular individual or group adopting the rule.<sup>438</sup>

In the functionalist account that evolutionary explanations always entail, one can have an efficiency view such as this, or a non-efficiency view, in which the phenomenon in question is explained by its function within a larger system but without the assumption of this function contributing to the fitness of that system. The rational choice tradition and Hayek adopt the former view, and North, in the later stages of his work (where evolution makes its appearance) seems to reject such a view, given his overall rejection of efficiency explanations of institutions and his emphasis on beliefs and irrational elements of human intentionality.<sup>439</sup> These are easily interpretable as distancing his view of institutional evolution from the contradictions in Hayek's approach identified by Geoffrey Hodgson: namely between Hayek's functionalist and fitness-maximisation based explanation of institutions as products of conscious choice and his equally evolutionary appeal to the unconscious operation, via routines and habit, of inherited cultural norms and traditions.<sup>440</sup> These need not inherently be contradictory, but require a consistent explanation connecting the two. Hayek's refusal to specify how the psychology of the individual relates to a group, and how the relevant psychology or rationality of either is defined when 'choosing' institutions or when following habitual norms and routines, makes him – as Hodgson points out – incapable of explaining either the formation of institutions or the evolutionary process that supposedly generates spontaneous order.<sup>441</sup>

---

<sup>436</sup> For a critique of the very possibility of an evolutionary theory as an improvement on neoclassical economics, see: Rosenberg 2000, ch. 9.

<sup>437</sup> Hodgson 1994b, p. 411-412.

<sup>438</sup> *Ibid.*, p. 422; Hayek 1967, p. 67-68.

<sup>439</sup> Cf. Vanberg 1986, where Hayek is criticized precisely for excessive functionalism and for neglecting the centrality of individual intentionality.

<sup>440</sup> Hodgson 1994b, p. 423.

<sup>441</sup> *Ibid.*, p. 423-424.

Yet North's attempt at escaping these contradictions ends up generating analogous ones. Firstly, as I have argued in the previous chapter, North is inconsistent in his explanation of how individuals relate to institutions. Explanations in terms of self-interest clash with explanations in terms of potentially irrational belief systems, and simultaneously game theoretical models relying on fairly strict criteria of common rationality and beliefs are joined with explanations in terms of indeterminate and undefined constraints on individual choice sets. In light of Hodgson's critique of Hayek's evolutionary theory, North's inconsistency in models of rationality can now also be better understood. Tellingly, in North the problem of efficiency in functionalist explanation also recurs. While North has argued for a long time against the neoinstitutional assumption he once shared, that institutions are necessarily economically efficient or the result of the pursuit of gains in a market-like setting (although without achieving full consistency in his rejection of this view), he nonetheless identifies an ideal type of society (the open access order) which has maximally adaptive properties, and so is both economically and evolutionarily 'efficient' in that sense (although he occasionally admits this is difficult to define<sup>442</sup>).

It is no coincidence that North's most frequent references to Hayek are just on this point. Hodgson points out that for Hayek's evolutionary free market argument to work, there must be evolutionary selection by means of the market, but *not* between the market and other economic forms, for in that case the market itself as an allocation system is no longer 'guaranteed' to be the most adaptive process in a given case. Hayek (rightly) denies on the basis of the blind nature of cultural evolution that one could determine one society to be 'good' as opposed to another, but simultaneously argues that the ideal type classical liberal society (which Hodgson, after Hayek, dubs the 'Great Society' vision) is the yardstick of evolutionary progress.<sup>443</sup> North engages in a similar double denial. While rejecting simplistic 'free market' arguments as insufficiently mindful of the suboptimality of most institutions and beliefs and the difficulty of moving away from such low level equilibrium traps, he *does* maintain that the ideal type classical liberal society – not a Great Society but an open access order – is objectively a society with the highest level of evolutionary fitness: it is most *adaptively* efficient. But by what criterion could North decide this? He asserts this, but can only justify it on the basis of the claim that market competition (in the economic and political sphere) equates to evolutionary competition. As with Hayek, this claim is not itself defended, nor is the possibility of evolutionary selection between markets and other arrangements permitted to play a significant role: at this level, as I have argued, North suddenly falls back onto ambiguous game theoretical argumentation.

I submit that only by violating his own assumption of a *real* evolutionary process in cultural evolution - which is blind, amoral, and knows only local optima - and substituting for it a *metaphorical* evolutionary process of the Victorian classical liberal (Spencerian) type can North defend this view. It is rather striking then that the maximally adaptive society turns out to

---

<sup>442</sup> North 2005, p. 162.

<sup>443</sup> Hodgson 1994b, p. 436.

be that of the contemporary Western world. North's claim regarding the Western world can not be saved by the argument that economic growth – serving as a measure of 'economic performance' – is the relevant measure of fitness for competing access orders, for economic growth in the Western world has been laggard and declining for decades in contrast to many societies described as 'natural states'.

Although there is no space here to pursue the argument systematically, the Eurocentric dimensions of North's historical account become clear when one considers how the achievement of the open access order is entirely credited to transformations internal to a small number of European (and European settler) societies. The impressive growth rates relative to these societies of e.g. the Soviet Union throughout much of its existence, or of the Asian tigers in more recent years, and the acknowledged irrelevance of economic growth at all in most of economic history; none of these prevent North from resuscitating a rather old-fashioned account of the 'rise of the West', be it now in terms of the evolution of adaptive efficiency.<sup>444</sup> But competition is not per se a measure of economic growth, economic growth is not an obvious analogue of biological fitness, and North's argument does not give us reason to believe in these identifications and analogies, for he never explicitly defends them. It is therefore hard to take this view seriously as more than a historically oriented justification of Whig (if not quite *laissez faire*) politics and of a development economics oriented to imitating Western institutional orders.

North differs strongly from Hayek in not believing that the price system itself is a sufficient mechanism to achieve the idealized adaptivity of the 'Great Society'; as he frequently avers, he differs with Hayek in that North sees 'social engineering' as necessary and therefore changes in institutions, political structures, and even belief systems as necessary for making societies more adaptively efficient (something Hayek denied was either possible or desirable).<sup>445</sup> But this difference, despite initial appearances and North's own rhetoric, turns out to be mostly a difference of preferred policy. The problems and inconsistencies in their respective theories of cultural evolution are fundamentally similar. Compare this to Hodgson's description of Hayek: "While Hayek rejects the suggestion that evolution automatically leads to progress, he also has a clear criterion by which advance may be judged: to the extent that rules consistent with the Great Society emerge, function, and overcome the assumed atavistic and collectivist instincts of humankind, then progress is deemed to be made."<sup>446</sup>

North, contrary to Hayek, is not so negative about the 'collectivist instincts', seeing them as indispensable: even suboptimal, rent-seeking political institutions and structures are better than none, in the face of uncertainty and pervasive violence. Moreover, North continuously

---

<sup>444</sup> Indeed, while North points to an alleged relative neglect of non-Western societies by economic historians, he argues for their inclusion in the narrative mainly in terms of their "arrested development": North 2005, p. 89. See also North, Wallis, and Weingast 2009, p. 241-243.

<sup>445</sup> E.g. North 2005, p 162-163.

<sup>446</sup> Hodgson 1994b, p. 436.

stresses that they are the results of the choices of intentional individuals and organisations whose choice and learning processes are themselves evolved; so that, contra Hayek, the possibility of an evolutionary explanation of institutional orders inherently requires an acceptance of the malleability of societies. This I think is what North means when he says we cannot do without social engineering, and also what he might mean by contrasting intentional evolution with Darwinian evolution.<sup>447</sup> But he is for that reason all the more worried about how difficult it is for them to make the desired form of social cooperation, a competitive market order, possible. Humans are certainly often enough in North's account far from the self-interested rational actors he criticizes in neoclassical 'anthropologies', but North is equally uncertain about whether they can really be sufficiently innately cooperative or even altruistic to make enduring large-scale cooperation and reduction of violence a likely result. So North hovers uneasily in between, condemned to a 'pessimistic anthropology' and unstable contractarian solutions. He is more aware of, and more worried by, the possibility of collapse and disorder or enduring illiberal and 'inefficient' economic and political systems than Hayek, and this colours his approach.<sup>448</sup>

The source of this problem is therefore not just the ideological and political component of classical liberalism and 'Smithian results' in North's work. As Hodgson rightly points out in the case of Hayek, this ideological argument is buttressed by a lack of specification of the required evolutionary mechanisms themselves: for example, in terms of answers to the four questions about replication, competition, and selection described above. In North's account, societies as such seem to appear and disappear depending on their adaptive efficiency in the longer run, as illustrated by his account of the decline and fall of the Soviet Union.<sup>449</sup> In this argument, whole institutional orders/artifactual frameworks are interactors, in some senses analogous to phenotypes, whereas North appears to suggest that their success or failure depends on the "evolution of beliefs", which would make the latter the replicators.<sup>450</sup>

But when we ask further questions, troubling ambiguities creep in. In the classic account of the ultimate beneficiaries of evolution supported by the work of Richard Dawkins, the salient fact of evolution is that those lineages which differentially survive and whose causal effect on the nature and continuity of further generations are the beneficiaries of evolution.<sup>451</sup> In the Northian case, this would mean we should expect in the Soviet case that as replicators, the ultimate beneficiaries of the selection process should be lineages of beliefs, ones that would give rise to subsequent 'types' of societies. North is quite sceptical of the idea that beliefs will change easily or rapidly in the face of changing environments.<sup>452</sup> We would therefore expect North to argue that the process of cultural evolution he describes is ultimately about the variable

---

<sup>447</sup> North 2005, p. 162, viii.

<sup>448</sup> Cf. Hodgson 1994b, p. 433.

<sup>449</sup> North 2005, ch. 11.

<sup>450</sup> *Ibid.*, p. 146.

<sup>451</sup> Hampe and Morgan 1988.

<sup>452</sup> E.g. North 2005, p. 62, 108; Hahn 1987, p. 324.

frequency of particular belief systems among the larger ‘population’ of societies or economic systems. But for North, the point of the Soviet illustration is to show that the Soviet Union was too internally rigid to adjust to the ‘environment’ when it changed: “Adaptive efficiency entails an institutional structure that in the face of the ubiquitous uncertainties of a non-ergodic world will flexibly try various alternatives to deal with novel problems that continue to emerge over time. In turn this institutional structure entails *a belief structure that will encourage and permit experimentation and equally will wipe out failures*. The Soviet Union represented the very antithesis of such an approach.”<sup>453</sup>

This does not suggest that the differentiating frequency of belief lineages throughout human history serves as the measure of evolution, i.e. that surviving belief lineages are the ultimate beneficiaries of evolution – as would be implied by his discussion of the ‘evolution of beliefs’ and how the survival of whole societies depend on it. Rather, this suggests that in addition, there is a different selection process going on where *internal* “failures” are “wiped out”, supported by a belief system that allows such selection to take place: here belief systems are not subject to selection, but determine *whether selection takes place*. This implies (although does not explicitly state) that we return rather to the Hayekian view of market competition as a process of selection, as is also suggested by North and colleagues’ discussion of the advantages of open access orders.<sup>454</sup> In such a selection process, the entrepreneurial firms, organisations, or individual agents or whatever the relevant competing units may be in political or economic markets are the interactors, and the replicators the societies (i.e. integrated artifactual frameworks of institutions and beliefs) that survive or ‘die’ by the ability of such ‘entrepreneurs’ to adjust to changing circumstances in well-functioning markets.

The point here is not to resolve this conundrum, but rather to point out that North is conflating two different levels at which selection takes place. Each has different units of selection and different replicators and interactors implied in the argument. North does not appear to be clearly aware of the importance of such identifications to the success of an evolutionary description of institutional change. It is certainly forgivable that he says little about the internal structure of organisations in economic markets (although more about political ones) and their role as replicators within the evolution of such organisations, as this is well covered by extant literature in evolutionary economics. It is less forgivable that he should, for the sake of joining a Hayekian argument about markets to a less Hayekian argument about the importance of beliefs and instituted orders so conflate various elements and levels of an evolutionary mechanism. But this is not yet all. For the foundation of each of these mechanisms, whether mutually compatible or not, is the evolved human individual and their ability or inability to achieve cooperation on a larger scale: the very possibility of a human ‘order’ under conditions of pervasive uncertainty, limited information, and the prisoner’s dilemma’s pessimistic implications for the achievement

---

<sup>453</sup> North 2005, p. 154.

<sup>454</sup> North, Wallis, and Weingast 2009, p. 146, 252.

of cooperation on the basis of self-interest. This is arguably the central problematic underlying *Violence and Social Orders*, even if most the discussion in that work concentrates on the operation of states and markets and their interaction, rather than on the origins of human sociality or cooperation per se.

This is where the link between North's Hayekian macro-level evolutionary theory and the implications of sociobiology must be made. For Hodgson, if we want to take the evolutionary account of the emergence and change of allocation systems seriously, we must see 'the market' itself (in all its varieties) as one of the possible outcomes of the process of selection, and as itself subject to whatever selection processes may operate in cultural evolution.<sup>455</sup> This point, already alluded to above, moreover implies that we must also be able to account for the origins of such a system in terms of its evolutionary ancestors, and explain how those ancestral forms could and did lead to those particular outcomes. Such reasoning is the bread and butter of much ethology today, and is equally important in applications of evolutionary theory to social life. This means we must be able to trace North's evolutionary subjects or agents at the macro- and meso-level - crudely put, societies and markets respectively - to their own origins in a consistent way. We must move now to evolution at the micro-level.

There may be no appeal to original creation: the story must be Darwinian all the way through. North seems to realize this insofar as he tries to derive institutional orders from beliefs that generate them, and equally market processes from institutional orders. The ambiguous "complex blend" of an individual's beliefs and interests I have discussed in chapter 3, jointly constituting North's model of rationality of the individual agent insofar as they determine or explain behaviour, is in this sense also the evolutionary microfoundation of North's NIEH theory. But it itself must be explainable in evolutionary terms, and all the problems of North's approach to rationality threaten to recur if he does not have an adequate account at this micro-level either. We must therefore now, finally but perhaps most fundamentally, take a critical look at how North treats the origins of human sociality (cooperation, altruism, and selfishness combined) in the context of cultural evolution/sociobiology, and how he selects from the arguments and evidence in this literature the model of human rationality - both in analytical isolation and interaction - he considers most appropriate.

### **Interpreting North's comments on sociobiology and cooperation**

I will argue here that North's account has the merit of rightly identifying one of the central problems in the sociobiological literature on the evolution of cooperation among humans, namely how such cooperation is possible and sustainable in larger groups where most will not be kin or even regularly interact, given limited information. North, I think rightly, seems to suspect, although this remains rather implicit, that this problem requires a solution in a better understanding of the conditions for the evolution of altruism and other-regarding behaviours and

---

<sup>455</sup> Hodgson 1994b, p. 431.



attitudes among humans. He also is keen, for example in his discussions of the evolution of cognition, to take human individuals seriously as evolved organisms and to examine the implications of this for the nature of human cooperation.

However, North does not develop this point particularly far. He does not pay sufficient attention to the implications this debate has, or rather should have, for his view of individual rationality and for his contractarian account of institutions. In the contractarian tradition he remains unduly pessimistic about the ability of cooperation to emerge in other ways than through instrumental reason, at least in larger groups, and views cooperation in a Hobbesian framework as dependent on power and institutions that will restrain violence and self-interest even within given societies or groups. As we have seen in my discussion of alternative models of rationality in the previous chapter, while his sophisticated approach to beliefs is laudable, his approach to preferences is more conventional. A rounded evolutionary picture also requires a re-examination of preferences. I will argue that the revisions to North's view of rationality I have defended in the previous chapter can better be reconciled with recent literature on the evolution of cooperation and altruism than North's own approach can – or rather that those parts of North's account that still assume individual rationality to be fundamentally self-regarding, non-altruistic, and only instrumentally cooperative be jettisoned in favour of those incipient tendencies in his work that take altruism and cooperation to be more fundamental. However, in order not to explain away the problem of cooperation, one must then place an explanation of such cooperative tendencies on a strong evolutionary and empirical footing.

Additionally, North's inadequate larger-level evolutionary theories prevent the necessary joining of such a macro-level theory of cultural evolution to the sociobiological foundations I believe the recent literature makes possible. To round off the substantive and to some extent immanent critique of North's work of the middle three chapters of this work, then, I will propose that the sociobiological school of gene-culture coevolution can show how such an integration between micro- and macro-level evolution, based on multi-level selection theory, might be possible in a more plausible manner. This would not quite form a theory of cultural evolution in its own right, a school still much in its infancy despite great advances in sociobiology over the last three decades or so; but it would point the way to how one might get past the aporias generated by North's muddled treatment of the units of selection problem and his Hayekian view of markets and competition.

I can here only point in the direction of an alternative approach to the problem of cooperation to the one North uses, partially complementary to *some* (but not all) of his statements on the subject. In a larger sense this would require the integration or application of two literatures. One is the empirical literature on cooperation and altruism from behavioural economics, often based on experiments in testing the predictions of conventional game theory.<sup>456</sup>

---

<sup>456</sup> E.g. Henrich et al. (eds.) 2004; Gintis et al. (eds.) 2005; Bowles and Gintis 2011; Fehr and Gächter 2000, 2002; Fischbacher, Gächter and Fehr 2001; Camerer and Thaler 1995.

The other is the sociobiological literature, often reliant on *evolutionary* game theory, on the origins and nature of altruism and cooperation in the human species.<sup>457</sup> As in the previous discussions, I intend no general overview of sociobiology nor of experimental economics, and the literatures in both areas have, in the three or four decades of their existence, already grown much too large to do even a cursory justice. All I can do here for now is sketch a discussion, prompted by North's own comments on the origins of cooperation and its implications for his work; as Hodgson and Knudsen point out, the application of generalized Darwinism is still very much a work in progress and depends on particular case studies. It is the overarching theory, the foundations, that have become increasingly well established.<sup>458</sup> Describing where North's work is a step in the right direction from this meta-theoretical viewpoint, and what larger theories might aid it where it goes wrong, is just one small contribution to this project, and thereby to its application to economic history.

North's theory of the evolution of cooperation can be summarized fairly effectively without too much loss of relevant detail as follows. For North, as pointed out before, Hoffman, McCabe and Smith's 1998 summary of experimental economics demonstrates that in small groups cooperation is relatively easily achieved because of the adjustment to conventional self-interested rational action in such settings. However, in large groups or under private information, such cooperation is not the game theoretically plausible outcome.<sup>459</sup> This constitutes, by my knowledge, one of only two occasions (the other I will discuss shortly) in any of his theoretical work that North explicitly discusses experimental economic evidence testing game theoretical specifications of rational choice. But this is not his only observation on the subject in general. On several occasions, North refers also to an *innate* ability or tendency of humans to cooperate in small groups. This and the reference to evolutionary psychology following the discussion of Hoffman et al. suggests that North's explanation of individuals' "modification of their strict self-interest behavior" is based on an evolutionary argument: humans as evolved to cooperate in such (and only such) settings. His repeated references to this innate small group cooperation makes it clear that this innate cooperativeness, however, is only relevant to such cases.<sup>460</sup> In larger settings, however, North seems to suggest that institutions, beliefs, and incentives, i.e. more conventional (non-evolutionary) models of rationality plus the normative power of culture, must do the job.<sup>461</sup> As an illustration of the latter kind of argument we need but observe the claim in *Violence and Social Orders* that, in the absence of third party enforcement, "cooperation by an

---

<sup>457</sup> This literature is vast and goes back to Darwin himself. Some works I have drawn on for the present discussion include: Lumsden and Wilson 1981; Sober and Wilson 1998; Boyd and Richerson 1985, 2005; Richerson and Boyd 2005; Field 2001; Hodgson 2013; Sterelny et al. (eds.) 2013.

<sup>458</sup> Hodgson and Knudsen 2010, p. 3.

<sup>459</sup> North 2005, p. 28-29.

<sup>460</sup> *Ibid.*, p. viii-ix, 46;

<sup>461</sup> *Ibid.*, p. 30, ix.

adherent organization's members must be, at every point in time, incentive-compatible for all members".<sup>462</sup>

Finally, North also makes a single but important observation about the implications of experimental economics in testing the rationality assumptions formalized in game theory, a little after his discussion of Hoffman et al.'s findings. Here he twice cites verbatim the results of the work of Joseph Henrich and his colleagues, who undertook experiments across a wide range of small scale societies with very different cultures to measure empirically the degree of correspondence to predicted utility maximizing behaviour in game theoretical settings. (This is the other such discussion in North's theoretical work I mentioned.) I will not cite these in their entirety again here, but the upshot is as follows: "the canonical model of the self-interested material payoff-maximizing actor is systematically violated. In all societies studied, [ultimatum game] offers are strictly positive and often substantially in excess of the expected income-maximizing offer, as are contributions in the public-goods game, while rejections of positive offers in some societies occur at a considerable rate. Second, preferences over economic choices are not exogenous as the canonical model would have it, but rather are shaped by the economic and social interactions of everyday life."<sup>463</sup>

As intimated in the previous chapter, empirical violations of canonical rationality in a systematic way suggest the need for a revision of the specification of rational action involved in such choice behaviour; if not in terms of some form of reconsideration of utility maximization, then at least its realignment to allow for more other-regarding preferences and utility beyond maximizing money returns (or indeed any returns) than is normally allowed. This should also change the way incentives and interests are represented in more general models of (economic) cooperation, something suggested by Henrich et al. themselves, but which we have seen North does not consistently do. Expanding the discussion about alternative rationality models of chapter 3 to our present concern with North's explanations of cooperative behaviour, we must note they also have consequences – as a corollary of the above – for the contractarian perspective of how cooperation can (and cannot) be achieved through conventional incentive models. It is therefore all the more striking that these conclusions do not seem to warrant any further discussion on North's part; he simply concludes they illustrate his point about the importance of learning and beliefs, and leaves it at that.<sup>464</sup>

Given this rather feeble treatment of the problem of cooperation, whether seen from behavioural economics or sociobiology, on the part of North, I will now try and sketch how the literature in these fields can present an alternative reading of this problem. As mentioned, this entails no more than a brief further exploration of the implications of research in these two domains than North has engaged in: i.e. the testing of the behavioural assumptions underlying

---

<sup>462</sup> North, Wallis, and Weingast 2009, p. 16.

<sup>463</sup> Henrich et al. 2003, p. 40, cited in: North 2005, p. 46-47.

<sup>464</sup> North 2005, p. 47.

conventional utility maximization and the evolutionary (especially evolutionary game theoretical) literature on other-regarding preferences and altruism, their evolutionary origins, and implications for human rationality. Finally, I will suggest that the most obvious way to integrate these perspectives into a theory of long-term cultural evolution in the sense that North has proposed, the most promising approach is not evolutionary psychology but gene-culture coevolution. North dismisses evolutionary psychology on the grounds of vast observed cultural variation<sup>465</sup>. Whether or not this is a fatal objection to the school of evolutionary psychology most associated with the work of John Tooby and Lana Cosmides, it is certainly not a problem for gene-culture coevolution-based approaches.<sup>466</sup> I think that this reappraisal is in principle very compatible with some of North's own lines of argument in his NIEH: particularly North's own desire to formulate an evolutionary theory in which institutions and cultural traits matter, as well as those instances where he is most critical of rational choice approaches and searches for alternatives beyond even Simonian 'bounded rationality'. But it does, to my mind, also imply jettisoning some of North's more Hayekian notions, as well as distancing us further from those aspects of North's work closer to neoinstitutional and self-regarding models of choice behaviour.

The central question here is: based on the discussions in *UPEC* and *Violence and Social Orders*, is North right to take as premise for his evolutionary account that humans have evolved to adjust their self-regarding rationality to allow small-group cooperation; but that in larger groups the free rider problem combined with the pervasiveness of violence will make cooperation (or 'order') only possible by means of a mixed Hobbesian-Lockean social contract, i.e. by means of the formation of elites with control over violence who distribute rents to create incentives for obedience? I submit that he is not, and for two reasons. The first is that based on recent literature in evolutionary game theory, there are strong reasons to suspect that no evolution of human sociality would have been possible at all – even in realistic small group settings - under 'contractarian' premises, that is to say, on the basis of evolutionary incentives to maximize fitness evolving only self-regarding behaviour. Therefore, the only possible evolutionary explanation for the observed reality of human cooperation and sociality in small and large groups is that at least to a significant extent humans have evolved as being from our earliest behaviourally modern beginnings as, in the words of Herbert Gintis and Samuel Bowles, 'a cooperative species'.<sup>467</sup> This implies that the contractarian perspective, in which human cooperation is explained on the basis of the interests of self-regarding actors to cooperate given correct incentives, is in all its forms implausible or at least radically incomplete.

Secondly, the results of game theoretical experiments demonstrate, as North himself briefly acknowledged, that the behavioural predictions of self-regarding utility maximization are not borne out in reality. Unlike North, however, I believe that these results significantly affect the plausibility of a (primarily) incentives-based account of the origins of institutions. While North

---

<sup>465</sup> *Ibid.*, p. 28-29.

<sup>466</sup> A useful comparison of different sociobiological schools, including these, is Laland and Brown 2002.

