Agent-Based Institutional Modelling: Novel Techniques for Deriving Structure from Behaviour

Christopher Konstantin Frantz

a thesis submitted for the degree of

Doctor of Philosophy

at the University of Otago, Dunedin,

New Zealand.

September 2015

Abstract

Institutions, as manifestations of social behaviour, are crucial elements of any society's structure and are widely studied in the area of New Institutional Economics as a determinant of a society's development and prosperity. Although taking promising steps by considering a wider array of behavioural influence mechanisms, to date analytical approaches primarily examine institutions from the perspective of stable equilibria, which are assumed as indicative for the existence of institutions. In this work we shift the focus from equilibria to an interactionist perspective, and apply Agent-Based Modelling (ABM) for the purpose of institutional analysis.

ABM offers the ability to represent institutions in great detail by using modelling metaphors that are in close alignment with real observable social structures and interaction patterns, as opposed to concentrating on rational strategy choices of structurally uniform individuals. With ABM we can leverage a sociological perspective and "grow" institutional structures from the bottom up, instead of just interpreting those from a bird's eye perspective.

This work not only applies ABM to institutional analysis, but also develops it further as a tool for the representation and systematic analysis of institutions. Essential contributions pointing in this direction include a refined generalised institution representation (nADICO) that captures the complexity of institutions. We further propose a continuous norm concept (Dynamic Deontics) that models the formation and change of normative understanding based on environmental stimuli. A final extension is the analysis of normative understanding on different sociological levels using Interval Type-2 Fuzzy Sets. These contributions enrich the computational social scientists' toolbox with modelling and analysis mechanisms that are a) sociologically grounded, b) facilitate a fine-grained

explicit representation of institutions, and c) permit the multi-level analysis of social systems.

To showcase ABM's applicability for institutional analysis, in the first part of this thesis we review a seminal case from the area of comparative economics, Greif's *Maghribī Traders Coalition*, a medieval trader collective that operated based on *informal means* without relying on contractual enforcement. Based on the detailed analysis of recent literature, we offer a refined interpretation of the coalition. We further introduce the Genoese society that Greif contrasted to the Maghribīs, since it primarily relied on *formal contracts* to govern cooperation.

With both scenarios as a basis, we review selected abstractions Greif puts forth for his original equilibrium-based model. Those include the assumed inherent secrecy among Genoese, in contrast to the openly sharing Maghribīs. We further relax the assumption of a closed trader coalition for the Maghribī side. A third aspect involves the neglected significantly contrasting role conceptualisations in both historical societies. To deepen our understanding of those aspects, we develop dedicated agent-based models to explore alternative explanations.

In the second part of the thesis, we move towards the greater goal of growing institutional structures from a behavioural, sociologically motivated, perspective. We introduce the aforementioned generalisable representations for institutions and continuous conception of normative understanding. We then apply those to explore the third aspect of the historical scenario and analyse how far the different societal role stratification could have influenced normative understanding. Using fuzzy concepts, we conclude with a multi-level analysis of normative alignment for both trader societies.

This work thus both *applies* Agent-Based Modelling, but also *enhances* it for the purpose of Institutional Modelling and Analysis.

Acknowledgements

Though written by a single author, most theses are hardly an accomplishment achieved in solitude. And as much as we are taught to identify the essence of our work and highlight its contributions, as little have we learned to thank everyone who has, in one way or another, contributed to its success.

I am most indebted to my supervisor Prof. Martin Purvis, whom I had the honour to work with well before this PhD began. He planted the seed for this research and help to shape and challenge my vision whenever I needed it. Particularly in the later stages of reviewing this work, his support did not know the boundaries of day and night, weekend and holiday! In conjunction with this I wish to extend my gratitude to Dr. Maryam Purvis who offered continuous support and developed valuable research collaborations that influenced the outcome of this thesis.

I would like to express my gratitude to my supervisor Associate Prof. Mariusz Nowostawski, who provided inspiration with respect to many technical aspects of this work. He never failed to encourage me to 'dig deeper' when exploring specific problems, which did not necessarily result in immediate excitement, but in maturing my approach to scientific inquiry.

I was further extremely fortunate to have Dr. Bastin Tony Roy Savarimuthu as my supervisor. His expertise was essential for the second part of this thesis. Beyond his domain knowledge his advice is what can be described as 'spot on'. I am particular grateful for his offer to read chapters of my thesis during the busy times of his daughter's birth.

A person that was central to wrapping up my work was my partner Noora whose consistent encouragement and reminders drove me to finish this long-lasting project. She was the continuous target of my research ramblings and challenging discussion partner, especially with respect to the sociological underpinnings of this work. Particularly in the later stages she relieved me from any imaginable responsibility in order to maintain my focus.

My parents Renate and Alexander gave me all the support that resulted in the 'delayed reward' of this thesis. They set the foundation and gave me every opportunity to experience and explore whatever I had interest in. These sentences do little justice to express the sheer endless support they have given and continue to give me. Not to be forgotten at this stage are the fond memories of involved topical discussions with my brothers Benedikt and Jonas.

The success of such work is not only defined by concentration, but also distractions. For this I hold all of my friends responsible. Those include my supervisors, whose initial supervisory relationships have become friendships. In addition to those I would like to acknowledge my friends Carl, Sharmila, Rexy, Tomek, Alexandra, Tony (Moore), alongside many others.

I would also like to express my thank you to my past and present fellow students whom I had the pleasure to share the place of study with, enjoy extensive topical discussions and receive advice from. This includes Toktam, Rasika, Surangika, Angrosh, Femi, Mehdi, Yasir, and Marzieh.

Furthermore, I was lucky to have Associate Profs. Holger Regenbrecht and Peter Whigham as conveners for my PhD meetings, both of which did not shy away from giving concrete advice on my research and continuously inquired about the progress. Their inspiring talks in postgraduate venues are deeply entrenched in my memory and should be considered a 'must go' for any postgraduate student. Further memorable input in the form of given talks or feedback to presentations came from Prof. Stephen Cranefield, Prof. Stephen MacDonell, and Prof. Michael Winikoff. Beyond this I would like to extend my 'thank you' to the entire Department of Information Science. The accessibility of staff members and diverse nature of expertise made it a great place of study.

A further person that should be explicitly acknowledged is Mark George, who is a brilliant source of ideas for technical solutions and whose insights make him a great discussion partner. Similarly, Nathan D. Lewis, who was a research assistant in the department until 2012, was a great sparring partner for bouncing back ideas.

Inasmuch as the Department of Information Science is characterised by its academic staff, the quality of the administrative staff should be emphasised. Stephen Hall-Jones, Gail Mercer and Heather Cooper were accessible with any possible inquiry and assisted greatly in tackling the bureaucratic hurdles in a PhD student's life. They own a great share in the timely submission of this thesis.

Though last, but by no means least, an essential 'thank you' goes to the University of Otago, whose support in form of a Postgraduate Scholarship made this research possible in the first place.

Contents

List of Tables

List of Figures

List of Algorithms

List of Acronyms

ABM	Agent-Based Modelling
ACE	Agent-Based Computational Economics
AML	Agent Modeling Language
AOSE	Agent-Oriented Software Engineering
BDI	Belief-Desire-Intention (Model)
CPN	Coloured Petri-Nets
CPR	Common Pool Resources
CW	Computing with Words
FOU	Footprint of Uncertainty
FS	Fuzzy Set
GoI	Grammar of Institutions
IA	Institutional Analysis
IAD	Institutional Analysis and Development (Framework)
IQR	Interquartile Range
IT2FLS	Interval Type-2 Fuzzy Logic System
IT2FS	Interval Type-2 Fuzzy Set
KS	Kolmogorov-Smirnov (Test)
LMF	Lower Membership Function
MAS	Multi-Agent System
MF	Membership Function
MWW	Mann-Whitney-Wilcoxon (Test)
NIE	New Institutional Economics
NorMAS	Normative Multi-Agent Systems
OCC	Ortony-Clore-Collins (Model)
ODD	Overview, Design concepts, and Details (Protocol)
OIE	Original/Old Institutional Economics
OWA	Ordered Weighted Averaging Operator
PD	Prisoners' Dilemma (Game)
RL	Reinforcement Learning
T1FS	Type-1 Fuzzy Set
T2FS	Type-2 Fuzzy Set
UMF	Upper Membership Function
UML	Unified Modeling Language

1
Introduction

1.1 Motivation

Institutions are omnipresent in our social environment. When individuals interact, behavioural patterns emerge, and, once established, impose constraints on individuals' future action choices. Manifestations of social behaviour or *institutions*, a term we will discuss in more detail in Chapter 2, serve various purposes, including the regulation of social behaviour (Greif, 2006), reduction of uncertainty (North, 1990), and efficiency increase (Williamson, 1998). Acting upon the society, their pervasive nature, in the form of conventions (Ullmann-Margalit, 1977), social norms (Ullmann-Margalit, 1977), and rules (North, 1990), is thought to define a society's developmental trajectory (see e.g. Greif (2006)). With a renewed interest in comparative economic studies, such as by Greif (2006) as well as Acemoglu and Robinson (2012), the quest for the secret of success is no longer exclusively ascribed to environmental conditions, such as available natural resources or geographical position.[1] Instead, the institutional environment is considered decisive, since it enhances and constrains opportunities for socio-economic development, a focus that is associated with the research field of *New Institutional Economics* (Williamson, 1975a), and has found revived attention by eco-

[1]This view is famously held by Jared Diamond (1997).

nomic scholars such as North (2005), Ostrom (2005b), North et al. (2009), Greif (2006) and Acemoglu and Robinson (2012).

Beyond the association with this particular field, the implications of institutions reach far into social scientific domains beyond economics, such as political science (Helmke and Levitsky, 2004) and sociology in general (Durkheim, 1933). Offering necessary analytical and explorative capabilities, the field of *Institutional Analysis* (North, 1990; Ostrom et al., 1994; Williamson, 1998) contributes to a systematic evaluation of institutions by drawing on tools such as *Game Theory* (von Neumann and Morgenstern, 1947; Tadelis, 2013), and, more recently, applying the notion of *Agent-Based Modelling* (Gilbert, 2008; Gilbert and Troitzsch, 2005), which is a central vehicle of exploration in the field of *Computational Sociology* (Macy and Willer, 2002; Epstein and Axtell, 1996; Epstein, 2007; Squazzoni, 2012).

The thesis presented in this volume puts specific focus on this latter aspect, promoting the use of agent-based modelling and simulation for the purpose of institutional analysis. Given its origin in the area of economics, traditional institutional analysis concentrates on the *economic* objective or impact of institutions (Williamson, 1998), the evaluation of which lends itself for formal institutions (see Subsection 2.2.3). However, recent examples of social unrest, such as seen in the 'Arab spring', the fall of the Ukrainian president (and subsequent civil conflict), protests around claims of rigged elections in Turkey and Malaysia – with the latter being economically thriving countries – highlight the importance of the *social* perspective to explain the emergence and manifestation of such collective action. Although such complex social phenomena cannot be reduced to the individuals' actions, they can neither be analysed without considering the individual perspective. What ties both aspects together are *interactions*, including direct interaction among individuals as well as influence processes emanating from social phenomena that act upon individuals. Consequently, an improved understanding of the emergence of such phenomena requires representations that are aligned with observable real-world structures and interactions. In addition, they also need to reflect the level of detail necessary to build models of complex social systems grounded in reality. Agent-Based Modelling offers such promises by providing representations for agency and structure, allowing the representation of complex social dynamics that are sometimes dismissed as secondary in purely economically motivated models (see e.g. Williamson (2000), Subsection 2.2.3). This is particularly important for cases such as social influence or subconscious action, in which the rationality assumption – which is increasingly dismissed by prominent figures such as Selten (2001) – can be challenged. Furthermore, characteristically the effects of informal decentralised interactions cannot be specified or predicted ex-ante, but only ‘emerge at runtime’. We thus see particular suitability of agent-based models to reflect the informal end of the institutional spectrum, such as social norms, that generally exist alternatively or are complementary (Helmke and Levitsky, 2004; North, 1991) to more straightforwardly tractable formal institutions such as legal norms, or rules (North, 1991, 2005).

A classical case from the area of comparative economics that showcases the effects of contrasting formal and informal institutional environments is the work of Avner Greif (1989, 1993, 1994, 2006). Greif contrasts a historical North African trader collective (the Maghribīs) (Goitein, 2000a; Goldberg, 2012c) that established cooperation in long-distance trade based on informal means, with contemporary Southern European traders (the Genoese) (Epstein, 1996) that drew upon formal mechanisms to govern cooperation. The nature of the respectively chosen institutional instruments, such as informal partnerships or formal contracts, Greif posits, influenced their ensuing economic success.

Using this scenario as a guiding example, we augment the historical perspective with recent literature that provokes revisions of Greif’s original conception, particularly with respect to structural aspects of the contrasted societies. In contrast to Greif’s original game-theoretical model, we apply agent-based modelling to explore selected aspects he subsumed as fundamental assumptions – an aspect that we see in part rooted in the chosen analytical method itself. Our intent is not so much to challenge the overall outcome of his work, but to show how agent-based modelling, with its structural representations, sociality, and the emergence metaphor, can supplement and enhance analysis based on traditional game-theoretical analysis.

Beyond the extension of the historical case analysis which is a focus of the first part of this thesis, in the second part (Chapter 5 onwards) we extend the application of agent-based modelling towards contributions to the field of normative modelling that are rooted in the scenario of interest but bear generalisability beyond the domain of this work. Particular focus lies on the endogenous perspective of institutional formation assuming a behaviourist perspective. We thus consider experience and interaction between individuals as fundamental formative processes of institutions and focus on their emergent nature, as opposed to encoding and pre-imposing institutions as part of individuals’ behaviour. Beyond a novel mechanism of institution formation, including a refined representation in the form of an extended version of Crawford and Ostrom’s Institutional Grammar (Crawford and Ostrom, 1995), we introduce general, yet practical techniques that enhance the institutional modeller’s toolbox with accessible means of multi-level analysis.

1.2 Research Questions

Two essential questions this thesis addresses are

- *How can agent-based modelling contribute to institutional analysis?*
- *How far can we leverage its strength, the representation of dynamic social processes, to support the modelling and analysis from a behavioural perspective?*

Addressing these questions we develop the following overarching thesis:

Agent-based Modelling provides appropriate modelling metaphors to

- *minimize the need for overly constraining high-level assumptions concerning complex social scenarios, and to*
- *understand endogenous processes of institutional emergence.*

We explore the prototypical comparative scenario of the Maghribī Traders Coalition (Greif, 2006) and review selected assumptions of previous analyses in the light of contemporary literature. As indicated in the previous section, we put particular focus on *informal institutional mechanisms*, such as social norms, given their crucial role for the functioning of societies, be it in the absence of, or complementary to, formal institutions such as laws (North, 2005).

Supporting the first sub-thesis, for the given scenario we illustrate how agent-based modelling can allow the relaxation of strong model assumptions and thus facilitate refined analyses of selected sub-problems, a theme that will accompany us throughout this work.

In the second part of this work, we enhance the capabilities of agent-based modelling by building on its strengths: the modelling of social structures and interactions. We introduce an approach that allows for an accessible representation of norm emergence in artificial societies and also fosters their analysis on different social levels.

1.3 Research Approach

1.3.1 Approach

This work is largely motivated from an information science perspective, i.e. the choice and application of technological and methodological means to address specific problems that cross disciplinary boundaries. In our case, this application is driven by existing analytical shortcomings in the context of a specific problem in the area of institutional analysis.

Given the problem's origin in the interdisciplinary field of economic history, we lay out the historical case based on available literature accounts, with focus on characteristics relevant to Greif's original conception and the ensuing analysis. We provide a comprehensive overview of social, economic, and institutional characteristics of the respective societies (Chapter 3). Doing so, we put particular emphasis on more recent literature findings that portray a far more detailed picture than was available at the time of Greif's initial analyses in the 1980s (Greif, 1989). Beyond picturing the historical perspective, we identify aspects of Greif's theses that have come under scrutiny by subsequent findings which we label as *Literature Debates* (see Section 3.2). We clarify the differing interpretations and take our own position, which we use as a basis to review three assumptions of Greif's earlier analyses from Chapter 4 onwards. Since the literature analysis bears refined structural and behavioural implications for the revised assumptions, we resort to agent-based modelling, since it offers conceptually aligned representations that foster a detailed reconstruction of historical happenstance. To review the initial two assumptions, we build dedicated agent-based models and provide a detailed evaluation (Chapter 4), which concludes the first part of the thesis.

In the second part, we emphasise a generative modelling approach and construct a generalisable comprehensive representation of structural aspects based on a refined variant of Crawford and Ostrom's Institutional Grammar (Crawford and Ostrom, 1995, 2005) (Chapter 5). It reflects evolving normative understanding based on experiential and social learning, which is informed by conceptual input from the area of social psychology and institutional economics. This conceptual contribution is applied to test the remaining last assumption of Greif's original work. For the remainder of this thesis, we develop and refine the normative model with focus on analytical features that are applicable beyond the scope of this work and its domain. In the spirit of interdisciplinary use, throughout this work we maintain particular focus on the accessibility of emerging normative understanding to the experimenter. We attempt to offer a pragmatic interpretation of institutions, while striving for their comprehensive structural representation.

Methodically throughout the thesis we thus perform a focal transition from an *application* of agent-based modelling in the first part towards *extending* agent-based modelling with features relevant for endogenous institutional analysis. This is reflected in the developed agent-based models and their evaluation. The models developed in the first part put stronger focus on the comparatively detailed representation of the scenario and the evaluation of the impact of refined assumptions. For the generalisable conceptual and methodological contributions introduced in the second part, the models put stronger focus on the exploration of principles, with limited emphasis on detailed representation. This aspect is challenged by the limited information base providing factual grounding for validation. Admittedly, validation is a challenging task for all models of a scenario situated in the twelfth century (including Greif's own models (Greif, 2006)), which, particularly for the Maghribīan side, has not been characterised by systematic data collection, but instead has been reconstructed from an extensive but fragmented corpus of letters (see Subsection 3.1.1). Like Greif (2006), the validation of model results thus mostly relies on exploration of parameter ranges and anecdotal evidence as opposed to hard and quantified information. Doing so, our approach shares characteristics of Edmonds' and Moss's KIDS[2] approach (Edmonds and Moss, 2005) that considers the admission of weak evidence for the purpose of modelling a target phenomenon. Consequently, the models we introduce opt for understanding, as opposed to the provision of an accurate representation[3] of the historical situation.

1.3.2 Method

The research method we are using throughout this work is the transdisciplinary approach of simulation that has its roots in the field of mathematics and computer science (Zeigler, 1976; Zeigler et al., 2000), and, inspired by the field of physics, found increasing application in social scientific disciplines ('sociophysics') and economics ('econophysics') in the form of *Social Simulation* (Troitzsch, 2009; Gilbert, 2004; Gilbert and Troitzsch, 2005). Its capability to explore scenarios of social entities by systematically 'growing artificial societies' (Epstein and Axtell, 1996; Conte et al., 1998) makes it appealing for problems that cannot be explored using traditional deductive approaches. Equation-based models that rely on ex-ante formalisation (and provability) are often associated with the introduction of strong assumptions. Inductive approaches, on the other hand, rely on instance level observations of the modelled phenomenon.[4] It is for that reason Axelrod suitably described simulation as the "third way of doing science" (Axelrod, 1997b, 2006). The intent of marrying elements of deduction and induction is to facilitate the emergence of macro-level phenomena that are not explicitly encoded in the underlying micro-level behaviour, and thus potentially bear novel unpredicted insights. Furthermore, in contrast to statistical models that rely on the availability of data to *develop models that correlate with collected measurements*, simulation lends itself well to the *reconstruction of processes and explicit social structure* (Gilbert

[2]Edmonds and Moss (2005) position the modelling principle 'Keep it Descriptive Simple Stupid' (KIDS) in contrast to the prevalent 'Keep it Simple Stupid' (KISS), the latter of which advocates modelling simplicity whenever possible. This comes at the risk of omitting relevant detail information, which is considered under the KIDS approach.

[3]See Axelrod (1997a) for the trade-off between those two simulation purposes. Greif (2012) similarly emphasises the explicit focus on the *understanding of institutions*; *not their accurate reconstruction*.

[4]It is challenging to consult medieval traders on the motivations for their behaviour.

and Troitzsch, 2005) which is useful when operating on a scarce data basis. In addition, recreating the underlying structural characteristics offers the potential to extend the inquiry beyond the original phenomenon.

1.3.3 Audience

This work is primarily targeted at researchers operating at the intersection of computer science and social sciences, including the area of economic history, with particular interest in institutional analysis and normative modelling. Given our intent to provide an accessible approach to institutional and normative modelling, the target audience of the conceptual and methodological contributions are social scientists. However, that does not imply a complete overlap with the targeted readership. As such, the thesis has been written for accessibility to readers with an applied computer science background and, ideally, a basic understanding of artificial intelligence and normative multi-agent systems, as well as for social scientists with interest in computational modelling. Given our intent to develop contributions in response to the modelling demands of the problem scenario, we emphasise a narrative, as opposed to formal approach, while organising chapters based on their relationship to research fields. The structural outline of the thesis in Section 1.5 allocates the essential chapters with their respective research fields. However, we will first provide a detailed overview of this work's contributions to the literature.

1.4 Contributions

We can loosely organise contributions based on the classification offered by Brinberg and McGrath (1985), who structured contributions by their *substantive* nature (i.e. contributing in the form of 'content' for the domain of research), *conceptual* nature (i.e. means by which we interpret substantive contributions, such as by introduction of novel terms and concepts characterising a novel phenomenon), and finally, *methodological* contributions. The latter describe additions or modification of the means by which we can study aspects of interest (Brinberg and McGrath, 1985).

Correlated with the nature of contributions, the organisation of the core chapters displays increasing abstraction as shown in Figure 1.1, starting with concrete incremental additions to the scenario of interest, and subsequently progressing towards the development of generalisable contributions to the field of institutional analysis. We describe the individual contributions in the following.

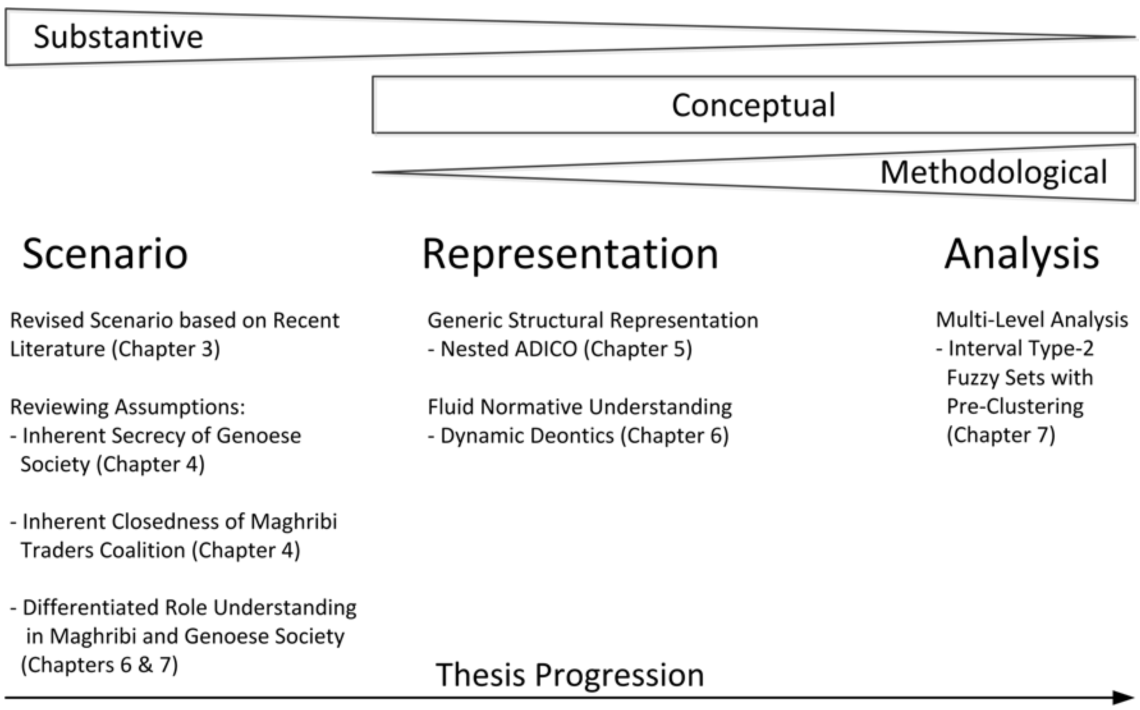


Figure 1.1: Contribution Overview

1.4.1 Substantive Contributions

Substantive contributions of this work concentrate on the field of new institutional economics, economic history and agent-based modelling by reviewing selected assumptions of seminal work in the light of recent evidence.

We provide an extended literature overview on the historical societies that considers recent historic findings, along with a critical discussion of inconsistencies (dubbed 'Literature Debates') with Greif's conception of the historical reality. We incrementally develop an interpretation of the historical case that takes the current literature base into account, thus providing an updated perspective on the Maghribī Traders Coalition, and consequently offering an interpretative basis for future explorative research (Chapter 3).

Based on the refined interpretation, and extending on Greif's work, in Chapter 4 we test three assumptions, namely

1. Genoese traders did not share information about their trade activities with fellow traders.

2. The Maghribī traders represented a 'closed' group that did not admit access to outsiders/newcomers.

3. The Maghribī trader society was modelled under the assumption of role specialisa-

tion, although in reality (and in contrast to Genoese traders) they operated under an integrated role conception.

We use agent-based models to explore how far those assumptions can be relaxed or refined, thus leveraging a complementary use of agent-based modelling for the purpose of institutional analysis for specific analyses. Moreover, the exploration of this relevant scenario introduces the Maghribī Traders Coalition to the research field of agent-based modelling and simulation.

1.4.2 Conceptual Contributions

Conceptual contributions are driven by modelling limitations of the omitted aspect of Greif's model, the differentiation of Maghribī and Genoese role understandings. We introduce a conceptually rich refinement of Crawford and Ostrom's institutional grammar (Crawford and Ostrom, 1995), called Nested ADICO (nADICO) (see Chapter 5), that facilitates the comprehensive representation of institutions and their characteristics (e.g. institution types, institutional regress – see Subsection 2.2.1). The refined variant offers a clearly specified syntax that reflects nesting characteristics along with an interpretation that is closely aligned with a sociological viewpoint and enables the representation of institutions from an endogenous perspective. We further provide an alternative approach for the representation of normative understanding, called Dynamic Deontics (see Chapter 6), that extends beyond the widely adopted static notion of norms and emphasises their dynamic and continuous nature instead.

Both contributions can find general application in the context of agent-based modelling and are independent of the trader scenario. Given our focus on informal institutions, the contributions bear relationships to the field of Normative Multi-Agent Systems (see Subsection 2.3.2) in particular.

1.4.3 Methodological Contributions

The methodological contributions largely complement the conceptual contributions, with systematic modelling approaches that operationalise and combine the proposed representations (introduced as part of the conceptual contributions) to evolve and analyse normative understanding. This includes the operationalisation of Dynamic Deontics (see Subsection 6.2.5) from a consequentialist perspective, utilising reinforcement learning (Watkins, 1989) (see Subsection 2.3.3) to develop normative understanding.

As a final step, we combine Dynamic Deontics with Interval Type-2 Fuzzy Sets (Zadeh, 1975a) (see Chapter 7) as a means to analyse dynamic normative understanding on micro-

(individual), meso- (group), and macro- (society) levels, thereby advocating the application of fuzzy concepts, particularly higher-order fuzzy sets, for the analysis of social concepts.

Similar to the conceptual contributions, the methodological contributions aim towards strong generalisability potential (see Chapter 8) in the field of agent-based modelling and offer specific contributions to the field of Normative Multi-Agent Systems (see Subsection 2.3.2).

1.5 Structure of the Thesis

This thesis is structured into eight chapters, with Chapter 2 providing the necessary sociological and domain-specific foundations along with an overview of the applied modelling techniques and computational tools.

The main part of the thesis is structured into two parts, consisting of Chapters 3 and 4 as well as Chapters 5 to 7, respectively.

Chapter 3 lays out the historical and institutional background of the trader scenarios, and includes the exposition of central debates about historical happenstance. Chapter 4 concentrates on Greif's fundamental model of institutional analysis of the Maghribī Traders Coalition in contrast to the Genoese trader society. We isolate the previously mentioned unrealistically rigid assumptions, which we relax and explore using two agent-based simulation models.

In Chapter 5 we introduce the extension of Crawford and Ostrom's institutional grammar (Crawford and Ostrom, 1995), facilitating a more refined institution representation. Chapter 6 contributes an associated dynamic norm representation, which we apply to refine the third limiting assumption (or conceptual omission) of Greif's model.

Chapter 7 concentrates on the interpretation of evolved normative understanding on different social levels of analysis using Interval Type-2 Fuzzy Sets, thus bearing strong relation to the area of institutional analysis.

Finally, in Chapter 8 we summarise the contributions, discuss the limitations, and provide an outlook on possible directions for future research.

Table 1.1 provides an overview of the central chapters of this thesis and associates them with the nature of their contribution (see Section 1.4) and related research fields. The table further refers to publications related to the chapters' respective contributions, followed by a comprehensive list of refereed publications in order of related thesis chapters.

Table 1.1: Overview of Core Chapters Structured by Nature of Contributions and Research Fields

Chapter	Contribution Type			Research Field			Publications
	Substantive	Conceptual	Methodological	Economic History	Institutional Analysis	Normative Modelling	
Chapter 3	●			●			
Chapter 4	●	○		●	●		(Frantz et al., 2014a, 2015a)
Chapter 5		●	○		●	●	(Frantz et al., 2013)
Chapter 6	○	●	○	○	○	●	(Frantz et al., 2014b, 2015b)
Chapter 7	○	○	●	○	●	●	(Frantz et al., 2014d)

● Strong relation to contribution type/research field.
○ Moderate relation to contribution type/research field.

Publication	**Related Chapter**
Frantz, C.; Purvis, M. K.; Nowostawski, M.: *Agent-Based Modeling of Information Transmission in Early Historic Trading*, Social Science Computer Review, 2014, vol. 32, no. 3, 393-416.	Chapter 4
Frantz, C.; Purvis, M. K.; Nowostawski, M.; Savarimuthu, B. T. R.: *Analysing the Apprenticeship System in the Maghribi Traders Coalition*, Multi-Agent-Based Simulation XV, Springer, LNCS 9002, 2015, 180-196.	Chapter 4
Frantz, C.; Purvis, M. K.; Nowostawski, M.; Savarimuthu, B. T. R.: *nADICO: A Nested Grammar of Institutions*, PRIMA 2013: Principles and Practice of Multi-Agent Systems, Springer, LNCS 8291, 2013, 429-436.	Chapter 5
Frantz, C.; Purvis, M. K.; Nowostawski, M.; Savarimuthu, B. T. R.: *Modelling Institutions using Dynamic Deontics*, Coordination, Organizations, Institutions and Norms in Agent Systems IX, Springer, LNCS 8386, 2014, 211-233.	Chapter 6
Frantz, C.; Purvis, M. K.; Savarimuthu, B. T. R.; Nowostawski, M.: *Modelling the Impact of Role Specialisation on Cooperative Behaviour in Historic Trader Scenarios*, Coordination, Organizations, Institutions and Norms in Agent Systems X, Springer, LNAI 9372, 2015, *in press*.	Chapter 6
Frantz, C.; Purvis, M. K.; Savarimuthu, B. T. R.; Nowostawski, M.: *Analysing the Dynamics of Norm Evolution using Interval Type-2 Fuzzy Sets*, 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), IEEE, 2014, vol. 3, 230-237.	Chapter 7

2

Foundations

This chapter presents the underlying foundations for our work and develops a common terminological understanding for the remainder of this volume. Our intent is not to provide a complete introduction of all relevant aspects, but instead to provide an overview that primes the reader and is refined in the relevant chapters.

In Section 2.1 we establish the sociological perspective of our societal understanding and discuss characteristics that are central to the modelling of artificial societies. Following this, in Section 2.2 we introduce the core concept of institutions from a high-level perspective and discuss characteristics and processes involved in institutional formation, and develop an interpretation that is formative for the understanding of this work. Finally, in Section 2.3 we prepare an overview of technology applied in specific parts of the thesis in order to avoid extensive excursus in the chapters of relevance.

At the end of this chapter we will summarise and associate the introduced foundations with their related chapters.

2.1 Sociological Foundations

An aspect that requires discussion when handling seemingly different concepts such as institutions and agent-based modelling are the sociological underpinnings they share. We thus discuss essential sociological concerns that allow us to position our approach when analysing societies and their institutions. Those include the debate of *structure vs. agency* and the related discussion of *micro- and macro-sociological perspectives*. We harmonise both perspectives by drawing on the differentiation between *objective and subjective perspectives* on social analysis in order to set the scene for this work.

2.1.1 Structure and Agency

A question that has entertained generations of sociologists is the relationship between *Structure* and *Agency*, with the discipline's founding fathers plotting out respective perspectives. Durkheim's (1964) conception of *social facts* as a basis for the study of society and the conception of a collective reality emphasised the precedence of structural aspects (*structural functionalism*). Parson's work (1961), which was heavily influenced by Durkheim (and also Weber) is often associated with structural functionalism. Weber's *methodological individualism* (1978a), in contrast, posits that all action originates from individuals, which generate all causal influence on the formation of society. The emphasis on the individual as central driver can be traced back to the economic domain, particular to Schumpeter (1909) who was the first to use the term 'methodological individualism'. Though preceding the terminology, this understanding is compatible with the analytical perspective assumed by John Locke and Jeremy Bentham (see Udéhn (2001)). Central figures associated with methodological individualism – the precedence of the individual – are Friedrich von Hayek (1942) and Karl Popper (1945).[1] Recent sociology has shown a tendency to reconcile both concepts, such as Giddens' *Structuration Theory* (Giddens, 1984) that recognises their interdependency and mutual formative influence. Giddens suggests a cycle of constant reproduction in which structure is both the means and the result ("duality of structure" (Giddens, 1979)) of social practices in which individual partake. Similar to Giddens, Bourdieu (1984) puts stronger emphasis on the individual and does not require the real existence of structure. He uses the concepts of *habitus*, *field* and *capital* to explore the relationship of agency and structure, which we will explore in due course. The essential aspect of *structuralists*, such as Giddens and Bourdieu, is thus the integrated perspective on agency and structure, implying that elements of 'structure' are internal to the 'agent'. *Post-structuralists*, such as Mouzelis (1991)

[1] A good overview of the advocates of the different camps is provided by Ritzer and Goodman (2008).

and Archer (1979), in contrast, maintain the equal conceptual weight of agency and structure, but conceive both as distinct entities that have mutual influence but exist in separation.[2] Figure 2.1 is a schematic representation of the different sociological doctrines with respect to their conceptualisation of agency vs. structure and their precedence.

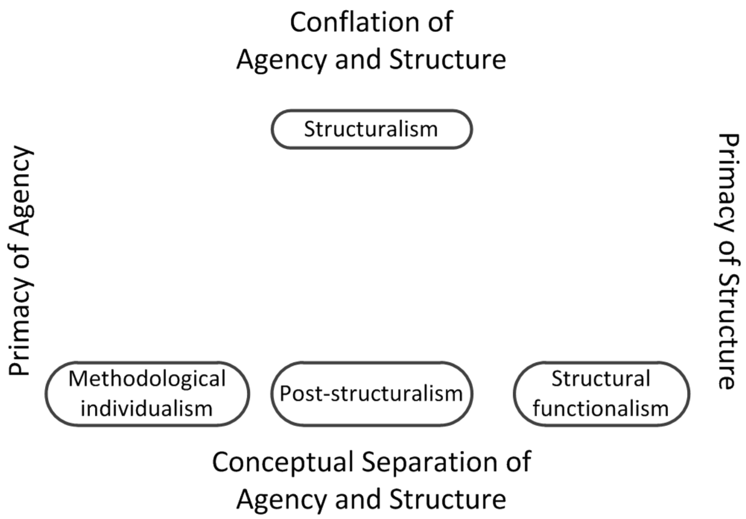


Figure 2.1: Sociological Doctrines on Agency and Structure

According to Bourdieu's conceptualisation of society – which we will use to inform our work –, individuals act in *fields* that describe roles and associated relationships. Operating in that field, and driven by notions of maintaining or building *capital* (Bourdieu describes four types of capital: economic, political, social, cultural) such as status or financial gain, the individual internalises expectations and practices associated with his role in the field, shaping its *habitus* (Bourdieu, 1985, 1990b). Linking structure and agency, habitus, as a result of the constant action within the field, is *embodied history*, thus affording a representation of external structure within the individual (Bourdieu, 1977), reifying its disposition (habitus) within the field. *Objectified history*, manifested in customs and laws, etc., can become "enacted" (Bourdieu, 1981) by an actor depending on its own disposition to do so, making habitus a 'strategy generating principle' (Parker, 2000). Its activation shapes the *field* that represents the space for agent interaction, and in consequence impinges on fellow social entities, shaping their habitus, in what Bourdieu describes as "dialectic of the internalization of externality and the externalization of internality" (Bourdieu, 1977). As part of the interplay of agency and structure, these internalisation and externalisation processes occur unconsciously (Bourdieu, 1984) since it would otherwise imply that all individuals' actions would be the result of deliberation, which is in contrast to Giddens' view of individuals as

[2] Parker (2000) offers an overview with particular focus on post-structuralist theories.

reflective actors.

Bourdieu explains this with his emphasis on "urgency of practice" (Bourdieu, 1981), suggesting that individuals have no choice but to act, an aspect that links the habitus with an agent's social environment. While bearing some degree of choice over their actions, individuals are ultimately constrained by their positions within the field, their history, and their belonging to social groups or classes, which drives them to act with a general focus on maintaining structural stability. The ability of individual actions to show effect is naturally constrained by their power, which again largely depends on structural aspects. In connection with the struggle for power in the field an individual acts in, Bourdieu's conceptualisation "frees us from the misplaced belief in illusory freedoms" (Bourdieu, 1990a).

This brief sketch highlights the integrated conception of agency and structure, with an embodiment of history (structure) within individuals, and subsequent externalisation, shaping others' habitus as well as objective history. It likewise explains that habitus is in a constant flux, and may include sudden changes or long-term adaptations (Navarro, 2006).

For this work, we adopt Bourdieu's interpretation as it bears compatibility with a consequentialist approach that is at the heart of Greif's work (1989; 2006), that is

- the path-dependent shaping of behaviour, and
- the motivation to increase (economic) capital.

At the same time his theory reflects the subjective (subconscious) understanding of institutions along with the urgency of practice which underlies structural constraints – aspects that are at the centre of our conceptual contributions (see Section 1.4). We do not seek to provide a complete representation of Bourdieu's conception, but consider the path-dependent nature of individuals' behaviour as fundamental to explain the emergence and manifestation of institutions, especially when embodied in the form of evolving social norms (as opposed to imposed rules, etc.).

This relationship between agency and structure leads us to a second, related issue of discourse in the field of sociology, the *different levels of analysis*.

2.1.2 Micro- vs. Macro-Level

Theories of multiple levels of social systems (and thus their analysis) are often conveniently conflated with the dichotomy of agency and structure. In contrast to the agency-structure discourse, which is conventionally entertained by European sociologists (see Archer (1988)), American sociology puts stronger emphasis on the different levels of analysis (see Ritzer and

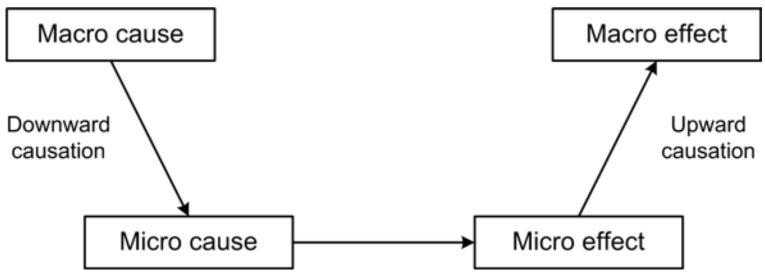


Figure adapted from Coleman (1990)

Figure 2.2: Coleman's Boat

Gindoff (1994)), an aspect rooted in works by Edel (1959) and Gurvitch (1964). Gurvitch identifies five horizontal layers of analysis (forms of sociality, groupings, social class, social structure, and global structures), along with cross-cutting vertical levels describing objective (e.g. ecological aspects) and subjective phenomena, such as collective values. Important work on the integration and interaction between both those levels has concentrated on the two outer horizontal layers, more commonly referred to as micro- and macro-layers. The micro-layer concentrates on elementary interactions of individuals, and the macro-layer emphasises macro-sociological phenomena, such as group structures, norms, bureaucracy, law and culture. An approach with strong traction in the field of social simulation (Squazzoni, 2008) is Coleman's boat (Coleman, 1990) (see Figure 2.2), which is similar to Hedström and Swedberg's (1998) conceptualisation. Coleman describes the dynamics of existing macro-level entities that influence interactions on the micro-level (*macro-micro situational mechanism/downward causation*), and, by means of a micro-micro action formation mechanism, manifest themselves in the macro-level (*micro-macro transformation mechanisms/upward causation*). Coleman's conception appears to support the pairwise conflation of agency/micro and structure/macro, as he insists that all action emanates from micro-level entities. For Coleman the macro-level represents an abstraction, "nevertheless an important one" (Coleman, 1990). In his view, a transformative process between macro-level phenomena does not exist without involving the micro-level, an interpretation that positions his approach with methodological individualism (see Subsection 2.1.1, Weber (1978a); Udéhn (2001))[3]

Coleman's view is in contrast to a *structuralist* perspective (e.g. Archer (1995)) that ascribes structural entities ontological statuses along with potential causal influence on micro-level entities.

[3]Udéhn (2001) refines that perspective, suggesting that Coleman's approach is associated with *structural individualism*, given its potential "structural effect" (Udéhn, 2001), beyond a "mere interaction effect" (Udéhn, 2001).

The dynamic processes underlying the discussed 'micro-macro link' bring us back to the discussion of *emergence*, an aspect that was explored implicitly as part of the introduction (see Subsection 1.3.2).

In the context of sociology, *Emergence*, or rather 'emergent properties'[4], can be defined as "stable macroscopic patterns arising from the local interaction of agents" (Epstein and Axtell, 1996). The essential characteristic is that the properties of social systems that result from social interaction cannot be reduced to the study of the individual social entities that participate in the interaction (Archer, 1995).

In the context of computational sociology, the concept of emergence comes in two qualities, or *orders of emergence* (see e.g. Squazzoni (2008)). *First-order* emergence describes a phenomenon or pattern that arises from generally decentralised interactions of individual entities and without control by any particular entities. Emergent elements are then the properties that have not been explicitly specified. Whether or not providing any added value for the participants, the 'emergents' are not perceived by the individual entities, who thus do not bear any mental representation of the phenomena. If at all, the observation of such phenomena is solely confined to the experimenter/s.

Second-order emergence or *Immergence* (Squazzoni, 2008), in contrast, differs in that it is perceived by the constituting entities, allowing its mental representation and thus consideration on the individual level. Whether individually or only as a collective, individuals can consciously influence the emergent phenomenon. Only second-order emergence closes the feedback loop between micro- and macro-layer, allowing individuals to react to macro properties. The representation of behaviour such as normative conduct – which is of relevance for this work – thus necessarily relies on some means to perceive and evaluate emerging social constructs.

This differentiation must not be confused with the philosophical differentiation between *weak* and *strong* emergence (Bedau, 1997), the latter referring to emerging entities that act autonomously from the underlying constituents. The limited scientific support of strong emergence, apart from filling explanatory gaps of contemporary science (Bedau, 1997), limits its practical relevance. It should nevertheless be highlighted that social systems, and social simulation in particular, can only represent weak emerging phenomena, given that social phenomena, such as norms, do not gain independence from the underlying constituents – their fate is tied to the individuals that form them.

Contrasting weak and strong emergence with first-order emergence and immergence (see

[4]For the epistemological implications of studying 'Properties' as opposed to 'Propositions' see Conte et al. (2014).

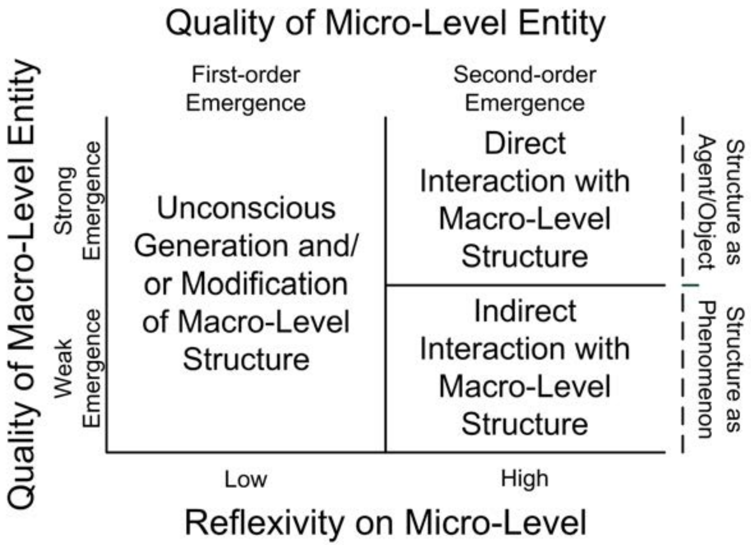


Figure 2.3: Differentiation of Emergence Concepts

Figure 2.3), the first pair describes the *quality of the emergent* (vertical axis in Figure 2.3), while first- and second-order emergence describe the *quality of the constituent*, in particular its reflexivity (horizontal axis in Figure 2.3). Theoretically, reflective agents ('high reflexivity' in Figure 2.3) could directly interact with strong emergents, assuming the latter would attain autonomous actorship qualities. Independent of an agent's cognitive abilities, weak emergence necessarily constrains individuals to indirect interaction with the emergent, be it by mere perception and/or manipulation or by adjustment of micro-level properties. In the absence of the ability to perceive the emerging phenomenon ('low reflexivity' in Figure 2.3), whether weak or strong, the absent feedback loop prevents the ascription of ontological meaning to the emergent by the constituents. Again, it is important to reinforce that this discussion is of theoretical nature, since no instance of strong emergence can be produced using simulation, given that such simulation would in itself be a reduction of the emergent (Bedau, 1997).

2.1.3 Micro/Macro and Subjectivity/Objectivity

Offering a more differentiated account on the micro-macro bifurcation than Coleman but less refined than Gurvitch, Ritzer (1979, 1981) claims that any micro-macro analysis necessarily requires the consideration of subjective and objective elements, as all micro- or macro-level phenomena fall into either category. Micro-level subjective phenomena are thus the mental aspects required to construct social reality, while objective manifestations are behavioural patterns, i.e. interactions, that are concepts with observable properties. In equivalence to

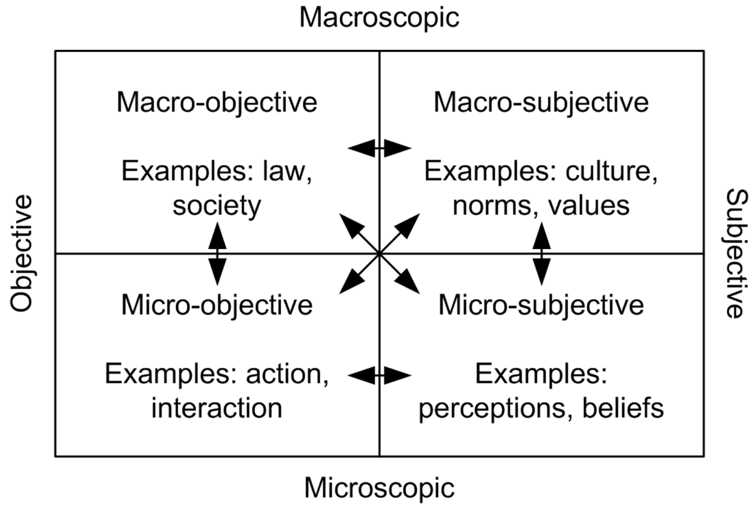


Figure adapted from Ritzer and Goodman (2003)

Figure 2.4: Ritzer's Model on Social Analysis

those operations at the micro-level, there are corresponding subjective and objective macro-level phenomena, including subjective aspects such as culture, norms and values, paired with objective macro-level entities, such as law, technology, or organisations (Ritzer, 1979, 1981). This quadrant approach, depicted in Figure 2.4, allows the modelling of interactions between all forms of phenomena. But beyond this, it can fundamentally address the lacking integration of the agency/structure and micro/macro differentiation by acknowledging the existence of structure both on micro- and macro-level (micro- and macro-objective), while allowing the ascription of agent-like formative characteristics to the subjective spectrum. However, structural artefacts can potentially be associated with any quadrant, an aspect we will touch upon in the following.

The integrative model provided by Ritzer is helpful, since it offers an integrated view with Bourdieu's embodied history (see Subsection 2.1.1), suggesting the internalisation of structural aspects in agents, while being reflected in a structural objectified history consisting of artefacts such as laws and culture, etc. On the micro-level, observable or otherwise perceivable objectified micro-interactions (e.g. interaction with other agents or observation of other agents' interactions) find their equivalence in an agent's social in-/externalisation processes which shape its disposition (habitus) in the social environment (field). In contrast, subjective micro-level aspects reflect fundamental internal concepts of agency, such as perceptions, beliefs, intentionality (aspects we will briefly allude to in Subsection 2.3.1) and expectations. With particular reference to the latter Castelfranchi (Castelfranchi, 2014) goes as far as to propose that the minds themselves *are* social institutions.

Though seemingly complicating the analysis, from a standpoint of institutional modelling in general, and normative modelling in particular, the differentiation between subjectivity and objectivity offers the ability to clearly differentiate norms and rules, generally characterised as informal and formal institutions (see e.g. North (1991, 2005)). We explore this aspect in more depth in the following Subsection 2.2.1, since it builds the foundation for the conceptual contributions introduced in Chapter 5 onwards.

2.2 Domain-Specific Foundations

The sociological backdrop sets the scene for the domain of our work, the modelling and analysis of institutions, which, as we will see, represents an intricate marriage of agency and structure, micro- and macro-sociological perspectives as well as subjective and objective dimensions.

2.2.1 Institutions

Definitions of institutions appear to be as numerous as their researchers. This is symptomatic for the hardly unified understanding across different fields of inquiry, but likewise the multitude of observable facets of institutions, such as their characteristics and purposes. However, as much as this harms a unified understanding, it shows the extent to which institutions are of relevance for the functioning and understanding of human societies.

Specific interpretations are generally roughly associated with the respective research field, an aspect we will explore as part of this section. First, however, we will provide an overview of commonly used definitions of institutions, and will use those to refine the institutional understanding for the scope of this thesis.

Definitions and Purpose

Definitions Institutions are often introduced by referring to North's popular slogan describing institutions as "rules of the game..." (North, 1990). However, North's full definition offers a richer contextualisation and characteristics: "Institutions are the rules of the game in society or, more formally, are the humanly devised constraints that shape human interaction." (North, 1990) As a central aspect this definition shows the human origin of such rules and their application to social systems.

North's second definition (1991) offers a more refined account of their characteristics[5] in

[5]"... They consist of both informal constraints (sanctions, taboos, customs, traditions, and codes of conduct),

which he differentiates between "informal constraints", under which he subsumes sanctions, customs and traditions, and "formal rules" which he sees in artefacts such as constitutions and laws. North describes the complementary role of "informal constraints" to support the formal rules, suggesting a precedence of formal rules. His definitions further describe the purpose of institutions, which he sees in the *reduction of uncertainty* (North, 1991).

Greif's institution account puts stronger emphasis on the input factors for the establishment of institutions, describing them as a "system of rules, beliefs, norms, and organizations that together generate a regularity of (social) behaviour" (Greif, 2006). Those input factors themselves can be of institutional nature, but Greif puts particular emphasis on their interrelated nature, i.e. 'the system'. The input factors he describes, such as rules, norms and organisations, are institutions themselves, supporting the path-dependence of institutional development (see also Greif (2006); Libecap (1989)).

Emphasising the cognitive aspects of institutions, Aoki describes institutions as stable "self-sustained systems of shared beliefs" (Aoki, 2001). He refines those beliefs as beliefs about behavioural expectations of fellow institution participants. As such he avoids the phenomenological description offered by North and Greif. Essential aspects of Aoki's definition are the assumptions of shared meta-beliefs and stability of institutions, an aspect also alluded to by Greif (2006).

A definition that puts stronger emphasis on the past- and path-dependence of institutions (an aspect also acknowledged by North and Greif) is Kreps' definition as "the product of long term experiences of a society of boundedly rational and retrospective individuals" (Kreps, 1990). Beyond the importance of history, Kreps' definition emphasises the importance of experience for the shaping of institutions.

Following these accounts, institutions thus have formal and informal characteristics and imply a regularity and stability of behaviour. They can further be derived from a set of input factors that can be institutions themselves, and be interpreted as a system of shared beliefs about expectations.

Purpose of Institutions Before exploring further characteristics of institutions, let us outline their purpose. As mentioned before, North (1991) sees the essential purpose of institutions in the reduction of uncertainty. A central figure in the context of New Institutional Economics, Oliver E. Williamson (1998), sees institutions' essential purpose in the reduction of transaction cost, and thus efficiency increase, a more refined view of von Hayek's perspective (1973). Both North and Williamson assume the view (originally proposed by

and formal rules (constitutions, laws, property rights). Throughout history, institutions have been devised by human beings to create order and reduce uncertainty in exchange." (North, 1991)

Alchian (1950)) that institutions with superior efficiency outcomes prevail in the long run and replace inferior configurations.

At this stage it is noteworthy to remark on the stability characteristic of institutions, an aspect Ayres (1944) considers intimately associated with human strivings to maintain the status quo, be it in wealth or status, etc. He sees institutions as a force countering the continuously changing social, economic and technological environment. Following this account, the reduction of uncertainty appears as a more general purpose of institutions than the (originally) economically motivated yielding for efficiency, since institutions can produce suboptimal efficiency outcomes, especially when considering unconsciously developed habit (Veblen, 1899) as a central driver of institution formation (see Subsection 2.2.1; see also Ostrom's (2005b) description of the institutional lock-in effect). North's view supports this essential appeal to stability as a means to reduce uncertainty, even though it may lead to unfavourable efficiency outcomes.[6]

Whether considered a purpose on its own or a characteristic supporting the purpose of uncertainty reduction, Young (1998) and Aoki (2001), as well as Kreps' definition (1990), argue that institutions reflect humans' limited cognition, or *bounded rationality* (Simon, 1955).

A further purpose is discussed by Knight (1992) who interprets institutions as a means to manage the (re)distribution of gains.

Beyond the exploration of further definitions and purposes, we can differentiate institutions based on their characteristics.

Characteristics and Manifestations

The definitions outlined before offer a glimpse into the rich nature of institution characteristics. Structuring those systematically, we will briefly discuss important characteristics of institutions that are either only partly explored or implicit to the definitions and purposes laid out up to this stage.

Formal and Informal Institutions A central differentiation of institutions is their stratification into formal and informal institutions, as suggested by North (1991). With particular focus on informal institutions, Kingston and Caballero (2009) propose three categories:

- *Written vs. unwritten* – This describes institutions by their non-/codification and/or public enforcement.

[6]North (1990): "the major role of institutions in a society is to reduce uncertainty by establishing a stable (but not necessarily efficient) structure to human interaction".

- *Ethical codes/moral norms* – Those describe the often inexpressible rules or principles that are innate to the individual (and, as such, its preferences) and guide the rule selection process, e.g. by exclusion of rule candidates considered unfair (Ostrom, 2005b).
- *Informal rules* – Socially adopted and not necessarily intentionally designed rules are followed by each individual, since its unilateral deviation is not rational. This understanding reflects the concept of a Nash equilibrium (Nash, 1950), in which only collective strategy change would result in an overall gain (see Subsection 4.1.1).

As explored as part of Kingston and Caballero's first category, formal institutions are understood as codified. This generally (but not necessarily[7]) implies their written nature, and legal enforcement. Examples for those include constitutions, statutes, laws, bylaws and contracts (North, 1991). Their establishment is likewise generally associated with an appointed (e.g. state) or collectively/self-assigned authority (e.g. organisation). Similarly, the specification of enforcement is part of the formal institution. Ostrom (1986, 2005b) offers a refined outline of formal institutions, or 'rules', along with a functional decomposition into elementary rule types, an aspect that is beyond the scope of this discussion.

Subjective and Objective Aspects An aspect of institutions that demands attention to develop unambiguous terminology, is their often complementary subjective and objective nature.

As indicated in Aoki's definition (2001), institutions inherently rely on the shared belief of adherence. This view is fundamental to Searle's (1995; 2005) concept of institutions that makes the mental representation a constitutive element for their existence. Parsons (1951), similar to Weber (1978b), indicates that many institutions withdraw themselves from conscious perception and would thus not even be accessible to reasoning, making their internal, subjective representation a central concern.

Those subjective aspects of institutions may or may not be objectified and represent themselves in visible patterns of interaction and social practices (Tuomela, 2002, 2003), or physical artefacts, such as contracts, laws, but also in social structures, such as organisations (e.g. governance bodies, firms (Williamson, 1975b; Schotter, 1981; Young, 1998)), schools, families, clans, and communities (Granovetter, 1985; Nelson, 1994).

This differentiation shows overlap with the formal/informal dichotomy, but objectified manifestations of institutions may not necessarily be formal (e.g. social practices, communities, etc.). A more central aspect of this understanding is that the subjective perspective is

[7]Consider the notion of unwritten contracts. We will discuss the formal/informal dichotomy in the context of our scenario (see Section 3.2).

the sine qua non of institutions. Without subjective elements, such as a normative backing, rules are what Ostrom (2005b) refers to as hollow *rules in form*, as opposed to *rules in use*.

An accessible example for its different subjective and objective manifestations is the institution of money (Menger, 1892). Its practical use is embedded in the subjective domain, with each institution participant ideally holding a conceptualisation for its purpose and handling, along with the inherent belief in its operational value (i.e. its use for the exchange of goods and services) and the short-term stability of its value. Its objective representation can be in the form of coins, paper, plastic (credit cards) or electronic (e.g. online banking), along with the institutions (here: macro-objective organisations) that support the belief (i.e. banks, governments, central banks). In this case the objective domain overlaps with the formal aspects of money, since regulation and recognition by the state perfected its use (Menger, 1892). However, historic examples of commodity money (e.g. shells (Menger, 1892), cigarettes (Radford, 1945), etc.), and also modern creations such as virtual currencies (e.g. Bitcoin (Nakamoto, 2009)), show the objectified nature without satisfying characteristics of legal enforcement. The central role of the subjective perspective gains salience in the light of eroding trust into its value or institutional function, as evidenced in cases of hyperinflation (e.g. post World War I in Germany (Cagan, 1956)) and/or limited availability (e.g. concurrent use of eight foreign currencies in Zimbabwe (Hungwe, 2014)).

Capturing the essential features of institutions, specifically

- their grounding in social behaviour (be it the interaction between individual parties, groups or collective behaviour),
- their stability characteristics (independent of efficiency considerations – potentially leading institutions to survive their original purpose), and
- their potential reflection in informal/formal and subjective/objective dimensions (thus transcending the behaviour they originate in),

we offer a refined synthesised definition of institutions as *manifestations of social behaviour*.

Complementing the differentiation into informal/formal and subjective/objective characteristics, institutions have been analysed from different perspectives that champion specific institution interpretations.

Analytical Perspectives on Institutions

Research into institutions can roughly be grouped into three groups as outlined by Crawford and Ostrom (1995). The association with particular applications given here is by no means exhaustive but reflects the general streams of institutional researchers.

The first understanding is *institutions as equilibria*. This approach is agnostic about the nature and the content of an institution. Instead it conflates institutions, independent from their concrete instantiation as norms or rules, in the form of equilibria, which offers an abstraction of the institution attributes, as long as the function is preserved. An example for this is the use of contracts to facilitate cooperation between contracting parties. Although increasing transaction costs for either party (e.g. for negotiation and drafting), it minimises the risk of defecting from transactions, since formalised obligations are subject to enforcement. Assuming functioning enforcement, the use of contracts (in the context of other available options) represents an equilibrium strategy for open societies, in which informal market interaction would fail to provide market stability (as of strong incentives for cheating), despite the reduction of associated transaction costs. This comparatively abstract perspective is particularly useful for theoretical and formal institutional analysis, since it makes institutions accessible to rational choice theory. Researchers that analyse institutions from the equilibrium perspective include Schotter (1981), Calvert (1995), Aoki (2001), Myerson (1999, 2004), Greif and Laitin (2004) and Greif (2006).[8]

A second group of institutional characterisation is the interpretation of *institutions as norms*. In contrast to the identification of equilibria as candidates for institutions, the normative perspective involves a focal shift onto the implications of interactions, patterns of interaction and the shared beliefs that establish or sustain a given institution. This involves the relationship between individual norm participants and development of norms at group level. Members of this group include Lewis (1969), Ullmann-Margalit (1977) and Coleman (1987). Empirical work on norms can roughly be grouped into (a) game-theoretical approaches (e.g. Schelling (1960); Lewis (1969); Ullmann-Margalit (1977); Coleman (1987); Fehr et al. (2002); Gintis (2009)) that generally represent norms as Nash equilibria or cooperative outcomes in Prisoners' Dilemma Games with focus on norm spreading, and (b) artificial intelligence approaches (Neumann, 2008; Mahmoud et al., 2014). The latter is most prominently represented in the area of Normative Multi-Agent Systems (Boella et al., 2007) that explores the spread of norms (e.g. Andrighetto et al. (2010b); Villatoro et al. (2011a)) as well as their different stages and functions in the normative lifecycle (Savarimuthu and Cranefield, 2011; Hollander and Wu, 2011b; Mahmoud et al., 2014), such as emergence or internalisation (which will be discussed in Subsection 2.3.2).

The final group focuses on *institutions as rules* and emphasises formal aspects, an approach grounded in the legal perspective, such as Hohfeld (1913), and the principles of institutional economics based on Commons (1968). Its view is commonly reflected in the

[8]See also Kingston and Caballero (2009) for an overview of the 'equilibrium view'.

field of New Institutional Economics (NIE) (Williamson, 1975a) (see Subsection 2.2.3 for a more detailed introduction), and consequently the analysis is centred around formal institutional constraints with particular focus on the cornerstones of transaction costs, property rights and contracts (Ménard and Shirley, 2014). Researchers associated with this perspective on institutions include Shepsle (1989), Shepsle and Weingast (1984), Ostrom (1986, 1990, 2005b), Williamson (1975b), North (1981, 1990, 2005), Knight (1992), North et al. (2009), and Acemoglu and Robinson (2012).[9]

Outlining those different views on institutions showcases the complexity that surrounds the concept of institutions. The categories identified by Crawford and Ostrom (1995) are only rough approximations, since comprehensive models allow a combination of different perspectives, e.g. an integrated economic and social perspective with emphasis on both rules and norms.

It is important to recall that the original threefold stratification offered by Crawford and Ostrom (1995) is centred around an economic and thus primarily game-theoretical perspective. The general popularity of rational choice approaches and associated analytical methods blurs the lines between the different perspectives,[10] in particular for rule and equilibria interpretations. Considering the stronger emphasis on 'social' approaches to institutions for the norms perspective, such as found in the context of Normative Multi-Agent Systems (see Subsection 2.3.2), we can alternatively structure the different perspectives based on the focal point of analysis.

Equilibrium-based approaches analyse existing institutions and develop a model representation (e.g. in the form of games) in order to generalise possible outcomes of such institutional constellations, without attending to a detailed structural representation of the institution itself. Norm-based approaches, in contrast, put stronger emphasis on the emergence, diffusion, in addition to the stability effects of institutions. They directly or indirectly evaluate how the 'shared belief' in a norm is maintained, which is reflected in cooperative outcomes based on informal means, i.e. either by self-enforcement or distributed enforcement. The rules-based approach, especially when considering its central application by NIE researchers, asks the question how far particular institutional configurations in the shape of rules and regulation affect efficiency outcomes (e.g. in terms of transaction cost). Thus,

[9] Note that the association of researcher and research field with analytical perspective is not crisp. For example, Greif (1989, 1993, 2006) can be associated with NIE, but he opts for an equilibrium-based method of analysis (see Subsection 2.2.3) that represents a preferable match for the problem domain since it involves the representation of *informal* institutions.

[10] Greif is an example for this case. His approach is centred around equilibrium-based analysis, but his various analyses involve a relatively detailed consideration of historical aspects reflecting formal institutional structures (e.g. his analysis of the power equilibrium of competing Genoese clans (Greif, 2006)).

while the equilibrium perspective concentrates on the exploration of possible *outcomes*, the NIE perspective can be associated with the exploration of outcomes with respect to *inputs* in the form of formal institutions, or rules. Besides input and outcome, norm-based approaches target the reconstruction of the social *processes* that mediate sustained cooperation or breakdown.

It is important to clarify that all approaches generally analyse outputs based on given input for the proposed model (which embeds the process perspective). However, the proposed allocation emphasises analytical focal points for the respective views, offering a more general differentiation between the discussed analytical perspectives on institutions.

2.2.2 Institutional Dynamics

To develop an understanding of the interaction between different institution types and their dynamics over time based on literature accounts, we explore the process of institution formation as well as institutional change.

The understanding laid out in this and earlier sections will then be synthesised into a prototypical institution formation process, the principle of which underpins the conceptual contributions offered in Chapter 5 onwards.

Institution Formation

Habit With the intent to describe the root of all institutions, Veblen recognizes the fundamental importance of the subjective dimension to drive institutional change in what he describes as "habits of thought" (Veblen, 1899). "Habituation" (Hodgson, 2006), driven by repeated behaviour in the context of comparable situations (Thomas and Znaniecki, 1920) is thought of as a rationalisation process in the light of limited cognitive capacity (Kingston and Caballero, 2009; Veblen, 1899). In Bourdieu's terms (see Subsection 2.1.1), it is part of developing a 'habitus' for the operation in a given 'field'. Habits may be consciously or unconsciously shaped (Hodgson, 2006) and may themselves rely on beliefs and attitudes that relate to behaviour undergoing the institutionalisation process, and thus ultimately involve a mental representation of the social reality (Searle, 1995, 2005).

Conventions The socialisation of behaviour can then shape collective regularities, with the lowest common denominator being collectively-adopted self-enforcing social behaviour, such as conventions, "informal rules" in Kingston and Caballero's (2009) classification, or "shared strategies" (Crawford and Ostrom, 1995, 2005). The most common example for such behaviour is the convention of road-side use, for which unilateral deviation would not provide additional gain but rather immediate sanctioning. Other examples include 'shaking hands' upon encounter in many Western cultures, or the convention that pouring of drinks is performed by young persons to senior individuals in many East-Asian societies. Non-compliance generally incurs no sanction but may signal disrespect, etc. The concept of 'folkways' (Sumner, 1906), such as traditions, has been used as an intermediate concept between conventions and norms, in which consequences of non-compliance, if at all, are of limited impact. Conventions can likewise be interpreted as *descriptive norms*, suggesting what one 'does' as opposed to prescribing what one 'should do' (*injunctive norms*) (Cialdini et al., 1990), the latter of which is generally expressed by obligations and prohibitions, aspects that are central to norms discussed in the following paragraphs. Summarizing, a central characteristic of conventions is thus that "everyone wants to conform" (Young, 1993).

Clarifying the 'Norm' In contrast to 'conventions', in literature the terms 'norms' and 'rules' are used in an inconsistent manner.[11] North (1991) and Ostrom (2005b) subsume those under the concept of institutions. Other researchers equate institutions with the concept of norms, and introduce sub-types, such as Tuomela's (1995) differentiation into *r-norms*, which are enforced by authority (and thus formal institutions/rules according to our understanding), *m-norms* (moral norms), which represent moral obligations, and *p-norms* (prudential norms), which are based on the assumption of rational behaviour. Beyond this Tuomela conceives *s-norms*, or social norms, which reflect the institution concept of relevance for us. For Tuomela (1995) social norms are both r-norms and s-norms, thus rules and what he refers to as *proper social norms*, which are characterised by mutual belief as constitutive characteristic. S-norms capture both conventions and group-specific norms, the latter of which thus apply to particular groups, not the society at large. However, in our conception conventions and social norms can apply at all levels of social organisation. From a legal perspective, norms are seen as rules imposed by an appointed authority (Verhagen, 2000; Posner and Rasmusen, 1999), largely consistent with what Tuomela sees as *r-norms*, and what we henceforth interpret as rules (consistent with North (1991) and Ostrom (2005b)).

Social Norms In this context we use the term 'social norm'[12] to signify conventions for which compliance is enforced, generally by fellow norm participants (Ullmann-Margalit, 1977). The necessity of enforcement can either lie in its importance, a 'norm as conduct',

[11] A good overview on the inconsistent understanding of conventions, norms and rules terminology is offered by Crawford and Ostrom (1995, 2005).

[12] For the remainder of this work we use the terms 'social norm' and 'norm' interchangeably.

or underlying ethical code (Kingston and Caballero, 2009), such as sanctioning jaywalking in order to avoid exposure of such behaviour to minors, or the diverging individual interest from collective goals, such as the protection of common pool resources from exploitation (Ostrom, 1990). In this case enforcement depends on the mutual ascription of belief in the norm's objective by norm participants, i.e. assuming an aligned collective aim (Hodgson, 2006) ('shared aim'), a process Tuomela (1995) named *collective intentionality*. Following this motivation, validation of such belief can then be sought by observing fellow norm participants' sanctioning behaviour. Sanctioning here implies not only direct sanctioning of the violator (Coleman, 1990), but also sanctioning of non-compliant sanctioners (second-order sanctioning) (Axelrod, 1986). Alternatively to sanctioning of violators, conformity of norm participants or sanctioners may be rewarded (Goode, 1978). Enforcement is consequently generally decentralised (Roland, 2004) (given that it relies on mutual ascription of intention); the number of sanctioners and means of sanctioning can differ. In any case, social norms require enforcement in order to be sustained (Coleman, 1990; Oliver, 1980).[13] For norm participants the incentives to follow a norm are manifold, including fear of authority (Axelrod, 1986), material sanctions (Posner, 2000), impact on status and reputation, avoidance of friction with fellow norm participants and emotions such as guilt and shame (Elster, 1989; Posner and Rasmusen, 1999), future impact on information sharing or multi-lateral sanctioning by seemingly uninvolved third parties (Posner and Rasmusen, 1999).

However, given the enforcement characteristics (decentralisation and second-order enforcement), the functioning of norms requires relatively low individual enforcement cost or sufficiently high stakes in the defended norm, an aspect that many sociologists since Durkheim have associated with the density of relationship networks (Durkheim, 1952; Simmel, 1902; Burt, 1982; Coleman, 1990). In essence 'everyone wants that everyone else wants everyone else to conform'.

Rules In contrast to generally informal social norms (if that implies 'neither written nor enforced'), in our conception 'rules' cross the informal/formal barrier, with rules offering a codified specification of conduct, consequences, and appointed enforcer. While the notion of collective intentionality has implicit characteristics and manifests itself in individuals' beliefs, rules are established based on collective action (Crawford and Ostrom, 1995; Ostrom, 2005b), or explicit declaration by an authority empowered to do so (Tuomela, 1995). Their establishment is purposeful if collective intentionality cannot be assumed (i.e. lack-

[13] Axelrod's Norms Game (Axelrod, 1986) represents a convincing empirical account for the necessity of second-order enforcement (i.e. sanctioning non-sanctioning norm monitors) to sustain norms.

ing mutually ascribed commitment to norm), if enforcement cost becomes sufficiently high (e.g. limited monitoring because of low density social networks), or if fair sanctioning is of importance (see e.g. Posner and Rasmusen (1999) for the legal perspective).

Institutional Regress Starting from 'habits', the extent to which institution objectives are shared (as 'shared aim' vs. 'shared strategy') is decisive for their manifestation. Habits themselves, however, are not the starting point of institutional formation, since they cannot be dissociated from exogenous influences that shape them. Thus the development of institutions does not occur in a void but relies on preceding institutions that shape emerging new institutions, a process referred to as *institutional regress* (Aoki, 2001; Hodgson, 2002). Institutions are faced with a similar precedence problem as the structure-agency problem discussed in the beginning of this chapter (see Subsection 2.1.1). If we follow Bourdieu's understanding of habitus as enacted history, an individual is at least in part shaped by preceding institutional processes, which themselves are grounded in individuals' interactions recursively depending on preceding institutions. Hodgson clarifies that an institution-free 'state of nature' (Williamson, 1975a) as suggested by Williamson cannot be assumed,[14] and describes this 'chicken-egg problem' of institutional development as *infinite institutional regress* (Hodgson, 2002).

Institutional Change

Beyond the formative aspects of institutional development and the resultant stabilising institutions (conventions, norms, rules), institutional researchers have developed diverse views on how institutional change, and in particular, how the interaction and adjustment of formal and informal institutions occurs over time.

In Williamson's (2000) NIE view informal institutions are excluded from analysis. He sees informal rules as largely exogenous factors, since he considers their change too slow to exert dynamics within the observational scope of New Institutional Analysis. In his Four-Layer Framework of Analysis (Williamson, 2000) (which we introduce in Subsection 2.2.3) Williamson suggests that informal rules 'embed' the more dynamically changing formal institutional environments.

North (1990), in contrast, emphasises the central role of consciously established formal rules that change the overall institutional environment. Informal institutions undergo a continuous change that follows the change of formal rules but dampens their effect. In North's

[14] Williamson sees markets as starting point of all development: "in the beginning there were markets" (Williamson, 1975a). However, Hodgson (1988) and Loasby (2000) clarify that markets are institutions themselves.

view informal and formal institutions are thus in a continuous interplay, with formal institutions driving intentional change that is gradually adopted by informal institutions, ultimately representing endogenous influence on future institutional adaptation.

Researchers, such as Roland (2004), emphasise the speed of the change process. "Fast-moving" formal institutions can undergo rapid changes (but not necessarily frequently) by means of centralised policy-making. The more continuous "slow-moving" decentralised informal institutions can promote conflicts with existing formal rules, which, as a consequence, are consciously adapted to resolve the dissonance. Ruttan (2006) introduces a more refined framework, in which he highlights the mutual interrelationship of institutions, culture, technology, avoiding a singular concentration of institutional influence. He shares, however, the emphasis on informal institutions (as opposed to formal institutions) as an important driver of change, but relativises the speed implication of informal vs. formal institutions. Instead Ruttan offers a refined differentiation of informal institutions in which he distinguishes arrangements that are relatively fast-moving[15] and slow-moving, such as cultural influence, which Ruttan refers to as "cultural endowments" (Ruttan, 2006).

Similar to Roland and Ruttan, Brousseau and Raynaud (2006) argue the importance of the informal perspective. In contrast to Roland, they do not conceptualise a delayed adoption of informal institutions by rapid adjustments in the formal domain. Instead they suggest a coexistence of multiple competing informal (private-order) institutions that are influenced by social and economic processes, which drive the emergence of a dominating institution that finds increasing adoption by participants. Brousseau and Raynaud believe that the increasing adoption limits the continued change of such institutions, making their manifestation as public-order institutions only consequent. In contrast to Roland, private-order institutions thus change their state, progressing from informal, or private-order institutions to formal ones, instead of coexisting. This change is ultimately facilitated in the 'formal domain'.

Our work seeks to explain economic behaviour by incorporating social influence. The understanding of institutional change in our work thus interprets informal institutions as central drivers of change, while committing to the path-dependence laid out by preceding formal and/or informal institutions. As such Roland's (2004) metaphor of tectonic pressure accommodates for the sudden change in the institutional environment as a result of ongoing institutional change. The elaboration of Ruttan (2006) shows that the simplified bifurcation into "slow" and "fast" does not account for the complex nature of informal institutions compared to well-defined formal institutions, since the informal institution concept itself is not easily decomposable because it represents an amalgamation of structurally diverse long-

[15]Ruttan makes reference to informal tenancy agreements in Phillipine villages.

term cultural and short-term behavioural adaptations. Brousseau and Raynaud (2006) take an evolutionary view and suggest a 'progression' from 'informal to formal' for informal institutions that withstand the competition of alternative behavioural regularities. Their evolutionary view is an exception; all other discussed views on institutional change emphasise the 'synchronisation' of informal and formal, but maintain a clear separation of formal and informal.

Synthesis

At this stage we summarise essential aspects of institutions, specific types and their interrelation in the form of an integrated formation process that lays out the institution understanding applied in this work.

Institutions can generally be stratified into informal and formal institutions that entail subjective and objective perspectives (see Subsection 2.2.1).

With respect to their content and characteristics of enforcement we can structure three elementary institution types as reflected by different schools of thoughts, or analytical perspectives (see Subsection 2.2.1), as summarised in Table 2.1.

Table 2.1: Overview on Selected Characteristics of Different Institution Types

Type	Content	Enforcement	Enforcement Cost
Convention	shared strategy	self-enforcing	none/low
Social Norm	shared aim	decentralised	low
Rule	declared/imposed aim	centralised	low-high

Conventions entail shared strategies, – strategies individuals *want to follow* since all others follow those – and are commonly invoked for game-theoretical analysis in which equilibria signify such shared strategies. As part of institutional change they can be transformed into norms (e.g. if compliance is harmed by conflicting norms) or rules (e.g. transformation of road-side use into traffic rules).

Norms extend this notion with aspects of enforcement, since the notion of a common aim (as opposed to a strongly aligned 'shared strategy') remains, but incentives for deviance increase. Enforcement originates in the norm participants themselves which requires a sufficient alignment in collective intentionality (see Subsection 2.2.1). If enforcement is challenged and the stakes sufficiently high, norms can be backed by formal rules with explicit specification of sanctions and appointment of sanctioners.

Rules represent a codified agreement on aim and strategies and may either derive from a convention or norm, or, if imposed, constitute norms that adopt a specified enforcement blueprint specified in the rule. In this conception rules are aspects of the objective domain (see Subsection 2.1.3).

Given the necessity to back rules with sufficiently aligned norms to assure compatibility of formal and informal institutions (see North (1991)) and make rules effective (see Ostrom's (2005b) *rules in form* vs. *rules in use*), enforcement of objective rules in our conception originates in constructed subjective norms that back the respective objective rules. Seeking terminological compatibility with Bourdieu's conception of history (see Subsection 2.1.1), we can define norms that derive from rules as *enacted rules*. As a bottom-line, enacted rules are at least supported by adopted enforcement norms that describe the obligations of sanctioners, who impose the *content* onto the norm subjects (i.e. the individuals to which the content applies and whose conduct is observed), whether or not the norm subjects adopt the normative content. Exemplifying this view, the effectiveness of rules associated with police officers' duties depends on the officers' enacting of such rules, i.e. *sharing the aim* of the rules' fulfilment in the normative space. Committing to the fulfilment of the rules would conflict with norms of corruption, the latter of which would render the imposed enforcement duties as 'rules in form' (Ostrom, 2005b) with the mere pretence for enriching those appointed to enforce them. Another aspect that shows the subjective dimension of rules is the case of graduated sanctioning, in which the determination of sanctions underlies the autonomy of the appointed enforcer. The inverse case, the compatibility of rules that emerge out of norms, is unproblematic with respect to such alignment, since the emanating rules represent the codification of already established norms.

Though this complementary view on norms and rules may not capture all peculiarities attached to either concept, it highlights the necessity of a subjective interpretation and enactment of objectified rules in the form of backing norms.

We summarise the institution understanding laid out here and in the preceding sections in an integrated institutional formation process (Figure 2.5). It is based on Bourdieu's (1977) integrated conception of agency and structure, which is compatible with Veblen's (1899) habit formation without necessarily requiring intentionality (Hodgson, 2006; Vanberg, 2002). We further apply Ritzer's (1981) analytical dimensions of social systems (micro/macro – subjective/objective) to maintain the differentiation of subjective perspective from objective perspective across micro- (agent) and macro- (institution) levels. Hodgson's (2006) *reconstitutive downward causation* explains the feedback process leading to the adjustment of habits based on internalisation of learning and enforcement. This subjectivity-centred formation model further reflects the dependence of macro-level phenomena (here: institutions)

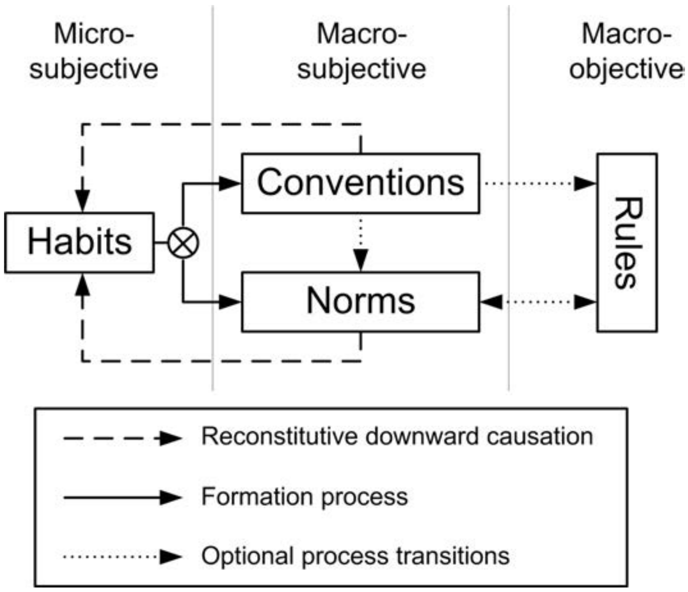


Figure 2.5: Integrated Institution Formation and Change Process

on the collective of individuals and their normative alignment, whether by force (top-down) or shared intentions (bottom-up).

The potential transitions between different institution types can be highlighted by examples. Conventions, such as road-side choice as a self-enforcing shared strategy, can be directly codified as rules, and even as such maintain their self-enforcing nature, since their status in the macro-subjective domain did not change. Similarly, the notion of language, represented as a self-policing convention (see Hodgson (2002)) can be formulated in terms of rules in order to govern its stability and unambiguous utilisation for the purpose of codification.[16]

Money, on the other hand, is not self-policing (Hodgson, 2002), and relies on enforced standardisation in order to undermine incentives for debasing and substitution by an exchange medium of inferior value. As such it bears normative characteristics and relies on collective enforcement, such as formally appointed enforcers that sustain the collective aim.

The inverse effect, inducing norms based on rules is an example for top-down enforcement, such as the case in the context of colonisation, be it the literal occupation of societies and imposition of rules, or Habermas' metaphorical colonisation (1987) as occupation of the lifeworld (public and private life) by steering influences (e.g. media). Similarly, assuming a conception from the computational domain, the norms of behaviour on the internet are imposed, shaped and governed by the fact that "code is law" (Lessig, 1999).

As mentioned before, institutions that have a representation in the subjective domain

[16]Bloomfield (1933) refers to dictionaries as an example of explicit definition, or formalisation, of language.

(whether imposed from rules, or emerged from previous institutions or habit) have indirect influence on future institutions by means of reconstitutive downward causation (Hodgson, 2006).

2.2.3 Institutional Analysis

Since our work explores specific institutions (Chapter 4) and provides contributions to the area of institutional modelling and analysis, we cannot omit a discussion of selected approaches to institutional analysis.

The concept of *Institutional Analysis* (IA) grew out of the economic subdiscipline of *New Institutional Economics* (Williamson, 1975a) which itself is a revision of *Institutional Economics*. Institutional Economics, or *Original Institutional Economics* (OIE), is associated with Veblen (1904) and Commons (1936) who recognised the multi-dimensional influence of legal environment, social norms, cognitive aspects, but also individuals, firms and the state, in short institutions, onto the development of markets (which are institutions themselves). However, only the systematic characterisation of exogenous and endogenous influence factors and governance structures, along with the integration of methods from neoclassical analysis offered an explanatory framework, *New Institutional Economics* (NIE). NIE could accommodate the impact of exogenous events, such as the case for the Great Depression, and offer a more general framework that removed the cultural bias of the earlier institutionalism (Groenewegen et al., 2010; Joskow, 2004; Rutherford, 2001). This latter research field was championed by nobel laureates such as Coase (1937, 1959), Williamson (1975a,b, 1998, 2000), Ostrom (1986, 1990, 2005b, 2008) and North (1990, 1991, 2005). More recent contributions to the field come from Avner Greif (2006), whose work is of central relevance in the context of this thesis, as well as Acemoglu and Robinson (2012). With particular focus on the earlier work of Coase, Williamson and North, Ménard and Shirley (2014) characterized NIE in the form of a "golden triangle", highlighting the essential relevance of *transaction costs*, *property rights* and *contracts*, a view that has been challenged for its overly simplifying account (see Hodgson (2014)).