<sup>467</sup> Bowles and Gintis 2011. Cf. Hodgson 2013.

also modifies this account substantially by ascribing a major role to the formation of beliefs, he gives no consistent evolutionary account of how such beliefs interact with Darwinian evolutionary pressures or with the operation of ‘rational’ behaviour. In chapter 3, I have suggested alternative models of rationality that could incorporate other-regarding behaviour and group identification in more plausible ways. Here, I will suggest that such models can be more readily integrated into an evolutionary account of long-term change that *takes culture seriously* than North’s own reliance on incentives and transaction costs. Put briefly, I think gene-culture coevolution and group rationality is more promising as an account of institutions and their sociobiological role than North’s reliance on exogenously given beliefs and New Institutional models of bounded rationality.

I do not wish to suggest that they are wholly mutually exclusive, for many of North’s theoretical tools remain useful: this is no argument against his definition of institutions, his claim that they matter for economic history, or his use of evolutionary concepts or path dependence, say. For a Northian evolutionary account to succeed, however, they need complementing by a gene-culture coevolution approach and a much closer attention to the functional role of ‘culture’ (in the broad sense of institutions, self-identifications, and cultural markers) in an evolutionary process of inheritance as well as in intentional behaviour. Jointly, these two claims suggest that we need to greatly expand rationality to include other-regarding forms of utility maximization as well as the role of emotions, identities, and other non-maximising aspects of choice behaviour, and that we must go beyond contractarian accounts of the origins of social institutions to a more rounded perspective of humans as ‘always already’ cooperative. Finally, cooperation – even altruistic – is not itself sufficient. The *content* of culture, including the role of commitment and its moral and emotional component, also matters, and this can perhaps lend a new perspective to North’s intuitions about the role of beliefs. To put it rhetorically, a Northian type evolutionary account needs more Hume and Darwin and less Locke and Coase. I will now expand on each of these arguments in order.

### **Social contract or cooperative species?**

As Alexander Field has pointed out in various critiques of New Institutionalism and similar approaches, any explanation of the emergence of institutions in terms of conventional models of rationality runs into serious problems of infinite regress. One can either explain institutions in terms of previous institutions, in which case these need explaining in turn. Or one can turn, as in the contractarian tradition, to a ‘state of nature’ in which the interaction between individuals will cause, spontaneously, the emergence of a particular original institutional order.<sup>468</sup> The latter is how I have interpreted, *in part*, the approach of North to the origins of institutions. In *Violence and Social Orders*, any society grown beyond the ‘small scale’ hunter-forager society will likely turn into a natural order. From the ‘foraging order’ where each individual knows all others, and

---

<sup>468</sup> Field 1979; Field 1981; Field 2001.

to which presumably Hoffman et al.'s findings apply, to the natural order is "a long process of development". Based on the work of Allan Johnson and Timothy Earle, North and colleagues argue that the transition from foraging orders to natural state orders takes place by a sort of Hobbesian process where 'chieftains' start controlling violence, a solution based on "the threat of coercion" and "by mutual interests".<sup>469</sup> The combination of coercion and incentives is what I mean by a Hobbesian-Lockean contractarian theory, and it is expressed, as I have argued in chapter 3, by North's reliance on game theoretical language (though not formalization) to suggest the operation of this kind of incentives in the formation of institutions.

Yet crucially, conventional game theory and state of nature contractarianism also fall afoul of Field's objections. As I have argued, one cannot have any conventional game theoretical presentation of the operation of individual interaction without already knowing the rules and payoffs applicable to that game. And these must therefore be given in advance. As Field points out, what this amounts to is that contractarian approaches to institutions, popular in economic theory, effectively smuggle the very institutions they seek to explain in through the back door, often implicitly.<sup>470</sup> The infinite regress remains in operation. For this reason, New Institutional approaches must depart, as some of its theorists have recognized, from a given combination of institutions and individuals; it cannot explain the origins of human sociality per se.<sup>471</sup> Indeed, no social theory can do so except by appeal to evolutionary models of the emergence of culture in behaviourally modern *homo sapiens*; and therefore any model of the path dependent inheritance of such institutions (and all 'culture' more broadly) must begin with well formulated evolutionary premises rather than contractarian assumptions about the emergence of institutions from incentives, including incentives to reduce violence.

Recent literature on the evolution of human cooperation has done just that, by studying the emergence of altruism in evolutionary game theoretical models. I will not review the methodology of evolutionary game theory here, but it is important to note the distinction between conventional game theory and evolutionary game theory in one respect. While the formal aspects are much the same, the difference is that where conventional game theory assumes *payoff maximization* on the part of individuals, and is therefore subject to critiques of the plausibility of this assumption based on Field's arguments as well as the testing of this assumption (described below), evolutionary game theory is based on *fitness maximization* assumptions in natural selection. This assumption is much more robust, since it is supported by a vast amount of empirical evidence from the field in evolutionary biology, and it is a central component of the dominant explanatory approach in ethology, which has been called 'adaptationism'.<sup>472</sup> Evolutionary game theoretical models therefore have the distinct advantage

---

<sup>469</sup> North, Wallis, and Weingast 2009, p. 52-53; Johnson and Earle 2000; Earle 1997.

<sup>470</sup> Field 1981; cf. Hodgson 2013, p. 38.

<sup>471</sup> Aoki 2001.

<sup>472</sup> See Dennett 1995 for a discussion; see also Williams 1966.

that their maximization assumptions give a solid explanatory grounding of a kind that conventional game theory, reliant as it is on rational choice analysis, does not.<sup>473</sup>

Now to return to altruism. Altruism is here defined as behaviour by an individual of a species enhancing the fitness of another individual at a fitness cost to the first individual.<sup>474</sup> The importance of altruism and prosocial dispositions in explaining human cooperative sociality is enormous, because (as the prisoner's dilemma would demonstrate) under normal circumstances a strong selection mechanism like natural selection would be expected to select against behaviour that favours others at the expense of oneself. If helping others at the expense of oneself by definition diminishes fitness, it will diminish *ceteris paribus* the likelihood of procreation, and therefore genes favouring such behaviour should not survive. Over evolutionary time, prosocial behaviour should not occur. For this reason, E.O. Wilson has called altruism "the central theoretical problem of sociobiology".<sup>475</sup>

In sociobiology, for a considerable time the most common explanation of the emergence of cooperation was twofold: 'inclusive fitness', also called kin altruism, and reciprocal altruism. Kin altruism as a concept was developed by William Hamilton to explain the altruistic behaviour of many organisms observed in nature. Hamilton pointed to the fact that maximization of genetic fitness would favour altruistic behaviour by individuals towards kin, i.e. other individuals that carry substantially similar genes. Even if the individual with altruistic genes is selected against, the fitness benefits accruing to the kin may outweigh this and in this manner selection may permit the emergence of altruistic behaviour within kin groups as an evolutionarily stable strategy.<sup>476</sup> Reciprocal altruism, in turn, is based on an idea similar to the tit-for-tat strategy discussed in chapter 3: namely, that in indefinitely repeated games cooperation rather than defection can be a winning move, as long as cooperation can be expected from the other player. To simplify somewhat, if the probability of reciprocation is high enough, selection will favour genetic makeup that behaviourally expresses itself as altruistic, since cooperating individuals in a species will then be cooperated with down the line, and so outcompete the selfish defectors who cannot expect assistance at a later point.<sup>477</sup>

As explanations of human cooperative sociality, however, these have not proven to be adequate. Kin altruism does not suffice because even in small bands of hunter-gatherer groups, groups are too large to sustain it: in groups over 10 or so people, most people will not be sufficiently genetically related to each other, and migration among humans has always been too common.<sup>478</sup> Therefore, reciprocal altruism, and later refinements of the idea such as reputation-

---

<sup>473</sup> See e.g. Gintis 2000.

<sup>474</sup> Sober and Wilson 1998, p. 17.

<sup>475</sup> Wilson 1975, p. 3.

<sup>476</sup> Hamilton 1964; cf. Maynard Smith 1964. In fact, the roots of this idea go back to much before: see Haldane 1932; Fisher 1930.

<sup>477</sup> Trivers 1971; Axelrod 1984.

<sup>478</sup> E.g. Boyd and Richerson 1988.

building (known as indirect reciprocity), have proven popular as explanations because they seemed the least bad of the possible alternatives: they fit well within the parameters of evolutionary game theory, and did not require departures from the ‘selfish gene’ model of natural selection, like group selection arguments did.<sup>479</sup> As at the time group selection was consigned to the scientific oblivion of heterodoxy, this was a strike in favour of reciprocal altruism. Indeed, evolutionary game theory and reciprocal altruism were often presented as alternatives to group selection, because in evolutionary game theory it was thought selection on individual phenotypes in a population sufficed to explain cooperation and no group selection arguments were needed. Elliott Sober and David Sloan Wilson even argue that the emphasis on using ‘cooperation’ rather than ‘altruism’ as the term for the relevant *explanandum* of human sociality is because the former sounds more contractarian, more a matter of self-interest.<sup>480</sup> Similarly, Anatol Rapoport, who invented the tit-for-tat strategy, has commented on how odd its interpretation as a matter of self-interest appeared to him.<sup>481</sup>

However that may be, group or multilevel selection has since made a comeback, not least due to the efforts of Sober and Wilson themselves.<sup>482</sup> Indeed, it has become an important component of more sophisticated approaches in contemporary evolutionary game theory – as Sober and Wilson point out, the difference is “a matter of perspective, not process”.<sup>483</sup> Moreover, it is a central component of gene-culture coevolutionary approaches to cultural evolution, as I will describe below. It is therefore indispensable for contemporary evolutionary alternatives to the contractarian-institutional approach. So what is multilevel selection theory, and how does it matter? Owing its origins to George Price’s equations describing the interaction of selection pressures at different levels of populations, multilevel selection describes the insight that selection does not just take place between genes in an individual (meiosis aside) and between individuals in a population, but also between groups (sub-populations) in a global population.<sup>484</sup> This higher order selection among groups affects selection within groups and vice versa, and their relative strengths depend on the relevant specifications of the model. A group is simply defined as “a set of individuals that influence each other’s fitness with respect to a certain trait but not the fitness of those outside the group”.<sup>485</sup> However, since within-individual selection is extremely limited due to the already ‘cooperative’ nature of our genome, the relevant units are generally individuals in a population or group and groups in a population, allowing populations themselves to evolve in that sense.<sup>486</sup>

---

<sup>479</sup> Sober and Wilson 1998, p. 55-67.

<sup>480</sup> *Ibid.*, p. 83-84. Dawkins 1982, p. 55 tellingly even speaks of ‘mutual exploitation’. See also chapter 3.

<sup>481</sup> Rapoport 1991, p. 92-93.

<sup>482</sup> I will use them interchangeably, as Sober and Wilson 1994 do; however, ‘multilevel selection’ is often favoured because it clarifies that selection takes place at all levels, not *just* group level.

<sup>483</sup> Sober and Wilson 1998, p. 86.

<sup>484</sup> *Ibid.*, p. 73-74.

<sup>485</sup> *Ibid.*, p. 92.

<sup>486</sup> *Ibid.*, p. 96-97.



The significance of this insight rests in the observation that selection pressures between groups can outweigh selection pressures within groups under particular circumstances. This means, in turn, that as long as groups are relatively stable (*vis-à-vis* migration, for example), certain traits – such as altruistic ones – can be evolutionarily successful in particular groups by their effects on that group’s fitness relative to other groups. This provides an obvious way to think about the advantages of e.g. altruistic sacrifice by individuals of a group for the benefit of that group’s competitiveness with other groups in the (relevant) global population of a species. This behaviour is observable, and therefore in need of explanation, in a wide range of species from ants to humans. However, this is not sufficient to explain altruism. After all, ‘selfish gene’ selection also still takes place, and even if group selection can favour altruism by its ability to enhance competitiveness against other groups, over the long run in-group competition would select against altruism and it would disappear, unless it could be shown to be stable at the level of individual selection also. As Sober and Wilson point out: “If groups remain isolated from each other, the global increase in the frequency of altruists will be transient and individual selection will ultimately run its course within each group. Altruistic groups must somehow export their progeny to other portions of the landscape for altruism to evolve.”<sup>487</sup> The classic argument for the evolution of altruism *within* groups is then, as we have seen, reciprocal altruism.

But reciprocal altruism has serious difficulties of its own: there is little evidence of its operation among any other species than humans.<sup>488</sup> This would point to the need for additional hypotheses in the human case. Moreover, in larger scale human societies, or if information is private and imperfect, or if migration levels are high, it is unlikely that reciprocal altruism can be evolutionarily stable as an explanation of cooperation, as the likelihood of repeat encounters with the same player would be low, fitness costs would be high relative to benefits, and it is sensitive to error rates. So a community of reciprocal altruists would be vulnerable to ‘invasion’ by rare individuals with genes ‘coded for’ defection.<sup>489</sup> It is therefore by no means clear that reciprocal altruism suffices as an explanation for observed behaviour, except if there is already strong assortment between reciprocal altruists (i.e. they form groups that interact more often with one another), just as kin altruism depends on assortment.<sup>490</sup> Finally, it is unclear how altruistic reciprocal altruism really is. As Bowles and Gintis point out, reciprocal altruism is often really “self-interest with a long time horizon”.<sup>491</sup> So while it therefore may account in part for cooperative behaviour, it is not a sufficient explanation for the many kinds of altruistic and self-sacrificing behaviours observed among humans, although it may be a useful tool when seen in the light of anthropology (about which more in the final chapter).<sup>492</sup>

---

<sup>487</sup> *Ibid.*, p. 95.

<sup>488</sup> Hammerstein 2003. However, see Carter and Wilkinson 2013 for a contrary view concerning vampire bats.

<sup>489</sup> Bowles and Gintis 2011, p. 63-70.

<sup>490</sup> Fletcher and Doebeli 2009.

<sup>491</sup> Bowles and Gintis 2011, p. 52.

<sup>492</sup> E.g. Moore 1984.

One solution to this problem has been a focus on indirect altruism, where reputation-building plays the main role. In such cases, long-term cooperative benefits may result from ‘reciprocal’ altruism even when individuals do not interact repeatedly with the same other individuals (as is likely in large groups), because altruistic behaviour signals a prosocial attitude. This prosocial attitude then invites cooperation on the part of strangers, and allows reciprocal altruism to be maintained despite the absence of a repeated game type situation.<sup>493</sup> Empirical studies confirm the relevance of such signalling in human interactions.<sup>494</sup> However, as with reciprocal altruism, the theory is not very robust to more restrictive assumptions. Under limited and/or private information, defectors may free ride on the effects of reputations by falsely acquiring a prosocial reputation at lower fitness cost than the actual co-operators would pay. In such situations, simulations indicate that indirect altruists are too vulnerable to invasion by defectors and it is therefore not a sufficient explanation for observed cooperative behaviour.<sup>495</sup>

Returning then to multi-level selection, we must see how this provides the possibility for a different approach to altruism than these more ‘individualistic’ accounts. The work of Bowles and Gintis, on which I here rely, is a fruitful investigation of these possibilities. Recall that group selection can work by producing ‘positive assortment’, i.e. altruists being likely to interact with other altruists, via the effect of group competition on competition between individuals. As Bowles and Gintis explain, “members of predominantly altruistic groups have above-average fitness and thus contribute disproportionately to the next generation... The same model applies to any process of selection based on the differential replication of traits over time.”<sup>496</sup> This is necessary because, as Hamilton’s Rule points out, selection can only favour altruistic traits if its bearers are more likely to interact with other bearers of this trait than chance.<sup>497</sup> In group selection models, *as opposed to* reciprocal ‘altruism’ models, altruists benefit from interactions with other altruists because of the positive assortment of altruists into groups and non-altruists into others. Such explanatory models are called models of inclusive fitness, and Bowles and Gintis helpfully distinguish them from ‘mutualist’ models such as reciprocal and kin altruism. Such positive assortment is particularly likely in human communities because of the assortative aspects of cultural markers and of communication through language, something I will discuss further below under gene-culture coevolution.<sup>498</sup> For now it is worth noting that positive assortment can take place genetically, but also culturally, by the intentional adoption through learning of favoured behavioural traits. The latter process is of course much faster.

What Bowles and Gintis have demonstrated is that applying more realistic, but therefore also more restrictive, constraints on the standard evolutionary game theoretical arguments for

---

<sup>493</sup> Sugden 1986; Alexander 1987; Panchanathan and Boyd 2004; Milinski, Semmann and Krambeck 2002; Binmore 2005. Ideas of signaling go back to Spence 1973 and Zahavi 1975.

<sup>494</sup> Smith and Bliege Bird 2000; Bliege Bird and Smith 2005.

<sup>495</sup> Engelman and Fischbacher 2009.

<sup>496</sup> Bowles and Gintis 2011, p. 52-53.

<sup>497</sup> *Ibid.*, p 48; Hamilton 1964.

<sup>498</sup> E.g. Eshel and Cavalli-Sforza 1982.

‘mutualist’ reciprocity, they find that such models cannot account for the evolution of human cooperation. As they argue, this is because “even presupposing extraordinary cognitive capacities and levels of patience among the cooperating individuals, there is no reason to believe that a group of more than two individuals would ever discover the cooperative Nash equilibria that the models have identified, and if it were to hit on one, its members would almost certainly abandon it in short order. Except under implausible conditions, the cooperative outcomes identified by these models are neither accessible nor persistent.”<sup>499</sup> It is known that in the case of indefinitely repeated games, if information is sufficiently perfect or public, error terms are sufficiently small, and the discount factor less than one, even amoral self-regarding players can achieve cooperation for any set of payoffs above a certain minimum – possibly the ‘folk theorem’ referred to by North allegedly supporting the role of markets.<sup>500</sup>

However, as Bowles and Gintis show, the corollary of this also holds: when groups are very large, error terms large, information private or very imperfect, or no common alignment of beliefs exists (i.e. the choices of other players are not known in advance), the folk theorem fails. Indeed, this finding is simply an implication of the observation, discussed in chapter 3, that except for the simplest types of games, no Nash equilibrium is likely to be found absent the criteria of common knowledge of rationality and common alignment of beliefs.<sup>501</sup> While some solutions have been found to resolve this problem, such as a learning mechanism for games, or resolving a monotone dynamic (that is to say, an evolutionary dynamic where higher payoff strategies increase their frequency in a population relative to lower payoff ones), none of these solutions can overcome the unrealistic assumption of common alignment of beliefs. Given North’s constant emphasis on the importance of beliefs and the difference they make, it is certainly in North’s spirit to agree with Bowles and Gintis that the Nash equilibrium is exceedingly improbable under realistic conditions such as disparate beliefs and private information.<sup>502</sup>

A more relevant and generally accepted solution is the correlated equilibrium, a suitable substitute for a Nash equilibrium in evolutionary models. What this essentially means is that some external and commonly known ‘agent’ interacts with the actual players of the game such as to coordinate their actions, by sending signals in such manner that if all players accept the signal, no player can improve their payoff by switching strategies.<sup>503</sup> The interesting thing for our purposes is that for human players, the most obvious form of the correlating device in concrete terms is the social norm, i.e. an *institution*. As Bowles and Gintis explain, a “cooperative equilibrium supported by social norms is one in which not only is the equilibrium evolutionarily stable, but also the social norms are themselves an evolutionary adaptation, stable against

---

<sup>499</sup> Bowles and Gintis 2011, p. 80.

<sup>500</sup> Fudenberg and Maskin 1986; Fudenberg, Levine and Maskin 1994; see North 2005, p. 29, discussed above.

<sup>501</sup> Hargreaves-Heap and Varoufakis 1994; Aumann and Brandenburger 1995.

<sup>502</sup> Bowles and Gintis 2011, p. 88-89.

<sup>503</sup> *Ibid.*, p. 89; Gintis 2009.

invasion by competing norms”.<sup>504</sup> This is necessary because otherwise the contractarian’s infinite regress would operate again: a norm may not be evolutionarily assumed without explaining away the problem. This means that individuals in a group that shares a norm must learn what the norm is (what behaviour it demands), learning that most in the group know the norm and follow it, and learning that the norm is common knowledge among those (i.e. that they all know this about each other as well).

For Bowles and Gintis, this role of social norms must be analytically prior to any explanations in terms of self-regarding individuals finding a cooperative solution, since absent such coordinating functions *no evolutionarily accessible and stable cooperative solution can be found under realistic conditions*. As they argue, this finding explicitly refutes the contractarian and ‘invisible hand’ approaches to the folk theorem, since such approaches are not robust to realistic restrictions on common alignment of beliefs and on the likelihood of error. Again, given the importance North rightly places on the role of varying beliefs and their independent explanatory role in behaviour, it is justified to conclude that to pursue a realistic evolutionary account of the emergence of institutions, North cannot resort to any ‘invisible hand’ or contractarian type explanations. Nor does the recognition of cognitive limits and bounded rationality suffice to compensate for that: the explanation must begin with seeing human individuals as *already prosocial*.<sup>505</sup>

As they argue: “The failure of the models underlying both the folk theorem and the fundamental theorem [i.e. welfare economics, MK] is hardly surprising, for they sought to explain cooperation among large numbers of self-regarding strangers without recourse to pre-existing norms and cooperative institutions, something that most likely never occurred in the history or prehistory of our species... the economic models of cooperation that assume pre-existing solutions to these problems thus do not accomplish their goal, namely explaining cooperation among amoral and self-regarding individuals... Economic theory, favoring parsimony over realism, has sought to explain cooperation without reference to social preferences, and with a minimalist or fictive description of social institutions. This research trajectory... may have run its course.”<sup>506</sup> This, then, requires a new microfoundation. But what kind of prosociality? And how is this to be integrated into a cultural evolutionary account such as North’s?

We know that evidence of prehistoric societies, such as can be deduced, is unlikely to be favourable to reciprocal or kin altruism as explanation of the emergence of prosocial behaviour and institutions.<sup>507</sup> In any case the paleontological evidence is controversial, and extrapolation from present-day hunter-gatherer groups is no less so. Bowles and Gintis, building on Boyd and

---

<sup>504</sup> *Ibid.*, p. 90.

<sup>505</sup> *Ibid.*, p. 91.

<sup>506</sup> *Ibid.*, p. 91-92.

<sup>507</sup> *Ibid.*, ch. 6; Boyd and Richerson 2005.

Richerson's theory of gene-culture coevolution (which I will turn to shortly), therefore applied simulations based on parameters deduced from both sources of evidence and varied these within that range, so as to obtain realistic and robust, if not guaranteed, results. On this basis, they have proposed a theory for the origins of what they call 'strong reciprocity', which consists of two components: a predisposition to cooperate, i.e. a form of innate altruism, *and* a willingness to punish noncooperators.

Where North sees the problem of violence as central and its reduction as the natural aim, the 'constructive' importance of violence in the form of punishment – physical, social, or emotional in nature – to the evolution of cooperation and institutions is equally paramount. Punishment is the key means by which the costs to free riders can be reduced, as North himself observes in the context of the problem of enforcement. Even wholly self-interested individuals may cooperate when punishment for not doing so is sufficiently likely and severe. But enforcement is not without costs, as North also observes; and this applies to evolutionary contexts also. In simple evolutionary models, the fitness costs of punishing to the punishers and the possibility of error selects against the likelihood of its evolution, just as with altruistic cooperation. But there is one core difference: the cost of punishment declines as more individuals engage in it, because the frequency of punishment will reduce cases of defection and thereby reduce the frequency with which the costs of punishment actually have to be incurred.<sup>508</sup> Because of group selection, more cooperative groups are likely to outcompete others; if punishment is therefore evolutionarily stable and 'accessible' over a wide range of parameter values, sustained by group selection, as Boyd et al. argue, this is one foundation stone for the evolution of institutions. Moreover, empirical evidence supports the observation that punishment can solve the free rider problem also due to its 'quorum' quality: individuals often punish as a group, only undertaking it if sufficient are disposed to punishment, and so solve the internal free rider problem among the punishers.<sup>509</sup> Finally, a major role is played by the punishment of nonpunishment. Once punishment can become stable and common, second-order punishment also becomes possible. As we shall see in the discussion of gene-culture coevolution, this is important, because second-order punishment, once common, can stabilize *any* norm.<sup>510</sup>

But one problem remains. For the quorum model of punishment to succeed, some degree of public information is required, otherwise coordination becomes impossible. Bowles and Gintis argue therefore that such mechanisms must be complemented evolutionarily by another: the development of the social emotions. The internalization of social norms into preferences, oughts into wants, takes place via the social emotions, particularly shame. For this to be a realistic evolutionary process, the process of socialization, i.e. the acquisition of norms, must be a stronger factor than the within-group selection against the fitness-reducing prosocial norm. This

---

<sup>508</sup> Boyd et al. 2003.

<sup>509</sup> Bowles and Gintis 2011, p. 149; cf. Boehm 2012; see also Tuck 2008 for an extensive analytical discussion of this point.

<sup>510</sup> Boyd and Richerson 2005, p. 91.

process is aided by the evolution of social conformism among humans, which I will discuss shortly: those very same cognitive properties North identifies as essential for understanding bounded rationality, namely those that serve as heuristics under conditions of uncertainty, also favour conformism with regard to the behaviour of others, by reducing the likelihood of making costly errors. The social learning processes of individuals can allow the internalization of norms, Bowles and Gintis show, under a wide range of realistic parameters for these reasons, since adopting the shared behaviour and beliefs of the group is more likely than not to be fitness enhancing. Given this fact, as long as the norm is not excessively fitness reducing, a fitness reducing norm may be adopted and internalized through the same processes that are used to acquire beliefs and to become ‘socially adjusted’ in the first place. The social emotions may have evolved to aid in this process, by overriding self-regarding intentionality in favour of anger, shame, guilt, and the like, which steer behaviour towards the individually fitness reducing but group fitness enhancing norms in question. Since the inducement of such emotions in defectors is likely to be cheaper in fitness terms than the use of physical punishment in each case, and since the internalization of such emotions saves on enforcement costs, groups in which shame and guilt evolve are likely to outcompete those in which this does not happen.<sup>511</sup>

The upshot of all these arguments is to demonstrate that the most likely evolutionary explanation for the evolution of institutions and the emergence and persistence of cooperation is, to a considerable degree, strong reciprocity or “genuine altruism: a willingness to sacrifice one’s own interest to help others, including those who are not family members, and not simply in return for anticipated reciprocation in the future”.<sup>512</sup> The experimental evidence, so quickly dismissed by North, gives substantial support to this view. Such experiments across a wide range of societies and cultures show that responses to various standard games, such as the ultimatum game and the dictator game, fluctuate wildly and (most importantly) in strong correlation with the institutional and political economic structure of the societies in question. In no case whatsoever was the postulate of maximization of expected returns actually observed.<sup>513</sup> There is considerable evidence to suggest that humans are strongly inclined to participate in collective projects and to punish free riders even at their own cost. Collective action problems are solvable (and solved) by the operation of group enforcement mechanisms that are not in the individual interest of any members but that depend on strong reciprocity and prosocial forms of rational action, i.e. forms of group rationality and identification that are stronger than individual utility maximization (at least of the conventional kind); and such mechanisms go beyond the markets or hierarchies dichotomy of the NIE tradition.<sup>514</sup>

These findings must be incorporated into any cultural evolutionary theory of economic history. More importantly at the methodological level, the salience of group level selection and

---

<sup>511</sup> *Ibid.*, ch. 10; p. 193.

<sup>512</sup> *Ibid.*, p. 199.

<sup>513</sup> Henrich et al. (eds.) 2004.

<sup>514</sup> Gintis et al. (eds.) 2005, ch. 13; Ostrom 1990.

of the interaction of norms, strong reciprocity, and punishment mechanisms as evolutionary stabilizers must be part of any explanation of how culture fits into models of rational action in history-writing. Finally, we must take the moral content of norms and emotions seriously, and this, too, suggests going beyond utility maximizing individuals. I will now expand briefly on these observations.

### **Gene-culture coevolution and the role of institutions**

Arguably there is a natural affinity between multi-level selection theory and gene culture coevolution, in that the latter builds on the methodological developments of the former and in turn has aided the evolutionary game theoretical advances represented by the work of Bowles and Gintis. Although the idea that selection pressures at phenotypical level and cultural dynamics interact via the mechanism of group selection goes back to Charles Darwin himself, contemporary gene-culture coevolution theory originates in the 1970s-1980s with the revolution in sociobiology. This revolution revived the evolutionary study of social life and culture and placed it on a more scientifically secure footing than the speculative or metaphorical evolutionary approaches of the Victorian age.<sup>515</sup> Here, I will draw on the works of Robert Boyd and Peter Richerson, who have done more than perhaps any other researchers to develop gene-culture coevolution into a potentially powerful explanatory framework to rival memetics, evolutionary psychology, and other sociobiological approaches – not least in emphasizing its flexible applicability to historical study.<sup>516</sup> This applicability, combined with the lack of theoretical development of memetics and the empirical problems with the modular view suggested by evolutionary psychology, cause me to see gene-culture coevolution as the most promising school of sociobiology today for the purposes of historical explanation.<sup>517</sup>

The core insight of gene-culture coevolution is that, as Boyd and Richerson put it, “cultural evolution is a population phenomenon”.<sup>518</sup> Relying on multilevel selection theory, gene-culture coevolution allows the examination of the interaction between population-level phenomena and individual-level phenomena, just as we have seen in Bowles and Gintis’ arguments regarding the origins of human cooperation (which are consciously compatible accounts). The difference is that Boyd and Richerson concentrate on the effects of one particular evolved tool particular to humans, the ability to intentionally develop culture and to acquire existing cultural traits through social learning. The effects of such traits on the formation of groups and thereby on group selection, as well as on the fitness of individuals acquiring (or rejecting) a particular cultural trait, and the modelling of the evolutionary interaction of these phenomena at genetic and

---

<sup>515</sup> An early forerunner is Campbell 1965. See also: Lumsden and Wilson 1981; Cavalli-Sforza and Feldman 1981; Durham 1991; Henrich 2004.

<sup>516</sup> Richerson and Boyd 2005; Boyd and Richerson 2005; McElreath, Boyd and Richerson 2003; Boyd et al. 2003.

<sup>517</sup> For a criticism of the school of ‘evo psych’, see Buller 2005. For a general critique of popular sociobiological models, see Lewontin and Levins 2007, ch. 26.

<sup>518</sup> Richerson and Boyd 2005, p. 241.

cultural level is the domain of gene-culture coevolution. This, too, is an extension of generalized Darwinian thinking.<sup>519</sup>

From the viewpoint of gene-culture coevolution, North is decidedly right to concentrate on the role of learning in accounting for the role of institutions. However, North's discussion of learning processes emphasizes the cognitive limitations of rationality in a rational choice context, and the importance of uncertainty for the acquisition of (potentially 'irrational' or wrong) beliefs. Gene-culture coevolution on the other hand makes learning a central bridge between the genetic explanation for the origins of human sociality on the one hand and the function of institutions and cultural markers, once they exist, on the other hand. This requires a more sophisticated understanding of what learning functionally entails. Moreover, it requires a reconsideration of the function of culture within a Darwinian perspective – one that can address North's objection to sociobiology that culture, being intentional in nature and highly variable, cannot be explained by generalized Darwinism.

First we must define culture. For Boyd and Richerson, culture (at least as far as this theory is concerned) is "information capable of affecting individuals' behaviour that they acquire from other members of their species through teaching, imitation, and other forms of social transmission". We are speaking here, then, of behaviourally visible traits that are acquired through learning processes, although the 'information' need not be (fully) conscious.<sup>520</sup> Culture is an evolved ability of humans as a species, and it performs powerful functions in this way indirectly shaping our evolved behaviour – functions we require social cooperative skills to perform, such as we have seen we have acquired through the processes described by Bowles and Gintis. Culture is adaptive because the behaviour of other humans can be acquired by learning, and it provides valuable information about the environment that need then not be independently rediscovered, over and over again, by individual reason. This is the major advantage culture (in particular via language) provides humans over every other species.<sup>521</sup> Culture is therefore, as North rightly emphasizes, an important proximate explanation of human behaviour, but a thoroughgoing Darwinism requires that its ultimate explanation always remain evolutionary in nature – and so does a consistent theory of cultural evolution. Even so, once culture is evolved in humans (as happened early on in our behavioural modernity) and the forces operating on its population statistical properties are understood, as I will summarize shortly, it plays an independent role in interaction with the effects of selection at the genetic level. Culture is not reducible to genes. Nor is culture simply part of 'the environment' with which our genes interact via our 'extended phenotype'. Knowledge, traits and beliefs are not part of the environment, and

---

<sup>519</sup> *Ibid.*, p. 247.

<sup>520</sup> *Ibid.*, p. 5.

<sup>521</sup> *Ibid.*, p. 12.



are not acquired from it in one go; moreover, in very similar environments there can be vast differences in cultural traits, beliefs, and technologies.<sup>522</sup>

The next step is to realise that culture has the required attributes to be subject to a process of Darwinian natural selection: it is heritable with sufficient stability, there is variety, and there is differential survival from one generation to the next. However, there is a major difference with genetic evolution, as North rightly intuited (with most social scientists, probably): intentionality plays a key role in cultural evolution. Variation in culture is not random, nor is retention. However, although the randomness of mutations plays a major role in evolutionary biology, it is not a requirement of Darwinian natural selection that the sources of variation be random. Nor is it inherently a problem that, *pace* Dawkins and Dennett, culture is probably not discrete nor faithfully transmitted. A suitably transformed theory of natural selection can cope with cultural descent with modification even if the modification is intentional, at least in general terms. In this sense it is somewhat similar to artificial selection, which is simply a special case of Darwinian theory. (Like all applications of Darwinian theory, gene-culture coevolution theory is general – auxiliary theories and data are needed for any meaningful results.) To the extent, moreover, that we accept North's argument that institutional change is mostly incremental, the applicability of natural selection is strengthened.<sup>523</sup>

What gene-culture coevolution can do for histories of the evolution of culture/institutions (economic in focus or otherwise) can be clarified by looking at the various forces that operate on its transmission. Drawing on Richerson and Boyd's analysis, we can usefully divide these into two types: forces of random variation and forces of non-random (guided) variation and inheritance. The former consist of cultural mutation and cultural drift, which respectively refer to any unintentional changes to cultural traits (they use the example of misremembering) and effects caused by the emergence or dying out of rare (new) traits in a small population. More important are the forces of guided variation: changes made in the process of learning and innovation. Learning, in turn, can be subdivided into different functional types of 'biased transmission', concentrating on by what mechanism the trait is acquired by an individual – intentionally or unconsciously, as in the case of evolved learning processes in small children.<sup>524</sup> These are threefold. The first is content-based bias, which depends as the name suggests on the content of the trait, and is acquired either through cost-benefit analysis of a boundedly rational type (this one best corresponds to North's model), or because of the attributes of the trait that

---

<sup>522</sup> *Ibid.*, p. 26-29.

<sup>523</sup> *Ibid.*, p. 50-54, 60.

<sup>524</sup> Note that 'trait' here has in this context the meaning of 'cultural variant', and therefore is subject to competition in the form of cognitive expression in specific human individuals, as Richerson and Boyd argue. The model therefore concerns the differential survival of such traits, which are therefore the equivalent of 'organisms' in the standard Darwinian model (the 'traits' of the traits are then those aspects that make them more or less likely to be culturally or cognitively expressed by humans). This can give rise to confusion: it is to be distinguished from most standard Darwinian models in which traits are generic attributes of organisms and affect the differential survival of those organisms through competition between *organisms*, not traits.

allow it to be better remembered or learned. The second is frequency-based bias, and refers to learning based on how common the trait is; this is also called conformist learning. Finally, there is model-based bias, in which a particular figure is treated as a ‘model’ for learning based on attributes that make them likely to be a good source of acquired information. A common form of such biased transmission is doing whatever the most successful or highest prestige figures do, or imitating those with traits marking them as similar to oneself.<sup>525</sup>

So much for inheritance. But if such traits are to be subject to a process of natural selection, they must also have differential survival. This means the traits must, in a sense, be in competition with one another. Boyd and Richerson suggest this happens in two ways: through the opportunity costs of time spent learning (and the limits on what one can learn), and through control of behaviour. If one must act, one act or another will be chosen, and one can only obey the imperative of one cultural trait at a time: one can know how to speak multiple languages, but one speaks only one on a given occasion, even if one can switch in mid-sentence. Given this, how does natural selection operate on culture? The usual requirements can now be straightforwardly applied. People vary because they have acquired different beliefs, values, or habits through social learning. Their trait variation affects the probability that others will, through social learning, adopt their trait. And finally, many cultural variants compete with each other to some sufficient extent. In such cases, natural selection will operate at the level of these traits, even if it is difficult to identify precisely what they are in discrete terms.<sup>526</sup>

The important implication of these forms of biased transmission and random change is that culture need not be adaptive in fitness terms, even though it is subject – as humans always are – to the processes of natural selection at genetic level. Unlike in the genetic case, but much like the case of rational choice, the process of descent with modification of cultural traits is therefore not inherently fitness maximizing.<sup>527</sup> While the combination of guided variation (decision-making) and biased transmission (learning) must be evolutionarily stable and able to outcompete alternatives, no individual cultural trait need be fitness maximizing, and indeed many are from a strict biological fitness viewpoint ‘maladaptive’.<sup>528</sup> (Boyd and Richerson argue that human socialization has exactly that overall evolutionary stability because the environment humans live in changes with just the right degree of slowness, but that argument need not concern us here.<sup>529</sup>)

The final dynamic to understand is the effect of the interaction of these forces with group formation. Groups form among humans by cultural markers (which need not be behavioural in nature) and particular traits (which are). As long as guided variation is error-prone and biased transmission plays a considerable role, and in particular as long as there is sufficient conformist

---

<sup>525</sup> Richerson and Boyd 2005, p. 69.

<sup>526</sup> *Ibid.*, p. 76.

<sup>527</sup> Boyd and Richerson argue for guided variation as utility maximizing, but this is not strictly necessary. See Boyd and Richerson 1985, ch. 5.

<sup>528</sup> Richerson and Boyd 2005, p. 150-151.

<sup>529</sup> *Ibid.*, p. 118.

bias and enforcement of norms through punishment, the amount of variation within groups will decline over time and the amount of variation between groups will increase. This, in turn, will strengthen the effect of group selection insofar as human groups compete with each other.<sup>530</sup> While the degree and dimensions of competition between groups is controversial, there is little reason to doubt that human groups historically often have competed, in particular if one defines groups as designated by the existing variants of the relevant cultural traits (language X vs Y, or Protestant vs Catholic, or whatever). Where this is relevant, the competition between groups will be competition between sets of cultural traits and markers, and the extinction of particular marker groups (physically, or by absorption into the winning group) means an increase in the frequency of the traits and markers of the more adaptive group. In this sense, there is a certain adaptive efficiency of culture, to put it in North's terms – a classic example being the Nuer organizational structure favouring its competitiveness vis-à-vis the Dinka.<sup>531</sup> There is some evidence that competition between groups, leading to 'cultural extinction', is common among small scale societies at least, although it is not conclusive.<sup>532</sup> Even so, we must be careful – North never specifies what competitive process allows for differential variation in 'efficiency' or 'performance' to make one society more adaptive than another, and this greatly weakens the claim to an actual selection process in his account. More than competition, diffusion from one group to another of group functional traits is a fast and plausible process through which differential variation of traits can operate on global populations (that is to say, supra-societal ones, not necessarily literally across the whole world). A salient example is the success of evangelizing religions in competition with non-evangelizing ones.<sup>533</sup>

The interaction between genes and culture on this perspective is useful for historians to integrate into their accounts. A few examples may serve to clarify what use this viewpoint has for constructing a narrative of cultural evolution in practical terms. For example, cultural markers interact with genetic evolution (which is fitness maximizing) to cause groups that differ by markers to increasingly fulfil different evolutionary niches so as to make their markers determinate for different economic strategies, habitats, etc., a process of immediate interest to economic historians.<sup>534</sup> The interaction between cultural traits and natural selection resulting from disease, or from famine, or other such forces, can play a potent role in the history of institutional survival and diffusion, although we must not simply assume that all traits are for that reason biologically adaptive. It has been argued this played a major role in the victory of Christianity over paganism in ancient Rome, insofar as Christians' mutual aid during times of plague increased their survival rate notably over that of pagans.<sup>535</sup>

---

<sup>530</sup> *Ibid.*, p. 162-163.

<sup>531</sup> Kelly 1985, p. 65.

<sup>532</sup> Soltis, Boyd and Richerson 1995.

<sup>533</sup> Richerson and Boyd 2005, p. 210-211.

<sup>534</sup> Boyd and Richerson 1987.

<sup>535</sup> Stark 1996. See however Bryant 1997 for a critique of this claim.

Of course, with such explanations the risk of multiplying speculative just-so stories is considerable. However this may be, what is more useful is to consider the tools that have been offered for cultural evolutionary history to succeed. Integrating strong reciprocity and its role in the origin of institutions with the persistence, change, and inheritance of institutions via gene-culture coevolution gives us a powerful conceptual apparatus that is both flexible and methodologically well-founded in the most robust functionalist theory that the social sciences have available, Darwinian natural selection. The evolution of human complex societies and the ongoing significance of traits and beliefs need then not be explained (solely) in terms of social contracts, boundedly rational choice, or market(-like) competition. Culture – institutions, beliefs, and markers – and the evolution of strong reciprocity and prosocial instincts like first and second order punishment go hand in hand as tools to explain both origin and change of human societies. As Boyd and Richerson argue, “in the short run, cultural evolution, partly driven by ancient and tribal social instincts and partly by selection among culturally variable groups, gave rise to the institutions we observe. In the longer run, cultural evolutionary processes created an environment that led to the evolution of uniquely human social instincts.”<sup>536</sup>

### **A word on morality and emotions**

Having discussed the evolution of strong reciprocity, the theory of gene-culture coevolution, and their significance for understanding the emergence and persistence of institutions, it is necessary to return for a moment to the problem of emotions. We have seen that Bowles and Gintis have offered an account of the evolution of the social emotions that saw them as evolutionarily advantageous in overriding self-regarding, but group fitness reducing behaviour, and as a time discounting heuristic. But what role should the emotions play in the model of rationality that emerges if we accept their account of strong reciprocity? To account for their evolutionary emergence is one thing, to explain their behavioural significance in economic explanation is another. Geoffrey Hodgson has argued that the emotions play an important role in morality, as opposed to altruistic behaviour. He distinguishes these along much the same lines as deontological and consequentialist ethics are distinguished in ethical philosophy: for him morality is about the rightness of the thing done, whereas other-regard is still utility maximizing insofar as it is about maximizing the effect of the thing done.<sup>537</sup> For him, moral values, with their inherent universalizing and other-regarding (or perhaps one should say other-incorporating) claims, are an important part of human behaviour but not reducible to preferences or beliefs in the conventional sense.<sup>538</sup>

Moral rules are therefore peculiar kinds of institutions: as Hodgson argues, moral rules are not “reducible to conventions. They become moral rules because many people believe in them as

---

<sup>536</sup> Richerson and Boyd 2005, p. 235.

<sup>537</sup> Hodgson 2013, p. 76.

<sup>538</sup> *Ibid.*, p. 79.

such and uphold them as more than matters of convenience, self-interest, or convention.”<sup>539</sup> In this sense they have something in common with the arguments for inclusion of commitments in rationality discussed in chapter 3. Hodgson’s main argument is that the point of morality, so to speak, is to act as a cultural trait akin to emotions, in that its purpose is to be *different* from any strategic or instrumental appeal to consequences, self-interest, or prudence, and to overcome the ability of humans as reasoning beings to override our learned behaviour when it is in our self-interest to do so. The social emotions play a major role in this account because strong feelings, such as would override any strategic decision-making processes, are the best supports for the persistence of such moral systems and values.<sup>540</sup> While I differ with Hodgson on the plausibility of founding this account on the strong modularity theory of the brain, as proposed by (some) evolutionary psychologists, the overall point is well-taken. Insofar Bowles and Gintis’ account of strong reciprocity is still consequentialist in nature, to some extent a consequence of the functionalist impetus of evolutionary accounts, it misses an important deontological dimension to human life that would overlook the powerful emotional drives and commitments involved in religion, patriarchy, or even notions of *ethnê*. These, too, are cultural traits that require incorporation into an account of institutional change, and these, too, are modifications to Northian transaction cost-reducing bounded rationality as much as to the neoclassical model.

## Conclusion

After this extended discussion, let me recapitulate the general argument of this chapter. Having given my reading of North’s theory of cultural evolution, I have wondered what function it serves in his overall argument. While it appears mainly to serve as a kind of backup argument to justify the general claims for the virtues of market competition for economic performance and for achieving cooperation, at the same time the turn towards evolutionary foundations is a very important theoretical step in its own right. In chapter 2, I contested North’s account of markets and market competition and whether they really could carry the analytical burden he placed on them. In chapter 3, I discussed North’s ambiguities and inconsistencies with regard to models of rationality and his difficulties with incorporating the growing role of *sui generis* beliefs – including their other-regarding component – into his model of agency. The appeal to evolutionary microfoundations has the potential, at least to some degree, to resolve both, and I suspect this is its main appeal for North.

An evolutionary account is both helpful and inevitable to overcome the infinite regress problem in joining instrumental reason and institutional structure in explaining social phenomena. Equally, an evolutionary account at the macro-level could provide North’s historical narrative with a powerful functionalist type of explanation, buttressed by the well-understood ‘trinity’ of Darwinian natural selection. North’s theory of cultural evolution can be read as attempting to take both these steps. Yet here, too, North appears conflicted: he does not want to

---

<sup>539</sup> *Ibid.*, p. 81.

<sup>540</sup> *Ibid.*, p. 120.

give up the central role of intentionality and culture, which he fears sociobiological approaches such as the Tooby-Cosmides school of evolutionary psychology would entail. The evolutionary and neurological descriptions of learning and cognition processes, the basis for the formation of beliefs, on my reading loom large in *UPEC* and elsewhere precisely because they allow North to concentrate on intentionality under uncertainty. In this way, he can place bounded rationality in its more conventional forms at the heart of his account of institutions, while seemingly avoiding the infinite regress as well as the pitfalls of evolutionary reductionism. Moreover, he wants to preserve the centrality to his account of the dual mechanism of markets and institutions to overcome, respectively, inefficiency/inflexibility and uncertainty-cum-transaction costs. For this reason, it comes naturally to formulate his evolutionary account in Hayekian terms, where the inheritance of tradition fills the gaps, as it were, that the free market cannot provide for on its own.

I applaud North's turn towards evolutionary thinking beyond the evolution of organisation and the firm: the best aspects of the Hayekian tradition is precisely this willingness to integrate micro- and macro-level evolutionary perspectives with the historical and theoretical dimensions of economics. However, there are serious problems with this tradition and with North's account as well, and these I have discussed at length in this chapter. I argue North cannot avoid the infinite regress problem merely by appeal to learning and intentionality, for he misunderstands the role of intentionality in a Darwinian account. He mistakenly believes that the importance of cultural variation and intentional action makes sociobiological theory less applicable. But this is not so, certainly not when other sociobiological approaches than the Tooby-Cosmides school are taken into account. Secondly, North's account of natural selection operating on societies via the adaptive properties of market competition is both implausible and incomplete, failing to rigorously describe the necessary prerequisites of a Darwinian process. Thirdly, North still fails to account in a systematic evolutionary manner how the human sociality that gives rise to our particular institutionalised and complex forms of socio-ethical reason comes about. As Boyd and Richerson rightly emphasize, decision-making forces (as they call them) are *derived* mechanisms: they must in evolutionary terms have their origin in other forces of selection. Which account of such origins makes sense therefore matters a great deal for what account of rationality is most plausible.<sup>541</sup>

Because he simply assumes away this central question in sociobiology, he also does not realize the significance the experimental economic evidence testing the relation between culture and utility maximisation has for his work, nor the full import of models of the evolution of his 'foraging orders', which he leaves to the controversial work of Johnson and Earle. Generally, North misses the tremendous opportunity evolutionary thinking offers to place his account of institutions in economic history on more firm foundations. In all evolutionary writing on the origins and function of certain sociocultural phenomena, it is a question of competing plausibility

---

<sup>541</sup> Boyd and Richerson 2005, p. 290.

arguments: no account can be proven with any degree of certainty, the purpose of sociobiological modelling is to establish which accounts are the least implausible. In such arguments, it does not suffice to simply argue that a particular account is incomplete or overly simple, since one is always abstracting from certain things and much of the work is still in the stage of establishing basic parameters. Rather, it is necessary to point to specific mechanisms identified in the model as inadequate, and/or to point to missing dynamics that would matter for the account.<sup>542</sup> I have therefore offered to substitute for North's incomplete and often implausible account an indication of where one could find the theoretical resources for a less implausible and more complete one, both at micro- and at macro-level.

To do this, I have sketched the integration of what I see as the two indispensable supports for such an approach, namely gene-culture coevolution and experimental/behavioural economics. The former allows for the integration of culture and intentionality with a Darwinian adaptationist account, so that the origin problem disappears and with it the threat of infinite regress. It also offers a way of viewing the relationship between historicity and functionalism in a more clear light. The latter, in turn, should be explored for its implications for sociality and rationality, i.e. the very Northian question of what institutions and choice behaviour have to do with one another. Since the predictions of utility maximisation and even minimum consistency of rationality are routinely violated in such experiments, and in such a way as to suggest that these violations are themselves derivative of sociocultural orders and the political economies over longer historical time of the various peoples studied, this evidence seems to me to point the way to how this Northian question might be answered.

North himself could not find an internally consistent answer, because of his unwillingness to abandon classical game theory and Simonian bounded rationality with their reliance on conventional forms of utility maximisation. But the gene-culture coevolution approach coheres very well with taking seriously other-regarding and group rationalities as well as the role of the emotions. As evolved behaviours, these both give rise to and interact with the necessary *forms* of social life, which are historically path dependent, arbitrary, and also dependent on the *content* of the institutions and beliefs in question. Here, functionalism and contingency can quite comfortably fit together in a historical and evolutionary account. But there is a price to pay: major Northian themes, such as the adaptive 'efficiency' of market competition, the use of indeterminate forms of bounded rationality in games as a contractarian explanation of the origins of institutions, and the neglect of sociobiology and anthropology must be given up. The result, however, would be a less implausible and more consistently evolutionary economic history, and without sacrificing the centrality of institutions and beliefs.

All this is of course not sufficient on its own, just as appeals to contracts or transaction costs are not sufficient on their own: specification by means of individual case studies is always

---

<sup>542</sup> *Ibid.*, p. 410-414.

needed. Moreover, the evolutionary perspective does not avoid the problem of non-ergodicity North rightly draws our attention to: the role of norms, arbitrary conventions, and group markers in cultural evolution and the considerable room in institutional ‘design space’, to put it in Daniel Dennett’s terms, give rise to historical path dependencies and ‘runaway dynamics’ that cannot be predicted in advance. Nor can an explanation of such historical dynamics be reduced to the maximisation of either utility or fitness.<sup>543</sup> But the larger scale, long-term view is aided by this perspective, one that is to my mind better grounded and more empirically plausible than the medley of social contract theory and evolutionary psychology offered by North. Finally, the sociobiologically integrated perspective has one additional interesting property: in emphasizing how second order punishment can evolve and be sustained, and how it can stabilize *any* norm, and in observing the importance of norms for the cultural evolution of societies, the gene-culture coevolution perspective paves the way for a more explicitly anthropological treatment of institutions in economic history as well as other historical accounts. Reconciling anthropology and economics also brings us back to the contribution of Polanyi. These themes I will take up in the next chapter.