Along with the shift from the focus on identifying the shortcomings of traditional economic analysis in the form of OIE, the integrated perspective offered by NIE led to the introduction of tools for a systematic analysis. We will briefly explore such models of IA, along with their focal objectives. We discuss Williamson's *Four-Layer Model* (Williamson, 2000), Ostrom's *Institutional Analysis and Design Framework (IAD)* (Ostrom et al., 1994), and finally, we will highlight Bates et al.'s *Analytic Narratives* (Bates et al., 1998) as a means of analysing economic scenarios. Following this, we will briefly contrast the different models and describe their relationship to this work.

Williamson's Four Layer Model

Williamson (2000) proposes a model based on four levels of social analysis as shown in Figure 2.6, each of which he associates with its respective research field/s.

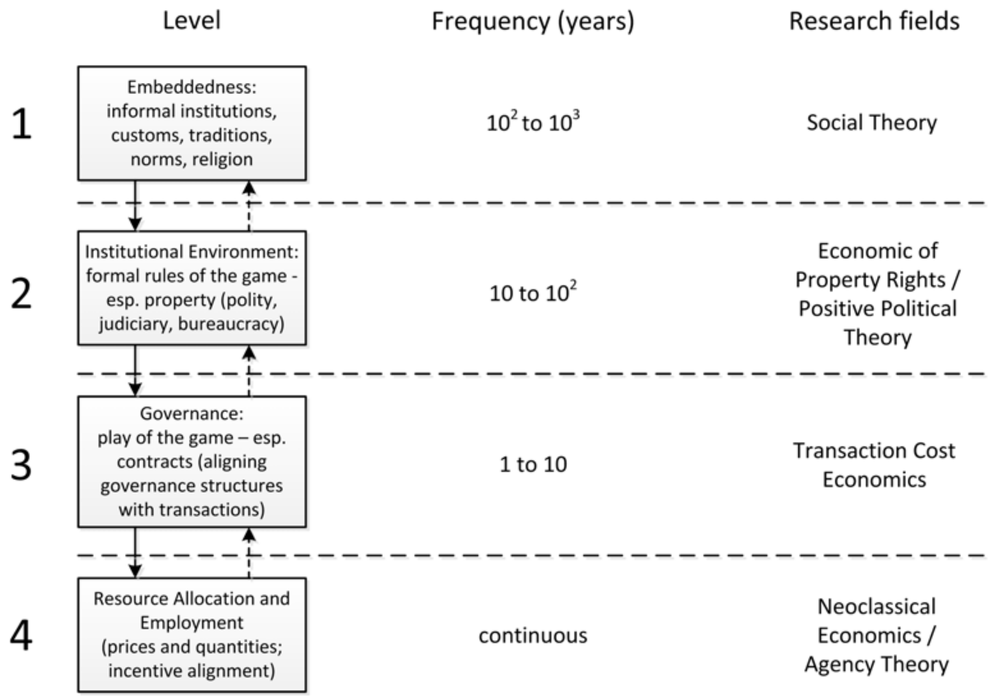


Figure adapted from Williamson (2000)

Figure 2.6: Four Layer Model

The highest level represents the embedding social and cultural foundations, essentially capturing aspects we have previously characterised as 'informal institutions', or North's *informal constraints* (North, 1991). In Williamson's view those are extremely slow-moving with a change frequency of 100 to 1000 years, and thus, other than being exogenous input factors, bear limited relevance for institutional analysis. On the second level exist the 'formal institutions', or North's 'rules of the game', that characterise the political and legal institutional environment, including enforcement mechanisms, such as commercial courts. On this level, change frequency ranges between 10 and 100 years. The third level of Williamson's model are aspects of governance, i.e. 'playing the actual game'. Concrete operative instruments, such as contracts, are allocated on this level, offering the basis for transactions, thus making this level subject of *Transaction Cost Economics*. Adjustments take place in a temporal frame of 1 to 10 years. The fourth and lowest level reflects the continuous market operation including tasks such as resource allocation and marginal analysis, thus covering approaches from the area of neoclassical economics. Williamson suggests a short-term focus on this level with a continuous adjustment. All levels exert influence on the next lower levels, and lower levels can likewise affect higher levels. In NIE, which Williamson allocates on the second and third level of the model, the effectiveness of contracts as means of regulating cooperation relies on the supporting institutions on Level 2 of his model, i.e. the underlying law, property rights, and enforcing entities. Those in turn are governed in the light of the informal constraints allocated on Level 1, and include cultural and social aspects, that influence the rule formation processes on Level 3 and in turn the manifestation of institutional instruments. In Williamson's model the exploration of the highest level (Level 1) is left to the area of social theory.

Institutional Analysis and Development Framework (IAD)

The Institutional Analysis and Development Framework (IAD), originally designed by Oakerson (1992), centres around the idea to model the governance of common pool resources, and has been subsequently extended by Elinor Ostrom and others (Ostrom et al., 1994; Ostrom, 1990, 2005b) into a comprehensive institutional framework as visualised in Figure 2.7.

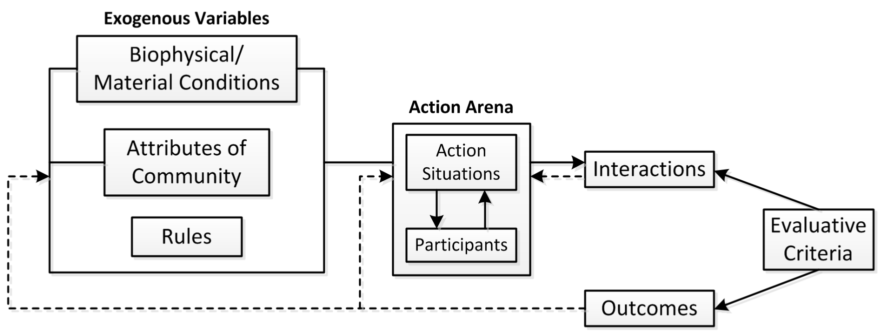


Figure adapted from Ostrom et al. (1994)

Figure 2.7: Institutional Analysis and Design Framework

The IAD in its current form concentrates on the representation of institutions with respect to their context, including physical and social environment as well as existing rules. Institutions of interest are then specified and analysed with respect to a given *Action Arena*, in which action situations and participants interact under the influence of *exogenous variables*, and produce outcomes that feed back both on the action arena and the environmental variables. The action situation holds a central role of analysis, requiring the specification of variables such as participants, positions, potential outcomes, action-outcome linkages, control of individual participants, types of generated information, as well as costs and benefits associated with the respective outcomes. (Ostrom, 2005b)

Analysis occurs across three interlinked levels, or tiers, that underlie their own respective rulesets, and are equivalent to levels 2-4 of Williamson's model. On the lowest *operational tier*, the application of *operational rules* constrains the interaction of individuals that have a direct impact on the world, including practical examples such as analysing the different performance of day care centers (Bushouse, 2011).[17] Individuals on the operational level act based on policy constellations emanating from the *policy tier* (or *collective-choice tier*) and are established based on *collective choice rules*. Actors on the policy tier act under the constraints of the constitutional tier that determines the participants and rules of policy-making processes by means of *constitutional choice rules*.[18] Apart from the different actors, roles and rules on the respective levels, Ostrom, similar to Williamson, posits the varying lifetime of rules on the different tiers, explaining gradual institutional change across all institutional layers.

The particular strength of the IAD framework is its wide range of cross-disciplinary applications, modelling diverse institutional settings, such as the analysis of coffee cooperatives in Cameroon (Walker, 1998), or the decentralisation of forest governance (Andersson, 2004). Central (though not exclusive) theme of applied scenarios is the management of Common Pool Resources (CPR), a central topic of Ostrom's own work (Ostrom, 1990).

Analytic Narratives

A final approach we briefly introduce under the header of *Institutional Analysis* is the concept of Analytic Narratives. *Analytic Narratives*, proposed by Bates et al. (1998), combine historical analysis based on narrative accounts with rational choice theory, thus making historical decisions accessible to the logical rigour of game-theoretical analysis. Historical decision points are extracted from narrative accounts and formulated as games in which selected parameters are assumed as given, thus complementing a macro-structural analysis of those scenarios. The abstract nature of this approach (compared to the previous approaches) is reflected in the interpretation of game-theoretical equilibria as indicators for institutions (Bates et al., 1998).

Compared to the previous approaches, analytic narratives offer the least structured account, leaving great autonomy on the part of the researcher with respect to the identification of historical decisions as well as events that affected the historical course. This autonomy

[17] See Ostrom (2005b, 2011) for further examples.

[18] Ostrom (2005b) further identifies a 'metaconstitutional level' that governs operations on the constitutional choice level, but it of limited relevance for practical applications of the IAD framework.

extends to the applied techniques, such as game-theoretical analysis. Greif (2006) in particular, approximates changing accounts of rationality in iterated games by introducing notions of "quasi-parameters" that are static for a given iteration, but can change as an outcome of the previous iteration, and thus affect future iterations.[19]

Summary

The highlighted approaches to institutional analysis vary in their scope of observation and function.

Williamson's Four-Layer Model provides an association of different levels of analysis with the research fields, with the first level concentrating on informal constraints, which Williamson sees as exogenous inputs for analysis on the lower levels, and which he considers an objective of inquiry in the field of social theory. The next increasing short-term-oriented lower levels describe formal rules that constrain actor behaviour and mediate the application of institutional instruments such as contracts, a central concern of Transaction Cost Economics and New Institutional Economics. The lowest level is reflected in traditional neoclassical economics.

Ostrom's IAD framework attempts to integrate the different levels (three levels in her conception, which are equivalent to Williamson's levels 2-4) into a functional framework that allows systematic description and analysis of institutions across all levels, reaching from the constitutional tier to the collective choice tier, and ultimately the operational tier. Though relying on game-theoretical exploration on all levels, the IAD seeks an integrated cross-disciplinary perspective on institutions, and offers the most refined approach that aims at application to concrete cases.

Analytic Narratives represent a cross-over of historical analysis based on literature accounts, strong contextualisation, interleaved with game-theoretical analysis of specific historical choices and decisions. This approach has its virtue for modelling scenarios that are characterised by relatively poor data coverage, and thus withdraw themselves from the more comprehensive specification necessary for the analysis using IAD. As such analytic narratives offer the most abstract and least structured account of institutional analysis, while bridging the link to hardly quantifiable historical scenarios, making those accessible to systematic analysis after all.

In the context of this work, the first model helps to differentiate the research fields concerned with institutions, especially the association of formal and informal institutions discussed in the previous Subsection 2.2.1. The IAD framework offers a flexible and com-

[19]We will explore this aspect in greater depth in Subsection 4.1.1.

prehensive mechanism for describing institutions, namely the institutional grammar ADICO (Crawford and Ostrom, 1995), which we will explore in the conceptual part of this thesis in Chapter 5 onwards. The analytic narratives bear great relevance for Greif's analysis of the Maghribī Traders Coalition (Greif, 1989, 1993, 2006), a scenario we will explore throughout the entire thesis (from Chapter 3 onwards). All mentioned approaches focus on a comparative-static analysis of institutions, and build on the assumption of rational selfish utility maximisers. Both aspects are in contrast to the dynamic perspective we will invoke in the conceptual part of this thesis.

2.3 Technological Foundations

With these essential sociological and domain-specific foundations, we now turn to the discussion of technological aspects that are central for the ensuing exploration.

2.3.1 Agent-Based Modelling & Simulation

In this work we apply *Agent-Based Modelling and Simulation*, subsumed as Agent-Based Modelling (ABM) for this work, as the central technique to model social systems. According to Gilbert (2008), an ".. agent-based model is a computer program that creates a world of autonomous heterogeneous agents in which each agent interacts with other agents and with the environment" (Gilbert, 2008). Compared to other micro-sociological and macro-sociological approaches to model human societies, ABM is a comparatively recent approach that focuses on the bottom-up interaction to represent observable social behaviour. It goes beyond Microsimulation (Gilbert and Troitzsch, 2005) which models individuals as sets of properties that are manipulated to show dynamics on population level in the form of aggregate data. Though modelling on the micro-sociological level, Microsimulation entails entities that do not directly interact and thus cannot exhibit emergent properties based on sociality. Models based on System Dynamics (Forrester, 1971) focus on the macro-sociological level, offering equation-based state transitions, and thus rely on explicit prior knowledge about the causal influences between system entities, an aspect that can well reflect physical systems with known variables, but may have limited applicability in social systems, where causal relationships are poorly understood (Gilbert and Troitzsch, 2005; Elsenbroich and Gilbert, 2014). In its essence, agent-based modelling links both micro- and macro-sociological levels by relying on the specification of prototypical individuals (agents) that carry characteristics relevant to the scenario of interest (such as behavioural rules and social relationships, personality, culture), and interact directly (e.g. via messages) and/or indirectly (e.g. via a shared environment) (Gilbert, 2008; Macal and North, 2010). They form artificial societies that resemble a society of interest with respect to phenomena on micro- and macro-level that can be grown from the bottom-up (Epstein and Axtell, 1996) (see discussion in Subsection 1.3.2). Emergent phenomena that may be hard to explore using equation-based models can then be tested with respect to their dependence on individual and social characteristics (Heath et al., 2009).

The agent concept itself is borrowed from the area of multi-agent systems, which puts stronger focus on cognitive aspects, and consequently has a long track record in developing the agent concept.

Jennings and Wooldridge (2000) offer a widely accepted definition of an agent as: "... an encapsulated computer system that is situated in some environment, and that is capable of flexible, autonomous action in that environment in order to meet its design objectives" (Jennings and Wooldridge, 2000).[20]

As implicitly referred to before, commonly ascribed characteristics for agents include:

- *Autonomy* – Agents act independently of other agents (executional autonomy), and may also be able to develop goals they intend to achieve (motivational autonomy) (Castelfranchi, 1995).

- *Reactive and Proactive Behaviour* – In addition to showing reactive behaviour based on invocation and perceived changes in their environment, agents can operate proactively in the pursuit of their goals.

- *Sociality* – Agents have means of social interaction, either by modifying their shared environment (e.g. via stigmergy or blackboards (Englemore and Morgan, 1988)) or via direct communication (Ferber, 1999).

Agents are instantiated as part of a multi-agent system (MAS) that, according to Ferber (1999), consists of an environment, objects, relationships amongst those objects and environment, as well as agents (which are specialisations of objects) that perform operations to act on the environment and other agents.

The listed characteristics are not exhaustive and can be extended with respect to the research field of interest. For example, the area of MAS puts stronger emphasis on the 'intelligent agent' concept that has cognitive capabilities based on symbolic knowledge representation and manipulation, such as the Belief-Desire-Intention (BDI) model (Bratman, 1987; Rao and Georgeff, 1995). From the perspective of distributed systems, in contrast, aspects

[20]For a selection of further definitions and their discussion refer to Franklin and Graesser (1996).

such as mobility are of greater concern (White, 1997; Fuggetta et al., 1998). In ABM the agent concept is more flexible and often more simplistic, since the focus is to model group or societal characteristics, which suggest stronger emphasis on greater number of agents as well as heterogeneity of agents. Consequently, stronger focus is put on behaviour as opposed to cognition (Elsenbroich and Gilbert, 2014).

The Agent Concept in Social Simulation Earlier, Carley et al. (1998) addressed this trade-off by associating simple agent architectures with an analytical emphasis on macro-level behaviour while suggesting that cognitive architectures are useful to investigate micro-level behaviour. However, the availability of increasing computational capabilities and demands for more realistic representations of behaviour spawned a debate in how far agent models that favour larger numbers of entities over their individual intelligence should emulate a wider range of humanoid capabilities and consequently adopt features of richer agent architectures (see e.g. Wellman (2015), Conte et al. (2012)). Aspects of this nature include incremental steps from hard-coded execution cycles towards autonomous behaviour based on sophisticated cognitive concepts such as beliefs (and associated knowledge-based communication), as well as expectations (which Castelfranchi (2014) sees as the basis for social institutions in the first place), ultimately facilitating the emulation of reflective behaviour. The promise of including advanced concepts lies both in the production of more realistic behaviour and detection of unexpected outcomes simpler agent models potentially prevent by design (Sun, 2009). This perspective provokes the consequent use of neurologically-inspired agent architectures (see Balke and Gilbert (2014) for an overview), since they represent the closest approximation of actual human behaviour. Candidates for this group are CLARION (Sun, 2003), ACT-R (ACT-R Research Group, 2015), and SOAR (Laird et al., 1987).

Contrasting this view, Balke and Gilbert (2014) remind us that a good model of human behaviour only captures the characteristics necessary for a given scenario of interest. As such the choice of an architecture, or even individualised composition of architectural components, is an intricate challenge and depends on the questions a modeller asks and the assumptions (s)he is willing to make.

To position our work in this discussion, let us borrow the example of second-order emergence as discussed in Subsection 2.1.2. Representing second-order emergence requires us to endow artificial entities with sufficient cognitive capabilities, so they can attain awareness of the emergent properties that can potentially influence their future behaviour (Gilbert, 1995; Squazzoni, 2008). The question is, however, how explicit this representation of emergent properties has to be. Does it a) afford a complete mental reconstruction of the emergence process itself, or b) is the representation of an outcome (e.g. the generated institution) sufficient as the case with EMIL-S (Lotzmann and Möhring, 2009), or c) is the mere access to simplified representations that enable social comparison (Festinger, 1954) (e.g. in the form of a simple metric) satisfactory?

This example shows that the selection of appropriate representations may not necessarily be addressed with a 'one size fits all' solution. This is particular relevant in the light of the limited extent to which cognitive architectures have considered social aspects, while social simulation has likewise made limited attempts to incorporate advanced cognitive concepts (Sun, 2007). However, though not primarily geared towards social interaction, specifically the philosophically-inspired BDI architecture has experienced considerable attention in the form of refined variants that put stronger emphasis on social influence. Examples for this include the BOID architecture (Dignum et al., 2000), in which agents consider obligation norms as part of their reasoning cycle. eBDI (Pereira et al., 2005; Jiang et al., 2007) considers affective components of human behaviour and consequently considers emotional responses to environmental stimuli as part of agents' reasoning. In contrast to the previous BDI variants that concentrate on the normative and affective domain, CoJACK (Evertsz et al., 2008) is an example of a BDI extension that emphasises the cognitive domain. It borrows cognitive parameters from ACT-R, which are controlled by a moderation layer that emulates the task-dependent physiological and affective constraints on memory access (and thus reasoning) that apply to human beings.

Observing the variety of extensions and context-dependent focus of different architectures, motivations to 'put the agent into agent-based modeling' (Wellman, 2015) are not intending to inspire the exclusive use of advanced rational or cognitive capabilities, such as BDI models or the above-mentioned neurologically inspired agent architectures. Instead they remind researchers to leverage complexity to a level that is relevant and appropriate for the scenario of interest (as is the case with most of the examples put forth by Wellman (2015)), such as the representation of social influence, be it by relationship networks, historical memory, or trust conceptions (elements that are of relevance in our work).

However, in the light of the diverse application scenarios of agent-based modelling, the agent representation appears secondary to the scenario representation itself, especially when involving larger numbers of agents. Consequently, features that are reflected in general-purpose agent architectures (such as the ones mentioned above), are introduced as intrinsic components of the specific model itself where necessary (such as learning or memory). This may – at least in part – be the reason why demands for unified execution environments have concentrated on the simulation platform itself (see e.g. Michel et al. (2009)), and did not extend to the presumption of a unified baseline agent architecture.

Instead modellers' primary emphasis lies on an approximate replication of the scenario of interest, generally involving iterative refinement and with the intent to explore scalability effects when involving large numbers of agents (an aspect we will explore in Subsection 4.2.2 of this work). Inasmuch as committing to a specific agent architecture and a potentially associated organisational meta-model (such as JaCaMo (Boissier et al., 2015)) promises high-level knowledge-centric modelling abstractions, it locks the modeller into the applied modelling paradigm, including the agent model and the associated environment representation.

Looking beyond the agent model itself, the infrastructural aspects are decisive for the choice of a runtime environment. The support for distributed processing, communication means, discovery services, and security, as found in general-purpose agent platforms such as JADE (Telecom Italia, 2015) and Cougaar (Raytheon BBN Technologies, 2015), is of limited use in the context of simulation. In fact the infrastructural weight can be incongruent to the lightweight agency notion applied in social simulations and consequently challenge the runtime performance (see (Frantz et al., 2010) for exemplary evaluations). More important, however, are the features that most of those platforms do not prioritise. Those include the fair scheduling of agents, central management of random number generation, a controlled environment (i.e. no interference by other platforms), and visualisation features in order to assure plausible and reproducible simulations.[21] Moreover, those platforms are not engineered with the intent to analyse macro-level phenomena, but focus on the accurate engineering of micro-level entities. An alternative group of agent platforms provides mechanisms to reflect the multi-level nature of systems, such as GAMA (Drogoul et al., 2013) and PADAWAN (Picault and Mathieu, 2011). However, such systems generally focus on the use with physical systems such as environment simulation (weather, flooding) that incorporate rule-based social entities without explicit cognitive features. The multiple system levels are explicitly known and engineered (potentially even within agents), and generally do not emerge from the agent interaction itself.

The combination of social simulation characteristics discussed in the preceding, namely the lightweight notion of agency with precedence of multitude (and associated macro-level effects) over micro-level features, scenario-specific representations, and essential platform features make platforms such as NetLogo (Tisue and Wilensky, 2004), Repast (North et al., 2013), and MASON (Luke et al., 2005) viable candidates, without preventing the use of cog-

[21] The varying feature sets of general-purpose MAS platforms and ABM-specific platforms are discussed in greater detail in (Frantz, 2010), specifically in Subsection 3.2.3.

nitive features[22], while allowing for the integration of those where necessary or purposeful.

This work is an example of employing cognitive concepts in a demand-driven scenario-dependent manner.[23] The introduced models capture aspects of social influence based on relationships and memory, but do not go as far as to allow reasoning about the agents' operations; in this work the use of ABM as a means of modelling is chosen to reduce and contextualise specific game-theoretical abstractions applied in earlier work (as presented in Chapter 4 onwards). The models attempt to offer a *refined institutional understanding* (for which we have evidence as shown in Chapter 3 onwards) but do *not* aim at providing a more accurate or complete model of the human per se.

However, even though not primarily geared towards the agent internals, the conceptual contributions of this work seek to narrow the gap between cognitive MAS and ABM by proposing higher-level institutional representations that are not tied to specific agent conceptions, and can be used to represent emergent as well as preimposed institutions. As such the introduced capabilities can be of practical use in both application domains, while mediating between simulated individual, group, or society and human observer.

Methodological Aspects In the effort to establish the scientific foundation for agent-based modelling, important methodological contributions include modelling guidelines (Gilbert and Troitzsch, 2005; Heath et al., 2009) as well as more comprehensive approaches that borrow from the field of software engineering (Ramanath and Gilbert, 2004). Further aspects include the challenges of verification and validation (Windrum et al., 2007; Moss, 2008; Klügl, 2008; Sargent, 2013). The problem of reproducibility (in the light of the what Michel et al. call the *engineering divergence phenomenon* (Michel et al., 2009)) is addressed by standardising model documentation based on the ODD protocol (Grimm et al., 2006, 2010). In the light of diverse simulation scenarios with specific agent conceptions the ODD protocol facilitates their documentation by structuring specifications into the components *Overview*, *Design concepts*, and *Details*, with seven further subcomponents for each component (including Purpose, State Variables/Scales, Process Specifications, Initialisation), intended to capture simulation specifications comprehensibly.

Potential application fields for ABM are manifold with research fields as diverse as anthropology (Heckbert, 2013), policy modelling and analysis (Matthews et al., 2007; Westerhoff and Franke, 2012), the wider area of economics (subsumed as Agent-Based Computational Economics (ACE) (Tesfatsion and Judd, 2006)) to simulations of emergency re-

[22]For example, though not equipped for direct communication per se, NetLogo offers an extension (Sakellariou et al., 2008) that introduces a BDI implementation as well as FIPA-compliant inter-agent communication.

[23]For an overview of the simulation environment refer to Appendix B.

sponse (Hawe et al., 2012). Dedicated venues for agent-based modelling include the Journal of Artificial Societies and Social Simulation (Gilbert, 2014) as well as the Multi-Agent-Based Simulation workshop series (Amblard et al., 2014).

For our work ABM represents the perfect conceptual match, since it allows us to represent the agency and structure characteristics introduced in Subsection 2.1.1, along with the processes that link micro- and macro-sociological layers. This argument for the suitability of the agent metaphor is supported by the inclusion of institutions, and norms in particular, given the existence of the dedicated subdiscipline of *Normative Multi-Agent Systems*.

2.3.2 Normative Multi-Agent Systems

Normative Multi-Agent Systems (NorMAS) represent a specialisation of Multi-Agent Systems that links the agent concept with social norms, offering a more realistic representation of social aspects when modelling target phenomena in artificial societies (Boella et al., 2007). Essential aspects of NorMAS include mechanisms that "... represent, communicate, distribute, detect, create, modify and enforce norms, and mechanisms to deliberate about norms and detect norm violation and fulfillment." (Boella et al., 2007)

Consequently, individual agents require some norm representation, potential to reason about and modify norms, learn norms from their social environment, transmit norms and enforcement violation behaviour, while having the ability to use norms to guide their behaviour (Hollander and Wu, 2011b). Interpreted as a specialisation of institutions (see Subsection 2.2.1), norms can act as shortcuts for an individual's reasoning and thereby increase the efficiency of its actions while facilitating cooperation, coordination, and social organisation (Epstein, 2001; Boella et al., 2007). Though generally followed, norms (or informal institutions in Subsection 2.2.1) can be violated without incurring legal enforcement, representing their essential difference from rules and constraints (Cialdini and Trost, 1998). A common understanding that underlies the norm concept in the NorMAS community is the necessity of enforcement, with sanctions representing an integral part of normative systems (Alchourròn and Bulygin, 1971; Boella et al., 2007).

The wide portfolio of work in the area of NorMAS has conventionally been structured by their association to stages of the normative lifecycle.[24] Savarimuthu and Cranefield (2011) propose four processes, including *norm creation*, *spreading*, *enforcement* and *emergence*. Hollander and Wu (2011b) offer a refined variant of the normative lifecycle that includes processes such as *norm internalisation* and *evolution*. Mahmoud et al.'s (2014) most re-

[24] Here we concentrate on lifecycle conceptualisations from the area of NorMAS, beyond the earliest approach offered by Finnemore and Sikkink (1998).

cent literature survey identifies the additional processes of *norm detection*, *assimilation* and *adoption*.

2.3.3 Reinforcement Learning

We provide a brief introduction into *Reinforcement Learning* (RL), since most of our experiments (in Subsections 4.2.1 and 4.2.2) and our approach to 'grow' normative understanding in the second part of the thesis (Chapter 5 onwards) builds on a behaviourist perspective. RL can be understood as a representation of *operant conditioning* (Thorndike, 1911; Honig and Staddon, 1977) in the context of behavioural sciences, in which the learner has the ability to influence events in addition to pure observation of events (such as Classical/Pavlovian conditioning (Pavlov, 1927)) (Sutton and Barto, 1998). For the purpose of our work it thus offers a purist approach to behavioural learning.

RL enables agents to learn from interactions with their environment in order to determine the optimal action choice for recurring environmental states. Its unsupervised nature and complete dependence on environmental characteristics makes it a prototypical representation of behaviourist learning. RL is commonly represented as a Markov decision process (Puterman, 1994), in which the action selection in a given state elicits a reward based on which the learner adjusts state-action selection (transition probabilities), and while successive states depend on the action selection in previous states, future state-action selection is independent of all previous state-action choices (Markov property). RL is a Markov Decision Process in which the transition probabilities are unknown a priori, i.e. the learner needs to infer those based on experience. The expected future rewards are expressed through value functions that accumulate and discount rewards for given state-action pairs (Sutton and Barto, 1998).

A particular challenge in this context is the trade-off between exploration and exploitation, i.e. the extent to which an agent continues to explore the state space in the search for highest possible rewards (exploration) at the price of losing out on expected rewards from the known state space (exploitation) (Kaelbling et al., 1996; Sutton and Barto, 1998). While settling on exploitation may incur high accumulated rewards in the short term, far-sighted learners can achieve greater accumulated rewards by shifting the bias towards exploration. This consideration depends on the size[25] and nature (static vs. changing) of the state space as well as available time steps.[26]

In the prototypical RL case the learner (agent) is placed in an environment and he per-

[25] Large state spaces may make it impossible or inefficient to explore the entire state-action space (*Curse of Dimensionality*), suggesting the use of approximation mechanisms for its reduction (Sutton and Barto, 1998).

[26] See Singh and Kearns (2002) for an efficient approach to address this trade-off.

ceives its state $s_t \in S$ at a given point in time t out of the set of possible S. Selecting an action $a_t \in A(s_t)$ out of all possible actions in that state $A(s_t)$, the agent receives a delayed reward r_{t+1} in the next time step $t+1$, along with updated state information s_{t+1} (Sutton and Barto, 1998). Figure 2.8 highlights this interaction.

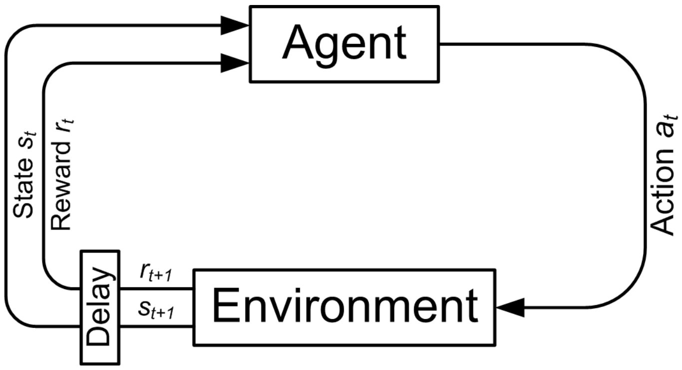


Figure adapted from Sutton and Barto (1998)

Figure 2.8: Agent-Environment Interface in Reinforcement Learning

Based on this interaction the agent adjusts its policy π_t, that is, the mapping of states to action selection probabilities for a given step t. For a more refined introduction and formal definition refer to Sutton and Barto (1998).

Q-Learning (Watkins, 1989; Watkins and Dayan, 1992) is a prototypical implementation of reinforcement learning, which we will use to describe essential reinforcement learning characteristics outlined before. At its essence, Q-Learning determines the *quality* of state-action combinations (Q-value) with respect to a given reward.

Apart from the choice of exploration vs. exploitation, generally represented by a static or changing parameter ε that decides whether learning actually takes place (as opposed to simply relying on previously learned information), Q-Learning uses three parameters:

- *Learning Rate* α – The learning rate, with $0 \leq \alpha \leq 1$, decides to which extent recently acquired feedback replaces previous information, thus indicating an agent's recency bias. A value of 0 would prevent the learner from internalising new information, while a value of 1 would imply the sole consideration of the latest information.

- *Discount Factor* γ – The discount factor, with $0 \leq \gamma \leq 1$, determines the extent to which the expected future rewards for state-action combinations are taken into account when calculating the Q-value for a chosen state-action combination. A value of 0 would prevent expected future values from being included in the calculation of Q-values for a given state-action pair, while a value of 1 would give expected future Q-values the same weight as the experienced reward for the performed action.

- *Initial Conditions* $Q(s_0, a_0)$ – The initially set Q-value is used for the initial value updates and, especially if set to optimistically high values (in addition to the parameterisation of ε), has influence on the learner's bias in the exploration vs. exploitation trade-off.

To contextualise the parameters, we discuss the central aspect of Q-Learning, the value update function, an annotated version of which is shown in Figure 2.9.

$$\underbrace{Q_{t+1}(s_t, a_t)}_{\text{updated Q-value}} = \underbrace{Q_t(s_t, a_t)}_{\text{old Q-value}} + \underbrace{\alpha_t(s_t, a_t)}_{\text{learning rate}} * [\underbrace{R_{t+1}}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} * \underbrace{\max_a Q_t(s_{t+1}, a)}_{\text{estimated Q-value for action choice in updated state}} - \underbrace{Q_t(s_t, a_t)}_{\text{old Q-value}}]$$

Figure 2.9: Q-Learning Value Update Function

Apart from an initial Q-value specified by the experimenter, learning involves the update of the existing (or the initial) Q-value ('old Q-value' in Figure 2.9). The learned information includes current feedback as well as an optimistic estimate for expected future rewards based on the updated state s_{t+1}. It is calculated based on the delayed reward for a chosen state-action combination R_{t+1} ('reward') and combined with the expected Q-value for the highest rewarding action in the next time step for the updated state s_{t+1} ('estimated Q-value for action choice in updated state'), discounted by γ (indicating the relative importance of estimated future reward), and reduced by the previous Q-value ('old Q-value') for the chosen state-action pair. Weighted by the learning rate (indicating the relative importance of learned information), the learned information is combined with the old Q-value in order to update the Q-value for the chosen state-action pair for action choices in future time steps ('updated Q-value').

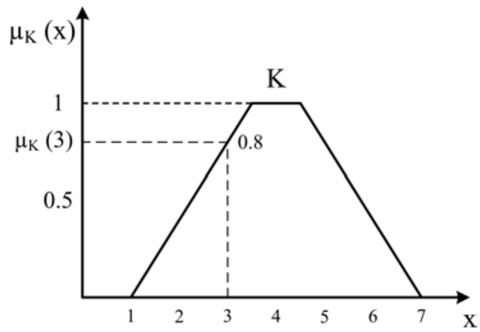
2.3.4 Fuzzy Sets

A final technique which we rely on in the second part of the thesis is the concept of Fuzzy Sets. Fuzzy Sets (FS) (Zadeh, 1965) offer a mechanism to operate with uncertain information. The need for a fuzzy representation to capture the complexity of social reality has been expressed in the *Incompatibility Principle* by Lotfi Zadeh, who first introduced the notion of Fuzzy Sets (Zadeh, 1965). The *Incompatibility Principle* (Zadeh, 1973) posits that increasing complexity is in inherent conflict with precision. This offers a central motivation to apply the FS concept in our work in order to structure fuzzy normative understanding (see Chapter 7).

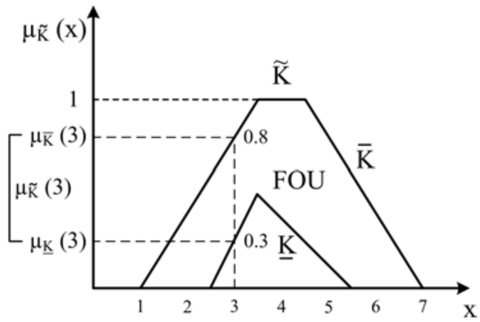
In contrast to classical set theory, where concepts are associated with sets using bivalent logic, this 'crisp' representation is relaxed and instead associations are described using *Membership functions* (MF) based on which a concept's or input value's *Degree of Membership* with a fuzzy set can be expressed.

An application of fuzzy sets that is central to Zadeh's motivation is *Computing with Words* (Zadeh, 1975a, 1996, 1999; Liu and Mendel, 2008b) (CW). CW seeks to represent the human capability to operate based on concepts without engaging in explicit computation. An example for CW involves the mapping of natural language terms such as 'small', 'tiny' and 'large' onto computational representations that reflect their relationship in a dimension such as 'size', which can then be used to draw inferences across one or more dimensions that can be translated back into natural language. Zadeh refers to this as the mapping of perception to measurement (Zadeh, 1999). The ambiguity of terms in natural language, limited precision and uncertainty across different users – the need to circumlocute terms with definitions in scientific papers displays this problem – makes FS a possible vehicle of representation.

We will explore this principle using a visual FS representation. In Figure 2.10a, the degree of membership (expressed as a value between 0 and 1) of the input value 3 with fuzzy set K, e.g. a natural language term, (for a given domain x) is $\mu_K(3) = 0.8$. An input can have membership with one or more fuzzy sets (within and across different dimensions) and have varying membership degrees, e.g. the varying and potentially overlapping understanding of the fuzzy sets 'small' and 'tiny', with 'tiny' generally perceived to be smaller than 'small'.



(a) Example of Type-1 Fuzzy Set



(b) Example of Type-2 Fuzzy Set

Figure adapted from Frantz et al. (2014d)

Figure 2.10: Type-1 vs. Type-2 Fuzzy Sets

However, a central limitation of the original Fuzzy Set concept, which we from hereon refer to as Type-1 Fuzzy Sets (T1FS), is that the degree of membership itself is a crisp value, such as 0.8 in the previous example. Paradoxically, the membership with a set is characterised by uncertainty, while T1FS are 'certain about this uncertainty'. This conflict has been well explored by Klir and Folger (1988), and the philosophical implications that challenge the scientific use of T1FS have been discussed by Mendel (2003, 2007a,b,c), Wu and Mendel (2009) as well John and Coupland (2007).[27]

As a consequence Zadeh introduced Higher-Order Fuzzy Sets, extending the original FS concept to Type-2 Fuzzy Sets (Zadeh, 1975a) (T2FS) (and more generally, Type-n Fuzzy Sets), which introduced second-order uncertainty, that allows one to express the degree of membership as a fuzzy value itself. The particular focus here is on *Interval Type-2 Fuzzy Sets* (IT2FS), as opposed to General Type-2 Fuzzy Sets that extend the fuzziness of the degree of membership into a third dimension. Instead of relying on a crisp membership function, IT2FS rely on an upper and a lower membership function that describe the degree of membership. Using the example in Figure 2.10b, the membership of input value with the fuzzy set $\widetilde{K}$ is thus described as an interval that captures the intersection with lower MF ($\underline{K}$) and upper MF ($\overline{K}$), resolving to $\mu_{\widetilde{K}}(3) = [0.3, 0.8]$ for the input value 3. The space between upper and lower MF is described as the *Footprint of Uncertainty* (FOU) and represents the extent of second-order fuzziness. As such, a T1FS can be considered a T2FS with identical upper and lower MF. As indicated before, since the upper and lower MFs of T2FS are themselves crisp the concept of higher-order fuzzy sets has been generalised to Type-n Fuzzy Sets (Zadeh, 1975a).

Since our interest is primarily in the generation of fuzzy sets for the purpose of analysis as opposed to decision-making, we omit a detailed introduction of fuzzy logic systems, and Interval Type-2 Fuzzy Logic Systems (IT2FLS) in particular. The foundations of T2FS as well IT2FLS are given by Zadeh (1975a,b); Mendel (2001); Mendel et al. (2006a); Mendel (2007a); Hagras (2007); Mendel (2007b,c); Liu and Mendel (2008b,a); Wu and Mendel (2014), along with work towards a standard for IT2FLS (Mendel et al., 2006b). Foundations along with an overview of the history of IT2FS have been provided by John and Coupland (2007).

Fuzzy sets, particularly T1FSs, have found adoption in diverse application areas. Those domains include fuzzy controllers (Long et al., 2014), data analysis (Höppner et al., 1999), production management (Mulaa et al., 2007), fire detection in battery compartments of hybrid cars (Dattathreya et al., 2012), and image processing (Acharya and Ray, 2005; Huang et al., 2010).[28] In the context of agent-based modelling and social simulation, T1FS have been applied to the modelling of personality traits (Ghasem-Aghaee and Ören, 2003; Ören and Ghasem-Aghaee, 2003), the representation of trust (Lesani and Montazeri, 2009) as well

[27] The limitations of T1FS have likewise been discussed in the context of applications (e.g. Hagras (2004, 2007), Mendel (2001), and Wu and Tan (2006)).

[28] For an overview of recent practical applications of fuzzy sets/logic, refer to Singh et al. (2013).

as a measure for social relationships (Sabeur and Denis, 2007; Hassan et al., 2007, 2011).

Recently, T2FS have received increasing attention with applications in the context of type-2 fuzzy logic controllers (Hagras, 2007; Cortes-Rios et al., 2014), robotics (Hagras, 2004), fuzzy clustering (Rhee, 2007), image processing (John et al., 2000; Huang et al., 2010; Choi and Rhee, 2009), video classification (Liang and Mendel, 2001), time-series forecasting (Karnik and Mendel, 1999), for the optimisation of inventory management and supply chains (Miller et al., 2011, 2012) in the context of logistics, the modelling of age-structured bird populations (Ramírez et al., 2011), and stock price prediction (Liu et al., 2012).[29]

2.4 Reflection on Relevance

In this chapter we have laid out factors that are relevant for the modelling of institutions and positioned our work in this context. We commenced with the discourse on agency and structure, with the key emphasis on the dynamic interaction between agency and structure without assigning clear precedence (see Subsection 2.1.1). For our work we assume Bourdieu's (1977) view on the strong interdependence of agency and structure, thus assuming institutions as structural representation embedded within agents. Though embedded in individuals, institutions may not be explicitly accessible by the individual (Parsons, 1951), and thus be absent from the individual's conscience. With respect to institutions as shared social concepts, the discussion of different levels of social organisation (micro-, meso-[30], and macro-level) in Subsection 2.1.2 structures the notion of institutions into different layers. It likewise highlights processes that interlink those levels (emergence and immergence), and which motivate the use of ABM as a suitable representation. In Subsection 2.1.3 we further borrow Ritzer's (1979; 1981) intersection of micro-macro stratification with the categories of subjectivity and objectivity, which provide the necessary backdrop for differentiating formal and informal institutions based on their subjectivity/objectivity (see Subsection 2.2.1), while furthermore laying out an integrated interpretation of institution formation processes in Subsection 2.2.2.

Turning towards the foundations for contributions of this work, in Subsection 2.2.3 we introduce different approaches to Institutional Analysis, such as Williamson's Four Layer Model (Williamson, 2000) that structures research fields related to institutions, and Ostrom's IAD Framework (Ostrom et al., 1994) that, based on its frequent invocation and refined

[29] A selection of recent IT2FS-related applications is provided by Sadeghian and Tahayori (2015).

[30] Without further exploration at this stage, with meso-level we mean intermediate levels of structural organisation, beyond the micro-level (e.g. individual interactions), but below the macro-level (society at large).

development, has established itself as the de-facto standard for IA. From the latter we will borrow integral elements, namely the Institutional Grammar (Crawford and Ostrom, 1995), as part of our own conceptual contributions in Chapter 5 onwards. A further approach we introduced is Bates et al.'s Analytic Narratives (1998), which their co-author Greif used for the analysis of the Maghribī Traders Coalition (Greif, 2006), the motivating scenario that accompanies us throughout this work. This aspect will be of particular focus in the upcoming chapters 3 (Scenario) and 4 (Greif's approach).

Providing the technological foundation, we introduced Agent-based Modelling and Simulation in Subsection 2.3.1 that offers a representational match of human societies and is applied throughout all experiments of this work. Relying on ABM and institutional analysis, in Chapter 4 we test specific institutional constellations that are in contrast to Greif's original scenario.

Chapter 5 shifts the focus towards conceptual modelling of informal institutions in the form of norms from a dynamic behavioural perspective, incorporating aspects of institutional formation and change (which we discussed in Subsection 2.2.2), in contrast to the dominating comparative-static perspective assumed in IA approaches introduced in Subsection 2.2.3. Though borrowing concepts from the area of IA (namely the Institutional Grammar (Crawford and Ostrom, 1995) in Chapter 5), the proposed concept offers a strong relation to the more specific area of Normative MAS (introduced in Subsection 2.3.2) by contributing a flexible institution representation structure with emphasis on a dynamic perspective (Chapter 6). Its operationalisation relies on Reinforcement Learning (introduced in Subsection 2.3.3) to reflect behavioural learning aspects.

In Chapter 7 we change the perspective from modelling of (informal) institutions to their analysis. We utilise Interval Type-2 Fuzzy Sets as a means of generalising the fuzzy normative understanding across different levels of analysis in agent societies, thus adding to the tool repository of IA.

Table 2.2 provides an overview of the relationship between chapters and the domain-specific and technological foundations introduced in this chapter.

With this overview, we shift our focus away from the general background and towards the introduction of the historical scenario that motivates our work.

Table 2.2: Overview of Core Chapters with their Relationship to Domain-Specific and Technological Foundations

Chapter	**Domain Foundations**			**Technological Foundations**			
	Institutions	Institutional Dynamics	Institutional Analysis	Agent-Based Modelling	Normative MAS	Reinforcement Learning	Fuzzy Sets
Chapter 3	●		○				
Chapter 4	●		●	●			
Chapter 5	●	●	○	●	●		
Chapter 6	●	●	○	●	●	●	
Chapter 7	●	●	●	●	●	●	●

- ● Strong relationship with chapter.
- ○ Moderate relationship with chapter.

3

Maghribī and Genoese Traders

Adapted from Frantz et al. (2014a)

Figure 3.1: Map of the Mediterranean Basin

A narrative that will accompany us throughout this work is what has been referred to as the *Maghribī Traders Coalition* (Greif, 1989, 2006) or *Maghribīs* in the context of comparative economics, and *Geniza Merchants* (Goldberg, 2012c) in the context of historical analysis. It describes a trader collective that performed long-distance trade along the North African coast (see Figure 3.1) and the Indian Ocean between the end of the 10th century to the 13th century. This collective's central feature was its cooperation based on informal means of normative enforcement, which is one central discussion point of this chapter. Their existence posed an interesting example of how (relatively) closed groups of individuals interconnected by mutual obligations could maintain cooperation in trade relationships that spanned across thousands of kilometres. Paralleling the Maghribīs, on the Northern side of the Mediterranean Basin (as shown in Figure 3.1) existed the growing Italian city states, such as Pisa, Venice and Genoa. Contrasting the Maghribīs' *informal* means of cooperation, the Italian traders relied on *formal* contractual enforcement of cooperation, providing suitable comparanda for institutional analysis from a comparative economics perspective as introduced by Greif (1989, 1993, 1994, 2006).

The core activity of both societies was the mediation of trade services across long distances. As schematically outlined in Figure 3.2, this generally involved a partner (Investor) that entrusted the sale of goods to a fellow confidant (Trade Delegate) that acted on behalf of the goods' owner outside his immediate reach. Upon realisation of the sale on the remote market, profits had to be truthfully returned to the investor who would then reward the fellow trader for his services.

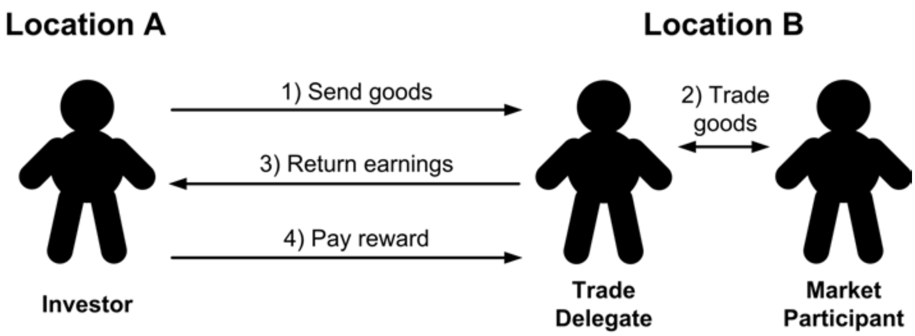


Figure 3.2: Schematic Overview of Long-Distance Trade Interactions

The essential challenge lay in the compliant conduct of such operations. How could investors assure that the seller of goods truthfully reported returns, so as to maintain long-distance trade as a profitable venture? This question can be approached by exploring the involved institutional instruments, an aspect that requires specific insights into the details of both trading societies. Previous explorations by Greif (1989, 1994, 2006) essentially ascribed the varying choice of institutional instruments to cultural traits, characterising Maghribīs as *collectivistic* and Italian traders as *individualistic*, respectively.

Though plausible on first sight, cultural traits represent a high-level abstraction that motivates a more refined exploration of the historical reality. It is for this reason that we cover both societies and their peculiarities in greater depth. Though inspiring this work, in the light of recent literature contributions Greif (2006)'s conception of the historical case has sparked a set of controversies, in particular with respect to the poorer documented Maghribī case. With particular focus on the latter, we will thus outline and discuss a selection of interpretational inconsistencies in order to develop a more comprehensive and consistent picture of the current state of research. Based on a meta-discussion of existing literature we assume our own position on selected issues of ongoing discourse. This discussion will highlight the historical happenstance and related research efforts in greater depth than necessary for the computational exploration pursued in the later chapters, but is necessary to offer a consistent understanding of the historical case and offer a basis to review some of Greif's assumptions. Even though not all of the discussed details find explicit introduction into agent-based models, the context offered by a comprehensive historical backdrop facilitates a foundation that avoids modelling decisions that are in conflict with historical reality, especially in cases where modelling is challenged by insufficient concrete information. As such this approach follows our spirit of utilising agent-based modelling as a means to explore specific hypotheses, as opposed to building more comprehensive models of high abstractions. Moreover, this synthesis of the current state of research on the Maghribī Traders Coalition from a comparatively unbiased position (from the viewpoint of an information scientist) and its contextualisation with Greif's original conception is a contribution in its own right and offers an accessible basis for further research.

The literature we consult on the Maghribīs can be broadly categorised into three groups, the first of which are the original Geniza studies (see Section 3.1), such as Goitein's seminal work, which is considered the most authoritative source of information on Maghribī traders (Goitein, 2000a), as well as Udovitch's (1962; 1967; 1970; 1977) interpretation and analysis of employed institution types. Their work is characterised by a very meticulous effort and detailed description to a level that led researchers such as Goldberg (2011) to attest them, and Goitein in particular, the ability to reconstruct the Maghribīs lives from letters. This stance may have been shared by Goitein himself, who considered himself a "sociographer" (Astren, 2012). For this categorisation, let us refer to them as *puristic field workers*, laying the ground for further research with the intent of sharing the utmost amount of detail while trying to avoid the development of preconceptions on the part of the reader or even attempting to apply any form of analytical methods (see e.g. Astren (2012)).

The second group, or rather individual, forms the *comparatist view*. Greif (1989, 1993, 1994, 2006) put the Maghribī traders into the spotlight of economic studies and applied the semi-formal method of Analytic Narratives (Bates et al., 1998) (see Subsection 2.2.3), that combines rational choice theory and game-theoretical analysis with historical accounts, to develop a crisp economic interpretation of the function of the institutional mechanisms employed by the Maghribīs. In his role as a comparative economist, Greif developed a sharp contrast to the Southern European Genoese traders that opted for contract-based formal institutional instruments to assure cooperation.

The third group of researchers considered in this overview include Goldberg (2005, 2011, 2012c,a), Ackerman-Lieberman (2007, 2012, 2014), and Edwards and Ogilvie (2012), many of which entered the arena of Geniza research in the past decade. Their works build on an extended corpus of commercial letters (see Goldberg (2005) and Ackerman-Lieberman (2007)), and, for the most part, challenge, revise and extend Greif's findings, and go as far as to modify Goitein's and Udovitch's interpretations of Geniza material. We will duly label them as *revisionists*. Note that this listing ignores the numerous additional studies that used the Geniza to further different subjects of study beyond the trader coalition, a selection of which is described by Cohen (2006).

This literature overview will be of particular relevance to the first part of this work in which we detach ourselves from an abstract comparatist perspective and explore questions specific to either specific historical trader society. Only for the second part of the thesis, we will step back and introduce modelling constructs and approaches that can capture a more general perspective on institutions with value for the sketched scenario and beyond. In doing so, we capture a broader analytical spectrum and leverage the power of agent-based modelling to navigate between the definition of comparatively fine-grained scenario specifications and the establishment of a more abstract comparative view without unnecessarily losing ourselves in scenario details.[1]

However, at this stage we first turn to the introduction of the historical background of the Maghribīs, before addressing the Genoese traders in Section 3.3.

3.1 Maghribī Traders

Before describing the Maghribī traders in more detail, we provide some background on the sources that describe this society to highlight the Maghribīs importance in the context of medieval Islamic Mediterranean trade, but also to emphasise the analytical challenges the information base poses.

[1]In this context Zerubavel's essay (Zerubavel, 1980) 'If Simmel were a fieldworker ...' describes the challenges sociologists face when immersed in their subjects of interest.

3.1.1 The Geniza

What we describe as Maghribī traders was in fact a very small subset of traders operating along the North African coast in the 10th century. However, it was their values and norms that led to disciplined implicit self-documentation which enabled Goitein (2000a) and other scholars to analyse their life in detail. The Maghribīs were Jews that operated in long-distance trade and intensively relied on letters for their communication. In alignment with strong beliefs in supernatural agency, letters generally contained references to God, both to reassure commonalities and sincerity (e.g. "... transaction which only a man like you was fit to carry through, may God reward you." (Goitein, 1973)) but also to activate compliance by referring to the omnipresence of God and divine intervention.[2] Jews were required to store documents that referred to God in what they called a *Geniza* (Hebrew for 'storage'), a special room in a synagogue (Greif, 1989). Despite the existence of multiple Genizas, the one of interest in the context of Mediterranean trade is the collection originally stored in the Geniza room of the Ben Ezra Synagogue in Fustat, now 'Old Cairo'. This Geniza was rediscovered in the 19th century and, with a century-long collection of important writings, is considered the most important information source on Medieval Mediterranean Islamic trading.[3] For this reason scholars often simply refer to it as 'the Geniza' or 'Cairo Geniza'.[4] The most extensive coverage, i.e. translation and interpretation, of the Geniza documents to date has been undertaken by Goitein (2000a) and Udovitch (1962, 1970). They represent the most authoritative accounts on the Geniza's interpretation, which is a cornerstone for contributions of many contemporary scholars, with Greif (1989, 1992, 1993, 1994, 2006) possibly being the most prominent figure to introduce the Maghribīs as a prime subject of study in the context of historical economics. More recently, Goldberg (2011, 2012c,a) advanced the topic by exploring controversial aspects of the interpretations and analyses by Goitein and Greif. Based on those sources, we will given an overview of the Maghribī traders, with a focus on the social and professional structures as well the institutional instruments at their disposal.

[2] An overview on societies that solved cooperation problems by invoking religious beliefs is given by Sosis (2005). Recent studies support the contemporary importance for the development of trust (Ruffle and Sosis, 2007), even when involving non-religious subjects (Ruffle and Sosis, 2010).

[3] Goldberg states: "Geniza merchant papers span the eleventh and twelfth centuries, and are the *only* substantial and coherent documentary records of extra-regional trade from the medieval Islamic Mediterranean" (Goldberg, 2012a).

[4] In fact Goldberg refers to the Maghribī traders as "Geniza merchants" (Goldberg, 2012a).

3.1.2 Social Structure

At the time of the Maghribī traders, North Africa was governed by the Muslim Fatimid rulers, whose empire extended from nowadays Morocco along the North African coastline up to Lebanon in the North and extending along the Red Sea in the South. Its capital was Fustat (nowadays known as 'Old Cairo'). Originally, the Maghribī traders were Jews that emigrated from the Baghdad-based Abbasid caliphate and joined Jewish trader communities located from the cities of Qayrawān (Kairouan) and al-Mahdiyya (Mahdia), located in the West (nowadays Tunisia) from a historical world view (see Figure 3.1). This led Goitein (2000a) to introduce the term "Maghrebis" (sic),[5] now more commonly referred to as Maghribīs, derived from the Arabic word for West (*al-Maghrib*) as a label. The Maghribī traders were *musta'ribun* – non-Muslims that adopted the values and customs of the Muslim society, including being a part of the 'nation' (*umma*) – and more so, they also remained an identifiable subset of the Jewish communities they were part of (Greif, 1994). A particular value they embraced, both being in a Muslim environment[6] as well as based on their own faith,[7] was to control and correct others' wrongdoing, thus sharing the emphasis on collective responsibility. Although the Maghribīs were integrated into the Muslim environment, they did not experience full equality compared to their fellow Muslim traders (Goldberg, 2011), which was reflected in the limited number of goods they were trading (40 out of 200 documented goods), and, most prominently, the prohibition to trade grain (Goldberg, 2011). Maghribīs generally traded within the boundaries of the Fatimid Empire and hardly operated on the Northern, European side of the Mediterranean Basin (with the exception of the Iberian peninsula, which for the most part was still under Muslim rule (*al-Andalus*), along with Sicily). This was so due to the harsh treatment of Jews in the European strongholds of Christianity, such as the anti-Jewish attitude that prevailed in cities such as Genoa (Epstein, 1996), along with rulings that introduced trade privileges for Southern European traders in

[5]In fact their labelling is an example for the interpretational differences among different historians. Greif adopts Goitein's term as Maghribīs, suggesting that merchants addressed themselves as such ("In their letters they refer to themselves as 'our people, the Maghribis, the travelers [traders]' ... " (Greif, 1989), similar in Greif (2006) and Greif (2012)), in which 'ashabana' (sic) is translated as 'our people' or 'coreligionists' (Greif, 1989); with reference to the term 'Maghribī traders' Goldberg (2012a) insists that "... merchants never refer to their group by this term ..." (Goldberg, 2012a), but instead used aṣḥābunā (translated as 'our associates') (Udovitch, 1977; Goldberg, 2012a). As mentioned above, in her work Goldberg utilises the neutral term 'Geniza merchants'. Originally inspired by Greif's comparative perspective and the widespread acceptance, we will use Maghribī traders, or simplified Maghribīs, throughout the remainder of this work.

[6]Words associated with Mohammed: "whoever sees a wrong, and is able to put it right with his hand, let him do so; if he can't, then with his tongue; if he can't, then with his heart, and that is the bare minimum of faith" (Cook, 2003).

[7]Greif (1989) expresses the compatible norm in Jewish society: "All Israel is responsible for every member."; see also Greif (2006).

remote trading ports at the price of banning Jewish trading activities.[8]

The outside pressures, both from their Fatimid environment as well as the hostile social environment of Southern Europe, leave us to suggest in line with Greif (1989, 2006) that their special identity, shaped both by being part of a religious minority in a Muslim environment as well as their common ancestry in Baghdad (in contrast to the existing members of the Jewish communities they had joined), was a central foundation of the mutual trust they exhibited when engaging in business transactions across the Fatimid Empire. Figure 3.3 offers a schematic representation of the social structure the Maghribī traders/Geniza merchants were embedded in. This common background allowed operation largely outside the unified but in many instances inefficient institutional environment. Those inefficiencies in law enforcement expressed themselves in the inability to track down individuals (Greif, 1989; Cohen, 2013) as well as the challenge to prove an individual's guilt, which generally depended on the existence of witnesses (Goldberg, 2012a). Further contributing factors included the at times incompatible legal interpretation,[9] let alone the time-consuming process of legal procedures, a claim supported by the documented involvement of heirs in the recovery of a deceased merchant's share (Greif, 1989). Nevertheless, on occasion Maghribīs engaged in formal business relationships with Muslim partners (Greif, 1989, 2012), despite the discrimination of Jews in Muslim courts (Goldberg, 2012a) and the ambiguous understanding of institutional instruments by Jews and Muslims (see Subsection 3.1.3).

The lacking adoption of conventional formal legal institutions, such as the commenda,[10] led Udovitch (1977) to suggest that business transactions based on individuals' ties did not have structural regularities. This aspect has subsequently been refined by Greif (1989), who adopted the understanding that individual relationships as well as their overall organisation were informal.[11] Based on the letters traders exchanged to share trade-related news Greif further identified the significant additional component of multilateral punishment for defecting trade partners (Greif, 1994, 2006). The essential form of defection was to misreport realised profits in remote trade locations or to perform trading to the disadvantage of the remote goods owner, such as prioritising one's own wares in sales. According to Greif, the defecting partner of such trust-based informal agreements would thus need to expect shunning by his

[8]Noteworthy in this context is the imperial decree Basil II granting Venetian trading privileges in Constantinople in 992 C.E., along with the commitment to exclude Jews from such activities (Linder, 1997). We extend this argument in a dedicated Appendix A under the theme 'Free Trade Debate'.

[9]Gil (1976) reports that for decisions made by Muslim courts the acknowledgement by Jewish courts could not necessarily be assumed; Cohen (2013) offers anecdotal evidence of explicit discouragement of litigants by Jewish authorities to take their legal matters to a Muslim court.

[10]Historians such as Harris (2009) consider Udovitch (1970) as authoritative with respect to the Arabic commenda.

[11]See e.g. Greif (1994, 2006).

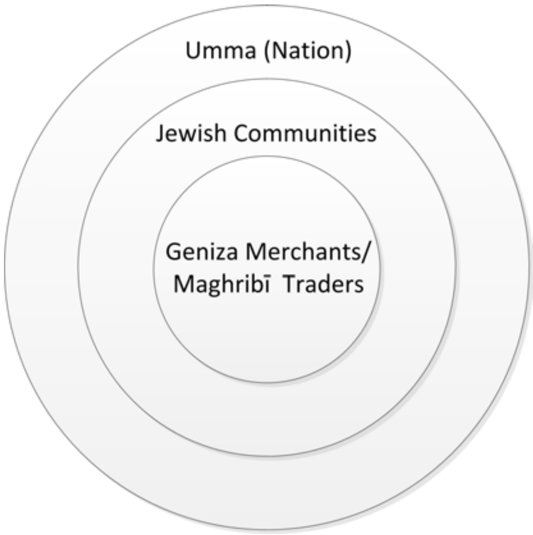


Figure 3.3: Social Structure of Fatimid Society

peers, the destruction of his reputation and being prevented from future employment within the coalition (Greif, 1989, 2006; Goldberg, 2012a), the commitment to which he considered grounded in the 'collectivistic' nature of the Maghribīs.[12] Using this ascribed collectivism as a basis, Greif (1994) considers wages, supposedly paid for their respective trade services,[13] to be lower compared to Southern European ones, since sufficiently high payment was not necessary to motivate cooperative behaviour – given the threat of permanent exclusion from the coalition for misconduct. This, in extension, would explain in part why Maghribīs did not engage in trade relationships with Southern European traders, who expected higher wages to ensure their compliance.

Thus, following Greif's view, instead of engaging in open contractual trade interactions, Maghribī traders were selective and largely relied on the employment of individuals from their 'in-group' to handle their remote trading activities in the context of long-distance trade. Moreover, even within this group, they were selective, making the selection dependent on recommendation and, ultimately, on another individual's jāh,[14] an aspect Goldberg (2012a) highlighted as an important regulatory mechanism. As such the relationship network and quality was not as unstructured as Udovitch's understanding (Udovitch, 1977) might have led one to believe; and though traders generally belonged to the social middle class (see Goitein (2000b), Gil (1997), Greif (1989), Goldberg (2012c)), the network did not have the structure of 'equals'. Connectedness among individuals largely varied depending on jāh, with hubs

[12]Recall the previous allusion to their collective social responsibility.

[13]This aspect is one of Greif's assumptions that will require further discussion at a later stage.

[14]In the widest sense, jāh can be interpreted as 'reputation', with a meaning covering aspects such as 'social rank', 'standing' or 'prestige' (Goldberg, 2012a).

built around central individuals.[15] The aggregate of an individual trader's relationships was understood as his aṣḥāb (his 'associates') (Goldberg, 2012a). The totality of all aṣḥābs were known as aṣḥābunā ('our associates') (Goldberg, 2012a). Building relationships with individuals having strong jāh not only included flexibility (such as speedy arrangement of trade), but in addition reduced monitoring cost. Individuals with limited influence had stronger incentives to engage in cheating, which led to the preferred choice of more established traders for business ventures, thus driving inequality of node interconnections.[16]

An important characteristic of this loosely coupled network was the trust invested into the value of relationships. In order to expect future benefit from one's network – effectively measured in the individual trader's jāh – it was fundamental to take care of one's connections. This 'trust in the system', or responsibility towards the fellow traders, was manifested by the obligation never to employ a known cheater for one's business, but also implied the continuous reporting about trade occurrences, even if not related to one's own business but observed in the common market places.[17] In consequence, unilateral reporting of cheaters, either directly or based on secondary accounts (e.g. passing on market place gossip), led to near immediate exclusion of that individual from trade interactions,[18] though rare instances of forgiveness and repayment have been documented (Greif, 1993).[19] Another mechanism to assure compliance of other traders was to maintain a notion of open books: revenue was not cleared on a per-transaction basis but only periodically, which gave considerable leverage to fellow aṣḥābunā members with the credible threat of rejecting the clearing of accounts as a potential reaction to cheating (see e.g. Greif (1989)).

Given that the identification of cheating was based on witnessing and gossip, traders were eager to preempt eventual suspicions of cheating, e.g. in cases where revenue expectations could not be met because of unexpected change of market prices at travel destinations. Proactively recovering or sharing the loss was a documented mechanism to avoid accusations of misdemeanour, given that one's jāh was at stake (see e.g. Greif (1989)). Consequently, the Maghribīs had very low cheating levels. From his (relatively small)[20] sample of 175 letters

[15]In her analysis Goldberg (2012c) emphasises two individuals in particular, namely Nahray Ibn Nissīm (who had more than 400 connections) and Yūsuf Ibn 'Awkal (who had the still considerable number of 150 connections), which shows the large variation even for well-connected traders.

[16]Goldberg (2012a) comes to the conclusion: "Geniza merchants seem to have preferred contracts in opposite proportion to the agent's natural incentive to provide good service."

[17]Goldberg (2012a) suggests that around 20 percent of non-formal business letter content, i.e. content beyond discussing issues concerning the writer's and receiver's business, was dedicated to such 'gossip'.

[18]See the trade conflict between Joseph and Samhun reported in Greif (1994), and the collective punishment of a Tunisian agent who failed to pay on time (Greif, 1989).

[19]This aspect is the essential core of Harbord's criticism (Harbord, 2006) and his subsequent extension of Greif's original game-theoretical model (see Chapter 4).

[20]More information on the different researchers' letter corpi is provided at a later stage (see Subsection

capturing 652 business ventures, Greif can only identify three cases of suspected cheating. The fact that 90 percent of all trade was performed via agents (Goldberg, 2012a) reinforces the claimed importance of reputation in that community.

To understand the high level of compliance and the individuals' interest to maintain their jāh, it is helpful to contextualise the established background with details on the institutions the Maghribīs employed. Given that the view on the employed institutional instruments is not undisputed and far from unified, we deem it important to provide an overview of the different available institutional instruments and their characteristics. In fact the Fatimid Empire's Islamic jurisprudence shaped an institutional environment whose instruments were roughly comparable with the formal institutions offered in the Southern European city states of Genoa and Venice (see Section 3.3).

3.1.3 Institutions

In this subsection we develop an overview of different institutional instruments available within the Fatimid Empire – and thus potentially at the disposal of Maghribī traders. Note that we do not describe all instruments in detail, but intend to provide an overview of the landscape of practically relevant institutional instruments in order to retrace their use by Maghribīs. Given that various institution types were comparable to the ones employed in medieval Europe, and Italy in particular, we further contextualise those where applicable.

The institutional environment offered a variety of different institution types related to trade partnerships.[21] The ones reflected in the Geniza can be broadly broken down into five relevant types. We will reflect individual instruments with respect to their formality, and extend the description with respect to their relative importance for the remainder of this work.[22]

- Sea loan – The *sea loan* was basically a regular loan given to afford seabound trade, the borrowed sum of which had to be paid back upon return, along with a fixed interest rate. As far as the Maghribīs are concerned, this mode of financing trade was hardly used; the Geniza only documents two cases in which sea loans were taken (Goitein, 2000b).

3.2.2).

[21] A central difference to European institutional instruments was the fact that a notion of 'corporation' is not recognised in the Islamic law (Schacht, 1983).

[22] Instead of utilising the original Arabic or Hebrew terms to describe the institutions, we borrow functional conceptual terms used by Greif (1989) to capture the conceptually overlapping specific terms used for the respective institutions.

- Factor Relationship – Using the concept of a factor (representant), individual traders offered a full representation for other traders, including the handling of sales – possibly compensating the services by commission (Greif, 1989), and going as far as to represent each other in front of trade courts. The factor often operated for multiple remote traders at the same time.

- Commenda – The *commenda*,[23] in analogy to the Italian mechanism, consisted of an investing inactive ('sleeping') member (in the Italian version known 'commendator' or 'stans' (Harris, 2009)) and an active partner ('tractator', 'socius portat' (carrying partner) (van Doosselaere, 2009) or 'tractor' (Harris, 2009)) that handled the trade operations, but did not invest himself and did not bear any risk associated with losses. The distribution of profit was fixed at the time of agreement. The Islamic terms for different variants include muḍāraba, qirāḍ algoyīm, and muqāraḍa (Schacht, 1983; Udovitch, 1967; Greif, 1989), all of which shared those characteristics. The concept offered variations[24] of distributions and types of investment (e.g. different proportions of involvement; labour vs. capital investment) (Udovitch, 1967), an aspect which found its rough equivalent in the European-style commenda that had forms of bilateral or unilateral commendas, with proportional payoffs being generally allocated in a 0.5/0.5 (bilateral commenda) or 0.75/0.25 proportions, along with the rare choice of $0.\overline{6}/0.\overline{3}$ distributions (unilateral commenda) between investing and travelling party (Harris, 2009; van Doosselaere, 2009). However, in contrast to the fixed distributions that did not accommodate different market conditions, the Islamic variants offered far wider variation of return constellations that could be adapted to market constellations, the most flexible being the qirāḍ algoyīm (Udovitch, 1970; Ackerman-Lieberman, 2014). However, legally any investment share must have been of monetary nature (Udovitch, 1970), not in material fashion, as possible with the in this respect more flexible Western commenda conceptualisation.[25] This minor but important difference led van Doosselaere (2009) to suggest that the ability to contribute materially (as opposed to only financially) gave rise to the extensive utilisation of the 'commenda' by Genoese. This was attractive for investors that were not generally involved with trade activities, or for one-off investors (e.g. churchmen) that did not have direct access to monetary funds but were attracted by the profitability of sea trade. The fixed pay-offs further allowed one-off non-professional and small-scale investors to invest at a calculable risk, while

[23] Though describing institutional instruments used in the Fatimid Empire, the large variety of specialisations is best subsumed using the Italian term since it shares the general characteristics.

[24] For an overview of different variants refer to Raymond and Lopez (2001).

[25] However, literature indicates deviating behaviour in practice (Goitein, 2000a).

making significant effectively guaranteed profits.[26] For the perspective of the travelling party, the commenda was attractive, as it removed all liability from his part, which made one-off travelling endeavours attractive for non-specialists, i.e. individuals whose skill set was not immediately associated with that of a ‘trader’ or ‘sailor’, but generally had practical manual character, such as artisans, gardener, etc. (van Doosselaere, 2009). This offers insight into the socio-economic stratification existing in the Genoese society, with tractators occupying the lower end of the wealth distribution – or as Byrne put it, they were “not men of great wealth or of high position” (Byrne, 1916).

While knowledge about the commenda’s origin is imprecise, historians allocate it in the Eastern Mediterranean[27] or Arabia.[28] Undisputed, however, appears to be the origin as a customary commercial principle, such as in pre-Islamic caravan trade (Udovitch, 1970), as opposed to being a generic legal concept (Udovitch, 1962, 1970; Pryor, 1977; van Doosselaere, 2009).[29] This view is supported by its wide spread across territorial boundaries, including the Indian subcontinent and Western Europe (Harris, 2009). However, its usefulness inherently relied on the backing by a legal system that would enforce sanctioning of contractual deviations (Edwards and Ogilvie, 2008).

To complicate matters further, besides the European and Arabic incarnation of the commenda, the Jewish commenda concept (Hebrew: ‘ʿeseq’, Aramaic: ‘ʿisqa’), though seemingly compatible, bore a significant difference to the qirāḍ algoyīm in that both commendator and tractator carried the risk of misfortune: part of the investment endowed to travelling merchants was interpreted as a loan that had to be recovered by the sales of goods in the target market, or paid by the travelling partner in case of incurring losses (Goldberg, 2012c; Ackerman-Lieberman, 2012).[30] The Arabic variant of the commenda thus appeared more flexible, but also allowed economic opportunism as

[26] van Doosselaere’s (2009) findings suggest that the profit for sea trade arrangements was between 20 to 110 percent.

[27] Here Pryor’s (1977) reference to the Byzantine ‘chreokoinonia’ and Jewish ‘ʿisqa’ , or ‘ʿeseq’ (explored in 12th book (‘Acquisitions’) of Maimonides’ Mishneh Torah) is essential.

[28] Most notably Udovitch (1962, 1970) supports this understanding (see also Harris (2009)).

[29] The striking similarity between Islamic law and the relation to observed unified but geographically spread trade practices/customs (“... in ... Ḥanafī commercial law ... the shades and hues of customary practice were more prominent, and determined the coloring of the law to a much large extent than has hitherto been thought. The Islamic modification ... of this material is sporadic, uneven, and in many cases, minimal.” (Udovitch, 1970)) led Udovitch (1970) to characterise Ḥanafī law – the most elaborate medieval Islamic legal framework – as the Islamic version of the European ‘law merchant’ (Milgrom et al., 1990) (see also Ackerman-Lieberman (2012, 2014)).

[30] Furthermore, the Jewish ‘ʿeseq’ did not require a written agreement; oral agreement was the dominant modus operandi (Ackerman-Lieberman, 2012).

it permitted the concentration of risk and profit on one partner (Ackerman-Lieberman, 2012). However, this presumed flexibility was at the expense of higher transaction costs: its generally written nature implied the fixation of various contextual aspects, such as traded commodities, which impaired the traders' flexibility to make ad hoc decisions to seize market opportunities (such as a spontaneous change of carried, sold, or bought goods) (Ackerman-Lieberman, 2012). The Maghribīs generally preferred more equal distributions of profit, even if investment proportions varied and both partners benefited from an opportunistic decision on the part of one trader.[31] This lowered the attractiveness of the legally-backed commenda for Maghribīs in an environment that fostered an ambiguous legal interpretation of this instrument, and whose explicit (i.e. written) nature was generally conceived as deterring (Goldberg, 2012a).

- Active Partnership – The *active partnership*,[32] known as 'shirka' (or as 'khulṭa' ('mixing') as a non-legal term (Goldberg, 2012a)) in Arabic,[33] or 'shuthafuth' in Hebrew, is a temporally limited venture partnership that allowed two or more partners to contribute in different shares, but share the authority over the common funds (Goitein, 2000b). Venture partners received profits and bore the risks in proportion to their investments. As a central contrast to the (Islamic) commenda, loss was thus born by all partners, as opposed to concentrating the risk on the investing partner. A second aspect was the multilateral nature of the agreement; partnerships could exist among more than two partners.

  A special aspect of the partnership was its existence in written and unwritten form. While the shirka implied a written agreement, the mu'āmala (Ackerman-Lieberman, 2007) did not require this.

- Formal Friendship – The last institution type that demands specific attention is the *formal friendship*, named 'sadaqa' ('friendship'), 'bida'a' ('goods') or 'ṣuḥba' ('association', 'companionship') (Goitein, 2000b; Udovitch, 1970; Goldberg, 2012a). It was a strictly 1:1 relationship that was based on agreements and as formal ritualised component, required a face-to-face meeting for its enactment. It did not entail a written component and its purpose lay in the handling of traders' respective dealings in differ-

[31] See Ackerman-Lieberman (2012) for a detailed analysis of profit-sharing impacts of opportunistic trading by individual partners.

[32] Previous work (Greif, 1989) characterises this institution type as a 'partnership'. However, to contrast it to the commenda, which represents a special case of a partnership, the 'sleeping-active partnership', we utilise the term 'active partnership'. Goldberg uses the term 'joint-active partnership' to describe shirka relationships.

[33] Further Arabic terms include 'kis wahid' ('one purse'), 'bayana' ('between us'), or 'lilwasat' ('in the midst') (Goitein, 2000b).

ent trading locations (Greif, 1989; Goldberg, 2012c,a). A mandatory component was the acceptance and storage of incoming goods for partners, even if the receiving partner chose not to follow instructions regarding the handling of those goods (such as selling them). This occurred *without* remuneration, i.e. no commission was paid (Goitein, 2000b), as a payment for such services (‘*khidām*’) among ‘equals’ (as trader considered themselves) would have been considered slavery-like (Goldberg, 2012c). This ‘friendship’ could be formally terminated unilaterally (under witnesses), but until then each participant was expected to provide mutual trade (or at least goods-handling) services, some of whose relationships lasted for a trader’s lifetime (Goitein, 2000b). With reciprocity at the core of this institution (Goldberg, 2012c) came the obligation to handle each others’ goods, but not to an unlimited extent as originally interpreted by Goitein (2000b) and Udovitch (1977). In fact, the aspect of maintaining a balance of services was important, including the weighing of services with the relative status difference of the partnering traders (Goldberg, 2012c).

Figure 3.4 provides a schematic overview of the different institution types and the discussed respective specialisations. Since the term ‘commenda’ (or ‘sleeping-active partnership’) and ‘active partnership’ are not used consistently in literature and occasionally referred to partnerships (see e.g. Greif (1989); Goldberg (2012a)), we group those accordingly.

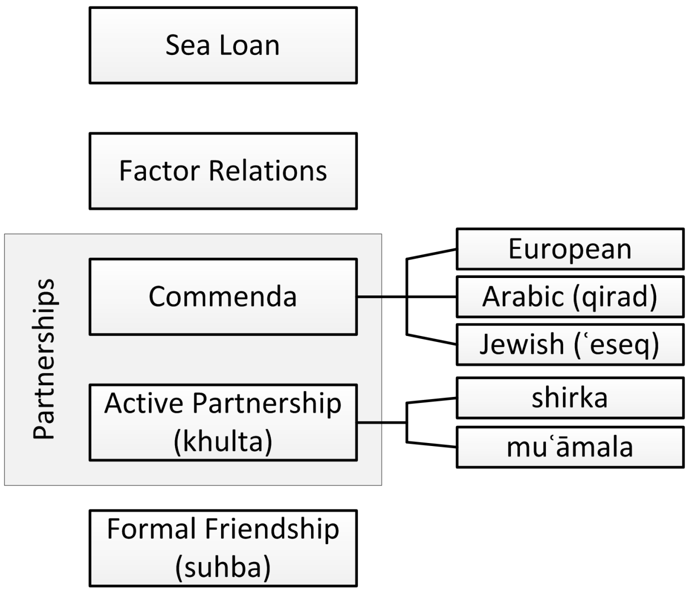


Figure 3.4: Overview of Institutional Instruments known by the Maghribī Traders

This overview provides us with a basis to retrace the usage of different institution types adopted by Maghribī traders. To develop a picture of the distribution of adopted institutions for commercial affairs, the possibly most authoritative source is Goldberg (2005, 2011, 2012c,a,b), since she performed the yet most extensive statistical analysis of the Geniza letters, enabled by its ongoing digitisation and translation efforts Goldberg participated in.[34] Based on her results she identified that more than 90 percent of all agency relationships were facilitated via fellow traders (Goldberg, 2012a), which confirms the importance of the relationship network the Maghribīs entertained. Further, around 93 percent of all relationships relied on institutions falling into the three categories commenda (qirāḍ algoyīm), active partnership (khulṭa)[35] and formal friendship (ṣuḥba), with the balance of relationships being driven by mostly informal intra-family arrangements (Goldberg, 2012a).[36] Goldberg's sample analyses reveal that the ṣuḥba was the by far most important relationship type, capturing around 67 percent of all agency relationships (Goldberg, 2012a). Compared to Goldberg, in his original studies Greif is comparatively imprecise about the actual distributions. However, his essential claim is that about half (Greif, 1989, 2008), or at least "[m]any ..." (Greif, 1989, 2006) agency relations did not have legal contracts.

Goldberg's most recent work (Goldberg, 2012c,a,b)[37] allows the most refined distribution of institutional instruments by Maghribīs as shown in Figure 3.5.

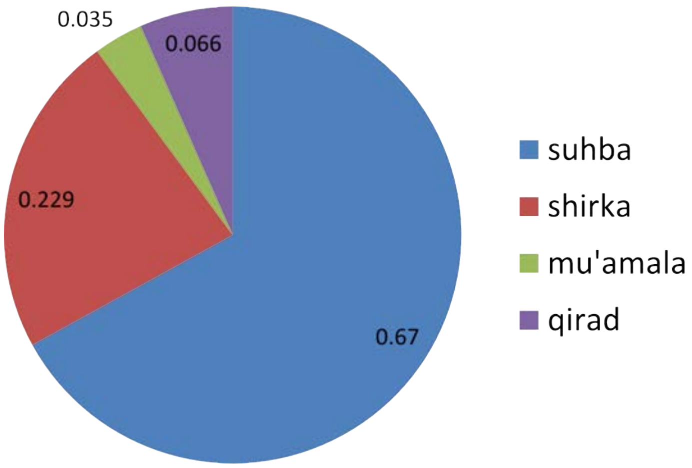


Figure 3.5: Distribution of Institutional Instruments used by Maghribī Traders based on Goldberg

Based on this distribution we can thus see the disproportional importance of the ṣuḥba (formal friendship) in comparison to partnerships (shirka and mu'āmala) and commenda-style

[34]A wide range of Geniza letters, though only in part translated, as well as the indices covering the various letter collections spread across different libraries, are available as part of the Princeton Geniza Project (2014).

[35]Maghribīs had a great preference for khulṭa relationships over qirāḍ algoyīms; the former were used in around 80 percent of all partnerships (Goldberg, 2012a).

[36]In her analysis, Goldberg reveals that the term khulṭa appeared six times more often than the legal term shirka; the frequency of khulṭa was further ten times higher than qirāḍ algoyīm and muḍāraba, indicating the relative importance of partnerships over commenda-related agreements (Goldberg, 2012a).

[37]Her earlier analysis arrived at a 0.75/0.25 distribution across formal friendships and partnerships (Goldberg, 2005).

(qirāḍ algoyīm) arrangements. In contrast to those findings, Greif attributed relevance to factor relationships in addition to “partnerships and friendships” (Greif, 1989). However, as far as Goldberg is concerned, both sea loan and factor relations were practically irrelevant.

3.2 Selected Literature Debates

3.2.1 Formality/Informality Debate

Neglecting accuracy of numbers, this finding challenges Greif’s suggestion (2006) that trade relationships were dominantly organised by informal means, a view that Goldberg suggests to be adopted from Udovitch’s insufficient description of institutional mechanisms employed by Maghribīs.[38] This is retraceable in the sharp contrast Greif draws based on the differing institutional characteristics of his comparanda, the individualistic Southern European city states with the collectivistic Maghribī traders (Greif, 1993, 1994, 2006). This interpretation carries through to the conclusions Greif (1989) draws, as he attests to the Maghribīs’ dominating use of “... partnerships, friendships, and factor relations.” (Greif, 1989), while suggesting that commenda relationships (read qirāḍ algoyīm) were hardly used. The latter aspects show considerable incompatibility with Goldberg’s results, possibly rooted in a mix of interpretative bias carried over from Goitein and Udovitch as well as the reduced letter sample at Greif’s disposal. A compatible conclusion both drawn by Greif and Goldberg is the limited relevance of family relationships to facilitate trade, and if existing, their mostly informal nature.[39] Although Greif was aware of the ṣuḥba and acknowledged its frequent use,[40] he seemed to have deemed it incompatible with his view on the trust-based informal relationship as he ascribes the ṣuḥba only weak, if any, reciprocity characteristics,[41] rendering it incompatible with reciprocity mechanisms the proposed informal enforcement implied. This appears to be surprising given that already Goitein considered the ṣuḥba the most common “informal business cooperation” (Goitein, 2000b) and even outlined the reciprocity characteristics.[42]

[38] Udovitch (1977): “The Italian merchant lived and breathed in a world of contract, of partnerships, agencies, commissions, and loans ... In the world of our eleventh- and twelfth-century Geniza traders this situation was reversed: informal ties were central, and formal ties, while important, were peripheral”.

[39] Goldberg (2012a): around 7 percent; Greif (2006): less than 12 percent.

[40] Greif (2006): “[The Maghribīs] mainly used partnership and ‘formal friendship.’ ”

[41] Greif (1989): “This exchange was not based on emotions, nor was it a reciprocal exchange; rather it was purely a business matter.”, albeit relativised by his concentration on an interpretation of strong reciprocity (see Greif (1989)).

[42] Goitein (2000b): “The nature of informal business cooperation ... is not easy to grasp. ... it seemed strange that a merchant should invest so much time and work in the mere expectation that his efforts would be properly reciprocated ...”.

Based on the literature evidence by Goitein, and Goldberg in particular, it is realistic to assume the ṣuḥba as the dominating institution type adopted by Maghribīs. However, this does not fully explain that Greif insists on informal relationships and consequentially dismisses the 'formal friendship' (ṣuḥba) as the institutional instrument of central relevance. Besides the different letter sample, a potential reason for this could indeed lie in the differing interpretation of formal vs. informal institutions as suggested by Goldberg (2012c). In her view the formality of institutions is characterised by obligations and specified rituals that lead to their formation and termination, which in the case of the ṣuḥba is reflected in the requirement to meet face-to-face and the unilateral termination under witnesses for this otherwise unwritten institution (Goldberg, 2012c). In contrast, Goldberg suggests that Goitein's, Udovitch's, and consequently Greif's views oversimplify their categorisations of the different institutional instruments and construe only written contracts as formal (Goldberg, 2012c). This suggestion appears retraceable and would restore the categorial view which would have allowed the characterisation of ṣuḥba relationships as informal. However, another conjecture is equally possible. Given that Goitein, and Udovitch in particular, were aware of the ways in which commercial laws were shaped by practice (Udovitch, 1970) (as opposed to an ex-ante specification of laws), institutional concepts such as the ṣuḥba appeared sui generis in that they were neither grounded in nor fully compatible with Jewish or Islamic law. Both legal frameworks did not support the reciprocal labour obligations, which were central to constituting the ṣuḥba (Goldberg, 2012c).[43] While property rights were enforceable (in the ṣuḥba the delegating party retained ownership over the delegated goods at all times (Goldberg, 2012c)), the service aspect, other than the violation of instructions on how to handle the goods, could *not* be regulated legally. In their work Goitein and Udovitch may thus not necessarily imply the informal nature as 'unwritten', but instead indicate the limited ability to *legally regulate* the underlying institutional construct. For example, when referring to the mu'āmala, Goitein (2000b) writes: "... mu'āmala ... would be used to describe a relationship as informal, not based on a legal instrument." He does not refer to the mu'āmala's unwritten nature (which was its outstanding feature compared to all other partnership types (see Subsection 3.1.3)) but highlights its *absence of legal backing*.[44] Knowing about the origin of Islamic and Jewish commercial law in often unwritten trade practices as well as the discussion of cases based on

[43]Ackerman-Lieberman (2014) supports the grounding of institutions in professional practice, but emphasises the *mediating* over the *decisional* role of Jewish courts, guiding (but not forcing) Jewish traders to follow and align their behaviour (and thus institutions) with established norms. If traders opted for a formal (here: legal) relationship, they generally opted for the Jewish variant (ʿeseq) and preferred oral agreements (Ackerman-Lieberman, 2012).

[44]A further more general example of Goitein, which Goldberg also refers to herself (Goldberg, 2012c): "... the Mediterranean trade, as revealed by the Cairo Geniza, was largely based, not upon cash benefits or legal guarantees, but on ... mutual trust and friendship." (Goitein, 2000b)

unwritten partnerships in front of courts, we suggest that Goitein's and Greif's interpretation of 'formal' as 'written' may not be entirely accurate, but rather supports their interpretation of 'formal' as 'legally enforceable'. In fact even the unanimously assumed formal institution with the most comprehensive legal backing, the qirāḍ algoyīm, was, when instantiated in its Jewish form (ʿeseq), mostly agreed upon in an oral fashion (Ackerman-Lieberman, 2012). In the context of the institution of interest, the ṣuḥba, the central aspects of the institution are not the property rights (which could be easier regulated through other institutions) or the sent instructions (which were enforceable (Goldberg, 2012c)), but the reciprocal obligations that arose from committing to such a relationship, i.e. the service-for-service principle, which could *not* be subject to (legal) enforcement (Goldberg, 2012c). Greif's work (from his perspective as an economic comparatist) followed the objective of modelling and rationalising the multilateral *enforcement* mechanism he saw in the Geniza letters.[45] The lacking evidence that the collective enforcement phenomenon was supported as a result of legal proceedings, despite the existence and his acknowledgement of a variety of legal partnership constructs (see Subsection 3.1.3; Greif (1989)), makes the equation 'formal = legally backed' appear more appropriate to describe the bifurcation Greif employs when ascribing different institution types formal or informal nature. Goldberg's understanding of formality is concerned with regularities regarding the handling of an agreement itself (such as formation and termination) as well as the importance to its stakeholders, here the Maghribī traders,[46] but not necessarily the legal system, even though some peripheral aspects could be handled by courts (such as violations of instructions in an agency relationship as well as property rights). Thus according to the conceptualisation of 'formal = legally backed', the enforcement mechanism for the unique and essential characteristic of the ṣuḥba, the reciprocal provision of services, is *informal*. Doing so, we can harmonise the differing instrumental understanding by Greif and Goldberg by characterising the ṣuḥba as the essential instrument that Greif failed to identify unambiguously, despite carrying the essential characteristics of absent legal enforcement and reciprocity (potentially supported by its misleading translation as 'formal friendship'), while Goldberg ascribed it strong formal characteristics based on the rituals involved in activation and termination, underemphasising the lacking legal enforceability of its essential characteristic – the reciprocal service obligation.

More than arguing for a particular definition and advocating a particular understanding,

[45] Greif (2012): "My work concentrated on enforcement per se and thus after verifying the limited reliance on court enforcement, I focused on reputation."

[46] Goldberg (2012c) : "The ṣuḥba was a system understood *by its participants* as formal and binding, but only one portion of the contract was binding in the legal system." (emphasis added). Similar in Goldberg (2012a): "Merchants understood the suhba as serious undertaking; it was begun and ended formally, often through the taking of oaths – but there was no contract under law."

it appears that the interpretation of formal vs. informal reaches beyond the discrete specification as either 'written' or 'unwritten', but is grounded in the perspective one assumes for institutional analysis, whether looking at structural aspects and regularities, such as of interest to the historian Goldberg, or inspecting it from a legalistic perspective with focus on the specification of enforcement mechanisms, as of central importance for the comparatist (or 'economic historian' as referred to by Goldberg (2012c)) Greif.

Table 3.1 provides a summarising overview over the essential institutions discussed in this and previous subsections, and differentiates those with respect to their written or unwritten nature as well as legal enforceability, as opposed to their 'formality'.

Table 3.1: Formality Characteristics of Maghribī Institutions

Institution	Form	Regulated Aspects	Legal Backing
qirāḍ algoyīm	written[a]	property ownership, partners' activities, redistribution of profit loss	yes
khulṭa	written and unwritten[b]	proportional contribution, redistribution of profit and loss	not a legal concept in itself[b]
mu'āmala	unwritten	customised, but not necessarily enforceable[c]	yes[c]
shirka	written	proportional contribution, redistribution of profit and loss	yes
ṣuḥba	unwritten	reciprocal service obligations, conditions for formation and termination	some[d]

[a] Often unwritten if Jewish variant ʿeseq is used.
[b] See respective legal variant (shirka/mu'āmala).
[c] In the absence of a written contract, courts assumed a standard partnership contract.
[d] Legal support as far as property ownership and goods handling instructions are concerned. No legal concept for sanctioning inaction or enforcing reciprocity.

3.2.2 Relationship Network Structure of Maghribī Traders

The discussion about formality or non-formality in the previous subsection may seemingly provide us with unnecessary level of detail. However, if we want to step beyond the boundaries of equilibrium-based institutional modelling (see Subsection 2.2.1) it is important to identify the concrete institutional instrument correctly in order to permit the modelling of properties beyond the rough conception of informal institutions Greif relied on.

However, more important than providing a refined model of the historical reality are the implications for the interpretation of the functioning of the Maghribī coalition at large, specifically the enforcement characteristics. In this subsection we thus identify how far the network structure that relied on the underlying institutional instrument could have supported enforcement characteristics, without relying on Greif's strong assumption of individuals' near selfless devotion to the coalition (Ackerman-Lieberman, 2012).

We have so far identified the ṣuḥba as the most relevant institutional instrument which satisfies Greif's understanding of private-order enforcement and the essential reciprocity characteristics he originally ascribed cultural characteristics. For the institution model he put forth Greif held the assumption that private-order enforcement, i.e. enforcement based on personal relationships, only functions within groups that are a) of limited size, b) have limited fluctuation, i.e. are closed groups in the ideal case, and c) have aligned interests, i.e. shared concern or kinship ties (see e.g. Cooter and Landa (1984); Grofman and Landa (1983); Boyd (1988); Kaplan and Gurven (2005); Greif (1994); North et al. (2009)). Limiting group size maintains the individuals' ability to know each other and perceive relationships as *personal*, thus sharing enough knowledge about each other to develop and maintain a notion of trust (North et al., 2009). This stands in opposition to *impersonal* relationships, commonly conceived as characteristic for formal (read: legal) institutions, in which enforcement authority is assigned to a third party in order to ensure a) the enforcement in cases in which the cost of cooperation (or enforcing cooperation) outweighs the rational selfish individual's marginal benefit from cooperation (Tragedy of the Commons (Hardin, 1968; Ostrom, 1990)), or b) the fair application of justice by imposing a fair and, in best case, impartial view that limits the litigants' involvement and prevents disproportionate or even repressive sanctioning (e.g. based on self-justice) (Rawls, 2005; Barry, 1995; Durkheim, 1933).

From his comparatist perspective Greif (1994) characterises the Maghribīs as a collectivistic society, to which he ascribes a 'segregated' society pattern in which the group members' interactions adhere to members of that same group (in-group) based on binding features such as ethnic, religious or familial identity (Greif, 1994), with individuals rarely transgressing group boundaries. In this conception the society thus consists of sharply defined but largely isolated sub-groups. In opposition to that exist individualistic societies, in this case the Genoese society to be described in Section 3.3, in which members interact and form relationships across still existing, though relatively weaker, group boundaries.[47]

In his work, Greif conceives the Maghribīs as a collectivistic closed group, unified by

[47] Conceptually the model Greif plots is compatible with Simmel's concept of social circles (Simmel, 1964), characterising societies by the extent to which different social relationships ('circles') overlap and shape group boundaries.

their ethnic background and beliefs, operating in an Islamic environment, but determined to only maintain trader relationships within their segregated group, avoiding relationships with out-group members such as Muslim traders where possible (Greif, 1994). In consequence, Greif construes a social model that is characterised by homogeneity and closedness. In his game-theoretical analysis he uses the assumption of collectivistic beliefs to model the preference for in-group traders as the shared collectivistic beliefs reduce uncertainty of enforcement. Uncertainty is represented by the concept of wages, which – given the dominating, though not exclusive, use of the remuneration-free ṣuḥba – is an example of Greif's inaccurate representation due to the lack of supporting evidence in the literature. However, as an abstract representation of cost it is an admissible utilitarian operationalisation of uncertainty. Notwithstanding, the reduced uncertainty in collectivistic societies (thanks to their shared beliefs) consequently results in reduced wages (read: reduced enforcement cost), which makes the involvement of outsiders unattractive, as they expect higher wages in order to act cooperatively (see Subsection 3.1.2).

This particular aspect, the limited interaction of Maghribīs with non-Maghribīs, has been challenged by a wide range of historians, often based on extended evaluations of Geniza letters.

Edwards/Ogilvie vs. Greif Edwards and Ogilvie (2008, 2009, 2012) offered a direct critique, centred around the rejection that the informal nature of enforcement mechanisms was decisive for the functioning of a coalition in the light of trials evidenced in the Geniza letters. In conjunction with this they challenge that the collectivistic features of Maghribīs were sufficiently different from the Genoese ones to use those cultural characteristics as a central assumption for Greif's argument. Both of these criticisms are based on cases from the Geniza itself. With great determination Greif meticulously defends his position (Greif, 2008, 2012), likewise on a case-by-case basis. Independent of this dispute, the most important benefit to the outsider is that Edwards and Ogilvie's claims provoke Greif to clarify, contextualise, and revisit some of the strong positions assumed in his original work, representing a central interaction with the revisionist Geniza researchers. This includes the consideration of Goldberg's more recent Geniza analysis (Goldberg, 2005) and the ex-post acknowledgement that the formal friendship (ṣuḥba) was the clearly dominatingly used institution type (Greif, 2008).

Similar to Edwards and Ogilvie (2012), and a result of his efforts in analysing the legal documents of the Geniza (complementary to Goldberg's effort concentrating on trade-related documents), Ackerman-Lieberman supports the suggestion that Maghribīs had stronger out-group relationships. He extrapolates the "openness" of commonly observed shared ownership of immobilia across confessional boundaries to commercial relationships (Ackerman-Lieberman, 2014). He writes: "each merchant sat at the center of a constellation of overlapping relationships, and the pool of potential partners was not restrictive." (Ackerman-Lieberman, 2014), alluding to a wider acceptance of new members than Greif's "guildlike coalition" (Ackerman-Lieberman, 2014) suggests. In describing the aṣḥābunā, Goldberg likewise emphasises the relative openness to connections with outsiders and newcomers as well as the existence of ṣuḥba and partnership relationships to Muslim merchants, but "... the way these men used it [the term aṣḥābunā] also expresses the sense that the Arab Jews were a central identity group among practicing merchants" (Goldberg, 2012c). However, the only evidential information on outsider relationships we are aware of comes from Greif himself, who only found two Maghribī-Muslim relationships in his original sample (Greif, 1994), and revised this to six relationships in an extended sample, suggesting a maximum of two percent of inter-religious trader relationships (Greif, 2012).

However, those findings bear little surprise, knowing that Greif's analyses were based on a limited corpus of documents only capturing around one third (175 (Greif, 1994)) of the around 695 relevant (out of more than 900 existing) trade-related letters (see Goldberg (2012b)).

A Refined Interpretation of the Maghribī Network Structure More important than this, however, is the understanding that Greif's analyses do not sufficiently explore the complex internal structure of the Maghribī trader network, an aspect that emphasises the stark contrast he builds between what he considers collectivistic and individualistic societies.[48] Greif's interpretation suggests a collective sense of commitment to the Maghribīs, mitigated by common acquaintances, religion and language (Greif, 1994), while rejecting their interpretation as a 'natural group' (Greif, 1989).

However, Greif assumed commitment of individual traders to the coalition in its entirety. He thus conceives it as a multilateral relationship network, in contrast to the dyadic nature of personal relationships as suggested by Udovitch (1977), a view that has been disputed (see also Ackerman-Lieberman (2012, 2014)).

Goldberg supports Udovitch's interpretation and identifies the relationship structures as far less cohesive, with dyadic relationships being more fundamental. Following this un-

[48] Greif (1994) uses the term 'segregated' to describe collectivistic societies who are characterised by sub-groups with comparatively well-defined boundaries (e.g. by ethnicity, family, etc.) as well as limited relationships to out-group members, and contrasts those to individualistic societies consisting of sub-groups with higher prevalence of inter-group links, which he considers 'integrated'.

derstanding, instead of interpreting the coalition as a coherent whole as done by Greif, the loosely coupled trader network (aṣḥābunā) emerged from structurally differing, more densely connected sub-groups (aṣḥābs). These aṣḥābs were overlapping networks of intimate 1:1 relationships, characterised by ṣuḥba relationships. Thus instead of concentrating the analysis on intra- and inter-coalition relationships, we believe that analysing the dominantly used institution, the ṣuḥba in conjunction with the associated network structure, provides grounds to relax Greif's assumption of collectivistic beliefs – an assumption that is at the centre of Greif's explanation for the functioning of informal enforcement.

Following the argument laid out in Subsection 3.2.1, the central characteristics of the ṣuḥba, the reciprocity characteristics, were not legally enforceable and thus informal. The reciprocal 'repayment' of 'service by service' (as opposed to money) will thus only be effective as long as the economic actors have sufficiently strong personal ties that credibly reassure commitment to honour mutual obligations.[49]

The dyadic, non-transitive nature of relationships defined a trader's aṣḥāb, concentrating his perception of the coalition on traders he had intimate dealings with. The overlapping individual aṣḥābs produced a social construct, the aṣḥābunā, as an emergent property. As such this virtual construct aṣḥābunā had no explicit representation. Udovitch (1977) describes the aṣḥābunā as "a constellation of individual relationships whose skeins could tie together a fairly large number of people; but those bonds were never expressed in terms of membership in a group abstractly defined." No individual had a global perspective on the entire coalition and neither did the aṣḥābunā have well-defined boundaries.[50] However, the fact that no individual had a global view of the coalition made it irrelevant whether the coalition actually had well-defined boundaries, as an individual's conceptualisation of the coalition was necessarily fuzzy. Furthermore, though historians have used the term 'our associates' (aṣḥābunā) to describe the entirety of the network, the individual trader's conceptualisation was ambiguous, as it could in principle equally refer to a subset of the entire network, such as a reference to the aṣḥābs of two communicating traders.[51] Goldberg (2012c) concludes: "Aṣḥābunā is an admirably and appropriately vague term."

Given Greif's understanding of the coalition, this offers a challenging position, as it loosens the coherent group structure and the assumption of well-defined boundaries. How-

[49]As mentioned at the beginning of this section, boundaries and size are an aspect that institutional economists, such as North et al. (2009), conceive as limiting factors for the functioning of informal institutions.

[50]Goldberg (2012c) notes: "The network was always in flux, both through addition, retirement, or death of individual merchants, but also through the creation, change in nature, and dissolution of bilateral ties."

[51]Recall that all evidence on the Maghribī traders coalition relies on letters between traders, which permits this conjecture.

ever, the recent findings offer potential for a much more pragmatic reinterpretation of the motivation for Maghribīs to honour informal commitments.

Although individuals did not have a global view of their trader network, they maintained individual groups of intimate ṣuḥba relationships which represented the 'personal glue' required to induce commitment, such as the necessary face-to-face meetings, which were generally witnessed. An individual's commitment would thus not need to extend to the broader aṣḥābunā, but could concentrate on a sub-group, its aṣḥāb, whose membership one could directly control.

The Benefits of Ṣuḥba Relationships Entering intimate mutual commitment relationships bore significant benefits for Maghribīs. They were trade professionals and committed to be involved in trade-related activities for a lifetime. As professionals they were aware of fluctuations of market prices and the importance of opportunistic adjustment to market conditions, aspects which formal instruments such as the qirāḍ algoyīm (the equivalent to the European commenda, see Subsection 3.1.3) would have prohibited, since details of trade interactions would have had to be specified ex-ante, leaving the agent handling the goods little autonomy to accommodate changing market conditions. Furthermore, contractually regulated partnerships (qirāḍ algoyīm or khulṭa), which for the most part prescribed material distribution but also individual obligations, demanded considerable attention and gave partnerships the character of exclusiveness, if not preventing partners entirely from engaging in other ventures (such as the case of the unlimited partnership) (Ackerman-Lieberman, 2014). Ṣuḥba relationships were much more flexible and, in principle, allowed the trader to enter an arbitrary number of such relationships. An individual trader's freedom to entertain additional ṣuḥba relationships was only constrained by his ability to honour the commitments, such as storing and handling goods, that arose from such relationships. Relationships based on the ṣuḥba thus offered the individual a greater level of autonomy than other instruments and enabled him to selectively focus on particular relationships in a demand-oriented manner (e.g. based on market conditions).

In the appreciation of long-term relationships with recurrent interactions, the remuneration-free services of the ṣuḥba were attractive to reduce transaction costs by avoiding the balancing of accounts associated with payment for services (beyond the accounting for the actual goods handling).

Based on those aspects we suggest that the system employed by the Maghribīs was not characterised by devotion to the 'greater good' of the coalition, but – in combination with the ṣuḥba as preferred relationship type – offered strong opportunistic benefits for individual (and in principle even individualistic) traders who could regulate their own aṣḥābs in response to market needs – but also based on the commitment of their aṣḥābs' members.

Monitoring among Maghribīs To this stage it is not clear, however, how individual aṣḥāb members' cooperation could be controlled in remote market places that were outside the reach of the other members of the aṣḥāb. To explain the functioning of monitoring, we allude to the information sharing aspects of the Maghribī trader network.

As part of the ṣuḥba, traders performed extensive communication to coordinate their trade dealings. While obligations from the non-transitive trade relationships did not transgress aṣḥāb boundaries, information did. In the absence of legal backing for the reciprocity elements of the ṣuḥba, traders needed to credibly reassure the compliant handling of service requests on their part in order to 'earn' reciprocal services and to show enduring commitment to the in-principle open-ended relationship. As reflected in the Geniza itself, letters were the central information medium of communication and business coordination. Traders used letters to exchange assessments of markets and business-related information (55 percent of content according to Goldberg (2012b)), but more importantly, they shared information about traders' conduct, which was found in around 75 percent of all letters, filling around 20 percent of letter content (Goldberg, 2012b).[52] Information both related to themselves (with 41 percent devoted to the writer's self-reporting), but also of third parties, i.e. other traders at their respective market places, with information in the latter covering nearly half (46 percent) of conduct-related letter content. Reports about third parties predominantly reflect negative conduct, generally challenging the quality of their work (around 75 percent) (Goldberg, 2012c), and were meant to advise the selection of potential future 'formal friends' by a committed and concerned business partner (see e.g. Goldberg (2012b)). Evaluating an agent's conduct, beyond the subjective evaluation, was a generally feasible task as all operations such as unloading and opening of goods at market places as well as trading were public affairs and accounted for (Goldberg, 2012a; Goitein, 2000b, 1973; Greif, 1989, 1994), since property rights remained legally enforceable even if services were not (see Subsection 3.2.1). Although instructional letters were in principle private, their actual operational nature was near public: other than goods transport that concentrated on the summer seasons, letters could be transported throughout the year, using caravans if waterways could not be operated on. Transporting the letters relied on various different transportation systems (e.g. sending with fellow traders or independent couriers), the choice of which was decisive for the privacy of information (Goldberg, 2012b). It was further common that letters were opened and

[52] A good overview on the distribution of letter content is provided by Goldberg (2012c).

read aloud in public to establish present traders as witnesses for delegated rights and obligations, given that only testimonies were admissible in courts; the letters themselves were not legal documents. However, even if contents were kept private, their receipt was a public affair (Goldberg, 2012b). Moreover, the confirmation of letter receipt was a central principle and provided assurance of successful delivery (Goldberg, 2012b). This was further enhanced by the frequent copying of letters and sending of redundant copies, particular when shifting from comparatively reliable land-based to the more environmentally exposed and thus less dependable (see e.g. Goldberg (2012c)) but certainly more favoured[53] sea travel. However, while redundancy increased reliability of delivery, it came at the price of reducing the privacy of letters further. So whether letter content was eventually private or not, the receipt was certainly public knowledge, and the sender could expect a confirmation of the receipt of his instructions.

Returning to the implications for informal enforcement, the focus on third parties and the practically public nature of letters, particularly if instructions were concerned, information was likely to be picked up and shared by Maghribīs that were present at the same trade location, ultimately transgressing aṣḥāb boundaries and spreading across the entire network of interlinked aṣḥābs (aṣḥābunā). Structurally speaking, in the network of trade relationships each individual was a *focal point* of its isolated aṣḥāb constructed of undirected links: trade service requests would never extend beyond the aṣḥāb's boundaries. With respect to information transmission each individual represented a *hub*, effectively broadcasting any incoming information across all traders that were interconnected by aṣḥābs. The virtual structure of the aṣḥābunā thus emerged from the information flow that connected the overlapping but well-defined individual aṣḥābs.

On the Role of Jāh In this system, the central currency for future commitment was one's reputation, jāh, which we can conceive of as a function of both the connectedness of an individual as well the duration of its cooperative membership with the network. Traders were under constant pressure to show their commitment and credibly convey the quality of services (which consumed around 41 percent of letter content), reports for which they needed to expect either silent confirmation (by not appearing in their letters), or, in best case, praise by other traders, which would bring the prospect of further trade relationships and thus consequentially a further jāh increase. Positive jāh would thus not only ensure continuous participation in the relationship network (with the benefits of a multitude of accessible markets and

[53]Goitein estimates the letter references to sea travel, as opposed to land-based travel, at proportions of 50 to 1 (Goitein, 1960).

low transaction costs), but it was the key to drive the development of one's trader network, in the best case extending to the comprehensive coverage of relevant market locations across the Islamic Mediterranean.[54]

Given the value of an individual's jāh, greater involvement in the network reinforced commitment, thus reinforcing cooperation. A strongly involved trader had the incentive to spread information about misconduct as well as the necessary connections to do so. The interest and impact of identifying cheaters was thus increasing with the announcing trader's jāh.

From an opportunistic perspective, an increasing jāh level had even more far-reaching effects. 'Earned obligations' against other traders depended not only on the services performed on their behalf, but were also anti-proportional to the status difference. A trader was thus motivated to develop a relatively higher jāh compared to his own aṣḥāb in order to procure services cheaply (i.e. with few reciprocal obligations), enabling him to attend to a greater number of relationships. Furthermore, from the perspective of opportunity cost, it would be cheaper to delegate 'owed services' (i.e. service obligations against him) across his aṣḥāb, or to agents of lower jāh, rather than performing laborious services himself. This effectively shifted his perspective from a *service provider* towards a role as a *service broker*, moving from an operating to a coordinating role. The gradual progression could have been an attractive incentive to 'buy into' the informal reputation-based reciprocal services system, at the price of being 'locked in' and thus enforce it as the value of jāh did not extend beyond the aṣḥābunā.

In fact literature (Goldberg, 2012c) supports the shift from a travelling agent towards a sedentary one, starting with the choice of a homebase after about 10 years after the onset of a career, and, potentially interleaved by further change of their bases (often across political boundaries, but generally within the Fatimid Empire), ultimately settling as a sedentary trader concentrating on coordinating, rather than operative tasks.

The Apprenticeship System To this stage, we can retrace that established traders had a significant motivation to act compliantly, and to enforce the compliance with increasing jāh level. However, other than relying on Greif's culture hypothesis, we cannot yet explain how the coalition would have dealt with newcomers, whether coming from the coalition's central identity group – Arab Jews – or from other ethnic and religious backgrounds. In-

[54] An example for this (and previously highlighted in Section 3.1) is Nahray Ibn Nissīm, who, as far as documented, had the most extensive relationship network, capturing more than 400 relationships, with a significant drop to the influential trader Yūsuf Ibn 'Awkal, who still maintained around 150 relationships (Goldberg, 2012c).

teracting with newcomers was highly unattractive, due to the absence of an incentive for compliance and thus uncertainty about their commitment to honour reciprocal obligations. Further, no knowledge of their actual trading skills was established, as traders were preferably looking for long-running service-centred relationships with considerable autonomy and thus influence on the partner's economic fate. The Geniza offers hints by providing traces of what we will refer to as an *apprenticeship system*[55] used to evaluate and control the entry of newcomers into the trader network. Descendants of Maghribī traders who wanted to enter a promising career in long-distance trade had to develop their own standing in the society. The benefits of informal relationships were only extended to individuals that showed credible long-term commitment and were considered suitable to further their prospective partners' interests. Though kinship relationships offered the necessary opportunities for a trader career, reputation was not automatically inherited but had to be earned.[56] Traders thus sent their sons[57] to fellow associates who could delegate increasingly challenging tasks to their 'apprentices' while monitoring them closely.[58] Offering apprenticeships offered the mentoring partner the benefit of obtaining free labour which could be applied for a wide range of tasks, including harvesting and packing, at the price of sustaining the apprentice's livelihood as well as higher monitoring cost.[59] In the course of the apprenticeship, the apprentice was entrusted with trade-related activities of increasing responsibility. While initially exclusively performing trade for the mentor, and assuming sufficiently developed trading skills, he was progressively exposed to other Maghribī traders, such as his mentor's aṣḥāb. This could include the introduction to established traders by his mentor[60] and enable him to shape his own ṣuḥba relationships and, in consequence, his aṣḥāb. This process could last more than a decade, and the admission to full tradership was not only a matter of compliance and loyalty,

[55]Goldberg (2012c) refers to it as "system of junior associates".

[56]Family relationships certainly reflected reputation, but it was seen as an "insurance policy" for potential private-order enforcement against the family of a shirking trader. Reputation gain, however, was attached to the individual, and thus had to be earned individually (Greif, 1989).

[57]As far as observable from the Geniza sources, females did not perform trader roles, other than managing a trader's homestead in her husband's absence. Given the Maghribīs' tolerance towards polygyny, it was possible for traders to engage in multiple marital relationships, which offered the convenience of maintaining presence at different trade bases (Ackerman-Lieberman, 2014; Friedman, 1986, 1982). Further support for the absent consideration of women in Geniza letters is provided by Goldberg (2012c).

[58]The work of junior associates found particular emphasis in shared letters, suggesting specific focus on those newcomers in trade locations (Goldberg, 2012c).

[59]Goldberg (2012a) states: "Geniza merchants seem to have preferred contracts in opposite proportion to the agent's natural incentive to provide good service."

[60]Such introduction was generally performed via letter sections dedicated to introduce other traders and suggesting the uptake of relationships, which were found in around four percent of letters (Goldberg, 2012c). However, the establishment of such relationship remained a private decision of the recipient and the recommendee. In any case the recommendation would be undertaken with great care and only once apprentices were considered skilled, since the recommendation put the recommender's jāh at stake.

but moreover, depended on the individual's skill; establishment as a trader was by no means guaranteed and bore option for failure (Goldberg, 2012c).

This elaborate mechanism introduced a significant barrier for the introduction of individuals into one's aṣḥāb. As a consequence of such an entry barrier, the time investment necessary to enter the system implied clear career objectives – a) the jāh established during apprenticeship, b) benefits associated with coalition affiliation (e.g. low transaction costs), and c) future prospects as a sedentary trader – that would have been engrained if full trader status was eventually attained.

Given this reinterpretation of the individuals' motivations to comply with informal rules, it may have not so much been rooted in collectivistic beliefs (without suggesting that those did not exist[61]), but rather a rational choice from the perspective of professional traders.

Self-Reinforcing Nature of the Network based on the Revised Interpretation Another aspect that challenges the assumption of collectivistically motivated commitment to the coalition is the atmosphere of fear that underlay the membership. Though compliant behaviour was the default, continuous reassurance of compliance was required, be it proactive sharing of profit (e.g. proactive sharing of profit from pepper sales (Goitein, 2000b)), the public announcement of letter content (Goldberg, 2012b) to display accountability for delegated obligations (as well their measurable fulfilment), or repeated assurance of acting in the interest of the sender of goods (especially if no explicit instructions had been provided).[62] The meaning of the informal nature of jāh was so predominant that it overruled court decisions. Even if accusations had been cleared in front of courts, simply the inflicted doubt made continuous operation in the trader network impossible (Goldberg, 2005). If a genuine 'esprit des corps' (Ackerman-Lieberman, 2012) and thus near selfless devotion to the coalition, as suggested by Greif's emphasis of collectivistic beliefs, had existed, these reassurances, along with the frequent reminder of timeliness in which trade should be performed (on the part of the delegating party), should not have been necessary – if not even counter-productive.

Commitment to the aṣḥābunā was thus driven by the selfish interest to further one's trade and was based only on a single currency one had not only to maintain but also to increase, given that it reduced the cost to procure others' services (as argued earlier in this section):

[61] The construct of reputation had collectivistic elements in that it its construction occurred individually, i.e. each individual needed to develop its reputation independently, while its destruction (e.g. by cheating) would also affect its kin and thus had a collective component (Greif, 1989).

[62] Goitein writes: "But even given free rein, some of correspondents were not sure of satisfying their friends. Our letters are therefore full of assurances that the writer had made every possible effort for this friend's affairs. One writer repeats three times that he never gives preference to his own interests over those of the addressee ..." (Goitein, 2000b).

one's jāh.

This focus on reputation initiated a reinforcement cycle. Given that jāh had only meaning (and thus value) within the coalition, it was in the interest of established members (who could cheaply earn obligations) to reinforce others' (and particularly traders of lower jāh) commitments by offering harsh non-gradual sanctions for misdemeanours ('one strike policy'), while luring them with decreasing transaction costs (free services, increasing accessibility) and increasing social (and thus economic) standing. This would ultimately motivate their commitment to the coalition as increasing jāh reduced the attractiveness of shaping prolonged links to outsiders. Sustained membership and aligned interests thus promoted a commitment to the institution aṣḥābunā without necessitating a collectivistic belief (i.e. commitment to the group and its members) per se. On an individual level this would be experienced as a progression from the fear of being expelled towards a selfish motivation to sustain the institution by monitoring (here: observation and exchange of gossip in letters), given the significant lifetime investment the development of jāh (and its worthlessness outside the aṣḥābunā) implied.

The emergent network structure and fuzzy understanding of the aṣḥābunā by individuals generated an effect that supports this interpretation of the trader coalition. Individuals would not likely have complete knowledge about another trader's relationships, and thus be deterred from cheating when facing the potential presence of an unknown monitor. The practical aspect of the aṣḥābunā as an institution was thus to reduce uncertainty about compliant trading on the part of the investor by increasing the uncertainty about the extent of monitoring on the part of the goods-handling agent. Psychologically this uncertainty could have effectively facilitated panopticistic self-monitoring (Foucault, 1977), making the potential cheater likewise a monitor of his own behaviour, and thus reducing the extent to which actual monitoring was necessary. This argument supports the idea that the refined decentralised structure of the aṣḥābunā could have had a significant impact on the strong extent of compliance among the Maghribīs.

This understanding does not revise the historical outcome directly. However, it does refine the characterisation of the trader society. Goitein's subtle admiration for his subject of study[63] ignores the fact that, while Maghribīs were indeed operating based on trust relationships and ('formal') friendships (i.e. ṣuḥba), they had considerable leverage against each other which surfaced in the negativity bias and potentially abusive tendencies expressed in

[63]Goitein alludes to the facilitation of trade based on the "... human qualities of mutual trust and friendship." (Goitein, 2000b).

letters[64] which required continuous reassurance of compliance.[65] Greif's ascription of compliance and multilateral punishment to cultural traits (Greif, 1994), on the other hand, may have been overly simplistic and did not reflect the intricate nature of the implemented institutions. The employed institutional instruments were developed with the intention to reduce transaction costs for individual ventures while maintaining a flexible relationship network in professional tradership, an aspect that could not have been supported by contemporary legal institutions or theory.[66] An essential part of the Maghribīs' success thus lay in the emergent complexity and self-enforcing nature of the trader network that played a pivotal role in establishing and maintaining an overall cooperative outcome.

3.2.3 Summary

At this stage we have provided a broad overview of the current state of knowledge with respect to the Maghribī Traders Coalition. We started with a general introduction of Maghribī characteristics, with particular focus on the social structure in Subsection 3.1.2, before providing a survey of the historically available institutional instruments along with their practical relevance for the Maghribīs in Subsection 3.1.3. Based on this background we entered the discussion of essential inconsistent interpretations of historical reality in Section 3.2, mainly between the comparatist Greif and more recent contributions by Goldberg and Ackerman-Lieberman. Contextualising the institutional instruments, we retraced the differing interpretation of formality, which led Greif and Goldberg to categorise the same observed institution, which – based on Goldberg's analyses – can be identified as ṣuḥba, as informal (Greif) and formal (Goldberg), with Greif overseeing the remuneration-free nature of this institution. In Subsection 3.2.1 we thus arrived at the conclusion that the essential difference lay in Greif's understanding of 'formal' as being grounded in *legal backing*, while Goldberg associated formality with aspects of enactment and termination of ṣuḥba relationships, along with the enforceability of (in our view secondary) property rights. In consequence we differentiated formality into *written/unwritten form* and *legal backing* for all institutional instruments previously discussed in Subsection 3.1.3.

[64] In the interest to induce stronger monitoring " ... merchants often wanted to get associates near the agent to provide some kind of oversight, either by direct request or by provoking a colleague to worry" (Goldberg, 2012b). Challenging a partner's compliance could further entail a strategic component: "Moderate abuse of one's fellows was only possible because it could be so easily and strategically forgotten" (Goldberg, 2012b).

[65] Goldberg (2012b) reports that nearly half of all conduct-related letter content (20 percent) were engaged with self-defense and self-praise.

[66] Islamic legal theory did not recognize the concept of a corporate person (Ackerman-Lieberman, 2014) and neither did it support labour contracts (and as such the reciprocal service obligations of the ṣuḥba) (see Subsection 3.2.1). See also Goldberg (2012c).

Extending from this clarification, in Subsection 3.2.2 we analysed the impact the refined institutional understanding – now based on ṣuḥba relationships – had on the interpretation of the Maghribī relationship network, which Greif construed as a coherent well-defined entity with multilateral enforcement. In the light of the intimate 1:1 ṣuḥba relationship, along with Goldberg's identification of sub-networks (aṣḥābs) derived from the individual dyadic relationships, we interpreted the overall network as a fuzzy entity that emerged from the relatively well-defined interlinked aṣḥābs. The intimate reciprocity relationships along with unknown global knowledge about the aṣḥābunā (and thus observers) as deterrant for non-compliance provide a credible alternative explanation for the effectiveness of information sharing and self-enforcing compliance. Following this conception, traders simply needed to honour their individual relationships by performing reciprocal services and share information, as opposed to serving the greater coalition as suggested by Greif. Supporting this view, we further highlighted the importance of status based on Goldberg's findings, and characterised the self-reinforcing effect of jāh in the light of its importance within the coalition but absent value outside the network. This aspect was further complemented with the flexible nature of the ṣuḥba, the application of which manifested itself in reduced transaction costs. A further component we discussed is the Maghribīan apprenticeship system which appeared to be a means of assessing and controlling access of newcomers to the network.

An additional aspect of interest the presumed importance of cultural difference between Maghribī and Genoese traders, namely the absence of political constraints that would have prevented the cooperation of both trader societies. Literature provides sufficient evidence to suggest that such constraints existed. However, in the light of the secondary role with respect to this work – our evaluations in Subsection 4.2.1 onwards rely on information explored in the previous debates – we have shifted this discussion into Appendix A.

To this stage we have provided an overview of the Maghribī society, with specific emphasis on providing a clear understanding on the historical happenstance. At this point we turn the Maghribīs' Southern European counterpart, the Genoese, but introduce those with lesser detail.

3.3 Genoese Traders

As alluded to in the previous section, the Maghribī coalition's characteristics were in striking contrast to other contemporary trader societies, with the Southern European Italian city states offering the sharpest contrast in societal structure, self-understanding, and institutional environment. In this section we want to give an insight into the Genoese society, beyond its role as a stereotype of Southern European city states, and furthermore, to contextualize some of the assumptions put forth for the comparative modelling of both societies by Greif. We will do this in greater brevity compared to the Maghribīs, given its better coverage and consistent treatment in literature, leaving fewer gaps that demand extended discussion.

3.3.1 Characteristics of the Genoese Society

Grown from a 'village' under Lombard and Carolingian rule in the 6th century, Genoa's history shows a gap up to 934 AD, where it fell victim to Muslim raids, with documented history continuing at 958 AD, with a charter documenting property possessions (van Doosselaere, 2009). Surrounded by non-arable mountainous rocky land, the limited agricultural opportunities explain Genoa's dedication to 'the sea'. Driven by a limited reach of the rulers' influence outside city boundaries and chronic economic shortages in the limited space at the disposal of its inhabitants, the Genoese lived off the sea, proffering opportunities other than just "bad fishing" (Epstein, 1996). The Genoese intensively engaged in piracy as a profitable enterprise, along with increasing, initially regionally concentrated, sea-based trading of grain and salt (Epstein, 1996). As Airaldi (1969) puts it, "Ianuensis ergo mercator" (Genoese therefore merchant), suitably describes the purpose of being 'Genoese'. However, their recognition as a maritime power only grew with their involvement in the First Crusade around 1097 AD that brought considerable economic direct benefit (based on the acquired spoils), but also stimulated Genoa's role as an important shipbuilding centre (van Doosselaere, 2009). The long-term benefit of their participation, however, lay in the establishment of trade posts in conquered lands. However, even prior to the crusade, the increasing Genoese assertiveness was displayed in expeditions serving the attack of al-Mahdiyya (in 1087 AD) alongside Pisa as a response to the Fatimids' attempt to capture Sardinia (in 1016 AD) with its strategic position in the Gulf of Genoa (Epstein, 1996; van Doosselaere, 2009).

The society, with its history of economic hardship and a fierce political environment was, as Lopez (1982) suggests, characterized by four central themes:

- strong religiosity, but clear separation of religious and business affairs
- "irrepressible individualism" (Lopez, 1982)
- strong emphasis on clan relationships
- openness towards newcomers in order to expand trade activities

Given the continuous contest between the two dominating Genoese clans, the Manecianos and Carmadinos (Greif, 2006), the Genoese employed a podestà (an administrator similar to the *doge* in Venice) that would serve the city for a one year (Greif, 2006; Epstein, 1996), and generally be an outsider that pledged neutrality and acted in Genoa's best interest. His role was to limit the conflict of either clan which had paralysed Genoa's progress in securing trade privileges in remote ports, initially leaving Genoa behind smaller cities such as Pisa, which did not face such inner political struggles (Ceccardi and Monleone, 29). However, in the absence of a powerful central government, a reliable political organisation concerned with outside affairs was important to establish maritime power to develop Genoa's role in the context of Mediterranean trade.

Individualism The competitiveness among the Genoese could in part be ascribed to effects of Church policies on family affairs. Following its lead, the nuclear family concept was encouraged, with the separation of larger kinship groups by disapproving of common practices, such polygamy and remarriage, and more importantly, the abolition of multi-generation households, all of which the Church perceived as drivers towards larger family units, preventing the Church from 'breaking into' these family structures and maintaining influence (Goody, 1983; Greif, 2006; Greif et al., 2012). Given the strong religious influence of the Church, this development spread through the European continent, an aspect that may have motivated Morris to suggest that by the end of the 12th century, Europe had already "discovered the individual" (Morris, 1972). Despite the continuous existence and importance of clans to further group interests in the challenging economic environment, in the Genoese case no clan managed to concentrate the entire political power. Greif (2006) suggests that breaking up kinship-based groups, and instead substituting them with group affiliations based on profession (e.g. guilds), drove an important institutional restructuring based on *task affiliation*, providing an advantage in the further specialisation and economic development, in contrast to collectivistic cultures, such as the Maghribīs, in which *clan affiliation* was the primary source of identity and standing. An indicator of that differing sense of affiliation among Genoese is observable in van Doosselaere's dataset of commercial relations (van Doosselaere, 2009), with the increasing uptake of names that described affiliation (e.g. *Phillipus Marcellius* (Philip the Butcher)) as an alternative to more traditional surnames based on lineage (e.g. *Bertoloto filius quondam Alberti* (Bertolotto, the son of the deceased Albert)), origin (e.g. *Obertus de Langasco*), or less common, physical features (e.g. *Boccanegra* (black mouth)) (van Doosselaere, 2009), a tendency we see still reflected in modern-day Western surnames, such as Smith, Charpentier, or Müller. Among the Maghribīs, as with other collective societies, such as Arabs and Chinese, surnames continued to represent lineage (e.g. Yūsuf Ibn 'Awkal (Yūsuf, the son of 'Awkal)), a convention that holds to the modern day.

The Use of Contracts among Genoese Along with the reliance on clan relationships, and despite their openness to newcomers, Genoese were known for the secrecy with which they performed trade operations. They hardly shared details of their operations, if they had the discretion to do so, or, when engaging in contractual agreements necessary for long-distance trade, delayed announcement of those until notary confirmation. Unlike the Maghribī case, private affairs and trade contracts were not openly shared until formalised by notaries that maintained cartularies of such agreements in order to facilitate enforcement in the case of disputes. The contracts themselves were mostly a public affair, which depended on the place of conducting such contracts and the number and nature of witnesses. However, the signing of such contracts also occurred privately, with knowledge only to the scribe, contracting parties and witnesses. In any case, formal contracts required witnessing by at least two 'sane' persons, the number of which depended on the importance of the witnessed transaction as well as the availability of witnesses in busy sailing seasons with more than thousand Genoese at sea at a given time (Epstein, 1996).[67] The employment of such contracts served the minimisation of uncertainty, reflecting a classical purpose of institutions. Breach of contracts was severely punished by Genoese law. Violators were publicly beaten and had their noses cut off so as to mark them as cheaters. Notaries that did not fulfil their duties of neutrality or ignored the necessity of witnessing, faced the mutilation of their hands (Epstein, 1994).

3.3.2 Institutions

According to Epstein (1996), the cartularies reveal that Genoese traders engaged in three types of contractual agreements to coordinate trade, including the *commenda*, the *societas* and the *sea loan*.

The *commenda*, as mentioned in Subsection 3.1.3, separated the obligations of two stakeholders into the commendator, who sponsored a business undertaking financially as an investor, and the tractator, who performed the laborious part of the operation, i.e. the actual trade. Profits were split between the parties, with the tractator conventionally receiving one quarter of the realised profits, and the investor received the bulk of the return (Epstein, 1996; van Doosselaere, 2009).

[67]The cartulary of Giovanni di Scriba who acted as scribe from 1154-1164 AD, the oldest available one of successive scribes, revealed between two and eight witnesses for individual contracts, with higher numbers of witnesses assuring compliance for more delicate matters (Epstein, 1994).

In the *societas* either member invested in the common venture with potentially differing shares, but the executing party would receive a greater return from the profit in order to remunerate him for his labour (Epstein, 1996). From the standpoint of these characteristics, the societas was comparable to the *active partnership* (shirka/khulṭa, see Subsection 3.1.3) the Maghribīs engaged in.

The *sea loan*, had the same intuitions as described in Subsection 3.1.3. It acted similar to a regular loan, which an individual could take up in order to generate profits exceeding the interest payment due upon return. Although of lesser importance than the commenda for Genoese trade, it offered a good indication of the high profit margins for long-distance trade, given that the interest rate for such loans lay between 40 and 100 percent (van Doosselaere, 2009).

The stratification into the different contract types is not unanimous among historians such as Epstein and van Doosselaere. In his relatively recent exposition (capturing over 7,221 commercial relationships between 1154 and 1300 AD), van Doosselaere's (2009) identifies only two contract types, namely the sea loan and the commenda, as prevalent in Genoese maritime trade relationships. But he adds that the commenda had been conceived under various different names, including *accomendacio*, *collegantia*, and – bearing similarity with Epstein's third type – *societas*.

However, this differentiation is of marginal interest, knowing that around 93 percent of contracts were commendae (van Doosselaere, 2009), supporting the perceived uniformity of contractual relationships Lopez (1975) described.

The dominant use of the commenda is hardly surprising, given the vast social stratification of the Genoese society. Greif's earlier analysis of Giovanni di Scriba's cartulary of 612 contracts revealed strong concentration of capital in very few families of the Genoese society. From the 180 investors only 37 families contributed 90 percent of the capital (Greif, 1993, 2006).

Business Operations of Genoese Genoese tended to control their capital in family firms, both to assure a fluent transition of leadership and to avoid the division of capital upon inheritance (Greif, 1994). Performing the actual trade was cumbersome and bore considerable risks. Delegating this part to wilful executors, made the use of the commenda particularly attractive. It concentrated the investment on family firms, but allowed individuals without monetary funds or foreigners easy access to temporary work, for whom such trade offered considerable payoff. The high level of uncertainty for long-distance ventures outside the institutional boundaries of Genoa based on unpredictable timing of realization of profits, and the extent to which those could be supported by a contractual backing of agreements, led to opportunistic partnership choice based on availability, rather than experience or previous collaboration, inhibiting expectations of reemployment (van Doosselaere, 2009) – or as van Doosselaere (2009) puts it: "... opportunistic behaviour was the norm." Given the limited chance for reemployment[68] the intrinsic motivation for compliant behaviour was limited, but the commenda contract along with its sanctions for cheating offered the necessary means to deter potential cheaters. Consequently, it was all but unusual that the tractatores of commendae, i.e. the ones performed the actual trade operations, had a broad range of occupational backgrounds, including artisans who could utilise long-distance trade to further their own reach. On the investing side, families offered shares to outsiders, effectively establishing a share market system that allowed anyone to invest in long-distance trade (Greif, 1994), inviting individuals of vastly differing backgrounds. For example, the cartularies reveal a considerable involvement of women but also a wide range of occupations (van Doosselaere, 2009), with exotic examples being churchmen that used trade as a means of investment (Kedar, 1976). The reliance on enforceable formal mechanisms introduced an abstraction between investors (commendatores) and operators (tractatores), with investors seeking the promise of high returns, while being unaware of and dissociated from their counterpart lifting the operation. However, as indicated before, from the perspective of operators, the formal nature of such trade arrangements offered a comparatively unconstrained accessibility to newcomers.

Given the near guaranteed profit,[69] opportunistic operators, interested in fast profits without any other investment than their labour, did not have any related trade training, were generally poor (de Roover, 1958), and did not enjoy a high standing in the Genoese society (Byrne, 1916). In contrast to the Maghribīs, whose apprenticeship system controlled access to the privileged trader coalition, Genoese traders did not have such entry barriers for trade participation. In fact a sample of apprenticeships existing in the later Genoese society (1451 - 1517 AD) (Epstein, 1996) does not show any trade-related apprenticeship other than for specialised skills for shipbuilding (such as rope makers, ship carpenters, etc.), for our context the most relevant being the trade of shipwrights who had to be taken aboard to monitor and document overseas transactions.

Social Stratification of Genoese Trader Society In contrast to the professionalised integrated role understanding (both delegating services and performing services for others)

[68]Commenda agreements were strictly concentrated on one trade interaction, with limited chance of reemployment (van Doosselaere, 2009). Based on his analysis of commenda relationship network for the period from 1154 to 1315 AD, van Doosselaere further "... confirms that the overwhelming majority of Genoese traders were occasional participants in long-distance trade." (van Doosselaere, 2009)

[69]Commendae typically facilitated profits between 20 and 110 percent (van Doosselaere, 2009).

exhibited by the Maghribīs (see Sections 3.1 and 3.2), the Genoese trader community was characterised by a comparatively rigid role stratification, with investors hardly ever engaging in trade contracts with family members,[70] while the tractatores provided trade activities without requiring experience. From his analyses Greif (2012) concludes that only 21 percent of Genoese traders operated in both roles during their lifetime. Thus they had generally advanced from accumulating capital by facilitating trade towards positions as investors who then invested into long-distance trade themselves, suggesting some flexibility of social movement. For the Maghribīs, on the other hand, the fraction of individuals acting both in investing and trade-operating roles lay at around 71 percent (Greif, 1994). The focus on role specialisation and the greater fluctuation of employed operators is further reflected in the asymmetry between investors and operators that engaged in commercial relationships. In his cartulary Giovanni di Scriba records 180 investors that, over time, employed 335 operators,[71] indicating a ratio of 1.57 (Greif, 1993, 1994, 2008; González De Lara, 2008).

This elaboration leads us to suggest that *Maghribīs were role-integrated trade specialists*, while *Genoese traders were characterised by stratified groups of non-specialists*. The need for the increased professionalisation among Maghribīs cannot only be attributed to the establishment of jāh in order to sustain informal cooperation and the apprenticeship system as means to separate wheat from sprout (see Subsection 3.2.2) alone. Professionalisation and choice of institutional instruments were driven by the greater price-sensitivity of more developed Eastern markets with fluctuating market prices and, associated with this, the need to adjust to such circumstances, an aspect that was even reflected in the qirāḍ al-goyīm, the Islamic equivalent to the commenda that offered more flexible profit distribution schemes (Ackerman-Lieberman, 2012). Medieval Europe, in contrast, was seller-centric. Goods coming from the Far East were in high demand and could be sold at the seller's dictate; the chronic scarcity experienced by Genoese did not require advanced bargaining skills or adaptation to market prices (Epstein, 1996; Ackerman-Lieberman, 2012).

Looking at the characteristics of both societies, we offer the reader a more differentiated perspective and soften the bifurcation into 'collectivistic' and 'individualistic' societies applied by the economic comparatist Greif, and instead look at particular characteristics that have not been explored in greater depth.

Candidates for such exploration include the inherent secrecy of Genoese traders that

[70]The cartulary of Giovanni di Scriba reveals around 6.45 percent of intra-family contract relationships. This general separation of investor and operator role was further supported by the considerable capital foreigners invested into such ventures (18.3 percent) (Greif, 1993, 1994, 2008; González De Lara, 2008).

[71]Note that for the purpose of crisp differentiation we employ the terms 'investor' and 'operator', in contrast to Greif's use of 'merchant' and 'agent'.

did not only rely on formal mechanisms to govern compliance but likewise private-order enforcement mechanisms based on their "clannishness" (Lopez, 1982), the central role of which made them different from other city states such as Pisa and Venice. For the Maghribīs, an aspect that has not been empirically appreciated is the apprenticeship system they employed to assure a 'professional' trader network. In Chapter 5 onwards we will then 'step back' and look at the compliance mechanisms from a high-level perspective, augmenting the *application* of agent-based modelling with conceptual and methodological contributions that can find general use beyond the scenarios explored in this work.

3.4 Summary

To this stage we have provided a profound overview over the essential characteristics of the respective societies, their social structures, characteristics and employed institutional instruments. Before turning to Greif's conception of institutional analysis and the prototypical model he applies to explain the societies' respective institutions from an economical perspective, let us summarise and outline the central characteristics of both societies.

The Maghribīs were characterised and unified by their Jewish heritage and operation in a benign Islamic environment. Despite their low numbers, they managed to facilitate long-distance trade along the North African coast based on informal means relying on dense networks of dyadic relationships that followed rituals of initiation and termination potentially continuing throughout a trader's lifetime. Services were 'paid' in reciprocal obligations. Failing to cooperate was sanctioned with removal from all trade networks, whose membership defined a trader's professionalism.

The Genoese, in contrast, developed as an independent city state in Southern Europe, and were characterised by a weak central government that could not prevent hostile invasions and shaped Genoese to be self-reliant. In an environment of constant scarcity and buyer-centric markets, Genoese were furthering their individual interests, increasingly shifting from piracy to trading activities, which they either saw as an investment or occupation. To sustain cooperation for one-off ventures in a rapidly growing society of vast social stratification and with constant influx of foreigners, Genoese nearly exclusively relied on meticulously documented contracts that offered a basis for public enforcement.

Table 3.2 compares the essential characteristics of both societies, many of which we will refer to throughout the remainder of this work.

With the essential background on both Maghribī and Genoese trader societies as well as the contemporary picture for the historical reality, in Chapter 4 we turn our attention to

Table 3.2: Characteristics of Genoese and Maghribī Trader Societies

Characteristic	**Genoese**	**Maghribīs**
Trade Relationships	mostly formal (commenda-based)	mostly informal (ṣuḥba-based)
Sanctioning/Monitoring	centralised	decentralised
Trader Roles	stratified in investor and operator	integrated trader understanding
Class Stratification	strong social stratification between investors and low status operators; broad range of operator backgrounds	largely homogeneous middle class
Lifespan of Relationships	short-term, usually limited to one venture	any length
Role Understanding	Long-distance trade as investment/job opportunity	Trader as life-long occupation

Greif's modelling approach – his operationalisation – beyond the essential assumptions that we laid out before (in Subsection 1.3.1). Following this, we apply ABM to provide refined scenarios for specific characteristics based on the knowledge basis developed in this chapter.

4

Analysing the Maghribī Traders Coalition

With the background of the central characteristics of the Genoese and Maghribī societies, in Subsection 4.1.1 we explore the game-theoretical evaluation Greif (2006) has chosen in his seminal work on the Maghribī Traders Coalition. In Subsection 4.1.2 we then discuss the applied modelling approach and his assumptions, which we initially explored in Subsection 1.3.1. We suggest that the high-level or unrealistic nature of those assumptions is, at least in part, driven by his chosen means of exploration. Turning towards the explorative part of the thesis, in Section 4.2 onwards we attempt to soften some of Greif's assumptions by exploring two specific aspects of the historical societies, namely the relaxation of the secrecy assumption of Genoese (Subsection 4.2.1) and the Maghribīan apprenticeship model (Subsection 4.2.2) as filtering mechanism for newcomers, using Agent-Based Modelling (see Subsection 2.3.1).

4.1 Reviewing Greif

4.1.1 Game-theoretical Foundation of Greif's Scenario

A central thesis of Greif's work is the suggestion that cultural aspects were central drivers to motivate the rare interaction between Maghribī traders and Genoese traders, and absence of any partnership that involved delegation of trade-related operations, such as operating the sales of goods overseas.

Game-Theoretical Foundations To explain the absence of interaction, Greif uses a game-theoretical approach to explain historical institutional developments, interleaving historical facts with rational choice theory, an approach dubbed *Analytic Narratives* and explored by Bates et al. (1998). However, instead of applying static games, Greif (1994, 2006) turns to dynamic games that allow the representation of repeated sequential decision-making, and attempts to identify Nash equilibria that he interprets as *indicators for the existence of institutions* (though not as institutions themselves (Greif, 2006)). Nash equilibria (Nash, 1950) describe a publicly known strategy constellation in a non-cooperative game for which unilateral strategy change will not result in additional gain for any player. In *weak Nash equilibria* unilateral strategy change may result in the same outcome; if there is no unilateral strategy that offers the same outcome (i.e. strategy adaptation reduces the player's gain), we observe a *strict Nash equilibrium*. Only multilateral strategy adaptation could provide additional gain for one or more players. To model the trader scenarios, Greif employs a repeated variant of the *Prisoners' Dilemma* (PD) game[1], the iterated prisoner's dilemma. The prisoners' dilemma is inspired by the metaphor of interrogation of physically separated crime suspects. Either is questioned and challenged to admit the crime or blame his accomplice (defect). He can further deny the crime (cooperate with accomplice). The pay-off for the independently chosen strategies are determined by a pay-off matrix (see Figure 4.1 for the general case, with P_1 indicating best possible pay-off and P_4 being the worst), which would resolve to mild sanctions for cooperation (both suspects get a minor sentence or fine – represented as P_2, P_2 in this example). If one cooperates (denies the crime) and his partner defects (denounces his partner), the cooperator receives a strong penalty (long prison sentence) (P_4 in this example), while his defecting partner gets a minor punishment or walks free (pay-off P_1). In the case of both defecting, i.e. blaming each other, both receive a medium sentence (P_3), worse than if cooperating (denying the crime) (P_2), but better than the worst possible

[1]The *Prisoners' Dilemma* had been originally conceived by Merrill Flood and Melvin Dresher at RAND Corporation and was formalised by Albert W. Tucker (see Poundstone (1992)).

outcome P_4. For selfish rational players and the absence of loyalty, the preferable strategy in a single-iteration prisoners' dilemma is to defect, since the highest achievable reward (here P_1) is to blame the other, even though the most beneficial global outcome would be mutual denial. Since any unilateral strategy shift would promise worse individual pay-offs, mutual defection represents the strict Nash equilibrium in this scenario.

		Player 1	
		cooperate	defect
Player 2	cooperate	P_2, P_2	P_4, P_1
	defect	P_1, P_4	P_3, P_3

Conditions:
$P_1 > P_2 > P_3 > P_4$

Figure 4.1: Example Pay-off Matrix for PD Games

Iterating this scenario allows either player to base his strategy on previous experience, with the ability to react to strategy choices (e.g. by reciprocation or retaliation). For the iterated prisoner dilemma the suboptimal equilibrium is continuous defection, and the ideal strategy being 'tit-for-tat', i.e. initially cooperating, followed by adoption of the counterpart's prior strategy.[2]

Applying the iterated PD, Greif can thus overcome the limitation of static games (a single iteration game), and add the dynamic adjustment of individual strategies based on strategy choice in preceding iterations – in the Maghribī case based on cheater information.

Greif further adapts the iterated PD as a sequential *One-Sided Prisoners' Dilemma Game* (Greif, 2006), as opposed to a simultaneous strategy choice, given the dependence of an operating agent's strategy choice on the investor's choice and vice versa. In this scenario, a trading agent's choice depends on being employed in the first place; an employing merchant bases his decision to continue employment on the compliance of his employed operator.

The repeated adoption of the same strategy combinations by either player provides the foundation for an equilibrium-based approach to institutional analysis. If the players' strategy choices stabilise over repeated iterations, it identifies a subgame perfect equilibrium (Greif,

[2]This strategy was proposed by Anatol Rapoport and emerged as the winner of Robert Axelrod's (1984) subsequent tournaments of prisoner dilemma strategies proposed by game theorists, and tested in all permutations for 200 iterations.

2006). The stable strategy combination over repeated games represents a Nash equilibrium, given that no unilateral strategy adaptation based on information about previous iterations can lead to additional gain. Greif interprets this equilibrium as an indicator for an institution, given the alignment with Nash's conceptualisation of social institutions (Myerson, 1999). Returning to the scenario at hand and using the One-sided Prisoners' Dilemma game, Greif proposes the 'efficiency wage model' (Greif, 1994, 2006) which is at the centre of his work and described in the following.

Greif's Efficiency Wage Model He conceives a scenario with M merchants (or, in our conceptualisation: investors) and A agents (which we call operators), whereby A > M. All entities operate infinitely and have a time discount factor β. During each time period, merchants can hire an agent (with each agent being limited to work for one merchant at any time). Unemployed agents receive some *reservation utility* ϕ_u. The selection of agents is randomised, but can be constrained to the pool of unemployed agents and agents that performed particular action sequences in the past. A merchant that does not employ an agent, receives a pay-off of K during each iteration.

If a merchant employs an agent, he offers a wage payment W. The agent can then decide whether to cheat or be honest. Cheating causes the employer's pay-off to be 0, while the agent's pay-off is $\alpha > \phi_u$. If the agent decides to cooperate, the merchant's gain is $\gamma - W$, with γ being the overall pay-off for cooperation. In this case the agent's pay-off is its wage W.

The parameters include further assumptions:

- Cooperation between merchant and agent is efficient and thus results in added value ($\gamma > K + \phi_u$).
- Merchants prefer K over the experience of cheating or paying $W = \alpha$, i.e. $K > \gamma - \alpha$.

Figure 4.2 shows the one-sided PD scenario in a tree structure as suggested by Greif, with each leaf node showing the pay-offs for the merchant and agent respectively.

To address a shortcoming of conventional games – static parameters – Greif utilises the notion of quasi-parameters (Greif, 2006) – parameters that are static for a given iteration, but change between iterations. Players can thus adapt their strategy choices based on previous outcomes, reflecting a simplistic notion of learning. Greif's model inherently builds on quasi-parameters to allow a representation of endogenous institutional change, making the iterated games not only a requirement to represent behavioural regularity, i.e. institutions, but also to enable a notion of endogenous institutional analysis. After each interaction, both

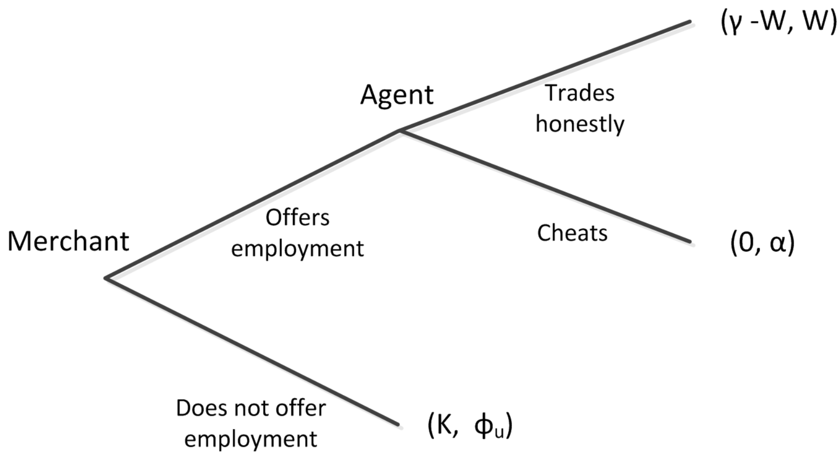


Figure 4.2: One-Sided Prisoners Dilemma Game

parties thus can decide whether to continue or terminate the relationship, and use previous knowledge to determine further actions. Apart from this intentional decision, with a probability σ ('forced separation'), the relationship ends due to unforeseeable exogenous factors. Likewise the adjustment of W can vary independent of the agent's previous conduct and is not directly dependent on external influences, such as courts or trade organisations. The history of interactions is shared knowledge.

With h_h as the probability of employment for an agent that was honest during the last employment period, and h_c as the probability for reemployment of a cheater, Greif (1994) formalizes the optimal wage W^* as $w(\beta, h_h, h_c, \sigma, \phi_u, \alpha) > \phi_u$, with w monotonically decreasing with β and h_h, while increasing with h_c, σ, ϕ_u and α.

In this model the wage plays a central role in inducing an agent's honesty, generally by being higher than the reservation utility. Wages that exceed the discounted lifetime utility of an unemployed agent (receiving the reservation utility ϕ_u) in comparison to the utility of an employed agent (receiving the wage W) motivate an agent's honesty. Consequently, this dynamic shifts if the gap between the discounted lifetime utility of a cheater versus non-cheater becomes $< W - \phi_u$, i.e. if the long-term pay-off for being unemployed is higher than the one for being employed. From an employer's perspective, cheating is sanctioned with immediate termination of the employment relationship.[3]

It is at this stage that Greif (1994) introduces the central bifurcation between collectivistic and individualistic societies, suggesting inherent communication of trade experience among collectivistic traders, while assuming inherent secrecy about trade performance in

[3] A model variant that removed the assumption of immediate relationship termination but instead introduced an option for compensation for cheating (as documented by Greif (1993)) has been introduced by Harbord (2006).

individualistic societies.

In his game-theoretical model, this is reflected in the consideration of an agent's past conduct prior to employment. In collectivistic societies with consideration of trade history, and the assumption that all merchants of a given society apply the same strategy, past cheaters are less likely to find employment. Traders that follow the individualistic strategy of non-communication are indifferent about an agent's past behaviour; the probability for reemployment does not depend on his past behaviour. From the perspective of the employee, his actions are never sanctioned, comparable to the one-shot game. In the absence of contractual enforcement, Greif's model thus suggests continuous defection.

From this inherently culture-based differentiation, Greif draws the conjecture that individualistic merchants need to offer a sufficiently high wage, thus maximizing the gap between wage and reservation utility, to induce cooperation with individualistic agents, whose past conduct is otherwise irrelevant for employment. In collectivistic societies, in contrast, a lower wage (and thus smaller gap between wage and reservation utility) is sufficient to sustain cooperation, given that cheating behaviour is public knowledge. The agent thus needs to consider his reputation along with earned wage when deciding whether to act compliantly.

Greif's model thus suggests two central aspects: The reliance on the informal information sharing was sufficient for the Maghribīs to sustain cooperation. In addition the wage level for compliant trading is lower compared to individualistic societies, in which agents would need to be incentivised by higher wages in order to maintain compliance. The higher wage level thus makes the employment of individualistic agents unattractive for collectivistic merchants since they need to pay higher wages compared to their fellow collectivistic merchants.

Following this line of thought, Greif explains why the Maghribī traders, despite the presumed absence of institutional boundaries, did not engage in long-distance trade relationships with Southern European traders as supported by historical evidence.

The underlying assumptions are thus twofold. The collectivistic society is of closed nature, and all members adopt the same strategy for trader employment. The individualistic model is an archetype of an open society, in which the past conduct of individuals, as with newcomers, is unknown. Greif's argument for the non-integration of both trader societies is thus inherently based on their cultural difference – secretive individualists vs. sharing collectivists.

Greif introduces further models to support his conclusions (see Greif (2006)). However, we are particularly interested in the assumptions that underlie Greif's efficiency wage model and their mapping onto historical social reality. In the following we discuss Greif's model in the light of the historical background we have provided in the previous chapter (Chapter 3).

4.1.2 Reviewing Greif's Model with Respect to Historical Reality

Greif's model of the Maghribī Traders Coalition and the contrasting individualistic society, along with more detailed information on the historical background, provides the necessary foundation for our work, which looks at various assumptions that are central to Greif's work.

The innovative aspect of Greif's modelling approach is to provide a formal model of the trader scenario, while enabling dynamic change of preferences necessary to model endogenously changing institutions – an aspect extending beyond the conventional use of game theory.

Underlying Assumptions Inasmuch as Greif's work overcomes some of the limitations of the comparative-static notion of game theory, such as substituting the static parameter set with quasi-parameters, it still relies on a strong abstraction from social reality for the two respective societies. In Greif's model this abstraction is subsumed by introducing relatively generic assumptions, such as cultural differences. This motivation is retraceable from the perspective of comparative economics, with its inherent intent of facilitating a high-level institutional comparison, which could be challenged by considering more fine-grained aspects of the respective societies.

In the context of his detailed and insightful historical analysis, in conjunction with more recent findings on the Maghribī case (see Section 3.1), some of Greif's model assumptions appear overly simplistic and invite further investigation.

We are particularly concerned with three central assumptions laid out by Greif (1994, 2006) which we alluded to in Subsection 1.3.1. At this stage we can fully contextualise those assumptions with respect to the historical background and the applied analytical method, since they accompany us for the remainder of this thesis. Greif's assumptions include:

- Cultural differences as a central differentiation between both societies, which is represented as either complete information sharing or non-communication;
- Strict assumption of 'closedness' for the Maghribī trader collective;
- Neglecting to consider the role-integrated nature of the Maghribī society vs. the role-stratified Genoese society.

Firstly, Greif assumes inherent cultural differences between the Maghribī and Genoese trader societies which he reduces to the bifurcation of communicating vs. non-communicating partners. Concerns about the generic nature of such assumptions have been discussed in Subsection 3.2.1, with particular focus on the critique offered by Edwards and Ogilvie (2012), who claim the limited relevance of cultural differences in producing different institutional outcomes. Even when committing to the cultural differences between both societies, Greif's translation into a dichotomy of non-communicators vs. communicators is indirectly challenged by Epstein (1994). As an essential source supporting the secrecy of the Genoese traders, Epstein softens that argument by emphasising the nuisance and uncertainty involved when engaging commercial courts to enforce contractual claims. We will explore this aspect further in Subsection 4.2.1.

The second assumption Greif posits for the persistence of the Maghribīs' coalition is their strict operation as a closed group (Greif, 1989, 1993, 2006) based on a common identity that limited involvement of outsiders in trade activities. More recent research into the Geniza records, in particular by Goldberg (2012c), has shown that the coalition was not as closed as previously assumed, but could also include 'outsiders'. Instead of entirely relying on a common identity (as members of the Jewish communities of al-Mahdiyya and Qayrawān), the persistent interaction with members of the aṣḥābunā was more decisive to gain increasing access based on established reputation (jāh), an aspect Greif acknowledges in his rebuttal to Edwards and Ogilvie's claims (Greif, 2012) (see Subsection 3.2.2). This overly rigid assumption is associated with Greif's conception of the traders coalition as a closely-knit group, and thus reliant on the individuals' commitment to the group in its entirety. As discussed in greater depth in Subsection 3.2.2, Maghribīs maintained networks of dyadic relationships within their individual aṣḥābs, from which the aṣḥābunā, or in Greif's terms, coalition, emerged. Given the composition of the coalition from smaller, more coherent sub-groups, the 'closed group' assumption can be relaxed, since the coalition in its entirety could likely not even be conceived by individual agents. We discussed this aspect at length in Subsection 3.2.2 and will explore it experimentally in Subsection 4.2.2.

The last assumption, possibly offering the strongest deviation of his model from provided historical facts, is Greif's failure to consider the differing role structures of both trader societies. The Maghribīs were characterised by what we can conceive as a role-integrated social structure, i.e. members of the aṣḥābunā acted both as what Greif calls 'merchant' and 'agent' (Greif, 1994, 2006, 2012) (see Subsection 3.3.2). In fact the system of reciprocal obligations based on the ṣuḥba inherently relied on this integrated perspective in order to satisfy outstanding commitments of service provision. The Genoese society, in contrast, satisfied the assumption of Greif's model – the differentiation into a merchant class and agent class – with the number of agents vastly exceeding the number of merchants (see Subsection 3.3.2). We believe that this deviation from historical reality bears a strong potential to redefine the understanding of why Maghribīs cooperated, and why Genoese traders had lower incentives to do so. This aspect will be the subject of exploration in Section 6.3 (based on the conceptual contributions introduced in Chapters 5 and 6).

Analytical Approach We believe that Greif's simplified representation is at least in part related to his analytical method and the comparatist perspective. This includes a) the demand for rigid bifurcated formal specification, b) a focus on a rational economic perspective without taking social factors sufficiently into account, and c) providing insufficient or incongruent modelling constructs for real-world entities such as specific institutional instruments.

The rigidity demanded by conceiving scenarios as games limits the degree to which other detailed social factors, beyond the obviously relevant economic aspects, can be represented and explored. This appears particularly striking when modelling complex aspects such as cultural differences.

Greif does his best to consider social aspects he recognises during the literature analysis in his model. In fact he devotes considerable space for the argument of his method as well as the appropriateness to represent institutional equilibria, supporting its suitability by the very fact that he intentionally ignores the roots of such institutions, whether based on learning or intentional design, given that his methods consistently reveal the equilibria he sees as indicators for an institution (Greif, 2006; Aydinonat, 2006). The limited representation of the detailed underlying social processes thus raises the question whether his method of analysis is appropriate to model truly endogenous institutions. Extending that question, apart from the iteratively adjusted quasi-parameters, how can his method reveal sufficient insight to suggest that an institution is a result of endogenous processes if it withdraws itself from inspection? In Greif's model social processes of learning and the relationship between micro- and macro-level – the very processes that drive an individual's and institutional environment's formation (see Subsection 2.2.2) – are of limited relevance. Evolutionary game theory that could accommodate the dynamic aspects better is dismissed as an alternative (Greif, 2006).

Though alluding to the *homo sociologicus*[4] in his work, both the chosen method and intention inherently reflect a *homo economicus*, following the assumption of perfect rationality and focusing on selfish gain, an aspect that has been challenged by economists such as Selten (2001).

Let us use one example to highlight the inaccurate representations resulting from the chosen analytical method. Greif's means of analysis forces him to conceive the provision of services as a 'wage' that is paid for an agent's 'employment' by a merchant. If not anything else, the literature background provided in the preceding Subsections 3.2.1 and 3.3.1

[4] See discussion in (Greif, 2006).

gives us insights into the fundamentally differing motivations of members of the respective trader societies. It is true that Southern European traders' profit orientation was doubtlessly reflected in money as meticulously accounted for by Giovanni di Scriba and his successors. But the *true currency* of Maghribīs, in contrast, was their reputation, their jāh, an aspect that is reflected in the remuneration-free provision of services, the value of which was, again, determined by the partners' respective jāh. Maghribī traders would abandon profits to prevent mere suspicion of non-compliance (see Subsection 3.2.2, Greif (1989)).

One could argue for the 'wage' concept as a possible operationalisation from an economic perspective, but from a historical perspective it is inherently inaccurate: paying a fellow trader for his services would have been conceived as 'slave-like' (Goldberg, 2012c). Yet, only the conceptualisation of wages allows Greif to suggest different wage levels for collectivistic and individualistic societies, an outcome that is the product of game-theoretical modelling. It seemingly explains the lack of cross-cultural cooperation, but builds on assumptions that represent a considerable gap from historical reality.

Using this example, we can see how, apart from simplified assumptions, the analytical method can impact the outcome. We do not suggest that the outcome or the modelling is inherently wrong, but given the absence of wage payment in the Maghribī society, the validity of such outcome is hard to support.

In this work – and in particular for the later part –, we put stronger emphasis on the social aspects and choose a modelling metaphor that we believe offers a closer representation of the modelling target, the scenario of interest. We believe Agent-based Modelling and Simulation (Epstein and Axtell, 1996) (see Subsection 2.3.1) is a suitable candidate to represent more refined scenarios. With its stronger emphasis on the emergent properties based on the social interaction among individuals and the separation into micro- and macro-level properties, it allows for a more realistic representation of social processes on different levels of analysis. Like Greif's approach, ABM is not flawless and bears the similar risk of overfitting the model to phenomena of interest and modelling at varying granularity levels. It further challenges the modeller not to lose reference to the 'bigger picture', especially when attempting to model comparative scenarios as intended. However, for the problem discussed here and the necessary relaxation of assumptions that deviate from historical facts, the refined modelling capabilities make it an appropriate choice.

In the following we will concentrate on the two initial assumptions, information sharing vs. non-sharing societies as well as the closed nature of the Maghribī society, and explore specific properties using agent-based models.

4.2 Exploring Selected Characteristics for Both Societies

In this section we explore scenarios that have found limited or no consideration in Greif's analysis, or were expressed in overly rigid assumptions. As a first aspect we relax the strict assumption of non-communication among Genoese traders, and replace it with refined information strategies in order to test how far the consideration of communication among the Genoese could have affected the cooperation outcome. This particular experiment has previously been published in Frantz et al. (2014a). We present it here in an extended form.

4.2.1 Experiment 'Informal Communication in the Genoese Trader Society'

Motivation

A central assumption of Greif's model is the assumption that Genoese traders – in opposition to Maghribī traders – did not share any information about their individual dealings until eventually codified in the form of contracts that documented trader relationships. However, this categorical view can be challenged, given that court procedures were cumbersome, time-intensive, and offered the risk of uncertain outcomes (Epstein, 1994). Epstein's reading leaves us with the impression that public-order enforcement would offer a 'backup' plan rather than being the norm. Only if informal means failed to resolve arising disputes or if the mere threat of engaging courts proved ineffective, drawing on legal instruments appeared as a necessary step. Greif (1994, 1998, 2012) himself supports the fact that private-order enforcement complemented the public-order enforcement prevalent in the Genoese case. With reference to Greif, González De Lara (2008) further contrasts the Genoese society as one relying on mixed private-/public-order enforcement, in contrast to the Venetian trader community that nearly exclusively relied on formal dispute resolution mechanisms, positing Genoa before the 13th century as an intermediate case of complementary use of private- and public-order enforcement, flanked by the Maghribī society on the private-order end, and Venice marking the public-order end of that continuum. Edwards and Ogilvie (2012) further highlight a general preference to solve conflicts informally, independent of the economy.

Another aspect that could have motivated the use of informal mechanisms was the potentially limited reach of public-order enforcement mechanisms, given the involvement of a considerable fraction of foreigners, sponsoring around 18.3 percent of the capital in trade activities (González De Lara, 2008; Greif, 2006). In his analysis of the Genoese commenda relationship network (from 1154-1315 AD), van Doosselaere (2009) further finds that in 10 percent of all commendae, multiple investors acted as co-investors. This fraction is arguably small compared to 86 percent of ties strictly enacted between investor and operator (van Doosselaere, 2009). Despite the presence of such formal ties, the existence of informal relationships appears likewise realistic, at least to the extent that informal negotiations would precede a formal relationship.

The "clannishness" (Lopez, 1982), a trademark of the Genoese society, offers ground for further claims to appoint fellow clan members for enforcement (private-order enforcement) before drawing upon public-order mechanisms for this purpose.

Leaving historical evidence and intuitions aside,[5] we can argue for the existence of communication from a rational perspective.

Sharing information is discouraged from an *individualistic* perspective where the sharer is indifferent about the requester's fate, as long as he does not receive information he could capitalise on for his own benefit, such as communicating the non-/compliance of employees. However, from a *selfish* perspective the bearer of information could benefit from sharing that information by sharing it in a strategically deceitful fashion. Instead of consistently responding to requested information in a truthful or wrongful manner, the respondent could strategically spread information that could be to the disadvantage of the recipient, assuming that the latter accepts it as truthful. Assuming the existence of a pool of operators[6] potential investors can source from, information about potential cheaters puts the bearer of such information into an advantageous position compared to his competitors that lack this knowledge. Intentionally misinforming a fellow investor about the experience with a specific operator thus goes beyond the *maximisation of one's own benefit*, i.e. maintaining secrecy about such information, but can also be used to *minimise one's fellow investors' information gain* by leaving them in doubt about the truthfulness of shared information, (mis)leading them from/to repeating the costly experience of being cheated.

Model

With this thought in mind, we construct a model that allows us to test how far *communication among Genoese investors* could have affected the cooperation outcome. Following Greif's analysis, the absence of communication led to a breakdown of cooperation when exclusively

[5] Naturally, private-order enforcement is evidenced by poorer documentation and relies, similar to Goitein's reconstruction of the Maghribī society from isolated narratives, on anecdotal accounts.

[6] As indicated before, the term 'operator' signifies the operating party of a trade relationship, i.e. the trader performing the actual trader, or tractator in the context of a commenda relationship. We use this instead of 'agent', a term that is confusing when choosing a software 'agent' as a modelling metaphor. The operator's counterpart, Greif 'merchant', will be referred to as 'investor'. The term 'trader' is used as substitute for either role.

concentrating on informal enforcement mechanisms. Apart from the extent to which communication is decisive, the cooperation outcome is also determined by the truthfulness of reporting. Randomised or consistent lying will not lead to compliance based on informal means as long as an individual does not adjust its belief about the purposefulness of truthful reporting. We thus make truthfulness of reporting a function of trust in others to allow its adjustment and reinforcement based on experiential input.

We reconstruct the essential elements of the Genoese trader community, with the most tractable historical evidence being the investor-operator ratio which González De Lara (2008), on the basis of Greif's sources (Greif, 1993, 1994, 2006), allocates at 1.57 (see Subsection 3.3.2).

Given the central role of the investor in this model, we will concentrate on its actions. Each investor initially picks an operator from the pool of available operators. If he has previously memorised this particular operator as a cheater, he continues to pick an alternative operator, an approach he can repeat *maxNumberOfMerchantPicks* times, before picking the highest memorised operator. This random choice of operators is grounded in the investors' inability to predict the availability of a preferred operator due to the duration of journeys and potential alternative contractual arrangements. Compatible with Greif's model (Subsection 4.1.1), and unlike Maghribīs, Genoese operators could not expect repeated employment despite existing contractual relationships. Not employing any operator, on the other hand, would misrepresent the opportunistic nature of Genoese investors.

Investors can determine the performance of their employed operators, but in addition they also keep a memory about the truthfulness of their fellow traders. They do so based on the probability p_{test} and determine whether a fellow trader's prediction matches the observation for a given employed operator, i.e. they test whether an operator reported as compliant turns to out act accordingly, and they store information both about the truthful reporting of the advisor as well as the operator of interest. If the requestee has experience with the subject of inquiry, he has the choice to advise truthfully or to lie about his experience. To maximise the likeliness of eventually receiving advice from fellow traders that had experience with the operator of interest, requests can be extended to other investors either until advice has been returned, or a maximum number of requests (*maxRequests*) has been sent. In the case of picking an unknown operator, inquiries about his conduct are sent and followed in any case, i.e. employment is established if advice is positive, and abandoned if advice identifies the candidate operator as a potential cheater.

Memory about both cheaters and fellow investors is finite in order to represent an agent's bounded rationality (Simon, 1955). For this reason, a parameterised number of last memory entries following *first-in first-out* semantics are available for decision-making, with oldest entries being overwritten by new information. Updated information about other operators or advising fellow investors that are already held in memory is merged and treated as new information. Memory influences the agent's decision-making, as it guides partner choice in cases where all picked candidate operators had been previously memorised as cheaters, and thus chooses an operator he had best experience with. Memory entries are represented as `<agent name>,<experience>` pairs, in which experience is represented as 1 for compliant behaviour and as -1 for cheating.

In contrast to the more complex investors, operators are instantiated as either cheaters or non-cheaters and react accordingly to incoming employment requests.

Figure 4.3 visualises the essential structure of the scenario laid out here. We dedicate a more detailed discussion to the intuitions that underlie the dynamic adjustment of trust levels in this scenario.

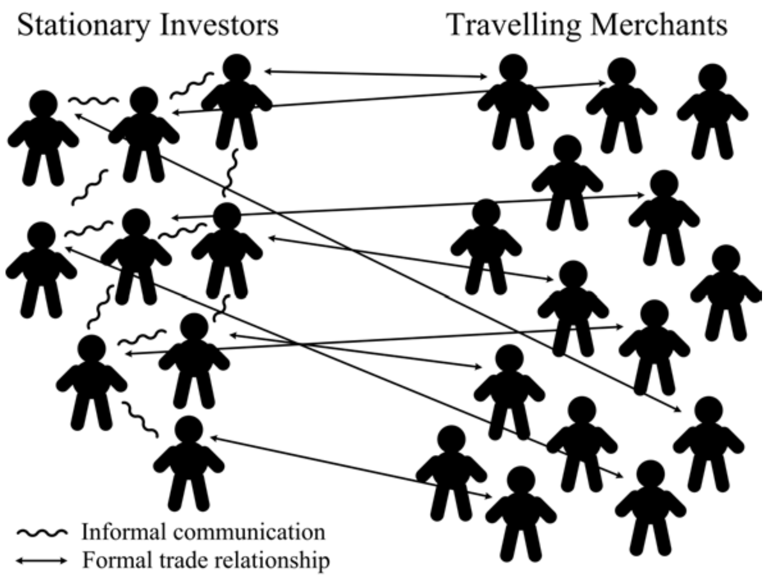


Figure adapted from Frantz et al. (2014a)

Figure 4.3: Communicating Traders Scenario

Dynamic Adjustment of Trust Recalling the central objective – to determine whether low levels of communication would suffice to establish trust that could have promoted the functioning of informal institutions – we require some representation of 'trust in the system', describing the extent to which an investor believes that other investors advise truthfully, loosely representing a notion of convention (see Subsection 2.2.2). We represent this using the variable $p_{truthful}$, which an investor dynamically adjusts based on his experience with fellow advisors and which in turn determines its tendency to report truthfully to future requests addressed to him. Inspired by Gambetta's trust definition (1988), we can interpret trust as an "individual's subjective probability that other individuals perform an action in compliance with the subject's expectations" (Frantz et al., 2014a). To determine such trust, the investor has to test the advisor's truthfulness, firstly to establish whether he can be referred to for future requests, and secondly to prevent a) misdeclaration of non-cheaters as cheaters (false positives) from reducing the employment pool, or b) extensive cheating experience (based on false negatives). The notion of trust represented here relates to the overall behavioural convention an individual investor perceives, but does not imply direct notions of reciprocity, such as 'tit for tat' (see Subsection 4.1.1), since the investors act based on $p_{truthful}$, not based on the memory they hold about individual fellow investors.