---

<sup>543</sup> *Ibid.*, p. 303-306.



## Chapter 5

### Polanyi's challenge revisited: historiographical and anthropological reflections on Douglass North's NIEH

#### Introduction: revisiting Karl Polanyi

In the course of this work, I have sought to examine the origins of Douglass North's NIEH and its development over time, concentrating on its historiographical and economic theoretical methods and their justifications. A guiding thread throughout my study of the various themes in North's conceptual apparatus – e.g. markets, game theory, evolution, institutions, states and social contracts – has been North's complex and often contradictory relationship to rival approaches. If my interpretation is plausible, North's NIEH has a peculiar character. On the one hand, in its increasing emphasis on the institutional, cultural, and even cognitive context of economic rationality and choice behaviour within and without well-defined markets, it seeks to move the mainstream of economic theorizing about historical change into a direction much closer to that of the other social sciences. His extensive critiques of neoclassical economics and its straightforward rational choice analysis, his increasing emphasis on the importance of beliefs, and his rejection of efficiency interpretations of institutions make this clear.

On the other hand, North is simultaneously unwilling to give up all of the analytical framework of the economic tradition from which he emerged. In particular, North's reintroduction of neoclassical thinking about the function and effect of market competition acts as a force pulling him back to the other direction. He is keen to insist on the classic welfare economic implications of well-ordered markets, sees such markets as necessary (if not sufficient) prerequisites for economic growth, and therefore takes a good deal of neoclassical consumer and growth theory and its larger micro- and macroeconomic implications essentially for granted. In his framework, 'the market' can only function properly if complemented by transaction cost reducing institutions and by sets of beliefs that make such institutions possible in turn. But that the full unfolding of 'the market' in its competitive structure with low barriers to entry is ultimately the *desideratum* of economic policy is not in dispute for North – to the extent that he ascribes to it not just growth-enhancing, but even evolutionary adaptive properties.

But beyond this, often North's way of presenting the behaviour of individuals in their interaction with their 'artifactual structures', whose importance the NIEH has done so much to stress, shows that North's thinking has not always gone much beyond the conventions of neoclassical theory. However much North outwardly rejects rational choice theory and insists on a bounded and constrained form of choice, even at times allowing preferences and beliefs to be intelligible and derivable only from a given social matrix, he equally when it suits him presents it as a matter of the 'players' of a 'game'. Similarly, as I argued in chapter 3, North cannot decide when self-interest is a valid assumption for the understanding of economic behaviour and when it is not. As keenly aware as he is of the limitations of such an approach when it comes to

understanding the institutional layer ‘underneath’ a market (or non-market) setting, he nonetheless resorts to such behavioural assumptions when to do otherwise would limit the relevance of ‘the economics approach’. Indeed, this if nothing else remains of the ‘economics approach’ that North wished to defend against the Marxist and Polanyian rivals, as he outlined in the 1977 paper I took as starting point of my investigation. As Geoffrey Hodgson has argued, "the case for the conquest of other social sciences and biology by neoclassical economists rests on the presumed universality of such ideas as scarcity, competition, and rational self-interest".<sup>544</sup> North is clear about his rejection of neoclassical economics; it is much less clear whether he can do without such concepts. This makes it dubious whether he can therefore avoid being pulled back into the same ‘economistic’ reasoning that Polanyi decried, and which has made previous economists’ forays into the broader social-historical sciences so difficult for their colleagues to accept.

In this concluding chapter, I aim to place my close reading of North’s NIEH literature in a wider context, in particular to highlight the anthropological and philosophical dimensions of his work and their implications. To this end, I will first clarify what I mean by the ‘pessimistic anthropology’ I have identified as underlying North’s NIEH project, and which I have alluded to in previous chapters. Secondly, I will discuss North’s confrontation with an anthropologically based approach – as opposed to an economic theory based approach – in the form of Karl Polanyi, his original foil. My critical observations about North’s incomplete turn away from what Polanyi would call the ‘economistic fallacy’ might make one think I am arguing for a return to a Polanyian perspective. But I will argue here why to my mind such an approach, while also providing valuable conceptual tools, is equally not sufficient for the purpose. To put it playfully, I am not convinced Polanyi himself fully answered Polanyi’s challenge either. In fact, at least on some interpretations of Polanyi’s work one may identify analogous problems to those of North, albeit from a reverse perspective: from a different point of origin Polanyi too takes conventional choice theory as foundational for understanding the ‘logic’ of markets, but limits the full operation of this logic in economic history to a particular institutional order. This ‘mirror image’, so to speak, can best be understood by placing the North-Polanyi encounter within the larger framework with which I began this study: the primitivism-modernism debate in economic history. A few words on this context are therefore in order before turning to Polanyi himself. Finally, I discuss some alternative approaches I believe to be in need of integration, or potentially helpful, for achieving the aims of the Northian research programme. With this I conclude the present study.

### **North’s pessimistic anthropology**

In my reading of North’s NIEH works, I have tried to show both the strengths and weaknesses of this approach and in particular of its conceptual apparatus, of the tools his NIEH provides for economic thought (especially, of course, economic history). In order to tease out these apparent

---

<sup>544</sup> Hodgson 2001, p. 235.

contradictions, I have sought to go beyond North's own programmatic statements alone, and to investigate what in practice becomes of his fulminations against neoclassical economics or against simplistic *laissez-faire* development economics when put to the test of his own theoretical development. The difficulty of North's project is, as I have argued, its indeterminacy owing to its ambiguous position between (1) a neoinstitutionalism that remains closely within the orbit of mainstream economics and (2) a quest for 'successive endogenisation', to borrow a term from Malcolm Rutherford, that converges with a Polanyian or other 'substantivist' approach.<sup>545</sup>

North, as I read him, wants to avoid both Scylla and Charybdis, but as a result threatens to sail nowhere at all, remaining adrift in indeterminacy and internal inconsistency. This is not because of any demerits of his 'endogenising' ambition in institutional economics, but rather because of his inability or lack of will to carry through the task of severing his ties with conventional economic thinking about economic institutions, including those that constitute and give rise to markets. Partially this is due to what I view as ideological constraints on his thought: in particular his political commitment to a Hayekian, classical liberal perspective on the virtues of market competition, strong individual property rights and small governments, and a well-developed civil society. To this I would add his commitment to the indispensability of taking the postwar framework of economic theory, with its well-developed microfoundations in choice and consumer theory and its extensive literature on the welfare implications of consumer choice in markets.<sup>546</sup>

On the other hand, in the course of my close reading I have also tried to find different sources of this tension, in what I have called the 'hidden anthropology' of North's approach. Here, I have pointed to the persistence throughout North's work, albeit in different forms over time, of his 'pessimistic anthropology' of human cooperation. From the beginning North's work has shown a decidedly contractarian view of the emergence of institutions. Initially explicitly so in his early work with Robert Paul Thomas, as in their discussion of the rules of the medieval manor<sup>547</sup>; later less explicitly so, but still observable in his discussions of the development of institutions and state (or elite) power as a response to pervasive violence and uncertainty.<sup>548</sup>

In this view, the rule of rent-seeking but violence-reducing elites is a second best option, with the achievement of cooperation through individuals in well-ordered markets a decidedly superior approach – even if both require an appropriate institutional order to function properly. North, Wallis and Weingast summarize this contractarian interpretation of property, state, and institutions clearly: "Although they are less robust to shocks than open access orders, they generate internal forces that provide for two of the basic tasks of all societies: stability and order. Natural states may appear to be corrupt according to the norms and values of open access orders, but that corruption is an inherent part of the operation of the social order... The key feature of

---

<sup>545</sup> Rutherford 1994, p. 44.

<sup>546</sup> For a critical discussion of this body of theory, see e.g. Mirowski and Hands (eds.) 2006.

<sup>547</sup> North and Thomas 1971.

<sup>548</sup> Of course especially so in North, Wallis, and Weingast 2009.

development within the natural states is the coevolution of institutional supports for organizations inside and outside the formal structure of the government. Fragile societies are able to secure more order through the proliferation of public organizations. These organizations need institutions to support and protect them and their flow of goods and services from opportunism. Similarly, the range of sustainable private organizations is linked to institutions that provide services to these organizations – such as contract enforcement – but also that provide credible commitments by the state not to expropriate the value created by the organizations.<sup>549</sup>

It is telling that the discussion of human evolution and its significance for the formation of institutions also begins with the ‘fact of violence’: namely by a brief discussion of Johnson and Earle’s work on the high levels of violence among foraging (hunter-gatherer) societies.<sup>550</sup> But it is too simple to see North’s anthropology as simply a reworking of Hobbes’ view that, of all forms of uncertainty, “worst of all [is] continual fear, and the danger of violent death”, and that this gives rise to ‘civil society’.<sup>551</sup> Rather, North’s anthropology is more subtle: it is based on the difficulty of cooperation in large groups, and therefore much more concerned with collective action problems than the Hobbesian framework. Firstly institutions, and then the market, provide both the means of coping with a changing, non-ergodic environment, as well as providing means for making cooperation in large groups possible. This cooperation in turn reduces the need for violence, as it is to everyone’s individual interest to participate in the civil society, even one based on rent-seeking and elite rule. It is therefore the function of beliefs and institutions to align incentives in such a way that individually noncooperative behaviour can sustain cooperation in a socially desirable way: the ‘Smithian result’ of market orders. This is, as I have argued, also the main implication of North’s singular discussion of behavioural economics and its anthropological implications: “Non cooperative outcomes are favored, however, where it is very costly to coordinate outcomes, in large groups, and even in smaller groups under private information. In large groups interacting through markets using property rights and a medium of exchange, and with disperse private information, non-cooperative interaction supports the achievement of socially desirable outcomes. Experimental studies have long supported this fundamental theorem of markets.”<sup>552</sup>

This contractarianism, with its emphasis on the transaction cost reducing – and therefore welfare increasing – benefits of even institutional orders based on exploitative elite rule (as a ‘least bad’ option) has a decidedly more Lockean or Humean than Hobbesian flavour at times, but even so it is identifiably based on an underlying pessimistic set of assumptions about the difficulties of achieving lasting cooperation. Both North’s evolutionary argument and his argument for the cooperation-enabling (or incentive-aligning) power of markets rest on this dark view of the effects of uncertainty and violence on human societies and their potential for success

---

<sup>549</sup> *Ibid.*, p. 269-270.

<sup>550</sup> *Ibid.*, p, 52-53.

<sup>551</sup> Hobbes 1651, ch. 1:13. See also Blits 1989.

<sup>552</sup> Cited in North 2005, p. 29. Note that North here uses ‘non-cooperative’ in its technical, game theoretical sense.

and survival. North's final words in *UPEC* serve, I think, well as a summary of how his NIEH framework, with its ambiguities between economic theory and the social context of markets, fit within a larger evolutionary-anthropological perspective deeply concerned with the ability of human institutions to get us lastingly beyond the 'state of nature': "What I have termed adaptive efficiency is an ongoing condition in which the society continues to modify or create new institutions as problems evolve. A concomitant requirement is a polity and economy that provides for continuous trials in the face of ubiquitous uncertainty and eliminates institutional adaptations that fail to resolve new problems. Hayek made this condition a central part of his argument for human survival... Economists have the correct insight that economics is a theory of choice. But to improve the human prospect we must understand the sources of human decision making. That is a necessary condition for human survival."<sup>553</sup>

For North, not only are 'open access orders' good for growth, they are economically *and* evolutionarily necessary to permit human survival on an on-going basis. Choice theory is the foundation for understanding the success or failure of institutional orders in achieving this aim. But even so, nothing is guaranteed: human frailty is so great, our decision-making in the face of uncertainty so limited, that 'even' choice theory may not get us far enough, and that even pro-market beliefs and institutions may fail. Viewed in this way, North's pessimistic anthropology and his optimism about the efficacy of market competition (and the ability of an 'economics approach' to comprehend it) constitute, if you will, the two faces of his NIEH. Each sustains the other but also contributes to generating the internal tensions within the theory.

### **North, Polanyi, and the primitivism-modernism debate**

Achieving North's ambitions, an institutional and evolutionary analysis of long-term economic change such as to explain the Great Divergence and to understand the rise and fall of institutional orders, may then require a different anthropological perspective to overcome these tensions. One possible way of approaching this problem would be to suggest a (re)convergence with Polanyian substantivism. After all, in his 1977 paper North found much in programmatic terms to agree with, such as Polanyi's emphasis on the *longue durée*, his acknowledgement that markets were absent or incomplete in many premodern societies, and his attention to the sociocultural context of economic behaviour and political structures.<sup>554</sup> At the time North wrote the paper, he had to invoke the ghost of Polanyi under the assumption that his influence among economic historians had waned to a mere wraithlike remainder. However, since the 1980s there has been a veritable 'Polanyi revival', with different interpretations and applications of Polanyi's thought gaining influence in social theory, and with Polanyi increasingly taking his place as a fixed star in the constellation of reference points for discussions of markets, global history, and embeddedness. The first international Karl Polanyi Conference held in Budapest in 1986 signalled the rediscovery of his works, and since then there has been a considerable expansion of secondary

---

<sup>553</sup> *Ibid.*, p. 169-170.

<sup>554</sup> North 1977.

literature on his work as well as biographical material.<sup>555</sup> Having discussed at some length the peculiarities and problems of North's NIEH, how should we now consider his relationship to Polanyi and the meaning of his 'challenge' with which I began the present work?

Recall how North acknowledges his debt to Polanyi several times in his early works. More than this, the importance of Polanyi's approach rests precisely in its difference from a purely economics-based one: Polanyi's anthropologism, so to speak, allows him to ask different questions than most economic historians did (or do), ones that North's NIEH is designed to take up. This is most clearly formulated when he discusses the very purpose of his NIEH, in an essay on the place of economic history within economics. As he writes: "while the new economic historians were busy capturing the field of economic history, they left the really interesting questions to the historians, the sociologists and the anthropologists. It is scholars in these fields who have been concerned with the structure of societies, with non-market forms of economic organization and distribution and with economic growth and decline. We left them with the interesting questions but without the tools to produce interesting theory. The result is that the followers of Moses Finley, Karl Polanyi and Immanuel Wallerstein point to the irrelevance of economic theory in analyzing the past... Yet the failure here is not that of economics. The fault lies with economic historians who have not been willing to extend economic theory to make it useful..."<sup>556</sup> Seen from this vantage point, it is clear that North is probably to be classified on the 'modernist' side of the divide – and yet perhaps the most substantivist modernist yet seen. Perhaps more than anything else this is what gives the NIEH its creative tension and makes it such an interesting development within economic historical theory. To understand to what extent a Polanyian approach can therefore complement the NIEH, we must put this Northian critique within the framework of the primitivism-modernism debate.<sup>557</sup>

Takeshi Amemiya has distinguished within the history of the primitivism-modernism debate – which he traces from the German Historical School to the time of Polanyi and Finley - two sub-debates: one concerning to what extent premodern economies were market dependent as opposed to oriented towards (surplus-producing) self-sufficient household economies, and one concerning whether neoclassical type methods are adequate for modelling premodern economic behaviour or whether such economies were 'embedded'. The former Amemiya describes as the primitivism-modernism debate proper, whereas the latter he identifies as an opposition between 'formalism' and 'substantivism' (terms used here in this specific sense). The latter opposition Amemiya specifies as follows: "typically, a formalist presupposes the existence of a well-developed market *and* assumes that consumers and producers seek only selfish interests [*sic*] and all the economic quantities are determined by the market equilibrium that equates supply and

---

<sup>555</sup> E.g. Block and Somers 1984; Polanyi-Levitt (ed.) 1990; Baum 1996; Hart 2008; Dale 2010; Hann and Hart 2011.

<sup>556</sup> North 1976, p. 463.

<sup>557</sup> Some representative arguments for what I call the modernist perspective are e.g. Ogilvie 2001; Silver 1992, 1995; Temin 2001, 2012. For a discussion of the debate, see Bang 1997; Van de Mieroop 1999. I discuss some of these below.

demand. A substantivist, on the other hand, believes that economic decisions are influenced or *constrained* by sociopolitical considerations and institutions.”<sup>558</sup>

Besides the primitivism-modernism opposition from economic history, this opposition uses ‘formalism’ and ‘substantivism’ in a sense perhaps not familiar to most economists; it is drawn from the anthropological literature, taking direct inspiration from Polanyi’s attempt at synthesizing both aspects of the debate on premodern economies.<sup>559</sup> Both of these oppositions are therefore recurrent aspects of the overarching debate I have given the name of primitivism versus modernism. Very crudely outlined, this debate originated with the German Historical School and the work of Karl Bücher, who argued – against his colleague Eduard Meyer – that the economy of ancient Greece was fundamentally based on householding rather than market exchange, and that this meant that the straightforward application of economic theories based on market society could not succeed.<sup>560</sup> This primitivist position formed the conscious basis of the work of Polanyi, Finley, and colleagues, who added to it the notion of ‘embeddedness’, which denied in fact the existence of a separate sphere of the economic in the premodern (especially ancient) world. They argued, be it in somewhat different ways, that because the various modes of economic integration, whether exchange, redistribution, or reciprocity, and the various forms of economic life, such as money and trade, had such radically different meanings in the premodern world compared to modern market societies that conventional economic theory was inapplicable and that a new way of thinking about the integration of social and economic institutions was necessary to understand these societies.<sup>561</sup>

Amemiya’s definition is not without its problems. For one, his definition of substantivism is here so broad that it is difficult to imagine who, outside the most die-hard proponents of rational choice economics, would not be a substantivist. Certainly Polanyi’s own definition of substantivism has been variably interpreted: the ambiguity of the concept is akin to the ambiguity of his closely related concept of embeddedness, in that in both cases one could provide a loose definition and a strict one.<sup>562</sup> One weak version of the substantivist thesis interprets Polanyi as having argued that substantivism means recognising the embeddedness of premodern economies, insofar as our model of economic behaviour in those societies should depart from the institutional context rather than from a pre-given rational choice (‘formalist’) concept alone. On this interpretation, Polanyi’s substantivism and the premises of institutionalism in economics, Old and New, are virtually identical.<sup>563</sup> A more strict definition is given in the interpretation of substantivism as implying additionally that each premodern economic system had its own inner logic, which must be drawn out by the means of economic anthropology and history. This

---

<sup>558</sup> Amemiya 2007, p. 57-59. Emphases added.

<sup>559</sup> Polanyi, Arensberg and Pearson 1957, p. 122ff; Rosser and Rosser 1995; Wilk and Cliggett 2007, p. 1-14; Elardo and Campbell 2006.

<sup>560</sup> Wagner-Hasel 2011; Humphreys 1978, p. 6.

<sup>561</sup> Amemiya 2007, p. 57; Dale 2010, p. 140-141.

<sup>562</sup> Cf. Dale 2011.

<sup>563</sup> Rosser and Rosser 1995, p. 96-97.

reading rests not just on the ‘fact’ of embeddedness, but also on strictly separating the logics of scarcity and subsistence in our understanding of ‘the economic’.<sup>564</sup>

Both can be defended in Polanyi’s work. He himself put it in his famous essay “The Economy as an Instituted Process” as follows: “the substantive meaning of economic derives from man’s dependence for his living upon nature and his fellows. It refers to the interchange with his natural and social environment, in so far as this results in supplying him with the means of material want satisfaction. The formal meaning of economic derives from the logical character of the means-ends relationship, as apparent in such words as ‘economical’ or ‘economizing’. It refers to a definite situation of choice, namely, that between the different uses of means induced by an insufficiency of those means. If we call the rules governing choice of means the logic of rational action, then we may denote this variant of logic, with an improvised term, as formal economics. The two roots meanings of ‘economic’, the substantive and the formal, have nothing in common. The latter derives from logic, the former from fact... the current concept of economic fuses the ‘subsistence’ and the ‘scarcity’ meaning of economic without a sufficient awareness of the dangers to clear thinking inherent in that merger.”<sup>565</sup>

We may say then that North certainly is committed to substantivism in the looser sense, the importance of institutional contexts for determining, in some sense, the nature and outcomes of economic behaviour. However, the stricter interpretation gives a less clear picture: North remains committed to scarcity and competition as facts about institutional orders and choice theory as the best way to model those social dynamics. Even so, he opposes the New Institutional way of interpreting choice in this context to the ‘instrumental rationality’ of neoclassical economics; with ambiguous results, as we have seen.<sup>566</sup> Now if we accept for the moment Amemiya’s schema, it is worth noting (as he does) that while the two aspects of the larger debate have historically been closely related, it is not strictly necessary for an Amemiyan ‘formalist’ to also be a ‘modernist’ and a ‘substantivist’ to be a ‘primitivist’ in this sense. Indeed, as he rightly notes, “a formalist is more likely to be a modernist, but not necessarily so. For example, one who believes that even the modern American economy should not be explained by utility and profit maximization may be said to be both a modernist and a substantivist with regard to the American economy”.<sup>567</sup>

Seen from this vantage point, the peculiarities and interest of North’s NIEH become clearer: because this perspective is of course precisely the one that North has been developing as part of his New Institutional critique of neoclassical economics. Without much explicit acknowledgement, at least after the papers of the late 1970s, North’s NIEH is a major new step in the larger debate Amemiya and others identify. Consider: North’s use of Polanyi as a foil for developing his own theory; the sometimes awkward, sometimes productive position of North’s

---

<sup>564</sup> Elardo 2013, p. 269.

<sup>565</sup> Polanyi 1957, p. 122-123.

<sup>566</sup> North 1993b; North 2005, p. 59; cf. North, Wallis, and Weingast 2009, p. 14-15.

<sup>567</sup> Amemiya 2007, p. 57.



microeconomic methodology between ‘oversocialized’ and ‘undersocialized’ perspectives (to borrow Mark Granovetter’s terminology) and between neoclassical economics and abandonment of choice theory<sup>568</sup>; his emphasis on the importance of institutional contextualization for all economies, modern or premodern, but also his insistence on the difference between the functioning market societies of the West and the dysfunctional part-market societies of the rest; and finally his willingness to concede to Polanyi the relative insignificance of market exchange in many premodern societies (“other allocation systems”), but not to concede to him the ability of a new economic theory, based on the fundamentals of the ‘economics approach’, to understand all economic orders, old and new. What I have identified as the tensions in his work, originating in his taking up of Polanyi’s challenge, amount – following Amemiya’s classificatory scheme – to the search for a theory that is *both modernist and substantivist* in economic history, in his looser sense. But on the other hand the more strictly one takes Polanyi’s meaning of substantivism as opposed to ‘formalism’, in the sense of presupposing a different economic rationality than that based on postulates of scarcity and competition, the less deeply substantivism actually runs in North’s theory.

North’s NIEH therefore from this perspective owes both many of its strengths and of its weaknesses to this new position in the historiographical debate that began with the German Historical School. Although space does not permit me an extensive discussion of this point here, it is worth noting that in this North’s closest commonality within the existing ‘highlights’ of this debate is to the work of Max Weber. The latter’s approach to the historical sociology of ancient societies also sought to combine, arguably, a ‘modernist’ acceptance of the applicability of mainstream economic theory with an institutional and premodern contextualization of the operation of economic rationality. This has been called a ‘third position’ in the debate, in between thoroughgoing primitivism à la Polanyi and equally thoroughgoing modernism as in the work of Meyer, or nowadays Morris Silver. In Weber’s approach, particularly in *The Agrarian Sociology of Ancient Civilisations*, the utility (or profit) maximizing rationality assumed in the ‘formalist’ perspective (once more in Amemiya’s terms) is contextual, dependent on the operation of market exchange with a profit motive by the individuals in question; it is not generalizable (in premodern societies) beyond this specific social-institutional context, and therefore does not characterize the premodern economy as such.<sup>569</sup> North does not often mention Weber in his work. But perhaps North’s ambiguity about the concept of rationality and his difficulties in operationalizing it in different institutional contexts, as I have shown in chapter 3, could at least in part be overcome by adopting this strength of Weber’s approach: a more historically specific and less indeterminate understanding of economic rationality, based on the presence or absence of profit-motivated market orientation in specific cases. Such an understanding could be implemented (or at least complemented) at the level of qualitative

---

<sup>568</sup> Granovetter 1985; Wrong 1961.

<sup>569</sup> Weber 2013 [1909]; Love 1991, p. 49ff.

modelling by the alternative approaches to rationality in microeconomics discussed in that chapter.

### **Revisiting Polanyi's challenge**

Having placed my critical interpretation of North's work in the wider context of these debates on primitivism, modernism, and substantivism, one cannot avoid the question to what extent North's original ambition – to address Polanyi's challenge – was actually successful. It is difficult to evaluate such a claim as a whole, not least because the theoretical differences between them may appear so strong that any commensurable judgement becomes impossible or simply partisan. Moreover, North's direct engagement with Polanyi fades into the background, not to say disappears, after the papers inaugurating the NIEH: his primary interlocutors are rarely anthropologically oriented, but rather other New Institutional economist, public choice theorists, and so forth. Yet on the reading of North's work as constituting a new phase, and a new attempt at a Weber-like 'third position', in the modernism-primitivism debate, contrasting North's strengths and weaknesses with Polanyi's becomes perhaps a more fruitful endeavour.

To some extent, the work of North and of Polanyi show definite complementarities. At the risk of repetition, I will briefly recapitulate them here. Where North has difficulty operationalizing the substance of a belief system in terms of its implications for economic behaviour and market structure in a particular case (other than open access orders), this has been the subject of extensive work by Polanyi, combining anthropology with economic history.<sup>570</sup> Both North and Polanyi propose a bifurcation in economic history expressed in terms of two fundamentally different institutional orders, each of which give rise to a fundamentally different operation of markets. For Polanyi, the difference was between premodern 'embedded economies' and the modern disembedded market economy; for North, the division extends into the present and is more geographically conceived, with Western (and perhaps Japanese) open access orders distinguished from natural state orders in all other cases, but the distinction is otherwise not so different from that between modern and premodern economies. (After all North's 1977 paper is concerned in particular with justifying the study of long-term premodern economic change.) Both, significantly, rely on at least some fundamentals of mainstream economic theory to explain the workings of markets in the 'modern' case: while Polanyi, as mentioned, famously rejected the applicability of economic theory to premodern embedded societies, he accepted the marginalist perspective of Carl Menger and his heirs as the correct model of the economy of market societies.

This is not to say there are no substantial differences as well. As I have mentioned, for North the fundamentals of choice theory, consumer theory and welfare economics remain applicable throughout economic history and regardless of institutional order, and must do, if we are to understand anything about economic change. Indeed, they come fully into their own in the

---

<sup>570</sup> In particular in: Polanyi 1977; Polanyi, Arensberg and Pearson 1957.

case of the open access order: North is quite explicit that neoclassical price theory is useful for describing capitalist markets, just not for historicity or for development purposes, and the more perfect the markets, the more relevant neoclassical economics is. It just does not tell you how to get there.<sup>571</sup>

Polanyi's critique of the fallacy of applying economic theory to what he viewed as societies based on other modes of integration than the market, and his definition of the 'modern' type of market as a price-making market, led him down a very different path of interpretation of premodern economic history than North. Polanyi was, if nothing else, a great exponent of the primitivist interpretation of premodern economies. By contrast North's NIEH framework is still based on the kind of universalization of a logic of competition, scarcity, and (sometimes) self-regard even within an institutional context – indeed in the formation of institutions themselves – that Polanyi, at least in the stricter reading of his substantivism, abhorred. It should be said that North does not explicitly support the main claims of the 'modernists', that markets were fairly developed in many premodern societies and that the standard economic categories and models apply to such economies equally as to present-day ones – although he is mostly silent on this point, his emphasis on differing institutional and belief systems and his acknowledgement of the limited nature of premodern markets is often more in line with Polanyi's.<sup>572</sup> But where Polanyi's perspective is about a revolutionary qualitative change in the nature of the economy, originating in Britain in the Industrial Revolution, North's marginalist-cumulative approach to institutional change goes so far as to deny the existence of political or social revolutions altogether.<sup>573</sup> North does acknowledge the Industrial Revolution as a break, but more conventionally as a break in the growth pattern, rather than a fundamental qualitative differentiation. All this fits the pattern of what I have described as North's combination of modernism with substantivism, in contrast to Polanyi's explicitly primitivist orientation.

So put briefly, they have much in common insofar as they are both institutionalists concerned with understanding the effect of institutions on economic behaviour and doing so from a *longue durée* perspective. Both also see competitive, price-making markets in liberal institutional orders as the natural domain of application of neoclassical thought, but insist on the importance of institutions to understanding the origins and direction of such societies. But in the context of the primitivism-modernism debate they take up fundamentally different positions. What makes their exchange so interesting is precisely that this commonality of ambition is joined to this major tension between them. So reinterpreted by North, Polanyi's challenge becomes the challenge of 'successive endogenisation': i.e. to explain the shift from societies with limited markets and different political and social structures to the 'open access orders' of the West by

---

<sup>571</sup> North 2003.

<sup>572</sup> In particular insofar North in his more 'Williamsonian' moments present hierarchies as often preceding markets as well as substituting for them: E.g. North 1981, p. 41. See also North 1977, p. 706.

<sup>573</sup> Contrast Polanyi 2001 [1944] with North 1995, p. 25.

digging into the underlying sociocultural and institutional orders and their autonomous change. Polanyi's theoretical starting point is economic anthropology; North's is economic theory.

All this is easily understandable from the context of the primitivism-modernism debate as outlined above. But what are the weaknesses of Polanyi's own conception, and to what extent has North succeeded in overcoming them by his NIEH? In his 1977 paper, North rather grandly suggested that such theoretical differences could not be overcome at that level, and rather required testable propositions to demonstrate the efficacy of a transaction cost economics-based approach over the Polanyian.<sup>574</sup> In this work, I have ignored this injunction, since North himself has ignored it. His work has produced little by way of immediately testable new hypotheses, but much by way of innovative new theorizing, and it is the latter that has established such influence as he has. We must therefore address this at the theoretical level: how does Polanyi's approach to 'successive endogenisation' compare to North's?

In my view, the weakness of Polanyi's approach was his conception of markets and their embeddedness, where his case for primitivism rested on a too narrow foundation to be historically sustainable. Decisive for Polanyi was the difference between markets that operated in an embedded way, regulated by social convention and as instruments of political power, and the autonomous operation of 'price-making markets'. This led Polanyi to extensive denials of the presence of market exchange on any significant scale in the premodern world, at least insofar as such exchange for him never (with the possible exception of the Hellenistic era) became sufficiently autonomous to allow the laws of supply and demand to determine the price level. In this way, and *only* in this way, he could deny for the premodern era the applicability of marginalist or neoclassical economics. As Gareth Dale summarizes Polanyi's position on premodern markets: "Those who depended for their livelihoods on incomes generated through buying and selling on markets were greatly outnumbered by peasants whose self-sufficiency insulated them from any meaningful market dependence. Because their survival was not market reliant they were under no economic compulsion to conform to market standards... Except where prices were regulated by custom or decree, peasants were likely to accept virtually any price for their wares... As a result, the supply-demand mechanism could not be said to function".<sup>575</sup>

This strong primitivist position espoused by Polanyi has not fared well under criticism. Even a close colleague and fellow-thinker of his like Moses Finley already warned him against underestimating the relevance of market exchange in the ancient world<sup>576</sup>; contemporary formalist economic historians have taken great pains to establish that, contrary to the substantivist thesis, factor markets were common from antiquity onwards. Moreover, in these markets, they have argued, individuals of whatever status have shown 'price-responsive'

---

<sup>574</sup> North 1977, p. 707, 715.

<sup>575</sup> Dale 2010, p. 145.

<sup>576</sup> *Ibid.*, p. 167-168.

behaviour: in other words, the internal logic of supply and demand, profit and loss, and individual optimisation that forms an abstract market can be read from the historical evidence. Morris Silver has argued the case for ancient Egypt, where he assures us recent papyri "open a window to the kind of world with which economic historians are familiar. Private property, 'economic man', money and coinage and markets for land and labor-power become visible, however dimly"; this as contrary to the substantivist position that "the postulate of wealth-maximizing used by contemporary economists is said to be utterly inappropriate to the "irrational," that is, nonutilitarian, ancients."<sup>577</sup>

Sheilagh Ogilvie's study of Bohemian peasants and their market behaviour has been equally keen to establish the relevance of utility maximisation and 'price responsiveness' as against - in this case - the substantivist view of peasant behaviour attributed to Alexander Chayanov. She reviewed peasant behaviour on the estate of Friedland in the 17th century and found all manner of price-responsive, profit-making activity. Quite in the style of the NIEH, she asks rhetorically: "In early modern Bohemia, as we have seen, markets were not very highly developed: prices and costs were hard to calculate, payments were often made in kind, information was imperfect, and risks were high. But when we observe such an economy, should we assume it is thus because its agents prefer to avoid money, markets, and gain? Or should we ask whether other, external constraints were preventing markets from working well? Markets perform poorly in the absence of supporting institutions... that disseminate information, prevent extortion, and enforce contracts."<sup>578</sup> The central concern of her paper is to refute the claim, central to primitivism (in Amemiya's classification), that "peasants make economic decisions in a way that differs from ours, and cannot be understood using standard economics".<sup>579</sup>

One may doubt whether either Silver or Ogilvie, both of whom refer to Polanyi (among others) as representing the opposing tradition, fully addressed the latter's thesis. Indeed Polanyi did not deny the existence of market exchange in antiquity as such, nor indeed the possibility of wealth-maximizing behaviour in such contexts – even aside from his view that the Hellenistic economic world was by his criteria practically capitalist. But however that may be, I think one can safely state that there is a long-established consensus that 'independent' market exchange and trade were more pervasive in the ancient world - and indeed even before the Bronze Age - than Polanyi's arguments would lead one to suppose, and that moreover it is not easy to sustain the radical claim that such markets were in every respect price takers or proceeding solely on the basis of customary prices and values, rather than determining relative prices through the process of exchange.<sup>580</sup> Although the significance of customary and state regulated pricing as opposed to

---

<sup>577</sup> Silver 2004.

<sup>578</sup> Ogilvie 2001, p. 442.

<sup>579</sup> *Ibid.*, p. 430.

<sup>580</sup> The empirical and theoretical literature on pre-modern markets in antiquity is vast. A valuable recent sample is Scheidel and Von Reden 2010. Some representative works critical of Polanyi's position are: for the Mesopotamian case, e.g. Algaze 2008, p. 14-24; for the classical cases, Manning and Morris 2005; Temin 2012; Silver 1983. Views more favourable to Polanyi's position, although without fully endorsing his claims on the empirics of markets and

'laws of supply and demand' increasingly seems to be a matter of degree rather than of kind in many premodern economic settings, one can certainly wonder whether the former aspect was *as* predominant, and as exclusionary of the formation of an independent motivational logic in market exchange, as Polanyi's argument would seem to require.

Indeed, within economic anthropology some have argued that the dualism of the 'market economy' and the economies based on reciprocity or redistribution does not help us understand the *differentia specifica* of markets and market behaviour in a wider context of what the economic is in a particular case. As Georges Dupré and Pierre-Philippe Rey critically observed, in Polanyian substantivism "the significance of the non-economic merely begs the question since the non-market economy is studied according to the criteria of the market economy... The relation between the two terms can be seen only as mutually exclusive... From the very moment when a single element of the 'essence' of the market is actualised in a non-market economy it then invades, at the same time destroying, the system in which it had manifested itself and replaces it by the market system."<sup>581</sup> It is this view of the essence and coherence of the market as an independent entity, if and when the 'supply-demand price mechanism' operates in them, that prevents a Polanyian viewpoint from comprehending the specificity of that mechanism to the institutional context of a society's economic life.

Here, a Northian contextualization in terms of institutions, beliefs, political structure, and even transaction costs allows for a broader acknowledgement of the presence and extent of profit-oriented market exchange than Polanyi would have allowed, without abandoning the substantivist premise. For this premise may still be maintained, if we abandon Polanyi's extreme primitivist case about the price-making character of markets. Polanyi's approach equally permits a more subtle interpretation of the problem, where considerable profit-oriented behaviour in premodern markets can be allowed, but where such behaviour remained a relatively insignificant part of the economy as a whole. Moreover, it never until modern times led to any systematic economic growth. As Ogilvie is also forced to acknowledge, Polanyi's argument about market *dependence* is a considerably stronger case than the argument about the absence of a Weberian economic rationality she attributes to Chayanov and that Polanyi arguably shared.<sup>582</sup>

With this in mind, let us return to where we began: North's 1977 paper on Polanyi's challenge. In North's 1977 paper we can find how, once this problem with Polanyi's perspective is acknowledged, the former's NIEH approach has if anything *converged* over time with the Polanyian. Recall that initially, North is happy to accept the claim that markets were often substituted for by other allocation systems. But, as he notes, Polanyi's definition of markets is

---

price-making in antiquity, are e.g. Cartledge 1983; Snodgrass 1991; Renger 2005. A middle ground between supporters and opponents of Polanyi's position on classical markets is argued for in Aubet 2013. See also Krul and Dale (2016 forthcoming).

<sup>581</sup> Dupré and Rey 1978, p. 184, 186.

<sup>582</sup> Ogilvie 2001, p. 442.

too restrictive, and ‘price-making’ is a matter of degree rather than all or nothing.<sup>583</sup> Even so, much “non-price allocation” in premodern (and modern) societies can be accepted as a matter of fact.<sup>584</sup> The core difference for him was that, according to North, such choices should be explained in terms of “economic forces”, most importantly transaction costs, rather than it being ascribed to economically “irrational” elements such as “custom, kinship arrangements” and other subjects of “social, cultural, and psychological” studies.<sup>585</sup> The original gambit on North’s part was therefore that, in order to meet Polanyi’s challenge, his NIEH could accept the Polanyian case regarding the limited extent of markets and the need for institutional explanation; but that transaction cost analysis could provide a different kind of substantivist analysis, one rooted in economic choices rather than in “ad hoc” studies of these “social, cultural, and psychological” factors in each society in question. North explicitly also argued that such an approach would work just as well for modern economies (i.e. from the 19<sup>th</sup> century onward) as for premodern ones.

Yet what we find, when considering the critical survey of North’s work I have undertaken here, is that far from substantiating such a case, North has become more and more Polanyian. From the efficiency approach to the medieval manor, with its straightforward neoinstitutionalist social contract model, North has moved to a deeper study of the variety of institutional orders, the importance of *sui generis* and stable belief systems, the relevance of cognitive theory for explaining persisting differences in worldviews, to fundamentally ‘social’ conceptions of rationality (at least some of the time), and to the essential role of political structures, especially states, in dealing with violence and disorder. All this time, transaction costs and the search for the ‘Smithian result’ – social benefit from the interaction of individual noncooperative behaviours – have been a guiding thread of his work. But equally, the more his work has developed, one can see that whether or not this harmonious result obtains, and whether or not transaction costs will become low enough to permit the kind of fully competitive and price-making markets North idealizes, has come to depend precisely on the ‘given’ sociocultural factors Polanyi saw as central to the explanation. In this sense, every acknowledgement of the inescapable importance of such thick anthropological concerns as belief systems, informal social norms, coordinating punishment and reciprocity, kinship networks, signaling traits and group rationality *as analytically prior to ultimate economic choices in terms of transaction costs* – such as the ‘choice’ of institutions ‘at the margin’ - is a concession to Polanyi’s explanatory framework, and a defeat for a project to overcome ‘Polanyi’s challenge’ through New Institutional Economics.

If my reading of North in this work is plausible, at least one inescapable conclusion is this: that for North’s project, and indeed any evolutionary-institutional approach seeking to overcome the weaknesses of Polanyi’s position in the primitivism-modernism debate, to be consistent and

---

<sup>583</sup> North 1977, p. 710-711.

<sup>584</sup> *Ibid.*, p. 710.

<sup>585</sup> *Ibid.*, p. 715, 714, 708.

successful, it will have to take anthropological foundations of economic life as central to the explanatory framework rather than departing from choice theory, precisely as Polanyi sought to demonstrate. This is not to say transaction cost analysis cannot be part of the theoretical toolbox, nor is this to rescue Polanyi's dubious claims about premodern market structure. It is merely to say that once the step towards 'successive endogenisation' is taken, and the economic historian seeks to integrate social institutions into understanding markets and economic change in the premodern world, economic theory alone can no longer suffice as explanatory foundation. To stick to such microfoundations, even if they take the form of bounded rationality and NIE rather than neoclassical premises, makes one unable to resolve either the infinite regress of institutional economics – each institution presumes a previous one – or the problem that the institutions regulating economic behaviour have their origins in nonmarket-like social phenomena, as the problem of 'markets versus hierarchies' illustrates.

One solution is the Williamsonian approach, to ignore it altogether and assume that "in the beginning there were markets"<sup>586</sup>; but if, like North, one does not accept this for the purposes of economic history (as opposed to economics), one must avoid the trap of internal inconsistency and indeterminacy North fell into. And while no theory can be complete and consistent, I would argue this discussion of the North-Polanyi encounter shows at least that the problems of North's NIEH could be overcome in principle – not by accepting Polanyi's empirical claims about the ancient world, but by accepting his argument for the necessity of an anthropologically based substantivism, one that does not fear the tools of economic analysis but does not take them, or the type of market exchange for which they were designed, as analytically primary in historical explanation.

### **What role for economic anthropology? A sketch of background and potentials**

Such a discussion raises inevitably the question what such an anthropological approach to institutionalist and evolutionary economic history would look like. It is of course impossible to fully develop an alternative theory here, as such a task could easily require many book-length works in its own right. A more fully fleshed out discussion will have to be postponed to another occasion. But in order to not simply provide criticisms, but also offer some directions for future research, I will simply sketch what I see as promising theoretical resources here. These should be seen as complementary to, not in lieu of, the various more specific theoretical alternatives to North's explanatory models suggested in the preceding chapters (i.e. on markets, rationality, and sociobiology).

Attempts to reconcile anthropology and economic thought, or at least to find reformulations of economic thought by using anthropological insights and methods, have a long and complex history. The primitivism-modernism debate mentioned above is one domain in

---

<sup>586</sup> Williamson 1975, p. 20. Cf. also Williamson's distancing himself from history and anthropology when asked about this claim: Hodgson and Gindis 2007, p. 378.



which economic anthropology and economic history have interacted, sometimes fruitfully, sometimes mainly polemically. The analogous debate for economic theory (as opposed to history, though not independent from it) is the formalism-substantivism debate in which Polanyi played just as much a key role as in the other. That both have come to be associated so strongly with the work of Polanyi, and that both debates are primarily concerned with the possibilities and obstacles to reconciling economic thinking, including economics itself, and anthropological theory, is once more evidence of the centrality of that figure to any such attempt at future ‘reconciliation’.

But Polanyi is certainly not the only one who has been concerned with finding a synthesis between these disciplines. The idea proposed here, to reconsider such possibilities of synthesis in the light of developments in institutional economics, economic anthropology, and evolutionary theory (including evolutionary approaches to culture and the economy), has a considerable pedigree. Already in the 19<sup>th</sup> century, the investigations of the first social anthropologists like Lewis Henry Morgan, Edward Tylor and Sir John Lubbock had a major influence on the political economy of Marx and Engels in the later stages of their intellectual trajectory, and there has been a Marxist interdisciplinary tradition bridging economics and anthropology since that time.<sup>587</sup> Equally, the Wilhelmine participants in the debates of the German Historical School alluded to above can be seen as forerunners of economic anthropology, just as they can be seen as forerunners of institutional economics. The work of Weber, following on this tradition, is another example of a conscious synthesis between methods of economics – albeit without formalisation – and the subject matter of historians and anthropologists, such as the interrelationship between class, religion, and economic institutional change. The formalist-substantivist debate in economic anthropology is a direct inheritance from these traditions, and both the Marxist and the German Historical School approaches influenced Polanyi greatly.<sup>588</sup>

In the 20<sup>th</sup> century, anthropologists have also in turn been influenced by the development of modern economic theory. Anthropologists like D.M. Goodfellow, Raymond Firth, and Melville Herskovits defended the rational choice and utility maximising approach of the economists of their day, in contrast to the disdain for instrumental rationality assumptions on the part of Malinowski and other founders of economic anthropology. For them, neoclassical economics was better than no economic theory at all, and all social institutional explanation needed to be explicable in terms of the behaviour and interests of individuals, something recognized by the economists.<sup>589</sup> But these reconciliations did not engage much with the historical comparative approach in which different forms of rationality or economic motivation could be compared, rather than a single type assumed; the major exception here is perhaps

---

<sup>587</sup> Krader 1974; Anderson 2010; Patterson 2009.

<sup>588</sup> Hart 2007.

<sup>589</sup> Chamlee-Wright 2015, p. 94-97.

Polanyi's most important forerunner, the French anthropologist Marcel Mauss. However, the latter's work never founded an influential 'school' of its own.<sup>590</sup>

Given the willingness to find a common ground on the part of early 'formalist' anthropologists such as Firth and Herskovits, one would perhaps expect a positive response from the economists' side. But this was not forthcoming. In a rare engagement with the theoretical writings of anthropologists, Frank Knight wrote a damning review of Herskovits' book. As Knight argued, no such reconciliation of disciplines between economics and the humanities was or would be possible: "...effectively uses inference from clear and abstract principles, and especially intuitive knowledge, as a method...[T]he conceptual ideal of economic behaviour is assumed to be, at least within limits, also a normative ideal, that men in general...wish to make their activities and organization more "efficient" and less wasteful...[T]he anthropologist, sociologist, or historian seeking to discover or validate economic laws by inductive investigation has embarked on a "wild goose chase". Economic principles cannot be even approximately verified – as those of mathematics can be, by counting and measuring."<sup>591</sup>

The substantivist side of the debate has been sufficiently mentioned. That 'formalist' anthropology, that is essentially to say neoclassically inspired anthropology, has nonetheless had a thriving career since has been more owed to the willingness of anthropologists to take over wholesale the methodologies of neoclassical economics than to any particular hospitality or encouragement on the part of economists. As Keith Hart notes, "formalist anthropologists, whose knowledge of the history of economics does not appear to have been strong, sacrificed the sensitivity to institutional context shown by leading economists such as Marshall in order to promote a universalizing rhetoric of 'maximizing individuals'. Eventually some of them found that a serious exposure to economics lent weight to their efforts; and formalism broke up into a number of specialist approaches drawing on information theory, game theory, cost-benefit analysis, rational choice, agricultural development and a host of other spin-offs from mainstream economics. By the 1980s many US universities were insisting that economic anthropologists should have a higher degree in economics rather than maintain the foolishness of the recent past."<sup>592</sup>

In recent years the institutional turn has nonetheless affected economic anthropology as well, and it has led some anthropologists to see the New Institutional Economics as a much better tool for achieving the dream of a more shared comparative research programme than the old neoclassical 'formalism' was.<sup>593</sup> Here Williamson, North, and other such writers have been sources of inspiration for the anthropologists – but, as I have pointed out, the reverse has been very rarely the case, with North much preferring the work of paleoanthropology and cognitive

---

<sup>590</sup> Hart 2007.

<sup>591</sup> Knight 1999 [1941], p. 111-113. Herskovits nonetheless continued to insist on the possibility and necessity of a joint research programme.

<sup>592</sup> Hart 2007.

<sup>593</sup> Acheson 1994; Ensminger 2002.

studies of hunter-gatherer societies to the kind of comparative social studies of economic anthropology, whether New Institutionalism in inspiration or not. Once again the effort seems to be foundering on the economists' unwillingness to venture beyond the established disciplinary boundaries and literatures.

Is a reconciliation therefore possible, even if desirable? I think so, but it can only be based on a thorough knowledge of both fields, including the historical comparative dimension that was the focus of Polanyi's work. Insofar the New Institutional turn shows that the 'new mainstream' of economics is searching for a way to integrate 'the social' in Polanyi's sense into its understanding of economic change, and insofar the aporias of even North's approach – surely the most sophisticated of attempts at synthesis on this basis – show that the combination of NIE theory with bounded rationality and the evolution of cognition is still not sufficient to achieve this objective, the way should be free for a meeting of the minds between anthropological traditions and other traditions in economic thought. Behavioural economics already demonstrates that such attempts at unification can develop exciting new insights that take us far beyond what was taken for granted on either side of the disciplinary divide. In experimental economics, anthropology and economics are increasingly joining in a research programme to test the implications of game theoretical analysis in the field, with an emphasis on understanding the components of the 'moral economy' of agents in strategic situations and cross-cultural comparison. Both have been preoccupations of economic anthropology since its inception. It is therefore all the more disappointing that an author like North does relatively little with this material.

If nothing else, social and economic anthropology since Mauss and Polanyi has been dedicated to the comparative exploration of the interaction between beliefs, social and economic institutions, and individual behaviour and self-representation. It seems wasteful and obtuse to sideline this wealth of insight altogether. The formalist-substantivist debate in anthropology can perhaps be bypassed now that the mainstream economic theory that inspired the 'formalist' side is itself convergent with the concerns and aims of the substantivists, be it in a preliminary and incomplete way, as the example of North shows. Equally, the 'formalists' were right to think that economic anthropology can learn much from the tools and methods of the economists, in particular now that the 'new mainstream' and the behavioural and evolutionary economic 'schools' have been breaking new ground in understanding the formation of institutions, the possibilities of cooperation, and the interaction between beliefs, norms, and economic behaviour. Finally, the economic historians can and should have a fruitful exchange with these strands of economic thought, for they need theory to frame their observations and research questions, even as they in turn contribute essential empirical material and analysis of long-term trends to the comparative study of economic institutions that emerges from this synthesis. The time for reconciliation, therefore, is now.