The truthfulness of such reporting not only influences the investor's understanding and drives his own response, but also bears an economic purpose. Establishing an institutional understanding, such as a convention of non-/cooperation based on experience, goes alongside a reduction of uncertainty about those conventions. With increasing trust in such conventions comes a potential increase in efficiency (Williamson, 1998), expressed in reduced transaction costs, such as the reduced need to monitor compliance within established relationships. In this case, we thus associate an increased trust level with an inversely related propensity of testing the truthfulness of advisers and thus increasingly rely on their advice. The probability of $p_{truthful}$ is thus conceptualised as a probability, of say 0.5, to which p_{test} is inversely related. Individual instances of advice established as correct or misleading increase or decrease $p_{truthful}$ by δ_{trust}.

Complementary to the trust investors gauge based on the actions of their fellow investors, the exchange of information related to individuals' conduct represents a reputation mechanism (see Conte and Paolucci (2002)) implicitly reflecting an operator's standing within the modelled society.

Algorithm 4.1 outlines the investor's execution cycle, and Algorithm 4.2 describes his reactions to incoming advice requests.

Table 4.1 shows the base parameter set that is applied to the model. A central figure that we derive from the cartulary information of Giovanni di Scriba explored by Greif (1993, 2006), which lists 612 contracts, is the involvement of 180 investors in contrast to 335 operators. The limited availability of information about actual values for further parameters (such as memory length and number of requests) forces us to perform a sensitivity analysis,[7] the resulting parameter values of which we describe in due course. Before turning to

[7]Since the sensitivity analysis described in Appendix C.1 relies on further performance metrics introduced

Algorithm 4.1: Genoese Investor Execution Cycle

```
// Establishing Employment based on Experience & Advice
Choose random operator;
if investor has memory about operator then
    if operator is memorised as cheater then
        if number of pick attempts < maxNumberOfMerchantPicks then
            Choose another operator randomly;
        else
            Choose memorised operator of which the investor has highest memory value (i.e. best experience with);
        end
else
    Ask for advice from other related investor;
    if no advice on operator OR experience with advising investor negative then
        Try other investor as long as requests for advice < maxRequests;
    Determine whether to test cheater behaviour and advice based on random boolean testing with probability $p_{test}$;
    if !testing then
        if advice negative then
            Choose new random operator;
    else
        // Prepare testing of advisors
        Memorise advisor and operator recommendation (whether indicated as cheater or not);
    end
    Employ operator;
end
// Evaluation of Employment & Advice
Observe outcome of employment;
if employment was based on advice then
    if operator did not cheat despite negative advice then
        Memorise operator positively;
        Memorise advisor negatively;
        Decrease $p_{truthful}$ by $\delta_{trust}$;
    else
        if operator cheated despite positive advice then
            Memorise operator negatively;
            Memorise advisor negatively;
            Decrease $p_{truthful}$ by $\delta_{trust}$;
        else
            // Operator behaved according to advisor's predictions
            Memorise advisor as positive;
            Increase $p_{truthful}$ by $\delta_{trust}$;
            if operator cheated then
                Memorise operator negatively;
            else
                Memorise operator positively;
            end
        end
    end
else
    // i.e. employment not based on advice but on random choice
    Memorise cheating operator negatively, non-cheating operator positively;
end
```

in the following, we allude to it at a later stage.

Algorithm 4.2: Investor Advice Response

```
Determine whether to respond truthfully (with probability p_truthful for truthful response);
if no information about operator then
    Do not give advice;
else
    if decided to respond truthfully then
        Answer request according to own information about operator;
    else
        Invert information about operator (report non-cheater as cheater and vice versa);
    end
end
```

this aspect, the limited information with respect to a significant parameter – the underlying network structure – leads to additional considerations.

Table 4.1: Simulation Parameters

Parameter	**Value**
Number of Investors (n_{inv})	180
Number of Operators	n_{inv} * 1.57
Initial Probability of Truthful Reporting ($p_{truthful}$)	0.5
Probability for Testing Advice (p_{test})	1 - $p_{truthful}$
Maximum Number of Advice Requests (*maxRequests*)	10
Length of Memory (i.e. Number of Entries) (*memoryEntries*)	40
Quota of Cheating Operators (*cheaterQuota*)	0.4
Trust incr/decrement (δ_{trust})	0.005
Mean Number of Relationships (*avgRelationships*)	40
Max. Number of Operator Picks (*maxNumberOfMerchantPicks*)	5

Relationship Networks among Genoese Investors Exploring the cartulary information, a striking feature is the strong concentration of capital on very few investors. From the 180 investors in Giovanni di Scriba's cartulary, only 37 noble families sponsored around 90 percent of the capital (Greif, 1993, 2006; González De Lara, 2008). This strong concentration of the network structure leads to the suspicion that topologies other than the random network could have been candidates for a more realistic representation of the hypothesised interrelationships. A systematic analysis for commenda relationships in the different cartularies has been performed by van Doosselaere (2009). However, in this context the increasing concentration of such relationships is of limited value, inasmuch as commenda relationships described the contracts *between investors and operators*, but not *between investors* – which we explore in this model.

In fact the structure could be guided by information hubs, which required a varying connectedness among all network nodes, and could have facilitated faster information spread, but also greater influence with respect to the overall cooperation outcome. Network structures of this type have been observed to reflect the hyperlink structure of the world wide web (Clauset et al., 2009). Alternatively, the investor network could have been structured by smaller communities of more densely connected individuals (comparable to the Maghribīs' aṣḥābs) that are interlinked by fewer relationships. Networks of the latter nature have been established as characteristic for human social networks (see e.g. Milgram (1967)), and could have likewise been representative for the clan structure that dominated the Genoese society (see Subsection 3.3.1, Greif (2006), van Doosselaere (2009)).

Given our intent to develop an overall – and not necessarily precise – understanding[8] of the scenario, we provide an extended view to address the valid concern that differences in the cooperation outcome based on the hypothesised communication could have been grounded in the underlying network structure.

Instead of simply relying on random contact between investors for advice requests, we additionally route requests through pre-generated networks of different topologies. Considered network topologies include:

- *Fixed randomly assigned relationships* (FXD), represented as *directed random networks*;
- *Fixed randomly assigned mutual relationships* (FXDM), represented as *undirected random networks*;
- *Small-world networks* (WS), generated using the Watts-Strogatz algorithm (Watts and Strogatz, 1998);[9] and finally,
- *Scale-free networks* (BA), generated using the Barabási-Albert algorithm (Barabási and Albert, 1999).

In FXD networks, individual investors maintain relationships to other traders based on a fixed number (here: *avgRelationships*) of initially assigned relationships. This scenario offers the extreme case of randomised interconnections among traders, the existence of which can be challenged as it implies that in many cases a requesting investor may not be considered as a potential advisor, despite mutual acquaintanceship. This exists in contrast to the

[8]We share this intent with Greif, whose modelling targeted *understanding*, as opposed to *accuracy*, with respect to the underlying quantitative information (Greif, 2012).

[9]Note that for the latter two network types we use widely adopted algorithms that carry the respective properties. Further more refined algorithms are discussed by Newman (2010).

FXDM configuration in which relationships are mutual, i.e. linked investors mutually consider each other as potential advisors. This network implies a reduction of links by half to establish the same number of advisor relationships.

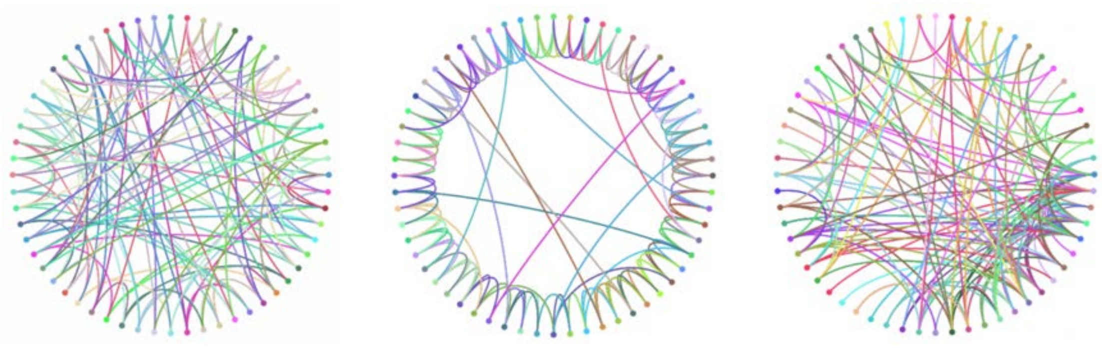
In the light of lacking sufficiently detailed information about the relationship network, the WS configuration tests the effect of investor groups that have tight relationships, with fewer links interconnecting those groups, or communities. The Watts-Strogatz algorithm (1998) initially constructs a ring lattice and connects a node with a specified number of neighbouring nodes (here: *avgRelationships*), before iterating over all generated links and rewiring them with another randomly selected node (to which the source node does not already have a link) with a given *rewiring probability*. To explore the potential implications of more strongly interconnected communities, for WS networks we will explore a range of different rewiring probability values.

Scale-free networks such as the ones produced using the Barabási-Albert algorithm (1999) generate a variation of node connectedness that follows a power-law distribution. This algorithm applies the concept of preferential attachment of newly created nodes with existing, more strongly connected nodes, thereby creating networks which are characterised by a few strongly connected hubs that interconnect a large number of nodes with weaker connectivity. Nodes have an average number of relationships of *avgRelationships*, making them comparable with all previously mentioned network types.

The latter network types, WS and BA, are modelled with undirected links in order to avoid isolating poorly connected nodes at the long tail of the link distribution. Given the unequal assignment of links (and algorithmic aspects such as the reassignment of those already assigned based on the rewiring probability for Watts-Strogatz networks), a node could potentially only be connected by a single incoming link, preventing access to any advisor.

To establish the degree to which different network structures affect the cooperation outcome, we evaluate the different network types with respect to a baseline scenario that does not involve any communication among investors (which we tag as *NoComm*). This baseline configuration effectively represents Greif's scenario of strict non-communication.

Figures 4.4 show the prototypical structure of the explored network types. The circular organisation emphasises the respective characteristics, such as the fully randomised structure of random networks (Figure 4.4a), the strong bias towards adjacent nodes in small-world networks (Figure 4.4b), and the bias towards strongly connected nodes (preferential attachment) in scale-free networks (Figure 4.4c).



(a) Random Network (b) Small-World Network (c) Scale-Free Network

Figure 4.4: Network Types

Measures of Performance The central interest is to identify whether latent communication based on advising requests among investors could have made a difference in excluding cheaters from trade relationships. The dependent variable of any measure is thus the fraction of employed potential cheaters, the baseline for which is 0.4 as defined in the initial parameters (see Table 4.1). In addition, we require metrics of performance that allow a more refined comparison of different parameter configurations – in this case, initial trust levels and network types.

Such measures include the extent to which cheaters, or *how many*, are excluded from future employment, which we label the *Effectivity* of a given parameter set. In addition to this quantitative measure we further measure the *Efficiency*, i.e. *how fast* this occurs.

To establish both measures, throughout a simulation run we measure the ratio of employed potential cheating operators, denoted as $q(r)$, with r representing the respective simulation round. Assuming that this ratio stabilises over the simulation run with minimal further change (an aspect that has been established during initial tests), the *effectivity* is thus represented as the difference between the initial fraction of potential cheaters and the final level of employed potential cheaters. Given that the fraction of excluded cheaters is a function of the established trust level, i.e. the extent to which information about cheaters is truthfully exchanged, we determine the cheater level as the mean value of the fraction of employed cheaters throughout all rounds following the establishment of a stable trust level. A trust level is considered stable, once the trust level reaches and remains within a tolerance $\delta_{tolerance}$ of 0.001 of the maximum trust level measured across the entire simulation run. With q_{min} indicating the minimal fraction of employed potential cheaters (and thus highest inter-investor trust level), the number of rounds to reach such trust level is defined as

$$r_{stableTrust} := min(\{r \mid q(r) \leq (q_{min} + \delta_{tolerance})\}) \tag{4.1}$$

Based on this, we can operationalise *effectivity* as the fraction of employed potential cheaters in the first round $q(1)$ reduced by the mean of employed cheater fractions following the establishment of a stable trust level divided by $q(1)$, and is formalised as

$$effectivity := \frac{q(1) - \frac{\sum_{r=r_{stableTrust}}^{r_{max}} q(r)}{r_{max} - r_{stableTrust}}}{q(1)} \tag{4.2}$$

Effectivity values thus range between 0 to 1, indicating the fraction of removed cheaters, with 1 signifying the exclusion of all cheaters from potential relationships, and 0 indicating failure to remove any potential cheater from employment.

In contrast to effectivity, the quality measure *efficiency* is directly derived from the number of simulation rounds necessary to reach a stable trust level. With r_{max} indicating the maximum number of simulation rounds and normalised to represent proportionally increasing efficiency, it is defined as

$$efficiency := 1 - \frac{r_{stableTrust}}{r_{max}} \tag{4.3}$$

In this operationalisation, 1 represents the highest possible efficiency, i.e. the immediate convergence to the stable trust level – or simply no change in the trust level throughout all simulation rounds; 0 indicates that the trust level did not converge to a stable level throughout all simulation rounds.

Having established those measures, we can evaluate simulation configurations both from a quantitative and qualitative perspective. Removing all cheaters from trade relationships only at the simulation end (high effectivity, low efficiency) may not have been as useful as removing fewer cheaters but doing so quickly (low effectivity, high efficiency).

At this stage we establish the parameter settings for the model parameters memory size (*memoryEntries*) and maximum number of advice requests (*maxRequests*) based on sensitivity analysis as described in Appendix C.1. The results identify the initial probability of truthfulness ($p_{truthful}$) as the essential independent variable. Since both memory size and number of requests have weak correlations with the outcome measures effectivity and efficiency (see Appendix C.1), we set the number of memory entries to 40, with the assumption that investors would have knowledge about around one fifth to one quarter of their co-investors. The maximum number of requests is further fixed at 10.

At this stage we can turn to the evaluation of simulation results for the introduced scenario. Information about the simulation runtime environment for this and all subsequent models can be found in Appendix B.

Results for the Genoese Model

For the simulation runs we varied the *initial truthfulness ($p_{truthful}$)* for all network types and ran each simulation for 10,000 rounds, a duration during which the level of trust stabilised for all configurations. During initialisation, an individual agent's $p_{truthful}$ is assigned based on a random draw from a normal distribution centred at $p_{truthful}$.

For each configuration we performed 30 runs and discuss the results shown in Table 4.2 based on the statistical means. In addition to the exploration of the different network types, we ran the simulation with the baseline scenario in which Genoese traders do not communicate at all (marked as network type *NoComm*). In this case removal of cheaters from the network fully depends on the individuals' memories, which allows us to discriminate memory from network effects.

A further aspect that demands attention is the parameterisation of the Watts-Strogatz algorithm, since it relies on the rewiring probability as an additional parameter in order to construct interlinked communities. In addition to simulation runs for all other network types, for WS networks we tested the results across the rewiring probability values 0.05 and 0.1 to 0.5 (with step size of 0.1) and all levels of initial truthfulness. The results are shown in Tables C.7 (Effectivity) and C.8 (Efficiency) in Appendix C.2. The impact of varying rewiring probabilities is reflected in the standard deviation (column σ) in Table C.7, which has a maximum value of 0.016 (for initial truthfulness of 0.5). Since this value is well below standard deviation measures for all network types (in Table 4.2), choosing any of the rewiring probabilities would have minimal impact on the overall simulation outcome, and no impact on the performance-based *ranking* of different network types. For the general representation of WS network results, we calculated their performance as the mean of the simulation results across all explored rewiring probabilities. In addition we tested the significance for all network configurations compared to the baseline (*NoComm*) using the Mann-Whitney-Wilcoxon test (MWW) (Mann and Whitney, 1947) (at a significance level of 0.05). The complete result tables including the individual rewiring probabilities and p values are provided in Tables C.11 and C.12 in Appendix C.3.

For low initial levels of truthfulness, we can observe moderate effectivity improvement for fixed network types with equal link distribution (FXD and FXDM) of around 5 percent, compared to the baseline scenario (*NoComm*). Significantly better performance is realised in networks with unequal link distribution (WS and BA), realising performance benefits of more than 15 percent for $p_{truthful}$ at 0.5. Looking at efficiency values refines that picture. However, efficiency ('how quick') is only meaningful in conjunction with effectivity ('how much'). An indicative example is the baseline scenario. Here, based only on individuals'

Table 4.2: Simulation Results for Informal Communication among Genoese Traders

Initial Truthfulness	**Network Type**	**Effectivity**		**Efficiency**	
		mean	σ	**mean**	σ
0.5	NoComm	0.195	0.019	1	0
0.5	FXD	0.25	0.035	0.02	0.019
0.5	FXDM	0.241	0.034	0.013	0.007
0.5	WS (mean)	0.361	0.039	0.016	0.006
0.5	BA	0.334	0.069	0.161	0.341
0.51	FXD	0.276	0.038	0.068	0.063
0.51	FXDM	0.26	0.037	0.027	0.018
0.51	WS (mean)	0.381	0.035	0.023	0.01
0.51	BA	0.36	0.043	0.068	0.18
0.55	FXD	0.364	0.033	0.531	0.105
0.55	FXDM	0.334	0.033	0.411	0.143
0.55	WS (mean)	0.419	0.029	0.161	0.085
0.55	BA	0.395	0.028	0.112	0.058
0.6	FXD	0.457	0.028	0.777	0.029
0.6	FXDM	0.414	0.03	0.748	0.03
0.6	WS (mean)	0.452	0.029	0.568	0.094
0.6	BA	0.424	0.029	0.379	0.119
0.7	FXD	0.558	0.028	0.909	0.015
0.7	FXDM	0.517	0.028	0.906	0.012
0.7	WS (mean)	0.508	0.028	0.874	0.023
0.7	BA	0.476	0.025	0.78	0.044
0.8	FXD	0.605	0.025	0.959	0.006
0.8	FXDM	0.581	0.025	0.957	0.007
0.8	WS (mean)	0.56	0.028	0.951	0.007
0.8	BA	0.52	0.031	0.909	0.026
0.9	FXD	0.614	0.025	0.977	0.003
0.9	FXDM	0.606	0.026	0.977	0.003
0.9	WS (mean)	0.593	0.027	0.976	0.003
0.9	BA	0.563	0.028	0.962	0.013

experiences (without notions of sharing), the exclusion of 0.195 of potential cheaters is realised immediately, resulting in an efficiency of 1. Looking at the different network types, efficiency of convergence for $p_{truthful} = 0.5$ is low, with BA networks showing the fastest convergence, reaching the maximum effectivity value (here: 0.334) after around 0.84 (i.e. 1 - 0.161) of all rounds, i.e. 8400 rounds. For all other networks, effectivity values increase throughout the entire runtime. This faster convergence is partially explained by the nature of scale-free networks, in which few well-connected nodes can quickly influence the entire network. Since $p_{truthful}$ is initialised at 0.5, the initial behaviour of social hubs (i.e. reporting cheaters truthfully vs. lying) is decisive for the overall performance, potentially leading to fast spread of cheater information, or delaying the spread by creating a bias towards lying. This aspect is reflected in the high standard deviation for efficiency in BA networks (0.341), compared to the low variation of efficiency for all other network types. In particular for BA networks, randomness shows strong influence on the simulation outcome.

In order to retrace the development across different initial values for $p_{truthful}$, Figure 4.5[10] shows a chart highlighting effectivity (represented as solid lines) and corresponding efficiency values (represented as dashed lines) as y values, with initial values of $p_{truthful}$ plotted on the x axis.

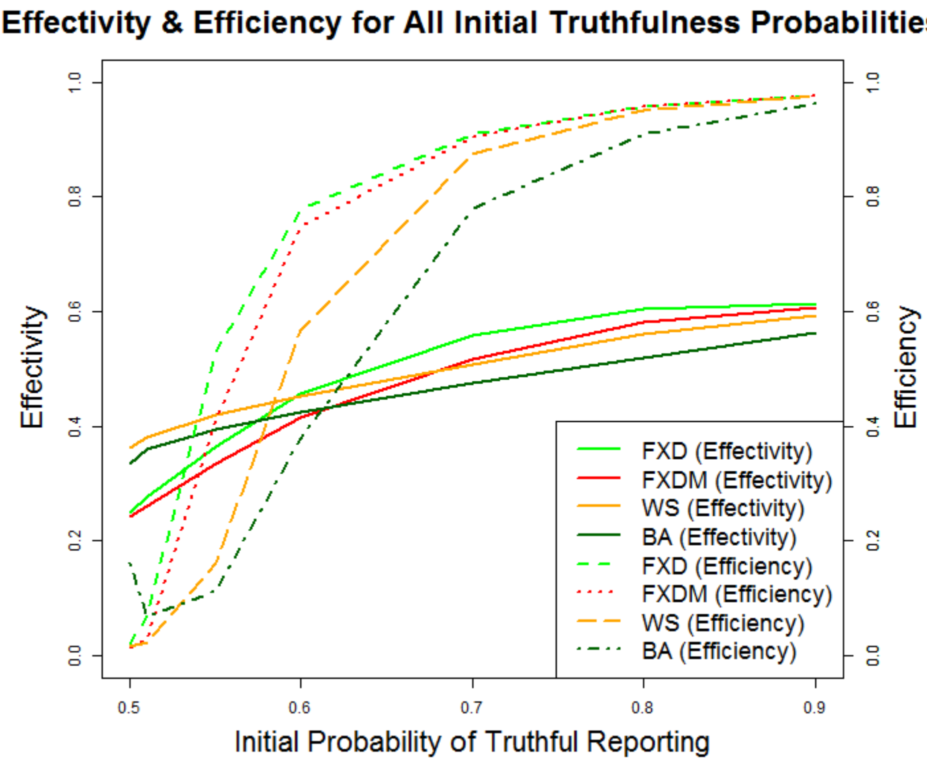


Figure 4.5: Effectivity/Efficiency Values across Range of $p_{truthful}$

The initially poorly performing FXD and FXDM networks show a strong increase for

[10] A variant that shows the individual results for different WS rewiring probabilities is provided in Figure C.5 in Appendix C.3.

increasing values of $p_{truthful}$, with a linear progression between $p_{truthful}$ = 0.51 and 0.6 before showing a decreasing growth, reaching an effectivity above 0.6 for $p_{truthful}$ = 0.9. Particularly noteworthy is the increase in efficiency for the FXD and FXDM network variants, especially between $p_{truthful}$ = 0.51 and $p_{truthful}$ = 0.55. Since all nodes are initialised with a clear bias towards truthful advice, cheater information penetrates the homogeneously connected network; recall that in BA networks the outcome depended on the truthful advice of the most densely connected nodes. Throughout all simulation runs, the directed network variant (FXD) performs better than its undirected counterpart (FXDM), since the former duplicates the links compared to the undirected network (FXDM), leading to greater spread of relationships across the network instead of relying on interlinked dyadic relationships. The networks of unequal link distribution (WS and BA) show a strong initial effectivity by exploiting the propagation within communities (WS) and through central hubs (BA). For $p_{truthful}$ values of > 0.51 WS and BA networks show a linear increase, with WS networks consistently showing a better performance than BA networks.[11]

Summarising, it is noticeable that networks with unequal link distribution, here small-world and scale-free networks, initially show a strong increase in effectivity, outperforming the random networks with fixed numbers of links. However, this effect is inverted for higher trust levels, breaking even at values of $p_{truthful}$ at around 0.6.

Complementing effectivity, efficiency is an important performance indicator. Once converging towards higher levels of effectivity, for FXD and FXDM efficiency values increase significantly, most notably when comparing efficiency values for $p_{truthful}$ = 0.51 and $p_{truthful}$ = 0.55. As indicated before, once a clear bias towards truthful cheater advice (here: $p_{truthful}$ = 0.55) is reliably established across a wider range of agents, cheater information is rapidly shared. WS and BA networks only arrive at comparable efficiency values for $p_{truthful}$ = 0.8. The reason for this lies in the structural characteristics of the networks. Since WS and BA exhibit a stronger variation of connectedness, it takes longer for cheater information to reach poorly connected nodes, leading to a continuous yet slow increase of information spread. This effect is particularly noticeable for BA networks, in which the power law distribution of links can lead a large number of poorly or even unconnected nodes that may hardly or never be reached. For this reason the BA effectivity remains significantly below all other networks, with an effectivity difference to WS between 0.02 and 0.04.

Returning to our original thread of thought, the results show that even for extreme levels of initial truthfulness, traders cannot be completely removed from employment relationships.

[11] Note that this effect is independent from the variation of rewiring probabilities, since standard deviations across all WS simulation runs are smaller than the difference between WS and BA means.

Even introducing limited communication and optimistically assuming initial levels of truthful reporting of 0.5 and higher would not have induced sufficient trust to sustain cooperation purely based on informal means. Though communication has an effect, an inquiry-based model (i.e. asking fellow investors) would not have been sufficient to facilitate sufficient information spread to crowd out cheaters.

In the light of this outcome, despite opportunistic parameter settings, we can gain additional support for this conclusion by reusing the model as a template to compare the performance of the communication pattern applied by the Maghribī traders. Thus, using an otherwise unchanged scenario we test how far the model can produce a different cooperation outcome based on the alternative, more proactive Maghribī communication patterns.

Testing the Model with Maghribī Communication Patterns

We have suggested, as a means of relaxing the rigid assumption of non-communication, that Genoese traders could have asked their fellow traders for advice before employing potential cheaters. We can abstractly describe this communication pattern as a *pull model* of communication. Inquiries are thus made based on individual demand and responses are provided in a reactive fashion. The Maghribīs, in contrast, entertained a different information regime. The credibility of an individual trader was tied to the proactive sharing of information about performed trades. Complementary to the Genoese mode of communication, we can describe this model as a *push model* of communication.

Maghribī traders proactively informed members of their aṣḥāb, i.e. fellow traders they held ṣuḥba relationships with, about observed trades (see Section 3.1; Goldberg (2012c)). This was not centrally motivated by the denunciation of cheaters, but rather a by-product of sharing information about current market prices and trading opportunities.[12] As indicated in Subsection 4.1.2, maintaining their place in the emerging total relationship network aṣḥābunā was the real 'currency' of a Maghribī trader. Short-term profits had no place in a professional network whose relationships were made for lifetimes.[13] Engaging in a ṣuḥba relationship came with a set of informal, but nevertheless strictly observed obligations (Goldberg, 2012a). Misinforming a partner, or otherwise raising suspicion of non-cooperation likely led to an exclusion that extended beyond a trader's individual aṣḥāb into the wider network (aṣḥābunā), since the suspect needed to assume that information links where redundant, and

[12] Recall that markets in the Middle East were more advanced and buyer-centric with considerable price fluctuation, an aspect that made the use of the ṣuḥba attractive because of the greater flexibility on the part of the operator (see Subsection 3.2.2) (Ackerman-Lieberman, 2012). The Southern European markets, in contrast, were seller-centric, largely following the price dictate of whoever had something to offer (see Subsection 3.3.1).

[13] Instances of ṣuḥba relationships reported by Goldberg (2012c) lasted for more than forty years.

at least in part unknown to him (see discussion of network structure in Subsection 3.2.2). The act of informing both on trader operations as well as observed behaviour of fellow traders[14] acted as a trust reassurance mechanism that demanded continuous reinforcement. A factor that could have potentially motivated this proactive communication among Maghribīs was their integrated role understanding, i.e. a trader's parallel operation as investor and operator, an aspect Greif's model has ignored and which we likewise ignore for this particular simulation. We will turn to this parallel operation consideration in Section 6.3.

However, even with the limited direct representation of the model, we can safely exclude one mode of communication that did not apply to the Maghribī case, the directed random network (FXD). Maghribīs necessarily maintained mutual relationships which were fundamental to assure reciprocal 'service for service' obligations, which were hardly fully balanced so as to keep one party obliged to reciprocate service provision and to stimulate the ongoing cycle of service provision. Furthermore, maintaining "open accounts" (Greif, 1989, 2006) incentivised compliance and bore direct material consequences, should a trader decide to defect, since he could not expect payment of outstanding debts from any member of his aṣḥāb.

To maintain the emphasis on the effectivity and efficiency of cheater propagation – not general information sharing – (and thus retain a conceptual compatibility to the Genoese scenario), we take a conservative stance and concentrate our extensions for the proactive communication scenario on *cheater reporting*; agents do not share of general information about each trade interaction. Consequently, Algorithm 4.3 highlights the relevant amendments to Algorithm 4.1.

Algorithm 4.3: Information Sharing in the Maghribī Model

…
if *operator cheated despite positive advice* **then**
    Memorise operator negatively and the receiver's entire aṣḥāb is informed;
    Memorise advisor negatively and the receiver's entire aṣḥāb is informed;
    Decrease $p_{truthful}$ by δ_{trust};
else
    // Operator behaved according to advisor's predictions
    Memorise advisor as positive;
    Increase $p_{truthful}$ by δ_{trust};
    if *operator cheated* **then**
        Memorise operator negatively and the receiver's entire aṣḥāb is informed;
    else
        Memorise operator positively;
    end
end
…

[14]Recall that as much as 20 percent of letter content was filled with such gossip (Goldberg, 2012c).

Results for the Maghribī Model

Similar to the Genoese case, for the Maghribī variant each configuration has been executed 30 times, the results of which are shown in Table 4.3.[15] Figure 4.6 further shows the effectivity and efficiency across different values of $p_{truthful}$.[16]

Table 4.3: Simulation Results for Informal Communication using Maghribī Communication Patterns

Initial Truthfulness	Network Type	Effectivity		Efficiency	
		mean	σ	mean	σ
0.5	NoComm	0.195	0.019	1	0
0.5	FXDM	0.635	0.264	0.699	0.191
0.5	WS (mean)	0.607	0.217	0.58	0.265
0.5	BA	0.588	0.197	0.497	0.279
0.51	FXDM	0.715	0.203	0.69	0.13
0.51	WS (mean)	0.696	0.127	0.553	0.169
0.51	BA	0.654	0.128	0.44	0.183
0.55	FXDM	0.79	0.031	0.736	0.018
0.55	WS (mean)	0.721	0.067	0.675	0.057
0.55	BA	0.694	0.03	0.416	0.134
0.6	FXDM	0.788	0.031	0.802	0.014
0.6	WS (mean)	0.72	0.067	0.755	0.045
0.6	BA	0.693	0.029	0.439	0.109
0.7	FXDM	0.787	0.032	0.892	0.013
0.7	WS (mean)	0.718	0.068	0.864	0.027
0.7	BA	0.69	0.03	0.747	0.092
0.8	FXDM	0.785	0.032	0.946	0.007
0.8	WS (mean)	0.717	0.068	0.931	0.014
0.8	BA	0.693	0.031	0.89	0.038
0.9	FXDM	0.784	0.032	0.971	0.003
0.9	WS (mean)	0.714	0.069	0.961	0.007
0.9	BA	0.713	0.032	0.951	0.016

In stark contrast to the Genoese case, in nearly all cases effectivity lies beyond 0.6 (i.e. 60 percent of all potential cheaters have been removed from trade transactions). In fact for all

[15] As before, we use the mean value across all WS configurations. The complete results for all rewiring probabilities are provided in Tables C.13 and C.14 in Appendix C.4.

[16] The corresponding chart for all WS rewiring probabilities can be found in Figure C.6 in Appendix C.4.

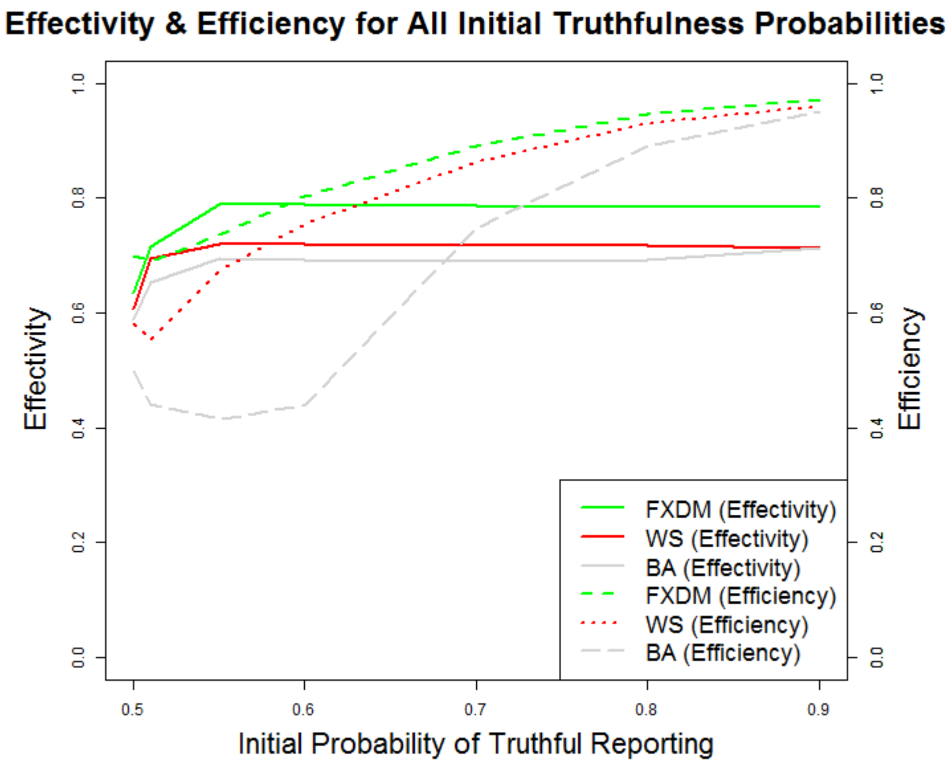


Figure 4.6: Effectivity/Efficiency Values across Range of $p_{truthful}$ for Maghribī Communication Pattern

but BA networks, effectivity peaks for $p_{truthful} = 0.55$ with levels at around 0.79 (FXDM) and 0.721 (WS). BA networks stagnate at around 0.69, and only peak for very high initial truthfulness levels (0.9), for which agents provide truthful information to effectively all requests, extending cheater information to poorly connected nodes. In fact for levels of $p_{truthful}$ beyond 0.55, effectivity values indicate a mild decline, an aspect that shows the limitations of our effectivity measure. The declining effectivity is an artefact of the rapidly increasing efficiency levels. Recall that the effectivity is calculated as a mean value of cheaters excluded from trade operations across all rounds following the establishment of the highest value. Since the number of excluded cheaters exhibits considerable variation throughout simulation runs, fluctuation around the determined convergence point (which is established based on the tolerance $\delta_{tolerance}$ from maximum fraction of excluded cheaters) obscures a minimal further convergence for higher efficiency values, leading to slightly reduced effectivity values.

These results do not come as a surprise. While the Genoese model relies on individual advice requests to inform employment decisions, the Maghribī variant proactively shares any observation with individual aṣḥābs. The penetration of all associated network links drives a comparatively comprehensive and rapid information dissemination.

Consequently, in the Maghribī communication variant the central finding is a high initial effectivity and efficiency (reaching values of 0.7 for FXDM) combined with a continuous performance increase. Instead of increasing the number of removed cheaters, in this communication model higher levels of initial truthfulness primarily reduce the number of iterations for convergence to a maximum fraction of excluded cheaters. Particularly striking are the results for $p_{truthful} \geq 0.8$, which show a near immediate convergence to a maximum level of information sharing based on the proactive communication among partners.

An exception to the general trend of rapidly increasing efficiency values represents the BA network. Though reaching nearly constant effectivity values, the scale-free network shows a comparatively slow convergence towards the final effectivity value, initially requiring more than half of the simulation time (i.e. efficiency values below 0.5 for $0.5 \leq p_{truthful} \leq 0.6$), before closing the gap to other network types by penetrating the long tail of poorly connected nodes more rapidly and comprehensively (note the effectivity increase for $p_{truthful} = 0.9$).

Discussion

Interpreting the Results With both results, a prototypical model of a mutually advising Genoese investor community, and its adaptation to a proactive Maghribī communication model, as far as the underlying assumptions of this model permit, we can discuss the results with respect to the assumption we wanted to relax – the inherent non-sharing policy of Genoese traders. We introduced a model that incorporates the 'selfish strategy', i.e. a strategy suggesting that individuals strive to maximise their utility at the expense of others (i.e. minimising others' utilities), in addition to the original 'individualistic strategy' of non-communication. Introducing a dynamic adjustment of honesty to permit the development of trust along with optimistic parameter settings (number of connections $= 40$, initial level of truthful reporting $>= 0.5$) facilitates at best around 60 percent of cheaters, suggesting a limited effectivity. The 'individualistic strategy' (*NoComm*), the model's equivalent to Greif's assumption, removes 20 percent of cheaters at best. This represents the baseline for any performance measure, since this strategy represents the exclusion of cheaters purely based on investors' individual memories. Introducing communication thus increases the exclusion by another 40 percent at best (for unrealistically high levels of initial truthfulness for a prototypical Genoese society). We can thus suggest that even when introducing the concept of advice requests, informal mechanisms, here represented as trust about mutual truthful advice, would not have sufficed to sustain a hypothetical cooperative Genoese trader society in which compliance could have been assured purely based on information sharing.

On the other hand, the 'push model of communication', representing the Maghribīs' mode of sharing, showcases how the obligations associated with the ṣuḥba relationship – which implied a monitoring component – , effectuated cooperation based on informal means.

Perhaps more interesting than the support of Greif's thesis, even with relaxed assumptions, is the observation of the performance of the different network types.

From an effectivity perspective the different network types offer a limited variation across all levels of initial truthfulness. However, results can be clustered into network topologies with equal connectedness of nodes (FXD and FXDM) as well the ones with wide distribution spreads (WS and BA).

For low initial trust levels ($<= 0.6$), the more centralised network types offered the highest effectivity, with an advantage of around 10 percent compared to the networks with homogeneous connectivity. The reason for this behaviour lies in the nature of the networks. As soon as information reaches the central social hubs of scale-free networks, it is rapidly distributed across the wider network. However, this initial performance advantage finds its limit when penetrating the weakly connected leaf nodes of the network. For longer runtimes information spread over nodes with equal connectivity (FXD and FXDM) fares better, since it assures slower but continuous and equal spread of information, penetrating the network more comprehensively, an aspect that is reflected in the nearly linear increase in efficiency over time, compared to the sigmoid shape of efficiency for small-world and scale-free networks.

Looking at higher levels of initial truthfulness (> 0.6), the effectivity for both groups of network types breaks even, suggesting better performance for random networks, followed by community-based small-world networks and scale-free networks. Generally, small-world networks fare better than scale-free networks, and only marginally worse than the undirected random network. While FXD and FXDM networks show negative acceleration across runs of increasing initial level of truthfulness, WS and BA networks perform nearly linear.

Given the limited realism of the FXD network to represent relationships in either scenario, overall small-world networks show the best performance compromise for lower and higher levels of initial truthfulness.

Mapping Societies onto Network Types Linking the historical societies to specific prototypical network types, we find that literature does not offer sufficient grounding to assign specific network types to the modelled societies, but offers pointers that allow us to suggest candidate network structures. For the Genoese society we recall the capital concentration of 90 percent, comprising a mere 37 families (see Subsection 3.3.2), and thus suggesting an unequal distribution of capital among investors. Further support to guide the specification of a possible candidate network with resemblance of the original relationships comes from van Doosselaere (2009). He analysed the Genoese commenda relationships across different cartularies. Deriving the network distribution, he found an increasing concentration of relationships with few investor families, suggesting a power-law distribution of commenda relationships. Inspecting van Doosselaere's coarse-grained cartulary network distribution of the earliest cartulary (1154-1164 AD) – the information of which we relied on for our model parameterisation – supports a potential approximation of a power law distribution, but is not fully conclusive (see Appendix C.5). However, this earliest cartulary by Giovanni di Scriba (1154-1164 AD) shows the strongest potential power law approximation, while later commenda networks exhibit reduced levels of centrality (see van Doosselaere (2009)). But the analogy to van Doosselaere's work has limitations. Given the clear role bifurcation into investors and operators as part of commenda contracts, van Doosselaere's analysis does not directly reflect relationships among investors, an aspect we modelled in this scenario since we were interested in the sharing of cheater information. Furthermore, his analysis focuses on *contractually formalised* relationships; in our model we represent hypothetical *informal* information sharing relationships. However, if such a network existed, one could hypothesise that the network, with few influential family firms at its centre and a long tail of low capital investors,[17] would likely bear resemblance to the commenda relationship network van Doosselaere analysed and similarly bear scale-free characteristics. As such the BA network is a possible candidate for its representation.

For the Maghribī society the document analyses performed by Goitein (2000b), Greif (1989, 1993, 2006), and Goldberg (2012c) provide insights that likewise point to network structures other than random networks. Goldberg (2012c) devotes extensive discussion to two traders, namely Nahray Ibn Nissīm and Yūsuf Ibn ʻAwkal, both of whom were tightly involved in the aṣḥābunā. Nahray Ibn Nissīm maintained around 150 documented relationships to other traders, and Yūsuf Ibn ʻAwkal even enjoyed as many as 400 relationships. This considerable variation in connectivity shows the most prominent cases of interconnection. With respect to the overall network structure, Goldberg (2012c) states:

> "The network of aṣḥābunā in the Ibn ʻAwkal and Nahray groups should thus not be understood as perfectly connected, monolithic, or composed of equally strong ties. [...] [The network had] enough connections to be mentioned as carrying out commercial services or having a partnership with more than one other merchant in the network. The network was always in flux, both through addition, retirement, or death of individual merchants [...]."

[17] Again, recall the strong capital concentration of Genoese (see Subsection 3.3.2).

The unequal distribution of network ties appears certain, as is the existence of central hubs that interlink more peripheral traders, which could point to an approximation of a scale-free network. However, it is important to consider two further aspects.

85 percent of the letters based on which Goldberg's quantitative analysis had been performed were destined to Fustat (Goldberg, 2012c), the capital of the Fatimid Empire, but also the location of the Ben Ezra Synagogue in which the Geniza was discovered. This suggests a bias of documented trade relationships, and will likely not be representative for the overall structure of the relationship networks, which spanned across homesteads and trade locations along the North African coast, as opposed to a concentration on a particular site. The history of the Maghribīs and their dominating origin from the two cities of al-Mahdiyya and Qayrawān (based on the common background of escape from Iraq, see Subsection 3.1.2) suggests strong local concentration of (inter-)connections within the Maghreb communities. Based on their informal nature and the close proximity of partners, those were unlikely to be reflected in the Geniza (or at least not in the remote Ben Ezra Geniza), if in any written form at all.

A second consideration is Goldberg's documentation of an apprenticeship system among Maghribīs in which newcomers were trained to participate in the skilled activity of trading. The initial training was characterised by the local concentration of task fulfilment for a mentoring established trader, and, once apprentices had built sufficient jāh, was followed by increasing expansion based on delegation (an aspect we will explore as part of the following Subsection 4.2.2). However, assuming that young Maghribīan traders initially shared a local focus, as soon as starting to enter the wider trader network, they would most likely engage with agents of similar (initially low) status, since their services were affordable. Recall that the remuneration for services by reciprocal obligations was determined based on the relative jāh difference of the partners (see Subsection 3.2.2). Engaging with traders of higher jāh was – beyond their limited accessibility in the first place – thus likely to be less profitable. Returning to the network structure, the existence of clusters of emerging trader generations suggests that intra-generational relationships were likely more dense than inter-generational relationships to older, more established traders.

Following these two considerations, along with the emergent nature of the aṣḥābunā as the combination of individual aṣḥābs, the network structure of the Maghribīs could thus have been characterised by a interconnection of more densely connected clusters based on status similarity (i.e. trader generations) and proximity. As an alternative to a scale-free topology, the Maghribī collective could have been characterised by small-world-style networks, reflecting the metaphor of interconnected communities with dense intra-community links.

Whichever represents the historical reality most accurately (in the light of a lack of further information), the results of our simulations show that the nature of the actual network structure is secondary to the proactive mode of communication the Maghribīs adopted.

Contextualising the Models Looking at the broader context, the model described here is an example for the path-dependent development of social institutions. By transplanting mechanisms of indirect reciprocity (as found in the Maghribī society) into a secretive Genoese society, we shift the paradigm from personal enforcement to communal enforcement (Greif, 1992). The results are compatible with work in the area of behavioural game theory that looks at cooperation from an evolutionary perspective, such as Bowles's and Gintis's (2004) influential model of strong reciprocity in Pleistocene societies. Their model postulates the existence of *strong reciprocators* that facilitate cooperation by punishing cheaters irrespective of their immediate selfish benefit. Our model of the Genoese society lacks such strong reciprocators, which is rooted in the limited will to share information about their individual ventures, and even less, to establish systematic private-order enforcement, which would have been costly, since newcomers could not expect repeated employment and thus had strong incentives to cheat. A cooperative outcome relied on the use of formal institutional mechanisms that reduced uncertainties of long-distance trade as a predictable investment opportunity for otherwise uninvolved traders. Maghribīs, on the other hand, did not rely on the power of strong reciprocators. Instead, they relied on the normative behaviour of collective sanctioning, which was by no means selfless behaviour but was incentivised by the carrot of maintaining a lasting cheater-free network and the stick of becoming subject to suspicion, should an observer fail to inform his partners.

The current model has various limitations. Networks are statically generated at the onset of the simulation, and relationships do not change over time (although investors may ignore fellow investors they consider liars). However, limited knowledge about its structure, beyond the presumed topological similarity to the commenda network, limits the benefit of such endeavours. This contrasts with Bravo et al.'s (2012) model which likewise represents an investor-trader scenario and, by testing various social structures, suggests that the continuous creation of ties supports the removal of cheaters from the network. They further suggest that the network structure itself is secondary to the ability to shape those networks dynamically. Villatoro et al. (2009) combine topology and memorised history of past interactions to explore their impact on the emergence of conventions. Their findings suggest that the ability of nodes to connect to distant nodes increases the speed of convergence, since conventions are widely distributed and local subconventions are avoided. This is compatible with earlier findings by Nakamaru and Levin (2004), who emphasise fastest spread of norms in scale-free networks, while avoiding coexistence of multiple social norms based on social learning. A further relevant finding by Villatoro et al. (2009) is the inverse relation of reward and memory size, decreasing the efficiency of convention emergence, an effect we did not explicitly explore. Villatoro et al. (2011b) further explore the convention dynamics by giving agents social instruments that allow them to refine their links based on observation and rewiring, with the major benefit of arriving at global conventions. Returning to an emphasis of economic impact and network structure, Skyrms and Pemantle (2000) support the importance of trust-based partner selection and associated shaping of network relationships to minimise the impact of free-riding. However, contrasting their models, the investor network described here does not consider explicit free-riding of investors. Instead, they are experiential learners that continuously adjust their trust towards the collective – not individual investors. Further work includes Zschache's (2012) proposal to apply notions of social comparison to evaluate fellow agents' performance in order to strategically shape the relationship network's structure. To introduce this social-psychologically inspired approach, traders would require more factual detail to display 'performance', which could potentially be derived from observable affluence (i.e. accumulated wealth). However, the direct inspection of trade relationships of third parties would then require a further relaxation of the secrecy assumption put forth for the Genoese.

Other aspects that lend themselves for refinement include a stronger differentiation between cooperation and defection during trade interactions by introducing behavioural expectations based on the situational trust level, i.e. developing an expectation for operators to cheat given dominating exposure to cheating. This would emphasise the behavioural perspective and, following the understanding of behaviour-based culture transmission (Parsons, 1967; Grusec and Kuczynski, 1997), offer an instrument to reconstruct the trade culture of the historical society. As such this would incorporate a stronger social-psychological component that emphasises individual experience in the evaluation of cooperation. It could further consider the asymmetric impact of cooperation and defection behaviour on reputation (Baumeister et al., 2001). For the Maghribīan case, a realistic refinement could involve the threat of detrimental punishment by supernatural agency (Johnson and Bering, 2006) as a central motivator for compliant behaviour.[18] For the Genoese the introduction of such extensions, however, would likewise imply that investors develop expectations among themselves, including the question whether voluntary truthful reporting could have been expected after

[18]Maghribīan writings continuously made reference to God, both to express high appreciation as well as threats, an aspect that is reflected in the nature of the Geniza itself: a storage of documents that contain the name of God (see Subsection 3.1.1).

all in a society that was known for its secrecy.

Looking at the more concrete context, agent-based simulations of historical trader societies, Ewert's and Sunder's model (2011) of the Hanseatic trade network reveals considerable parallels. Their model is a comprehensive reflection of documented Hanse characteristics, including the focus on kinship relationships and the closed nature of the network and reciprocal services, elements that bear similarity to the Maghribī network. Ewert's and Sunder's evaluations concentrate on the expansion of the trade network into the Baltic region and explore the resulting change in network structure along with the impact of regional trade privileges.

Summary and Contribution As far as this investigation into Greif's assumptions is concerned, we believe that the optimistic relaxation of Greif's first assumption – the non-communication among Genoese traders – in combination with a theoretical absence of formal contractual enforcement, *would not have resulted in a cooperative outcome*. However, the agent-based model offers a more differentiated insight into structural aspects of the society than the original game-theoretical model could capture. Examples for these refined modelling features are the differentiated modes of communication (*pull* vs. *push* model) and the reflection of the institutional instrument ṣuḥba with respect to compatible relationship network structures (affording the exclusion of the directed network (FXD) from push model analysis). However, even with relaxed assumptions, similar to Greif we suffer from the absence of more detailed micro-information and thus considerable abstraction. It is thus important to recall that this model is geared towards 'understanding' as opposed to the provision of an 'accurate representation' of historical reality.

As part of this model, we contrasted the communication patterns the Maghribīs employed to sustain cooperation based on informal means, an aspect that relied on the assumption of a 'closed' group based on Greif's assumption of individuals' devotion for the greater coalition, a view that has been subject to the previous literature challenges (see Subsection 3.2.2). A candidate mechanism that could have supported the relaxation of the 'closed group assumption', while maintaining compliance based on informal means, could have been the apprenticeship system the Maghribīs employed (see Subsection 3.2.2). Its consideration promises further insights into the structural make-up of the Maghribī relationship network, an aspect we experimentally explore in the upcoming Subsection 4.2.2.

4.2.2 Experiment 'Apprenticeship System in the Maghribī Traders Coalition'

One of Greif's fundamental assumptions to explain the functioning of the Maghribī Traders Coalition is its closed nature and inherent motivation to share information of value with the wider collective (Greif, 1994). Under a regime of repeated interactions among a fixed number of group members defined by a common cultural identity (history, language), the threat of informal enforcement, such as the removal from trader networks and taking regress against kin, thus becomes credible driver for cooperative behaviour.

He bases this suggestion on the limited observation of informal relationships to outsiders.[19] However, Greif's letter sample captured 175 letters, while the more recent analyses by Goldberg capture as many as 1500 letters, including 900 trade-related ones (Goldberg, 2012b). This challenges Greif's conception of the Maghribīs as a fully coherent and closed group, leading Goldberg to suggest that he confused the 'plurality' of members originating from al-Mahdiyya and Qayrawān with the 'totality' (Goldberg, 2012c). Goldberg's analyses further suggest a larger extension to other Jewish settlements in the West of the Fatimid Empire than previously assumed, loosening the 'closed group' assumption further.

Motivation

A possible institution that could account for a relaxation of the 'closed group' assumption without softening the group's boundaries is the documented existence of an apprenticeship system among the Maghribīs (as introduced in Subsection 3.2.2). This could deepen our understanding in much the same way as knowledge about the ṣuḥba as essential coordination instrument drove the reinterpretation of the structural nature of the network as loosely coupled. To recall, Maghribī traders were not able to simply join the aṣḥābunā based on cultural relationships or recommendations, but instead had to show commitment to the trader collective by passing a period of unremunerated work, which Goldberg characterised as "'junior associate' system" (Goldberg, 2012c), and which we, despite absence of explicit formal recognition (see Goldberg (2012c)), refer to as an *apprenticeship system*. In order to develop sufficient commitment of newcomers to the system of interconnected aṣḥābs, those newcomers, often sons of established traders, offered unremunerated work to another established trader, who would instruct and endow them with travelling and trading obligations of increasing complexity. This mentoring included delegation of concrete tasks, such as procuring goods from the country-side, but also sharing of trade-related skills as well as secondary

[19] For the observed sample he reports six external traders (Greif, 1994).

aspects such as managing what we would interpret as 'work-life balance' (Goldberg, 2012c). Over time the apprentice's relationships would extend from an intimate mentor/mentee focus to the wider trader community based on the exposure to different market places and the mentor's introduction to fellow associates. The mentee incrementally developed his own trade relationships using his own initial capital, given that it was the mentor's obligation to maintain the mentee's basic livelihood during the apprenticeship period. The overt function of such a system was to groom qualified newcomers that could reliably and flexibly contribute to a system that relied on advanced trading skills and members that would identify themselves with the relationship network they had been nurtured in. Accounts of this progressive development of traders indicate that apprenticeship relationships could last longer than a decade and bore the option of failure (Goldberg, 2012c).

However, besides this documented purpose of grooming qualified traders, let us look at its implication on the cooperation problem we are interested in. The more subtle purpose of the apprenticeship system – implicit or not – was its role as an entry barrier into a network of compliant traders. The apprenticeship system thus had to establish the cooperativeness of the newcomer, opening the network's boundaries and making it permeable for outsiders – thus adding prospects of growth – , but introducing a considerable threshold for entering and a disincentive for ever leaving the trader network (an aspect that can be represented as an investment). Assuming the duration of a trader's operation in his profession to last around 40 years, a decade of 'apprenticeship', characterised by intensified monitoring, would signify a considerable commitment, especially when considering the absence of remuneration until the apprentice was to engage in his own trade ventures. Furthermore, established traders had a strong incentive to deal with traders that were members of the aṣḥābunā, for which they could expect the honouring of the ṣuḥba relationship that often implied skilful autonomous handling of trade in the opportunistic interests of the goods owner at the target market location (Goldberg, 2012c). Aṣḥābunā membership and previous experience are thus retraceable motives for the establishment of trade relationships. However, membership in itself was likely not sufficient to qualify for a trade relationship between aṣḥābunā members.

At this stage it may be helpful to look at two principles Podolny (1994) offers in the context of organisational theory. In the face of economic uncertainty, Podolny posits, an entity concentrates on partners it had previous experience with, and beyond this, prefers partners that are of similar status. Appreciating a partner's similar status (or jāh) was of practical benefit for Maghribīs. Engaging with a merchant of higher jāh was expensive. Given that the price for reciprocating services was roughly based on the relative jāh difference (Goldberg, 2012c), the delegating partner of lower jāh thus had to perform more services to 'repay' the incurred obligations. Given this principle, it may have been attractive to engage with traders of lower jāh in the first place. However, jāh was considered as a representation of experience and described the extent to which a trader could be expected to act compliantly, given his increasing 'investment' in jāh and the associated sunk cost.[20] We believe that Maghribī traders thus had a preference for partners that had a fairly similar status.

With this background in mind, let us construct a basic model of the Maghribī apprenticeship system and explore the extent to which this particular institution could have substituted the conceptual need for a closed trader network to sustain cooperation. The has been published in Frantz et al. (2015a) and is presented here in an extended form.

Model

The historical accounts suggest that the apprenticeship network's structure relied on a trader's jāh as the pivotal variable to shape network relationships, a mechanism that, we believe, had a filtering function that excluded cheaters from the network. In contrast to our earlier exploration of the Genoese communication behaviour (see Subsection 4.2.1), the jāh concept gives us the necessary regulative to adopt a generative approach and let the relationship network emerge, as opposed to predefining it at the simulation onset. This is in line with Bravo et al. (2012)'s suggestion that the actual topology is secondary to a dynamic generation of the network.

In this model, agents perform actions (described in Algorithm 4.4) and react to those of other agents (Algorithm 4.5). At this stage recall that the Maghribī society did not have separate 'investors' and 'operators'; agents acted in both roles simultaneously.

At the beginning of the simulation, agents are initialised with a jāh level of zero and randomly pick another agent, and, if the target is unemployed and was not employed previously, offer themselves as an apprentice. If accepted by the requestee, an agent commits to an apprenticeship lasting between $apprenticeshipDuration_{min}$ and $apprenticeshipDuration_{max}$ simulation rounds. If not accepted as apprentice, the agent opportunistically attempts conventional employment if the jāh difference to the requestee permits. Alternatively, if previously employed, the agent will offer himself as a regular trade partner.

The conditions for the acceptance of apprenticeship and partnership requests vary. If the requestee does not maintain another apprentice at the current stage, he will employ the requester as apprentice, independent of his jāh level. The apprentice will then be at his avail and not continue to offer partnerships to others until his apprenticeship finishes. Conven-

[20] Goldberg (2012a): "Geniza merchants seem to have preferred contracts in opposite proportion to the agent's natural incentive to provide good service."

tional partnerships between two established traders (in which both partners act as a potential senders of goods and as trade operators), however, are only assumed if both partners are within the *tolerance range* of their respective jāh levels. The tolerance threshold is defined as $tolerance_{jah}$, with

$$\textit{jāh - jāh * lowerJahDifference} \leq tolerance_{jah} \leq \textit{jāh + jāh * upperJahDifference} \tag{4.4}$$

, and *lowerJahDifference* as well as *upperJahDifference* lying between 0.0 and 1.0.

If in a ṣuḥba relationship, agents trade with a fraction *tradeQ* of all agents they have relationships with. Successful trade results in mutual jāh increments (by *jahIncrement*), representing an increase in experience and standing.

Besides trading, agents observe the behaviour of other agents, expressed as the fraction *observationQ* of *all* other agents. The monitoring of fellow traders' conduct (or *proactive reporting norm* – represented in the 'push model of communication' in previous experiments in Subsection 4.2.1) was an integral function of the Maghribī traders collective and appeared to be exhaustive in the light of the redundant use of letters (i.e. redundant sending using both sea- and land-based transport) and their public acknowledgement as well as written confirmation of receipt (see Subsection 3.2.2; Goldberg (2012b)). Performing this observation in a randomized manner abstractly represents the distributed nature in which trade (and thus monitoring) was pursued. Note that the targets of observation did not necessarily need to have any relationship with the observer or even his partners. However, due to the specifics of the applied institutional instrument and distributed dissemination, observations could nevertheless pass to one's partners. The sender could not assume that none of his partners did have a ṣuḥba relationship with the monitored, since the ṣuḥba was a strictly private 1:1 matter and was not necessarily known to third parties. Informing partners about market observations was an obligation that signified credibility and trust, reinforcing the informer's standing. We believe that this feature, the emergent structure (the aṣḥābunā) of overlapping partially unknown relationship networks (aṣḥābs), had strongly reinforcing features, since the observer's reputation was at stake, should his aṣḥāb receive information he had intentionally withheld via other observers unknown to him.[21] As part of this obligation to share information, we further assume that all information received by a trader is passed on to his entire aṣḥāb, leading to a penetration of the aṣḥābunā. Signifying the trader's commitment to its aṣḥāb, announcements to one's aṣḥāb are thus rewarded with an increase in jāh. If mentoring an apprentice,

[21] At this stage recall the detailed discussion in Subsection 3.2.2 that offers a refined understanding of the aṣḥābunā as an emergent construct of overlapping dyadic relationships, preventing individual members from perceiving the coalition as a structure with crisp boundaries.

we consider the mentor to observe his mentee's conduct in any case, given the assumption of tight interaction between mentor and mentee, including smaller tasks that offered limited opportunity and reason for cheating.

At the end of each execution cycle, an agent may die as determined by the probability $p_{deathBelowExpectedAge}$, if his *age* is below *expectedAge*, and the probability $p_{deathBeyondExpectedAge}$ if older. Using this mechanism we can specify an expected age but at the same time accommodate the random passing of a trader, smoothing the collective's generational transition. Each dying agent is replaced by a newcomer who needs to pass through the phase of apprenticeship in order to become an established member, unless within the potential future employer's permissible jāh range. The latter case is relevant to represent the control case of full permeability of the coalition, in which employers employ irrespective of newcomers' experience (represented as jāh).

Algorithm 4.4: Agent Execution Cycle

```
Agent picks random other agent;
if unemployed and not previously employed then
    Offer oneself as apprentice;
    if accepted then
        Commit to apprenticeship for random duration between
            apprenticeshipDuration_min and apprenticeshipDuration_max rounds;
    else
        if randomly picked agent is within acceptable range of own jāh then
            Offer oneself as employee;
    end
else
    if randomly picked agent is within acceptable range of own jāh then
        Offer oneself as employee;
end
Trade with tradeQ of employed agents;
Increase own jāh by jahIncrement for each trade;
Randomly choose observationQ number of agents from all agents (incl. eventual apprentice);
foreach agent in observed agents do
    Check if agent has cheated;
    if agent has cheated then
        Memorise agent as cheater;
        Announce to other agents in own aṣḥāb;
        Increase own jāh by jahIncrement;
end
Increment age each round;
if age ≤ expectedAge then
    Check for death with probability p_deathBelowExpectedAge;
else
    Check for death with probability p_deathBeyondExpectedAge;
end
```

Complementary to the execution cycle outlined before and shown in Algorithm 4.4, the agents' reactions to other agents' actions are summarised in Algorithm 4.5 and described in the following.

If an agent receives an employment request, its response depends on the state of the requester. If the requester seeks initial employment as an apprentice and the requestee does not currently maintain another apprentice, the apprentice is accepted. Otherwise the establishment of the partnership depends on a compatible jāh difference. Requests from cheaters are ignored, since agents only deal with members of their aṣḥāb (from which cheaters are removed upon announcement).

Trading requests by aṣḥāb members are processed in any case and responded to with trading and an increase in jāh as experiential gain. However, if the receiving trader has been initialised as cheater, he cheats with a given probability $p_{cheating}$.

Finally, if an agent receives notifications about cheating by fellow traders, he forwards this information to his entire aṣḥāb (apart from the sender of the original notification). This models the information propagation obligation for all entertained ṣuḥba relationships. Forwarding information is associated with an increase in jāh.

Algorithm 4.5: Agent Reactions

```
if receiving employment request then
    if requester seeks initial employment and receiver has no other apprentice at current stage then
        Employ irrespective of jāh difference;
    else
        if requester within jāh range and not known as cheater then
            Employ requester;
        else
            Reject request;
        end
    end
if receiving trade request then
    if recipient is cheater then
        Cheat requester with probability p_cheating, otherwise trade fair;
    Trade;
    Increase own jāh by jahIncrement;
if receiving cheater notification by others then
    Memorise declared cheater;
    Terminate eventual relationship to cheater;
    Share with own aṣḥāb (excluding agent who sent cheater announcement);
    Increase own jāh by jahIncrement;
```

At this stage it is worthwhile to reflect on the chosen jāh representation. In this model, jāh is represented by a monotonically increasing value, only bounded by the agent's lifetime. In fact the representation increases the jāh even in the case of cheating. An alternative would be a more complex representation of reputation with accurate situational values. Candidates would include the use of *endorsements* as introduced by Cohen (1985) and adapted to the context of social simulation by Alam et al. (2010). Endorsements are intended to offer a lightweight model that can substitute some of the features offered by the processing-intense cognitive architectures based on the Belief-Desire-Intention model (BDI) (Bratman, 1987). Endorsements can be used to established "cognitive trajectories aimed at achieving information and preferential clarity over an agent or object from the perspective of the endorsing agent itself" (Alam et al., 2010). This concept encompasses aspects such as context, frequency, and latency (Alam et al., 2010) of reinforcement to allow continuous adjustment of endorsements. However, our model represents a singular context, here economic relationships, and disregards the multitude of further social relationships (e.g. kinship relationships), for which the use of endorsements would be purposeful. Furthermore, the notion of fluctuating jāh is of limited relevance, given that a detected cheater is excluded by definition. Adjusting the jāh would be secondary to his 'tagging' as cheater, which would disqualify him from future employment in the first place, as long as his defection is propagated throughout the relationship network. Reducing the jāh would limit our ability to isolate the effect from excluding cheaters based on their 'cheater tag' as opposed to an incompatible (low) jāh value. Given our central intent to explore the effect of an *apprenticeship system* on removing cheaters by means of proactive communication (which is supported by literature), introducing endorsements would involve additional assumptions that are difficult to support. However, for a comprehensive model of the trade society and the social relationships extending to the respective communities, endorsements would be able to unfold the implications a detected defection could have on the social network of the cheater (e.g. exploring the social effects of private-order enforcement).

Returning to the model description and the motive of our experiment, exploring the effect of the apprenticeship model, we need to specify various parameters. Agents are initialised as either cheaters of non-cheaters (the former cheat based on $p_{cheating}$), and start with a unified initial jāh level. Further parameters, most of which have been introduced before, and their underlying intuitions are described in the following. In many cases we rely on optimistic guesses and parameter sweeping for selected parameters to explore the generalisable properties of the institution 'apprenticeship system'. Table 4.4 provides an overview of the different parameters and indicates their default values and range of variation.

To establish empirical grounding where possible, in our simulation ten simulation 'rounds' represent one year. The lifetime of a trader only considers the period in which he is economically active, including his apprenticeship and further trade activities. In our control case, agents have immediate access to the market without prior apprenticeship. Based on anecdotal evidence for ṣuḥba relationships of more than 40 years duration between Nahray Ibn Nissīm and one of his associates (Goldberg, 2012c), we specify the *expectedAge* with 400 rounds. As mentioned before, we smooth the transition between trader generations by significantly increasing the death probability once reaching this age (from $p_{deathBelowExpectedAge}$ to $p_{deathBeyondExpectedAge}$). The possible premature death reflects incidental cases and consequentially a dissolution of ṣuḥba relationships, an aspect comparable to Greif's notion of 'forced separation' (Greif, 1994) (see Subsection 4.1.1).

Table 4.4: Simulation Parameters

Parameter	(Default) Value	Parameter Variations
Number of Agents	400	200
Fraction of Cheaters	0.4	
$p_{cheating}$	0.5	0.1 - 0.9; step size: 0.1
expectedAge	400 rounds	
$apprenticeshipDuration_{min}$	50 rounds	
$apprenticeshipDuration_{max}$	100 rounds	
lowerJahDifference	0.2	0.1 - 1.0; step size: 0.1
upperJahDifference	1.0	0.1 - 1.0; step size: 0.1
tradeQ	0.5	0.1
jahIncrement	1	
observationQ	0.025	0.025 - 0.3; step size: 0.025
$p_{deathBelowExpectedAge}$	0.0001	
$p_{deathBeyondExpectedAge}$	0.001	

As mentioned before, apprenticeship relationships could last for more than a decade. As the lower boundary for apprenticeships we use 5 years, which is reflected in the minimum and maximum durations (see $apprenticeshipDuration_{min}$ and $apprenticeshipDuration_{max}$). To represent the tighter interaction cycle, we initially assume 6 trade interactions per year, a value we vary as part of the model exploration.

A last aspect to be supported is the number of agents. Literature reports a wide range of values, with the lower boundary at around 330 (Greif, 1994) and a more recent figure, suggesting around 550 active traders (Greif, 2012). Those numbers are not fully conclusive as they are based on Greif's sample of all available Geniza documents, for which future research is likely to provide refinements. For the default value we chose the intermediate value of 400.

For the remainder of the parameters, such as the number of cheater, $p_{cheating}$ and the permissible jāh differences, we did not find evidence to support particular values. Particularly the cheating parameters have relatively high default values in order to amplify the effect of the apprenticeship system, and furthermore, to exclude eventual cultural bias towards non-cheating on the part of the Maghribīs (e.g. by extrapolating the limited cases of observed cheating reported by Greif (2012) to the potentially more numerous cases of performed cheating).

The grounding of the model shows the practical problem of working based on rather weak anecdotal evidence and its extrapolation into model parameters. This aspect highlights the importance of the exploration of different parameter ranges. This naturally limits the ability to provide a historically accurate and realistic representation, but it allows us to analyse the apprenticeship system's function, identifying whether, or under which conditions, it could have managed the trade-off of permitting newcomers, while maintaining relatively tight group boundaries. This would promote the apprenticeship system as a substitute to Greif's 'closed group' assumption (see Subsection 4.1.2), but also highlight generalisable findings that characterise conditions under which the employment of apprenticeship systems appears purposeful or not.

To establish the effectiveness of the apprenticeship system, we measure the *number of cheaters that maintain relationships with non-cheaters* as dependent variable. We ran the simulation for 20,000 simulation rounds and measured the mean number of such relationships across all simulation rounds. As a control case – as indicated before –, we initialised each configuration in a variant that does not require newcomers to pass through the apprenticeship phase.

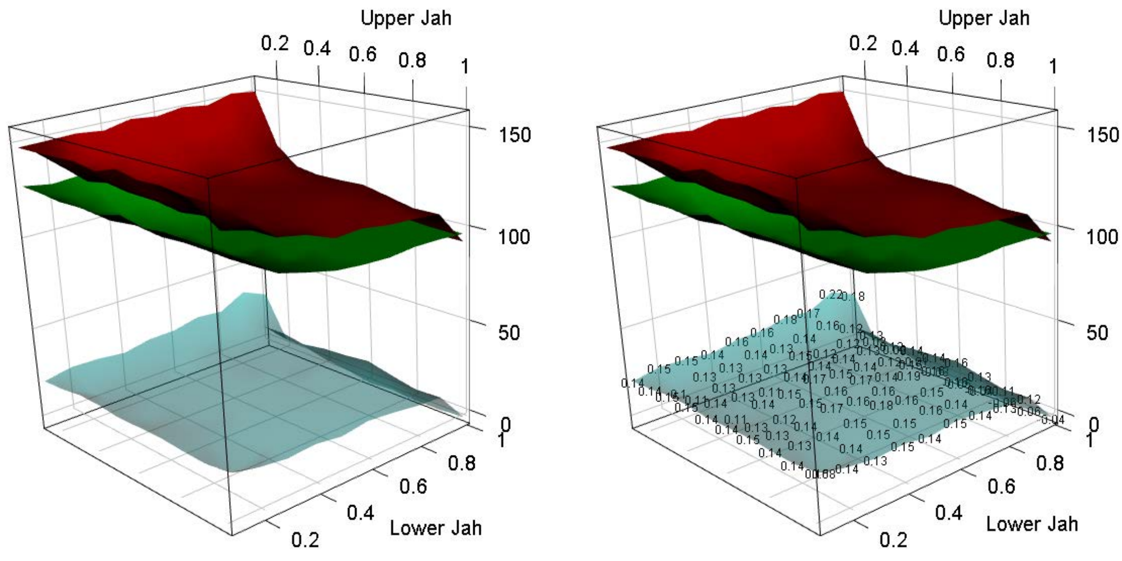
Analysing the Apprenticeship Model

In order to analyse the performance for the apprenticeship model as a means to constrain cheating, we vary selected parameters systematically. We concentrate the initial analysis on the effect of narrowing and widening the tolerance for jāh differences to form partnerships.

Upper and Lower Jāh Differences A central assumption of our model is that traders prefer partnerships with traders of similar jāh level. We thus explore the extent to which the tolerance of traders of higher jāh levels (*upperJahDifference*) and lower jāh levels (*lowerJahDifference*) affects the inclusion of cheaters in partnerships. We systematically vary both tolerance levels from 0.1 (implying an acceptable relative jāh difference of 10 percent) to 1.0 (no jāh-based partnership constraints) in step-wise increments of 0.1.

To facilitate the discussion, we show the results of all parameter variations in the form of surface plots (for this case in Figure 4.7), in which the number of cheaters are shown on the vertical axis,[22] with the green surface representing the apprenticeship model results, and the red surface reflecting the control case without any apprenticeship concept (i.e. immediate full employment). The difference between both models is plotted as a transparent blue surface showing the inverted absolute performance difference, i.e. the extent to which the apprenticeship model reduces the number of connected cheaters in comparison to the apprenticeship-free model.

[22] The axis description has been intentionally omitted to improve legibility of charts.



Red surface: without apprenticeship system
Green surface: with apprenticeship system
Blue surface: inverted absolute cheater reduction by apprenticeship system
The right chart additionally shows the relative cheater reduction for each data point.

Figure 4.7: Number of Cheaters with Relationship to Non-Cheater(s) for Varied Upper and Lower Jāh Differences

Table 4.5: Statistical Measures of Relative Differences and Absolute Values for Upper Jāh vs. Lower Jāh Difference

Relative Difference				Correlation[a] of Mean Relative Diff. with	
Min.	Max.	Mean	σ	Upper Jāh Diff. (p)	Lower Jāh Diff. (p)
-0.057	0.216	0.128	0.054	-0.2 (0.584)	-0.00606 (1)

Absolute Values				Correlation[b] of Absolute Values with	
Min.	Max.	Mean	σ	Upper Jāh Diff. (p)	Lower Jāh Diff. (p)
95.628	128.505	109.265	10.363	-1* (2.2e-16)	-0.975* (2.2e-16)

[a] Calculated as Spearman's ρ between input parameter and mean value of inverted relative cheater reduction (of apprenticeship model compared to apprenticeship-free control model) across all parameter combinations that include input parameter value, along with p value (α = 0.05). * indicates significance.
[b] Calculated as Spearman's ρ correlation coefficient between input parameter and mean absolute cheaters (in apprenticeship model) across all parameter combinations that include input parameter value, along with p value (α = 0.05). * indicates significance.

A variant of the plot that additionally shows the relative difference for each data point is provided alongside any chart, with an enlarged variant shown in Figure D.2 in Appendix D.[23] In addition to the visual representation we provide a statistical overview in Table 4.5, which shows the relative reduction of cheaters (labelled 'Relative Difference') in the apprenticeship-based model compared to the apprenticeship-free model. To differentiate relative performance benefit of the apprenticeship model from its absolute performance, we additionally provide metrics for absolute cheater numbers in the apprenticeship model variant (labelled 'Absolute Values'). For each variant we further provide correlation coefficients of output values with respective input parameter variations.[24] The data underlying the produced surface plots and the statistical evaluation can be found in Table D.1[25] and are calculated as mean values across 10 runs for each input parameter combination. Interactively navigable versions of all surface plots related to the apprenticeship model are available under (Frantz, 2014).

Highlighting the general properties of results, we can observe the effect of very high overall numbers of cheaters, with a theoretical maximum of 160. As mentioned before, those values have hardly been realistic, but they amplify the effect of the simulated apprenticeship model.

For nearly all combinations of upper and lower jāh difference values, we can observe a better performance of the apprenticeship variant, with 0.11 to 0.22 fewer connected cheaters for all parameter combination other than *lowerJahDifference* = 1.0 (a case which we discuss below). Increasing values of jāh difference reflect the 'openness' of the relationship network, with lower jāh difference indicating the tolerance of established traders to newcomers, and upper jāh difference defining the willingness of newcomers to engage with established traders. The results reveal a general tendency towards reduced numbers of connected cheaters when increasing 'openness' (see correlations for absolute values in Table 4.5) – an unexpected result at first glance. At the same time the apprenticeship variant offers a nearly constant improvement in reducing cheater levels (see correlations for relative difference in Table 4.5).[26] Passing an apprenticeship thus limits the cheaters that 'make it' to full tradership which lets them engage in relationships with established traders. The only exception is the boundary case of full permissiveness of partners with lower jāh, which leads to a better performance of the apprenticeship-free model (apprenticeship model performs worse by nearly 0.06). For the extreme case of unconstrained tolerance towards newcomers (*lowerJahDifference* = 1.0) the apprenticeship is not effective, because newcomers can

[23] The appendix further contains a full tabular overview (in Table D.3) of the individual data points.

[24] Correlations are based on Spearman's ρ since our results do not follow a normal distribution (as established based on the Shapiro-Wilk test (Shapiro and Wilk, 1965)) but show a monotonic relationship.

[25] Tables for all other tested input parameter combinations are likewise provided in Appendix D.

[26] Increasing openness has an insignificant negative correlation with relative performance difference of the apprenticeship system.

bypass the apprenticeship period and immediately engage with established traders. Given that they do not underlie the constraint of a 1:1 relationship to a single mentor, they can instead form multiple relationships at the same time.[27] Assuming a constant cheating probability, more frequent interactions increase the likeliness of being detected and flagged as cheater. This unconstrained tolerance towards lower status traders is overturned if permissible upper jāh differences are reduced, representing the unwillingness of newcomers to engage with established traders. The extreme case of not tolerating higher status relationships (lower jāh: 1.0, upper jāh: 0.0) effectively separates trader generations 'from the bottom' and realises segregated trader generations, preventing cheater information to permeate the separated trader networks.[28] However, separating the communities reactivates the apprenticeship system's effectiveness: newcomers enter apprenticeship relationships with established traders, since agents are indifferent about jāh differences when forming apprenticeship relationships.

Summarising the overall findings, open systems drive the identification of cheaters based on the increasing level of interaction with newcomers. However, this 'openness' needs to remain bounded to make controlled assessment of newcomers effective.

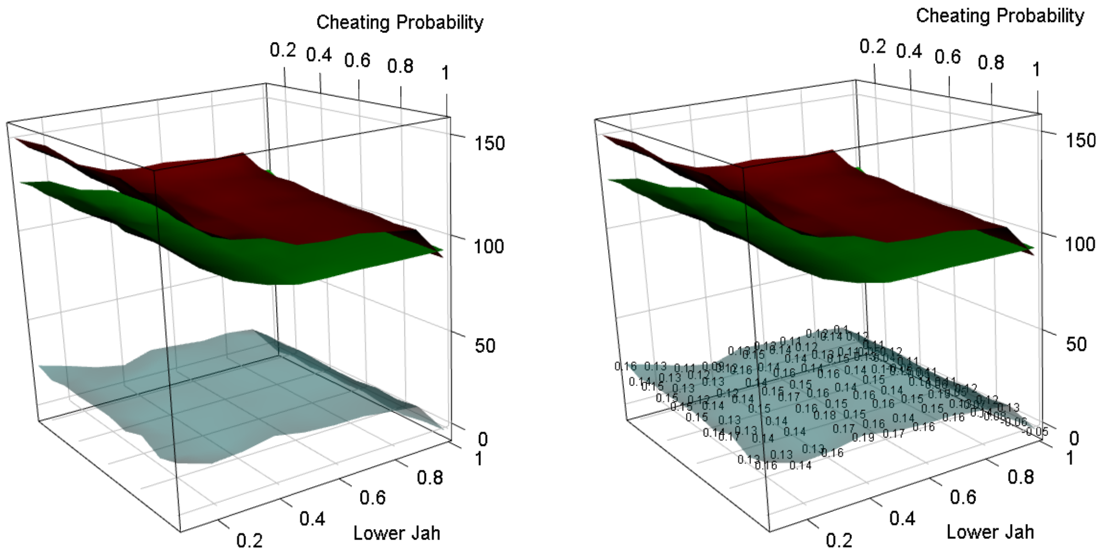
Identifying the openness towards newcomers, i.e. lower jāh difference, as more decisive for the outcome, could the cheating probability of cheaters, and thus the chance of being detected, have impacted the overall outcome?

Cheating Probability and Lower Jāh Difference Analysing the interaction between the cheating probability ($p_{cheating}$) and the lower jāh difference (see Figure 4.8 and Table 4.6), similar to the previous configuration we can observe that an increasing acceptance of newcomers drives their detection. The benefit of the apprenticeship system is near constant for all acceptable jāh difference values < 1.0. As expected, increasing the probability of cheating improves the detection and removal of cheaters from the system (significant strong correlation: -1). Similarly, varying *lowerJahDifference* shows a significant improvement of relative performance of the apprenticeship system (0.915).

A key effect of varying the cheater probability is the increase of cheater detection during earlier simulation rounds, leading to a reduction of mean cheating levels. However, due to the long-lasting nature of apprenticeship relationships, cheating is eventually detected, even if the probability is low. From an institutional perspective, differing cheating probabilities are thus accommodated by sufficiently long apprenticeship periods.

[27] Recall that this flexibility was a key economic benefit of the informal ṣuḥba compared to more formally regulated instruments such as the commenda (see Subsection 3.2.2).

[28] This effectively represents a small-world network consisting of communities without interconnecting links.



Red surface: without apprenticeship system
Green surface: with apprenticeship system
Blue surface: inverted absolute cheater reduction by apprenticeship system
The right chart additionally shows the relative cheater reduction for each data point.

Figure 4.8: Number of Cheaters with Relationship to Non-Cheater(s) for Varied Cheater Probability and Lower Jāh Difference

Table 4.6: Statistical Measures of Relative Differences and Absolute Values for Cheating Probability vs. Lower Jāh Difference

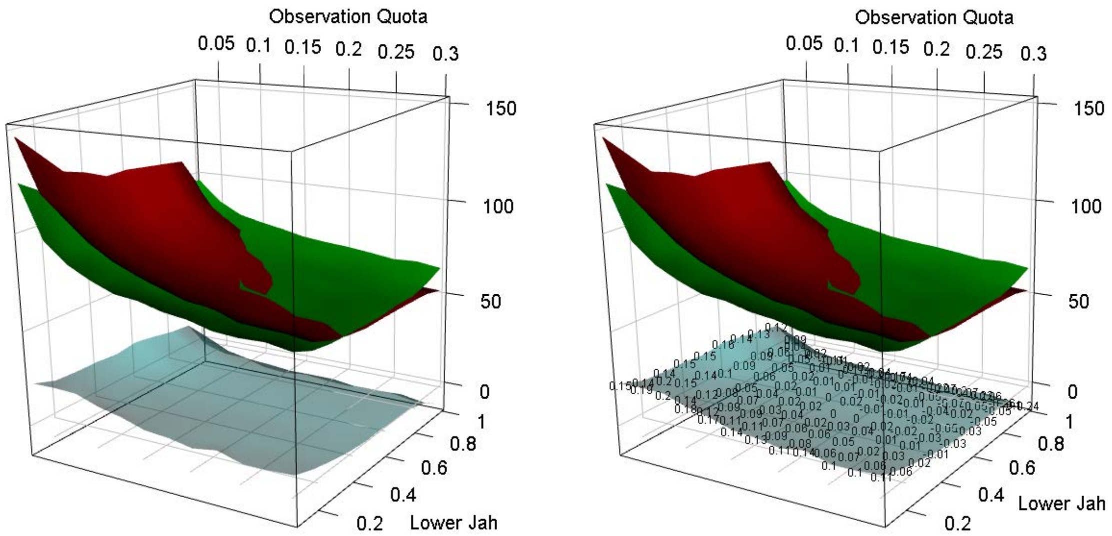
Relative Difference				Correlation of Mean Relative Diff. with	
Min.	Max.	Mean	σ	Cheating Prob. (p)	Lower Jāh Diff. (p)
-0.082	0.186	0.119	0.062	0.915* (0.00046)	-0.539 (0.1133)

Absolute Values				Correlation of Absolute Values with	
Min.	Max.	Mean	σ	Cheating Prob. (p)	Lower Jāh Diff. (p)
92.256	127.533	104.027	10.277	-1* (2.2e-16)	-0.951* (2.2e-16)

Another parameter that is indicative for the functioning of the coalition is the observation quota (*observationQ*), since it represents the monitoring function, which is particularly decisive in informal institutions.

Observation Quota and Lower Jāh Difference The combination of lower jāh difference and the fraction of fellow agents observed by a trader during each round (*observationQ*) allows us to model both the 'openness' of a given group as well as the extent of monitoring.

As before, the lower jāh difference is varied between 0.1 to 1.0 (step size: 0.1), while the observation quota ranges from 0.025 to 0.3 (step size: 0.025). The parameter range choice is driven by pragmatic concerns, given that the observation of as many as 0.3 of all fellow agents during a trade interaction appears unrealistically optimistic. The results displayed in Figure 4.9 show the expected effect of increased monitoring in combination with increasing lower jāh difference, with an observation quota of 0.3 reducing mean cheater numbers to around 65 for closed groups and to approximately 58 cheaters for more open groups (see Table D.6 for absolute numbers). However, as explored with respect to upper and lower jāh difference, in the extreme case of unconstrained access (*lowerJahDifference* = 1.0), the ability to bypass the apprenticeship system limits its effectiveness, increasing cheater levels to around 63.



Red surface: without apprenticeship system
Green surface: with apprenticeship system
Blue surface: inverted absolute cheater reduction by apprenticeship system
The right chart additionally shows the relative cheater reduction for each data point.