The anthropologist David Graeber's discussion of value versus values is perhaps helpful here. As he points out, one of the weaknesses of the substantivist perspective was precisely its

inability to study questions of value and commensurability as an integral part of their project. The substantivist approach remained at the level of empiricism and providing useful categorisations like the Polanyian trinity, but it did not theorise value: “it is one thing to say 'societies' have different ways of distributing goods. It is another to explain what particular members of the society in question think they're doing when they give gifts, or demand bridewealth, or exchange saffron for ivory in a port of trade.”<sup>594</sup> In order to make this clear, I will have to have recourse to some examples from the practice of economic anthropology. These examples concentrate on the various roles material goods and their allocation play in different societies, and the (in)commensurability between them, as an illustration of the relationship between the allocation systems themselves. The relationship between people appears, in the study of economics, as mediated by material goods and their distribution. As Arjun Appadurai has suggested, the valuation of goods in different allocation systems is what determines their commensurability or incommensurability. Drawing on Georg Simmel, another major exponent of the German Historical School, he suggests that goods (or commodities), the objects of material culture, can only be valued according to the same norm or standard of value if they are exchanged according to that standard. In other allocation systems regarding these goods, the valuations may be different depending on the social system in question, and therefore commensurability is absent. This is to my mind one of the main tasks of economic history reconstructed along the lines suggested in this paper: “exploring the conditions under which economic objects circulate in different *regimes of value* in space and time”.<sup>595</sup>

A similar important anthropological concept, too often missing in institutionalist discussions of economic history, is commensurability. What I mean by this term is the measurability or comparability of one good or aspect of economic and social life in terms of another. It is characteristic of modern society that almost anything should be seen as commensurable with anything else via the medium of money. Yet this is not so for much of premodern life, and indeed not for all of modern life. Economic anthropology can teach economic historians much, I think, about the importance of studying commensurability relations to understand the embeddedness of economic life in structures of meaning, belief, and function (willed or unintended). The very nature of the subjects of economic activity, such as goods or labour, depends on those structures within which they are made comparable with others. As Maurice Godelier said: "economic activity then appears as activity with many different meanings and functions, differing each time in accordance with the specific type of relations existing between the different structures of a given society. The economic domain is thus both external and internal to the other structures of social life, and this is the origin and basis of the different meanings assumed by exchanges, investments, money, consumption, etc., in different societies... The existence of a currency thus has not the same meaning in a primitive economy [*sic*] as in a

---

<sup>594</sup> Graeber 2001, p. 11.

<sup>595</sup> Appadurai 1986, p. 4. Emphasis original.

Western commodity economy. One and the same reality may take on different and unexpected significances through belonging to different social wholes."<sup>596</sup>

This may move us beyond Polanyian embeddedness itself. As Godelier wrote on another occasion, reviewing Polanyi's theory of the embedded economy: "At no point does Polanyi really ask himself *why* it is that the economy occupies a particular position in a given society, and why it functions, 'embedded' or not, within kinship or political or religious relations... The notion of 'embeddedness' itself gives rise to problems which deserve further discussion. For one cannot conceive of any form of economy being compatible with any form of kinship, religion or government, and vice versa".<sup>597</sup> The mental model of the world is, following Godelier, then to be seen as a property of the particular forms of economic activity – both production and allocation – in question. "Once one sets oneself the task of building a comparative theory of the shifting place of the economy in society, one is duty bound to pose the question... of the role of economic relations, of their effects upon societies' functioning and evolution. The fundamental problem is then to know up to what point and through what mechanisms the economic relations – the relations of human beings with each other in the production and redistribution of their material means of existence – determine the functioning of this evolution."<sup>598</sup>

Nor is it merely a matter of 'property rights'. An interesting example is the view of the classical scholar, Richard Seaford, who suggests that the need for the community as a whole, mediated by the temples, to pay for public works, and other common interests such as hiring mercenaries, required use of temple treasure. Although Seaford's work is controversial, and I cannot here judge to what extent it will stand up to future scrutiny, it may serve at least as a practical illustration of how one might think about institutions and economic behaviour in an 'anthropological' way. The distribution of sacrificial meat had coordinated and constituted the *polis* as a community hitherto, and established relations of commensurability between individuals in social 'exchange', something akin to citizenship (if not necessarily on an equal basis). But for these kind of payments, durability as well as commensurability, and the public faith in the payment that underpins this commensurability, is required. This in turn suggests the need for a more permanent and generalizable form of payment than such meats. So, for Seaford, as the *polis* developed more coherently, the iron spits – cheap, easily replaceable, durable, lightweight, and commonly possessed – that the meat had been roasted on substituted as payment for the meat itself, essentially forming the first money as a means of payment. It is for this reason that the word for such a spit, *obolos*, became the word for the cheap form of coin in ancient Athens (and indeed was so in modern Greece until it introduced the Euro). These spits, in turn, were replaced by coins of precious metal, such as the Athenian silver and the Ionian electrum.<sup>599</sup>

---

<sup>596</sup> Godelier 1978, p. 63, 87.

<sup>597</sup> Godelier 1986, p. 189. Emphasis added.

<sup>598</sup> *Ibid.*, p. 190-191.

<sup>599</sup> Seaford 2004, p. 110-111.

The advent of the silver *obeloi* (and *drachmai*) used as money in archaic Greece ultimately led to a commensurability between two different kinds of allocation systems and of two different kinds of mentality about the social function of economic activity that had pervaded Homeric times: “on the one hand the aristocratic circulation of prestige objects (treasure, sometimes of precious metal) – as interpersonal gifts or sometimes ransom or prizes – stored temporarily or permanently in private houses, and on the other hand the egalitarian sacrificial distribution of meat that integrates a group or community.”<sup>600</sup> Money can only arise when this commensurability between goods or labour measured in it is already established or coming into being, and in turn expresses the commensurability between these economic spheres, for example as it increasingly came to exist in the archaic and classical periods of Greece.

In other words, the commensurability of money in the market depends on the political structure of the society in question, and at an even deeper level, on the 'constitutive' language games - incorporated in habituation and institutions - of that society. The mental component of market activity, even in straightforward monetary exchange, must be traced back to its social, political, and religious roots. The particular interrelation of these that characterises a society and its economy is the embeddedness of that society, viewed anthropologically. Different media can be used for the different purposes of money in this sense whenever commensurability between different systems of value is not complete. Therefore in such societies, as the anthropological literature demonstrates, the economy is an essentially divided entity that nonetheless in its entirety reproduces the society in question. Indeed, the fungibility of money is perhaps never 'total' at least until the advent of the modern market society, which could indeed be defined in part by the (seeming) completeness of this commensurability and thereby of this fungibility.<sup>601</sup>

Such brief sketches can only indicate how the ways of seeing of economic anthropology, such as understanding the institutional and belief structure underlying exchange – so important for North's approach – in terms of value and commensurability, could contribute to an institutional-evolutionary economic history. In the evolutionary dimension, besides the potentials of gene-culture coevolution approaches to institutional evolution mentioned in the previous chapter, one could take up another cue from evolutionary economics: One such avenue is proposed by Geoffrey Hodgson in his discussion of social formations and their structuration. He points to the possibility of identifying, similar but not identical to the Marxist 'mode of production' approach, the interaction between the dominant deep level socio-cultural habits and norms of thought and the prominent 'provisioning institutions' of that society. In this way, we can unite the evolution of institutions with differentiation in types of economies according to their dominant production and distribution relations.<sup>602</sup> Combined, this seems to come close to the 'reproduction perspective' proposed by Godelier, i.e. identifying the evolution of the inner 'logic'

---

<sup>600</sup> *Ibid.*, p. 111.

<sup>601</sup> A point made by Von Reden 2010, p. 4-5.

<sup>602</sup> Hodgson 2001, p. 336-345.

of social relations and of the sphere of the economic within them, but enriched by contemporary insights in evolutionary theory.<sup>603</sup>

Finally, as an integrative possibility one could consider the potentials of the 'deep history' (also called 'big history') that has become an exciting new way of presenting global history. Such an approach seeks to integrate the evolutionary scale of human history with the long-term institutional scale as well as the more everyday level of history of much more specific periods and places: a thoroughgoing institutional, evolutionary, and *longue durée* approach based on the interaction between different timescales and causal levels. As Andrew Shryock and Daniel Lord Smail write in their introduction to an interdisciplinary collection on this theme, human evolution has until relatively recently been something difficult for the social sciences to fully incorporate into the conception of history itself. It nonetheless gives us information not just about the development of hominin species at the phenotypical level, but is a constantly relevant factor in the very long-term patterns and changes in human social and cultural history.<sup>604</sup> As they put it, "histories can be written from every type of trace, from the memoir to the bone fragment and the blood type... The ongoing merger of history and social science has produced an intellectual world in which most scholars realise that intentions are social products, and the grounds for their production are largely beyond the control of individuals and their desires. In this realisation, the methodological distinctions that once separated history from anthropology and archaeology all but disappear."<sup>605</sup>

The promise of deep history, then, is to connect the very long term timescale of evolutionary change with the 'historical' level of the formation and change of social institutions, by the mediation of the physical human needs, produced goods, and not least the evolved mental life of humans as a *zoon politikon*. This allows us to dispense with "the unhelpful assumption that the deep past is best understood in relation to a fixed human nature or universal behavioural tendencies (such as 'economising', 'rational choice', or 'kin selection')" or "the belief that certain cultural forms, such as 'ethnicity', are quintessentially modern and that similar processes of group identification are not found in the past."<sup>606</sup>

Such concerns may seem remote from the world of recorded history or the domain of economic theory. But I suspect that they are not, and that future research will show the convergence of this level of analysis with the concerns of historians, anthropologists, and economists. If at the lower level the anthropological approach to changing regimes of production, value, commensurability, and institutions unites economic history with institutions as emergent properties of human sociality; and if at the middle level of abstraction gene-culture coevolution forges an intimate connection between the socially determined and variable habits and institutions of social life and the pressures of Darwinian selection on humans in general; then

---

<sup>603</sup> Cf. Godelier 1978, p. 98, 105, 107.

<sup>604</sup> Shryock and Smail (eds.) 2011, p. 12.

<sup>605</sup> *Ibid.*, p. 13.

<sup>606</sup> *Ibid.*, p. 13.

at the highest level of abstraction our very human needs and means of production, our foodstuffs and our clothing, our responses to climate and to the material environment, and our very sociality, need and regard for others are part of a unified evolutionary *and* socio-economic history stretching back into deep time. "We demonstrate", Shryock and Smail report, "how humans have used food and kinship to create worlds that, by comparison with other primate standards, are highly dependent on an awareness of past and present. As social projects, these shared substances are media of 'kinshipping', a tactic for moving through time and space that requires networks of relationships and exchange... Kinshipping is possible only if... a formative relation preexisted and continues to define the new and particular. It has no point of origin. Likewise, the coevolutionary spiral, which envisions two genealogies entwined and feeding off each other, displaces metaphors of genesis, revolution, and the biblical fall."<sup>607</sup>

There is no denying that sketches of this kind are highly abstracted and operate on a theoretical level not immediately operationalisable in specific economic historical cases. To that extent, I do not seek – as has been fashionable for some time – to reject North's NIEH approach as a grand narrative that must fail as all grand narratives do. I am on the contrary very appreciative of North's efforts in constructing a grand narrative seeking to integrate institutional economics with a substantivist quest to answer Polanyi's challenge, and insofar as I criticize his work and propose alternatives, it is to build a better and more plausible narrative, but not a less grand one. To some extent, this is the natural consequence of the philosophical and anthropological lens through which I have viewed North's NIEH in this work, and with this historiographical focus in mind I have consciously avoided engaging in much empirical discussion on the merits of any particular claim about concrete historical events in North's account, with only occasional exceptions. The task of applying these insights regarding the promises and pitfalls of North's approach to more specific case studies falls onto further research, as does further exploration of the philosophical and historical context and implications of the main themes discussed in this work.

---

<sup>607</sup> *Ibid.*, p. 17, 19.



## Conclusion

Allow me now, as is conventional, to reflect on the argument of this present work as a whole. I have begun this discussion of Douglass North's NIEH by trying to locate it in the context of the primitivism-modernism debate. More specifically, I have tried to understand the *differentia specifica* of this approach, even as contrasted with other NIE approaches, by reading North through the lens of his engagement with the challenge of Karl Polanyi and what this has meant for the direction in which he has developed his theory. In some respects this reading has not differed so greatly from existing interpretations of the work as a whole, such as the observation that his work has gone through several stages: from a more rational choice contractarian 'efficiency approach' to institutions to a much more thoroughgoing search for the preconditions of effective and growth-enhancing markets and the moral orders that sustain them. The latter stage has shown a tremendous expansion of the scope of his work, incorporating a number of subjects and problems normally left by economic historians to anthropologists, sociologists, or even cognitive scientists. This process of incorporation I have described as the 'successive endogenisation' of more and more elements of the social-institutional 'background' into North's theory of the effect of institutions on economic behaviour and performance.

For some, this process has been a general success, perhaps even the way forward for the reconciliation of old and new institutionalisms, or of economics and other social sciences: such commentators speak of a "fruitful attempt at integration" of different disciplinary approaches, or even "a powerful example of how persistent and well placed confidence and hard work can productively transform the status quo".<sup>608</sup> Others have seen in it a case of deceptive heterodoxy, an attempt to reconcile the irreconcilable differences of mainstream economic theory and the domains of the other social sciences. In this perspective, North's work represents a prime example of "economics imperialism" or an opportunistic attempt at reclaiming relevance for a 'neoclassical' economic history to impress the colleagues in the economics departments.<sup>609</sup>

I have sought to argue for neither of these positions, but rather to find a middle ground. For me, this has meant taking North seriously and at his word as a critic of neoclassical economics (in his own terms) and as searching for a way to provide, to paraphrase Sally Humphreys, a 'new theory of comparative economic history' that would do better than Polanyi's.<sup>610</sup> But it has also meant an opportunity to assess to what extent he has succeeded in doing so. It is not easy to make such a judgement, especially not about such a large, variable, and sweeping 'grand narrative' as that of North. For this reason, and to make my critical observations applicable, I have sought to understand what criteria for North distinguished his approach from Polanyi's and what he expected they would contribute to improving on the latter's theory. This I have found in the importance of the 'economics approach', which he so clearly contrasts with Polanyi's

---

<sup>608</sup> Groenewegen 2001, p. 812; Ménard and Shirley 2014.

<sup>609</sup> Milonakis and Fine 2007; Boldizzoni 2011.

<sup>610</sup> Humphreys 1969, *op. cit.*

anthropological foundations in his 1977 paper. But this raises new questions: what is this approach, and how has North sought to make it work?

By examining in his work three important themes, or methodological components – markets, rationality, and evolution – I have sought to clarify what this ‘economics approach’ has come to consist of over time in North’s NIEH. I say ‘over time’ because, as mentioned, it has by no means been a static model to apply straightforwardly, but has undergone many shifts of emphasis and structure. In the first chapter I have sought to present these in general terms, and then to examine them more closely in the subsequent thematic chapters. The real nature of this ‘economics approach’ and how its superior explanatory value is to be demonstrated has proven remarkably elusive: North’s early confidence that his propositions would, by a straightforward application of transaction cost economics, prove to be more testable and therefore more scientific has not really been borne out in practice. Indeed, as North has sought to reveal more and more of the iceberg of beliefs, institutions, forms of cooperation and ‘irrational’ behaviour underlying the peak of behaviour in market exchange observable above the waterline, it has become less rather than more clear what work is really done for his theory by his recurrent appeals to the merits of choice theory and its supposed scientific foundations.

Daniel Ankarloo has made the point in this context that North’s approach to markets rejects the universality of the market to emphasize the importance of an institutional approach, but then seeks to reintroduce this market-like nature of individual behaviour in economic and political institutional structures through the back door. I agree with Ankarloo that North has a tendency to do this, but rather than simply critiquing North in this way, I have used this as another stepping stone for understanding North’s NIEH. I have sought to understand why this strange ambiguity, not to say inconsistency, should appear. Partially, its origins seem to me ideological: North’s commitment to the ‘economics approach’ seems as much a commitment to a contractarian liberalism, inspired by Hayek and a certain (mis-)reading of Adam Smith, as a claim about methodology. Central for North is to safeguard the claim that markets, when provided with the correct institutional supports in norms and political structures, have a dual benefit: enabling peaceful cooperation, where otherwise violence would prevail, and enabling economic performance or economic growth, where otherwise there is stagnation and rent-seeking. This premise recurs throughout as a *quod erat demonstrandum* as well as as a point of reference, which I have referred to – using North’s own terminology – as the ‘Smithian result’ (though one may doubt its relevance to Adam Smith’s work).

One must keep in mind in this regard that North not only seeks to answer the challenge of Polanyi’s primitivist substantivism, but – more in the background – sought to provide an alternative to the Marxist explanatory framework as well. Before even becoming a neoclassical economic historian, North was (briefly) a Marxist, and he has argued that this continued to shape

his concerns for explanations over the *longue durée*.<sup>611</sup> His ideological relationship to the liberal institutionalism of Max Weber is less easily defined. Although North's work bears similarities to that of Weber in terms of his search for a middle ground between 'oversocialized' and 'undersocialized' views, Weber's influence is more directly felt within the 'new economic sociology' of Granovetter (from whom this dichotomy is after all derived) than directly in North's work.<sup>612</sup> In North's NIEH works Weber is mainly cited in the context of his claims about the origins of the spirit of capitalism and his equally famous definition of the state as the holder of the monopoly of violence.<sup>613</sup> Even so, more research is needed on the comparison between Weber and North in light of their shared concerns with contextual rule-following behaviour and the importance of beliefs for constraining such behaviour.

In any case a purely ideological or political explanation is not sufficient: one must also explain the work in terms of its own scientific content, and one cannot simply dismiss North on political grounds alone. Proceeding to analyse North's discussions of rationality, game theory, and individual choice, I found a recurring similar ambiguity. On the one hand there is a clear desire to found his work on the (allegedly) solid foundations of conventional choice theory, be it in a more boundedly rational form, expressed for example in his frequent use of terms derived from game theory; and simultaneously he insists on the fundamental sociality of preferences, the independent causality of beliefs, and rejects rational choice explanations as inadequate. I have sought to explain his turn to evolution as one attempt to deal with this ambiguity, by recasting the 'efficiency' of competitive markets in correct institutional orders as a case of evolutionary adaptiveness rather than an unexamined causal relationship between a classical liberal order and economic prosperity.

The latter relationship, I think, is difficult to maintain without accepting a rather simple version of a Panglossian neoclassical view of the market, precisely the sort of model that North had set out to reject. One can add to this utility of evolutionary theory the importance of evolutionary (and anthropological) analysis to his deeper examination of the *origins* of institutions – a stage even beyond seeing institutions as the origins of market performance. Even so, North sought once more to apply this evolutionary perspective in an ambiguous way, distancing himself from 'Darwinian' approaches and seeking a rather more 'Lamarckian' approach instead by insisting on the difference made by intentionality. As I have argued, this distinction is not as plausible conceptually as he seems to suggest. Moreover, in any case he succeeded neither in substantiating his claim about open access orders in history nor in doing justice to evolutionary theory, anthropology, or experimental economics: all fields I think it is of

---

<sup>611</sup> North 1986; North 1995, p. 252; North 2009, p. 160; Klein and Daza 2013, p. 525-526. Cf. Heijdra, Lowenberg, and Mallick 1988 on contrasting Marxism and New Institutionalism. Of course North has in turn had his share of criticisms from a Marxist viewpoint, such as Ben Fine, Dimitris Milonakis, and Daniel Ankarloo – see the discussion in chapter 2.

<sup>612</sup> E.g. Swedberg 1998.

<sup>613</sup> North 1993, p. 7, North 2005, p. 43; North, Wallis, and Weingast 2009, p. 17, 22.

essential importance to integrate into the process of ‘successive endogenisation’ if a Northian type analysis is to accomplish its aims.

This then brings me back to the larger framework of primitivism versus modernism. As an overarching explanation of the ambiguities and problems of North’s work, I propose not just the political-ideological commitment to an ‘ordoliberal’ society. More importantly, I suggest that North has sought to find a middle ground within this debate that would, in a sense, allow him to have it both ways: to have the scientific prestige and welfare economic implications of conventional microeconomics and its view of the market, but also the substantivist claim about the dependence of the functioning of markets on a particular institutional order. He would apply, as the modernists would do, the premise that economic theory – in this case New Institutional Economics – is capable of explaining economic behaviour and institutions in all periods of history, but he would also allow – with the primitivists – that often market exchange was not the main allocation system in the past and that the choice between market exchange and other forms of social integration depended on sociocultural and ideological motivations rather than on an assumed individual orientation towards gain or profit. But in both of these cases, he has veered back and forth as has suited his argument or his leanings at a given time, and the result is a powerful and intriguing, but ultimately internally inconsistent and methodologically conflicted grand narrative.

Ultimately, I believe that for a Northian project to succeed, North’s own attachment to a more conventional microeconomics of markets and a view of political-institutional orders as market-like cannot be maintained. North has from the beginning sought to demonstrate that a well-ordered market can do what the more liberal wing of neoclassical economics has claimed it can do, but *for different reasons* than those economists argued: in this he is not so different from his immediate inspirations, Hayek and Coase. However, in giving up the conventional reasons in defence of the efficacy of markets and the applicability of neoclassical economics to the study of markets (and other allocation systems) in history, he gave up the standard ways of defending these claims about well-ordered markets. Moreover, at the level of scientific analysis – rather than normativity – his rejection of neoclassical views of economic behaviour and rationality combined with this contractarian, classical liberal view of the emergence of order and cooperation in historical societies threatened to leave him stranded in a methodological quicksand, with little solid ground to base an argument on. North’s greatest strengths, his innovative and at times brilliant application of a wide range of tools from institutional economics, cognitive science, public choice theory, game theory, and so forth, derive on my reading from his search for a new foundation. But as I have tried to show, while he has come far, he has not (yet) found it.

As a ‘third position’ between primitivism and modernism in economic history, overcoming the challenge of Karl Polanyi, Douglass North’s NIEH does not, on my account, succeed just yet. But this does not mean North’s toolset should be cast aside: for example, the usefulness of thinking in terms of institutions as rules, transaction costs, uncertainty, or the problem of

cooperation in the face of perennially potential violence, does not depend on the specific explanatory frame North's NIEH provides. And on the whole, the larger ambition expressed in the 'challenge of Karl Polanyi' remains an important advance in economic historical theory. By this formulation, and his attempt at answering it by means of his NIEH theory, North has made the primitivism-modernism debate and the inadequacy of conventional economic theory for overcoming it central to the discussion in economic history. His progressive insight into the need for a 'sociologically holist' explanatory approach to produce a plausible alternative, and the realization that this entails incorporating a great deal of analysis on topics like norms, beliefs, power, identity, and cultural markers, not coincidentally precisely those dimensions generally disfavoured by conventional 'economistic' (in Polanyi's terms) or 'formalist' (in Amemiya's) explanations in economic history, constitutes a tremendous step forward for convergence in the historical social sciences.

Of course, North's functionalist gambit has its own risks, such as just-so stories, teleology, and dubious appeals to evolutionary mechanisms. Some examples of these we have already seen in chapter 4. Moreover, if my argument in this work is correct, North himself has not been able to fully develop this project beyond the attachment to the microfoundations of economic theory and the liberal contractarian 'pessimistic anthropology' that have been persistent features of his theoretical career. A different approach might draw on the resources I have identified, such as heterodox microeconomics, gene-culture coevolution, theory in economic anthropology, and a different view of markets. It could do this without giving up on the contributions of North in terms of the NIEH's toolkit, as described above. It could do so without abandoning the Northian insights about the importance of institutions, beliefs, and the evolution of cooperation. But most importantly, it could do so without giving up on answering the challenge of Karl Polanyi – and perhaps make some headway in finding a resolution to the primitivism-modernism debate.

## Bibliography

- Acheson, James (ed.). 1994. *Anthropology and Institutional Economics*. Lanham, MD: University Press of America.
- Alchian, Armen A. 1950. 'Uncertainty, evolution, and economic theory'. *The Journal of Political Economy* 58:3, p. 211-221.
- Aldrich, Howard E. et al. 2008. 'In defence of generalized Darwinism'. *Journal of Evolutionary Economics* 18:5, p. 577-596.
- Alexander, Richard D. 1987. *The Biology of Moral Systems*. New York, NY: Aldine De Gruyter.
- Algaze, Guillermo. 2008. *Ancient Mesopotamia at the Dawn of Civilization: The Evolution of an Urban Landscape*. Chicago, IL: Chicago University Press.
- Amemiya, Takeshi. 2007. *Economy and Economics of Ancient Greece*. Abingdon: Routledge.
- Anderson, Kevin B. 2010. *Marx at the Margins*. Chicago, IL: University of Chicago Press.
- Ankarloo, Daniel. 2002. 'Using Karl Polanyi as a stepping stone for a critique of the new institutionalist orthodoxy'. CRIC workshop paper, University of Manchester.
- 2006. 'New institutional economics and economic history'. Historical Materialism workshop paper, University of London.
- Ankarloo, Daniel and Giulio Palermo. 2004. 'Anti-Williamson: a Marxist critique of new institutional economics'. *Cambridge Journal of Economics* 28:3, p. 413-429.
- Aoki, Masahiko. 2001. *Towards a Comparative Institutional Analysis*. Cambridge, MA: MIT Press.
- Appadurai, Arjun (ed.). 1986. 'Introduction: Commodities and the Politics of Value', in: *The Social Life of Things: Commodities in Cultural Perspective*, p. 3-63. Cambridge: Cambridge University Press.
- Arnsperger, Christian and Yanis Varoufakis. 2006. 'What is neoclassical economics?'. *PAE Review* 38, p. 2-13.
- Arrow, Kenneth J. 1974. 'General economic equilibrium: purpose, analytic techniques, collective choice'. *The American Economic Review* 64:3, p. 253-272.
- Aubet, Maria Eugenia. 2013. *Commerce and Colonization in the Ancient Near East*. Cambridge: Cambridge University Press.
- Aumann, Robert and Adam Brandenburger. 1995. 'Epistemic conditions for Nash equilibrium'. *Econometrica* 63:5, p. 1161-1180.
- Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York, NY: Basic Books.

- Bang, Peter Fibiger. 1997. 'Antiquity between "primitivism" and "modernism"'. Aarhus Centre for Cultural Research Working Paper 53-97.
- Bang, Peter F., Mamoru Ikeguchi and Hartmut G. Ziche. 2006. *Ancient Economies, Modern Methodologies: Archaeology, Comparative History, Models and Institutions*. Bari: Edipuglia.
- Bates, Robert H. 1988. 'Contra contractarianism: some reflections on the New Institutionalism'. *Politics & Society* 16:2, p. 387-401.
- 2014. 'The New Institutionalism', in: Galiani and Sened (eds.), p. 50-65.
- Baum, Gregory. 1996. *Karl Polanyi on Ethics and Economics*. Toronto, ON: McGill-Queen's University Press.
- Becker, Gary S. 1981. *A Treatise on the Family*. Cambridge, MA: Harvard University Press.
- Becker, Marcus C. 2004. 'Organizational routines: a review of the literature'. *Industrial and Corporate Change* 13:4, p. 643-677.
- Beckert, Jens. 2009. 'The Great Transformation of Embeddedness: Karl Polanyi and the New Economic Sociology', in: Hann and Hart (eds.), *Market and Society: The Great Transformation Today*, p. 38-55. Cambridge: Cambridge University Press.
- Benham, Alexandra and Lee Benham. 1998. 'The Costs of Exchange'. Ronald Coase Institute Working Paper no. 1.
- Binmore, Ken. 1994. *Game Theory and the Social Contract vol. 1: Playing Fair*. Cambridge, MA: MIT Press.
- 1998. *Game Theory and the Social Contract vol. 2: Just Playing*. Cambridge, MA: MIT Press.
- 2005. *Natural Justice*. Oxford: Oxford University Press.
- 2007. *Game Theory: A Very Short Introduction*. Oxford: Oxford University Press.
- 2015. 'Why all the Fuss? The Many Aspects of the Prisoner's Dilemma', in: Martin Peterson (ed.), *The Prisoner's Dilemma*, p. 16-34. Cambridge: Cambridge University Press.
- Bliege Bird, Rebecca L. and E.A. Smith. 2005. 'Signaling theory, strategic interaction, and symbolic capital'. *Current Anthropology* 46:2, p. 221-248.
- Blits, Jan H. 1989. 'Hobbesian fear'. *Political Theory* 17:3, p. 417-431.
- Block, Fred. 2001. 'Introduction', in: Karl Polanyi, *The Great Transformation: The Economic and Political Origins of Our Time* [1944]. Boston, MA: Beacon Press.
- 2008. 'Polanyi's double movement and the reconstruction of critical theory'. *Papers in Political Economy* 38. URL : <http://interventionseconomiques.revues.org/274>

- Block, Fred and Margaret R. Somers. 1984. 'Beyond the Economistic Fallacy: The Holistic Social Science of Karl Polanyi', in: Theda Skocpol (ed.), *Vision and Method in Historical Sociology*, p. 47-84. Cambridge: Cambridge University Press.
- Boehm, Christopher. 2012. 'Costs and benefits in hunter-gatherer punishment'. *Behavioral and Brain Sciences* 35:1, p. 19-20.
- Boettke, Peter. 1990. 'The Theory of Spontaneous Order and Cultural Evolution in the Social Theory of F.A. Hayek'. *Cultural Dynamics* 3:1, p. 61-83.
- Boldizzoni, Francesco. 2011. *The Poverty of Clio: Resurrecting Economic History*. Princeton, NJ: Princeton University Press.
- Bowler, Peter J. 2014. 'Herbert Spencer and Lamarckism', in: Francis and Taylor (eds.), *Herbert Spencer: Legacies*. Abingdon: Routledge.
- Bowles, Samuel and Herbert Gintis. 2011. *A Cooperative Species: Human Reciprocity and its Evolution*. Princeton, NJ: Princeton University Press.
- Boyd, Robert and Peter J. Richerson. 1985. *Culture and the Evolutionary Process*. Chicago, IL: University of Chicago Press.
- 1987. 'The evolution of ethnic markers'. *Current Anthropology* 2, p. 65-79.
- 1988. 'The evolution of reciprocity in sizable groups'. *Journal of Theoretical Biology* 132, p. 337-356.
- 2005. *The Origin and Evolution of Cultures*. Oxford: Oxford University Press.
- Boyd, Robert et al. 2003. 'The evolution of altruistic punishment'. *PNAS* 100:6, p. 3531-3535.
- Brown, Vivienne. 1997. 'Mere inventions of the imagination: a survey of recent literature on Adam Smith'. *Economics and Philosophy* 13, p. 281-312.
- Brownlow, Graham. 2010. 'Structure and change: Douglass North's economics'. *Journal of Economic Methodology* 17:3, p. 301-316.
- Bruni, Luigino and Stefano Zamagni. 2013. 'Introduction', in: Bruni and Zamagni (eds.), *Handbook on the Economics of Reciprocity and Social Enterprise*, p. 1-8. Cheltenham: Edward Elgar.
- Bryant, Joseph M. 1997. 'Review: The Rise of Christianity'. *Sociology of Religion* 58:2, p. 191-195.
- Buchanan, James A. 1975. *The Limits of Liberty: Between Anarchy and Leviathan*. Chicago, IL: University of Chicago Press.
- 1989. *Explorations in Constitutional Economics*. College Station, TX: Texas A&M Press.
- Buchanan, James A. and Gordon Tullock. 1962. *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor, MI: University of Michigan Press.



- Bücher, Karl. 1893. *Die Entstehung der Volkswirtschaft*. Tübingen: H. Laupp.
- Buller, David. 2005. *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature*. Cambridge, MA: Bradford Books.
- Camerer, Colin and Richard H. Thaler. 1995. 'Anomalies: ultimatum, dictators and manners'. *Journal of Economic Perspectives* 9:2, p. 209-219.
- Campbell, Donald T. 1968. 'Variation and Selective Retention in Socio-cultural Evolution', in: Barringer, Blanksten and Mack (eds.), *Social change in developing areas: A reinterpretation of evolutionary theory*, p. 19-49. Cambridge, MA: Schenkman.
- Carter, Gerald G. and Gerald S. Wilkinson. 2013. 'Food sharing in vampire bats: reciprocal help predicts donations more than relatedness or harassment'. *Proc. Biol. Sci.* 280:1753, 20122573.
- Cartledge, Paul. 1983. 'Trade and Politics Revisited: Ancient Greece', in: Hopkins, Garnsey and Whittaker (eds.), *Trade in the Ancient Economy*, p. 1-15. Berkeley, CA: University of California Press.
- Cavalli-Sforza, L.L. and M.W. Feldman. 1981. *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton, NJ: Princeton University Press.
- Chamlee-Wright, Emily. 2015. 'Markets as an Extension of Culture', in: Grube and Storr (eds.), *Culture and Economic Action*, p. 88-122. Cheltenham: Edward Elgar.
- Cliteur, Paul B. 1990. 'Why Hayek is a conservative'. *ARSP* 76:4, p. 467-478.
- Coase, Ronald H. 1984. 'The New Institutional Economics'. *Journal of Institutional and Theoretical Economics* 140:1, p. 229-231.
- Colander, David. 2000. 'The death of neoclassical economics'. *Journal of the History of Economic Thought* 22:2, p. 127-143.
- Colander, David, Richard P.F. Holt and J. Barkley Rosser. 2004. 'The changing face of mainstream economics'. *Review of Political Economy* 16:4, p. 485-499.
- Crawford, Vincent P., Uri Gneezy and Yuval Rottenstreich. 2008. 'The power of focal points is limited: even minute payoff asymmetry may yield large coordination failures'. *American Economic Review* 2008 98:4, p. 1443-1458.
- Cronk, Lee. 1988. 'Spontaneous order analysis and anthropology'. *Cultural Dynamics* 1, p. 282-308.
- Dale, Gareth. 2010. *Karl Polanyi: The Limits of the Market*. Cambridge: Polity Press.
- 2011. 'Lineages of embeddedness: on the antecedents and successors of a Polanyian concept'. *American Journal of Economics and Sociology* 70:2, p. 306-339.
- David, Paul A. 1985. 'Clio and the economics of QWERTY'. *The American Economic Review* 75:2, p. 332-337.

- Davis, Ann. 2008. 'Endogenous institutions and the politics of property: comparing and contrasting Douglass North and Karl Polanyi in the case of finance'. *Journal of Economic Issues* 42:4, p. 1101-1122.
- Davis, Lance E. 1966. 'The new economic history II: professor Fogel and the new economic history'. *Economic History Review* 19:3, p. 657-663.
- Davis, Lance E. and Douglass C. North. 1970. 'Institutional change and American economic growth: a first step towards a theory of institutional innovation'. *The Journal of Economic History* 30:1, p. 131-149.
- Dawkins, Richard. 1982. *The Extended Phenotype*. Oxford: Oxford University Press.
- 1986. *The Blind Watchmaker*. Harlow: Longman.
- De Marchi, Neil (ed.). 1992. *Post-Popperian Methodology of Economics : Recovering Practice*. New York, NY: Springer Science.
- Demsetz, Harold. 1983. 'The structure of ownership and the theory of the firm'. *The Journal of Law & Economics* 26:2, p. 375-390.
- Dennett, Daniel C. 1995. *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York, NY: Simon & Schuster.
- De Soto, Hernando. 1989. *The Other Path: The Invisible Revolution in the Third World*. New York, NY: Harper & Row.
- Didry, Claude and Caroline Vincensini. 2011. 'Beyond the market-institutions dichotomy: the institutionalism of Douglass C. North in response to Karl Polanyi's challenge'. CNRS working paper: halshs-00601544.
- Diebolt, Claude. 2012. 'Where are we now in Cliometrics?'. *Historical Social Research (Cliometrics)* 37:4, p. 309-326.
- Dopfer, Kurt (ed.). 2005. *The Evolutionary Foundations of Economics*. Cambridge: Cambridge University Press.
- Dupré, Georges and Pierre-Philippe Rey. 1978. 'Reflections on the Relevance of a Theory of the History of Exchange', in: David Seddon (ed.), *Relations of Production*, p. 171-208. London: Cass.
- Durham, William H. 1991. *Coevolution: Genes, Culture and Human Diversity*. Stanford, CA: Stanford University Press.
- Earle, Timothy. 1997. *How Chiefs Come to Power*. Stanford, CA: Stanford University Press.
- Eggertsson, Thrainn. 1990. *Economic Behavior and Institutions*. Cambridge: Cambridge University Press.

- Elardo, Justin A. 2013. 'Formalism/Substantivism', in: R. McGee and R. Warms (eds.), *Theory in Social and Cultural Anthropology: An Encyclopedia*, p. 269-271. Thousand Oaks, CA: Sage.
- Elardo, Justin A. and Al Campbell. 2006. 'Choice and the Substantivist/Formalist Debate: a Formal Presentation of Three Substantivist Criticisms', in: Donald C. Wood (ed.), *Choice in Economic Contexts*, p. 267-284. Bingley: Emerald Publishing.
- Elster, Jon. 1982. 'Marxism, functionalism, and game theory'. *Theory and Society* 11, p. 453-482.
- (ed.) 1986. *Rational Choice*. New York, NY: New York University Press.
- 2000. 'Rational choice history: a case of excessive ambition'. *American Political Science Review* 94, p. 685-695.
- Endres, Anthony M. and Christine Rachel Woods. 2006. 'Modern theories of entrepreneurial behavior: a comparison and appraisal'. *Small Business Economics* 26:2, p. 189-202.
- Engelmann, Dirk and Urs Fischbacher. 2009. 'Indirect reciprocity and strategic reputation building in an experimental helping game'. *Games and Economic Behavior* 67:2, p. 399-407.
- Ensminger, Jean (ed.) 2002. *Theory in Economic Anthropology*. Lanham, MD: AltaMira.
- Eshel, Ilan and L.L. Cavalli-Sforza. 1982. 'Assortment of encounters and evolution of cooperativeness'. *Proc. Natl. Acad. Sci. USA* 79:4, p. 1331-1335.
- Fehr, Ernst and Simon Gächter. 2000. 'Fairness and retaliation: the economics of reciprocity'. *Journal of Economic Perspectives* 14:1, p. 159-181.
- 2002. 'Altruistic punishment in humans'. *Nature* 415:6868, p. 137-140.
- Field, Alexander J. 1979. 'On the explanation of rules using rational choice models'. *Journal of Economic Issues* 13:1, p. 49-72.
- 1981. 'The problem with neoclassical institutional economics: A critique with special reference to the North/Thomas model of pre-1500 Europe'. *Explorations in Economic History* 18:2, p. 174-198.
- 2001. *Altruistically Inclined? The Behavioral Sciences, Evolutionary Theory, and the Origins of Reciprocity*. Ann Arbor, MI: University of Michigan Press.
- Fine, Ben and Dimitris Milonakis. 2003. 'From principle of pricing to pricing of principle: rationality and irrationality in the economic history of Douglass North'. *Comparative Studies in Society and History* 45:3, p. 546-570.
- 2009. *From Economics Imperialism to Freakonomics: The Shifting Boundaries Between Economics and Other Social Sciences*. Abingdon: Routledge.
- Finley, Moses. 1985. *Ancient History: Evidence and Models*. New York, NY: Viking.

- Fischbacher, Urs, Simon Gächter and Ernst Fehr. 2001. 'Are people conditionally cooperative? Evidence from a public goods experiment'. *Economics Letters* 71:3, p. 397-404.
- Fisher, Ronald Aylmer. 1930. *The Genetic Theory of Natural Selection*. Oxford: Clarendon.
- Fletcher, Jeffrey A. and Michael Doebeli. 2009. 'A simple and general explanation for the evolution of altruism'. *Proc. Royal Soc. B* 276, p. 13-19.
- Fogel, Robert W. 1964. *Railroads and American Economic Growth: Essays in Econometric History*. Baltimore, MD: Johns Hopkins University Press.
- 1966. 'The new economic history: its findings and methods'. *Economic History Review* 19:3, p. 642-656.
- Fudenberg, Drew and Eric Maskin. 1986. 'The folk theorem in repeated games with discounting or with incomplete information'. *Econometrica* 54:3, p. 533-554.
- Fudenberg, Drew, David Levine and Eric Maskin. 1994. 'The folk theorem with imperfect public information'. *Econometrica* 62:5, p. 997-1039.
- Fukuyama, Francis. 2000. 'Social Capital', in: Harrison and Huntington (eds.), *Culture Matters: How Values Shape Human Progress*. New York, NY: Basic Books.
- Freedman, David A.; with David Collier, Jasjeet S. Sekhon and Philip B. Stark (eds.). 2010. *Statistical Models and Causal Inference: A Dialogue With the Social Sciences*. Cambridge: Cambridge University Press.
- Galiani, Sebastian and Itai Sened (eds.). 2014. *Institutions, Property Rights and Economic Growth: The Legacy of Douglass North*. Cambridge: Cambridge University Press.
- Geertz, Clifford. 1988. *Works and Lives: The Anthropologist as Author*. Stanford, CA: Stanford University Press.
- Gemici, Kurtuluş. 2008. 'Karl Polanyi and the antinomies of embeddedness'. *Socio-Economic Review* 6:1, p. 5-33.
- Geneakoplos, John, David Pearce and Ennio Stacchetti. 1989. 'Psychological games and sequential rationality'. *Games and Economic Behavior* 1, p. 60-79.
- Giddens, Anthony. 1979. *Central Problems in Social Theory: Action, Structure, and Contradiction in Social Analysis*. London: Macmillan.
- Gintis, Herbert. 2009. *The Bounds of Reason: Game Theory and the Unification of the Social Sciences*. Princeton, NJ: Princeton University Press.
- Gintis, Herbert et al. 2005. *Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life*. Cambridge, MA: MIT Press.
- Godelier, Maurice. 1978. 'The Object and Method of Economic Anthropology', in: David Seddon (ed.), *Relations of Production*, p. 49-126. London: Cass.

- 1986. *The Mental and the Material*. Transl. Martin Thom. London: Verso.
- Gould, Stephen Jay. 2002. *The Structure of Evolutionary Theory*. Cambridge, MA: Harvard University Press.
- Graeber, David. 2001. *Towards an Anthropological Theory of Value: The False Coin of Our Own Dreams*. New York, NY: Palgrave Macmillan.
- Granovetter, Mark. 1985. 'Economic action and social structure: the problem of embeddedness'. *American Journal of Sociology* 91:3, p. 481-510.
- 2005. 'The impact of social structure on economic outcomes'. *Journal of Economic Perspectives* 19, p. 33-50.
- Greif, Avner. 1993. 'Contract enforceability and economic institutions in early trade: the Maghribi traders' coalition'. *The American Economic Review* 83:3, p. 525-548.
- 2012. 'The Maghribi traders: a reappraisal?'. *The Economic History Review* 65:2, p. 445-469.
- Groenewegen, John. 2001. 'North's Theory of Institutional Change', in: Philip Anthony O'Hara (ed.), *Encyclopedia of Political Economy: L-Z*, p. 810-812. London: Routledge.
- Groenewegen, John, Frans Kerstholt and Ad Nagelkerke. 1995. 'On integrating old and new institutionalism: Douglass North building bridges'. *Journal of Economic Issues* 19:2, p. 467-475.
- Gudeman, Stephen. 2008. *Economy's Tension: The Dialectics of Market and Economy*. New York, NY/Oxford: Berghahn.
- Gul, Faruk and Wolfgang Pesendorfer. 2008. 'The Case for Mindless Economics', in: Caplin and Schotter (eds.), *The Foundations of Positive and Normative Economics: A Handbook*, p. 3-42. Oxford: Oxford University Press.
- Hahn, Frank H. 1987. 'Information, dynamics and equilibrium'. *Scottish Journal of Political Economy* 34:4, p. 321-334.
- Haldane, J.B.S. 1932. *The Causes of Evolution*. Princeton, NJ: Princeton University Press.
- Hamilton, William D. 1964. 'The genetic evolution of social behaviour'. *Journal of Theoretical Biology* 7, p. 1-16.
- Hammerstein, Peter. 2003. 'Why is Reciprocity So Rare in Social Animals? A Protestant Appeal', in: Peter Hammerstein (ed.), *Genetic and Cultural Evolution of Cooperation*, p. 83-93.
- Hampe, M. and S.R. Morgan. 1988. 'Two consequences of Richard Dawkins' view of genes and organisms'. *Studies in History and Philosophy of Science A* 19:1, p. 119-138.
- Hann, Chris and Keith Hart (eds.). 2011. *Market and Society: The Great Transformation Today*. Cambridge: Cambridge University Press.

- Hardin, Russell. 1982. *Collective Action*. Baltimore, MD: Johns Hopkins University Press.
- 2003. "The Free Rider Problem", in: Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Spring 2013: <http://plato.stanford.edu/archives/spr2013/entries/free-rider/>.
- Hargreaves-Heap, Shaun P. and Yanis Varoufakis. 1995. *Game Theory: A Critical Introduction*, First Edition. London: Routledge.
- Hart, Keith. 2007. 'A short history of economic anthropology'.  
<http://thememorybank.co.uk/2007/11/09/a-short-history-of-economic-anthropology/>
- 2008. 'Karl Polanyi's legacy'. *Development and Change* 39:6, p. 1135-1143.
- Harriss, John. 2002. 'Institutions, Politics and Culture: A Case for 'Old' Institutionalism in the Study of Historical Change'. LSE DESTIN Working Papers Series 02-34.
- Hausman, Daniel. 1992. *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.
- Hayami, Yujiro and Masahiko Aoki (eds.). 1998. *The Institutional Foundations of East Asian Economic Development*. IEA Conference Vol. no. 127. New York, NY: St Martin's Press.
- Hayek, F.A. 1944. *The Road to Serfdom*. Chicago, IL: University of Chicago Press.
- 1952. *The Sensory Order*. Chicago, IL: University of Chicago Press.
- 1960. *The Constitution of Liberty*. Chicago, IL: University of Chicago Press.
- 1967. *Studies in Philosophy, Politics and Economics*. London: Routledge & Kegan Paul.
- Heijdra, Ben J. and Anton D. Lowenberg. 1986. 'Duhem-Quine, Lakatos, and research programmes in economics'. *Journal of Interdisciplinary Economics* 1:3, p. 175-187.
- Heijdra, Ben J., Anton Lowenberg and Robert J. Mallick. 1988. 'Marxism, methodological individualism, and the New Institutional Economics'. *Journal of Institutional and Theoretical Economics* 144:2, p. 296-317.
- Henrich, Joseph. 2004. 'Cultural group selection, coevolutionary processes and large-scale cooperation'. *Journal of Economic Behavior and Organization* 53, p. 3-35, 127-143.
- Henrich, Joseph et al. 2003. 'In search of homo economicus: experiments in fifteen small-scale societies'. *American Economic Review* 91:2, p. 73-79.
- (eds.). 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford: Oxford University Press.
- Hirshleifer, Jack. 1987. *Economic Behavior in Adversity*. Chicago, IL: University of Chicago Press.

- Hobbes, Thomas. 1651. *Leviathan: Or the Matter, Forme and Power of a Commonwealth Ecclesiasticall and Civil*. London: Andrew Crooke.
- Hodgson, Geoffrey M. 1991. 'Economic evolution: intervention contra Pangloss'. *Journal of Economic Issues* 25:2, p. 519-533.
- 1994a. 'The evolution of socioeconomic order in the move to a market economy'. *Review of International Political Economy* 1:3, p. 387-404.
  - 1994b. 'Hayek, Evolution, and Spontaneous Order', in: Philip Mirowski (ed.), *Natural Images in Economic Thought: Markets Red in Tooth and Claw*, p. 408-447. Cambridge: Cambridge University Press.
  - 1998. 'The approach of institutional economics'. *Journal of Economic Literature* 36, p. 166-192.
  - 2001. *How Economics Forgot History: The Problem of Historical Specificity in the Social Sciences*. London: Routledge.
  - 2003. 'The mystery of the routine: the Darwinian destiny of an evolutionary theory of economic change'. *Révue Economique* 54:2, p. 355-384.
  - 2004. 'Darwinism, causality, and the social sciences.' *Journal of Economic Methodology* 11:2, p. 175-194.
  - 2006. 'What are institutions?'. *Journal of Economic Issues* 40:1, p. 1-25.
  - 2007b. 'Evolutionary and institutional economics as the new mainstream?'. *Evol. Inst. Econ. Rev.* 4:1, p. 7-25.
  - 2013. *From Pleasure Machines to Moral Communities: An Evolutionary Economics Without Homo Economicus*. Chicago, IL: University of Chicago Press.
- Hodgson, Geoffrey M. (ed.) 2007a. *The Evolution of Economic Institutions*. Cheltenham: Edward Elgar.
- Hodgson, Geoffrey M. and David Gindis. 2007. 'An interview with Oliver Williamson'. *Journal of Institutional Economics* 3:3, p. 373-386.
- Hodgson, Geoffrey M. and Thorbjørn Knudsen. 2008. 'In search of general evolutionary principles: why Darwinism is too important to be left to the biologists.' *Journal of Bioeconomics* 10:1, p. 51-69.
- 2010. *Darwin's Conjecture: The Search for General Principles of Social and Economic Evolution*. Chicago, IL: University of Chicago Press.
- Hoffman, Elizabeth, Kevin A. McCabe and Vernon L. Smith. 1998. 'Behavioral foundations of reciprocity: experimental economics and evolutionary psychology'. *Economic Inquiry* 36:3, p. 335-352.

- Hogarth, Robin M. and Melvin W. Reder (eds.). 1986. *The Behavioral Foundations of Economic Theory*. *Journal of Business* supplement.
- 1987. *Rational Choice: The Contrast Between Economics and Psychology*. Chicago, IL: University of Chicago Press.
- Hough, Jerry F. and Robin Grier. 2005. *The Long Process of Development: Building Markets and States in Pre-Industrial England, Spain and their Colonies*. Cambridge, Cambridge University Press.
- Hull, David L. 1980. 'Individuality and selection'. *Annual Review of Ecology and Systematics* 11, p. 311-332.
- 2001. *Science and Selection: Essays on Biological Evolution and the Philosophy of Science*. Cambridge: Cambridge University Press.
- Humphreys, Sally C. 1969. 'History, economics, and anthropology: the work of Karl Polanyi'. *History and Theory* 8, p. 165-212.
- 1978. *Anthropology and the Greeks*. London: Routledge & Kegan Paul.
- Ingrao, Bruna and Giorgio Israel. 1990. *The Invisible Hand: Economic Equilibrium in the History of Science*. Cambridge, MA: MIT Press.
- Johnson, Allen W. and Timothy Earle. 2000. *The Evolution of Human Societies*, Second Edition. Stanford, CA: Stanford University Press.
- Jongman, Willem 2012. 'Formalism-Substantivism Debate', in: R.S. Bagnall (ed.), *The Encyclopedia of Ancient History*. Oxford: Blackwell.
- Jongman, Willem. 2014. 'Why modern economic theory applies, even to the distant Roman past'. *Theoretical Roman Archaeology Conference Proceedings 2013*, p. 27-36.
- Kanel, Don. 1985. 'Institutional economics: perspectives on economy and society'. *Journal of Economic Issues* 19:3, p. 815-828.
- Kelly, Raymond Case. 1985. *The Nuer Conquest: The Structure and Development of an Expansionist System*. Ann Arbor, MI: University of Michigan Press.
- Kennedy, Gavin. 2009. 'Adam Smith and the invisible hand: from metaphor to myth'. *Econ Journal Watch* 6:2, p. 239-263.
- Klein, Daniel B. and Ryan Daza. 2013. 'Douglass C. North'. *Econ Journal Watch* 10:3, p. 525-532.
- Knight, Frank. 1999 [1941]. 'Anthropology and Economics', in: *Selected Essays by Frank Knight*, vol. 2, p. 107-125. Chicago, IL: University of Chicago Press.
- Knight, Jack and Douglass C. North. 1997. 'Explaining economic change: the interplay between cognition and institutions'. *Legal Theory* 3:3, p. 211-226.