Figure 4.9: Number of Cheaters with Relationship to Non-Cheater(s) for Varied Observation Quota and Lower Jāh Difference

The statistical results in Table 4.7 show that monitoring has a very strong impact on the effectiveness of the overall ability to remove cheaters from trade interactions ('Absolute Values'), while *lowerJahDifference* shows an insignificant impact.

However, while increased monitoring drives reduction of cheaters, its impact on the functioning of the apprenticeship system is adverse. Increasing observation is negatively corre-

Table 4.7: Statistical Measures of Relative Differences and Absolute Values for Observation Quota vs. Lower Jāh Difference

Relative Difference				Correlation of Mean Relative Diff. with	
Min.	Max.	Mean	σ	Observation Quota (p)	Lower Jāh Diff. (p)
-0.257	0.204	0.023	0.097	-1* (2.2e-16)	-1* (2.2e-16)

Absolute Values				Correlation of Absolute Values with	
Min.	Max.	Mean	σ	Observation Quota (p)	Lower Jāh Diff. (p)
57.390	124.628	72.104	12.751	-1* (2.2e-16)	-0.454 (0.1909)

lated with the effectiveness of the apprenticeship system, since higher frequency in combination with extensive observation by full traders makes cheater control upon immediate admission to the entire trader network more efficient (with the apprenticeship variant performing worse by nearly up to 0.26).

Increasing the lower jāh difference drives inter-generational trader relationships along with faster identification and denunciation of cheaters. Increasing the monitoring likewise leads to a faster identification of cheaters, with a lower benefit of the apprenticeship system, given the more frequent observation by external observers (i.e. other agents that merely observe cheaters without having an apprenticeship relationship with them). For sufficiently high tolerance towards lower jāh levels (> 0.4) and higher observation levels (> 0.15), the apprenticeship system performs worse than directly interacting with newcomers without prior apprenticeship period. This observation reinforces the suggestion that the apprenticeship system's benefit was the early identification of cheaters based on relatively few interactions, highlighting the efficiency aspects of that institution (Williamson, 1998).

However, the analysis of the underlying implications of varying 'tolerance towards newcomers' and 'monitoring' reveals another aspect: the reduction of uncertainty (North, 1991). Assuming a risk-averse trader collective with relatively limited openness towards newcomers (i.e. low permissible relative jāh difference), the apprenticeship system concentrated the task of cheater detection onto a self-selected subset of the entire trader coalition, namely the apprentices' mentors. Ideally, 'full traders' within the coalition could thus rely on a relatively cheater-free environment once traders had passed the apprenticeship phase.[29] We suggest that the apprenticeship system converted the effortful task of cheater identification into a profitable endeavour, in which traders could 'invest' by offering their expertise and involving in intensive monitoring at the risk of identifying a cheater or having an unskilled

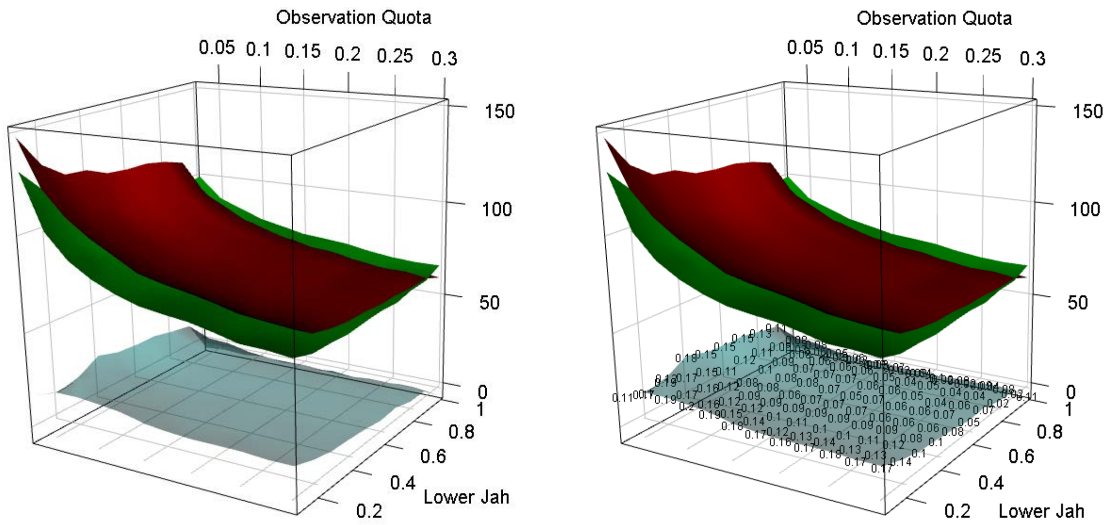
[29]Recall that our parameterisation does not reflect this due to the intentionally high level of potential cheaters.

newcomer as their apprentice, both of which would incur loss. The incentive for potential mentors to engage in such system was the procurement of cheap trade services, given that the mentor (at least initially) did not need to 'pay', i.e. reciprocate the apprentice's services, other than sustaining his living.

In effect this mechanism provided an entry barrier to newcomers and addressed (if not solved) the problem of sustaining the collective's 'Commons', a cheater-free environment in which traders could concentrate on flexible market-oriented professional arrangements at the benefit of the entire coalition, instead of bearing the constant risk of attracting cheaters. The announcement of cheaters by mentors – in contrast to the Genoese individualistic strategy that discouraged from sharing information in order to maintain competitive advantage (see Subsection 4.2.1) – was assured by attaching the mentor's reputation to the apprentice's conduct. Furthermore, using the institution 'apprenticeship system' and attaching newcomers to their mentors, access to the coalition could be governed in a decentralised manner, as opposed to relying on multilateral approval of newcomers, thus representing the more fluent characteristics and flexibility Goldberg (2012c) ascribed the trader network. Though the simulation does not explicitly point us to the institutional characteristics, rebuilding the apprenticeship system allows the exploration of boundary cases, and reveals its essential purpose of governing the Maghribīs' Commons – their relationship network.

Reducing Interaction Frequency To support our claim that the apprenticeship system's relative benefit increased with limited interactions but unchanged comprehensive monitoring of apprentices, we reduce interaction frequency, which limits monitoring opportunities. In this model interaction is represented as the trade quota *tradeQ*, i.e. the number of partners a trader interacts with during each round. We initially used an optimistic value of 0.5, which we reduce to 0.1. For apprentices (which are not affected by the trade quota) we reduced the number of interactions from six interactions per year to 1, thus a rather pessimistic value. The results of this parameter modification are shown in Figure 4.10 and Table 4.8.

The results support the suggestion that in environments with limited interaction frequency, the advantage of unconstrained interaction partners melts and the apprenticeship system generally outperforms direct employment without preceding apprenticeship, with exception of the extreme case of an inherently open society (with *lowerJahDifference* = 1.0), allowing immediate employment of newcomers. For the explored parameter space the advantage of the apprenticeship version lies at a mean of nearly 8 percent. This is to be compared to a value of just above 2 percent for the variant with higher interaction frequency, with only a moderately higher mean number of cheaters (around 75 compared to 72 for the



Red surface: without apprenticeship system
Green surface: with apprenticeship system
Blue surface: inverted absolute cheater reduction by apprenticeship system
The right chart additionally shows the relative cheater reduction for each data point.

Figure 4.10: Number of Cheaters with Relationship to Non-Cheater(s) for Varied Observation Quota and Lower Jāh Difference with Reduced Interaction Frequency

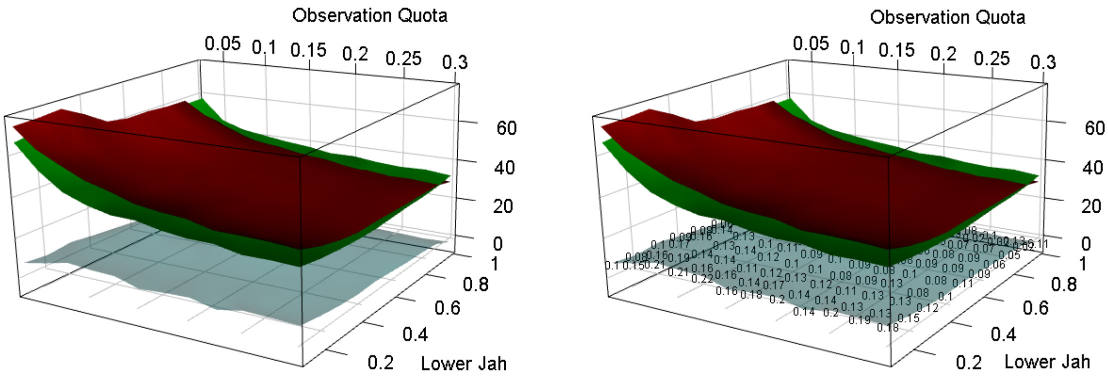
Table 4.8: Statistical Measures of Relative Differences and Absolute Values for Observation Quota vs. Lower Jāh Difference with Reduced Interaction Frequency

Relative Difference				Correlation of Mean Relative Diff. with	
Min.	Max.	Mean	σ	Observation Quota (p)	Lower Jāh Diff. (p)
-0.111	0.195	0.079	0.069	-0.881* (0.00019)	-1* (2.2e-16)

Absolute Values				Correlation of Absolute Values with	
Min.	Max.	Mean	σ	Observation Quota (p)	Lower Jāh Diff. (p)
59.310	131.380	74.625	12.970	-1* (2.2e-16)	-0.352 (0.279)

high-frequency variant).

Reducing the Number of Traders A question that lies at the heart of informal enforcement is its scalability. Since the actual number of traders is uncertain, it appears purposeful to test the relative performance difference for varying trader numbers. Implicitly this tests the refined network structure that is decomposed into more elementary aṣḥābs with overlapping links. Modifying the previous scenario (*tradeQ* = 0.1), we reduced the number of traders by half to now 200 agents. The results, observable in Figure 4.11 and Table 4.9,[30] show a mean improvement in cheater removal by the apprenticeship system of around 0.006 (0.085 for 200 agents; 0.079 for 400 agents), suggesting nearly linear scalability of effectiveness for the explored range.



Red surface: without apprenticeship system
Green surface: with apprenticeship system
Blue surface: inverted absolute cheater reduction by apprenticeship system
The right chart additionally shows the relative cheater reduction for each data point.

Figure 4.11: Number of Cheaters with Relationship to Non-Cheater(s) for Varied Observation Quota and Lower Jāh Difference (Reduced Interaction Frequency, 200 Agents)

Table 4.9: Statistical Measures of Relative Differences and Absolute Values for Observation Quota vs. Lower Jāh Difference (Reduced Interaction Frequency, 200 Agents)

Relative Difference				Correlation of Mean Relative Diff. with	
Min.	Max.	Mean	σ	Observation Quota (p)	Lower Jāh Diff. (p)
-0.164	0.217	0.085	0.075	-0.413 (0.1845)	-1* (2.2e-16)

Absolute Values				Correlation of Absolute Values with	
Min.	Max.	Mean	σ	Observation Quota (p)	Lower Jāh Diff. (p)
28.519	67.823	38.116	8.698	-1* (2.2e-16)	-0.357 (0.3128)

Although monitoring is decisive for the absolute reduction of cheaters (significant correlation of observation quota of -1), it plays little role for the relative performance of the apprenticeship system in the light of unchanged interaction frequency (insignificant correlation

[30]Equivalent to other respective configurations, an enlarged data-point surface plot and underlying tabular data is provided in Appendix D.5.

of -0.413). Instead the accessibility of established traders to newcomers (lower jāh difference) shows greater relevance, since it – as in all previous cases – defines the extent to which the apprenticeship system can be bypassed, or alternatively how quickly newcomers can engage with established traders. However, in contrast to its effect on the relative performance lower jāh difference has insignificant impact on the absolute change in cheater numbers.

Overall the observations for reduced trader numbers suggest that, in the explored range, the apprenticeship system's effectiveness is relatively robust against changing coalition size.

Discussion

Reconstructing the Maghribī relationship network around the assumption that jāh was of central importance to the network participants, and furthermore, that traders preferred partnerships with fellow traders of similar standing, we explored the central properties of the institution 'apprenticeship system'. It 'privatised' the problem of identifying cheaters from the collective of established traders to a self-selected subset of 'skill investors' that developed new traders both from a perspective of compliance but also trading. Their return was an established relationship with promising emerging traders and, while honing their skills, benefiting from lesser remunerated trading services. The risk lay in the cost associated with the time invested in developing a non-cooperative or untalented newcomer, and the loss associated with being cheated. For the collective this mechanism offered an effective way for established traders to rely on a cheater-free Commons. Non-compliant traders were under strict monitoring during their apprenticeship and could thus be filtered prior to entering the 'full trader' status giving them full access to the collective's relationship network collective. This prevented those from destabilising the effectiveness of the informal instruments, such as the ṣuḥba, whose adoption and distributed enforcement was crucial for the coalition's efficient market operation. The apprenticeship system thus reduced the uncertainty of engaging in a trade relationship with a cheating trader, while, at the same time, increasing the efficiency of market operations by admitting not only cooperative but likewise talented traders to the collective, an aspect we did not explicitly explore in this model. The system's effectiveness, i.e. the ability to exclude cheaters, – relative to the direct involvement of newcomers in trading – largely relied on the extent to which the collective was open to newcomers (assuming that newcomers are willing to participate based on a sufficiently high *upperJahDifference*) and number of trade interactions, with lower levels of openness and trader interactions providing a better relative performance. A secondary finding is the limited relevance of the cheater probability, given that the long duration of apprenticeships would eventually lead to detection of cheaters.

Based on those observations, we suggest that the apprenticeship system is a possible substitute for the 'closedness' assumption held by Greif (see Subsection 4.1.1). It explains the inclusion of newcomers and their intergenerational exchange with relatively low cheater levels while maintaining clear group boundaries. The 'closed group assumption' put forth by Greif offers a rigid simplification that cannot capture the fluid nature of the 'coalition' (see Goldberg (2012c)) and its persistence of the coalition over multiple trader generations. As such the apprenticeship system represented a suitable measure of cheater control matching the characteristics of the Maghribī trader society. For the Genoese society, in contrast, an apprenticeship system for the purpose of cheater detection would have hardly been useful, given the limited chance for repeated employment of operators by investors, along with the strong willingness to engage with opportunistic and potentially inexperienced newcomers – as long as the relationship was formally secured. This should not be confused with an assertion for the absence of skill-based apprenticeships in medieval Europe. In fact, Genoese were employing a comprehensive set of maritime apprenticeship schemes[31] for different crafts, generally with tangible products as outcomes, such as oar-making or ship carpentry. However, trading did not appear to be a tangible 'skill'; the facilitation of long-distance trade services did not require documented experience or quality assessment, other than honouring contractual commitments.

The model explored here concentrates on the apprenticeship system's effectiveness in identifying and excluding cheaters from the trader coalition. With this relatively puristic intent, we did not consider detailed aspects that would have possibly offered a more realistic representation of human behaviour, but would have distorted the objective of this model. One aspect we have discussed before is the simplistic jāh representation, with a monotonically increasing value attached to the individual trader. This objectified representation exists in contrast to alternative concepts, such as endorsements (Cohen, 1985; Alam et al., 2010), which emphasise the differentiated nature of reputation and its fluctuation. However, we maintained the simplified representation in order to isolate cheater propagation from changing employment levels based on fluctuating reputation levels. Furthermore, the model does not reflect aspects such as experiential and social learning. Newcomers cheat based on parameterised cheating probabilities and do not adjust this behaviour based on feedback or observation (e.g. observing other cheaters' fate, or adopting other traders' norms). A final aspect is the unambiguous detection of cheating behaviour, an aspect that serves the objective of understanding the apprenticeship system, as opposed to providing an accurate representation, which – as mentioned previously – is challenged by the lack of information in the first

[31] A listing of recognised skills has been documented by Epstein (1996).

place. Whether observed as apprentice or fully established trader, in this model cheating can be determined with full certainty.

More generally, modelling the historical Maghribī Traders Coalition is fundamentally associated with the concept of social control, the "capacity of a society to regulate itself according to desired principles and values" (Janowitz, 1975), in order to sustain cooperation. While for Genoese a formal institutional environment appeared necessary, the Maghribīs posed an interesting case. Their compliance appeared seemingly wilful, leading Greif to ascribe Maghribīs a propensity for sharing on a cultural basis, such as relying on the effects of gossip for social control (Giardini and Conte, 2012). An alternative perspective could be based on the existence of indirect reciprocity (Nowak and Sigmund, 1998, 2005) (i.e. expectation of 'repayment' of cooperate acts by third parties) or strong reciprocators (Gintis, 2000; Bowles and Gintis, 2004) (i.e. reciprocal acts despite absence of reward), which is compatible with what Ackerman-Lieberman (2012) described as 'esprit de corps' with reference to Greif's conceptualisation of the individuals' devotion to the coalition. However, following the spirit of our work – the reduction of assumptions – we retract from this strong assumption of multilateral commitment. Instead the adherence to the partnership obligations laid out by the ṣuḥba (trading and reporting), and thus to the immediate partners, in combination with the unknown observers (interlinked by potentially unknown ṣuḥba relationships), is sufficient to explain Maghribīan cooperation. Beyond this, our experiments suggest that the apprenticeship system employed by the Maghribīs did not rely on selfless commitment to the 'Greater Good'. Instead, the apprenticeship system had entrepreneurial characteristics, since it allowed for an investment in promising newcomers on the part of self-selected mentors, in response to the carrot of long-lasting cheap services and the stick of bearing the cost of attracting cheaters or untalented apprentices.

4.3 Consequences of our Experiments

Returning to the greater picture on the Genoese and Maghribī societies, we addressed two essential assumptions that had been insufficiently explored in previous work. At this stage, we summarise the results and offer a refined perspective on the respective societies, with focus on the Maghribī society in particular.

Summarising the Experiments The first experiment challenged the degree to which strict non-communication was firstly realistic, given that a strategy of misinformation may have been more promising in deceiving fellow traders and establishing a competitive advantage.

Furthermore, inherent secrecy would have hardly been helpful to explain the evidently existing private-order enforcement which relied on a minimal level of communication. The results suggest that even when accommodating optimistic initial levels of trust, cheater control purely based on informal means, i.e. by mutual advice, would not have sufficiently reduced cheater levels to promote a cooperative outcome – independent of the assumed underlying network structure. From this perspective, relaxing the assumption of non-communication thus does not challenge Greif's results; Genoese would not have cooperated solely relying on informal means.

The second experiment shifted the view to the contrasted Maghribī traders coalition that operated purely based on informal means. In contrast to the bifurcation into 'open' and 'closed' society, as assumed by Greif, we softened the 'closedness' assumption in order to accommodate further aspects of historical reality, such as the coalition's persistence over multiple trader generations and the documented existence of an apprenticeship system. Based on experimental exploration, we suggest that the apprenticeship system itself could have represented an institutional mechanism that acted as an entry barrier to the cheater-free trader coalition, both serving the purpose of quality assurance of the privileged members of the coalition, while at the same time filtering non-cooperators. Based on this mechanism Greif's 'closed group' assumption can be relaxed and instead supports the view of the Maghribī Traders Coalition as a 'relatively closed' group,[32] without making concrete quantitative claims, since historical cheater levels (or more exactly, levels of cheating intent) are unknown.

The model further captures a revised view on the structural aspects of the relationship network. We model the overall network (aṣḥābunā) as an emergent property of intimate associate networks (aṣḥābs) of overlapping dyadic relationships, in which the partial ignorance about relationships (and thus unknown information network), along with a strong emphasis on reputation, motivated compliance. Those structural aspects offer a refined explanation of high compliance levels among Maghribīs, as opposed to interpreting the coalition as an explicit transparent multilateral relationship network relying on private-order enforcement based on a shared devotion towards the group.

Shifting from the thesis of inherent cultural difference towards a perspective that integrates the recent contributions by Geniza researchers discussed in Chapter 3, let us summarise the overall picture and plot out a refined scenario.

[32]Recall the inversely related effectiveness of the apprenticeship system for increasing openness towards newcomers (see Subsection 4.2.2).

A Refined Interpretation of the Maghribī Traders The Maghribīs were a minority in a large empire expanding across nearly the entire Mediterranean. Though culturally distinct from their environment, the rulers of the surrounding empires gave them sufficient autonomy to conduct long-distance trade, at that time a challenging coordination problem. But the Maghribīs converted their seemingly weak position into a comparative advantage by drawing on communal ties that allowed them to operate based on informal institutional mechanisms. Those bore the significant advantage of flexibility (i.e. relative autonomy of partners to react to market fluctuations; multitude of selectively activated partnerships across different trade locations) compared to conventional formal institutional instruments such as the ʿeseq (Jewish commenda equivalent – see Subsection 3.1.3); the ʿeseq required clear ex-ante specification of obligations (clear instructions, distribution of profits, inability to flexibly reallocate resources during contract execution), and thus limited the reaction of changing circumstances in developed Eastern buyer markets. Maghribīs relied on an exclusive network that was held by a mix of instruments to govern controlled access (apprenticeship) and monitoring (proactive sharing of information as part of ṣuḥba relationship), with a structure that offered great scalability potential while reinforcing compliance by obscuring a global perspective of the coalition for individual members. The long-running nature of ṣuḥba relationships along with strategies of delayed clearing of accounts acted as further reinforcement and lowered transaction costs.

Whether by intent or happenstance, the Maghribī trader network showed a flexible adjustment to its institutional, social and economic environment. The same aspects help to understand why Maghribīs were unlikely to penetrate al-Rūm ('The Land of the Romans') and be active in European markets. The Maghribī network was adapted and dependent on a benevolent homogeneous institutional environment offered by the Fatimids, a united empire whose internal borders were fluid (see Goldberg (2012c)). Their minority status did not give Maghribī sufficient political influence or resources to establish themselves autonomously or even drive technological advancements. From the perspective of the institutional environment (see Appendix A) medieval Europe was in the opposite situation of the Fatimid Empire, with its particularised sovereignties and aggressive social environment characterised by war, and thus hardly beneficial for minorities. The political uncertainties of continuous conflicts and religious discrimination limited the security of minorities. Small sovereignties of limited reach and initially limited trade privileges[33] challenged long-term economic relationships, a cornerstone of the Maghribīan recipe for success. However, the seller-centric nature of Euro-

[33]Compared to Venice and Pisa, Genoa was comparatively late to establish exclusive trade privileges in remote trade locations (see van Doosselaere (2009)).

pean markets allowed the profitable operation based on less flexible institutional instruments (with the European commenda showing very limited flexibility with respect to profit distribution in contrast to its Arab relative, and certainly compared to the ṣuḥba) that concentrated on short-term relationships and inherent dependence on public-order enforcement mechanisms such as commercial courts. This thesis of environmental adaptation is supported by the fate of the Maghribīs. After being challenged by Italy's naval domination of the Mediterranean, Maghribīs were successively forced to shift their trading activities towards the Indian ocean, an endeavour that was ultimately terminated by the institutional environment they depended on: the Muslim Egypt rulers forced them to withdraw from long-distance trade indefinitely (Greif, 1993), putting an end to the Maghribī Traders Coalition.

Though cultural aspects, here collectivism vs. individualism, especially with respect to the influence of the Church in Europe (see Subsection 3.3.1), are not negligible in the historical development, their relevance was possibly limited compared to the concrete political, social and economic environment that shaped, enabled, and constrained both societies.

Benefits of Applying Agent-Based Modelling The results of the experiments do not challenge historical happenstance, the 'what', but they offer refined explanations of the 'why' and 'how' for the historical case. Here, the utilisation of agent-based modelling offers the advantages of hypothesising and exploring refined configurations by offering modelling metaphors that match the social reality. This offers the benefit of adapting the modelling detail according to available information, at the risk of filling information gaps with modelling intuitions and varying levels of modelling detail. Game theory offers a consistent and well formalisable approach on a high abstraction level, leading modellers to opt for strong assumptions (here: cultural differences) as a subsumption of the more complex set of interdependent influences (cultural influences, political environment, different stages of market development/professionalisation, social stratification of traders, etc.) that act in social scenarios. Along with sharp bifurcations necessary to maintain the comparative nature (here: Genoese vs. Maghribīs), such models risk the loss of scenario-specific peculiarities (such as the apprenticeship system). We believe that the more amenable nature of agent-based models complements the rigid character of game theoretical models, filling in the 'grey areas' by modelling refined sub-aspects or testing alternative hypotheses that are hard to model based on the rationality assumption (see related discussion in Subsection 4.1.2).

Turning to the Third Assumption To this stage we have explored the trader scenario with respect to two assumptions. The third assumption, the failure to consider the differentiated role stratification among Genoese in contrast to Maghribī traders, i.e. the assumption that role conceptualisations were identical in both societies, has not been explicitly addressed in the previous models.

From a sociological perspective it has been undisputed that labour specialisation changed societies' structure from integrated overlapping comprehensive role understandings by relying on jobs of limited specialisation, to more stratified and specialised role concepts. The emerging depersonalised functional interdependence of specialised skilled individuals – without the flexible assumption of different occupations – generally drives productivity increase and thus theoretically overall welfare, finding great support by Smith (1776) and Babbage (1835), central architects of the industrial revolution. However, the concept of specialisation has been recognised as far back as the fourth century BC.[34]

However, this stronger interdependence afforded an emphasis on what we previously introduced as formal institutional instruments (e.g. contracts) and a supporting environment (e.g. courts). The arising social solidarity is well reflected in what Durkheim (1933) contrasts as *mechanical* vs. *organic solidarity*. While both affect the overall benefit, he sees *mechanical solidarity* as the foundation built on aligned world views as well as common moral and cultural background, well reflected in the moralistic underpinnings of criminal law. It is a fundamental characteristic of kinship-based societies with stronger adherence to absolute authority, along with penal law and repressive sanctions (Durkheim, 1933; Merton, 1994). In an *organic form of solidarity*, the overall benefit, as with Smith's increase in productivity, is an emergent phenomenon of interdependent specialists. Since common values, based on the divergent understanding are of lower relevance and gradually replaced by an 'organ-like' interdependence of individuals, Durkheim (1933) sees it as a fostering driver for the establishment of civil law. This shifts from an emphasis on punitive measures towards the reestablishment of the status quo by applying restitutive sanctions, such as the redistribution of property, and complementing the absolute legal standards with domain-specific law that reflects the specialisation in different fields of economic activity.

Thus, assuming a behavioural perspective (one of the underlying premises of this work) and adopting a perspective of individuals as experiential learners, shaped by feedback they receive from their physical and social environment, one can propose that a differentiated role understanding, as entertained by Genoese traders, was less likely to assure compliance if the exhibited roles by themselves had conflicting economic interests. This perspective is reinforced by the strong social stratification that contrasted affluent Genoese investors

[34]Xenophon (1914) (ca. 350 BC) highlights the inability for individuals to perform all tasks necessary for survival, but can do so when interacting with others that specialise in specific tasks. Ibn Khaldun (1967) (ca. 1400 AD) made a similar observation, emphasising the productivity increase based on cooperation.

from opportunistic operators of low social standing.[35] The unified role understanding and thus integrated interest of the Maghribīs, in contrast, both influenced by their homogeneous middle class affiliation and intimate reciprocity relationships, in our view likely contributed to the cooperative outcome of their operations.

We thus hypothesise that the different environmental stimuli were relevant to drive traders into cooperation or non-cooperation. Analysing this aspect opens a social-psychological dimension of the scenario that has not found consideration in the previous economically motivated work. The more comprehensive consideration of cognitive aspects and their influence on social processes, beyond the focus on the psychological domain, as demanded by Parsons (1976), is an understanding we share when turning towards the modelling of institutions from a dynamic perspective. However, in order to maintain a comparative perspective, i.e. the ability to contrast Genoese and Maghribīs, we will step back onto a higher abstraction level that limits the modelling of simulation detail, but not so far as to detach us from the relevant social concepts such as roles.

[35]Without discussing this aspect in more detail, the social stratification caused by the establishment of influential family firms is compatible with Marx's (1977) differentiated view on economic and social division of labour, with the latter being driven by interests related to power relationships as opposed to actual economical necessity.

5

Towards a Dynamic Perspective on Institutional Modelling

To support the claim that the differentiated role understanding influenced decision-making and had impact on the cooperative outcome, we require an extended frame of observation that helps us understand how far the internalisation of experiential feedback could have shaped the individuals' habiti (see Subsection 2.1.1). In the previously developed scenarios, we introduced models of informal communication patterns (see Subsection 4.2.1) or concrete institutions such as the Maghribīan apprenticeship system (see Subsection 4.2.2) and observed the macro-level effects. In this second part of the thesis, we put a stronger emphasis on how individuals may have developed institutional understanding based on played roles, thus maintaining the subjective institution perspective (see Subsection 2.1.3). Focusing on a dynamic perspective, we show the development of institutional understanding over time, an aspect that reinforces the usefulness of agent-based modelling as a means to represent the micro-macro interactions in institutional environments. Our approach attempts to narrow the gap between the *legalistic view* on norms as imposed or perceived duties and the *interactionist/emergentist perspective* that emphasises the emergent nature of norms based on the interaction among individuals (Conte and Castelfranchi, 1999; Castelfranchi, 2003). However, to determine the development of informal institutions as drivers for norm innovations, we intend to marry the expressiveness offered by *structured unified representations that capture the informal and informal range of institutions (i.e. conventions, norms, and rules)* with the *flexibility required to express emergent institutions*. The modelling of normative understanding, as one form of institutional understanding, based on experiential (or other forms of) learning along with the theme of minimising introduced assumptions will accompany us for this second part of this volume. In this conceptual chapter, we introduce an institution representation derived from Crawford and Ostrom (1995) that lends itself to reflect different types of institutions, governing the informal spectrum of conventions and norms, as well as formal rules. Though our work primarily focuses on informal institutions, we aim to provide a generalisable representation that not only describes different institution types but also reflects the transitions between them.

5.1 Unifying Institutional Representations

5.1.1 Criteria for Institution Representation

The choice of our institutional representation is driven by the challenge that the representation should manage the trade-off of sufficient representational detail and covering the strata of different institution types, i.e. the trade-off between usability and reusability. Criteria thus involve:

- *General Nature of Representation* – The structure needs to be general enough to capture institutions from arbitrary application contexts (e.g. the representation should not be tied down to traffic rules, etc.) and of arbitrary type (e.g. conventions, norms, rules). To make it useful for the evaluation of institutional scenarios, we require the consideration of two related sub-conditions:
  - *The structure can be used by agents* to represent dynamic institutional understanding independent of the context, i.e. agents can use the structure to identify and refine institutional understanding over time.
  - *The structure is accessible to experimenters*, i.e. human experimenters can interpret the developed institutional understanding independent from the application context. A general representation in the form of binary strings, for example, would thus not be sufficient.

- *Flexibility of Representation* – The structure must lend itself to flexible use by agents at runtime, while satisfying the comprehensive representation of institutions. Central aspects include:
  - *Runtime Instantiation and Modification* – The institutional representation can be invoked and used by agents at runtime.
  - *Transition between different Institution Types* – The representation needs to be able to accommodate the transition between different institution types, such as the shaping of conventions to norms or rules, at runtime.

Since our central interest lies in the behavioural emergence of institutions – deriving from the society itself as opposed to being explicitly imposed, the central focus lies on informal institutions, which, in the field of multi-agent systems and agent-based modelling, is associated with Normative Multi-Agent Systems (NorMAS), aspects of which we discuss in the following. Later, we will complement this understanding with insights from the field of institutional analysis.

5.1.2 Norm Representations in Normative Multi-Agent Systems

Relating a range of definitions from the norms literature, including Coleman (1990), Elster (1989) and Ullmann-Margalit (1977), Savarimuthu et al. (2013a) identify three essential characteristics of norms, namely

- Expectations associated with behavioural regularity,
- Norm enforcement mechanisms, and
- Norm spreading mechanisms.

In the field of NorMAS a variety of representations has been applied over time, with the most coarse differentiation being the *implicit* or *explicit* nature of the representation (Savarimuthu, 2011). With *implicit norm representations*, the existence and meaning of norms is not immediately accessible, but relies on the experimenter to infer it from the interpretation of the simulation outcome. With *explicit representations* on the other hand, norms and their content are directly expressed in a dedicated structural representation. Hollander and Wu (2011b) structure norm representation types into four strands, namely deontic logic, rule-based systems, binary strings, and game theory. We adopt and refine this structure based on accounts of norm representation found in the literature. We organise the representations by their descriptive weight, starting from implicit representations, such as game-theoretical equilibria and move towards approaches of increasing descriptiveness, such as rule-based representations. The continuum of identified norm representations along with the sub-types is shown in Figure 5.1 and introduced in the following.

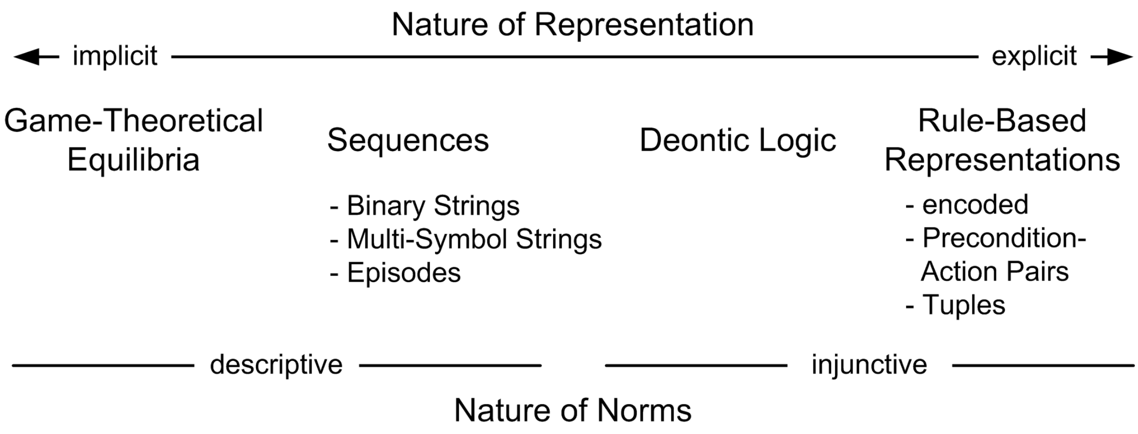


Figure 5.1: Norm Representation Continuum

Game Theoretical Equilibria

The analysis of norms and institutions in game-theoretical approaches relies on the identification of preferred strategies that can be analytically derived based on agents' action choices and associated pay-offs (Bicchieri, 2006; Epstein, 2007; Ullmann-Margalit, 1977; Young, 1998). The stability of those choices in the form of equilibria can then be indicative for the existence of norms (Mukherjee et al., 2007). The same approach is assumed in the context of institutional analysis based on Analytic Narratives (Bates et al., 1998) (see Subsection 2.2.3), which is the very approach Greif (2006) relies on for the representation of institutions. Since the identification of such equilibria requires analysis and interpretation on the part of the modeller without finding explicit structural representation, equilibria thus represent the implicit end of the representation continuum. In the context of institutional analysis, equilibria are associated with the notion of conventions or descriptive norms (see Subsection 2.2.1), thus describing observable rational behaviour without prescribing it. Further approaches applying a game-theoretical or 'game theory'-style representation include Shoham and Tennenholtz (1997); Perreau de Pinninck et al. (2008); Sen and Airiau (2007); Urbano et al. (2009); Sen and Sen (2010); Brooks et al. (2011); Sugawara (2011); Yu et al. (2013); Airiau et al. (2014); Franks et al. (2014). With 'game theory'-style we refer to the specification of cooperation/defection feedback in terms of pay-offs, since the representation of norms similarly relies on determined equilibria, even if the 'game' operation metaphor is not explicitly invoked.

Sequences

Sequence representations emphasise the dynamic perspective on norms, indicating their activation and salience over time but generally bear no or limited information about the actual content of the norm. Though norms are represented by primitive symbols, their interpretation relies on the ascription of meaning by the experimenter. With particular focus on the purpose of exploring norm emergence and dynamics over time, different forms of sequence representations can be observed.

- *Binary Strings* – Binary strings, i.e. strings consisting of zeros and ones, signify the momentary existence or absence of norms respectively, and according to Hollander and Wu (2011b), find particular application in the analysis of the transmission and emergence of norms.[1] Binary strings are only capable of indicating the activation of one specific norm, but cannot describe the existence of multiple concurrently activated norms.

- *Multi-Symbol Strings* – Multi-symbol strings (e.g. 'AUIFSD') facilitate the representation of activated alternative norm choices, with individual norms represented as symbols from a problem-specific set of alternative choices.[2] Similar to the case of binary strings, this approach is particularly useful to reflect changes in event sequences over time, but, in contrast to the former, extends this to multi-norm representation.

- *Episodes* – Norms can further be represented as recurring episodes in action or event sequences. By decomposing norms into the associated actions, this representation not only reflects multiple coexisting norms (as with representations based on multi-symbol strings), but adds semantic weight by operating on a more fine-grained level that allows the expression of normative content, such as the involved activities. Examples for this include action episodes (e.g. Savarimuthu et al. (2013b)) and event episodes (e.g. Savarimuthu et al. (2010)).

Deontic Logic

Deontic logic (von Wright, 1951a; Wieringa and Meyer, 1993), as a special case of classical logic, offers an approach to combine the dimensions of 'possibility' and 'necessity' by introducing primitives for the representation of obligations, permissions and prohibitions. A

[1] This representation is used by Caldas and Coelho (1999); Epstein (2001); Flentge et al. (2001); Nakamaru and Levin (2004); Galan and Izquierdo (2005); Hollander and Wu (2011a).

[2] Examples include Savarimuthu et al. (2007, 2008).

central characteristic is the interdefinability (von Wright, 1963) of the different deontic primitives, such as the ability to define a prohibition (*Pr*) by the logical combination of inverted obligation ($\sim$*Ob*) and inverted permission ($\sim$*Pe*), i.e. $Pr := \sim Ob \; \& \sim Pe$. Applications with a focus on deontic logic centre around the reasoning capabilities and generally rely on simplistic symbolic action representations (such as the primitive symbolic representation found in 'Episodes'). Despite relying on simplistic action representations, approaches based on deontic logic facilitate the transition from a focus on the descriptive norm perspective (i.e. the description of observable normative behaviour) to an injunctive norm understanding (Cialdini et al., 1990), the latter of which features a prescriptive component that describes the desirability of behaviour. Example applications are discussed in Boella and van der Torre (2006), Aldewereld et al. (2006), and Panagiotidi et al. (2014). However, in the light of increasing consideration of domain-specific contexts, approaches based only on deontic logic are increasingly replaced by rule-based representations that embed deontic primitives.

Rule-Based Representations

The representation of norms in the form of some notion of rules is most commonly found. Rule representations generally involve context-dependent behavioural prescriptions, along with potential consequences. Observing the literature, three main approaches can be identified: behavioural encoding, precondition-action pairs, and tuple representations. The latter two approaches can often hardly be distinguished, since precondition-action pairs can be interpreted as abstract variants of tuple-based representations. Furthermore, application examples for the latter two generally involve a combination of deontic logic and structural representation. In many instances rule-based representations can be interpreted as an extension to representations based on deontic logic. However, the introduction of detailed structural representations of normative content exists in a trade-off with puristic reasoning capabilities.

The first type of rule-based representations is not explicitly expressed and handled by agents, but it is explicitly specified by the modeller and *encoded in an agent's behaviour*. Since the behavioural encoding generally prevents the option for violation of such norms, these rather bear characteristics of fixed behavioural strategies or constraints. Representations of this kind have generally found application in earlier models of normative behaviour.[3]

In their most general form explicit rule-based norm representations consist of *precondition-action specifications*, in which an action consequence is activated under matching preconditions. The activated action (e.g. a violation), can satisfy other rules' preconditions, a

[3] Examples include Schelling (1978); Castelfranchi et al. (1998); Saam and Harrer (1999); Shoham and Tennenholtz (1992); Conte and Castelfranchi (1995).

means that allows one to represent the complexity of norm interdependencies. Example models that use explicit, predefined rule specifications include Villatoro and Sabater-Mir (2009) and Santos and Pitt (2014). Abstract representations that are used by agents to infer norms are used by Morales et al. (2013, 2014) as well as King et al. (2014), all of whose works feature the use of deontic operators. Testerink et al. (2014)'s work puts particular emphasis on norm interdependencies.

Though following a precondition-action specification, the representation can be of declarative plan-based nature and geared towards the implementation in intelligent agents and electronic institutions (Noriega, 1997; Rodríguez-Aguilar, 2001; Esteva et al., 2001; Esteva, 2003).[4] Declarative approaches that put particular weight on the deontic aspects of norms include da Silva (2007); Odelstad (2007); Kollingbaum et al. (2007); Campos et al. (2009). Declarative approaches with specific emphasis on temporal aspects such as norm expiry is explored by Riveret et al. (2007); Cardoso and Oliveira (2009); Criado et al. (2010).

Extended norm representations based on *tuples* or *classes* provide the structural scaffolding to capture a wider range of norm properties of general or customised application-dependent nature. Such properties include temporal and contextual preconditions for norm activation, post-conditions, associated actions, norm participants and potential sanctions.[5] As mentioned in the context of deontic logic approaches, the use of rules and deontic logic is increasingly interlinked, and such is the case for the works listed above. In those cases the nature of the norm is generally expressed using deontic primitives, identifying the norms as obligations, prohibitions or permissions. As such, most rule representations are not puristic but conflate elements from a deontic logic representation with a more refined description of general aspects such as actorship, action and context. Exceptions from the blended use of deontic concepts include Lacroix et al. (2009) and Boissier et al. (2011).

Specialisations that lie outside the categorisation of rule-based representations (but could be translated into those) are event-action trees. Those are used in EMIL-S (Lotzmann and Möhring, 2009) to represent potentially interlinked micro-level rules, and are explicitly represented in a normative board (Andrighetto et al., 2010a) (whose alignment on the macro-level is interpreted as normative behaviour). More complex logic representations related to model checking integrate temporal aspects into norm and institution representations as discussed by Viganò and Colombetti (2008). Examples for approaches that put particular focus

[4]Examples include Artikis et al. (2005); Vázquez-Salceda et al. (2004); López y López and Luck (2004); García-Camino et al. (2007); Aldewereld et al. (2007); Oren et al. (2009); Hübner et al. (2010); Panagiotidi et al. (2013); da Silva and Zahn (2014); Dybalova et al. (2014).

[5]Examples include Kollingbaum and Norman (2006); Gâteau and Boissier (2007); Montealegre Vázquez and López y López (2007); Oren et al. (2010); Ghorbani et al. (2013a); Jiang et al. (2013); Haynes et al. (2014); Li et al. (2014).

on expectation modelling and monitoring include work by Cranefield (2006, 2007, 2014).

Delineating the extreme end of norm representations, norm modelling languages, such as the Agent Modeling Language (AML) (Trencansky and Cervenka, 2005) and NormML (da Silva Figueiredo et al., 2011), rely on the Unified Modelling Language (UML) (Object Management Group, 2014) to provide metamodels for norm specifications. A comprehensive overview of rule-based norm specifications related to organisational metamodels and development methodologies in the context of Agent-Oriented Software Engineering (AOSE) is given by da Silva Figueiredo et al. (2011), whose work offers a suitable starting point to identify a general representation of institutions.

Summarising this brief overview, especially in the more recent NorMAS and electronic institution literature,[6] the focus on rule-based norm representations is dominant.

In our view many rule-based approaches have an association with what Savarimuthu and Cranefield (2011) describe as intentional *norm creation* in the context of the normative life cycle. Associated notions include *offline design*, *imposition by a leader*, or introduction by a *norm entrepreneur* (Savarimuthu and Cranefield, 2011).[7] In contrast to the 'intentional view' on norm development, which we see in closer proximity to the legalistic view on norms, exists the emphasis on norm emergence, which emphasises the interactionist view on norms.

Hollander and Wu (2011b) structure the view on *Norm Emergence* into three major categories, namely the *game-theoretical analysis of norm emergence*, the *influence of sanctions on norms emergence*, and last, *measuring the effect of norms transmission across participants*. Reflecting on different norm representations, the emergentist perspective is primarily associated with what we characterised as game-theoretical and sequence representations.

However, a few exceptions exist within the rule-based representation domain. Villatoro and Sabater-Mir (2009) propose a model of a gathering society, in which agents explore different precondition-action combinations to arrive at a subset of stable explicit rules. Corapi et al. (2011) propose a learning mechanism in which agents can refine declaratively specified norms based on inference processes. Further exceptions include the work around the EMIL project (Andrighetto et al., 2010a; Lotzmann and Möhring, 2009), whose essential purpose is to express norms in an accessible manner and model the two-way interactions of emergence and immergence (see Subsection 2.1.2). Possibly the most relevant exception from the offline norm design are rule-based approaches that establish norms at runtime. Examples include

[6] We see contributions to the COIN workshop as particularly indicative for this trend.

[7] The different approaches are analogous to what Mahmoud et al. (2014) describe as *Offline Design*, *Norms Autonomous Innovation* and *Social Power Mechanism*. Hollander and Wu (2011b) concentrate on *Offline Design* and *Autonomous Innovation*.

recent works on norm synthesis (Morales et al., 2013, 2014) that feature the identification of precondition-action statements along with deontic primitives to describe generalised norms.

5.1.3 Structural Components of Unified Institution Representation

As indicated before, a promising starting point for the provision of a general institution representation – which is the intent of this chapter – is the work by da Silva Figueiredo et al. (2011). Beyond the provision of a concrete norm representation, da Silva Figueiredo et al. (2011) propose an essential set of properties that should find consideration for any comprehensive norm representation, and – in our view – are likely applicable to institutions in general. We borrow the proposed properties and characterise those in the spirit of this work:

- Deontic concept – The deontic concept, borrowed from deontic logic (von Wright, 1951a; Meyer and Wieringa, 1993; Wieringa and Meyer, 1993), describes the nature of a norm describing behavioural restrictions in the form of obligations, prohibitions and permissions.
- Involved entities – Norm participants need to be explicitly described.
- Actions – Norms are related to specific actions they constrain or prescribe.
- Activation constraints – Norms may only be active during specific periods and may involve explicit activation and/or deactivation.
- Sanctions – Norms generally involve sanctions for non-compliance. Although many specifications do not explicitly describe sanctions, their nature as direct, indirect, material or emotional has been explored widely. An overview of aspects related to the sanctioning process is provided by Balke and Villatoro (2012).
- Context – Norms or institutions are applicable in particular contexts. Those may be implicit as part of the scenario, or explicit and entail organisational context (see e.g. Jiang et al. (2013)) and environmental context.

A representation that offers strong structural similarity with the aspects highlighted by da Silva Figueiredo et al. (2011) is Crawford and Ostrom's *Grammar of Institutions* (1995; 2005) (GoI), which is an integral part of Ostrom's institutional analysis framework IAD (Ostrom, 1990; Ostrom et al., 1994). Along with the compatible structural representation, it extends beyond norm representation and reflects what we seek: a generic institution representation that includes conventions, norms, and rules – capturing essential concepts from the subjective institution spectrum (see Subsection 2.1.3). Beyond this, the representation is intended to be aligned with a natural language human understanding of institutions, an aspect that is supported by its use in the context of policy-coding (see Subsection 5.2.4). Its use in the context of agent-based systems thus offers an accessible interface between agents and human modeller – an aspect that we specified as a requirement in the beginning of this chapter. In the following we will introduce the grammar and contextualise it with work within and outside the area of multi-agent systems.

5.2 Crawford and Ostrom's Grammar of Institutions

The Grammar of Institutions (GoI) (Crawford and Ostrom, 1995, 2005), or *ADICO* as we will refer to it interchangeably, was originally conceived to marry different approaches to institutional analysis, differentiated as equilibria ('shared strategies' or conventions), norms and rules (as discussed in Subsection 2.2.1), and to offer a general representation for institutions from a human-centred perspective.

5.2.1 Grammar Components

The ADICO grammar consists of five components. Those include (Crawford and Ostrom, 1995, 2005):

- ***A****ttributes* – Attributes describe the attributes and characteristics of social entities (which can be individuals or groups) that are subject to the institutional statement (e.g. shared strategy, norm, rule). If not specified explicitly, all individuals (or members of a group/society) are implied.

- ***D****eontic* – The deontic component uses a deontic primitive to describe an *obligation* (Ob) (represented by the term 'MUST', 'OBLIGED'), *permission* (Pe) (expressed using 'MAY', 'PERMITTED'), and finally a *prohibition* (Pr) ('MUST NOT', 'FORBIDDEN'). In Crawford and Ostrom's conception it captures the aspects of deontic logic, most importantly the interdefinability (von Wright, 1963) referred to above (see Section 5.1).

- *A****I****m* – The aim describes an action or outcome associated with the institutional statement. The only constraint put on an aim instance is that the action or outcome it describes must be physically possible, so as to avoid the expression of shared strategies, norms or rules that cannot be fulfilled, thus their non-/compliance not be determined (Crawford and Ostrom, 2005; von Wright, 1963).

- *Conditions* – Conditions capture the circumstances under which the statement applies. This can include spatial, temporal and procedural elements, such as preconditions that have to be met in order put the institutional statement into effect. The Conditions component can also be used to express linkages between institutional statements, such as execution dependency (Crawford and Ostrom, 2005). Similar to the Attributes, if not further constrained, conditions default to "at all times and in all places" (Crawford and Ostrom, 1995).

- ***Or** else* – The 'Or else' part describes consequences that are associated with the violation of the institutional statement, i.e. the combination of all other components used in that statement. In Crawford and Ostrom's grammar, this component has a constitutive role in classifying statements as rules, leaving potential violators with information about the threatened consequence. Note that the 'Or else' component has the special ability to manipulate other institutional statements' deontic primitive, in order to express changing rules for a violator, such as the introduction of an obligation or prohibition (e.g. change of a previous permission to participate in economic interaction to a prohibition as a consequence of cheating). Crawford and Ostrom specify three requirements for an 'Or else' statement: It needs to be an outcome of a decision-making process by a collective that has the power to do so (which need not necessarily be a government). It further requires the 'Or else' component to be supported by another norm or rule statement that modifies the assigned deontic under the condition that the first rule is violated. Finally, it requires the specification of a rule that specifies the responsibilities of a monitor. The 'Or else' component can further be expressed as an institutional statement itself.

5.2.2 Institution Types

The grammar components, representing the acronym ADICO, are used in varying combinations to describe what Crawford and Ostrom refer to as *institutional statements* that reflect specific institution types. Institutional statements are defined as "the shared linguistic constraint or opportunity that prescribes, permits, or advises actions or outcomes for actors (both individual and corporate). Institutional statements are spoken, written, or tacitly understood in a form intelligible to actors in an empirical setting" (Crawford and Ostrom, 1995), thus representing both the explicit and tacit character of institutions. Components of the ADICO grammar can be combined to represent different institution types, which we list in the following.

Conventions The combination of the subset of <Attributes, Aim, Conditions> (`AIC`) describes 'shared strategies', which in our understanding equate conventions and descriptive norms.[8] Using an example statement inspired by our Maghribī trader scenario, and annotating the relevant expressions with the respective grammar components, a *convention* can be:

Traders (A) trade fair (I) under any circumstances (C).

From a normative perspective conventions can be interpreted as *descriptive norms* (Cialdini et al., 1990), since they describe observable behaviour without prescriptive implications.

Norms In order to describe *norms*, the Deontics component is added to the convention, constructing an `ADIC` statement:

Traders (A) *must* (D) trade fair (I) under any circumstances (C).

Beyond the earlier description of behaviour, clear expectations are associated with the Trader role, making `ADIC` statements *injunctive* norms in contrast to *descriptive* norms.

Rules To describe *rules*, all of the grammar components are used to describe institutional structures that include sanctions for non-compliance, thus constructing `ADICO` statements:

Traders (A) must (D) trade fair (I) under any circumstances (C),
or else observers must report this deviation (O).

The notion of rules in ADICO is equivalent to the understanding of formal institutions, i.e. enforced by an appointed monitor. Only in this case the institutional grammar explicitly describes consequences. As indicated before, the 'Or else' statement can be a nested institutional statement in itself.

5.2.3 Operationalisation

To make this general-purpose grammar accessible for formal analysis, Crawford and Ostrom (1995, 2005) also propose a game-theoretical operationalisation of the syntax that introduces the notion of "deltas" to represent an individual's utility (and thus opting for a consequentialist view of norms (Anscombe, 1958)). Deltas are associated with the deontic of an in-

[8] Refer to Subsections 2.2.1 and 2.2.2 for our understanding of conventions as a 'shared strategy' among institution participants.

stitutional statement, and they stratify the utility of complying with norms (or rules) into punishments and rewards associated with its compliance (δ^o), as well as those associated with violating a prescription (δ^b). They further differentiate those two kinds of deltas by effects of

- *internal nature* ($\delta^{o,i}$), originating in the individual itself (such as a positive emotion ("warm glow" (Crawford and Ostrom, 2005)) or feeling of guilt, etc.), and
- *external nature* ($\delta^{e,i}$), reflecting the impact of external sources (e.g. physical or material threats or rewards by third parties) on the overall delta associated with the deontic.

Crawford and Ostrom align their stratification into 'internal' and 'external' deltas with Coleman's (1987) differentiation of "internalized norms" and "externally sanctioned norms".[9]

5.2.4 Applications

Institutional Analysis

The ADICO grammar is a constituent part of the IAD Framework (Ostrom et al., 1994; Ostrom, 2005b) (see Subsection 2.2.3). It provides the essential rule representation on the different analysis levels of the framework, such as the expression of operational rules, collective-choice rules, and constitutional choice rules. Based on its structure Ostrom and Crawford (2005) introduce further refined rule classifications (e.g. boundary rules, choice rules, aggregation rules) commonly found in institutions. This semi-formal description of institutions makes them accessible for economic analysis, such as using the previously mentioned game-theoretical approach. It also affords its use in a wide range of domains, such as the following.

Policy-Coding

Beyond the analytical use in the IAD framework, the ADICO grammar's structural representation facilitates the preceding step – the extraction and translation of policies into rules. For that purpose Basurto et al. (2010) introduce explicit instructions for the systematic analysis and coding of institutional statements using the *Institutional Grammar Tool* (as they refer to the GoI). Siddiki et al. (2011) refine this to react to challenges such as differentiating between aim and conditions sufficiently by introducing an o**B**ject component into the then *ABDICO* grammar.

[9] A detailed overview on the notion of "deltas", including examples, is given in Crawford and Ostrom (2005). A discussion, beyond the elaboration by Crawford and Ostrom, is offered by Schlüter and Theesfeld (2010).

Political Science

An example for the application of the grammar on different governance levels is given by McKenna (2013), who instantiates the grammar to model local political institutions by developing a typology of respective roles and rules.

Agent-Based Modelling

The first application of the ADICO grammar in the context of ABM was introduced by Smajgl et al. (2008). They proposed an agent architecture that embeds conceptual representations of the essential grammar components in order to build models that allow the consideration of existing ADICO rules as part of individuals' deliberation process. Individuals produce endogenous rule changes based on social comparison and a changing pay-off structure. The authors' approach requires a tight integration of grammar and agent architecture. Agents necessarily operate with ADICO *rules*, i.e. the most elaborated institution type, and do not reflect the different informal institution types *shared strategy* and *norms*. Extending on this work, Smajgl et al. (2010) apply their approach to the modelling of endogenous rule changes in the context of water use.

Multi-Agent Institutional Analysis

Important recent contributions that use the grammar in more depth include Ghorbani et al.'s MAIA (Multi-Agent Institutional Analysis) framework (Ghorbani et al., 2013b) that represents a comprehensive attempt to translate Ostrom's Institutional Analysis and Development Framework (IAD) (Ostrom, 1990; Ostrom et al., 1994; Ostrom, 2005b) into an agent-based operationalisation for the participatory modelling of socio-technical systems. It provides a guided approach and associated toolset that facilitates the modelling and analysis of specific scenarios, with the intent to produce executable simulation code for further exploration. As in Ostrom's original framework, the ADICO grammar plays the central role as a descriptive mechanism for formalising institutions. Moreover, Ghorbani et al. (2013a) explored the notion of shared strategies as a fundamental statement type and provided a nuanced differentiation into *common*, *shared* and *collective strategies*, and further included the consideration of temporal aspects.

Normative Multi-Agent Systems

An approach from the area of NorMAS includes Jiang et al.'s (2013) conceptualisation of Norm Nets to capture the organisational context of norms (using the grammar's norm and rule syntax), and reflect this across multiple nesting levels. They have conceived the concept of Norm Nets to provide the contextual frame for individual expressions. They showcase this approach to translate an EU policy into a multi-perspective representation operationalised using Coloured Petri-Nets (CPN) (Jensen, 1992). Similarly to Ghorbani et al.'s work, they assume a descriptive macro-level perspective on institutional analysis.

Observing the diverse areas of application and despite its wide use in the context of institutional analysis, the grammar's flexible and general nature has found surprisingly little attention in the computational domain beyond the mentioned approaches. This could in part be rooted in selected limitations, which have been met with suggestions for refinement, if not the grammar's reinterpretation. Schlüter and Theesfeld (2010) suggest amendments, some of which we share and address in this chapter, since they are in line with our objective: to leverage the grammar's use for a *bottom-up emergentist perspective*. All other previously mentioned approaches (with exception of Smajgl et al.'s approach (2008) that exclusively concentrates on ADICO rules) focus on a *descriptive perspective* that affords the a priori specification of institutional statements, as opposed to growing them from the bottom up.

5.2.5 Limitations

The differentiation between *shared strategy* (`AIC`), *norm* (`ADIC`) and *rule* (`ADICO`) purely based on syntax establishes a rigid and clear differentiation between those institutional statement types, which is truthful to the institutional understanding laid out in Subsection 5.2.2. However, a particularly striking problem, and well-acknowledged by Crawford and Ostrom (2005), is the *lacking specification of consequences for norms*. From the perspective of policy-making as well as coding, the bias towards rules for an explicit and clear description of sanctions is retraceable. However, this comes at the price of applying different measures for different institution types inasmuch as it leads to an incomplete description of norms, ascribing a relevance of sanctions only to rules that require explicit collective decision-making. Although collective decision-making is possible on different political levels, trimming norms by sanctions suggests the misconception of a natural progression of institutions from conventions to rules, risking to interpret norms as 'some interim state' on a convention's path towards becoming a fully fledged rule. On the contrary, if introduced rules are in conflict with existing norms, their adoption is likely to find strong resistance, especially when global laws are intended to supersede norms established on a local level, and worse, if those institutions are transferred from different cultures (Acharya, 2004; Thakur, 2001).[10] In this context, Schlüter and Theesfeld (2010) express the suggestion that the omitted sanction specification might lead to an interpretation of norms as "less serious" (Schlüter and Theesfeld, 2010) institutions. More importantly, we think, the limited representation challenges the original intent of the grammar, its strong generalisability, an aspect that Ostrom particularly highlights with respect to rules (Ostrom, 2005c). A weak norm representation challenges the grammar's ability to model a society's social institutions comprehensively – especially for societies that dominantly or completely rely on informal enforcement.[11] However, it is safe to say that this is far from the intent of the authors, since Ostrom's life achievement is explicitly centred around the decentralised self-governing of common pool resources, leveraging normative potential in the pursuit of conscious rule shaping.

An argument supporting Crawford and Ostrom's view is the often lacking (or for the observer, unknown) clear specification of norm sanctions in societies, though those nevertheless exist. Another stance in the defence of the authors is the concentration on an exogenous view, thus modelling norms from a) a macro-perspective (of the respective community) and b) an objectified outsider perspective who observes the existence of norms but may not able to identify associated sanctions,[12] especially when carrying consequences that are internal to norm participants (e.g. emotional) or the respective societies (e.g. reputation-based, such as loss of face). However, in order to model the onset and establishment of such institutions, an endogenous perspective cannot be neglected, since initially invisible social sanctions can become candidates for codified sanctions, and may thus surface, should the establishment of a rule occur.

To integrate the internal and external perspectives on institutions, the rigid categorisation of norms and rules can be revised in order to increase the inclusiveness of different institution types and to remove the bias towards formally specified rules. In fact, if we accept the emergence of institutions as rooted in habitual human behaviour (Veblen, 1899; Sumner, 1906; Parsons, 1951; Bourdieu, 1977) as put forth for this work (see Section 2.1), we must attempt to include the individual subjective perspective as a basis for socialisation processes (e.g. formation of customs and conventions, see Subsection 2.2.2) that shape increasingly prescriptive institutional constructs (e.g. norms and rules). The current GoI conceives in-

[10]Refer to the discussion in Subsection 2.2.2 on the necessary reflection of rules in norms as *enacted rules* for their sustained success.

[11]An example for this is our scenario of the Maghribī Traders Coalition that thrived based on informal institutions (see Section 3.1).

[12]This is not meant to be a criticism of the approach, since the outsider perspective may often be required to crystallise commonly followed institutions, which, especially in the form of conventions and norms, may be tacit to the institution participants.

stitutions from an economics perspective, assuming rationality for the analytical approach, which involves the choice of the unified analytical instrument game theory in order to complement the unified representation of institutions.[13] However, in line with our view that Greif's high level assumptions are in part produced by the use of game theory (Subsection 4.1.2), we believe that Ostrom's focus on 'games' obscures the greater potential of the grammar, especially with respect to the informal and subconscious dimension. We see this reflected in the concentration on rules and their refined classification as a central aspect of analysis in the IAD framework (Ostrom, 2005c; Ostrom and Crawford, 2005). The focus on conscious agency implies the avoidance of subconscious processes such as habit formation (see Parsons (1951); Waller (1988); Subsection 2.2.2). The game-theoretical operationalisation addresses that partially by the use of internal deltas as the internal valuation of norms in terms of pay-offs (with its potential to represent emotions, etc.), which takes an essential step towards the representation of the subjective perspective (Crawford and Ostrom, 1995, 2005). This is based on Ostrom's suggestion that the intrinsic valuation is inherently individual and hardly generalisable (Ostrom, 2005a), a view that is also supported by Searle (2005). However, though individualised within a group of concern, those valuations can vary for different cultures (Henrich, 2000), such as the culturally varying predisposition for maintaining reputation (see Hofstede et al. (2010)). Since deltas can bear higher-level generalisability components, we argue that the meaning of Crawford and Ostrom's quantified deltas, especially when paired with observable consequences, can, at least in part, be common knowledge transcending the individual, and – if attempting to provide a *general* 'Grammar of Institutions' – find generalised representation in the institution description, i.e. the institutional statement, itself.

Furthermore, the sanctions themselves[14] carry integral institutional information, since they can indicate a society-specific choice of sanctions. This offers grounds for refined institutional analysis since it not only facilitates the inference of social, economical or political meaning of an institution in the context of the society,[15] but can indicate the developmental state of a society.[16] Considering normative sanctions can thus increase the precision with which the grammar can characterise institutions in general, and norms in particular. Doing so, a refined 'Grammar of Institutions' can integrate internal and external viewpoints

[13]This conceived mapping is well reflected in conventions' conception as 'shared strategies'.

[14]Posner and Rasmusen (1999) differentiate six different sanction types found in norms: Automatic sanctions, Guilt, Shame, Informational sanctions, Bilateral costly sanctions, Multilateral costly sanctions.

[15]For example, material, informational, emotional sanctions in the Maghribī and Genoese societies had vastly different influence on individuals' behaviour, such as the significance of status (jāh) among Maghribīs in contrast to the material focus of the Genoese society (see Chapter 3).

[16]Recall from Section 4.3 that Durkheim's (1933) mechanical vs. organic solidarity permits the characterisation of societies' development based on the sanctions they employ.

by extracting generalisable institutional knowledge. Extending the generalisability primarily attributed to rules to the normative perspective facilitates a systematic and comprehensive "growing" of institutions from the bottom up, while facilitating their analysis across multiple social levels, such as individuals, groups, and the society at large – an aspect we address in Chapter 7. We thus expect the grammar's ability to represent normative sanctions directly, as opposed to transforming and attaching them in a form of rationalised pay-offs to the individual's instantiation of an institution representation. Even if the formation is driven by subconscious processes, on a social level the outcome can be observable, whether rational or not. The established institutions are thus stable, though not static, observable social artefacts with potential manifestations in the objective domain.[17]

Summarising, the GoI offers a sound structural starting point for a general institution representation. However, to operationalise it for a dynamic use in the context of agent-based systems (and normative multi-agent systems in particular), we require various refinements that improve its versatility, which we discuss in the following.[18]

5.3 Nested ADICO (nADICO)

In the attempt to extend the grammar from a discrete view on institution types towards a dynamic emergentist perspective,[19] we will introduce refinements that address the limitations of the original grammar and thus extend its generalisable use.

Any modification towards a more continuous representation arguably blurs the neat categories the ADICO grammar provides, but introduce refinements that offer the potential to capture institutions in greater detail, to help explain the transitions, and to take steps towards a better integration of external and internal perspective (the original ADICO grammar emphasises the external perspective on institutions). One step in this direction is taken by Schlüter and Theesfeld who differentiated the notions of norms expressed in the grammar.[20] To emphasise the transitions between different institution types, while providing a shift from objectified understanding to a shared subjective perspective on institutions (and norms in particular), we propose a refined variant of the institutional grammar, called *Nested ADICO*

[17] See Ritzer's integrated view on social levels of analysis and objectivity vs. subjectivity in Subsection 2.1.3.

[18] The concept of nADICO has been published in Frantz et al. (2013) and is presented here in an extended form.

[19] Note that a dynamic institution understanding is not to be confused with the historically laden understanding of *institutionalisation*, involving the admission to psychiatric wards, prisons (Foucault, 1977), etc.; institutions which Goffman refers to as "Total Institutions" (Goffman, 1961).

[20] The differentiation of norms into *personal* and *social norms* was already raised by Crawford and Ostrom (1995) themselves, but only systematically developed by Schlüter and Theesfeld (2010).

(nADICO).

The essential refinements include

- an *explicit syntactical representation of nested institutional statements*,
- the *explicit representation of consequences for normative statements* (using the 'Or else' component), and consequentially,
- a *refined differentiation between norms and rules*.

nADICO intends to offer a more comprehensive representation of institutions that manages the trade-off of capturing greater levels of detail and interdependencies between different institution types, while providing a standardised structural scaffolding accessible to agents. The syntax involves the specification of multiple nested levels of institution representation in order to allow agents to express not only *which* institutions exist, but furthermore, *what motivates* their emergence and stability, an aspect that can be reflected in associated consequences. Applying structural uniformity on all levels assures the required computational tractability and provides a blueprint for agents to store, shape, and communicate institutional understanding. Conceptually the notion of nesting in itself is not novel, since Crawford and Ostrom (1995, 2005) themselves considered the *Or else* to be an institutional statement, but it has not been systematically established. We further offer a notion of horizontal nesting in the form of statement combinations, an aspect not previously explored.[21]

5.3.1 Vertical Nesting

As vertical nesting we understand the substitution of the *Or else* component with an institutional statement of either institution type (convention, norm, rule). We call the leading institutional statement a *monitored statement* and the trailing one a *consequential statement*, since the latter describes, or more precisely, prescribes consequences for an actor's non-compliance with the monitored statement. A structured nesting approach bears the benefit of recursive nesting, enabling the representation of "monitoring the monitor". This is in equivalence to the institutional characteristics, and reflects the interdependency of institutions across different levels, a concept we refer to as *structural institutional regress*. The conventional interpretation of institutional regress (Aoki, 2001; Hodgson, 2002) as the temporal path-dependence of institutions (i.e. determining how far preceding institutional environments shape contemporary environments) ultimately relates back to the discussion of

[21] Jiang et al.'s Norm Nets (2013) allow a similar representation by combining different Norm Nets based on logical operators, but do not make this as part of the institutional grammar specification itself.

institutional dynamics in the context of the Structure-Agency debate (see Subsection 2.2.2). However, especially from a constitutional perspective – and in addition to the temporal institutional regress – structural institutional regress represents the governance *by* and *of* monitoring entities, an essential component of the concept of 'Rule of Law'[22], thus allowing a structural reflection of an institutional environment. Similar to temporal institutional regress, structural institutional regress can in principle be infinite. Like the 'chicken-egg problem' of the absence of precedence for the initial institution, i.e. non-existence of an institution-free state (see Subsection 2.2.2), we face the challenge of identifying a termination condition for structural institutional regress.[23]

In equivalence to the statement nesting levels, we can characterise different institutional roles. Let us apply and extend our previous example statement used in the context of the original grammar in Subsection 5.2.2 with a loose representation of ṣuḥba enforcement characteristics (see Subsection 3.2.1) for the purpose of clarification.[24]

Traders (**A**$_1$) must (**D**$_1$) trade fair (**I**$_1$) under any circumstances (**C**$_1$),

OR ELSE ——————————— ***2nd Level*** ———————————

observers (**A**$_2$) must (**D**$_2$) report the deviation (**I**$_2$) under any circumstances (**C**$_2$),

OR ELSE ——————————— ***3rd Level*** ———————————

fellow traders (**A**$_3$) may (**D**$_3$) consider the violating observer a cheater himself (**I**$_3$) in any case (**C**$_3$).

Decomposing the syntactical components, the previous statement thus has the structure `ADIC(ADIC(ADIC))`.

Grouping the overall statement into interlinked statement pairs (here: two pairs consisting of monitored and consequential statement), actors' roles can be defined with respect to the frame of reference. For example, the actor of the *first-order monitored statement* (in this case A_1 ("Traders")) is the potential *first-order violator*. The *first-order consequential statement*'s actor (A_2: "observers") is the *first-order sanctioner*, while at the same time being potential *second-order violator* (of a *second-order monitored statement*), monitored by a *second-order sanctioner* (A_3: "fellow traders") and so on.

This clear sequential nesting structure reflects the interdependency of different institutional levels, but it is hardly able to capture the complexity observable on specific levels, since actions can co-occur and combinations can have different institutional implications.

[22]See Belton (2005) for a discussion of varying respective definitions. In the context here we allude to the accountability of the monitoring party (such as appointed enforcers or governmental officials).

[23]Possible answers could relate to the representation of 'trust in the system', with a terminating consequence in the form of a descriptive norm à la "... , or else people stop trusting the system of governance".

[24]In this and all following examples we use indices to identify a statement component's nesting level (e.g. A_2 indicating an *Attributes* component on the second level).

Using the running example, sanctions can involve multiple consequences at the same time. Second-order sanctioners can thus be obliged to label second-order violators as cheaters *and* denouncing them at the same time.[25]

5.3.2 Horizontal Nesting

In addition to the vertical nesting that substitutes the *Or else* component with an institutional statement, we thus introduce a notion of statement combinations that allows us to represent institutional complexity on a given vertical nesting level.

For this purpose we draw on elementary Boolean operators expressing logical conjunction 'and' (Operator: `and`), inclusive disjunction 'or' (Operator: `or`), exclusive disjunction 'either or' (Operator: `xor`) as well as negation 'not' (Operator: `not`).

Applying the operators to our trader example, we can express not only concurrent actions/consequences but also notions of gradual sanctioning:[26]

> Traders ($\mathbf{A}_1$) must ($\mathbf{D}_1$) trade fair ($\mathbf{I}_1$) under any circumstances ($\mathbf{C}_1$),
>
> **OR ELSE** ———————— ***2nd Level*** ————————
>
> observers ($\mathbf{A}_{2a/b/c}$) must ($\mathbf{D}_{2a}$) report violators ($\mathbf{I}_{2a}$) under any circumstances ($\mathbf{C}_{2a}$)
>
> **and**,
>
> depending on severity ($\mathbf{C}_{2b/c}$), must ($\mathbf{D}_{2b/c}$)
>
> **either** fire them ($\mathbf{I}_{2b}$)
>
> **or** retaliate against their family ($\mathbf{I}_{2c}$),
>
> **OR ELSE** ———————— ***3rd Level*** ————————
>
> fellow traders ($\mathbf{A}_3$) may ($\mathbf{D}_3$) consider the violating observer a cheater himself ($\mathbf{I}_3$) in any case ($\mathbf{C}_3$).

The syntax of this statement is thus `ADIC((ADIC and (ADIC xor ADIC))ADIC)`.

In addition to the explicit specification offered by the operators `and` and `xor`, applying the inclusive disjunction `or` allows us to express cases in which actions or sanctions may or may not co-occur.[27] In our example, we can thus describe *potentially co-occurring* sanctions:[28]

[25]A practical example from the domain of traffic violations is the concurrent imposition of fines *and* demerit points.

[26]In this example we extend the numeric level index with characters that differentiate individual statement associations on a given nesting level. Note that different statements can share components.

[27]The negation operator allows the inversion of statements, such as the prescription of inaction, an aspect that completes the comprehensive nature of the nADICO feature set but is secondary compared to the combinatorial characteristics of the remaining operators. We will thus explore it implicitly in the context of further examples.

[28]Recall that the use of the *Conditions* component is optional, and if not specified, does not constrain the applicability of the statement, defaulting to "at all times and in all places" (Crawford and Ostrom, 1995).

Traders ($\mathbf{A}_1$) must ($\mathbf{D}_1$) trade fair ($\mathbf{I}_1$) under any circumstances ($\mathbf{C}_1$), **OR ELSE** ——— ***2nd Level*** ——— fellow traders ($\mathbf{A}_{2a/b}$) may ($\mathbf{D}_{2a}$) denounce them ($\mathbf{I}_{2a}$), **or** may ($\mathbf{D}_{2b}$) exclude them from future trade interactions ($\mathbf{I}_{2b}$), **or** their trade partners ($\mathbf{A}_{2c}$) may ($\mathbf{D}_{2c}$) reclaim their goods ($\mathbf{I}_{2c}$).

The corresponding nADICO expression is thus `ADIC(ADIC` *`or`* `ADIC` *`or`* `ADIC)`.

Summarising this overview, to this stage we have described the different nesting principles in a narrative fashion. However, we can clarify the different principles schematically. The use of logical operators affords comprehensive institutional representation on a given level (horizontal nesting). Its purpose centres around the *representation of detail*. In contrast, the use of *Or else* signifies the transgression of institutional levels, offering the semantic capabilities to describe *interdependencies of institutions across different levels* (vertical nesting). The exemplary statement `ADIC((ADIC(ADIC` *`or`* `ADIC)` *`and`* `ADIC(ADIC` *`xor`* `ADIC)))` can thus be visually represented as shown in Figure 5.2.

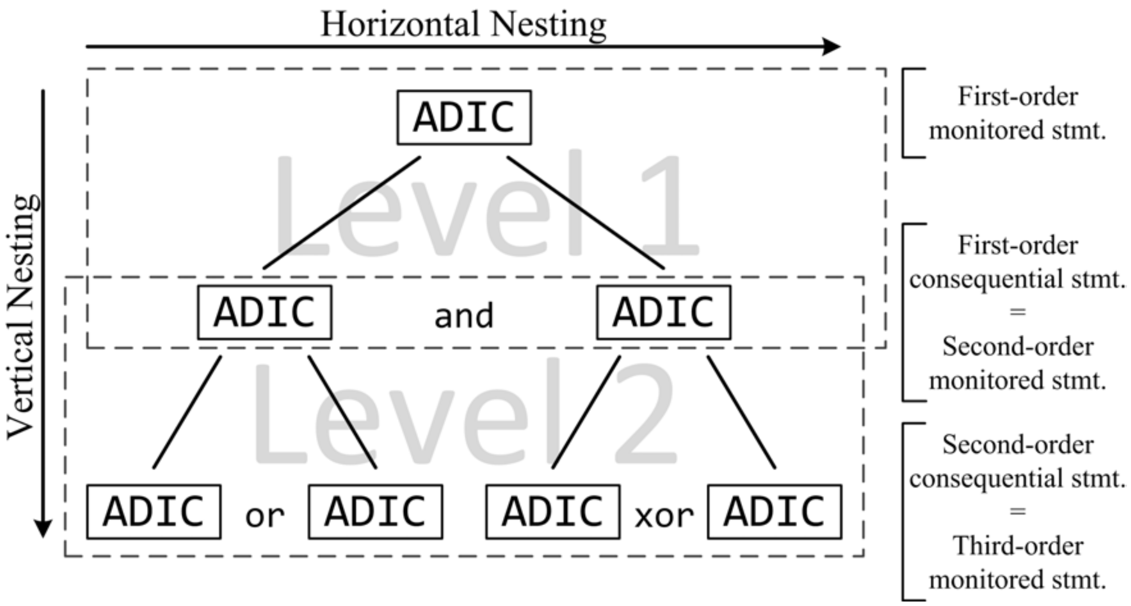


Figure adapted from Frantz et al. (2014b)

Figure 5.2: Nesting Characteristics of nADICO

Although our exemplary elaboration on horizontal nesting has concentrated on the consequential statement, the combination of institutional statements can likewise occur within first-order monitored statements:

- `(ADIC` *`and`* `ADIC)(ADIC)` requires the satisfaction of the actions (or outcomes) specified in *both* statements to prevent the activation of the shared consequence;

- `(ADIC` *`or`* `ADIC)(ADIC)` requires the satisfaction of *one or both* statements to prevent a given consequence;
- `(ADIC` *`xor`* `ADIC)(ADIC)` exclusively allows the satisfaction of *either* single monitored statement.

As mentioned, statements on all levels can combine multiple operators, such as `(ADIC` *`and`* `(ADIC` *`xor`* `ADIC))(ADIC)`. Additionally, individual monitored statements can have their own consequential statements, beyond the compound consequence they share when combined with further statements, such as for the statement `(ADIC(ADIC)` *`and`* `ADIC(ADIC` *`or`* `ADIC))(ADIC)`.

Deriving an example expression from our scenario, we can express the ṣuḥba obligation to store goods and the alternative choice of trading or not trading as follows. The non-compliance of individual sub-statements can have specific sanctions (rejection of future goods acceptance if storing is rejected; announcement as cheater if not trading fair), but the individual is bound to comply with the elementary statements, or otherwise expelled from the trader coalition.[29]

Traders ($\mathbf{A}_{1a/b/c/d}$) must ($\mathbf{D}_{1a}$) store received goods ($\mathbf{I}_{1a}$) under any circumstances ($\mathbf{C}_{1a}$),

OR ELSE ——— ***2nd Level*** ———

fellow traders ($\mathbf{A}_2$) may ($\mathbf{D}_2$) **not** accept their goods in the future ($\mathbf{I}_2$),

---

and

either may ($\mathbf{D}_{1b}$) **not** trade those goods ($\mathbf{I}_{1b}$)

or must ($\mathbf{D}_{1c/d}$) trade those goods ($\mathbf{I}_{1c}$)

and return the realised profit ($\mathbf{I}_{1d}$),

OR ELSE ——— ***2nd Level*** ———

the goods sender ($\mathbf{A}_2$) must ($\mathbf{D}_2$) announce them as cheater ($\mathbf{I}_2$),

---

OR ELSE ——— ***2nd Level*** ———

all fellow traders ($\mathbf{A}_2$) must ($\mathbf{D}_2$) expel them from the coalition ($\mathbf{I}_2$).

The structure of this statement can be abstracted as

`(ADIC(`*`not`*`(ADIC))` *`and`* `(`*`not`*`(ADIC)` *`xor`* `(ADIC` *`and`* `ADIC)(ADIC)))(ADIC)`.

[29]Consequential statements for individual horizontally nested monitored statements are represented by indented *OR ELSE* blocks.

5.3.3 Syntax Specification and Potential Operational Refinements

To complete our discussion of syntax beyond nesting principles, we need to consider the implications for the simplest form of institution statements, `AIC` statements – 'shared strategies', that is *conventions* or descriptive norms. Apart from their unnested use, alternatively to the use of sanction-less norm statements `AIC` statements can likewise be used to terminate nested institutional statements, describing observable behaviour as opposed to prescribing it. This can be purposeful since – as far as the specification goes – the terminating statement is not explicitly sanctioned/enforced, potentially giving an explicit prescription (such as obligations or prohibitions) limited purpose. While conventions cannot have a vertically nested consequence, they can be combined (horizontally nested) with other institution types. A special case is their combination with other conventions in order to express co-occurring behaviour without implying prescriptiveness (e.g. `AIC` ***and*** `AIC`).

Figure 5.3 shows the complete nADICO syntax specification in the Extended Backus–Naur Form (ISO, 1996).

```
attributes   =    "A" ;
deontic      =    "D" ;
aim          =    "I" ;
conditions   =    "C" ;
aic          =    attributes , aim , conditions ;
adic         =    attributes , deontic , aim , conditions ;
and          =    "and" ;      (* Conjunction *)
or           =    "or" ;       (* Inclusive Disjunction *)
xor          =    "xor" ;      (* Exclusive Disjunction *)
not          =    "not" ;      (* Negation *)
ws           =    " " ;        (* Whitespace *)
LB           =    "(" ;
RB           =    ")" ;
nadico       =    adic | (* Individual Norm Statement without Sanction *)
                  [ not , LB , ( nadico | aic ) , RB ] | (* Negation *)
                  [ nadico [ , LB , ( nadico | aic ) , RB ] ] | (* Vertical Nesting *)
                  [ LB , ( nadico | aic ) , ws , ( and | or | xor ) ,
                                 ws , ( nadico | aic ) , RB ] ; (* Horizontal Nesting *)
statement    =    aic | nadico ; (* nADICO Statement (including Top-Level Conventions) *)
```

Figure 5.3: nADICO Grammar in EBNF

The representation provided here is abstract and allows scenario- or domain-specific refinement of grammar components. This can involve the decomposition of the *Attributes* component into sets of *individual* and *group/social markers*. The *AIm* can be refined with respect to specific action properties, such as *action object* and *action target*. The *Conditions* component effectively reflects aspects of activation, i.e. the conditions under which a rule is activated (other than its activation as a consequential statement for a violated monitored statement). This may not only consider place and time but also include the fulfilment of preceding institutional statements (or a combination of those).

A special case in this context is the *Deontic* component, whose rigid representation based on deontic primitives can likewise be substituted. We will turn to this aspect in the following Chapter 6 by introducing the notion of *Dynamic Deontics* in an attempt to facilitate the representation of a continuous norm understanding beyond rigid *must*s, *must not*s and *may*s.

Figure 5.4 provides exemplary aspects for potential component refinements.

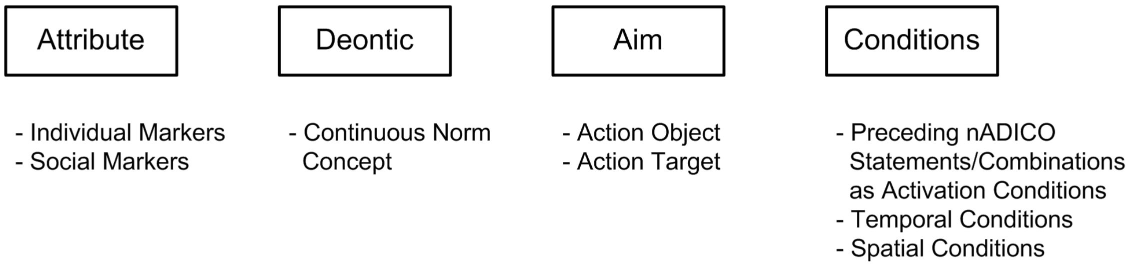


Figure 5.4: Potential Refinements for nADICO Operationalisation

5.3.4 Refining the Differentiation between Norms and Rules

Since nADICO introduces sanctions into the norm representation, it sacrifices the ability to syntactically differentiate between norms and rules. Consequently, we rely on alternative means to discriminate norms from rules in nADICO. In the following we discuss these differentiation criteria and additional refinements we introduce to substitute the syntactic differentiation in the original grammar. Here, the essential qualifications of rules laid out by Crawford and Ostrom (1995, 2005) provide a starting point for such refined differentiation.

Collective Action One essential meta-characteristic is the collective action process that constitutes a given rule. Individuals thus follow a decision process, which, depending on governance style, provides mechanisms to formalise and activate rules within the society of concern. A possible approach is majority-based voting or consensus-based approaches.[30] Whatever the nature of the process, we maintain the requirement of an explicit collective action process to constitute rules.

Nature of Monitor A second aspect is Crawford and Ostrom's requirement of monitors for rules. However, in line with our position on sanctions as part of the norm representation, norms are similarly characterised by the existence of a monitor (though not necessary explicit or appointed as in the case of rules). As explored earlier, the understanding put forth for this

[30]The field of computational social choice has explored a wide range of related approaches. An introductory overview is provided by Chevaleyre et al. (2007).

work is that norms, in the light of diverging means (strategies) of achieving a common shared aim (see Subsection 2.2.2), rely on enforcement by a monitor. However, the characterisation of decentralised norm enforcement, in contrast to centralised rule enforcement, allows us to utilise the quality or *nature of the monitor* as a differentiation criterion.

Here Schlüter and Theesfeld's (2010) characterisation of different monitor types is helpful, since it allows us to associate different monitor with institution types. An internal monitor can be associated with personal norms,[31] while informally assigned monitors (such as the totality or a self-appointed subset of all institution participants) can be associated with social norms. On the other side of the divide, formally appointed monitors, such as an organisation's or society's leader(s), are indicative for the rule nature of an institution. A comprehensive overview of these associations is provided in Table 5.1.

Table 5.1: Monitor Types

Monitor Type	**Institution Type**
Internal Monitor	Personal Norm
Social Monitoring Informally Assigned Monitor(s)	Social Norm
Formally Assigned by Private Entities Formally Assigned by Legislative Body	Rule

This table is an extension of the original overview provided by Schlüter and Theesfeld (2010) and adapted from Frantz et al. (2013).

Differentiation of Monitor and Sanctioner We can further the refinement by focusing on the characteristics of rules in societies that are governed by formal institutional environments. Recalling the Durkheimian differentiation between mechanical and organic solidarity (Durkheim, 1933) (discussed in Section 4.3), the functional interdependency of specialised individuals, along with diverging world views with limited common moral ground, promotes a shift of judicial perspectives. It affords a progression from an assumption of absolute authority and penal law (associated with repressive sanctioning) towards a legal system that operates based on civil law systems (often based on secular principles) and emphasises impartial restitutive sanctioning with focus on fair redistribution (Merton, 1994; Barry, 1995).[32] From a justice perspective, fair and impartial enforcement is associated with a clear specification of laws and – particularly for the democratic case – separation of pow-

[31] Schlüter and Theesfeld's essential contribution is the differentiation between personal and social norms based on the nature of Crawford and Ostrom's deltas, an aspect that is secondary in this discussion.

[32] See Barry (1989) for an overview on different positions on impartiality and fairness.

ers (Montesquieu, 1949), which in practice is reflected in an *explicit specification of monitor and sanctioner* as well as *differentiated appointment of both*. In our view, the more systematic differentiation between monitor and sanctioner beyond the integrated conceptualisation as monitor in the original grammar affords a more truthful representation of institutional reality. Even if sanctioner and monitor are embodied in unity, a refined grammar would require their clear specification as a central criterion for the constitution of the institution type 'rule'.

Nature and Activation of Sanction An additional aspect beyond the nature of monitor and enforcer is the discussion of the sanction itself. An aspect that potentially led Crawford and Ostrom to refrain from the inclusion of sanctions as part of the norm specification is the potentially diverse nature of sanctions and the uncertainty about their activation. Sanctions associated with a given norm violation may not be openly communicated and known outside the society of concern, leading to an objective observation of the normative behaviour without making the observation of sanctions accessible to outsiders. On the other hand, however, it is precisely this uncertainty about the invocation, their nature and multitude that can make norms effective. Assuming the example of jay-walking, an explicit specification makes both sanctions and sanctioner public knowledge (Posner and Rasmusen, 1999; Posner, 2000): assuming its formal prohibition, an individual knows what to expect when jaywalking in presence of traffic police. Assuming the absence of such appointed enforcer, in contrast, the behaviour of fellow pedestrians witnessing a jaywalker bears uncertainty (*Do* bystanders react? (Invocation of sanctions); *How* will they react? (Nature of sanctions); *How many* bystanders will react? (Multitude of sanctions)). Thus, in contrast to rule-based enforcement, we argue, an essential characteristic of norms is the limited predictability and thus uncertainty about enforcers and sanctions (Posner and Rasmusen, 1999). As an additional criterion beyond the clear specification of monitor and sanctioner, we can differentiate rules based on the nature of the well-defined[33] sanctions in contrast to the near arbitrary choice of sanctions for norm violations.[34]

Summarising, we see the explicit vs. fuzzy nature of specification of a) *sanctioner* – ideally separated into monitor and sanctioner – and b) *consequences* as essential differentiation criteria for norms and rules, beyond the requirement of collective action for the formation of rules. As such, the refined differentiation bears a stronger sociological grounding as opposed

[33]Note that even for gradual sanctioning in the context of rule enforcement, the extent of discretion on the part of the enforcer can, at least to an extent, be predicted since the sanctions are public knowledge (e.g. demerit points for speeding).

[34]Note that in the case of normative enforcement sanctioning individuals defend their stake in the violated norm (if they are self-appointed enforcers), as opposed to maintaining an impartial perspective, with the former aspect being related to Durkheim's justice understanding in the context of mechanical solidarity.

to relying on syntactic differentiation.

However, even though the differentiation may not be overtly obvious, the logical operators introduced in the context of horizontal nesting can be indicative for the existence of norms or rules. Without drawing a crisp boundary (e.g. for cases where normative sanctions consistently co-occur), if prescribing combinations of sanctions, the conjunction (*and*) as well as exclusive disjunction (*xor*) indicate rules; however, the fuzzy nature of selective combination of sanctions based on the inclusive disjunction (*or*) reflects the uncertainty of (co-)occurrence of sanctions, thus being indicative for norms.

Concluding this discussion, the differentiated characteristics of norms and rules are summarised in Table 5.2.

Table 5.2: Differentiation of Norms and Rules in nADICO

Characteristic	**Norms**	**Rules**
Specification of Monitor	unspecified / fuzzy	clear specification
Specification of Sanctioner	unspecified / fuzzy	clear specification
Assignment of Monitor and Sanctioner	informal	formal
Relationship between Monitor and Sanctioner	often unified entity, not explicitly specified	unified or separated, clear specification
Sanctions	fuzzy knowledge about invocation, nature and multitude	explicit specification
Nature of Monitor	see *Monitor Types* specified in Table 5.1	
Combination Operators	*or*	*and*, *xor*

This table has been adapted from Frantz et al. (2014b).

5.4 Discussion

In this chapter we have developed a generic structural representation of institutions, selected requirements of which we laid out in the beginning of this chapter. Central requirements included its generic nature, the ability to represent different institution types and its role of interfacing between agents (that use the structure to express normative understanding) and researchers (that interpret it). We further required a flexible representation that is geared for runtime instantiation as well as fluent transition between different institution types.

Since our work is related to the area of normative multi-agent systems, we provided an overview over norm representation types adopted in this field, and observed a general tendency towards the use of rule-based representations (see Subsection 5.1.2). However, as pointed out by Hollander and Wu (2011b) in their survey, rule-based representations find primary use in off-line specification of norms, a regularity from which we could only find few exceptions. Interactionist or emergentist approaches dominantly rely on what we dubbed 'Sequence' representations in Subsection 5.1.2.

A suitable scaffolding for norm representation has been laid out by da Silva Figueiredo et al. (2011) in their attempt to extract the essential features of norm representation languages in the context of introducing NormML. Comparing those features to Crawford and Ostrom's institutional grammar specification – which emerged out of the area of institutional analysis as opposed to NorMAS – we can find a strong conceptual alignment as shown in Table 5.3. Noteworthy exceptions for matching equivalents are the activation constraints, which in Crawford and Ostrom's grammar largely rely on meta-processes of institution formation, such as collective action in the case of rules, or are encoded in the grammar's *Conditions* component. However, for our case, the use of a representation that fosters bottom-up institution formation, activation aspects should surface during the formation process itself. A more relevant deviation is the original grammar's lacking specification of sanctions for norms.

Table 5.3: Properties of Norm Representations (da Silva Figueiredo et al. vs. Crawford and Ostrom)

da Silva Figueiredo et al. (2011)	**Crawford and Ostrom (1995)**
Deontic Concept	Deontic
Involved Entities	Attributes
Actions	Aim
Activation Constraints	Conditions*
Sanctions	Or else**
Context	Conditions

* The Conditions components carries spatial, temporal and procedural aspects of norms. For rules collective action processes determine activation.
** Crawford and Ostrom (1995) assign sanctions only for rules; not for norms.

However, the built-in quality of representing a wide range of institution types makes Crawford and Ostrom's grammar a suitable starting point for further refinement in the form of nADICO as introduced in Section 5.3.

Essential refinements include the specification of nesting of institutional statements in order to represent *structural institutional regress*, i.e. the representation of higher-order enforcement based on recursively backed institutions. In addition to this notion of vertical nesting, we introduce statement combinations on given nesting levels (horizontal nesting), the combination of which can capture complex institutions in great detail.

A second essential aspect is the introduction of sanctions for norm statements, which affords a refined differentiation between norms and rules, since the original grammar relies on their syntactic disambiguation. By introducing sanctions, we widen the grammar's scope, introducing a subjective internal perspective – in contrast to the external perspective in the original grammar. This offers a more accurate representation of *why* a particular norm emerged (i.e. what drives its emergence), as opposed to only observing the behavioural norm itself. Along with this focus, we offer a more comprehensive representation of institutions by separating norms and rules based on the specification (and thus nature) of sanctioner and sanctions, beyond the collective action requirement for rules set out in the original grammar. The logical operators introduced as part of the horizontal nesting capability provide additional syntactical support for the differentiation. This approach arguably gives up the neat differentiation of the original grammar, but replaces it with a realistic and comprehensive representation of institutions based on sociological instead of syntactic grounding.

Our nesting approach does not exist in solitude. As described in Subsection 5.2.4, Jiang et al. (2013) have formalised a nesting of normative statements by introducing the notion of Norm Nets that combine norms within the scope of a given context. Their operationalisation does not differentiate between the various institution types but concentrates on the representation of what we would interpret as rules; the authors use the term 'norm' in the legalistic sense. Using CPN, their approach can verify the execution of codified Norms Nets.

Earlier work by López y López and Luck (2004); López y López et al. (2003) proposed an interlocking norms structure, in which so-called *primary norms* are tied to the context of *secondary norms*, with violation of normative goals of the secondary norm triggering the interlocked primary norm. In their concept, the multi-level interdependency of norms is implicit. Their approach concentrates on the interlocking of context, which in nADICO syntax would reflect an interlocking on the *Conditions* component as a means of activation, an aspect that is reflected in potential operational refinements of nADICO (see Subsection 5.3.3).

Work by Grossi et al. (2006) emphasises the multi-level nature of institutional enforcement in the context of electronic institutions. Doing so, they decompose *substantive norms* (norms that describe the desired behaviours governed by the institution) into elementary *sub-institutions*, namely *check norms*, i.e. norms that monitor behaviour, and *reaction norms*, i.e. norms that describe reactions to observed norm violations. This differentiation is analogous to our understanding of explicit specification of a monitor (which is the actor in Grossi et al.'s check norm) and a sanctioner (which is the actor in their reaction norm). However, since Grossi et al.'s concept is applied to the domain of electronic institutions, their norm terminology reflects the legalistic institution understanding, the equivalent of what we understand as rules in the context of nADICO. Specifying norms and rules comprehensively, Grossi et al. (2006) adopt a global (off-line) perspective on institutions, in contrast to nADICO, where the emerging institutions (conventions and norms in nADICO terminology) may not provide the structural clarity from their onset but crystallise when progressing towards more explicit institution types.

At this stage, we have provided an institution representation that satisfies the structural requirements (labelled as *General Nature of Representation*) specified at the onset of this chapter in Section 5.1 (domain-independence, institution-type-independence, interpretation by humans and agents). However, the second set of requirements that address the dynamic nature of the representation (under the label *Flexibility of Representation*), including aspects such as runtime instantiation and runtime transition between different institution types, has yet to be addressed. These aspects will be of central concern in the upcoming chapter.

6

Modelling Institutional Dynamics

Working towards an operationalisation of mechanisms to 'grow' institutions, we shift our focus from a mere structural representation to aspects related to the formation of normative understanding, raising the question of how individual social learners independently develop a shared understanding of behavioural norms.

In the context of the institutional representation developed in the previous chapter, a central role is associated with the institutional grammar's *Deontic* component. In alignment with large parts of the NorMAS work (Boella et al., 2007), it is operationalised using the conventional tripartite differentiation of deontic logic (Wieringa and Meyer, 1993) into primitives that represent obligations (*must*s), prohibitions (*must not*s), and permissions (*may*s).

6.1 Towards Dynamic Normative Understanding

The conceptual soundness of conventional deontic logic (von Wright, 1951a; Wieringa and Meyer, 1993), and the interdefinability in particular, makes its use compelling for stabilised obligation and prohibition norms. However, if we step back and assume a long-term perspective on norms, such as in the context of institutional analysis (as discussed in Subsection 2.2.2), we need to acknowledge that the process of institution formation has a dynamic component. Suggesting that institutions change with a 'flick' from a permissive *may* to a prescriptive *must*, as done in conventional deontic logic, accommodates the formal domain of institutional understanding,[1] which is well reflected in the meta-rules that govern rule changes in the original institutional grammar (see Section 5.2 as well as the comprehensive discussion by Ostrom (2005b)). However, in the informal domain,[2] we find a continuous moulding process in which seemingly unconstrained conventionally exercised behaviour can gain a prescriptive character, creating institutional facts that constrain future behaviour.[3] The reflection on the traditional static deontic concept becomes important in the light of a) an increasing understanding of norms as a means to induce socially desirable behaviour (see e.g. Cialdini et al. (1990); Cialdini and Trost (1998)), and b) the recognition that the conventional "boolean norms" (as Ghose and Savarimuthu (2013) name those) are insufficient to capture more complex behaviour.[4] The demand for a more dynamic perspective on normative understanding is further shared by Schlüter and Theesfeld (2010), but more importantly, by Ostrom herself (Kinzig et al., 2013). They recognise fast- and slow-changing norms (which Kinzig et al. (2013) characterise as *fluid* and *viscous* norms),[5] in addition to norms that persist despite outliving their purpose and continue to maintain influence on the institutional environment, the society, and thus avoidably onto culture.

To support our point, let us finish the discussion based on a set of examples that highlight the continuous shift of normative understanding.

Smoking, a behaviour glorified by media and associated with a 'right of passage' well into the 1990s has been incrementally outlawed in Western countries both in media (prohibition of advertising, smoking ban in movies) and public perception, to an extent that a vast amount of public spaces are now considered smoke-free.[6] This paints a picture of a continuous shift from a (at least in the display of media) somewhat desirable behaviour towards increasing prohibition and removal of smokers from the public eye (Gutman, 2011).

A movement that experienced a continuous change into the opposite direction is the normative view on homosexuality. While accepted in ancient Greece, and largely outlawed during the past centuries (with Alan Turing as a noteworthy victim of legal prohibition),

[1]Refer to the different positions on institution change discussed in the beginning of this work (see Subsection 2.2.2).

[2]For our position in institution formation and change refer to Subsection 2.2.2.

[3]This is in line with Bourdieu's understanding of "urgency of practice" (Bourdieu, 1984) – whether conscious or not – that shapes the institutional environment and "frees us from the misplaced belief in illusory freedoms" (Bourdieu, 1990a) (see Subsection 2.1.1).

[4]Ghose and Savarimuthu (2013) introduce the concept of *Optimization Norms* to describe expressions such as "doing one's best", i.e. cases in which the fulfilment cannot be determined based on Boolean truth values.

[5]Various positions on institutional change have been laid out in Subsection 2.2.2.

[6]Graham (1996)'s analysis supports that trend, with smoking levels among the male Northern Europeans peaking at levels of 70-90% in 1950 before declining to levels of 30-40% until 1990.

from the 1960s and 1970s onwards minorities acknowledged their sexual orientation and moved their discrimination into the public eye. This can be interpreted as a shift from a barely legal permission towards increasing tolerance and acceptance by the wider society. Furthering this direction, an increasing number of countries are permitting legal admission to marriage (Stanford Encyclopedia of Philosophy, 2011).

Another example is the traditionally unregulated notion of child spanking, which had been considered a permissible means of education and nurturing. However, norms surrounding physical sanctioning of minors have shifted towards the domain of prohibition and transcended from a normative rejection to an increasing legal criminalisation in Western countries (International Encyclopedia of the Social Sciences, 2008; Global Initiative to End All Corporal Punishment of Children, 2015).

With these examples in mind, we can progress towards the construction of a mechanism that reflects the dynamic nature of institutions using a concept which we refer to as *Dynamic Deontics*.[7]

6.2 Dynamic Deontics

Given the interpretation of norms as implicitly shared representations, they are subject to subjective perception and evaluation by norm participants. Those representations, even when modelling the same behaviour, will unavoidably vary across individuals (Searle, 2005) and cultures (Henrich, 2000). This motivates a more nuanced interpretation that not only reflects situational motivators or deterrants for complying with a norm – represented as deltas in the Grammar of Institutions –, but shows the subjective importance an individual associates with a given institution. We can observe this latter aspect in the use of language (e.g. 'You *must* be home before it is getting dark' vs. 'You *should* be home before it is getting dark').

Attempting to represent dynamically changing normative understanding, while assuring a reproduction of essential institutional features, our dynamic norms concept has the following properties:

- A continuous concept of deontic primitives,
- an individualised dynamic normative understanding, and
- stability of normative understanding.

[7]This work has previously been published in Frantz et al. (2014b). In the following we present it in an extended version.

6.2.1 Continuous Notion of Deontics

A special case that highlights the flipside of the rigid conventional deontic primitives is the permissive primitive *may*. In contrast to the unambiguous injunctions *must* and *must not*, its value to describe behavioural conventions is very limited. Apart from constituting *the right* to take an action or constituting that action itself (Crawford and Ostrom, 2005; Searle, 1969), it is imprecise about associated duties and thus bears limited predictive value for individuals' actions, thus relying on individuals' situational evaluation.[8]

As a consequence, we refine the 'intermediate' permissive deontic (here operationalised as *may*[9]) and allocate it on a continuous scale (an aspect von Wright (1951a) had considered as a possibility) delimited by the 'polar' deontics, i.e. the prescriptive deontic (represented as *must*) and its proscriptive counterpart (*must not*). Figure 6.1 visualises this principle.

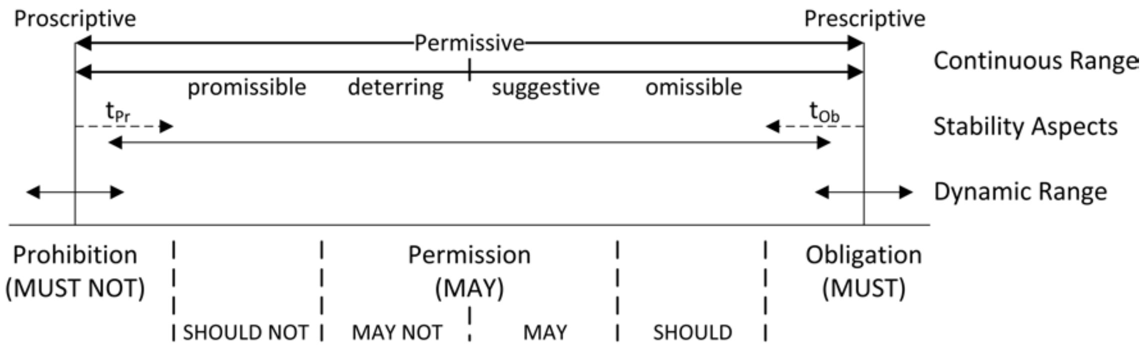


Figure adapted from Frantz et al. (2014e)

Figure 6.1: Dynamic Deontics

Starting from a centred indifferent permissive view, we can characterise the range towards the prescriptive extreme by its increasingly suggestive nature, and in the complementary direction, the range towards the proscriptive extreme by its deterring nature. The normative understanding of a given action can thus be situationally evaluated. Making the arising normative understanding accessible to the modeller, we can compartmentalise the respective ranges between midpoint and extremes and allocate descriptive labels. We exemplify this by assigning labels to characterise compartments with respect to their prescriptiveness. The permissive deontic centre could be bordered by *suggestive* (*may*) and *deterring* (*may not*)[10]

[8]The semantic incongruence between the permissive deontic that describes *rights* and its invariants that describe *duties* has been noted by Crawford and Ostrom (2005).

[9]The operational terms associated with obligation, prohibition and permission are interchangeable with alternative terms, such as 'forbidden' or 'obliged', etc. However, for the sake of unambiguous terminology in the context of this work, we will use the terms *must*, *must not*, and *may* to signify obligation, prohibition, and permission.

[10]Note here that the label *may not* is used to reflect the continuity of the deontic range (preceding the stronger

compartments, reflecting what Sumner (1906) described as *folkways* that exhibit a potentially unconscious either weakly or non-sanctioned "societal force". Deontic compartments that impose stronger behavioural compliance could describe the *omissible* (*should*) and *promissible* (*should not*).[11] Only at the extremes would we find the conventional prescriptive and proscriptive injunctions. In this operationalisation the choice of number and nature of compartments as well as deontic terms (or simply *deontics*) is systematic but arbitrary. Compartments could be of different sizes or be asymmetrically allocated along the deontic range (such as narrowing compartments of greater prescriptive weight). We touch upon the implications of this aspect in the context of the operationalisation (Subsection 6.2.5). Similarly, the chosen terms exemplify potential labelling with increasing prescriptiveness. Those aspects are subject to future research. Under the assumption of symmetry of range and compartments, this conceptualisation offers a starting point for the exploration of the principal dynamics of individualised normative understanding.

6.2.2 Dynamic Deontic Range

The individualised normative understanding is not only reflected in the continuous allocation of norms on the deontic range, but also in the range itself. We posit that an individual's normative frame of reference is similarly shaped based on the individual's personal experience, such as its own history, social ties (e.g. family, peers), moral dictates (e.g. by religion or culture), etc. During the individual's developmental process and throughout life, the deontic range will vary. During the early phases of development, a person's deontic range will be narrow and unbiased, providing limited grounds to contextualise novel experience, leading to experiences lodged around deontic extremes, such as moral views based on preimposed cultural beliefs. Only with greater exposure to other influence factors (different social groups, cultures, etc.) will the deontic range evolve, i.e. widening the permissive range between the extremes of obligation and prohibition, and thus reflect more nuanced perspective on "oughtness" (Turner, 1991). An example is the development of differentiated views on sexual orientation that allow the individual to contextualise or revise religiously imposed moral views. However, this dynamic deontic range does not expand infinitely, but in the course

should not) and the terminological symmetry of the compartments (*may* vs. *may not*). The term as used here is to be understood as a mild disinclination from permission ("You may not cross the road when facing red traffic lights."), as opposed to expressing that some environmental state is unlikely to occur ("The celebration may not happen, because it is supposed to rain.").

[11] As *omissible* we understand an obligation from which we can deviate in exceptional cases. Complementary, the *promissible* characterises a prohibition that can be exceptionally foregone. Our invocation of the term "omissible" deviates from traditional scheme of normative status, in which it signifies a non-obligation (McNamara, 1996).

of an individual's lifetime can likewise contract based on subsiding external stimuli (such as reduced interaction with the outside world) or rigidly adopted viewpoints. The dynamic deontic range thus offers the experiential backdrop for the formation of normative understanding, reflecting how external stimuli shape the individual's world view and habitus over time (see Subsection 2.1.1) – represented as the dynamic deontic range. Based on the deontic range the individual contextualises and gives meaning to novel information (i.e. allocating it on the deontic range). Similar to the issue of compartmentalisation of the deontic range, the dynamic nature of the deontic range brings up challenges for further exploration. How far does an agent look back to remember his extreme experiences? Does it reflect its entire history (potentially applying a recency bias), or only operates across a sliding window of past experiences, or is it based only on situational experience? We extend the discussion on those challenges in the context of the operationalisation in Subsection 6.2.5.

6.2.3 Stability

Reflecting the manifestation of social behaviour over time, individualised normative understanding does not continuously oscillate but entails inertia properties that resist spontaneous rapid changes (see Subsection 2.2.1), and underlie internalisation processes that are activated by sustained reinforcement.[12] Stability is an essential characteristic of institutions (Ayres, 1944; Aoki, 2001; Scott, 2008), since it facilitates their purpose, namely reducing uncertainty (North, 1990) based on the ability to predict others' behaviour, and increasing efficiency (Williamson, 1998) by reducing the cognitive load of deliberation in the light of limited cognitive abilities (Young, 1998; Aoki, 2001) (see Subsection 2.2.1).

Utilising our example of sexual orientation in the context of Britain, we can retrace the slow-moving nature of institutional change: In the UK the influential Wolfenden report (Committee on Homosexual Offences and Prostitution, 1957) challenged the view on homosexuality as a disease and sparked a public debate and change in perception of homosexuality. However, it took further ten years until the perceptual shift in the informal domain was reflected in the decriminalisation in the Sexual Offences Act 1967 (Act of Parliament, United Kingdom, 1967), with further 'normalisation' only established at the end of the last century.[13]

Reflecting the notion of stability and inspired by the concept of hysteresis (Ewing, 1889), i.e. the delayed effect of past inputs on current outputs (e.g. an enacted force on a subject),

[12] Recall the process of reconstitutive downward causation discussed in Subsection 2.2.2.

[13] The Sexual Offences (Amendment) Act 2000 unified the age of consent to 16 for both homosexual and heterosexual behaviour. Previously acts incrementally reduced the age of consent for homosexual behaviour (1967: to 21 years; 1994: to 18 years).

we propose that normative understanding becomes increasingly change-resistant when progressing towards either polar extreme (prohibition or obligation). Consequently, the extremal deontics exhibit the strongest change resistance, which, once stubbornly entrenched, require extended countering reinforcement to return into the more dynamic permissive range. Concentrating on the 'stickiness' of the deontic extremes, we introduce tolerance regions around the deontic extremes – denoted as t_{Pr} and t_{Ob} in Figure 6.1 –, which we use to measure the prolonged penetration of the deontic range boundaries, indicating the shift from permissive to pre- or proscriptive normative understanding. Complementarily, prolonged deviation signifies a deontic shift towards permissiveness. We leave the representation of viscosity along the deontic range itself to the chosen operationalisation and implementation of the approach.

6.2.4 Discussion

Before turning to the operationalisation of the approach and application to the remaining challenge from our motivating scenario, we clarify and contextualise our approach.

The concept of Dynamic Deontics introduced here is an approach that emphasises an accessible individualised norm understanding. The mechanism per se operates on the level of the individual, which we believe offers a puristic approach to grow normative understanding in diverse societal make-ups (e.g. homogeneous vs. heterogeneous constellations). Based on the flexible yet unified norm representation, normative understanding in all its diversity can be inferred based on global observation without presuming universally shared understanding a priori. The approach is agnostic with respect to specific learning or norm sharing mechanisms, leaving the choice and operationalisation to the modeller of specific scenarios. In principle, learning mechanisms can be of arbitrary nature, including individual experiential learning, imitation, social learning (Bandura et al., 1961), or even direct communication (the latter could be facilitated using the previously described nADICO (Chapter 5) as a message container). Nevertheless, the analysis of global normative understanding is not explicit and requires additional mechanisms such as statistical aggregation.[14] However, we believe that our concept offers a fundamental alignment with the shaping of norm understanding in social reality, allowing us to retrace principal outcomes for specific scenarios.

This approach presented here is not to be confused with Dynamic Deontic Logic (Meyer, 1988) that utilises dynamic logic to solve the ambiguous understanding of state ('ought-to-be') and action ('ought-to-do') in standard deontic logic (von Wright, 1951b), with focus on the latter (norms over actions). With our approach, 'dynamic' refers to the dynamic (viscous

[14] An accessible means of inspection and analysis will be explored in Chapter 7.

and continuous) nature of normative understanding; it is not a reference to the employed logic. Since our approach does not introduce an explicit logic representation, the inconsistencies of standard deontic logic are secondary in this context. Our pragmatic approach unambiguously describes the gradual un/desirability of evaluated actions.

Since Dynamic Deontics support a behaviourist perspective, an essential aspect is an appropriate operationalisation before we revisit our historical case and explore the impact of differentiated trader role concepts on emerging normative understanding within both trader societies (as introduced in Chapter 3).

6.2.5 Operationalisation

Since we assume a behavioural perspective, agents are conceived as experiential learners, with reinforcement learning (RL) techniques representing the matching metaphor. In this work we operationalise Dynamic Deontics using Q-Learning (Watkins, 1989; Watkins and Dayan, 1992) as introduced in Subsection 2.3.3, since it offers a purist interpretation of behavioural learning and operates unsupervised, which assures strong generalisability potential. However, as mentioned before, the abstract concept of Dynamic Deontics per se is agnostic with respect to a specific operationalisation.

Operationalising Dynamic Deontics using Reinforcement Learning With respect to our approach, Q-Learning provides a natural mapping for most of the discussed features. Representing social interactions as action-reaction sequences along with an associated reward, we can use the expression of quality (i.e. Q-values) of state-action combinations (here: action-reaction combinations) as delimiters of the deontic range. We thus associate the minimum and maximum Q-values with the prohibition (*must not*) and obligation (*must*) ends of the deontic range. The range between these values is compartmentalised as conceptually visualised in Figure 6.1, associating the terms *should not*, *may not*, *may*, and *should* with the respective compartmental value ranges. During its initial learning phase, the agent thus continuously expands its deontic range, leading to a widening of the deontic compartments. The stability of the emerging deontic range is in part determined by an *exploration probability*, i.e. the probability with which an agent chooses to explore state-action transition probabilities, as opposed to exploiting (and reinforcing) developed policies based on previous learning. A second aspect that determines the width of the deontic range is the chosen *discount factor*, describing the extent to which previous knowledge (i.e. the existing Q-value for a state-action combination) is discounted when updating the respective Q-value, which assures the narrowing of the deontic range, should reinforcements subside.[15] As a third element of stability, we assume a sliding window approach of past experiences to determine an agent's deontic range. Thus the minimum and maximum values are calculated as the mean of past minimum and maximum Q-values. To model the transitions to the extreme deontics (*must not* and *must*), we specify the number of rounds for which an updated Q-value penetrates, or, for the inverse case, deviates from a tolerance zone (t_{Pr} and t_{Ob}) around a deontic extreme. We use stability thresholds (in number of rounds) to parameterise the establishment ($th_{establish}$) and destruction ($th_{destruct}$) of a prohibition or obligation.

At this stage it is important to emphasise that the presented operationalisation using RL takes a pragmatic approach to norm interpretation, reflecting a purist behavioural perspective, which bears limitations. Norms, or more precisely, normative strength is solely derived from experiential and observational reinforcements, and thus both reflects the pay-offs associated with given actions as well as their frequency of invocation. Agents do not reflect on norms or reason about compliance or violation; the normative understanding is embedded within agents' accumulated action experience, which is a result of the agents' interaction with their physical and social environment.

Though being limited with respect to normative reasoning, this approach is non-intrusive and prevents confounding influence of the normative framework on the agents' actions; instead it captures behavioural regularities, which we interpret as emerging institutions in the context of our evaluation.

Evaluating Deontic Terms Apart from the mapping of Dynamic Deontics to reinforcement learning, reflecting an agent's normative understanding, we are also interested in *why* he does so. To make this accessible to the experimenter, we propose a mapping for Q-values to deontic terms, which in extension can be used to replace the institutional grammar's deontic component.

Using an illustrative example shown in Figure 6.2, we explore this mapping approach. Assuming the existence of a prototypical Q-table that contains action-reaction pairs as state-action representations, with actions being an individual's action and the reaction being a transactional counterpart's reaction, here represented as potential reactions to cheating behaviour. Each action-reaction combination is mapped to their respective accumulated Q-value. The value range of the entries can be mapped on the deontic range.[16] However, the

[15] For this exemplified operationalisation we simplify our account and set the learning rate to 0.5, thus assuming equal importance of novel and historical knowledge.

[16] Note that this example simplifies the mapping, since the min. and max. values of the deontic range are determined based on mean values of a parameterised number of past min. and max. Q-values.

operationalisation raises questions with respect to a realistic approximation of norm understanding. We explore associated challenges based on two exemplary configurations:

1. Symmetric compartmentalisation between minimum and maximum values, with the centre being the midpoint of the deontic range (deontic centre).

2. Zero-centred asymmetric compartmentalisation, with a predefined deontic centre of 0 and equally-sized compartments along the proscriptive and prescriptive sub-ranges.[17]

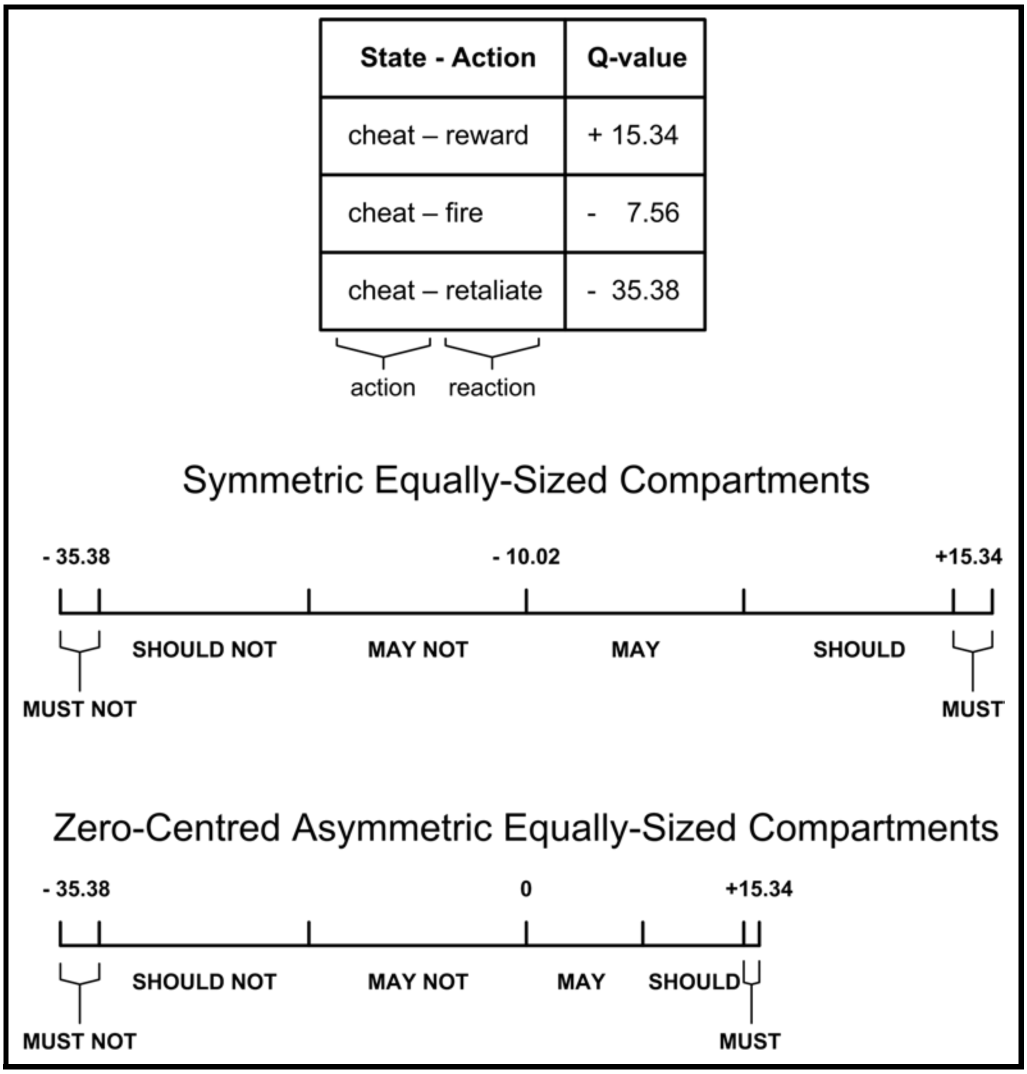


Figure 6.2: Operationalising the Deontic Range based on Q-Learning

[17]To clarify, 'equally-sized' in this context implies compartments of equal size along the prescriptive range (i.e. from the centre to the prescriptive extreme), or the proscriptive range (i.e. from the centre to the proscriptive extreme); it does not imply equal size across the entire range (i.e. it is asymmetric), in contrast to the symmetric configuration, in which all compartments across the deontic range are of equal size.

For the purpose of the ensuing exemplary exploration, we ignore several operational details laid out above, such as the delayed adaptation of the deontic range and ignorance of further reinforcements other than for the evaluated actions. For the Q-table shown in Figure 6.2 an individuals' association with cheating (Action 'cheat') is defined by the mapping of extremal values, here for the reactions 'retaliate' and 'reward', onto the most extreme deontic compartments *must not* and *must*. However, the interpretation becomes more problematic with respect to the intermediate sanction 'fire'. In the first configuration, it would resolve to the deontic term *may*. However, in the second case it would evaluate to *may not*. Being related to the scaling of compartments, we are facing a central philosophical issue, tapping into the conflicting stances of subjectivism and objectivism: Do we assume subjective autonomy of the individual (i.e. individual beliefs)? Extending this, can we permit the incongruence of individual human understanding independent of objectively specified feedback? Under the closed world assumption, the feedback specified by the developer represents the objective 'truth' or 'facts'. In our example we specify feedback under the assumption that values < 0 represent pain and values > 0 elicit pleasure. Subjectively, however, an individual could associate pain with values > 0 if those represent the 'most negative' experience ever attained, thus debasing the objective facts. In our example in Figure 6.2, a symmetric configuration implies an association of neutrality with -10.02, implying feelings of reward for feedback values $>$ -10.02.

The alternative 'objectivist' configuration permits individuals to develop differentiated understanding, but bases it on an objective, i.e. global, specification of neutrality. This likely removes the quality of symmetry, and, more pressingly, demands for a differentiated perspective on deontic compartments, since it potentially leads to a very skewed and strongly biased normative evaluation (note the narrow compartments in Figure 6.2). That, in extension, begs further questions for a psychologically accurate representation of compartments, reflecting effects such as negativity bias (Baumeister et al., 2001) – the greater impact of negative feedback on individuals. Are compartments equally sized? Do proportions change with increasing prescriptiveness of compartments (i.e. is the compartment *may* narrower or wider than *should*?)? Do compartment sizes vary for the pre- and proscriptive side of the deontic range (i.e. do *should not* and *should* have the same size?)?

Considering our intent, the development of a tool for the use by social scientists with a focus on retracing the emergence of subjective understanding of institutions from a behavioural perspective, suggests a preference for the first position, the 'subjectivist' configuration.[18] However, to anticipate problems that may challenge an objectified evaluation,

[18]See Diesing (1966) for a dated yet insightful overview on both positions in the context of the social sci-

we will consider both deontic range configurations for experimental evaluation, especially since we cannot claim objective precision for any outcome other than developing a general understanding. To develop a refined understanding, we explore the different deontic range configurations in the context of the experimental evaluation in Section 6.3.

Aggregating Deontic Values based on the Institutional Grammar Up to this stage, we have explored the operationalisation of the deontic range for individual actions and associated reactions. However, to generalise a normative attribution on the action level, we require the integration of related individual action-reaction pairs.

For this purpose we draw on the previously introduced nested institutional grammar, nADICO, since it provides the scaffolding for the pragmatic structural representation of institutions as shown in Figure 6.3.[19] The *Attributes (A)* component holds actor-related information, such as identifier, or, in our case more relevant, role descriptions, while the *Deontic (D)* component describes duties or permissions in the form of deontic primitives, or, as we have introduced in the previous paragraphs, deontic values that can be resolved on the deontic range. The *AIm (I)* component maps well onto the action concept, while the *Conditions (C)* component effectively represents context, an aspect that is secondary at this stage.

Component	A	D	I	C	O
Mapping	Actor	Duty	Action	Context	Consequence

Figure 6.3: Mapping of Actions onto the nADICO Structure

Using the structural concept of levels as introduced in the context of nADICO, with vertical nesting representing consequences (*O* in Figure 6.3) on different levels, we can associate common actions (in our example in Figure 6.2: 'cheat') with multiple reactions on a nested level. Applying the previously introduced mapping to our example with specific focus on Actor/Attributes (A) and Action/AIm (I), Figure 6.4 exemplifies the corresponding nesting structure.

A	D	I	C	O			
actorA	*	cheat	*	actorB	*	reward	*
				actorB	*	fire	*
				actorB	*	retaliate	*

Figure 6.4: Mapping Actions and Reaction on the nADICO Structure

ences.

[19] Alternative mappings for the individual nADICO components have been outlined in Subsection 5.3.3.

Since all reactions share a common action, they are integrated using nADICO's horizontal nesting operators (see Subsection 5.3.2). For our example as well as the later exploration we do not assume pre-established rules, or notions of collective action. We can further assume that reactions occur independently, and for the normative case potentially concurrently, thus suggesting the use of the `or` operator to combine the different consequences. However, the operationalisation likewise applies to the `xor` operator, since the concrete consequence is unknown. The remaining challenge relates to the question of how to aggregate the individually accumulated feedback values. To do so, one could assume a variety of strategies associated with the individual's personality that act as a filter on the interpretation of normative understanding.[20] A selection of possible strategies is given in Table 6.1.

Table 6.1: Deontic Determination Strategies

Rational Strategy	A rational individual attaches the *mean value of all sanctions* to the common action.
Pessimistic Strategy	A pessimistic individual would fear the invocation of the *worst possible sanction* for a given action and associate this value with the action.
Optimistic Strategy	An optimistic individual would expect the invocation of the *best possible sanction* for a given action.
Opportunistic Strategy	An opportunistic individual would associate *extremal sanctions*, i.e. • the worst possible sanction value for an overall undesirable action, and • the best possible sanction value for an overall desirable action with the evaluated action.

We will operationalise all the mentioned strategies. However, for our evaluation we will concentrate on the least biased strategies, i.e. the rational and the opportunistic ones.

Let $stmt_l$ describe a nADICO (monitored) statement on the level l, and $stmt_{(l+1),i}$ represent the i^{th} (consequential) statement on level $l+1$. $d(stmt_l)$ is the deontic of statement $stmt_l$; $count_l$ holds the number of statements on level l, and $c_{deonticRange}$ represents the centre of the deontic range.

Rational Strategy For the rational strategy the leading monitored statement's deontic ($d(stmt_l)$) is expressed as the mean of all consequential statements' deontics.

[20] Note that personality traits are not necessarily inherently *personal* but could be culturally determined (Hofstede et al., 2010).

$$d(stmt_l) := \frac{\sum_{i=0}^{count_{(l+1)}} d(stmt_{(l+1),i})}{count_{(l+1)}}$$

Pessimistic Strategy As indicated before, the pessimistic strategy concentrates on the most negative consequential deontic value:

$$d(stmt_l) := min(d(stmt_{(l+1),0}), \ldots, d(stmt_{(l+1),count_{(l+1)}}))$$

Optimistic Strategy Complementing the pessimistic strategy, optimistic agents choose the most positive consequential deontic value:

$$d(stmt_l) := max(d(stmt_{(l+1),0}), \ldots, d(stmt_{(l+1),count_{(l+1)}}))$$

Opportunistic Strategy For the opportunistic strategy the individual associates the extremal deontic values of consequential statements with the monitored statement's deontic.

Doing so, we initially identify the *deontic bias*, i.e. the direction (left or right from the centre $c_{deonticRange}$) towards which the deontic values point overall. We determine this bias from the sum of all consequential deontic values. Depending on the direction, i.e. an overall negative or positive connotation with the given action, we identify the respective extremal value associated with the nested reactions (max. value for prescriptive direction; min. value for proscriptive direction):

$$extremeDeontic(stmt_l) := \begin{cases} max(d(stmt_{(l+1),0}), \ldots, d(stmt_{(l+1),count_{(l+1)}})), & \text{if } (\sum_{i=0}^{count_{(l+1)}} d(stmt_{(l+1),i})) \\ & > c_{deonticRange} \\ min(d(stmt_{(l+1),0}), \ldots, d(stmt_{(l+1),count_{(l+1)}})), & \text{otherwise} \end{cases}$$

The extremal deontic value ($extremeDeontic(stmt_l)$) is only used if the sum of the consequential deontics is not located exactly at the centre of the deontic range (which would imply that the deontics of the nested statements cancel each other, or none exist). In that case, the deontic range centre itself describes the monitored statement's deontic. The derived deontic value is thus

$$d(stmt_l) := \begin{cases} c_{deonticRange}, & \text{if } (\sum_{i=0}^{count_{(l+1)}} d(stmt_{(l+1),i})) = c_{deonticRange} \\ extremeDeontic(stmt_l), & \text{otherwise} \end{cases}$$

Converting Causal Relationships to Social Consequences To this stage we have aligned the aggregation process with nADICO's nesting capabilities. However, so far we did not consider a semantic mismatch between nADICO's institutional statements and the statement structure derived based on deontic value aggregation.

Q-Learning reflects subjective causal associations of action and reward. Translating the third example entry from Figure 6.2 (<cheat - retaliate>), we can reconstruct an agent's understanding as "Cheating is undesirable *because* when doing so I experience retaliation.", which we use to construct expectations about future reactions. Q-Learning state-action pairs (modelling action and respective reaction by another agent) represent "because", or "on the grounds of" relationships, suggesting an *associative activation* of action and reaction. nADICO, on the other hand, attaches a social consequence to an injunction imposed on the actor, using the *Or else* component to express social consequences of non-adherence ("..., otherwise ..."), thus indicating *exclusive activation* of either monitored or consequential statement. In addition to the semantically differing relationship between monitored and consequential statement, nADICO expresses individual atomic statements (i.e. decomposed monitored and consequential statements) from the perspective of the respective actor. For the translation of the Q-Learning interpretation into nADICO statements this implies a shift from first to third person (i.e. the interaction counterpart) for the consequential statement. Instead of expressing "... when doing so I experience ...", we thus say "... otherwise the cheated trader ...". As such the nADICO representation facilitates an applied account of empathetic perspective-taking, predicting an actor's future behaviour based on past conduct. For the previously developed value aggregation and attached sanctions, this implies a necessary semantic transformation from an 'ego-centric causal relationship' to an 'actor-centric social consequence'. Table 6.2 summarises the essential semantic differences between the relationship of state-action pairs in Q-Learning and social consequences in nADICO.

Table 6.2: Causal vs. Social Consequence Semantics

Representation	*Statement Perspective*	*Relationship*	*Activation of Consequence*
Q-Learning	first person	causal	associative
nADICO	third person	social consequence	exclusive

We reflect this transformation by *inverting* the consequential statements' deontic values along the deontic range. Taking different deontic range configurations into account (such as asymmetric value ranges), the inversion is proportional to the absolute range value of the deontic's original direction. From a perspective of deontic compartments, this translates to an inversion from *should not* to *should*, *may not* to *may*, *must not* to *must*, and vice versa.

Figure 6.5 exemplifies the proportional inversion from *should not* to *should* for the different deontic range configurations.

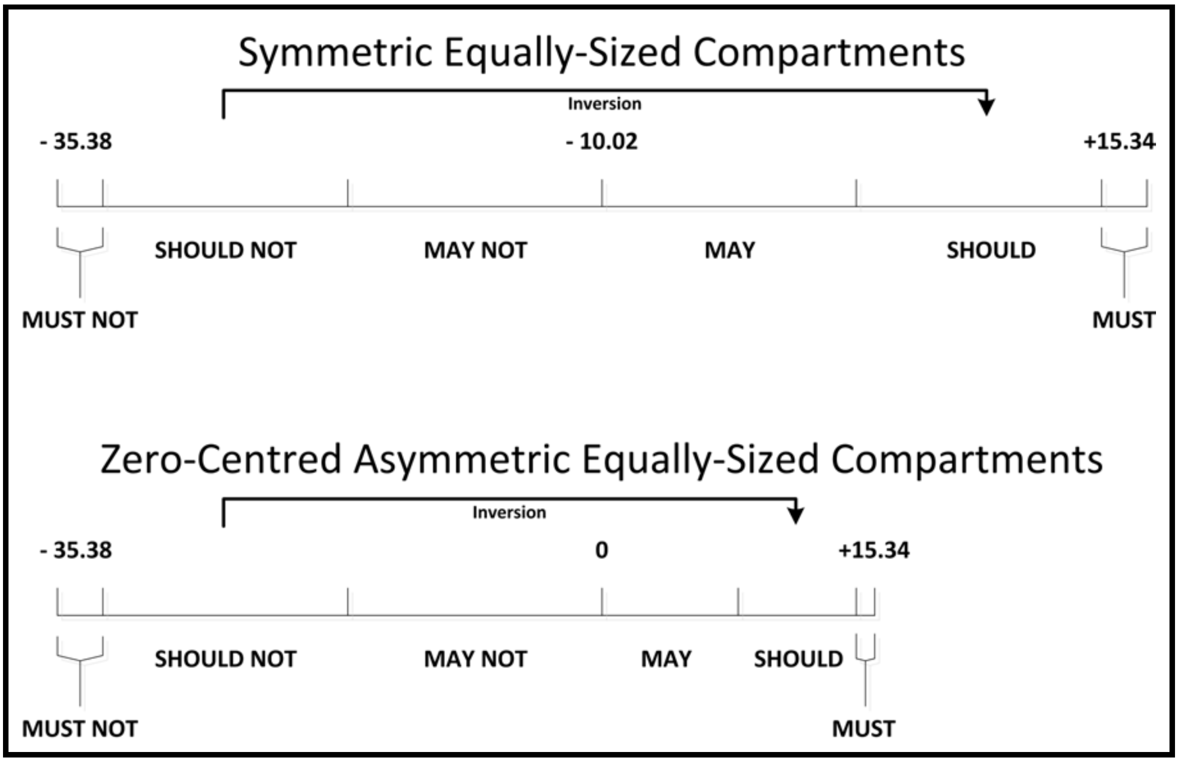


Figure 6.5: Proportional Inversion of Deontics

Using our running example with deontic values from Figure 6.2, we schematically visualise the derivation of the monitored deontic and the inversion of consequential deontic values (under the assumption of opportunistic strategy on a zero-centred asymmetric deontic range) as shown in Figure 6.6. As mentioned before, the horizontally nested consequential statements are combined using the `or` operator.

We can then translate the derived nADICO statement into natural language. Assuming that the deontic value of the monitored statement resolves to *must not*, we can say "I (actorA) (`A`) must not (`D`) cheat (`I`), or else (the cheated trader) actorB (`A`) *must not* (`D`) reward me (`I`), or *may* (`D`) fire me (`I`), or *must* (`D`) retaliate against me (`I`).", thus translating into an `ADIC(ADIC or ADIC or ADIC)` statement.[21]

[21]Recall that the use of the *Conditions* component is optional (see Section 5.2).

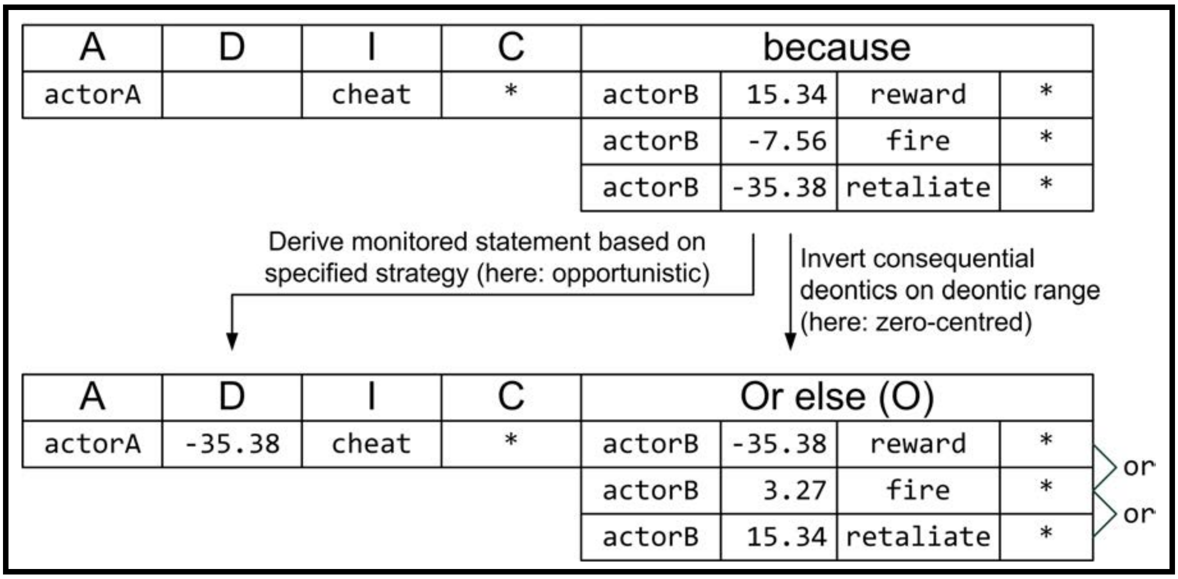


Figure 6.6: Deriving nADICO Statements with Inverted Consequential Deontics

6.2.6 Summary

In this section we have introduced the concept of *Dynamic Deontics* as a mechanism to model the development and continuous adjustment of normative understanding based on social interaction. We operationalised it using reinforcement learning to emphasise a behavioural perspective of norm formation, as opposed to specifying such understanding a priori or impose it otherwise. As part of its operationalisation, we integrated the approach with the institutional grammar introduced in Chapter 5. The combined use of nADICO and Dynamic Deontics enables a comprehensive and dynamic institution representation that manages the trade-off of expressing general institution characteristics (unified institution structure) while incorporating individualised understanding based on agents' personality (evaluation strategies) and history (dynamic deontic range). Besides introducing the concept, we highlighted selected concerns that require future attention, such as the nature of deontic compartments (width, symmetry, progressive change) and the mapping of Q-values onto the deontic range (along with associated philosophical and psychological implications).

At this stage we have provided the conceptual base for the exploration of the final question addressing an assumption previous work has made with regards to the analysis of the Maghribī Traders Coalition (see Chapter 4): Could the *integrated* role understanding of the Maghribīs – in contrast to the stratified role understanding among Genoese – have shaped an aligned world view that primed them for cooperation?

6.3 Experiment 'Evolving Norm Understanding under Maghribī and Genoese Role Conceptualisations'

With the conceptual compromise of pure representation and experimenter accessibility, we target the last part of our substantive investigations related to the Maghribī Traders. This work has previously been published in Frantz et al. (2015b) and is presented here in an extended form.

6.3.1 Motivation

To recall, the Maghribī traders were North African traders that managed to sustain cooperative behaviour in the context of long-distance trade without relying on formal institutional instruments.[22] Their counterpart, the Southern European Genoese traders relied on formal instruments such as contracts to govern their behaviour, since their open society challenged the effectiveness of informal cooperation.

In our work we reviewed seminal work in this area and reviewed the extent to which a relaxation of previous assumptions could have affected the cooperation outcome. Greif's game-theoretical account (Greif (1989, 2006), see Subsections 4.1.1 and 4.1.2) modelled Maghribīs as inherent information sharers, while Genoese traders were reportedly secretive. Relaxing the assumption of non-communication among Genoese traders using experimental exploration with ABM (see Subsection 4.2.1), we support Greif's view and suggest that enhanced information sharing would *not* have made a sufficient difference in the cooperation outcome to warrant a claim that informal mechanisms could have been sufficient to assure cooperative behaviour in the individualistic Genoese society.

A second assumption put forth by Greif was the conception of the Maghribī Traders as a coalition that was inherently closed to outsiders. We explored recent literature accounts that documented the existence of an apprenticeship system that Maghribīs used to instruct and monitor newcomers before joining the wider coalition. Our experimental evaluation (see Subsection 4.2.2) explores properties that such apprenticeships could have had, and suggests that the apprenticeship system could have well functioned as a mechanism to manage the trade-off of giving controlled access to outsiders while limiting cheating within the group, and thus replacing the rigid assumption of closedness.

A last, and in our view significant, detail Greif's approach ignores is the differentiated role understanding among Maghribīs and Genoese traders. Maghribī traders had a compar-

[22] A discussion of the differentiation of 'formal' and 'informal' is provided in Subsection 3.2.1.

atively uniform background, being members of a Jewish middle class most of which were emigrants from Iraq that had settled in the Western Mediterranean basin, and leveraged their strong in-group ties to trade throughout the Fatimid Empire and beyond (see Section 3.1 for greater detail). The Genoese traders, in contrast, established a society which was characterised by strong occupational specialisation and a relatively unconstrained influx of outsiders.[23] To established Genoese, trade was largely seen as a promising investment, ideally providing the funds for trade ventures without taking an active part in the operation. The investing 'commendatores' (which we refer to as 'investors') left this task to 'tractatores' (which we name 'operators') who were contractually endowed with performing the laborious part of the agreement against the payment of a fixed fraction (usually one quarter) of the realised profit.[24] Since the initial financial investment was largely borne by one party and profit was nearly guaranteed,[25] performing the mercantile activity solely relied on the availability and willingness of candidates. Thus, performing long-distance trade attracted opportunistic newcomers under the prospect of achieving significant earnings based on a one-off commitment. Consequently, trade relationships among the Genoese were characterised by a considerable class stratification.[26] While the investment part was concentrated with upper class family firms,[27] trade operations were generally performed by poor individuals of low status (de Roover, 1958; Byrne, 1916).

This differentiated role understanding, driven by a vastly differing social stratification – the comparatively homogeneous Maghribīs vs. heterogeneous Genoese – along with different societal characteristics (relatively 'closed' vs. 'open' societies) leads to the suggestion that an integrated role understanding among Maghribī and a stratified role understanding among Genoese traders could have contributed to the determination of a(n) (un)cooperative outcome.

Though documented, Greif did not consider this differentiation in his analyses. In the following we thus conceptualise an experiment that addresses this hypothesis from a behavioural perspective. Taking a comparatively puristic approach,[28] we employ the representational mechanisms introduced in the previous chapter (Chapter 5) and the current chapter.

[23] For a more comprehensive characterisation of the Genoese refer to Section 3.3.

[24] Recall our extensive discussion on the different forms and characteristics of the 'commenda' in Subsection 3.1.3.

[25] Recall the prospective profit of 20 to 110 percent (van Doosselaere (2009), see Subsection 3.3.2).

[26] Only 21 percent of Genoese traders operated both as investors and operators throughout their lifetime, generally with a shift from early operational experience, followed by an investor perspective (Greif, 2012).

[27] Recall that for the period of 1154-1164 AD 37 families sponsored 90 percent of the capital (Greif, 1993, 2006; González De Lara, 2008). See also Subsection 3.3.2.

[28] In contrast to the earlier scenarios in which we modelled the entire trade interaction, we will concentrate on the impact of behavioural learning on the cooperation outcome.

6.3.2 Model

By acting in and experiencing both roles, we posit that Maghribīs could have developed an integrated role understanding that would have shaped an aligned world view, i.e. manifested behavioural regularities and thus institutions, within the trader society. Ignoring further external effects, this could have supported the emergence of an overall cooperative outcome. For Genoese, the differentiated perspectives drove individuals to further their selfish ends characterised by their specific roles. The emerging divergence of interest, defined in the respective role understanding, then led to a social configuration that required external, formal means to assure compliance. This position is certainly reductionistic, since it ignores external factors that determine overall cooperation outcome (environmental factors, openness of societies, etc.). However, we see our argument in line with the suggestion that overlapping role understanding, an aspect associated with pre-modern societies (Durkheim (1933), see Section 4.3), could been a factor important enough not to be ignored in a systematic comparative analysis of both societies.

Feedback Specification Our model concentrates on evolving normative understanding from a behaviourist perspective. Given its explorative nature, our model is intentionally simple and explicit about the feedback associated with individual actions. Instead of modelling trade interactions in full detail, we concentrate on the essential decisions of long-distance trade operators to return realised profits truthfully in combination with potential reactions on the part of the investors. Operators, which we assume to hold goods they have been sent by investors, have two actions at their disposal and must decide whether to be compliant and return realised profits (cooperate) or to withhold the profit (defect). The investor responds to this action choice by applying a corresponding reaction based on his[29] own experience. Each combination of action and reaction leads to a individualised feedback for the respective party, an overview of which is provided in Table 6.3.

Exemplifying its interpretation, the compliant action of trading fair (on the part of the operator) and being rewarded as a reaction (by the investor), represented as the combination 'TRADE FAIR - REWARD', leads to a positive pay-off for both parties. To reflect the expected nature of such outcome and to set the stakes for cooperation high, we represent this with conservative low values (+1 for each party). To reflect the notion of gradual sanctioning, reacting investors can explore a set of different reactions with varying impact on either party. For example, trading fair and being fired (combination 'TRADE FAIR - FIRE') leads

[29]Recall that the choice of the male form is historically accurate, since the involvement of women as travelling parties in medieval long-distance trade has not been documented (see the discussion in Subsection 3.2.2).

Table 6.3: Action Reaction Feedback Combinations

Action-Reaction Combinations		**Utility from Actions**	
Action (Operator)	**Reaction (Investor)**	**Operator**	**Investor**
TRADE FAIR	DON'T REWARD	-1	+1
TRADE FAIR	FIRE	-2	-1
TRADE FAIR	RETALIATE	-3	-1
TRADE FAIR	REWARD	+1	+1
WITHHOLD PROFIT	DON'T REWARD	0	+1
WITHHOLD PROFIT	FIRE	-1	+1
WITHHOLD PROFIT	RETALIATE	-2	+1
WITHHOLD PROFIT	REWARD	+2	-2

to a negative feedback on the part of the operator, since he has been fired despite his honest conduct. For the investor it implies the loss of a compliant trader, whereas merely withholding the reward ('DON'T REWARD') is to the investor's advantage without necessarily losing an operator (e.g. under the strategic pretence of dissatisfaction[30]). Applying a more rigid sanction ('RETALIATE') results in even stronger negative feedback for the operator.

On the flip-side, if operators choose to withhold profit and are fired (combination 'WITHHOLD PROFIT - FIRE'), it creates mild negative feedback (since the reaction is appropriate for the dishonest conduct). For the reacting investor, all sanctions (apart from the case of rewarding a non-compliant trader) are associated with a constant positive reward, reflecting an appropriate choice for cheating behaviour and the assumption that 'deserved punishment' can elicit feelings of reward in the part of the sanctioner (de Quervain et al., 2005). Receiving a reward despite cheating results in a strong positive feedback on the part of the operator and negative feedback on the part of the investor, since either party makes a disproportionate gain (cheater gains withheld profit *and* reward) and loss (investor loses profit *and* reward).

To emphasise behavioural aspects beyond individualised experiential learning, we allow agents to perform social learning (Bandura et al., 1961) based on the observation of other agents' interactions. Furthermore, reflecting additional features existent in the Maghribī society, we consider a notion of norm enforcement, in which third parties can observe and themselves apply a reaction to the observed party's action using their own experientially and socially acquired knowledge. Features of experiential learning, social learning and norm enforcement thus act complementary to establish essential aspects of norm formation. While social learning represents behavioural internalisation processes, enforcement represents the

[30]Recall that Maghribīs strategically challenged fellow traders' compliance (see Subsection 3.2.2; Goldberg (2012b)).

externalisation of acquired behavioural knowledge which norm enforcers use to influence the enforcement target's behaviour. This effectively represents a closed micro-macro feedback loop by which individual habits are modified based on the reconstitutive downward causation of norms (see Subsection 2.2.2).

Though differing from Greif's method, our approach matches the intuitions that underlie Greif's modelling approach: "A structure made up of institutionalized rules and beliefs enables, guides, and motivates the self-enforcing behavior that reproduces it. Most individuals, most of the time, follow the behavior that is expected of them" (Greif, 2006). While reinforcement learning facilitates the reinforcement and thus reproduction of behaviour, the combined use of nADICO and Dynamic Deontics allows a comprehensive representation not only of how the individuals act, but foremostly, of how they perceive the institutional environment based on the given representational structure.

With this underlying motivation in mind, we introduce the model operationalisation in the following.

Agent Execution Cycles To reflect the different societal configurations in the implementation, we construct varying execution cycles, with a first one that does not differentiate between the different role perspectives (Algorithm 6.1) – thus representing the Maghribī configuration. Complementing this, the second variant (Algorithm 6.2) considers the same set of activities, but ties their execution to the instantiation in either the investor or the operator role – thus reflecting the Genoese setup.

Algorithm 6.1: Agent Execution Cycle – Maghribī Version

```
// Pick two agents: 1st agent for learning or enforcement, 2nd agent for operator action execution
Pick two random other agents;
Decide whether to explore or exploit in this round;
if exploring then
    Pick random action from action pool;
    // Social Learning
    Observe action of first randomly chosen agent and internalise action-sanction combination
        along with valence (not actual reward value);
else
    Pick action with highest Q-value from action pool;
    if norm enforcement activated then
        // Norm Enforcement
        Sanction action taken by first randomly chosen agent using sanction with highest
            Q-value;
        Memorise feedback from sanction choice;
end
Execute picked action and apply to second randomly chosen agent;
Memorise reaction and make action-reaction combination visible to other agents;
Update deontic range (see Algorithm 6.4);
```

Algorithm 6.2: Agent Execution Cycle – Genoese Version

```
Initialisation: Assign either investor or merchant role;
Decide whether to explore or exploit in this round;
if exploring then
    if is merchant then
        Pick random action from action pool;
        // Social Learning
        Observe action of randomly chosen agent and internalise action-sanction combination
            along with valence (not actual reward value);
else
    if is merchant then
        Pick action with highest Q-value from action pool;
    if is investor & norm enforcement activated then
        // Norm Enforcement
        Sanction action taken by randomly chosen agent using sanction with highest
            Q-value;
        Memorise feedback from sanction choice;
end
if is merchant then
    Execute picked action and apply to randomly chosen agent;
    Memorise reaction and make action-reaction combination visible to other agents;
Update deontic range (see Algorithm 6.4);
```

The interaction between two agents is modelled starting from the operator's perspective: initially, Maghribī agents (Algorithm 6.1) randomly choose fellow agents. The first agent will be of relevance for the internalisation and externalisation processes, such as social learning and enforcement aspects. The second agent is the target of the trade interaction. For the Genoese case (Algorithm 6.2) only one agent is chosen as an individual, and depending on its role, it will either perform social learning (operator) or enforcement (investor). Having chosen agents, an individual determines whether to explore new actions or exploit its existing policies developed based on previous experience (see the discussion on the exploration-exploitation trade-off in the context of reinforcement learning in Subsection 2.3.3).

If choosing to explore (which we select based on a Boolean value chosen with a probability of 0.1 for the value 'true'),[31] the agent picks a random action from the set of actions and engages in social learning. It does so by observing the first randomly chosen agent's last experienced action-reaction combination (and the associated feedback). To abstract from the actual feedback (in order to soften the precision of observational learning compared to direct experiential learning), the observer internalises the *valence* associated with the feedback from an operator's perspective, which we simply represent with a negative (-1) or positive value (+1).[32]

[31] We summarise all parameters in Table 6.4.

[32] Neutral feedback signals indifference. However, in this case any feedback has positive or negative valence.

Alternatively, if exploiting, an agent picks the action with the highest Q-value from the action pool. The chosen action, whether picked randomly or based on highest Q-value, is then used for the trade interaction with the second randomly chosen fellow agent. Switching to the target's perspective (shown in Algorithm 6.3), the action target (which is a fellow trader in the Maghribī case or an investor in the Genoese variant) reacts according to its situational focus on exploration or exploitation based on its own execution cycle. If exploring, it applies a randomly chosen reaction, or, if exploiting, it applies a reaction that promises the best possible feedback based on past experience (i.e. the reaction that offers the highest Q-value with respect to the imposed action). Both the acting and reacting party then memorise their respective feedback (as specified in Table 6.3), and the acting party makes the action-reaction combination visible in order to allow potential third parties to observe (for social learning) and react (for norm enforcement) (see Algorithms 6.1 and 6.2). If norm enforcement is activated, third parties can, as part of their execution cycle, apply their own reaction/sanction to an observed agent's action, simulating multilateral norm enforcement. Feedback from the chosen enforcement reaction is internalised by the sanctioner.

Algorithm 6.3: Agent Reaction – Maghribī Trader/Genoese Investor

```
Incoming action by other agent
if exploring (determined in execution cycle) then
│  Pick random reaction;
else
│  Choose reaction (for incoming action) based on highest Q-value for action-reaction
│      combination;
end
Perform chosen action against action sender;
Memorise feedback and associate with action-reaction combination;
```

The remaining operations in the execution cycles described in Algorithms 6.1 and 6.2 are related to the Dynamic Deontics operationalisation and include updating the deontic range boundaries, checking for norm transitions based on continuous reinforcements, and the application of a discount factor to all Q-values (see Subsection 2.3.3), simulating the contraction of the deontic range if no further reinforcement occurs. The operational details are described in Algorithm 6.4.

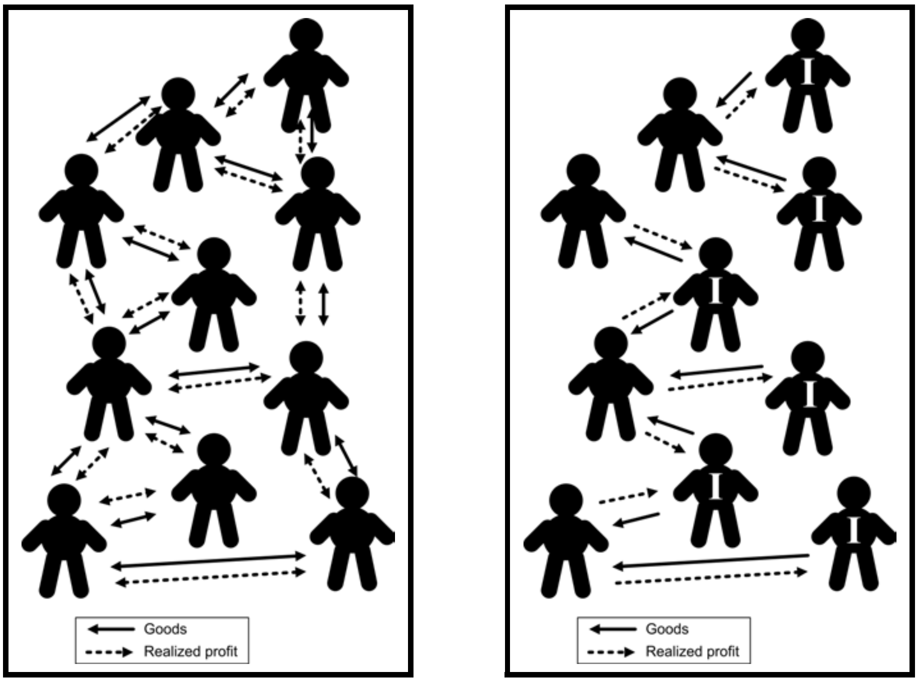
Scenario Configurations Individual agents hold a single memory representation, independent of the number of roles they play. All action or reaction feedbacks associated with these roles (i.e. feedback from operator and/or investor perspective as specified in Table 6.3), whether from two roles, such as in the Maghribī case, or from a single role, such as in the Genoese case, are fed into the individual's memory instance. Whether playing multiple roles

Algorithm 6.4: Deontic Range Update

```
// Update deontic ranges limits
Upper limit of deontic range (max_deonticRange) ← mean value of maximum Q-values across past
    length_history rounds;
Lower limit of deontic range (min_deonticRange) ← mean value of minimum Q-values across past
    length_history rounds;
// Check for stability
foreach action in actions do
    // Check for establishment of prohibition
    if deontic value for action ≤ (min_deonticRange + t_Pr * (max_deonticRange - min_deonticRange))
    for ≥ th_establish rounds then
        Register action as prohibition;
    else
        // Check for release of prohibition
        if deontic value for action > (min_deonticRange + t_Pr * (max_deonticRange - min_deonticRange)) for ≥
        th_destruct rounds then
            Release prohibition action;
    end
    // Check for establishment of obligation
    if deontic value for action ≥ (max_deonticRange - t_Ob * (max_deonticRange - min_deonticRange))
    for ≥ th_establish rounds then
        Register action as obligation;
    else
        // Check for release of obligation
        if deontic value for action < (max_deonticRange - t_Ob * (max_deonticRange - min_deonticRange)) for ≥
        th_destruct rounds then
            Release obligation action;
    end
end
// Contract deontic range
Apply discount factor to Q-values;
```

or a single role, individuals develop an integrated normative understanding based on the Dynamic Deontics concept introduced in the previous section. From a conceptual perspective this bears further implications, which we highlight with reference to the Figures 6.7a and 6.7b. In our model, during each execution round, interaction is initiated by operators. Since Maghribī traders operated both in investor and operator roles, *all* agents perform one interaction with a second agent, and each agent needs to expect its invocation as an investor (see Figure 6.7a). The Genoese society, in contrast, has been modelled with specialised roles by initialising 50 percent of the agents as investors (labelled *I* in Figure 6.7b) and the other half as operators. In order to assure a comparable number of interactions (and thus reinforcements) for both scenarios, we double the number of traders for the Genoese scenario. Thus if initialising the Maghribī variant with 100 agents, the corresponding Genoese variant will be initialised with 200 agents.

The basic model parameters shown in Table 6.4 (beyond the specified feedbacks in Table



(a) Maghribī Role Constellation (b) Genoese Role Constellation

Figure 6.7: Schematic Overview of Role Interactions

6.3) include the number of agents, which are initialised as discussed before. The tolerance zones for the extreme deontics (i.e. prohibition and obligation) are parameterised at 5 percent of either end of the deontic range. For analytical purposes, we likewise introduce a tolerance zone of 5 percent around the center of the deontic range. Thresholds for establishment of prohibition and obligation norms are established at 100 for establishment and 200 for destruction, suggesting the delayed shift from prescriptive to permissive norms. The history length for upper and lower Q-values that are averaged to calculate the deontic range boundaries are parameterised at 100 rounds. Finally, we specify a memory discount factor of 0.99 and an exploration probability of 0.1 as indicated before.

Table 6.4: Simulation Parameters

Parameter	**Value**
Number of agents	100 (Maghribīs); 200 (Genoese)
Tolerance zone around extreme deontics (t_{Pr}, t_{Ob})	0.05 of deontic range amplitude
Norm establishment threshold ($th_{establish}$)	100 rounds
Norm destruction threshold ($th_{destruct}$)	200 rounds
Deontic range history length ($length_{history}$)	100 rounds
Memory discount factor	0.99
Exploration probability	0.1

The introduced model allows us to specify four possible societal configurations:

- Scenario 1 – Role Integration without Norm Enforcement
- Scenario 2 – Role Integration with Norm Enforcement
- Scenario 3 – Role Specialisation without Norm Enforcement
- Scenario 4 – Role Specialisation with Norm Enforcement

Aligning the scenarios with historical reality, we find that the Maghribī configuration is in closest alignment with Scenario 2, which reflects an integrated role understanding along with norm enforcement. For the Genoese traders, Scenario 3 offers the most likely configuration, since it relies on a specialised role understanding without additional norm enforcement. However, to improve comparability and to address the interesting question of whether Genoese norm enforcement could have made a difference, we explore all possible societal configurations.

We performed simulation runs over 10,000 rounds, since test runs reveal a stabilisation of the established normative understanding within that period for all configuration. Each configuration has further been randomly seeded and executed 30 times. The initial exploration further revealed challenges in the context of the symmetric deontic range configuration (see Subsection 6.2.5). We will explore this aspect in the context of discussing the results of the zero-centred deontic range configuration. For all configurations we explore both the rational and opportunistic agent strategy (see Subsection 6.2.5).

6.3.3 Results

We organise the results section by societal configurations and initially explore the impact of the varying Dynamic Deontics configurations with respect to their ability to provide a useful representation of the agents' evolving normative understanding. For a retraceable discussion we provide time-series charts of individual representative simulation runs, before providing a summarising statistical overview. We explore normative understanding from the perspective of the operators based on the two potential action choices, 'trading fair' or 'withholding profit'.

Scenario 1 – Integrated Role Understanding, No Norm Enforcement, Symmetric Deontic Range Exploring the two proposed deontic range configurations – the symmetric as well as the zero-centred asymmetric one (see Subsection 6.2.5) – we initially ran the scenario using the symmetric deontic range configuration along with a rational trader setup in which the deontic value was determined based on the mean of all experiential values (see Subsection 6.2.5). As observable in Figure 6.9a, around 80 percent of all traders believe that not trading fair is mildly preferable (*may not*), complemented with a remainder that is largely indifferent about the desirable behaviour. However, withholding profits (Figure 6.9b) appears even less desirable (similarly, 80 percent interpret it as a *should not*, with the remainder opting for *may not*). On the first view this appears as a contradiction, since both actions are undesirable. However, given the individual's "urgency of practice" (Bourdieu, 1981), i.e. the necessity to act, individuals can experience predominantly negative feedback for all action choices. However, in this particular case the reason for the low appeal of trading is caused by a conceptual problem of the symmetric deontic range configuration. We explore this by inspecting an individual agent's situational deontic range as shown in Figure 6.8.

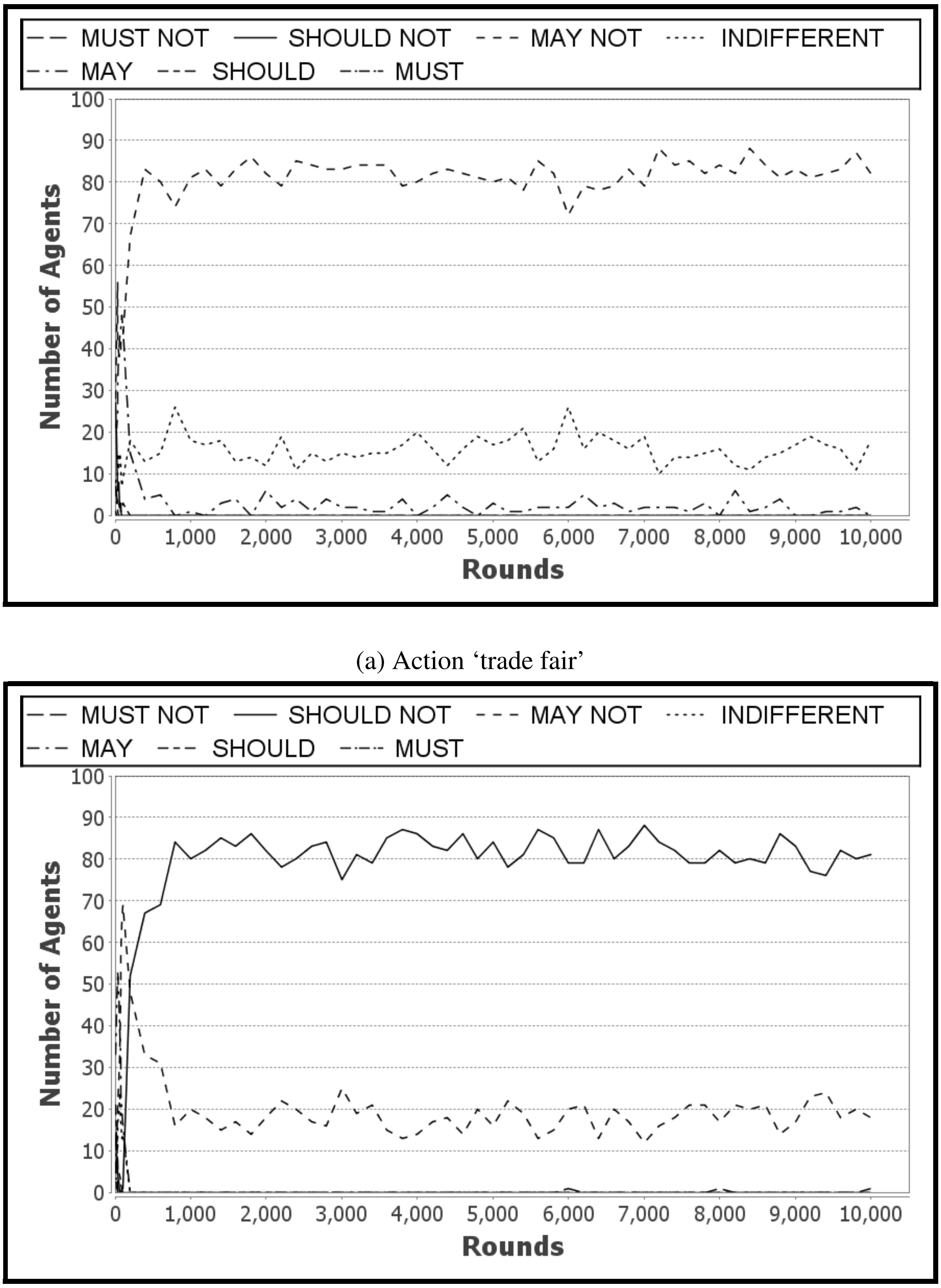
```
MUST NOT: -40.739162
SHOULD NOT: to 3.1241875
MAY NOT: to 46.987537
INDIFFERENT: 46.987534 (Tolerance: 42.6012 to 51.373867)
MAY: to 90.85089
SHOULD: to 134.71425
MUST: 134.71423

L0: A=Maghribi, D=MAY NOT (20.430897), I=trade fair, C=*,
 O=(
  (L1 (p: 0.25336322): A=*, D=MUST (inv) (-37.627388), I=don't reward, C=*, O=(null)) OR
  (L1 (p: 0.027721157): A=*, D=SHOULD (inv) (-5.4913974), I=fire, C=*, O=(null)) OR
  (L1 (p: 0.68752545): A=*, D=MUST NOT (inv) (131.21716), I=reward, C=*, O=(null)) OR
  (L1 (p: 0.031390134): A=*, D=SHOULD (inv) (-6.3747835), I=retaliate, C=*, O=(null)))

L0: A=Maghribi, D=SHOULD NOT (-1.3753093), I=withhold profit, C=*,
 O=(
  (L1 (p: 0.03802817): A=*, D=SHOULD (inv) (-3.449765), I=retaliate, C=*, O=(null)) OR
  (L1 (p: 0.14859155): A=*, D=SHOULD (inv) (-0.976019), I=fire, C=*, O=(null)) OR
  (L1 (p: 0.8133803): A=*, D=SHOULD (inv) (0.299856), I=reward, C=*, O=(null)))
```

Figure 6.8: Situational Symmetric Deontic Range of Selected Agent

The top part of Figure 6.8 highlights the boundaries for the individual deontic compartments along the deontic range, starting with the lower boundary of the deontic range towards the upper boundary. The bottom part of the figure shows the derived normative understanding for individual actions in the form of nADICO statements, with `L0` indicating first-level statements and `L1` signifying the (vertically) nested statements used to derive the monitored deontic (`D`) of the top-level statement. The attribute `p` (probability) in nested statements indicates the distribution of past invocations across different consequences. Inspecting the



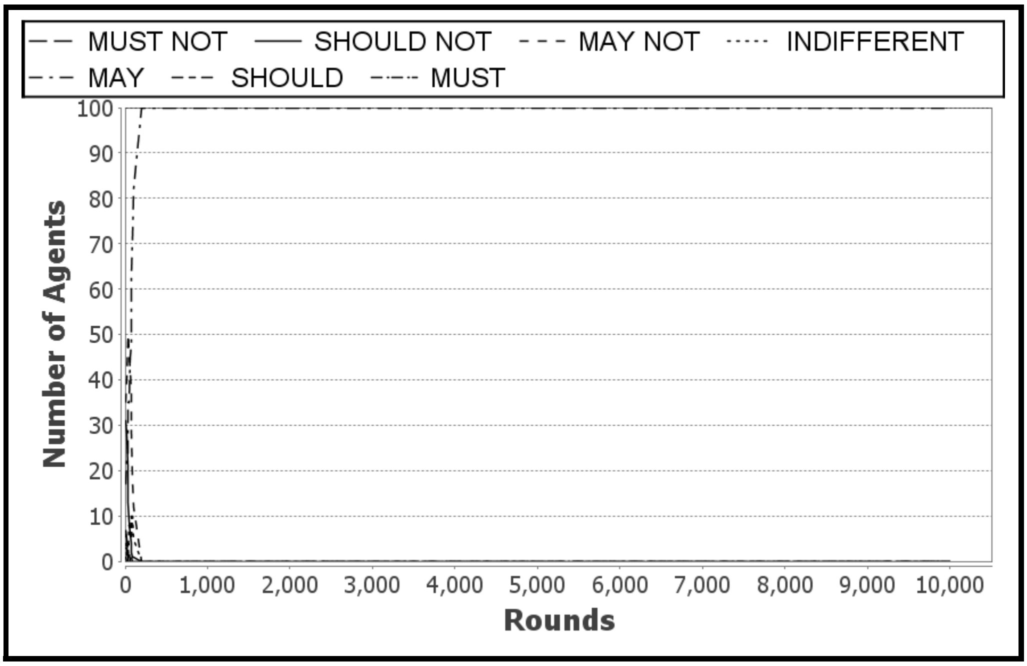
(a) Action 'trade fair'

(b) Action 'withhold profit'

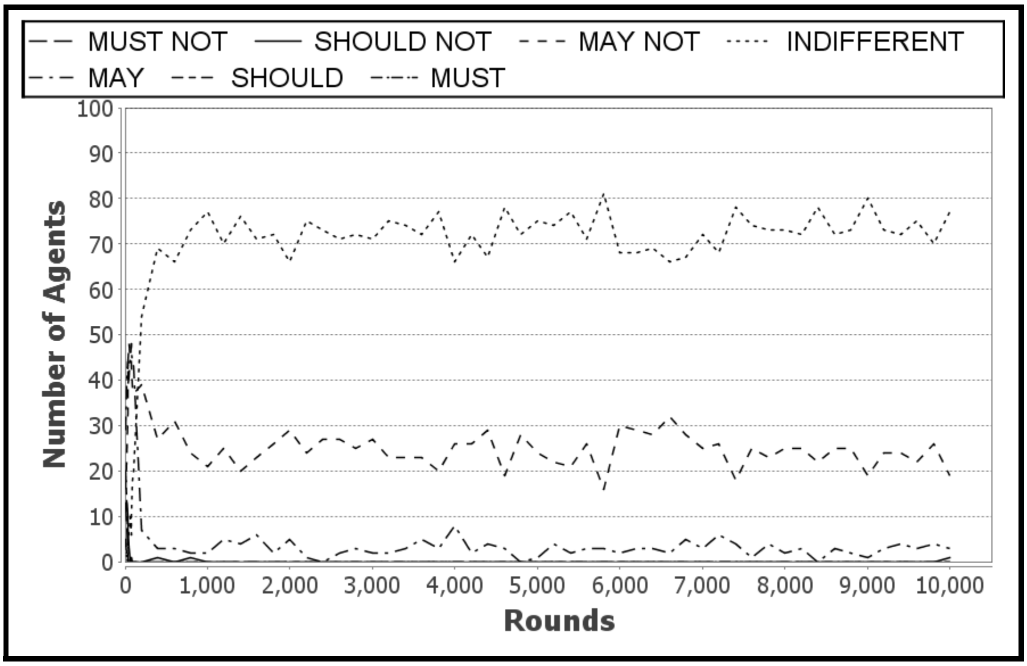
Figure 6.9: Role Integration, No Norm Enforcement, Symmetric Deontic Range, Rational Traders

action `trade fair`, we can observe the strongest reinforcement for the action-reaction combination <`trade fair` - `reward`> with a value of around 131 (associated with the inverted deontic *must not*), indicating a strong reinforcement of compliant trading (and rewarding). Alternative action-reaction combinations such as <`trade fair` - `retaliate`> have a comparatively low reinforcement (here around -6). However, since the symmetric deontic range spreads across the entire range of the agent's experience (from around -40 to around 134), the deontic centre lies at around 46.98. Consequently, for this configuration the calculated mean (around 20.43) is allocated to the compartment *may not*. This suggests that a symmetric deontic range configuration, particular when involving vastly different reinforcement levels, should not be applied without further refinement of the individual deontic compartment's configuration. For the remainder of this discussion, we thus concentrate on the zero-centred asymmetric deontic range configuration, which enforces an objective grounding for the deontic compartments (see discussion in Subsection 6.2.5).