- Krader, Lawrence (ed.). 1974. *The Ethnological Notebooks of Karl Marx*. Assen: Van Gorcum.
- Krippner, Greta et al. 2004. 'Polanyi symposium: a conversation on embeddedness'. *Socio-Economic Review* 2, p. 109-135.
- Krul, Matthijs and Gareth Dale. 2016 (forthcoming). 'Markets in Ancient Greece: The Challenge of the New Institutionalism', in: Gareth Dale, *Reconstructing Karl Polanyi: Excavation and Critique*. London: Pluto Press.
- Lacher, Hannes. 1999. 'The politics of the market: re-reading Karl Polanyi'. *Global Society* 13:3, p. 313-326.
- Laland, Kevin N. and Gillian R. Brown. 2002. *Sense and Nonsense: Evolutionary Perspectives on Human Behavior*. Oxford: Oxford University Press.
- Lamb, Robert. 1974. 'Adam Smith's system: sympathy, not self-interest'. *Journal of the History of Ideas* 35:4, p. 671-682.
- Langlois, Richard H. 1986. 'The New Institutional Economics: An Introductory Essay', in: R.H. Langlois (ed.), *Economics as a Process: Essays in the New Institutional Economics*, p. 1-25. Cambridge: Cambridge University Press.
- Lewontin, Richard C. and Richard Levins. 2007. *Biology Under the Influence: Dialectical Essays on Ecology, Agriculture, and Health*. New York, NY: Monthly Review Press.
- Lilly, Gregory. 1994. 'Bounded rationality: a Simon-like explication'. *Journal of Economic Dynamics and Control* 18:1, p. 205-230.
- Lloyd, Elisabeth. 2012. 'Units and Levels of Selection', in: Edward Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Winter 2012 edition. URL: <http://plato.stanford.edu/archives/win2012/entries/selection-units/>
- Loasby, Brian. 1999. *Knowledge, Institutions, and Evolution in Economics*. London: Routledge.
- Love, John R. 1991. *Antiquity and Capitalism: Max Weber and the Sociological Foundations of Roman Civilization*. London: Routledge.
- Lumsden, Charles J. and Edward O. Wilson. 1981. *Genes, Mind and Culture: The Coevolutionary Process*. Cambridge, MA: Harvard University Press.
- Mäki, Uskali. 1993. 'Economics with Institutions', in: Mäki, Gustafsson and Knudsen (eds.), p. 3-44.
- Mäki, Uskali (ed.). 2002. *Fact and Fiction in Economics: Models, Realism, and Social Construction*. Cambridge: Cambridge University Press.
- Mäki, Uskali, Bo Gustafsson and Christian Knudsen (eds.). 1993. *Rationality, Institutions, and Economic Methodology*. New York, NY: Routledge.
- Manning, J.G. and Ian Morris (eds.). 2005. *The Ancient Economy: Evidence and Models*. Stanford, CA: Stanford University Press.

- Mantzavinos, Chris, Douglass C. North and Syed Shariq. 2004. 'Learning, institutions, and economic performance'. *Perspectives on Politics* 2:1, p. 75-84.
- Marciano, Alain. 2009. "Why Hayek is a Darwinian (after all)? Hayek and Darwin on social evolution". *Journal of Economic Behavior and Organization* 71:1, p. 52-61.
- Maynard Smith, John. 1964. 'Group selection and kin selection'. *Nature* 201, p. 1145-1147.
- McElreath, Richard, Robert Boyd and Peter J. Richerson. 2003. 'Shared norms and the evolution of ethnic markers.' *Current Anthropology* 44:1, p. 122-129.
- Meikle, Scott. 1996. 'Modernism, economics, and the ancient economy'. *Proceedings of the Cambridge Philological Society* 41, p. 174-191.
- Ménard, Claude and Mary M. Shirley. 2014. 'The Contribution of Douglass C. North to New Institutional Economics', in: Galiani and Sened (eds.), p. 11-29.
- Menger, Carl. 1985 [1883]. Tr. Francis J. Nock. *Investigations into the Method of the Social Sciences*. New York, NY: New York University Press.
- Metcalf, J. Stanley. 1998. *Evolutionary Economics and Creative Destruction*. London: Routledge.
- Meyer, Eduard. 1913. *Geschichte des Alterthums*. Stuttgart/Berlin: J.G. Cotta.
- Milinski, Manfred, Dirk Semmann and Hans-Jürgen Krambeck. 2002. 'Reputation helps solve the 'tragedy of the commons''. *Nature* 415, p. 424-426.
- Miller, Roger L., Daniel K. Benjamin and Douglass C. North. 2013. *The Economics of Public Issues*, 18<sup>th</sup> edition. New York, NY: Prentice Hall.
- Millward, Robert. 1983. 'Review: Structure and Change in Economic History'. *The Economic Journal* 93:372, p. 963-965.
- Milonakis, Dimitris and Ben Fine. 2007. 'Douglass North's remaking of economic history: a critical appraisal'. *Review of Radical Political Economics* 39:1, p. 27-57.
- Mirowski, Philip S. 2002. *Machine Dreams: Economics Becomes a Cyborg Science*. Cambridge: Cambridge University Press.
- Mirowski, Philip and D. Wade Hands (eds.). 2006. *Agreement on Demand: Consumer Theory in the Twentieth Century*. Durham, NC: Duke University Press.
- Moore, Jim. 1984. 'The evolution of reciprocal sharing'. *Ethnology and Sociobiology* 5, p. 5-14.
- Morgan, Jamie (ed.). 2016. *What is Neoclassical Economics? Debating the Origins, Meaning, and Significance*. Abingdon/New York, NY: Routledge.
- Morgan, Mary S. 2012. *The World in the Model: How Economists Work and Think*. Cambridge: Cambridge University Press.

- Morrison, Margaret. 1999. 'Models as Autonomous Agents', in: Morgan and Morrison (eds.), *Models as Mediators: Perspectives on Natural and Social Science*, p. 38-65. Cambridge: Cambridge University Press.
- Mueller, Dennis C. 2003. *Public Choice III*. Cambridge: Cambridge University Press.
- Nadeau, Robert. 2013. 'Neoclassical economic theory', in: *Encyclopedia of Earth*.  
<http://www.eoearth.org/view/article/154813>
- Nelson, Richard R. and Sidney G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, MA: Harvard University Press.
- 2002. 'Evolutionary theorizing in economics'. *Journal of Economic Perspectives* 16:2, p. 23-46.
- North, Douglass C. 1974. 'Beyond the new economic history'. *The Journal of Economic History* 34:1, p. 1-7.
- 1976. 'The place of economic history in the discipline of economics'. *Economic Inquiry* 14:4, p. 461-465.
  - 1977. 'Markets and other allocation systems in history: the challenge of Karl Polanyi'. *Journal of European Economic History* 6, p. 703-716.
  - 1978a. 'Structure and performance: the task of economic history'. *Journal of Economic Literature* 16:3, p. 963-978.
  - 1978b. 'Review: The Livelihood of Man'. *Business History Review* 52:3, p. 398-399.
  - 1979a. 'Review: New Directions in Political Economy: An Approach from Anthropology'. *The Journal of Economic History* 39:4, p. 1088.
  - 1979b. 'A framework for analysing the state in economic history'. *Explorations in Economic History* 16:3, p. 249-259.
  - 1981. *Structure and Change in Economic History*. New York, NY: W.W. Norton.
  - 1983. 'Private property and the American way'. *National Review* July 8, 1983, p. 805-812.
  - 1984. 'Government and the cost of exchange in history'. *The Journal of Economic History* 44:2, p. 255-264.
  - 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.
  - 1992. 'Transaction Costs, Institutions, and Economic Performance'. ICEG Occasional Papers, no. 30.
  - 1993a. 'Toward a Theory of Institutional Change', in: Barnett, Hinich and Schofield (eds.), *Political Economy: Institutions, Competition, and Representation*, p. 60-70. Cambridge: Cambridge University Press.

- 1993b. 'What do we mean by rationality?'. *Public Choice* 77:1, p. 159-162.
  - 1993c. 'Economic performance through time'. Lecture to the memory of Alfred Nobel, 1993. [http://www.nobelprize.org/nobel\\_prizes/economic-sciences/laureates/1993/north-lecture.html](http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/1993/north-lecture.html)
  - 1995. 'The Paradox of the West', in: R.W. Davis (ed.), *The Origins of Modern Freedom in the West*, p. 7-34. Stanford, CA: Stanford University Press.
  - 1997. 'Cliometrics – forty years later'. *American Economic Review* 87:2, p. 412-414.
  - 1999. 'Hayek's Contribution to Understanding the Process of Economic Change', in: Viktor Vanberg (ed.), *Freiheit, Wettbewerb und Wirtschaftsordnung: Hommage zum 100. Geburtstag von Friedrich A. von Hayek*, p. 79-96. Freiburg im Breisgau: Haufe.
  - 2005. *Understanding the Process of Economic Change*. Princeton, NJ: Princeton University Press.
- North, Douglass C. and Robert Paul Thomas. 1970. 'An economic theory of the growth of the Western world'. *Economic History Review* 23:1, p. 1-17.
- 1971. 'The rise and fall of the manorial system: a theoretical model'. *The Journal of Economic History* 31:4, p. 777-803.
  - 1976. *The Rise of the Western World*. Cambridge: Cambridge University Press.
- North, Douglass C. and Barry R. Weingast. 1989. 'Constitutions and commitment: the evolution of institutions governing public choice in seventeenth-century England'. *The Journal of Economic History* 49:4, p. 803-832.
- North, Douglass C., John Joseph Wallis, and Barry R. Weingast. 2009. *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*. Cambridge: Cambridge University Press.
- Northcott, Robert and Anna Alexandrova. 2015. 'Prisoner's Dilemma Doesn't Explain Much', in: Martin Peterson (ed.), *The Prisoner's Dilemma*, p. 64-84. Cambridge: Cambridge University Press.
- Ogilvie, Sheilagh. 2001. 'The economic world of the bohemian serf: economic concepts, preferences, and constraints on the estate of Friedland, 1583–1692'. *The Economic History Review* 54:3, p. 430-453.
- Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, MA: Harvard University Press.
- 1993. 'Democracy, dictatorship and development'. *American Political Science Review* 87:3, p. 567-575.
- Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge, MA: Harvard University Press.

- 2005. *Understanding Institutional Diversity*. Princeton, NJ: Princeton University Press.
- Panchanathan, Karthik and Robert Boyd. 2004. 'Indirect reciprocity can stabilize cooperation without the second-order free rider problem'. *Nature* 432, p. 499-502.
- Pannenberg, Wolfhart. 2004. *Anthropology in Theological Perspective*. London: Bloomsbury.
- Patterson, Thomas C. 2009. *Karl Marx, Anthropologist*. Oxford: Berg.
- Pelikan, Pavel. 1987. 'The Formation of Incentive Mechanisms in Different Economic Systems', in: S. Hedlund (ed.), *Incentives and Economic Systems*, p. 27-56. New York, NY: New York University Press.
- Pettit, Philip. 2002. 'Rational Choice, Functional Selection, and 'Empty Black Boxes'', in: Mäki 2002, p. 231-256.
- Peukert, Helge. 2001. 'Bridging old and new institutional economics: Gustav Schmoller and Douglass C. North, seen with old institutionalists' eyes'. *European Journal of Law and Economics* 11, p. 91-130.
- Polanyi, Karl. 1947. 'On belief in economic determinism'. *Sociological Review* 37:1, p. 96-112.
- 1968. Ed. George Dalton. *Primitive, Archaic and Modern Economies: Essays of Karl Polanyi*. Garden City, NY: Anchor Books.
- 1977. Ed. Harry W. Pearson. *The Livelihood of Man*. New York, NY: Academic Press.
- 2001 [1944]. *The Great Transformation: The Economic and Political Origins of Our Time*. Boston, MA: Beacon Press.
- Polanyi, Karl, Conrad M. Arensberg and Harry W. Pearson (eds.). 1957. *Trade and Market in the Early Empires: Economies in History and Theory*. Glencoe, IL: Free Press.
- Polanyi-Levitt, Kari (ed.). 1990. *The Life and Work of Karl Polanyi: A Celebration*. Montréal, QC: Black Rose Books.
- Powell, Walter W. and Paul J. DiMaggio (eds.). 1991. *The New Institutionalism in Organizational Analysis*. Chicago, IL: University of Chicago Press.
- Powell, Walter W. and Jeannette A. Colyvas. 2008. 'Microfoundations of Institutional Theory', in: R. Greenwood et al. (eds.), *Handbook of Organizational Institutionalism*, p. 276-298. Thousand Oaks, CA: Sage Publications.
- Rapoport, Anatol. 1991. 'Ideological commitments in evolutionary theories'. *Journal of Social Issues* 47:3, p. 83-99.
- Reiss, Julian. 2013. *Philosophy of Economics: A Contemporary Introduction*. Abingdon: Routledge.

- Renger, Johannes. 2005. 'K. Polanyi and the Economy of Ancient Mesopotamia', in: Clancier, Joannès, Rouillard and Tenu (eds.), *Autour de Polanyi: Vocabulaires, Théories et modalités des Échanges*, p. 45-65. Paris: De Boccard.
- Richerson, Peter J. and Robert Boyd. 2005. *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago, IL: University of Chicago Press.
- Richter, Rudolf. 1996. 'Bridging old and new institutional economics: Gustav Schmoller, the leader of the younger German historical school, seen with neoinstitutionalists' eyes'. *Journal of Institutional and Theoretical Economics* 152:4, p. 567-592.
- 2015. *Essays on New Institutional Economics*. Berlin: Springer.
- Rizvi, S. Abu Turab. 1994. 'The microfoundations project in general equilibrium theory'. *Cambridge Journal of Economics* 18, p. 357-377.
- Robbins, Lionel. 1932. *Essay on the Nature and Significance of Economic Science*. London: Macmillan.
- Roberts, William Clare. 2010. 'Pessimism and anti-state politics: from human evil to institutional particularism'. Conference paper, Canadian Political Science Association, June 2010.
- Rodrik, Dani. 2000. 'Institutions for High-Quality Growth: What They Are and How to Acquire Them'. NBER Working Paper no. 7540.
- Rodrik, Dani, Arvind Subramanian and Francesco Trebbi. 2002. 'Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development'. NBER Working Paper no. 9305.
- Rosenberg, Alexander. 1992. *Economics: Mathematical Politics or Science of Diminishing Returns?*. Chicago, IL: University of Chicago Press.
- 2000. *Darwinism in Philosophy, Social Science and Policy*. Cambridge: Cambridge University Press.
- 2008. *Philosophy of Social Science*. Third Edition. Boulder, CO: Westview Press.
- Rosser, Marina Vcherasnaya and J. Barkley Rosser. 1995. 'A comparison of comparative economic anthropologies'. *History of Economics Review* 23:4, p. 96-107.
- Rostovtzeff, Michael Ivanovitch. 1957. *The Social and Economic History of the Roman Empire*. Oxford: Clarendon Press.
- Rostow, Walt W. 1982. 'Review: Structure and Change in Economic History'. *Business History Review* 56:2, p. 299-301.
- Rutherford, Malcolm. 1994. *Institutions in Economics: The Old and New Institutionalism*. Cambridge: Cambridge University Press.
- Sawyer, K.R., Clive Beed and H. Sankey. 1997. 'Underdetermination in economics: the Quine-Duhem thesis'. *Economics and Philosophy* 13:1, p. 1-23.

- Scheidel, Walter and Sitta von Reden (eds.). 2002. *The Ancient Economy*. New York, NY: Routledge.
- Schelling, Thomas C. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
- Schotter, Andrew. 1981. *The Economic Theory of Social Institutions*. Cambridge: Cambridge University Press.
- Schubert, Christian and Georg von Wangenheim (eds.). 2006. *Evolution and Design of Institutions*. Abingdon: Routledge.
- Seaford, Richard. 2004. *Money and the Early Greek Mind*. Cambridge: Cambridge University Press.
- Selten, Reinhard. 1975. 'A reexamination of the perfectness concept for equilibrium points in extensive games'. *International Journal of Game Theory* 4, p. 25-55.
- Selwyn, Benjamin and Satoshi Miyamura. 2014. 'Class struggle or embedded markets? Marx, Polanyi, and the meanings and possibilities of social transformation'. *New Political Economy* 19:5, p. 639-661.
- Shryock, Andrew and Daniel Lord Smail (eds.). 2011. *Deep History: The Architecture of Past and Present*. Berkeley, CA: University of California Press.
- Silver, Morris. 1983. 'Karl Polanyi and markets in the ancient Near East: the challenge of the evidence'. *The Journal of Economic History* 43:4, p. 795-829.
- 1992. *Taking Ancient Mythology Economically*. Leiden: Brill.
- 1995. *Economic Structures of Antiquity*. Westport, CT: Greenwood Press.
- Simon, Herbert. 1959. 'Theories of decision-making in economics and behavioral science'. *The American Economic Review* 49:3, p. 253-283.
- 1972. 'Theories of Bounded Rationality', in: McGuire and Radner (eds.), *Decision and Organization*, p. 161-176. Amsterdam: North-Holland.
- 1976. 'From Substantive to Procedural Rationality', in: Spiro J. Latsis (ed.), *Method and Appraisal in Economics*, p. 129-148. Cambridge: Cambridge University Press.
- 1986. 'Rationality in Psychology and Economics', in: Hogarth and Reder (eds), p. 209-224.
- Smith, E.A. and R.L. Bliege Bird. 2000. 'Turtle hunting and tombstone opening: public generosity as costly signaling'. *Evolution and Human Behavior* 21, p. 245-261.
- Smith, Vernon L. 1998. 'The two faces of Adam Smith'. *Southern Economic Journal* 65:1, p. 2-19.
- Snodgrass, Anthony. 1991. 'The Place and Role of the Annales School in the Approach to the Roman Rural Economy', in: J. Bintliff (ed.), *The Annales School and Archaeology*, p. 77-103. Leicester: Leicester University Press.

- Sober, Elliott and David Sloan Wilson. 1994. 'A critical review of philosophical work on the units of selection problem'. *Philosophy of Science* 61:4, p. 534-555.
- 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.
- Soltis, Joseph, Robert Boyd and Peter J. Richerson. 1995. 'Can group-functional behaviors evolve by group selection? An empirical test'. *Current Anthropology* 36:3, p. 473-494.
- Spence, Michael. 1973. 'Job market signaling'. *The Quarterly Journal of Economics* 87:3, p. 355-374.
- Spread, Patrick. 2016. *The Evolution of Economies: Money-Bargaining, Institutional Change and Industrial Revolution*. Abingdon/New York, NY: Routledge.
- Stanfield, James Ronald. 1999. 'Disembedded economy', in: Philip Anthony O'Hara (ed.), *Encyclopedia of Political Economy A-K*. London/New York, NY: Routledge.
- Stanford, Kyle. 2013. 'Underdetermination of Scientific Theory', in: Edward Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Winter 2013 edition. URL: <http://plato.stanford.edu/archives/win2013/entries/scientific-underdetermination/>
- Stark, Rodney. 1996. *The Rise of Christianity*. San Francisco, CA: Harper.
- Steele, David Ramsay. 1987. 'Hayek's theory of cultural group selection'. *Journal of Libertarian Studies* 8:2, p. 171-195.
- Sterelny, Kim et al. (eds.). 2013. *Cooperation and its Evolution*. Cambridge, MA: MIT Press.
- Stinchcombe, Arthur L. 1980. 'Is the prisoner's dilemma all of sociology?'. *Inquiry: An Interdisciplinary Journal of Philosophy* 23:2, p. 187-192.
- Stoelhorst, Jan-Willem. 2007. 'The naturalist view of universal Darwinism: an application to the evolutionary theory of the firm', in: Geoffrey M. Hodgson (ed.), *The Evolution of Economic Institutions: A Critical Reader*. Cheltenham: Edward Elgar.
- Sugden, Robert. 1986. 'New developments in the theory of choice under uncertainty'. *Bulletin of Economic Research* 38:1, p. 1-24.
- 1989. 'Spontaneous order'. *Journal of Economic Perspectives* 3:4, p. 85-97.
- Swedberg, Richard. 1998. *Max Weber and the Idea of Economic Sociology*. Princeton, NJ: Princeton University Press.
- Temin, Peter. 1991. 'A Market Economy in the Early Roman Empire'. MIT Economics Working Papers Series 01-08.
- 2012. *The Roman Market Economy*. Princeton, NJ: Princeton University Press.



- Tooby, John and Leda Cosmides. 1992. 'The Psychological Foundations of Culture', in: Barkow, Cosmides and Tooby (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, p. 19-136. Oxford: Oxford University Press.
- Trivers, Robert L. 1971. 'The evolution of reciprocal altruism'. *The Quarterly Review of Biology* 46:1, p. 35-57.
- Tuck, Richard. 2008. *Free Riding*. Cambridge, MA: Harvard University Press.
- Tullock, Gordon. 1980. 'Efficient Rent Seeking', in: Buchanan, Tollison, and Tullock (eds.), *Towards a Theory of the Rent-Seeking Society*, p. 97-112. College Station, TX: Texas A&M University Press.
- Ullman-Margalit, Edna. 1978. 'Invisible-hand explanations'. *Synthese* 39:2, p. 263-291.
- Vanberg, Viktor. 1986. 'Spontaneous market order and social rules'. *Economics and Philosophy* 2:1, p. 75-100.
- 1994. *Rules and Choice in Economics*. New York, NY: Routledge.
- Van de Mieroop, Marc. 1999. *Cuneiform Texts and the Writing of History*. London: Routledge.
- Van den Bergh, Jeroen C.J.M. and John M. Gowdy. 2003. 'The microfoundations of macroeconomics: an evolutionary perspective'. *Cambridge Journal of Economics* 27:1, p. 65-84.
- Van der Spek, R.J., Bas van Leeuwen and Jan Luiten van Zanden (eds.). 2015. *A History of Market Performance: From Ancient Babylonia to the Modern World*. Abingdon: Routledge.
- Varoufakis, Yanis. 1991. *Rational Conflict*. Oxford: Blackwell.
- Veblen, Thorstein. 1898. 'Why is economics not an evolutionary science?'. *Quarterly Journal of Economics* 12:4, p. 373-397.
- Von Neumann, John and Oskar Morgenstern. 1944. *The Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.
- Von Reden, Sitta. 2010. *Money in Classical Antiquity*. Cambridge: Cambridge University Press.
- Vromen, Jack J. 1995. *Economic Evolution: An Enquiry Into the Foundations of New Institutional Economics*. London: Routledge.
- Wagner-Hasel, Beate. 2011. *Die Arbeit des Gelehrten: Der Nationalökonom Karl Bücher (1847-1930)*. Frankfurt-am-Main: Campus.
- Wallis, John Joseph and Douglass C. North. 1986. 'Measuring the Transaction Sector in the American Economy, 1870-1970', in: Engerman and Gallman (eds.), *Long-Term Factors in American Economic Growth*, p. 95-162. Chicago, IL: University of Chicago Press.

- Wang, Ning. 2003. 'Measuring Transaction Costs: An Incomplete Survey'. Ronald Coase Institute Working Paper no 2.
- Weber, Max. 2013 [1909]. *The Agrarian Sociology of Ancient Civilizations*. Transl. R.I. Frank. London: Verso.
- Weintraub, E. Roy. 1979. *Microfoundations: The Compatibility of Microeconomics and Macroeconomics*. Cambridge: Cambridge University Press.
- 1993. 'Neoclassical Economics', in: David Henderson (ed.), *Fortune Encyclopedia of Economics*. New York, NY: Warner Books.
- Wilk, Richard R. and Lisa Cliggett. 2007. *Economies and Cultures: Foundations of Economic Anthropology* (Second Edition). Boulder, CO: Westview Press.
- Williams, George C. 1966. *Adaptation and Natural Selection*. Princeton, NJ: Princeton University Press.
- Williamson, Oliver E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York, NY: Free Press.
- Wilson, Edward O. 1975. *Sociobiology: The New Synthesis*. Cambridge, MA: Harvard University Press.
- Winter, Sidney G. 1986. 'Comments on Arrow and on Lucas', in: Hogarth and Reder (eds.), p. 427-434.
- Witt, Ulrich (ed.) 1993. *Evolution in Markets and Institutions*. Würzburg: Physica.
- 2003. *The Evolving Economy: Essays on the Evolutionary Approach to Economics*. Cheltenham: Edward Elgar.
- Wrong, Dennis H. 1961. 'The oversocialized conception of man in modern sociology'. *American Sociological Review* 26:2, p. 183-193.
- Zahavi, Amotz. 1975. 'Mate selection – a selection for a handicap'. *Journal of Theoretical Biology* 53, p. 205-214.