Scenario 1 – Integrated Role Understanding, No Norm Enforcement, Zero-Centred Deontic Range Exploring the same scenario for rational traders with a zero-centred deontic range configuration, we can observe a different picture. For the role-integrated Maghribī case the derived deontic for the action `trade fair` resolves to a *may* (see Figure 6.10a). Complementing this, the action `withhold profit` (see Figure 6.10b) is considered neutral by around 70 to 80 percent of agents. A smaller fraction of 20 to 30 percent has a stronger understanding and considers withholding profit undesirable (*may not*), along with around 5 percent that consider cheating as an option (*may*). The reason for the relative moderate understanding of `withhold profit` lies in the relatively infrequent invocation of this action and thus poor reinforcement.



(a) Action 'trade fair'



(b) Action 'withhold profit'

Figure 6.10: Role Integration, No Norm Enforcement, Zero-Centred Deontic Range, Rational Traders

Activating opportunistic understanding, i.e. choosing extremal consequences, produces a different outcome as shown in Figure 6.12a and 6.12b. Around 90 percent of traders believe that they *must* trade fair; only around 10 percent believe they *should* trade fair. With respect to cheating (see Figure 6.12b) fractions of 40 percent oscillate around the belief that they *may* and *may not* withhold profits. This is complemented by 10 to 20 percent of indifferent traders. Around 10 percent have a stronger understanding, believing they *should not* withhold profits.

To clarify the comparatively weak normative understanding for cheating behaviour – similar to the case of rational agents (Figure 6.10b) – we show a situational extract for an opportunistic individual in Figure 6.11.

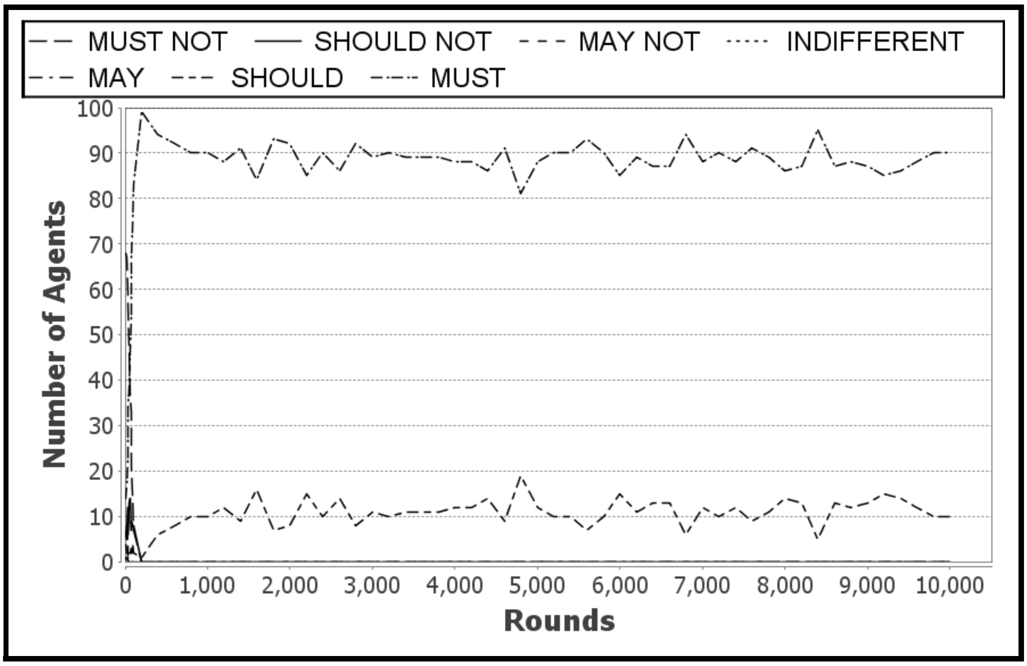
```
MUST NOT: -7.9449825
SHOULD NOT: to -3.9724913
MAY NOT: to -0.39724913
INDIFFERENT: to 2.8031073
MAY: to 28.031073
SHOULD: to 56.062145
MUST: 56.062145

L0: A=Maghribi, D=MUST (55.728825), I=trade fair, C=*,
 O=(
  (L1 (p: 0.68187976): A=*, D=SHOULD NOT (inv) (30.607388), I=don't reward, C=*, O=(null)) OR
  (L1 (p: 0.026679307): A=*, D=MUST (inv) (-8.683846), I=fire, C=*, O=(null)) OR
  (L1 (p: 0.2665222): A=*, D=MUST NOT (inv) (55.728825), I=reward, C=*, O=(null)) OR
  (L1 (p: 0.024918742): A=*, D=SHOULD (inv) (-4.7680025), I=retaliate, C=*, O=(null)))
L0: A=Maghribi, D=MAY (5.276766), I=withhold profit, C=*,
 O=(
  (L1 (p: 0.032447048): A=*, D=INDIFFERENT (inv) (0.022168726), I=don't reward, C=*, O=(null)) OR
  (L1 (p: 0.04731861): A=*, D=MAY (inv) (-2.7757525), I=retaliate, C=*, O=(null)) OR
  (L1 (p: 0.14781433): A=*, D=MAY NOT (inv) (5.276766), I=fire, C=*, O=(null)) OR
  (L1 (p: 0.77242): A=*, D=MAY (inv) (-0.8907324), I=reward, C=*, O=(null)))
```

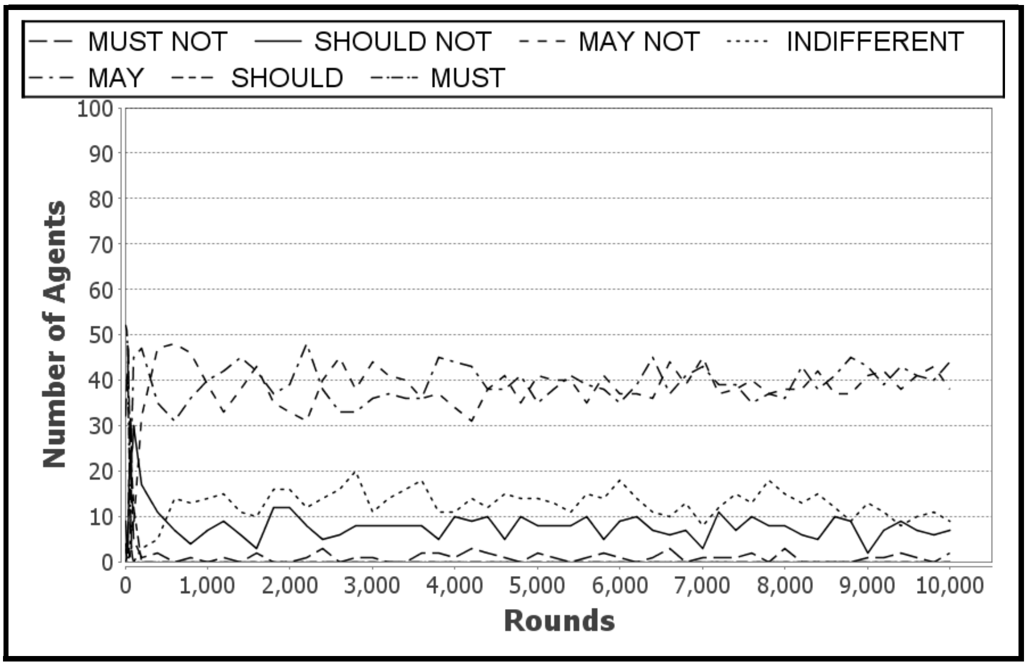
Figure 6.11: Situational Zero-Centred Deontic Range of Individual Opportunistic Agent

This extract shows an aspect that has considerable impact on the deontics evaluation, namely the reinforcement of given actions. Reflecting the notion of Bourdieu's "urgency of practice" (Bourdieu, 1981), action choice implies reinforcement and thus the development of an action understanding. Beyond the accessible and retraceable normative understanding, we can observe the comparatively weak reinforcement levels for the action `withhold profit`, indicating its infrequent invocation in comparison to the action `trade fair`. For the action `withhold profit` this leads to a relatively fuzzy normative understanding close to the centre of the deontic range, in contrast to the clear tendency of desirability for the action `trade fair`.

Another noteworthy aspect is the effect of the multi-perspective action reinforcement based on the integrated role conception. As such Maghribīs are motivated to act compliantly (`trade fair`) based on the prospect of being rewarded (`reward`: 55.72). This positive reinforcement from not being paid (`do not reward`: 30.60) appears counter-intuitive at first.



(a) Action ‘trade fair’



(b) Action ‘withhold profit’

Figure 6.12: Role Integration, No Norm Enforcement, Zero-Centred Deontic Range, Opportunistic Traders

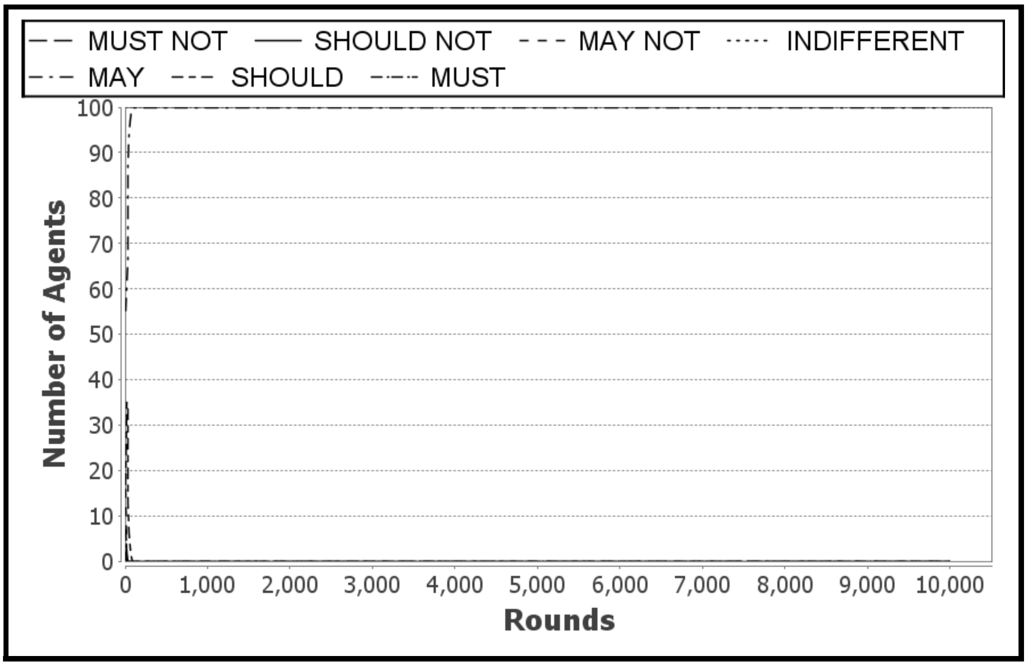
However, from the perspective of an investor, fair trading by operators without remunerating them is desirable. From his investor perspective the agent thus appears to advocate non-cooperative behaviour for this specific trace. Nevertheless, the integrated overall reinforcement for rewarding fair trading (from both investor and operator perspectives) outperforms the non-cooperative investor perspective, thus promoting the overall convergence to cooperative trading. We can find similar investor-biased traces in the reinforcements associated with `withhold profit`, such as the preference to fire cheating operators (`fire`: 5.27) and the negative association with rewarding cheating operators (`reward`: -0.89).

Scenario 2 – Integrated Role Understanding, Norm Enforcement, Zero-Centred Deontic Range Activating norm enforcement in addition to the integrated role conception, for rational traders,[33] we arrive at a crisp understanding that fair trading is the preferred action (see Figure 6.13a). Similar to the previous case, strong levels of compliance lead to limited reinforcement of violating behaviour, thus constraining the agent's opportunity to develop a clear understanding, an aspect that is reflected in the weak tendency towards rejection of cheating (see Figure 6.13b). From the behavioural perspective this leads to the suggestion that agents hardly shape normative understanding about cheating at all.

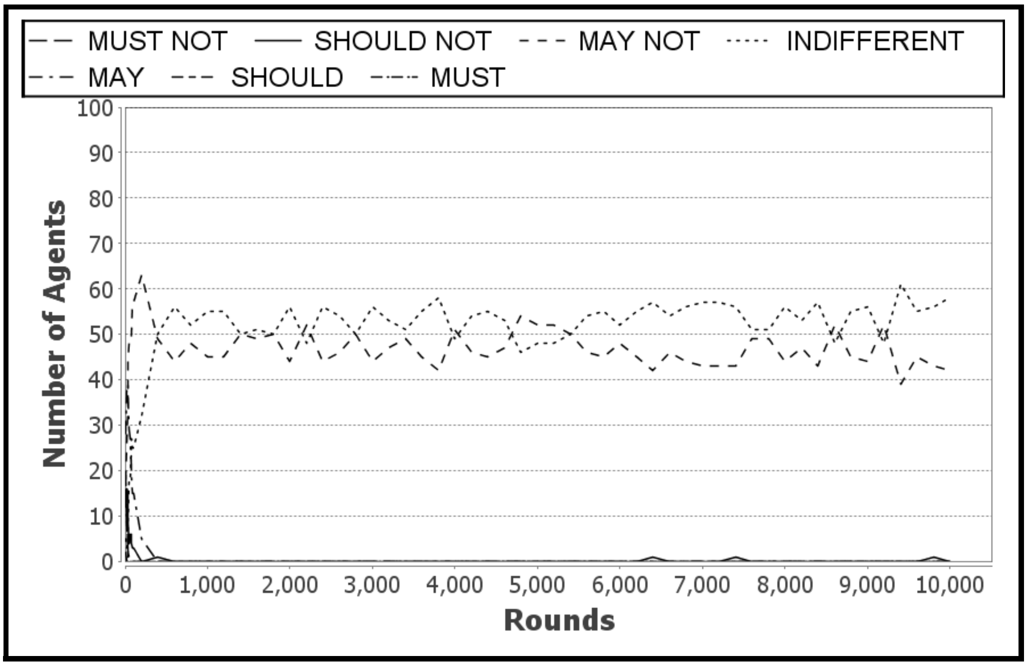
The activation of opportunistically determined normative understanding shows a nearly identical outcome for fair trading (Figure 6.14a). However, for withholding profit (Figure 6.14b) we can observe a more diverse understanding, spanning across all deontic compartments left of the deontic centre, with a minor exception of a minimal fraction who associate the deontic *may* with cheating behaviour. The reason for the diverse landscape lies in the interpretation of opportunism; each agent opts for the strongest reinforced consequence, with around 40 percent maintaining indifference, while 30 percent opt for *may not*. Based on their experiences and observation, around 20 percent have the understanding that they *should not* cheat. The remainder are split between between *must not* and *may*. Agents that opt for *may* generally made positive experience by being rewarded despite invoking cheating behaviour.

Recall that this scenario with its integrated role understanding and norm enforcement is the prototypical configuration for the Maghribī society. Looking at the overall results, independent from the individual strategies (rational and opportunistic) and concrete figures, we can say that agents show a strong tendency to interpret cooperative behaviour as desirable.

[33]Refer to Table 6.1 in Subsection 6.2.5 for an overview of the deontic determination strategies.

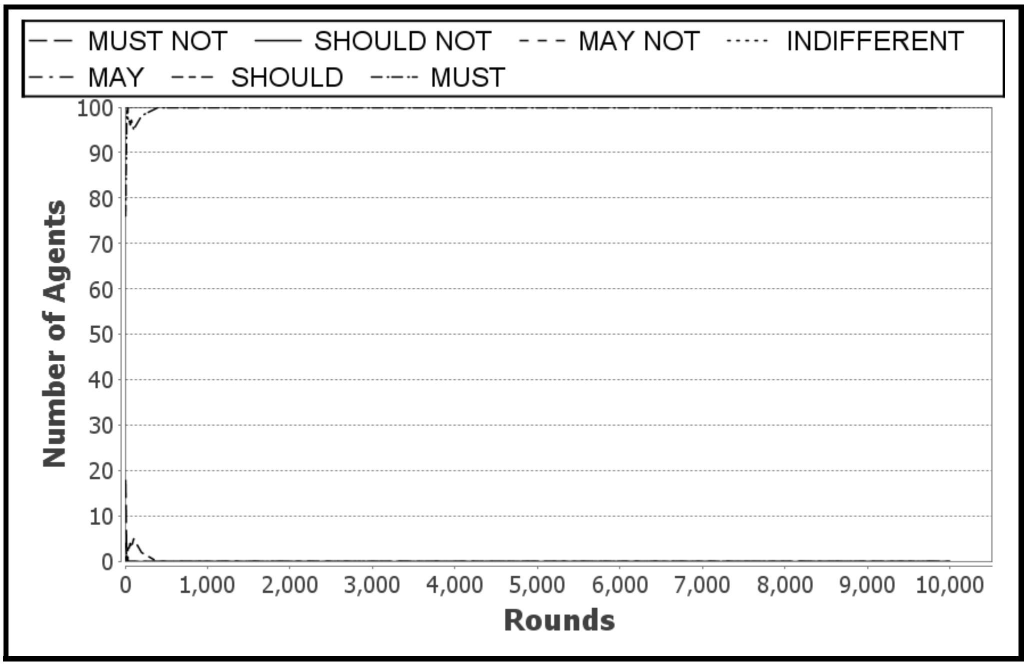


(a) Action 'trade fair'

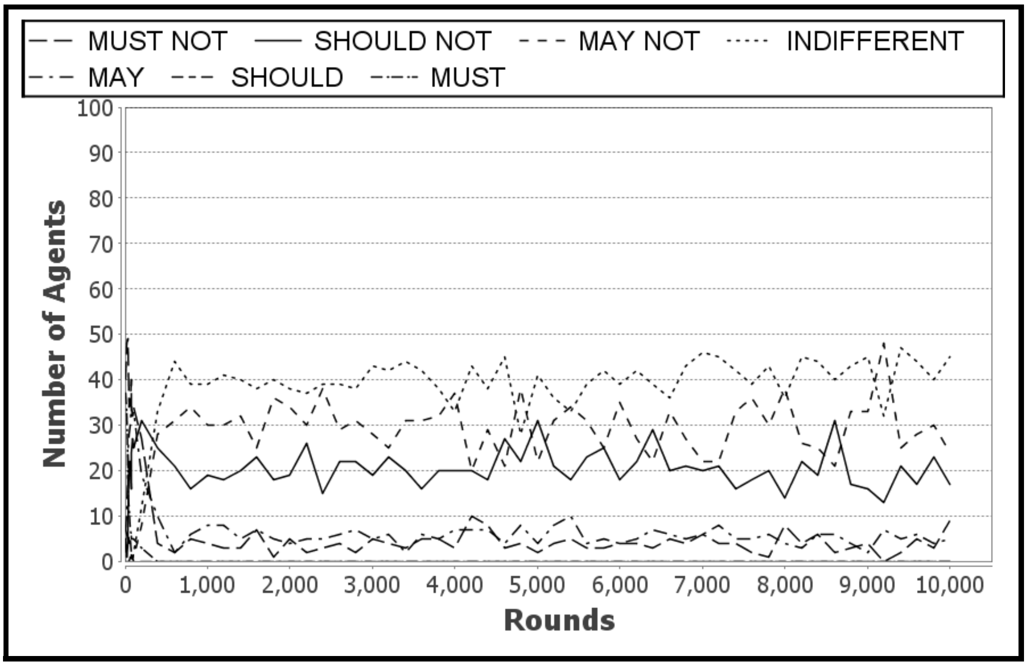


(b) Action 'withhold profit'

Figure 6.13: Role Integration, Norm Enforcement, Zero-Centred Deontic Range, Rational Traders



(a) Action ‘trade fair’



(b) Action ‘withhold profit’

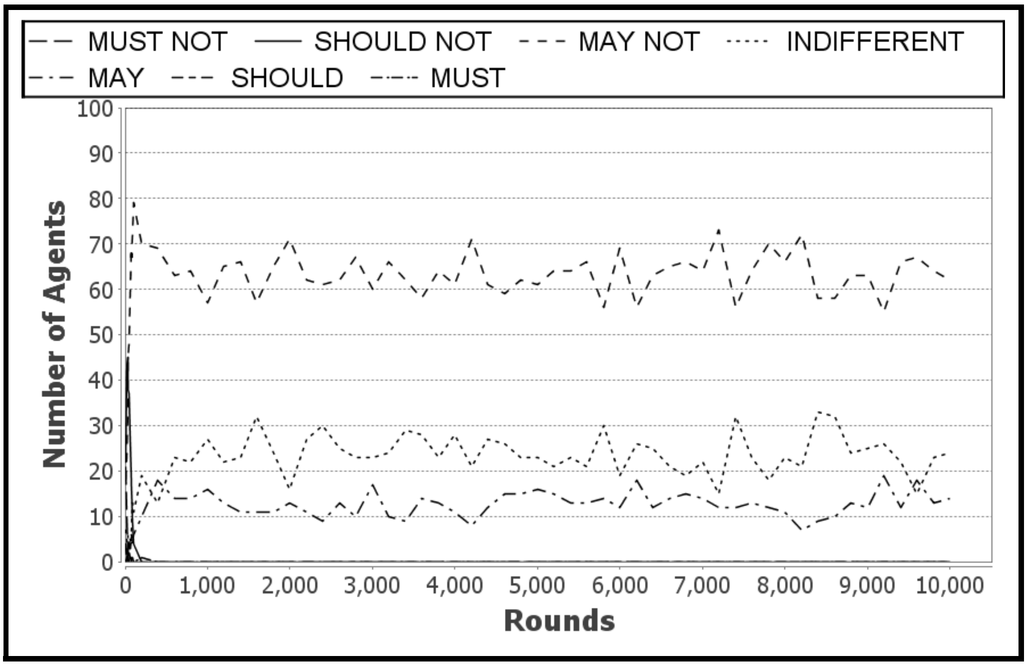
Figure 6.14: Role Integration, Norm Enforcement, Zero-Centred Deontic Range, Opportunistic Traders

Having explored emerging normative understanding for the integrated role concept, we now turn to the exploration of the stratified role configuration that, within the scope of this model, offers closer resemblance to the Genoese society in which traders either act as investors or operators.

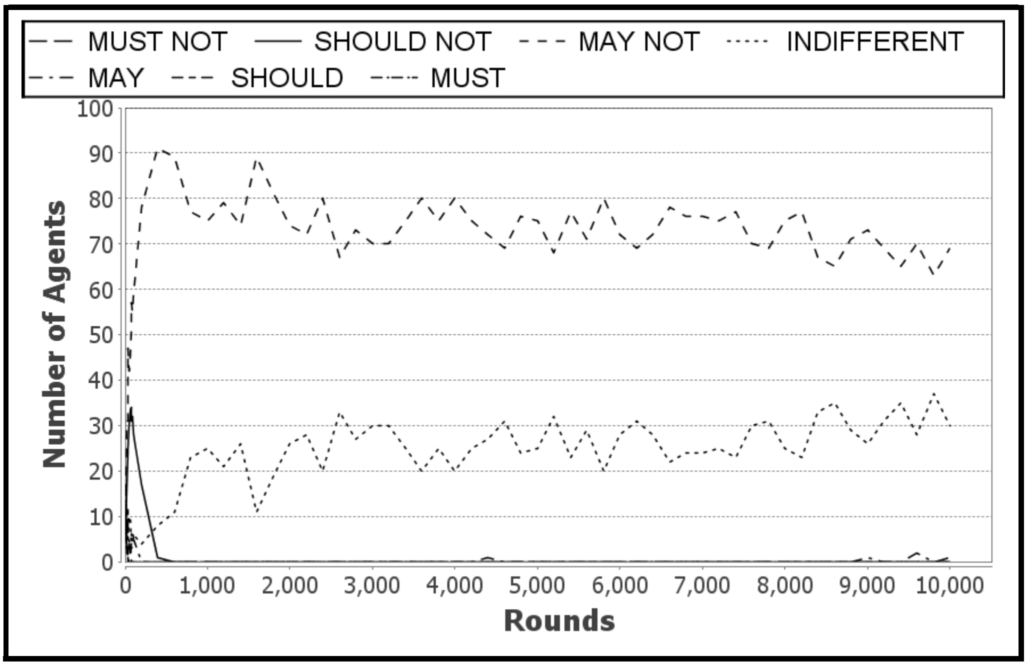
Scenario 3 – Differentiated Role Understanding, No Norm Enforcement, Zero-Centred Deontic Range Introducing role differentiation without additional norm enforcement, we arrive at a sharply contrasting picture. For the rational trader setup (see Figure 6.15a), a majority of around 60 to 70 percent categorises fair trading as *may not*. This is complemented by 10 to 20 percent of traders who believe that fair trading is desirable, along with an additional 20 to 30 percent who are indifferent about the desirable behaviour. With respect to withholding profit (see Figure 6.15b), a fraction of initially 80 percent associates the deontic *may not* with that action, and over time converges towards 70 percent of traders. This reduction is complemented by increasing numbers of agents that are indifferent about cheating behaviour.

We can explore the diverse and in parts contradicting normative understanding developed by operators when observing the independently developed normative understanding of investors. However, since investors have a larger repository of actions at their disposal, we opt for a representation that allows for the concurrent display of norm understanding. Since deontic terms as used in this context allow ordinal scaling, we use a Kiviat-inspired chart that shows the distribution of action associations across deontic compartments as shown in Figure 6.16. Each leg represents the fraction of investors that associate a given reaction with the deontic compartment.

Inspecting the distribution, we can observe a relatively broad distribution of traders that opt for *not* rewarding traders (with 0.35 opting for *may*, 0.19 opting for *should*, 0.45 opting for *must*, and 0.01 being *indifferent*). At the same time, 20 percent of investors believe they *should* reward merchants; the remaining 80 percent of investors believe they *must* reward them. Comparable fractions consider firing and retaliation mildly desirable (*may*). The results thus show that despite overwhelmingly non-compliant behaviour, a considerable fraction of both merchants and investors believe in cooperative trading, driving the diverse normative understanding identified for traders. The evaluation of larger numbers of actions highlights a central characteristic of Dynamic Deontics. Since associations with actions (or reactions in this case) are built based on reinforcement, seemingly conflicting normative understanding can emerge, such as highlighted in the case of 80 percent considering rewarding an obligation, with fractions of similar size (0.78 and 0.79) considering retaliation as mildly appropriate (*may*). However, the development of associations with actions is not logically related, i.e. an increasingly positive association with a given action, such as `do not reward`, does not automatically reduce valuation with another seemingly conflicting action `reward`. Instead, individual actions' associations are indirectly integrated based on subsiding reinforcements, which reduces the salience (Cortell and Davis Jr., 2000) of a given normative understanding over time. Situationally this can lead to seemingly conflicting understanding as showcased above, an aspect we deem realistic for continuously adapting normative



(a) Action ‘trade fair’



(b) Action ‘withhold profit’

Figure 6.15: Role Differentiation, No Norm Enforcement, Zero-Centred Deontic Range, Rational Traders

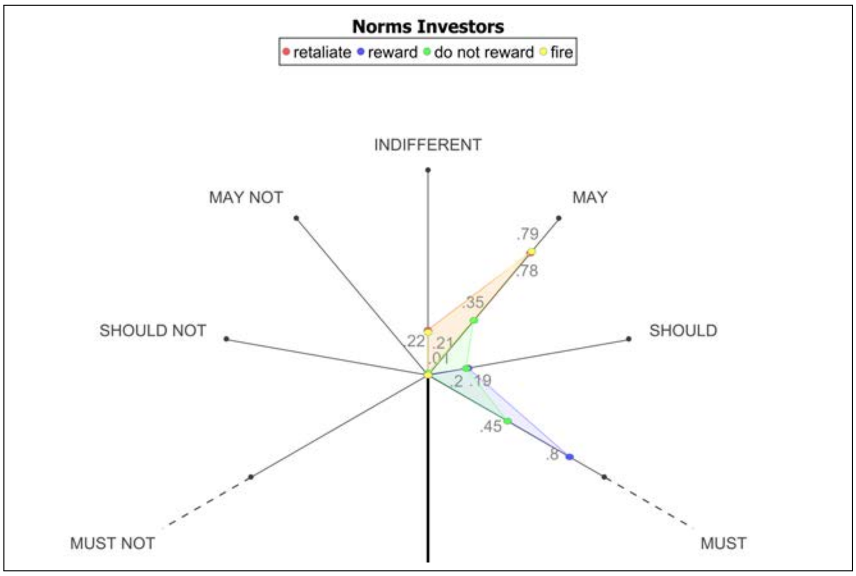


Figure 6.16: Distribution of Actions across Deontic Compartments for Investors

understanding which Dynamic Deontics intend to reflect.

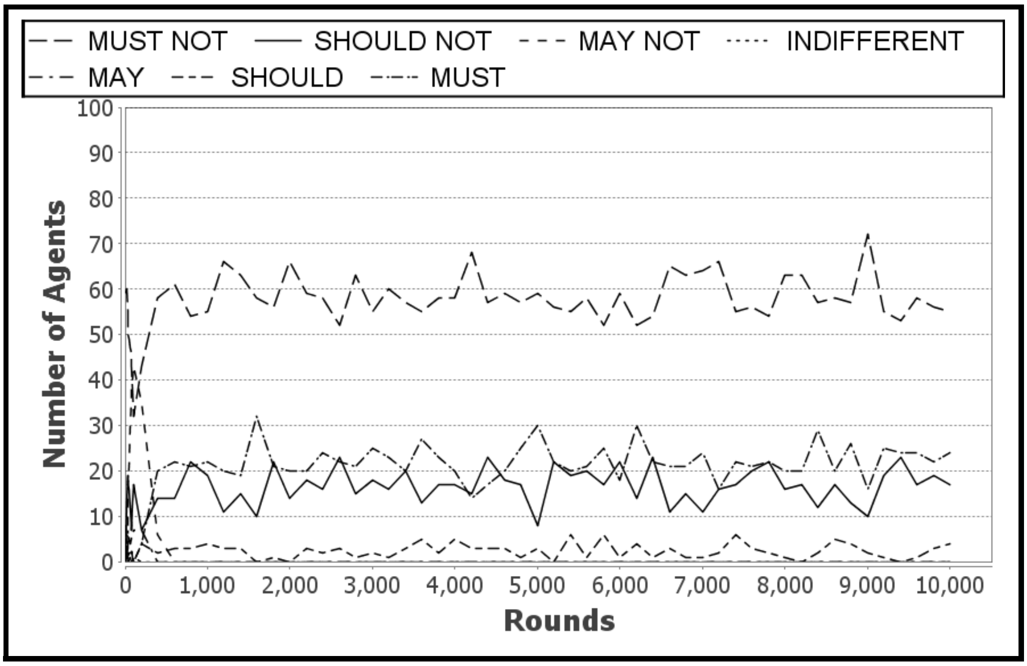
Returning to the operator perspective, activating the opportunistic trader setup (see Figure 6.17a), we can observe that around 60 percent of traders see fair trading as a *must not*, with around 20 percent opting for *must* and slightly less for *should not*. Less than 5 percent believe they *should* trade fair. As observed before, the opportunistic setup offers a more differentiated picture, since it emphasises extremal behavioural expectations. In the case of `withhold profit` (see Figure 6.17b) between 80 and 90 percent favour *may not*, while the remainder is indifferent.[34]

Overall, whether concentrating on a rational or opportunistic aggregation strategy, these observations suggest that cooperation and fair trading is not desirable, an aspect that supports the suggestion that a differentiated role understanding could have shaped individuals' experience, supporting the emergence of diverging normative understanding. For an integrated role conceptualisation, as shown before, the shared experience in an otherwise unchanged scenario produces a predominantly cooperative normative understanding as long as individuals perform different roles at the same time.

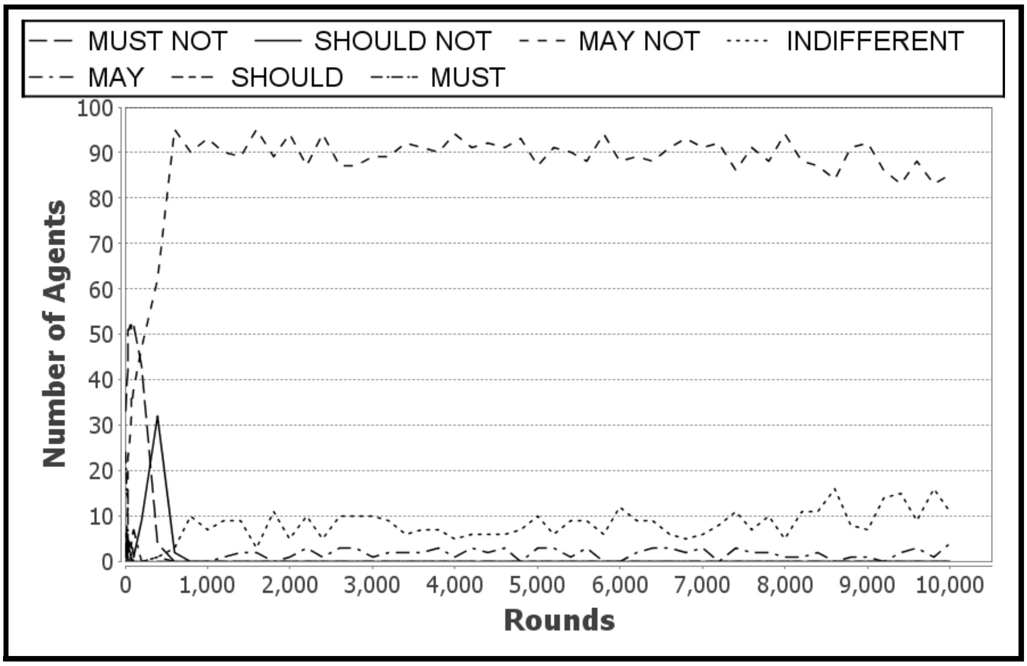
For the purpose of completing the various configuration combinations, we further explore the differentiated role understanding in combination with normative enforcement, suggesting that third party investors reacted to observed behaviour based on their own experience, thus reinforcing the acting individual's feedback (since it could expect multiple reactions to its action). Although private-order enforcement among Genoese is documented (see Subsection

[34] Although the chart may suggest further convergence, the values stabilise within the shown boundaries.

4.2.1), its level of enforcement was far from comparable with the systematic enforcement the Maghribī traders employed (see Subsection 3.2.2).



(a) Action 'trade fair'



(b) Action 'withhold profit'

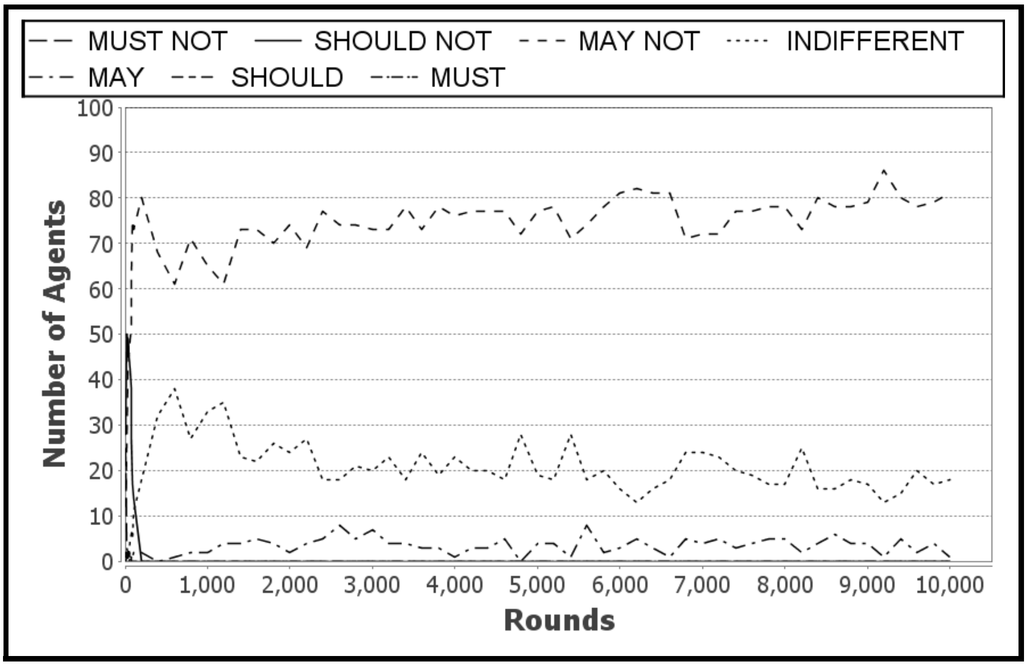
Figure 6.17: Role Differentiation, No Norm Enforcement, Zero-Centred Deontic Range, Opportunistic Traders

Scenario 4 – Differentiated Role Understanding, Norm Enforcement, Zero-Centred Deontic Range In the final case, for the action `trade fair` (see Figure 6.18a), we can observe an increasing regard for deterrence with a continuous tendency towards *may not*. Looking at the action `withhold profit` (Figure 6.18b), in contrast, we can observe a shift from an initial emphasis on *may not*, complemented by a diversifying perspective, with an increasing fraction of indifferent traders as well as a smaller fraction of individuals that consider cheating desirable (*may*). This observation leaves us with an ambivalent impression, since trading appears increasingly undesirable. However, this reaction obscures the more complex nature of the interdependencies. In our simulation we observe an increased bias towards enforcement: for every action agents can expect multiple reactions of diverse nature – one from his trade partners, a second one from an observing outsider. With negative consequences dominating the reactions, agents develop a negative understanding about the desirability of fair trading (increasing *may not* in Figure 6.18a). Along with this, traders' views on cheating decrease from an initially dominantly rejecting perspective towards a stronger emphasis on interpreting withholding profit as desirable (i.e. increasing fractions for compartments *indifferent* and *may* at the expense of *may not* in Figure 6.18b).

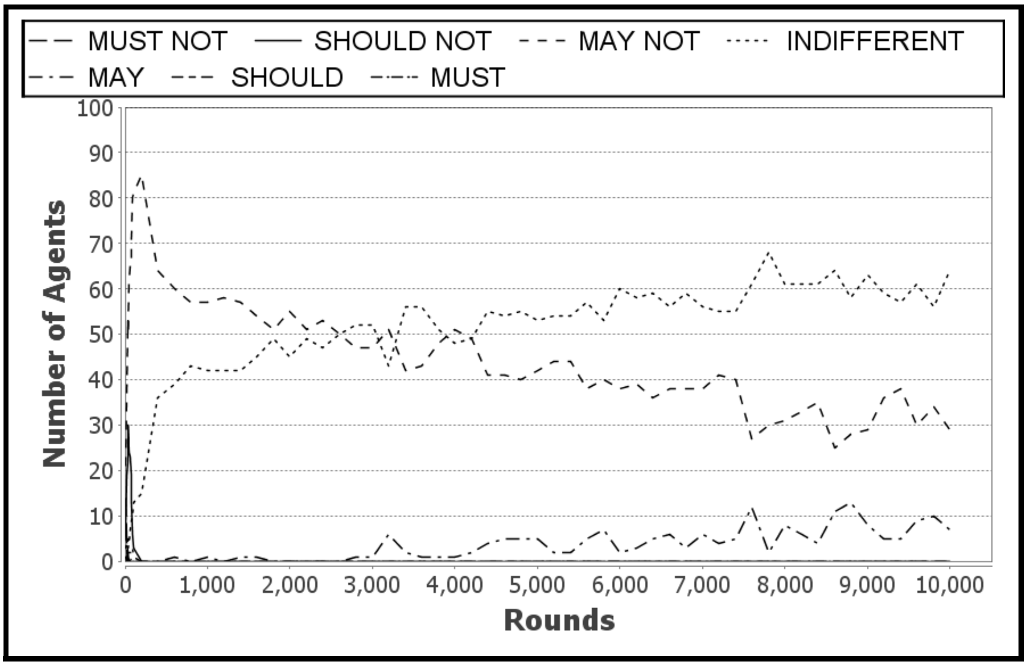
As with the previous scenarios, the case of opportunistic normative understanding provides a more differentiated picture. Around 50 percent of all traders associate compliant behaviour with *must not* (see Figure 6.19a). A lesser fraction of around 30 percent associates *may not* with fair trading, while comparable fractions of 10 percent favour *should not* and *may* respectively.

Complementarily, withholding profits shows increasing attraction, with decreasing numbers of agents for *may not* and a wide distribution among all other deontics, and fractions between 10 and 20 percent maintaining *should not* and *must not*, along with increasing fractions for *may* and *indifferent* understanding. The association of cheating with *must* oscillates between 5 and 10 percent.

Overall, the activation of norm enforcement results in a more moderate (i.e. less rejecting) normative understanding for fair trading, compared to the previous scenario that showed higher levels of rejection of fair trading (*must not* at around 60 percent and *should not* as well as *may not* at around 20 percent in Figure 6.17a, compared to 50 percent of *must not*, around 10 percent for *should not*, and around 30 percent for *may not* in Figure 6.19a).

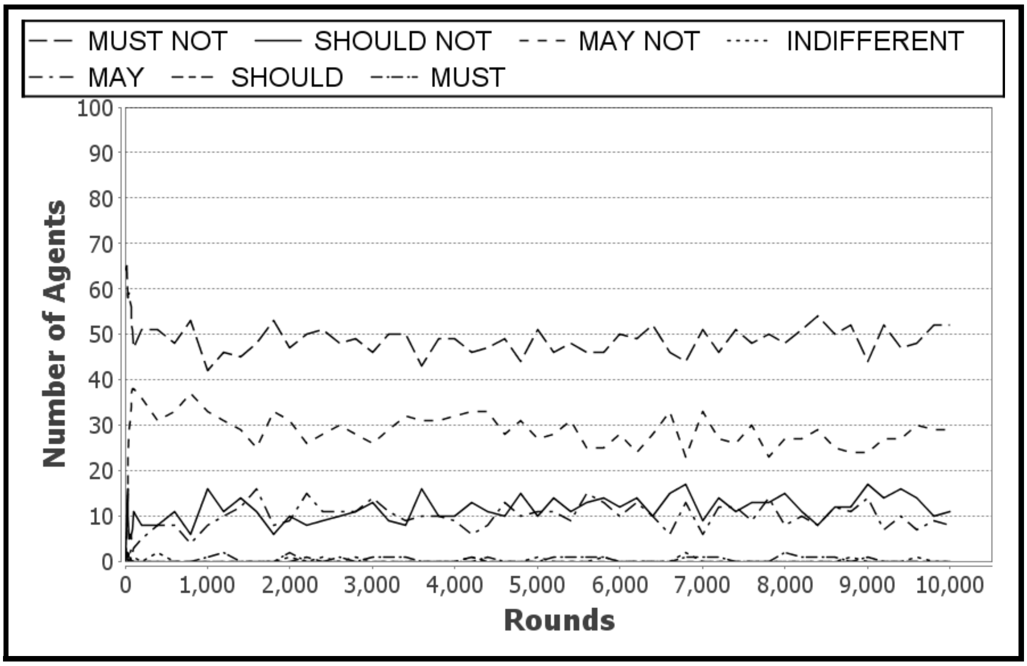


(a) Action 'trade fair'

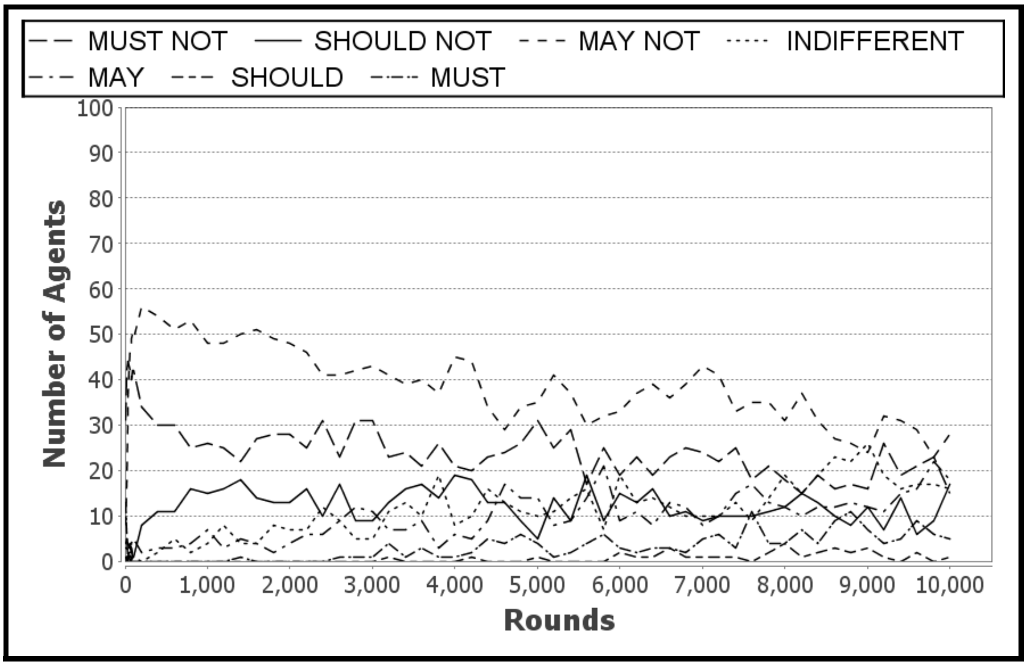


(b) Action 'withhold profit'

Figure 6.18: Role Differentiation, Norm Enforcement, Zero-Centred Deontic Range, Rational Traders



(a) Action ‘trade fair’



(b) Action ‘withhold profit’

Figure 6.19: Role Differentiation, Norm Enforcement, Zero-Centred Deontic Range, Opportunistic Traders

Looking at the investor side of things (Figure 6.20), the activation of norm enforcement leads to a revised understanding. Compared to the previously wider spread across different reactions, the choice of the reactions `fire` (previously 0.79, now 0.33 for *may*) and `retaliate` (previously 0.78, now 0.36 for *may*) has been concentrated in the compartment *should* (now 0.63 and 0.64 respectively). Likewise the reaction `do not reward` shifted its concentration toward *should* (previously 0.35 for *may* and 0.45 for *must*, now 0.59 for *should* and 0.39 for *must*). Although the results appear counter-intuitive at first – since increased enforcement should drive compliance –, the adverse effect of enforcement on the enforcer is the feedback from the sanction which he applies in addition to the regular trade interaction; as much as the operator can expect two reactions for a given action, the enforcer internalises feedback for those additional imposed reactions. Effectively, the enforcement of sanctions increases their salience on the part of the enforcer, leading to a shift of the deontic distribution.

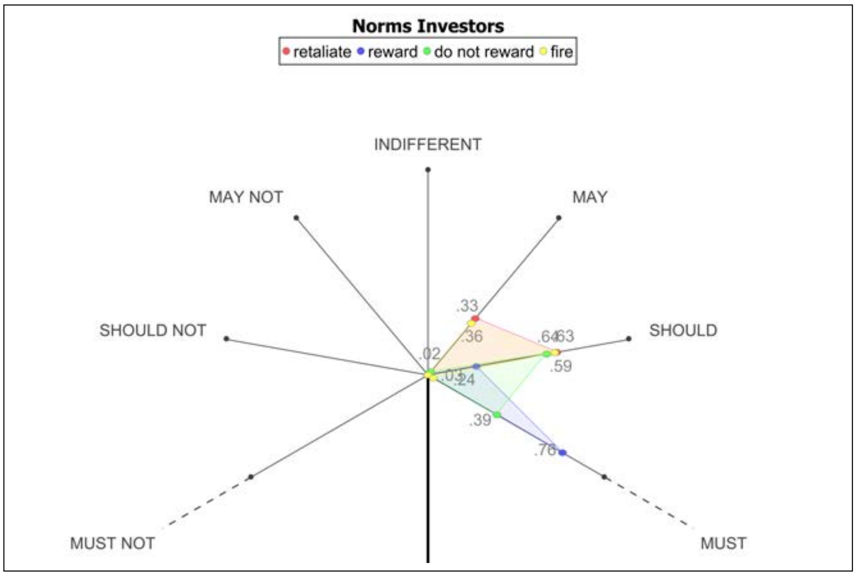


Figure 6.20: Distribution of Actions across Deontic Compartments for Investors with Activated Norm Enforcement

A statistical evaluation of the results for 30 simulation runs per configuration is provided in Appendix E. The results support the essential observations which we described in the form of an exemplary discussion of individual simulation runs. In order to test the sensitivity of the model against changing numbers of agents, we additionally performed the configuration runs with twofold agent numbers. The results for the default configuration (100 Maghribīs; 200 Genoese) and twofold numbers (200 Maghribīs; 400 Genoese) are captured in Tables E.1 and E.2. Comparing the relative distribution of actions across the deontic compartments for both actions (see Tables E.5 and E.6), the impact of twofold numbers of agents on the results is insignificant.[35] For a detailed discussion refer to Appendix E.

6.3.4 Discussion

In this section we have introduced experiments that offer an abstract representation of the trader scenario, in which operators determine their behaviour based on experience and social learning. To reflect the different societal configurations – which we believe could have been influential in the differing cooperation outcomes –, scenarios are differentiated by the nature of role configurations (either integrated or stratified) and the extent of enforcement by third parties. For the operationalisation we utilise the concept of Dynamic Deontics, introduced in Section 6.2. The concept facilitates an accessible representation of an emerging norm understanding relying on a purely behavioural perspective. The experiments described in this section thus likewise explore the hypothesis of interest, while at the same time they provide insights about the usefulness of different configurations of the Dynamic Deontics concept itself.

For the different trader societies we find that a shared experience from a role-integrated perspective, as reflected in Scenario 1 and 2, supports cooperative behaviour if we assume that individuals have an integrated learning mechanism that internalises experience from all roles they play. Activating norm enforcement, thus providing additional bias from the perspective of the reacting investor, increases the compliant understanding.

Separating different roles into investor and operator, as done in Scenarios 3 and 4, leads to a singular perspective on the trade scenario, selfishly reinforcing behaviour that is beneficial from the perspective of either role without producing an integrated view. However, activating norm understanding for the Genoese perspective does not lead to dominantly cooperative understanding; it does not rule out cheating behaviour but leads to a stronger concentration on moderate reactions (compare Figure 6.16 and 6.20).

Under the given, and in our view, conservative assumptions, the results suggest that the nature of role understanding could have promoted different cooperation outcomes in both societies. This aspect has been neglected in previous analyses (see Subsection 4.1.2) and thus extends the exploration of the Maghribī Traders Coalition from a behavioural perspective.

The model relies on a set of core assumptions. Those include the assumption of a unified experiential learning mechanism that considers all actions an individual takes irrespective of the roles it plays. We further expect individuals to be social learners that internalise observations of third-party interactions. Moreover, the outcomes rely on the intuitions expressed

[35]For all comparisons, the Mann-Whitney-Wilcoxon test (confidence interval of 0.95) reports p values of at least 0.503.

in the action-reaction feedbacks associated with different combinations, which reflect the anecdotal accounts that are at the basis of this thesis. In saying this, the current experiments have an arguably more limited documented grounding compared to the earlier experiments (see Subsections 4.2.1 and 4.2.2). However, the specified feedback values have been chosen conservatively, especially with respect to the moderate feedback for compliant behaviour (+1 for either party), and in the light of considerable monetary pay-offs for compliant trade in the historical case.[36] The limited literature base prohibits more refined specifications beyond the intuitions expressed in this configuration.

For the infrastructural aspects, here the Dynamic Deontics concept, we explored a set of configuration variations, including the symmetric compartmentalisation as well as the zero-centred asymmetric compartmentalisation. The symmetric configuration offers a representation of a subjective normative understanding without a shared reference point within the society. However, in simulations this puristic representation offers problems with respect to plausibility as explored in the beginning of the results subsection (Subsection 6.3.3). Since our simulation scenario relies on an objectified representation of feedback (with zero marking the neutral reference point), the zero-centred configuration offers retraceable simulation outcomes, since its valuation is aligned with the applied feedback specification. A further aspect includes the method by which overall deontic compartments (and thus terms) are derived. We introduced a set of different strategies by which feedback for individual action-reaction combinations can be aggregated by actions. In our experiments we explored this for the *rational* and *opportunistic* strategies (see Subsection 6.2.5). The rational strategy determines the overall deontic value based on the mean of individual action-reaction combinations. The opportunistic strategy, in contrast, associates the deontic value with extremal experiences. Both strategies offer different insights about the normative understanding, with the rational strategy offering moderate conservative outcomes, while the opportunistic strategy provides more diverse extremal perspectives. However, as supported by the statistical evaluation (see Appendix E), the variation of the aggregation strategy did not alter the general simulation findings (dominating cooperation among Maghribīs; defection among Genoese).

We ignored alternative strategies, such as the pessimistic and optimistic perspective, since both introduce a uniform bias towards prohibitions or obligations (*must not*s and *must*s). However, the additional strategies offer potential for the representation of different agent personalities using Dynamic Deontics, an aspect that is beyond the scope of this work.

From a sociological perspective the model reflects the effects of differentiated world

[36] Recall that Genoese trade typically resulted in profits between 20 and 110 percent (van Doosselaere (2009), see Subsection 3.3.2).

views in the Durkheimian understanding (see Section 4.3), with Maghribīs showcasing stronger alignment compared to the Genoese case, leading to an informal cooperative outcome. The provided evaluation mechanisms reflect this, but have limitations with respect their expressiveness and accessibility. Dynamic Deontics facilitate a dynamic view on the emerging normative understanding, the observation of which is supported by the assignment of deontic compartments. Although the time-series representation, as used in this work, highlights the progression of understanding over time, we note that with increasing numbers of deontic compartments, the analysis loses its clarity, especially for confusing outcomes with a wide spread across deontic compartments such as shown in Figure 6.19b. A challenge to the time-series representation is the limited focus on given actions, which challenges its use for the exploration of the investor perspective. This produces a trade-off: clarity of representation decreases with an increasing number of deontic compartments, while a more fine-grained compartmentalisation facilitates a refined interpretation of the agents' normative understanding. For large numbers of compartments, only statistical analysis is sufficiently insightful from a macro-level, but limits the inspection of individual agents.

From the perspective of an experimenter, it is thus desirable to offer an analytical representation that allows inspection on different social levels, while, at the same time, provide a better understanding about the detailed characteristics of the individual deontic compartments. Relating this to the problem of interest, can we provide an alternative representation that is visually accessible but also provides us with measures of 'alignment' that describe a society's world view? We will make the introduction of an accessible representation based on novel analytical instruments a subject of the following chapter.

7 Analysing Emerging Normative Understanding

7.1 Motivation

At this stage we have explored how behavioural processes could have shaped a differing normative understanding among and within societies. Evaluation based on statistical analysis, though doubtlessly necessary to support experimental outcomes, generally concentrates on the macro-level, aggregating and generalising outcome measures. But especially when addressing complex social problems from an explorative perspective, an experimenter may be inclined not only to get an aggregate overview but also to dynamically inspect individual properties of interest (such as individual deontic compartments). Doing so on different aggregation levels, or levels of social analysis, is particularly relevant for norms (and institutions in general), since those are inherently associated with the social structures (such as social groups or societies) that underlie them.

We thus propose a tool that abstracts from the observation on action level (as done in the previous Chapter 6) and augments the previous simulation scenario with visual explorative capabilities that allow us to inspect the overall normative understanding from a more refined perspective, addressing questions with respect to *width* and *positioning* of deontic compartments along the deontic range. We further want to develop a measurable understanding of *how aligned individuals' normative understanding is* – both for specific deontic compartments as well as across different levels of social aggregation. More important than the numeric analysis of such aspects is their accessible representation that supports the understanding and communication of experimental findings.

The extension we are proposing thus facilitates

- the *inspection and analysis of normative understanding*
  - *across deontic compartments*, but also
  - *within deontic compartments*, as well as
- *analysis across different levels of social organisation*, namely micro-, meso- and macro-levels.

Beyond the analysis on greater level of detail, the outlined objectives reflect representations that have conceptual meaning from a sociological standpoint, such as the abstractly expressed 'world view' referred to in Section 4.3 and the previous chapter (Chapter 6). The solution proposed in the following has been published in Frantz et al. (2014d) and is presented here in an extended form.

7.2 Applying Interval-Type 2 Fuzzy Sets to Social Analysis

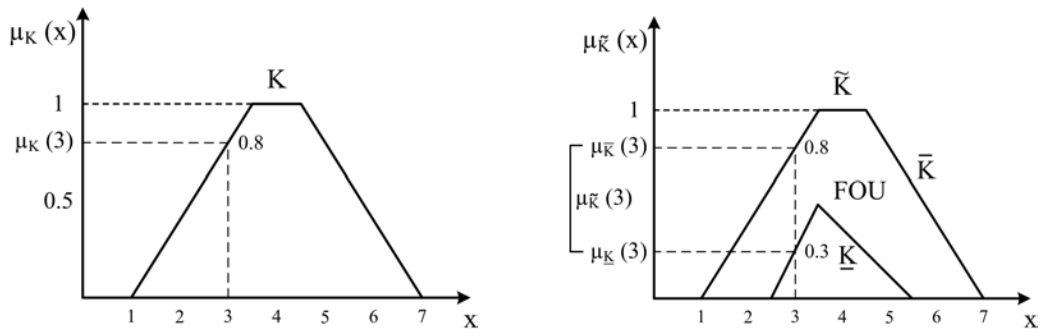
Inasmuch as Dynamic Deontics facilitate a refined representation of normative understanding and are able to reflect fluid and viscous norms (Kinzig et al., 2013), the deontic compartments are likewise an abstraction from the continuity of normative understanding, reducing it to commonly shared categories. This may appear as an oversimplification at first. However, human language, such as the labels we attached to the individual deontic compartments in Section 6.2, applies similar mechanisms to communicate seemingly uniform conceptualisations of terms, resulting in ambiguity, vagueness and context-dependence – leading Chomsky to famously suggest that language is not designed for communication (Chomsky, 2010). Emphasising the psychological dimension, Szalay and Deese (1978) suggest the importance of shared subjective experience to develop a shared lexicon. Supporting the experiential perspective, Bloomfield (1933) highlights the primacy of inferring words' meanings from use, as opposed to explicit definition in dictionaries. In the context of inter-cultural communication, Szalay (1981) further suggests that the correct interpretation of a message ultimately depends on the extent to which sender and receiver share experience. Applying this understanding to the context of institutions, particularly to informal ones that are transmitted by language as well as action (such as folkways or norms), we can utilise representation mechanisms from the area of Soft Computing to foster a more meaningful interpretation. A particular application field is the area of Computing with Words (Zadeh, 1975a, 1999; Mendel, 2003), which recognises the imprecision of words and instead uses mechanisms such as fuzzy sets to operate with imprecise information, an aspect we briefly introduced in Subsection 2.3.4.

Since norms arise from complexity of human interaction (or, in our case, artificial agents), we believe that the underlying concepts can be equally applied to normative understanding, whose complexity – similar if not greater compared to human language – exists in conflict with precise representation, as expressed in Zadeh's *Incompatibility Principle* (Zadeh, 1973). We adapt the notion of fuzzy sets for the purpose of analysing the norm understanding agents develop using Dynamic Deontics.

Recalling the essential characteristics of Fuzzy Sets introduced in Subsection 2.3.4, their essential difference from conventional *crisp* sets is that associations with a given fuzzy set are characterised by a *degree of membership* ranging between 0 and 1, whereas crisp sets determine membership in a Boolean fashion. Translating this concept to our problem, we can use fuzzy sets to support the determination of a unified societal representation of deontic terms. Reusing the previously introduced Figures 7.1a and 7.1b, a fuzzy set enables us to describe what *should* means with respect to its allocation on the continuous deontic range for individual agents, groups, or the society in its entirety. For the exemplified (type-1) fuzzy set in Figure 7.1a, with the entire deontic range as the domain and the fuzzy set $\widetilde{K}$ representing the deontic *should*, we can identify that for an aggregation of interest (e.g. a group) *should* ranges from 1 to 7 along the deontic range. With the membership function describing the certainty, we can derive that for values between 3.5 and 4.5 the understanding of *should* is unambiguous across a given group – resolving to a degree of membership of 1. At the same time the understanding becomes increasingly fuzzy in the shoulder areas of the fuzzy set, reducing the certainty to which an input value, such as 3, is understood as *should* to a degree of membership of 0.8.

However, the use of this characterisation of Fuzzy Sets (known as "Type-1 Fuzzy Sets" (T1FS)) limits the interpretation for our context, the determination of aligned normative understanding, since the only measure of alignment is the gradient of the shoulders of a given fuzzy set. Beyond the philosophical challenge that the membership function for the description of a *fuzzy* set itself is *crisp*,[1] it prevents the inspection of how aligned individuals

[1] We discussed this paradoxon in Subsection 2.3.4.



(a) Example of Type-1 Fuzzy Set (b) Example of Type-2 Fuzzy Set

Figure adapted from Frantz et al. (2014d)

Figure 7.1: Type-1 vs. Type-2 Fuzzy Sets

are within those shoulders, i.e. how certain the membership function itself is.

For this purpose we apply Interval Type-2 Fuzzy Sets (see Figure 7.1b), which introduce a second membership function, thus describing the overall fuzzy set based on an upper and lower limit of certainty. This replaces the crisp representation of membership with more refined boundaries that specify the minimal certainty of membership of an input value (lower membership function), and thus essentially add precision to the seemingly well-defined macro representation of membership offered by the upper membership function. With respect to our intent to facilitate a multi-level representation of normative understanding, this offers a significant benefit. By introducing upper and lower levels of certainty we can not only derive an aggregate measure of certainty, but also integrate it with the micro-perspective. The *minimal common ground* of normative understanding for a given deontic range value (resolving to 0.3 for input value 3 in the example in Figure 7.1b) explains how representative the derived macro-level representation of normative understanding is with respect to individual normative understanding. This allows for a transparent fuzzy set generation process but also provides insight about the micro-level alignment of normative understanding represented by the Footprint of Uncertainty (FOU). We will revisit this aspect in the context of the operationalisation.[2]

7.3 Operationalisation

For the purpose of operationalisation, Dynamic Deontics, as used in our work, provide a matching representation, since deontic terms are expressed as intervals on the deontic range. Since our approach bears analogy with the principle of Computing with Words, we borrow

[2]For a more detailed general introduction to Fuzzy Sets refer to Subsection 2.3.4.

elements from Liu and Mendel's Interval Approach methodology (Liu and Mendel, 2008b) for the generation of IT2FS. Their methodology offers a suitable starting point, since it defines data pre-processing steps as well as membership function generation steps. Since their approach deals with inconsistency and erroneous human input, we can adapt it for the use with software agents that share some of those problems. For this purpose we constructed a Java-based software module, the components of which are shown in Figure 7.2.

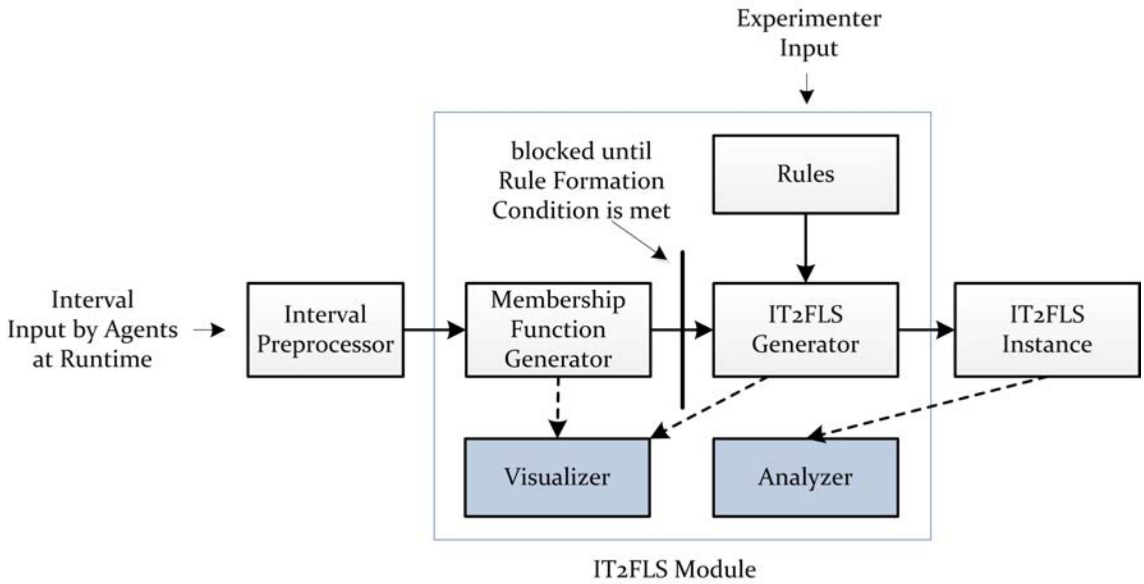


Figure 7.2: IT2FLS Software Module

The module builds on elements from Liu and Mendel's Matlab implementation and extends it with a software interface for the runtime invocation by software agents. We introduced refined preprocessing steps (which we introduce in this chapter) along with visual inspection capabilities that allow the experimenter to retrace the membership function generation process. Beyond this, the software module allows an experimenter – or an agent for that matter – to specify rules that activate a consequent for combinations of fuzzy sets associated with given input values. A related application that uses consensus-based decision-making to specify rules for gradual sanctioning has been proposed by Frantz et al. (2014c). However, this work concentrates on the provision of an accessible representation of shared normative understanding – in this case generated fuzzy sets are not used for decision-making on the part of the agents.

Interval Preprocessing Integrating the module with our simulation scenario, agents feed their situational deontic compartments, with each compartment specification consisting of the tuple <`deontic term, left boundary, right boundary`> into the interval preprocessor which performs basic consistency checks such as assuring that all agents' intervals have been received without duplicates. An essential requirement for the generation of membership functions is a fundamental overlap of input intervals, so that a common understanding can be inferred in the first place. Particularly for input by autonomous entities, such as humans or software agents, basic data preprocessing is necessary to identify outliers in order to arrive at a subset of sufficiently overlapping intervals that can be meaningfully integrated. Let us retrace the membership function generation process with respect to an example in Figure 7.3 (with membership functions shown in red and green colour).

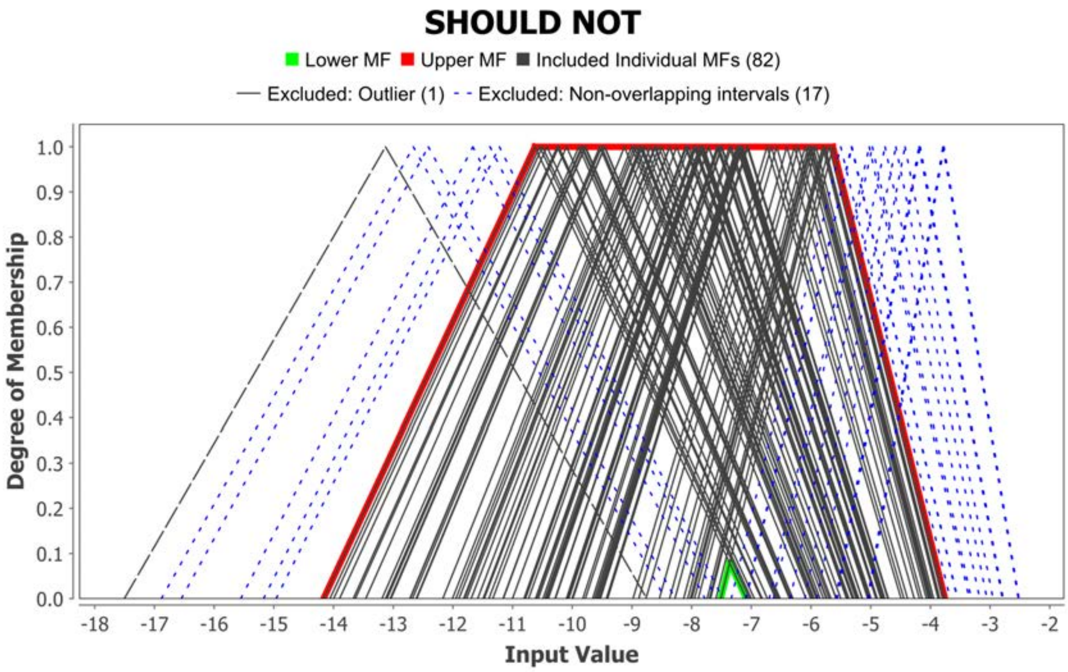


Figure 7.3: Interval Preprocessing and Membership Function Generation

To extract the essential set of intervals to establish a lower membership function, as shown for the extreme example in Figure 7.3, Liu and Mendel (2008b) apply a plausibility check and a set of statistical preprocessing steps. The first step includes a basic plausibility check that filters invalid intervals, such as intervals whose left and right boundaries are inverted. This problem, which is likely caused by invalid coding by human subjects, is of limited relevance for our context (with software agents as subjects), since we can control it by design. The second step involves the identification and removal of outliers. For this purpose Liu and Mendel (2008b) suggest the exclusion of all intervals *i* either of whose boundaries lie to the left or right of a zone that extends by a factor of 1.5 to either side of the interquartile range (IQR) (the range of the central 50% of all intervals), i.e.

$$i > Q(0.75) + 1.5 * IQR; i < Q(0.25) - 1.5 * IQR \tag{7.1}$$

The introduced visualiser makes the preprocessing transparent as exemplified for the deontic term *should not* in Figure 7.3. In our example, only one *Outlier* exists outside the zone of 1.5 to the left side of the interquartile range.

A further step includes the identification of intervals that lie outside a tolerance zone of a subset of all remaining intervals, which is statistically determined based on a given confidence level 1-γ that a given proportion 1-α of all intervals is included (Liu and Mendel, 2008b; Walpole et al., 2012). Values of 0.05 for γ and α thus express 95% confidence that 95% of all intervals are included, with the remainder being discarded from membership function generation. This approach rests on the assumption that the input intervals approximate a normal distribution. However, we cannot make this assumption for our input data, as is visible from exemplary distribution plots for all deontic compartments for both Maghribī and Genoese scenarios provided Section F.1 in Appendix F. In fact the primary value of our approach lies in the *avoidance of any assumed distribution*, since the analytical contribution lies precisely in the representation of *varying degrees of alignment* for normative understanding. For this reason we exclude this processing step from our configuration.

The final step introduced by Liu and Mendel (2008b) effectively represents the necessity of establishing an elementary *footprint of certainty*, in which individuals have a shared understanding. Thus all intervals that do not intersect with *all* other remaining intervals (unless the intervals have the same boundaries) are removed from further processing. In our example in Figure 7.3 this excludes a further 17 intervals from the consideration for membership function generation (marked as *Non-overlapping intervals*).

Summarising the employed preprocessing steps, the incremental filtering process assures that the remaining intervals do not include extreme outliers and are overlapping in order to drive the identification of a minimal commonly shared understanding for a subset of intervals (here: 82) which represents the basis for the lower membership function as observable in Figure 7.3.[3]

Generating Membership Functions Based on the remaining intervals we generate the upper and lower membership function. Our emphasis lies on a generic visual representation that allows us to retrace the generation process and the resulting fuzzy set. Representing all

[3]For further mathematical details on the individual processing steps refer to Liu and Mendel (2008b).

remaining intervals, the upper membership function (UMF) is determined by the union of all admissible input intervals. It thus represents the widest possible normative understanding for a given deontic term within the social group of interest. The intersection of the intervals with the leftmost right boundary value and the rightmost left boundary value describes the lower membership function (LMF). The LMF captures the value range in which agents show the strongest alignment with respect to the evaluated term. The certainty to which an individual's understanding is aligned with the totality of all considered agents is described by the degree of membership `[LMF(x),UMF(x)]`. Ignoring aspects such as weighing, this interval can be defuzzified into a crisp (type-1) degree of membership by taking the mean of lower and upper degree of membership (`[LMF(x),UMF(x)]/2`). From a sociological viewpoint type-1 membership functions can be used to represent the macro-level perspective, informing the experimenter about the *overall understanding* of the certainty with which a given input value represents a given deontic. However, the type-1 membership function does not carry any information about *how unified* the society is with respect to this degree of membership. Type-2 fuzzy sets can contextualise a given aggregate normative understanding with respect to the quality of its alignment. As exemplified in Figure 7.4 this provides the observer with the immediate insight that agents have a comparatively aligned normative understanding with respect to the right boundary, but have vastly diverging interpretations with respect to the left boundary of what *may not* entails.

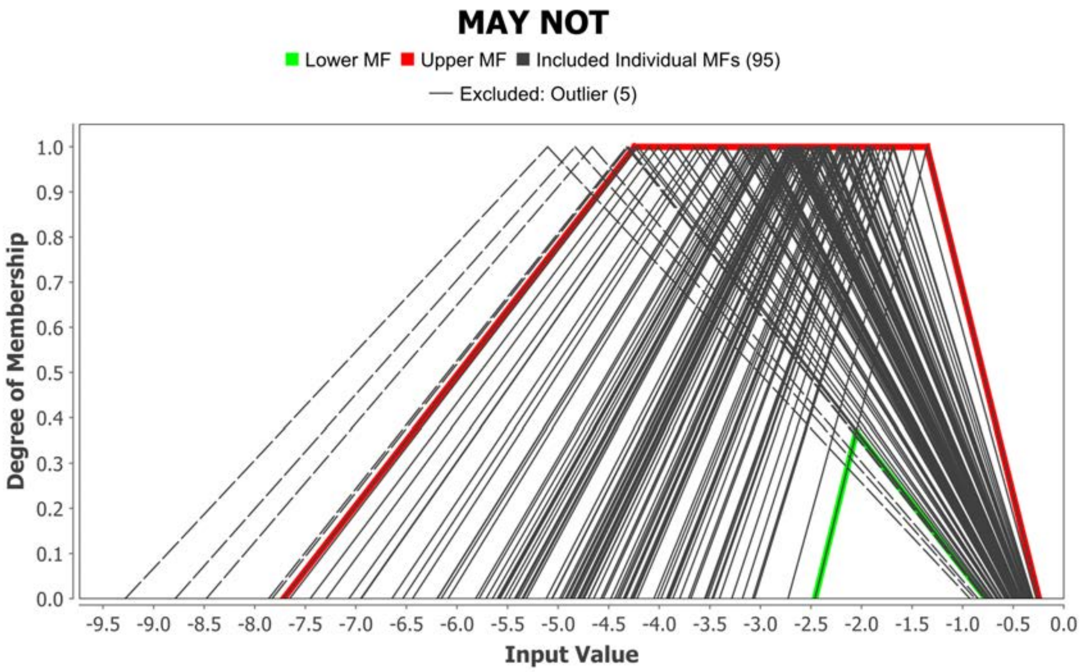


Figure 7.4: Generated Interval Type-2 Membership Functions for *may not*

However, since we use agents as a source for individual input intervals, we need to clarify a central assumption made for the individual inputs. We assume that agents themselves produce symmetric type-1 intervals with full certainty at the intervals' centre and linearly decreasing certainty towards the outer boundaries. Though the agents can relay situationally crisp understanding of the boundaries of the individual compartments, their dynamic construction is based on the unpredictable expansion or contraction of the deontic range, a process that takes place at the end of each round. Thus, anticipating unknown change of the deontic range, the assumption of linear uncertainty is an admissible approximation since the boundaries of deontic compartments are likely to change within short time frames, while the overall allocations of deontic compartments change within longer time frames.

To make the results not only visually accessible, we introduce quantitative measures that capture the quality of the generated MFs with respect to analysed input intervals. Since preprocessing steps potentially affect the number of intervals considered for the actual membership function generation, the generated fuzzy set can thus be described with respect to the fraction of input intervals that have ultimately been included in its generation. This quantitative measure of *Representativeness* is thus the fraction of included intervals from all collected input intervals:

$$Representativeness := \frac{\textit{count(collectedIntervals) - count(filteredIntervals)}}{\textit{count(collectedIntervals)}} \tag{7.2}$$

The second aspect which we capture is the alignment of normative understanding, which quantifies the *Quality* of a given fuzzy set. Ideally, agents should develop strong alignment with respect to the certainty to which a value represents a specific deontic term, thus resulting in an identical upper and lower membership function without any FOU. With micro-level understanding being in full alignment with macro-level understanding, this would reflect a type-1 membership function. The *Alignment* can thus be plausibly expressed as the relative difference in area between lower and upper membership function:

$$Alignment := \frac{\int LMF}{\int UMF} \tag{7.3}$$

Equivalent to *Representativeness*, an *Alignment* measure of 1 indicates identity of UMF and LMF; 0 indicates a lower membership function resolving to 0 for all inputs (`LMF(x) = 0`), marking the greatest possible dispersion of intervals within the boundaries of the UMF.

Both measures are of complementary value and can exist in a trade-off, since reduced *Representativeness* – in the worst case to one interval – increases the agents' *Alignment*.

Conversely, high levels of *Alignment* are only meaningful if the generated membership functions are based on a fraction of intervals that is representative with respect to the observed agent society.

7.4 Exploring the Historical Scenarios

At this stage we apply the introduced extension to our model from Section 6.3. To recall, the essential question concerned the degree to which a differing role understanding could have driven cooperative behaviour among Maghribī and deviating behaviour among Genoese traders. We arrived at the conclusion that, if assuming a behaviourist perspective, a role-integrated understanding likely drives compliant behaviour, and differentiated role understanding drives deviating behaviour.

To provide an overview of the agents' normative understanding along with the measures of representativeness and alignment, we integrate generated upper and lower membership functions for selected deontic compartments along the deontic range. Since the main focus lies on the wider intermediate deontic compartments, in contrast to the narrow extremal compartments (*must not*, *must*) and the central compartment (*indifferent*), the exploration concentrates on *should not*, *may not*, *may* and *should*. We explore all scenarios based on representative simulation runs. Visual representations of membership functions are further contextualised by tables, since the narrow nature of membership functions challenges the interpretation at times. A statistical overview over 30 simulation runs for each configuration discussed in the following is provided in Appendix F.3.

7.4.1 Maghribī Scenario

In closest approximation of historical reality, we use the Maghribī scenario with activated norm enforcement (referred to as Scenario 2 in Section 6.3), and merely extract the agents' situational boundaries of the deontic compartments to generate IT2FS as described above. Figure 7.5 shows the stabilising normative understanding for the Maghribī scenario, with the introduced measures *Representativeness* and *Alignment* as well as the MF boundaries shown in Table 7.1.

The chart reflects the different deontic compartments, with the left two ones (*should not* and *may not*) being characterised by a very narrow nature, which is caused by the use of the zero-centred deontic range configuration (that enforces zero as the centre of the range) in combination with poor negative reinforcement levels. The compartments for the terms

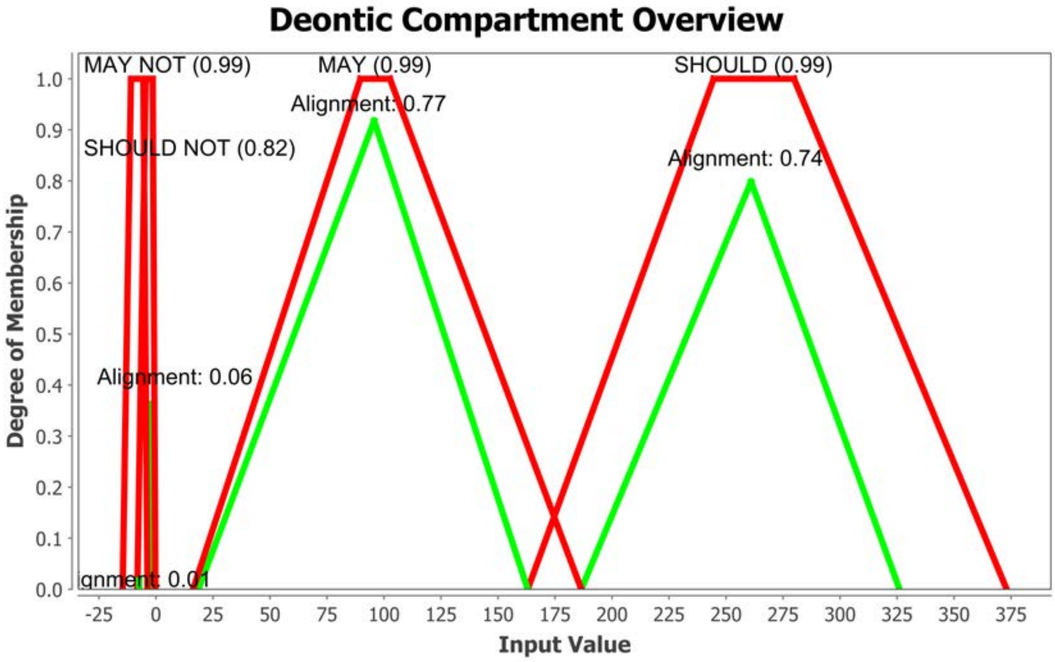


Figure 7.5: IT2FS Representation of Deontic Compartments for Maghribī Scenario

Table 7.1: Macro-Level Compartment Measures (Maghribīs)

Compartment	Representation	Alignment	Min. Value	Max. Value
should not	0.82	0.01	−14.6	−3.7
may not	0.99	0.06	−8.16	−0.26
may	0.99	0.77	16.3	186.53
should	0.99	0.74	163.04	373.07

may and *should* integrate nearly all input intervals,[4] with only one interval excluded in either case (*Representativeness*: 0.99). The alignment of the input intervals is relatively high for *may*, indicating a similar understanding of *may* by all considered agents. On the other hand, with a value of 0.74 this understanding is less aligned for the deontic *should*. More challenging is the analysis of the proscriptive compartments, since both have, despite their narrow (and visually undistinguishable) shape, very low levels of alignment. Since the aggregated view provides a poor comprehensive insight into individual generated fuzzy sets – especially in light of widely varying fuzzy set proportions –, we augment the deontic compartment overview with a dashboard view, shown in Figure 7.6. This allows the inspection of different deontic compartments along with all individual input intervals in order to retrace the generation process.

[4]The Maghribī base scenario is configured with 100 agents.

Using this dashboard view, we can identify the individual compartments and understand the nature of the deontic compartments *should not* and *may not* (see lower charts in Figure 7.6). In both cases the alignment varies between the different boundaries, with *may not* being delimited by the midpoint on the right and a relatively large variation of the left boundaries, leading to a relatively small LMF and an alignment value of 0.06. As before, nearly all input intervals have been considered for the generation of the membership functions (*Representativeness* of 0.99). The case of *should not* shows an even greater variation, with the filtering of 16 non-overlapping intervals in order to establish a LMF after all (*Alignment*: 0.01). The low levels of alignment are in part caused by the comparatively small ranges of both compartments, with *may not* showing a range of below 8, and *should not* a range of around 11, both of which are caused by the limited reinforcement of negative pay-offs since agents generally act compliantly. The complementary compartments *may* and *should* reflect the observation on the macro-level (Figure 7.5), with only one interval[5] being filtered based on its identification as outlier (`> Q(75) + IQR * 1.5`). For both compartments the very strong alignment is observable.

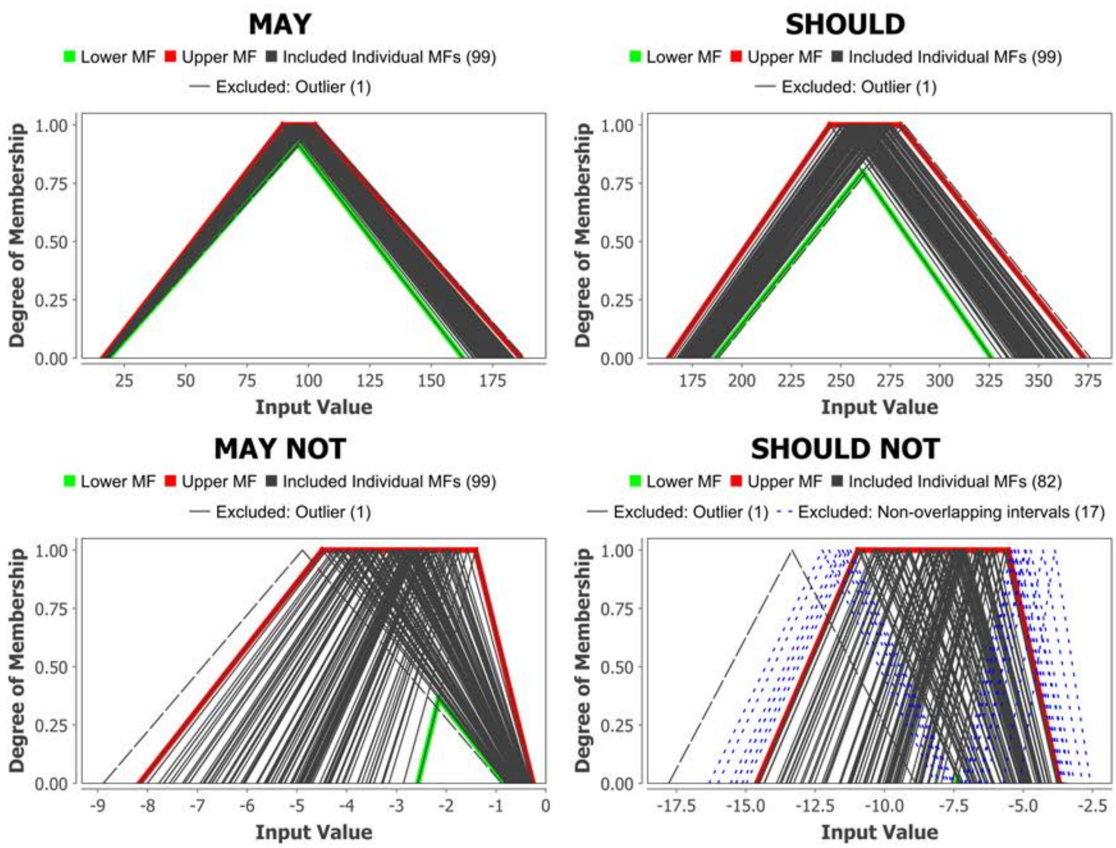


Figure 7.6: Dashboard View for Maghribī Scenario

[5]Upon closer analysis the interval is associated with the same agent for both compartments.

We believe that the translation of deontic ranges into IT2FS in combination with a dashboard visualisation makes the exploration of simulation results very accessible without the need for extensive statistical evaluation. This approach allows the experimenter to navigate both on macro-level (Figure 7.5) as well as retracing the interrelation with the micro-level (Figure 7.6) that allows the interpretation of outcomes.

Knowing the representation of the Maghribī scenario, we can proceed to a meaningful discussion by comparing it to the Genoese variant.

7.4.2 Genoese Scenario

The Genoese scenario, characterised by a differing role understanding and without norm enforcement (Scenario 3 in Section 6.3), offers a different overall normative landscape as shown in Figure 7.7. In contrast to the Maghribī scenario, the distribution of compartments across the deontic range is more balanced, ranging from around -23 to 48. A striking observation is the lower levels of *Representativeness* compared to the Maghribī scenario.

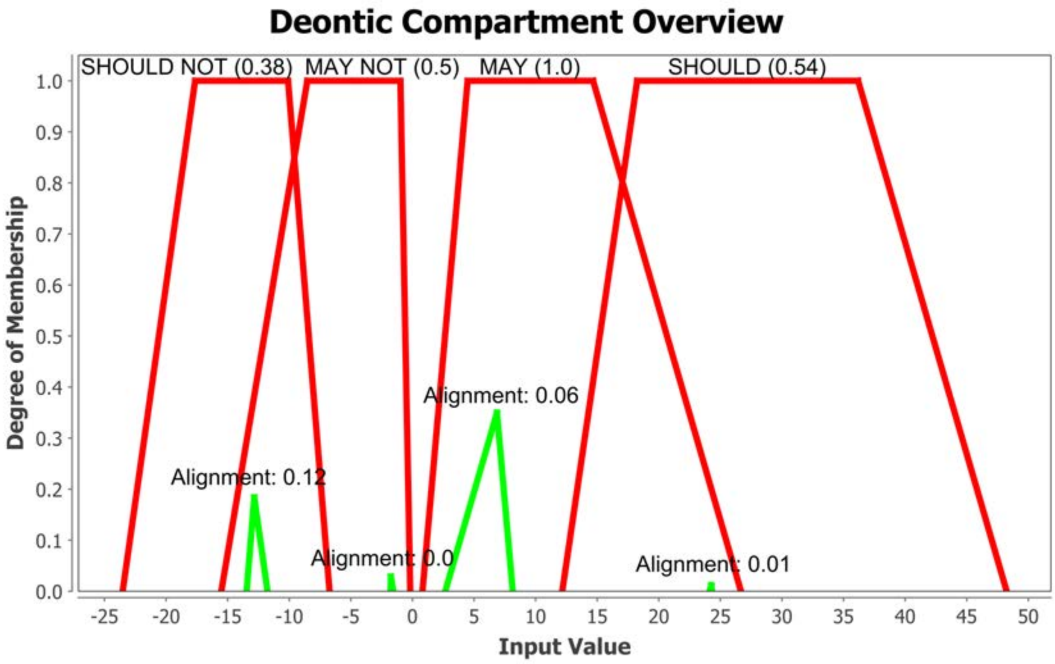


Figure 7.7: IT2FS Representation of Deontic Compartments for Genoese Scenario

The generated proscriptive compartments *should not* and *may not* involve levels of 0.38 and 0.5, respectively, while the level of *Representativeness* for the compartment *should* lies at 0.54. The only exception is the compartment *may*, which is generated using all input intervals. Along with the limited *Representativeness* of the generated membership functions, we can observe the low levels of *Alignment*, reaching levels of 0.12 at best. Although the low levels of *Alignment* follow the previous results (in Section 6.3) that suggested a diverse normative understanding within the society, the low levels of *Representativeness* point towards extensive filtering of intervals prior to fuzzy set formation. We can thus retrace this process using the respective dashboard perspective as shown in Figure 7.8.

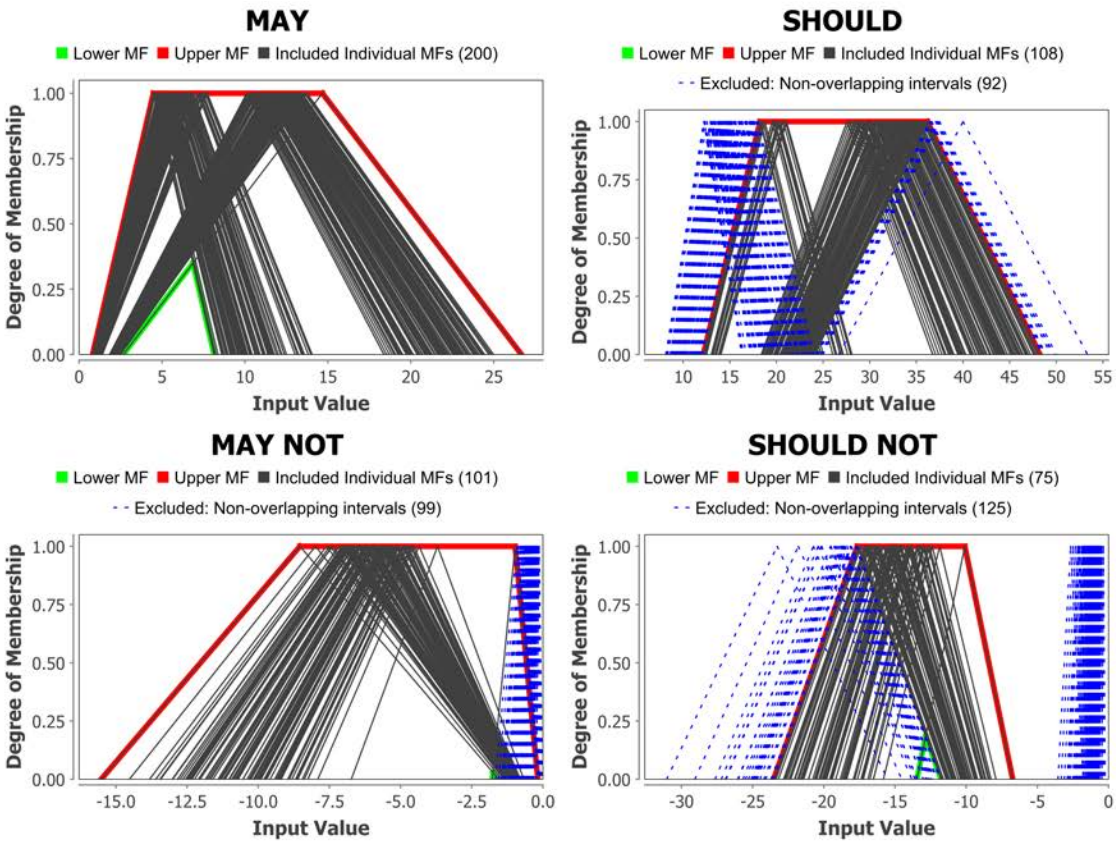


Figure 7.8: Dashboard View for Genoese Scenario

The dashboard view highlights the limited consideration of input intervals. In all cases, intervals are roughly organised in two clusters along the deontic range. The two clusters reveal the differentiating norm understanding among investors and operators, since both shaped their understanding based on different experiences, with operators exclusively exploring action-reaction combinations from an actor's perspective, and the investor experiencing the actee perspective, an aspect we discussed in Subsection 6.3.3.

This example highlights the challenges when using fuzzy sets for the purpose of generating an overall normative understanding, since the generation process attempts the integration of *all* input intervals, i.e. all individuals' understandings. The generation mechanism attempts to establish membership functions around the most central interval, leading to the exclusion of the respective second cluster (since they are non-overlapping intervals), along with potential further individual intervals, such as the case for the compartment *should*. The only exception represents the compartment for *may*. In this case all intervals of the right cluster overlap with the ones of the left cluster, preventing the left cluster's exclusion from membership function generation. The reason for this lies in the objectified deontic range configuration which assures a zero-centred deontic range, leaving the left interval boundaries of the compartment *may* with limited flexibility.

Despite those challenges, comparing the resulting membership functions with the Maghribī scenario, Interval Type-2 Fuzzy Sets offer an accessible representation of what can be described as the societal *normative landscape*. It captures alignment and nature of normative compartments, beyond the time-series-based action distribution across deontic compartments which we explored in the context of the original experiments (Section 6.3).

So far, and similar to the original conceptualisation by Liu and Mendel (2008b), we have assumed that the individuals' understandings (in the form of intervals) *overlap to some extent*. However, applying this concept to the abstract notion of normative understanding, especially when measuring its divergence based on differentiated experience, this assumption does not appear to be truthful to society in reality, since societies may simply not showcase an aligned understanding, an aspect we can clearly observe from the large number of filtered intervals in our simulation case. Consequently, we remove the assumption of minimal common understanding for macro-level analysis, and thus trade *Alignment* for *Representativeness*: intervals do not need to overlap in order to be considered for membership function generation. The resulting membership functions are shown in Figure 7.9, along with the associated dashboard perspective in Figure 7.10. The central measures are shown in Table 7.2. For all cases but for the compartment *may*,[6] no lower membership function could be established, leading to a maximum FOU for those compartments. Only for the compartment *may* can we thus observe a moderate alignment of 0.06.

Table 7.2: Macro-Level Compartment Measures (Genoese)

Compartment	**Representation**	**Alignment**	**Min. Value**	**Max. Value**
should not	1.0	0.0	−31.04	−0.24
may not	1.0	0.0	−15.52	−0.02
may	1.0	0.06	0.81	26.67
should	1.0	0.0	8.14	53.34

Comparing the macro-measures of both the Maghribī and Genoese societies[7] in Table

[6]Recall that for the case of the compartment *may*, similar to *may not*, the zero-centred deontic range configuration prevents wide divergence of the left boundary.

[7]The measures for the Maghribī society shown here are determined without filtering of non-overlapping

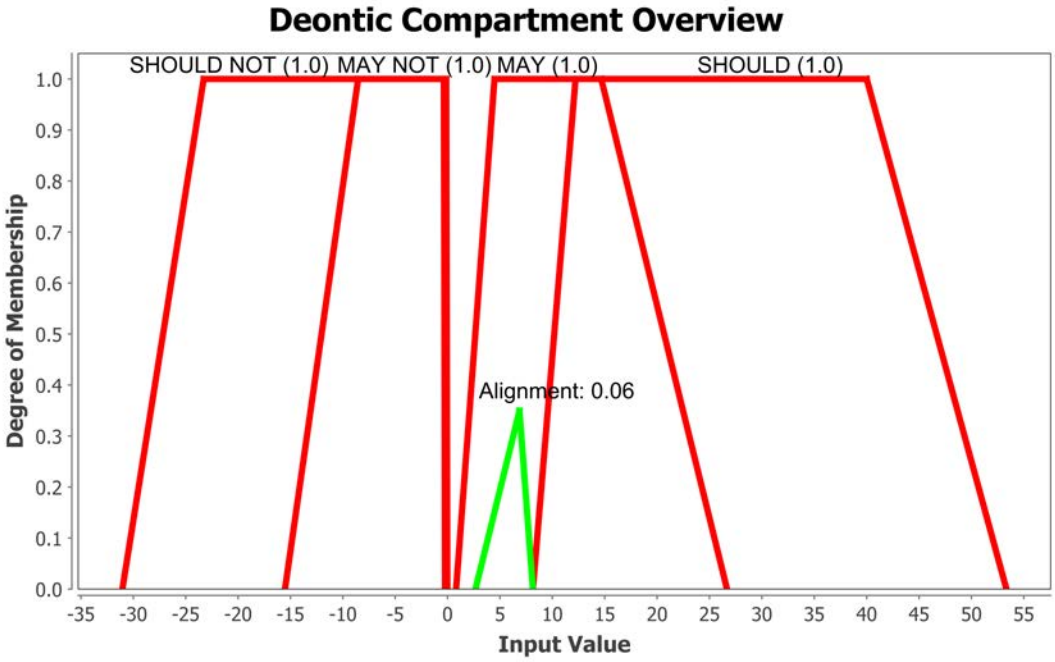


Figure 7.9: IT2FS Representation of Deontic Compartments for Genoese Scenario (Macro-Level)

7.3, we can observe that with comparable levels of *Representativeness* Maghribīs show significantly higher levels of alignment compared to the Genoese society for prescriptive deontic compartments.[8]

Table 7.3: Macro-Level Compartment Measures for Maghribī and Genoese

	Compartment	Representation	Alignment	Min. Value	Max. Value
Maghribī	*should not*	0.82	0.01	−14.6	−3.7
	may not	0.99	0.06	−8.16	−0.26
	may	0.99	0.77	16.3	186.53
	should	0.99	0.74	163.04	373.07
Genoese	*should not*	1.0	0.0	−31.04	−0.24
	may not	1.0	0.0	−15.52	−0.02
	may	1.0	0.06	0.81	26.67
	should	1.0	0.0	8.14	53.34

Knowing the clustered nature of the individual input intervals for the Genoese, the large FOU of the macro-view has limited informational value other than providing information

intervals – in contrast to the previous results in Table 7.1. The complete macro-measures for the Maghribī case are shown in Appendix F.2.

[8]Refer to Appendix F.3 for a statistical overview of all scenarios explored in this subsection, along with a significance test (in Appendix F.3.3) comparing Maghribī and Genoese society setups.

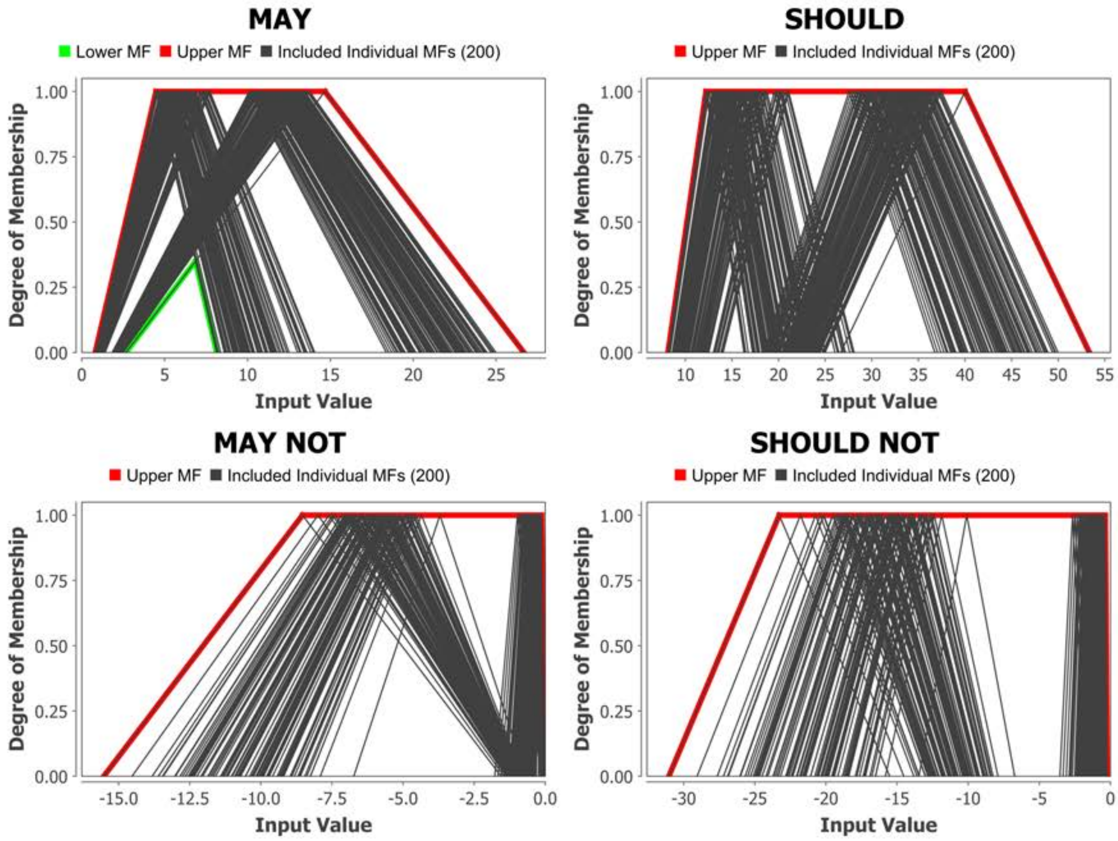


Figure 7.10: Dashboard View for Genoese Scenario (Macro-Level)

about the allocation of the compartments. We will therefore refine the membership generation process to capture a more detailed societal representation.

7.5 Analysing the Meso-Level

The observable clusters in the dashboard view (see Figure 7.10) provide a close approximation of the social clusters to the roles individuals play. We believe that the filtering of non-overlapping intervals – which we previously abandoned as a consequence of the limited representativeness – offers a starting point for further refinement by giving the experimenter facilities for cluster analysis that can be refined based on visual inspection. Since societal configurations can vastly differ, a supervised approach offers the necessary flexibility to reflect the normative stratification into different societal sub-groups, reflecting the sociological meso-level.

The coarse-grained clustering based on filtering of non-overlapping intervals has various limitations that prevent reliance on it as the sole application for meso-level analysis:

- *Primary cluster determination based on mean of input intervals* – The original algorithm uses the mean position of all intervals to define the primary interval cluster, thus picking the cluster closest to the center of the compartment range. Firstly, this constrains from the selection of a particular cluster of interest (see next item). Secondly, in scenarios with a polarised social stratification, the selection of the central cluster can oscillate between either extremal cluster and thus appear arbitrary.

- *Single strategy for clustering* – From the perspective of the experimenter, it would be preferable to have more sophisticated strategies to choose clusters of interest, especially in complex social scenarios that showcase wide stratification.

- *Inability to detect multiple clusters* – Since the filtering mechanism is not geared towards clustering per se, it can only differentiate between filtered and non-filtered intervals, without determining the relationship between excluded intervals. However, for the purpose of social analysis we need to consider an arbitrary number of social groups.

To address these limitations, we preempt membership function generation process with an additional pre-clustering step that generates interval clusters which can then be selectively used as input for the data pre-processing and membership function generation steps introduced in Section 7.3.

7.5.1 Pre-Clustering Intervals

For this purpose we use the density-based clustering algorithm DBSCAN (Ester et al., 1996), which operates unsupervised apart from the specification of two parameters, which, applied to our context, are the *maximum permissible distance for which intervals are considered clustered (ε)*, and the *minimum number of intervals to constitute a cluster*. In contrast to the binary measure of inclusion offered by filtering non-overlapping intervals (i.e. overlapping or not overlapping), we introduce a continuous distance measure that describes the mutual overlap (based on the smallest relative overlap) of two intervals with 0 indicating a full overlap of intervals (or identical midpoints, even if their widths might vary) and 1 indicating no overlap. The distance measure for intervals (with interval properties being *left boundary*, *right boundary* and *midpoint*) is described in Algorithm 7.1.

Based on the clustering threshold we can detect an arbitrary number of clusters, one of which can then be further analysed with respect to other metrics, such as alignment and boundaries. However, this does not solve the problem of cluster selection for inspection. For

Algorithm 7.1: Distance Measure for Interval Pre-Clustering

```
Input: intervalOne, intervalTwo
// Determine the intervals' order and calculate overlap accordingly
if intervalOne.getMidpoint() < intervalTwo.getMidPoint() then
    if intervalTwo.getLeft() < intervalOne.getRight() then
        rightToLeft = 1 - (intervalOne.getRight() - intervalTwo.getLeft()) /
        intervalOne.getIntervalWidth();
        leftToRight = 1 - (intervalOne.getRight() - intervalTwo.getLeft()) /
        intervalTwo.getIntervalWidth();
        // return max. (i.e. lowest) relative overlap
        return max(rightToLeft, leftToRight);
    else
        // no interval overlap
        return 1;
    end
else
    if intervalOne.getMidpoint() > intervalTwo.getMidpoint() then
        if intervalOne.getLeft() < intervalTwo.getRight() then
            rightToLeft = 1 - (intervalTwo.getRight() - intervalOne.getLeft()) /
            intervalTwo.getIntervalWidth();
            leftToRight = 1 - (intervalTwo.getRight() - intervalOne.getLeft()) /
            intervalOne.getIntervalWidth();
            // return max. (i.e. lowest) relative overlap
            return max(rightToLeft, leftToRight);
        else
            // no interval overlap
            return 1;
        end
    else
        // midpoints must be identical
        return 0;
    end
end
```

this purpose, we introduce cluster selection strategies that pick clusters according to their orientation on the deontic range as well as number of intervals of a given cluster. We group the strategies as follows:

- Maximum/Minimum Number of Intervals – Under these strategies, the clustering module selects the cluster with the highest or lowest number of intervals.
- Left/Rightmost cluster – This strategy describes the use of the left- or rightmost cluster.
- Interactive – In this mode the experimenter can pick a cluster to be analysed at runtime.

The pre-clustering and cluster selection steps are integrated with the remaining membership function generation process as schematically shown in Figure 7.11. Note that the pre-clustering as well as the previously discussed filtering of non-overlapping intervals can be selectively de/activated. Depending on the activation, subsequent steps can be superfluous (such as filtering after clustering).

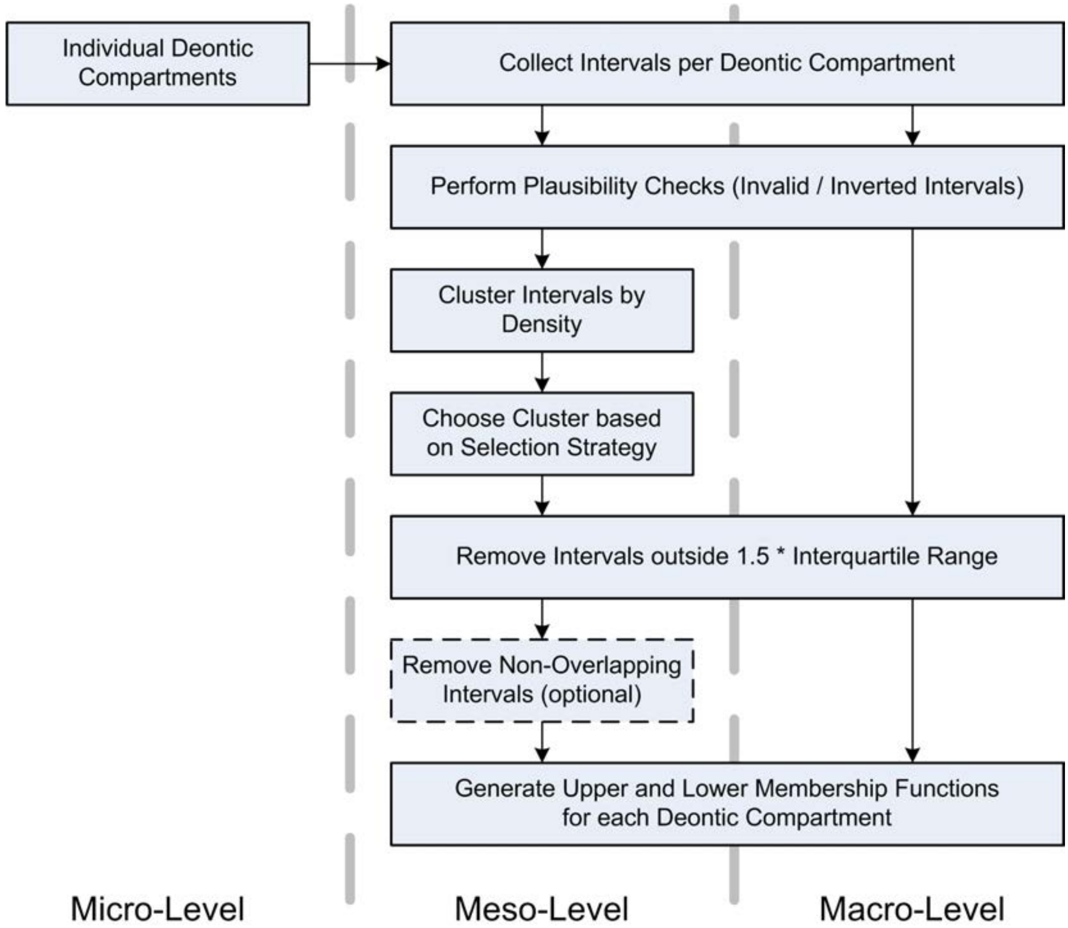


Figure 7.11: Interval Type-2 Membership Function Generation for Analysis on Different Sociological Levels

We organise the schema based on the level of observation. The micro-level perspective merely requires the inspection of individual intervals, which can be directly observed in the dashboard view. The macro-level view requires the deactivation of filtering non-overlapping intervals, while the meso-level perspective is sensitive to the experimenter's intentions, such as the granularity of clustering as well as the selection of inspected clusters, and thus represents the most refined process.

7.5.2 Exploring the Meso-Level of the Genoese Scenario

Left Cluster We showcase this using the Genoese scenario, identifying measures for the respective clusters. Applying a distance threshold of 0.2 based on visual inspection of the dashboard perspective (see Figure 7.13)[9] of the interval distribution and selecting the left-most cluster (since our scenario generally produces two clusters – investors and operators), we arrive at membership functions shown in Figure 7.12 and selected metrics given in Table 7.4. As with the previous cases, we explore the scenario based on exemplary simulation runs and provide a comprehensive statistical overview across 30 simulation runs in Appendix F.3.

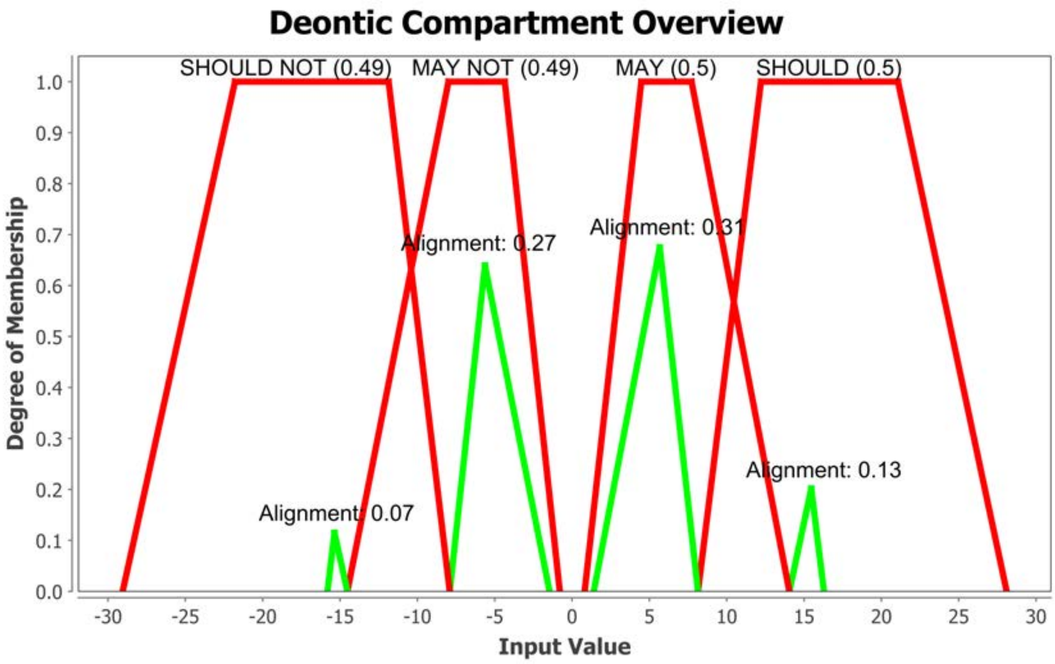


Figure 7.12: IT2FS Representation of Deontic Compartments for Genoese Scenario (Left Cluster)

Table 7.4: Genoese Deontic Compartment Measures (Left Cluster)

Compartment	Representation	Alignment	Min. Value	Max. Value
should not	0.49	0.07	−29.05	−7.91
may not	0.49	0.27	−14.53	−0.79
may	0.5	0.31	0.81	14.05
should	0.5	0.13	8.14	28.1

As a measure of *Representativeness*, we show the relative importance of the cluster with respect to the overall input intervals, since it offers greater informational value. Most clusters thus reflect around 50 percent of the society, and in this case shows the Genoese operators who experience a wide range of feedback, leading to a relatively broad stratification of

[9]Intervals belonging to identified clusters are coloured accordingly. Upper and lower membership functions are plotted for the cluster that has been identified based on the specified cluster selection strategy.

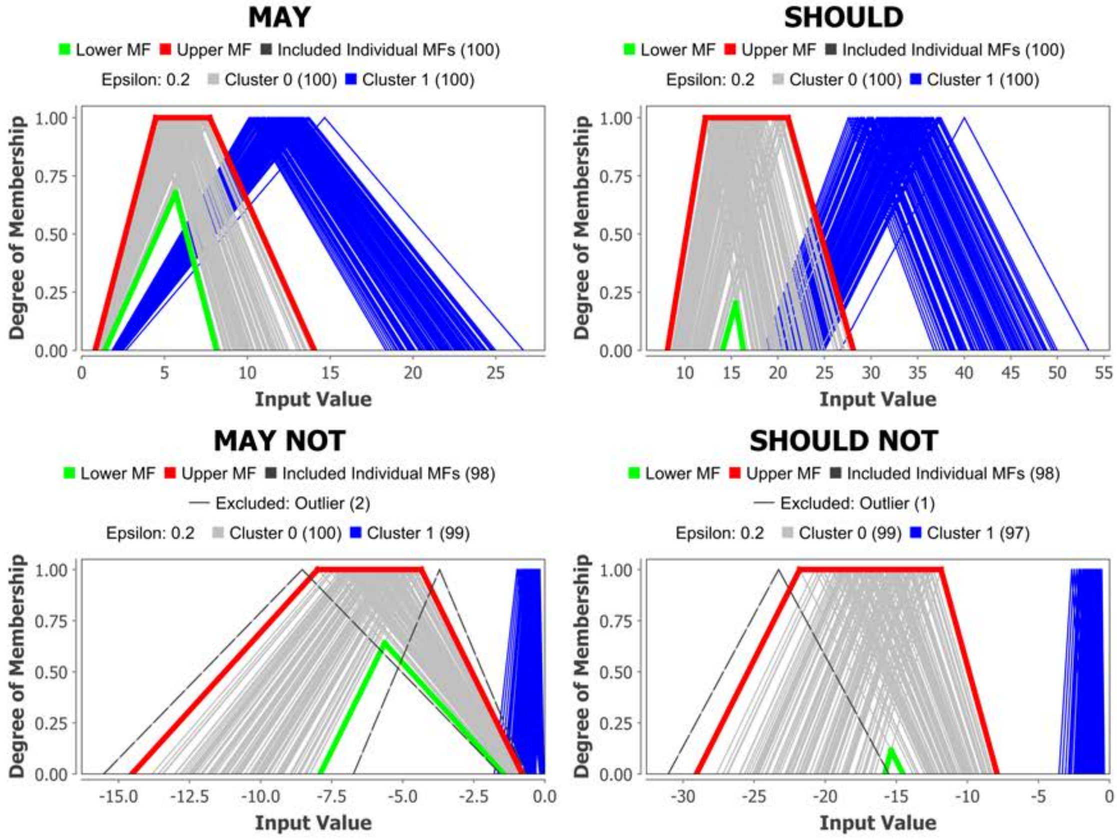


Figure 7.13: Dashboard View for Genoese Scenario (Left Cluster)

compartments. Compared to the macro-perspective, the group focus shows comparatively high levels of alignment, peaking in the compartment *may* with 0.31, with *may not* and *should* showing a common ground of 0.27 and 0.13, respectively. The compartment *should not* shows a relatively poor alignment (0.07), which is caused by the diverse negative feedback received over time, an aspect that is clarified by the dashboard view (Figure 7.13). Beyond the poor intra-cluster alignments, the compartment *should not* for operators shows a clear separation for the right cluster, thus highlighting poor inter-group alignment, while all other compartments show at least a minimal overlap. The proscriptive side of the deontic range thus shows the greatest divergence of normative understanding.

Right Cluster Complementing the left cluster, the right investor cluster can be explored separately, with generated membership functions shown in Figure 7.14, metrics in Table 7.5, and dashboard view in Figure 7.15.

We can observe that the proscriptive compartments *should not* and *may not* are very narrow (value ranges of around 3 for *should not* and 1.75 for *may not*). Within those narrow boundaries intervals further show poor *Alignment*, pointing to poor reinforcement levels

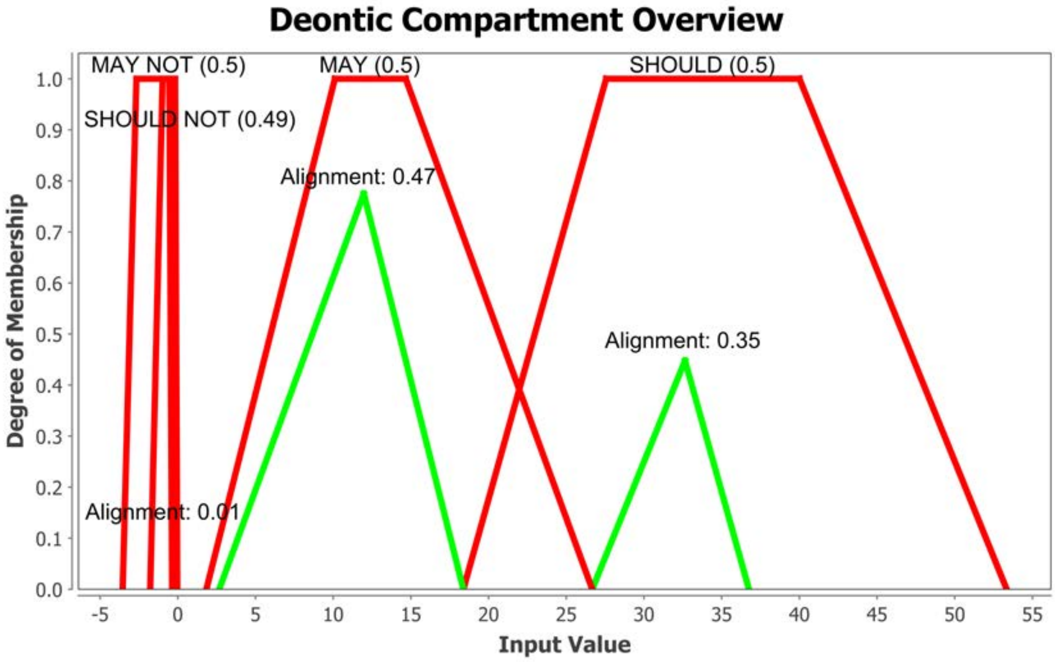


Figure 7.14: IT2FS Representation of Deontic Compartments for Genoese Scenario (Right Cluster)

Table 7.5: Genoese Deontic Compartment Measures (Right Cluster)

Compartment	Representation	Alignment	Min. Value	Max. Value
should not	0.49	0.0	−3.55	−0.36
may not	0.5	0.01	−1.77	−0.03
may	0.5	0.47	1.84	26.67
should	0.5	0.35	18.38	53.34

for those compartments. *Alignment* levels for the compartment *should not* lie at 0.0, indicating the absence of common understanding. The compartment *may not* shows a similar *Alignment* of 0.01. The prescriptive compartments show stronger reinforcement along with comparatively high levels of *Alignment*. In all but one case all agents have been considered for MF generation (*Representativeness* of 0.5). Compared to operators, investors have a stronger *Alignment* along the prescriptive side of the deontic range. Operators' understanding is more widely spread based on their explorative role of initiating transactions, while investors merely 'react'.

Observing the dashboard perspective, the visual identification of different clusters shows the benefit of the introduced clustering approach. It allows the fine-tuning of parameters for clusters that would otherwise be missed by more coarse-grained approaches, such as the original filtering mechanism that attempts to integrate all input intervals into a single

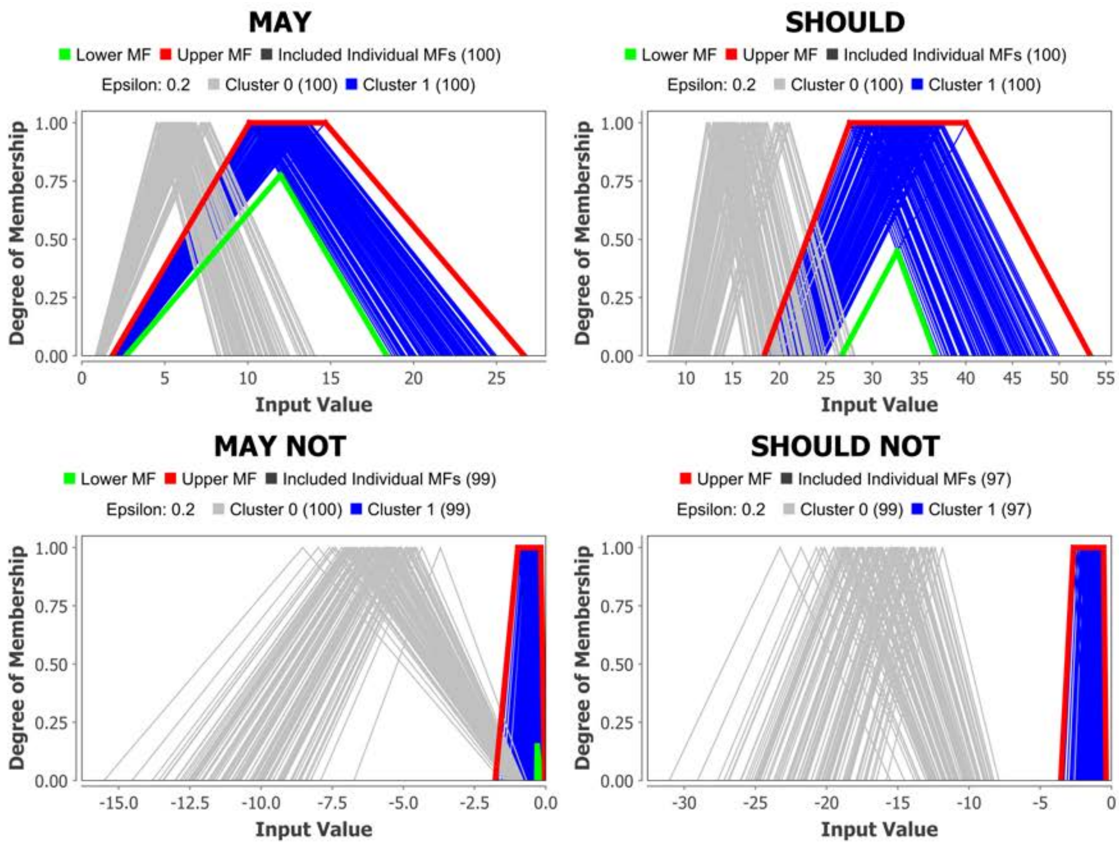


Figure 7.15: Dashboard View for Genoese Scenario (Right Cluster)

membership function without considering the varying levels of interval density. As such the original filtering mechanism was not able to differentiate clusters for the compartment *may* (compare Figure 7.8), because the left boundaries of all intervals are in close proximity but show strong divergence in their right boundaries. This inspection capability would likely be of stronger value for more complex cases with more than two clusters, an aspect that this approach accommodates by offering a variety of selection strategies.

Using a systematic pre-clustering of intervals prior to IT2FS MF generation gives control to the experimenter: not only can one identify homogeneous interval clusters, but perform a supervised analysis of such groupings. From a sociological perspective this permits the identification of sub-groups with respect to dimensions of interest – here: normative understanding. Beyond the mere identification, we can explore how uniform the understanding is within the respective groups. For our example case that identifies two social groups, namely operators and investors, this leaves us with the insight that operators have a far more differentiated experience. Operators have developed an understanding for both obligatory and prohibited actions (as seen by the wider expansion of the deontic range). Beyond this, operators have a more diverse normative understanding which is seen by the relatively poor alignment levels *within* the individual compartments. Investors, on the other hand, have primarily gathered experience on the proscriptive part of the deontic range, thus have developed an understanding 'what they should do', but only to a limited extent 'what they should *not* do'. Furthermore, their understanding is more aligned within the investor group, thus showing stronger homogeneity compared to the operator group.

7.6 Discussion

7.6.1 Summary

In this chapter we extended the exploration of Chapter 6 with additional analytical facilities based on Interval Type-2 Fuzzy Sets (Zadeh, 1975a; Liu and Mendel, 2008b). The original experiment built on the concept of Dynamic Deontics, providing a vehicle for the abstract representation of normative understanding based on experiential learning. We applied this mechanism to the question of how far an otherwise unchanged model drives differing normative understanding based on the comparison of integrated and differentiated role conceptualisations. The findings suggest significantly varying outcomes for Maghribī and Genoese societies, leaving the impression of stronger normative alignment for the former society.

In order to provide a more accessible interactive exploration of the experimental results, compared to statistical analysis or visualisation based on time series, in this chapter we introduced a structured process for the formation of membership functions that integrates individual norm understanding based on Interval Type-2 Fuzzy Sets. Its purpose is to explore the allocation of normative understanding on the deontic range, while providing insight into the extent of alignment. Beyond developing a macro-perspective of the individual normative understanding, we further introduced an interactive toolkit that allows the runtime exploration of intervals on different aggregation levels. From a sociological perspective this leverages analysis not only on the micro- (individual intervals) and macro-levels (society at large), but makes the intra-society group relationships on the meso-level accessible. In combination with the introduction of quality metrics for established membership functions, such as *Representativeness* and *Alignment*, this offers a conceptual mapping onto the notion of *Normative Alignment* of given groups or societies. Contextualising this with our scenario, we can thus not only identify the stronger *Alignment* of the Maghribī society on the macro-level (see Section 7.4), but further decompose the Genoese society into its subgroups (see Subsection 7.5.2) and explore the contrasting normative landscape and levels of *Alignment*. Investors generally show high levels of *Alignment* for the prescriptive part of the deontic range, compared to the operator perspective that showed greatly varying normative understanding within proscriptive compartments.

Though largely explorative, the introduced mechanism offers a more detailed perspective on the experiments of interest and materialised what could be understood as an aligned world view for the Maghribī case, and expressed the diverging world view within and between groups observed in the Genoese society. Though geared towards the representation of sociological concepts, the mechanism is generic and can be applied to problems within and outside the domain of agent-based modelling.

7.6.2 Limitations and Related Work

In this work we have not fully exploited the computational feature set of fuzzy sets, and IT2FS in particular. A particular value of their application lies in the use of established fuzzy sets across multiple dimensions, such as normative understanding and wealth, by specifying rules that trigger potential consequences. The potential extensions are manifold. Employing some of these, a collective of agents could not only express their normative alignment in generated membership functions but also assign gradual sanctions for violations. However, the focus of this work lies on the provision of analytical mechanisms that facilitate the exploration of emerging normative understanding on different levels of aggregation.

Looking at the intersection of T1FS and ABM, we can find a wider range of approaches. Although demands to use fuzzy sets to model social concepts are not new (e.g. Cioffi-Revilla (1981); Epstein et al. (2003)), fuzzy sets have only recently found stronger recognition, such as for the representation of personal traits (Ören and Ghasem-Aghaee, 2003), addressing fuzzy concepts such as romance (Situngkir, 2007), partnership selection in negotiations (Ren et al., 2007), motivation (Arredondo et al., 2013), trust modelling (Aref and Tran, 2014), the comparison between fuzzy and crisp values in BDI systems (Vu et al., 2013), and the modelling of poverty levels in Tijuana (Márquez et al., 2011). Along with calls for a wider recognition of fuzzy sets in ABM (Sabeur and Denis, 2007; Hassan et al., 2007), the most related approach that uses fuzzy sets to structure a continuous social domain we have found in the work by Hassan et al. (2011). They use Type-1 fuzzy sets to model friendship and partnership dynamics in alignment with the European Value Study using the introduced Mentat model. Their model reflects shifting relationships with a fuzzy progression from stranger, via acquaintance, towards friend and partner, relying on the Ordered Weighted Averaging Operator (OWA) (Yager, 1993) to derive individuals' relationship distances.

The use of IT2FS in the context of ABM or other forms of social modelling has been comparatively limited. The predominant (but not exclusive) application field remains engineering, covering application fields of robotics (Hagras, 2004), fuzzy control (Hagras, 2007; Cortes-Rios et al., 2014), clustering (Rhee, 2007; Hwang and Rhee, 2007), image processing (John et al., 2000; Huang et al., 2010), but also extending into domains such as business operations (Miller et al., 2011, 2012; Liu et al., 2012), and biology (Ramírez et al., 2011).[10] However, for the context of normative modelling or institutional analysis, we are not aware of any previous applications of IT2FS.

7.6.3 Concluding the Exploration

Contrasting this chapter with preceding ones, this chapter provides an analytical contribution, enriching the modeller's toolbox with mechanisms that align with the observation of phenomenae of interest, while bridging the conceptual gap between simulation results and sociological interpretation. Reflecting on the structure of this work, the contributions of this chapter provide the analytical features that augment the representation of normative understanding (Section 6.2) for a concrete simulation model (Section 6.3), thus completing the analytical stack schematically shown in Figure 7.16.

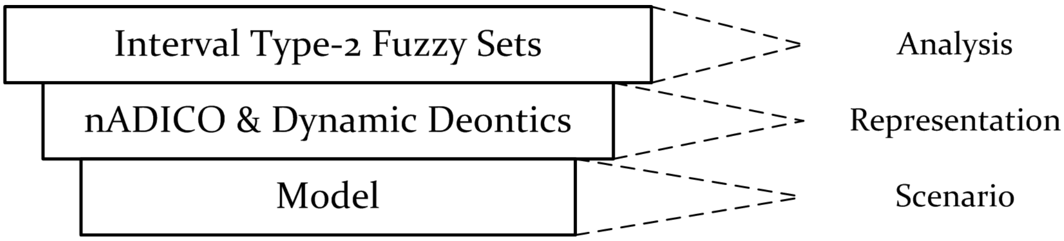


Figure 7.16: Technological Stack of Behavioural Institutional Analysis

For the purposes of analysing the problem of interest, the Maghribī Traders Coalition, this chapter's contributions conclude our outlook on institutional analysis as far as applied in this work. Following the scenario of concern, the notion of institutional analysis employed here is concentrated on the informal domain and assumes a behavioural perspective. At this stage, the individual elements that constitute our approach to institutional analysis (representation and analysis) are loosely coupled and leave the modeller great freedom with respect to applications. More refined analytical guidelines would extend this initial step into the direction of one could call *Behavioural Institutional Analysis*. Despite the necessity of further refinement, we believe the current contributions carve out directions towards a more realistic analysis of institutions in human societies. Firstly, we offer a *uniform representation* (nADICO) that captures various institution types that is accessible to both human modellers

[10]For a broader focus on the work related to fuzzy sets, both T1FS and T2FS, refer to Subsection 2.3.4.

and software agents and applicable on arbitrary levels of social analysis (Chapter 5). Secondly, we give up rigid norm conceptualisations and facilitate a more realistic[11] modelling of *dynamic and continuous norm understanding* (Dynamic Deontics) that can reflect long-term processes of institutional change and uses nADICO for its operationalisation (as described in Chapter 6). Finally, this chapter contributes the necessary *analytical mechanisms that reflect the soft and fluent boundaries of normative understanding, while accommodating the multi-level nature of social systems* (as introduced in Section 2.1). The combined application of those components as done in this work thus sketches an integrated approach to Behavioural Institutional Analysis that provides mechanisms to model but also to analyse complexity in an accessible manner.

Returning to the sociological discussion in Chapter 2, the introduced analytical stack fosters a systematic representation of emergent institutional understanding by extracting it from individual agents and providing a macro-level aggregation as visualised in Figure 7.17. Conceptually this approach displays first-order emergence (Squazzoni, 2008). But even though we leverage an explicit multi-level understanding, it is not fed back into the simulation model itself (depicted as a dashed line in Figure 7.17). In consequence this institutional knowledge is not directly accessible by the micro-level entities and thus cannot influence their decision-making – characteristics that would constitute second-order emergence (Squazzoni, 2008).

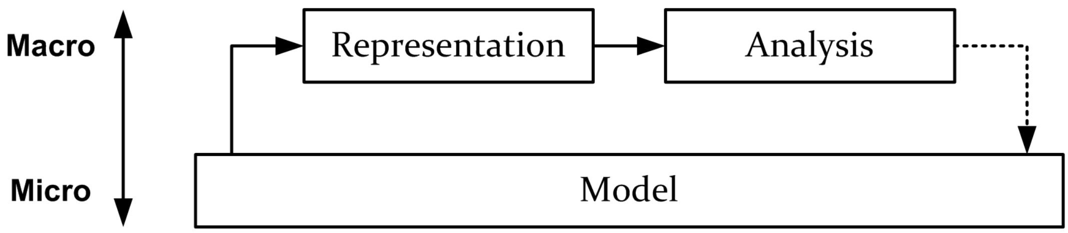


Figure 7.17: Emergence Processes in Behavioural Institutional Analysis

The reasons for this are twofold. Firstly, the institutional representation is of general character, and thus applicable to agent architectures of varying complexity. Consequentially, the experimental evaluation as presented here relies on a relatively primitive agent conception that does not extend beyond simple sensory and memory capabilities, and thus does not provide sophisticated cognitive abilities necessary to process institutional knowledge (as discussed in Subsection 2.3.1). Secondly, though the models could have captured these capabilities in principle, introducing a feedback mechanism from the detected institutional representations would have challenged the puristic interpretation of role stratification vs. integration put forth as part of our exploration in Section 6.3 onwards. The incorporation of

[11] Refer to Subsection 2.2.2 for different positions on institutional change.

the institutional representation would make the introduced representations part of the model itself, and as such part of the underlying assumptions, an aspect this work considers with great care. At this stage the use of the representational and analytical components is non-intrusive and does not directly reflect the reconstitutive downward causation we considered essential to drive the reification of institutional understanding based on the internalisation of habits (see Subsection 2.2.2).

Even though not directly influenced by the aggregate institutional constructs proposed in the work, some of the scenarios presented in Section 6.3 onwards do exhibit characteristics of second-order emergence. Scenario configurations that reflect the Maghribīan regime of norm enforcement accelerate the establishment of cooperative outcomes, since agents sanction other agents' behaviours based on their own experience (including trade experience, observational learning and being sanctioned themselves). This notion of norm enforcement thus reflects a reaction to the institutional reality as internalised by the agent, and thus represents reconstitutive downward causation (see Subsection 2.2.2). This viewpoint is compatible with Castelfranchi's interpretation of minds as social institutions (Castelfranchi, 2014) themselves, since those capture the experiences and thus expectations of behavioural manifestations attached to what we interpret as institutions.

However, the consequent next step for future exploration is to close this feedback loop to establish a systematic account of second-order emergence that involves the introduced higher-level institutional structures as means to inform agents' decision-making capabilities.

Beyond the reinforcing nature of second-order emergence, in principle this link further facilitates the modelling of top-down enforcement of institutions specified at the macro-level. This is of particular relevance for the objective domain (see Subsection 2.1.3), an aspect we do not reflect at this stage. Although the introduced institution representation in principle permits the operationalisation of rules, we do not model the collective action necessary to instantiate rules, but concentrate on a passive observation of the informal institution spectrum. Furthermore, our approach does not explicitly represent constructs such as organisations, markets and the State, all of which are important players in a more comprehensive approach to institutional analysis (Hollingsworth, 2000).

8

Discussion and Outlook

8.1 Synthesis

Motivation We now draw this work towards a closure. We have looked at a scenario from the area of comparative economics that explores the path-dependent institutional development of two societies, the Southern European Genoese and the North African Maghribīs that operated based on formal and informal institutions, respectively. In his seminal work, Avner Greif (1989, 1993, 2006) applied game-theoretical analysis as a means of identifying equilibria he considered indicators for institutions. He further postulated that the respective societies' institutional paths either limited, or enabled, their long-term success. In his work he ultimately considers cultural factors decisive for the differing institutional outcomes, an aspect that has been criticised for its generic nature (Aydinonat, 2006; Goldberg, 2012c). More than the outcome, we are concerned with various rigid and in our view overly simplistic assumptions he posits for his game-theoretical models in order to create bifurcated experimental setups. Assumptions of central interest include the

- inherent secrecy of the Genoese trader society – in contrast to the inherently talkative nature of the Maghribīs,

- inherent closedness of the Maghribī trader society – in contrast to the inherent openness of the Genoese to newcomers, and
- the existence of separated role understanding for models of both societies – which represents an inaccurate modelling simplification, since only Genoese actually operated in role stratification.

Literature Background In order to address those assumptions, we have drawn on a detailed overview of background literature (Chapter 3), with particular focus on the Maghribī trader society, since its literature base builds on a century-spanning collective effort of translating and interpreting documents of the Cairo Geniza, in contrast to the comparatively well-documented Genoese society. Displaying the challenge of historical analysis, we have highlighted and assumed positions in selected literature debates (Section 3.2) that are of relevance for the above-mentioned assumptions. Based on the resulting inclusive picture, we have plotted a refined understanding of the respective trader societies. With the established background knowledge, we employ the interactionist metaphor of ABM – instead of the rational game-theoretical approach employed by economic historians such as Greif –, to develop models that address the specific assumptions mentioned above. Agent-based Modelling enables us to shift the emphasis from rational selfish strategy choices framed in prototypical game structures, to the focus on modelling agent interactions. Doing so, we have traded rigid assumptions, such as the closed nature of the Maghribī Traders Coalition in the original game-theoretical model (see Subsection 4.1.2), for a detailed exploration of sub-aspects of the original scenario. An example is the function of the apprenticeship system with respect to cheater detection and its potential to relax the assumption of a closed trader coalition. We will recall the considered assumptions in the following.

Reviewing Assumptions Reviewing the first assumption, the inherent secrecy of the Genoese trader society, we introduced a model of informal information transmission in a modelled Genoese trader society in Subsection 4.2.1. To test whether inquiring for cheater advice among fellow investors – in the light of an open society – could have been sufficient to sustain cooperative outcomes, we built a request-based 'pull model of communication': Investors 'ask' fellow investors and adjust their own likeliness of truthful reporting according to the perceived truthfulness of the advisor's response. Even though optimistically parameterised, the model shows limited ability to remove cheaters from future trade interactions effectively. To address the limited information about a historical investor relationship network, we tested this model across four different topologies and measured *the extent* to which cheaters could be removed (effectivity) and *how fast* this level could be reached (efficiency). To provide a comparable case for the Maghribī scenario, we explored the impact of proactive sharing by conceptualising a 'push model of communication' which led to significantly better outcomes based on the comprehensive sharing of cheater information within individuals' aṣḥābs. This leads us to suggest that whether sharing advice or not, Genoese investors could not have successfully operated purely based on informal means.

Extending our exploration to the second assumption – the inherent closedness of the Maghribī Traders Coalition –, we built on literature evidence that posits the existence of a "system of junior associates" (Goldberg, 2012c), which we tagged as 'apprenticeship system'. Hypothesising that the principle of a closely monitored apprenticeship period could have offered means to control access of newcomers instead of being inherently closed to outsiders, in Subsection 4.2.2 we constructed a model that allowed us to explore the apprenticeship system's effect on the cheater level in the coalition. Instead of claiming historical accuracy, we argue that we have developed a better understanding of the institution 'apprenticeship system' by exploring the impact of selected parameters, such as cheating probability, openness to relationships with status differences, trade interactions and number of agents, in order to provide a more comprehensive picture on the functions and dependencies such a system could have had. We arrived at the conclusion that the apprenticeship system was particularly effective for low levels of openness towards newcomers (based on status difference) as well as a low frequency of trade interactions (which made the detection of cheaters harder) and showed near linear scalability for tested values. This supports the suggestion that the apprenticeship system could have played a central role in maintaining a comparatively cheater-free trader community.

Turning to the last assumption, or rather simplification – the presumed separation of roles for both societies –, we faced a specific challenge. While previous experiments concentrated on the characteristics of a specific society, Genoese or Maghribīs respectively, the last aspect opens a comparative perspective for both societies. Like Greif, we were thus challenged by the need to assume a sufficient level of abstraction in order to establish a common baseline for both scenarios. Since we have limited access to the individuals' actual historic behaviour, we concentrated on a behavioural perspective and necessarily retreated to the minimal assumption of experiential learning, which we believe, is shared across many different cultures and held for medieval traders as it does for modern societies. However, doing so required a set of conceptual additions.

Institution Representations For the establishment of a uniform representation of experience, we concentrated on conceptual contributions in the form of a general-purpose institution representation (Chapter 5) that makes action dependencies accessible as 'nested consequences'. Operationalising a continuous norm understanding, we introduced the concept of Dynamic Deontics (Chapter 6) that extracts normative understanding from knowledge an agent gathers based on experiential and social learning.

Using these conceptual contributions, we constructed an abstract representation of the trade scenario in which individuals' experiences are defined by the roles that they play (Section 6.3). The results suggest that an integrated role understanding could have been a contributing factor to drive compliance within a collective that was characterised by shared stakes in long-distance trading. In this scenario, as in historical reality, Genoese investors and operators follow the selfish interest of short-term enrichment, leading to an understanding that cheating is desirable. Though not challenging historical reality per se, the conceptual contributions and the model introduce a behavioural dimension into the analysis in order to address a detail that previous research has neglected, thus broadening the discussion around influence factors on the historical development.

Analysing Normative Understanding Augmenting the introduced representations of normative understanding with accessible means of analysis, we applied the notion of Interval Type-2 Fuzzy Sets to establish what can be interpreted as a normative landscape. In addition to the analysis of individual intervals or the aggregate exploration on the macro-level, we introduced a supervised clustering approach an experimenter can refine at runtime in order to carve out the aggregation levels of interest, thus leveraging meso-level analysis. Applying this tool, the evolving normative understanding can be reviewed with respect to quantitative measures, with particular focus on the normative alignment within given agent societies. Exploring the historical scenarios, the Maghribī case is characterised by a comparatively high level of alignment on the macro-level, while the Genoese normative landscape is heterogeneous and can only accommodate unified views for individual groups on the meso-level.

Returning to the Bigger Picture Returning to the discussion of the overall contributions, let us review the essential argument made in our research, namely the ability of ABM to provide a more refined perspective on institutional analysis by 'digging deeper' into concrete scenarios and representing properties that may fall victim to high-level abstractions.

In our view Greif's work is an example of such nature. His work produces a convincing account of the historical reality of Maghribī and Genoese society. His writing offers an impressive narrative and is flanked by detailed accounts on the historical happenstance and interpretation that persuades the reader to 'buy into' his bifurcated conception of the contrasted societies. This can lead readers to overlook the at times overly abstract assumptions Greif puts forth for his game-theoretical analysis, which breaks with the level of detail he offers when introducing both societies. In Greif's defence we admit that the technology available at the time of his investigation, the end 1980s, may have challenged a comprehensive agent-based model, a challenge that we face even with current computational capabilities.[1] However, more than challenging Greif's work, the presented work shows how agent-based modelling is a venerable *addition* to the analytical investigator's toolbox. We thus *do not argue for an alternative approach* to the historical analysis, but rather advocate the potential to *augment* abstract formally grounded game-theoretical analysis with detailed behavioural modelling based on the agent metaphor for specific sub-scenarios. In this particular case it highlights how the rationalised representation of historic happenstance can be complemented with an interaction-centric account that revisits and fills gaps the formal approach fails to reflect appropriately. Examples for such coarse generalisations include conceiving the Maghribīs as an inherently closed collective, or the failure to consider a historically existent diverse role structure for both societies.

The Challenges of Analysing Historical Scenarios Although increasing the level of detail with which we can model scenarios, our analysis – similar to Greif's – highlights the challenge of working based on weak historical accounts. Similar to Greif's work, we rely on a set of basic assumptions on which our models build, in conjunction with largely anecdotal accounts that drive modelling decisions. Nevertheless, although we cannot provide clear answers as to what was decisive for the Maghribīs' cooperation based on informal institutions, our results suggest that rigid assumptions of closedness and cultural background are not the only possible explanations. Relaxing these central assumptions, we open the space for potential alternative explanations. As such, we can argue that the Maghribīs' apprenticeship system may have been sufficient to filter newcomers and thus sustain cooperation while giving access to outsiders as suggested by Goldberg (2012c). Likewise, assuming a behaviourist perspective in an otherwise identical scenario, the integrated role understanding of the Maghribīs (as opposed to role stratification in Genoese society) could have been decisive for the cooperation outcome – as an alternative to the assumption of collectivistic traits. The revised understanding of the Maghribī relationship network structure (see Subsection

[1] For the apprenticeship model (Subsection 4.2.2) the processing of an individual scenario across the parameter range for two selected variables (for a statistically significant number of runs) required up to three weeks of computational runtime on modern desktop hardware (Intel i7 8-core, 12 GB RAM).

3.2.2)[2] based on the obligation of individuals towards their respective aṣḥābs (i.e. groups of reciprocal partners), as opposed to the aṣḥābunā at large (i.e. the collective in its entirety) as assumed by Greif, offers further support for alternative explanations, beyond the reference to a collectivistic cultural background. Assuming a comprehensive observation network (as supported by the intensive sharing of letters (Goldberg (2012b); Subsection 3.2.2) and the absence of global knowledge about aṣḥābunā membership, individuals had strong incentives to report profits and cheating behaviour truthfully, since they may not have known all partners' reciprocity relationships and thus information sources. Instead of acting based on dedication to the coalition grounded in collectivistic beliefs, traders could have simply been compliant based on fear of detection of violations. Concluding this reasoning, and since all approaches rely on weak factual grounding, we believe that the best possible approximation of historical reality is offered by a) minimising assumptions that underlie the respective model, and b) by considering known social aspects when remodelling human societies. This is aligned with our argument for ABM's flexibility to explore gaps in historical research that cannot be sufficiently explored by rational analysis alone.

8.2 Summary of Contributions

Concluding this work, we will summarise the provided contributions. Doing so, we retrace the contributions based on their differentiation as substantive, conceptual and methodological. The structure of this thesis roughly reflects the progression from substantive to methodological contributions as shown in Figure 8.1.

8.2.1 Substantive Contributions

Starting in Chapter 3, we provided a detailed literature overview of the explored societies, with particular focus on the Maghribī society, the perspective on which is fragmented by the various scholars that have interpreted the historical sources. We provided a detailed account on conflicting views in literature in Section 3.2, dubbed 'Literature Debates', in which we laid out the different positions, contextualised those with recent additions to the research field, in particular work by Goldberg (2012c) and Ackerman-Lieberman (2014), and assumed a position with respect to our own work. Central to our concerns are those involving the discussion around the formal or informal nature of selected institutional instruments such as the ṣuḥba (Formality/Informality Debate), which we discriminate not by

[2]We applied this understanding for the apprenticeship model experiments.

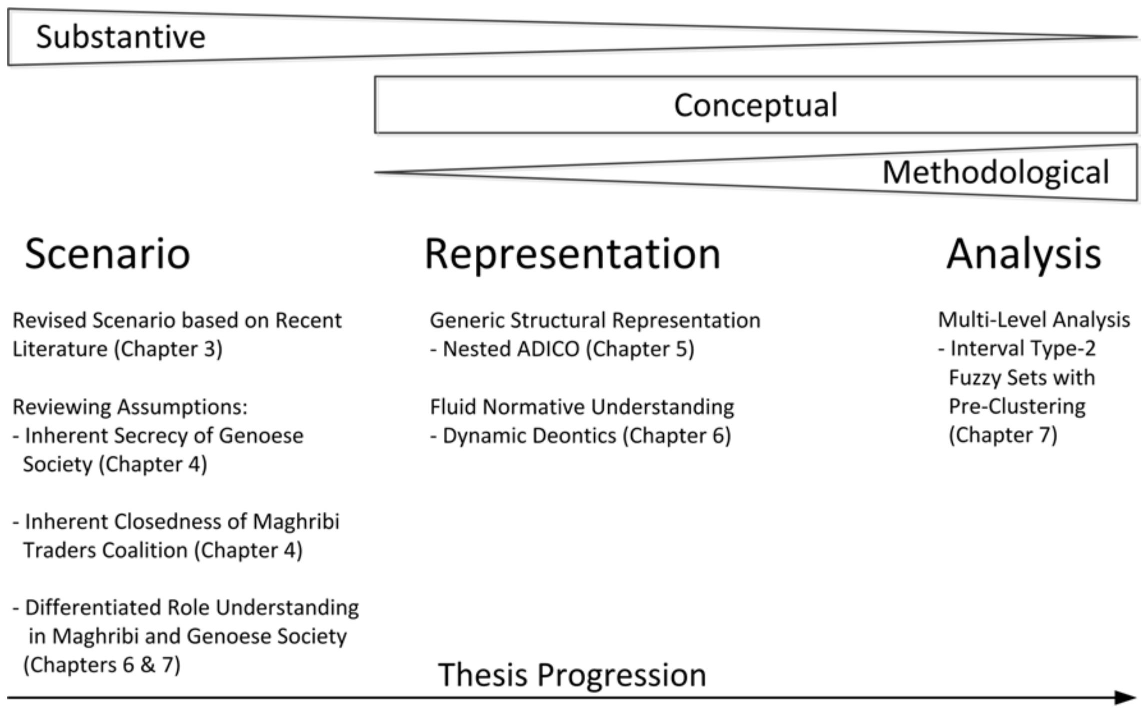


Figure 8.1: Thesis Structure

their unwritten nature or existence of well-defined rituals, but by their legal enforceability (Subsection 3.2.1). Based on Goldberg's work (2012c), we highlight the ṣuḥba as the essential institutional instrument employed by Maghribīs, an aspect Greif was not specific about in his original work. The concept of the ṣuḥba – the intimate 1:1 relationship between traders – as the primary relationship type afforded a reinterpretation of the network structure of the Maghribī Traders Coalition (Subsection 3.2.2). Instead of conceiving the aṣḥābunā as a well-defined group of like-minded individuals with crisp boundaries, as done by Greif (1989, 2006), more recent findings by Goldberg (2012c) and Ackerman-Lieberman (2012, 2014) led us to suggest that only sub-networks, the individual traders' aṣḥābs, had a well-defined structure. The aṣḥābunā itself was a virtual structure that emerged from the interlinked aṣḥābs. This revised perspective removes the assumption of a self-less 'esprit de corps' (Ackerman-Lieberman, 2012) towards the aṣḥābunā, and reduces the barrier for a self-reinforcing institution, since individuals only require selfishly motivated commitment to one's aṣḥāb, as opposed to the coalition at large.[3] Although we do not directly contribute to the primary original Geniza research, the contemporary overview is not detached from specific perspectives assumed by historians (e.g. Goldberg), economic historians (e.g. Greif),

[3]A further aspect of relevance is the discussion of institutional constraints that affected medieval Mediterranean trade. Since the latter is of limited direct relevance for our experimental evaluations, we shifted this discussion into Appendix A.

and legal scholars (e.g. Ackerman-Lieberman). Instead we merge the different perspectives under the leading theme of explaining the cooperation among the Maghribī Traders, marrying the economic enquiry initiated by Greif, with the detail-laden information base offered by Goldberg, Ackerman-Lieberman, and others, to provide an updated picture on the complex historical reality of the Maghribī Traders Coalition.

Beyond the literature contribution, our experiments contribute to the field of comparative economics, clarifying that the *cooperation outcome of the Genoese society could not have been sustained based on informal communication* alone. Though supporting the currently prevailing analytical perspective put forth by Greif (2006), our contribution lies in the relaxation of the assumption of secrecy among Genoese (Epstein, 1994). Results are tested across various network topologies, for which we further suggest a mapping for the discussed societies.

A further substantive contribution is the *analysis of the apprenticeship system in the Maghribī Traders Coalition*. Since this is only recently documented, it has not been explored with respect to its institutional function. We put forth the hypothesis that one purpose was to prevent cheaters from joining the trader coalition, and support this claim by systematically exploring a model that builds on societal properties (e.g. status, network structure, relationship duration) extracted from the literature (see Subsection 3.2.2). The results identify the openness towards newcomers (acceptable lower jāh difference) as a central parameter to determine the apprenticeship system's effectiveness in removing cheaters. The network structure based on loosely coupled aṣḥābs further promises scalability potential.

The final topical contribution addresses a central simplification made in previous research by introducing the consideration of a society-specific role understanding into the discourse, whose potential influence on emerging behavioural equilibria (i.e. institutions) has yet been neglected. Our results suggest that from a behavioural perspective, an integrated role understanding could have primed the Maghribīan society for cooperative behaviour, while the Genoese role stratification may have motivated conflicting interests that required governance by formal mechanisms. We explored the normative alignment of both societies and on different social levels (e.g. analysis of sub-groups in the Genoese scenario). Under the assumption of individuals as experiential learners, our work shows that social aspects such as a differentiated role conceptualisation are significant enough to require consideration in further explorations of the comparative scenario.

More generally, this work has explored the Maghribī Traders Coalition from a social perspective, beyond its primary exploration from an economic perspective. This expands the discourse of Maghribī Traders into a wider domain for detailed analysis beyond the broad question of cooperation. Using the agent metaphor for our exploration, we have further introduced this important scenario into the context of agent-based modelling, beyond the economically motivated game-theoretical analysis.

8.2.2 Conceptual Contributions

The conceptual contributions include the introduction of a generic institution representation, coined nADICO (Chapter 5), that builds on previous work in the area of institutional analysis and political science. nADICO can reflect various types of formal or informal institutions and represent their interdependencies. Specific contributions lie in its extended ability to represent norms by introducing the explicit notion of sanctions, and leveraging a sociological, as opposed to mere syntactic differentiation between norms and rules. Its particular strength lies in the comprehensive representation of different institution types as well as their complexity, including aspects such as structural institutional regress based on systematic nesting capabilities, while maintaining interdisciplinary accessibility.

In addition we present a continuous norm representation, named Dynamic Deontics (Chapter 6), that reflects the emergence and change of normative understanding over time. It emphasises a behavioural perspective and merges social aspects with subjective norm understanding as well as individual properties, while being indifferent about the means of norm sharing (e.g. by experience, observation, communication, etc.), and thereby offers strong generalisability potential.

Based on both concepts, nADICO for the norm representation and Dynamic Deontics for the dynamic norm understanding, we introduced an operationalisation based on reinforcement learning that marries expressiveness and accessibility with a continuous norm understanding that can accommodate social and individual characteristics (such as strategies associated with norm interpretation).

8.2.3 Methodological Contributions

An aspect of methodological value is the application of Interval Type-2 Fuzzy Sets from a sociological perspective (Chapter 7), specifically the multi-level analysis of normative understanding and alignment. We provided a comprehensive description of a refined systematic fuzzy set generation process along with a software module that facilitates clusters of norm understanding on different social levels. We incorporated density-based clustering to enable the supervised detection of social clusters on the intermediate level of social analysis, and introduced quantitative measures for the outcomes.

8.3 Limitations

The provided contributions come with limitations, which we discuss following the same structural layout that we used for contributions in Section 8.2.

8.3.1 Substantive Contributions

Experiment 'Informal Communication among Genoese Investors'

The experimental exploration of informal communication among investors (Subsection 4.2.1) is subject to a set of further extensions, namely the consideration of dynamic network formation and adaptation (see e.g. Bravo et al. (2012)). The current conceptualisation is confined to a static network generation at the simulation onset. A dynamic generation would further allow the more realistic trust-based adaptation of relationship networks (Skyrms and Pemantle (2000)) or the consideration of social comparison (Zschache (2012)).

The experiment could further be refined by making operator behaviour dynamic. In the current setup cheaters always cheat, while non-cheaters are compliant. Furthermore, the concept of trust is modelled as a singular value that describes the agent's likeliness to respond truthfully to any incoming request. However, more realistically, trust should be associated with individual fellow investors, as opposed to being of global nature. Alternatively, the model could apply a refined differentiation of societal trust and individual-centred reputation.

Experiment 'Apprenticeship System in the Maghribī Traders Coalition'

The Maghribī apprenticeship model (Subsection 4.2.2) can experience refinement from multiple perspectives. In the current model, the detection of cheating can be established with full certainty. A further aspect is, similar to the model of informal communication among Genoese investors, the static nature of the agents' behaviour. Cheaters do not act strategically but follow the parameterised cheating probability. Beyond this a refined conceptualisation of jāh, or status, could be introduced; in the current model jāh is monotonically increasing. Instead, it could find representation in the form of endorsements (Alam et al., 2010), which would offer a more accurate representation of the social nature of jāh.

At the current stage, our investigation of the apprenticeship model focuses on the filtering function of cheaters, which are detected based on their close monitoring. However, we do not consider skill assessments in order to prevent traders from entry to the coalition. With the intent of providing a more realistic representation we would need to consider this aspect.

Another aspect that is neglected in this reconstruction of the Maghribī trader society and the apprenticeship system is the exclusive focus on the ṣuḥba as the prototypical informal institution. However, further investigations could explore the complementary use with other, formal institutional instruments (such as the shirka) and the effect on the functioning of the apprenticeship system and coalition at large. This is particularly relevant since formal institutional instruments had evidently been employed in addition to the ṣuḥba to a non-negligible extent (see Figure 3.5 in Subsection 3.1.3).

Generally, however, the refinement of all models relies on future contributions from the field of Geniza research, such as providing statistical information about general cheater levels and more accurate structural information about the trader networks.

Experiment 'Evolving Norm Understanding under Maghribī and Genoese Role Conceptualisations'

The exploration of the behavioural impact of role conceptualisations (Section 6.3) has several limitations. It features a puristic modelling approach that abstracts feedback of trade interactions into pay-offs. Though concentrating on the experiential establishment of normative understanding, the abstraction of interactions into pay-offs exploits the conceptual richness of the agent concept to a far lesser extent than the previous models. However, a refinement would require deeper knowledge about the historical activities. This instance of the model concentrates on the claim that the development of normative understanding can depend on different role conceptualisations; accuracy is not of central concern. However, alternative pay-off configurations could be explored, but those would benefit from the availability of more accurate historical information.

Looking at the application of the Dynamic Deontics concept, the model only relies on its use to develop normative understanding. Though the concept offers a schematic representation, our agents do not reflect or act upon this understanding. In this scenario, the Dynamic Deontics operationalisation process extracts and translates fine-grained experiential knowledge into a more comprehensive representation without feeding it back into the decision-making process. While this does not exploit the full potential of the Dynamic Deontics concept, with respect to our model it is intentional, since it shows how the Dynamic Deontics concept can be attached to existing models and used as a vehicle for interpretation without necessarily becoming an integral part of the model itself.

8.3.2 Conceptual Contributions

Shifting to the conceptual aspects, we put particular focus on the institution representations developed in this work. Since the institutional grammar and the continuous norm concept have been developed from an application-/demand-oriented perspective (i.e. to represent aspects relevant to our case study), they have not been explored to their full possibilities. Though able to capture and differentiate formal and informal institutions, the institutional grammar (Section 5.3) has not been operationalised for rule formation processes that afford collective action. We further did not systematically exploit the full potential of the introduced horizontal and vertical nesting capabilities for the representation of institutional complexity.

Limitations with respect to the continuous norm representation (Chapter 6) include the limited grounding of deontic compartments. Open questions with respect to the representational aspects of the Dynamic Deontics concept can be organised into three categories:

- Shape – How are compartments allocated along the deontic range? Are they equally-sized, do they change progressively, or do they have individual sizes? Do they overlap?
- Orientation – Are there alternative orientations of compartments beyond the orientation at the centre? Can we offer a better representation of subjective experience (which the symmetric orientation failed to do)?
- Labelling – Which terms most appropriately reflect the meaning of the deontic compartments?

Overall, all representations concentrate on the subjective institution perspective. Our approach neglects what we characterised as the objective domain of institutions (see Subsection 2.1.3), such as markets, organisations, or the State, be it as a precursor or as a result of interactions. Instead we concentrate, as far as the built-in assumptions permit, on a relatively puristic endogenous perspective on institution formation.

8.3.3 Methodological Contributions

The multi-level analysis of results using Interval Type-2 Fuzzy Sets concentrates on the analytical perspective but, similar to the case for Dynamic Deontics, does not find application beyond the extraction of normative understanding. The membership function generation process is geared towards accessibility, not fully-fledged fuzzy sets with rule-based evaluation. Experimenters can derive what a particular input value 'means' with respect to a given level of aggregation. However, at this point information is not fed back to the observed agents, and can thus not be used to model collective action processes (e.g. to establish formal institutions). Similarly, the analysis concentrates on the singular dimension of normative understanding without considering further domains (e.g. wealth, social status, etc.), which would be of particular use when relying on the generated fuzzy sets for decision-making.

8.4 Future Directions

While many of the limitations directly point to future improvements, the explorative nature of the later contributions gives those a particularly strong generalisability potential.

8.4.1 Nested ADICO

The nested institutional grammar offers a wide-ranging potential for future use, beyond the application in this work. Though we concentrated on its use in conjunction with Dynamic Deontics, its operation is independent from the latter, leaving a variety of directions for future exploration.

Generic Institution Representation

In the field of normative multi-agent systems, an immediate extension is its operationalisation for the representation of formal institutions (such as rules), as well as the informal end of the spectrum (conventions). Furthermore, it could be applied to the representation of complex institutional scenarios that highlight interdependence and great level of detail in order exploit the potential of nADICO's nesting capabilities. Beyond this, the institutional grammar represents a candidate for a generic institution representation that can find use for the internal representation, but also serve as a message container for the communication of institutions beyond the purely normative domain.

Domain-dependent Customisation

Its generic nature further makes it a starting point for application- or domain-dependent refinement in order to capture specific aspects. This could include a more refined action representation, e.g. including *action object* and *action target*,[4] and customised representations of context, catering for applications as diverse as environmental sensing or behaviour specification in the area of robotics.

[4] This aspect has been discussed in Subsection 5.3.3.

Behaviour Generalisation Processes

nADICO's versatility extends down to the atomic level. The format is agnostic of the representation of individual action instances or high-level institutions. Accompanied with a generalisation process, individual action statements could be transformed into institution representations. The structural elements could thus be used as a generic action representation extending from a singular action via observed conventions up to the level of norms and rules.

8.4.2 Dynamic Deontics

At the current stage, the concept of Dynamic Deontics potentially bears the greatest potential for further exploration and grounding, with central aspects outlined in the previous section (Section 8.3). Refinements can generally be assorted into the sociological and computational domain.

Sociological and Cultural Grounding

Establishing a solid grounding of terms, compartment allocation, nature, and range adjustment processes requires extensive empirical studies. Doing so, Dynamic Deontics can serve as a unified vehicle to represent the degree of prescriptiveness for associated studies, ultimately mitigating cross-cultural and -lingual representations.

Human-Computer Interaction

Establishing a sociological grounding would establish a reusable repository of normative labels that could be drawn upon in *human-computer interaction* for the purposes of mapping the computational representations of prohibitions, obligations, and permissions into human-readable statements, and vice versa.

Emotions

Apart from terminological mapping, the application of Dynamic Deontics, especially when combined with nADICO as a rich action representation, can allow the mapping of processed actors, actions, and events along with associated feedback onto computational *representations of emotions* (e.g. using the Ortony-Clore-Collins (OCC) model (Ortony et al., 1988)).

Personality and Culture

In line with this approach, the varying operationalisation strategies for Dynamic Deontics (rational, opportunistic, optimistic, pessimistic) allow for the abstract representation of personal or cultural traits, such as uncertainty avoidance or short-term vs. long-term orientation (Hofstede et al., 2010). Such individualisation could further extend to strategies involved in expanding and contracting the deontic range based on personality types. Overall, Dynamic Deontics provide an integrated mechanism that considers social (or societal) and personal characteristics.

Moral Agents

Apart from the problem of labelling, the Dynamic Deontics concept offers a potential generic representation of how we understand and interpret the world in terms of beliefs, thus following a subjectivist perspective. The expanding and contracting deontic range points to what an agent perceives as 'good', or desirable, and 'bad'. Following this interpretation, the Dynamic Deontics can contribute to a pragmatic representation of morality in agents (Himma, 2009; Wallach and Allen, 2010).

8.4.3 Multi-Level Analysis with IT2FS

Similar to the previously mentioned concepts, the use of fuzzy sets for the purposes of multi-level analysis has the potential for more general application. This concerns the use with further simulation cases that produce interval-based output, whether in combination with Dynamic Deontics or by other means.

Multi-Dimensional Inputs

An aspect that has not been explored is the use of multi-dimensional fuzzy inputs, such as the representation of social structures (e.g. social proximity between interacting parties) as further input, an aspect that could enrich the context-sensitive use of normative understanding. This could of particular relevance for sociological inquiry, since the challenge within this discipline lies in the analysis of complex relationships amongst a wide range of factors (e.g. social influence, demographic characteristics).

Feedback into Individuals' Decision-Making

As mentioned in Section 8.3, the current application is of purely interpretative nature; the derived understanding is not fed back into the agents' reasoning process. Closing this feedback loop would enable the representation of second-order emergence as discussed in Subsection 7.6.3, as well as to impose norms or rules in a top-down fashion.

In the context of normative multi-agent systems, the nature of fuzzy sets lends itself well to the representation of gradual sanctioning. Individuals could thus evaluate input values against the established fuzzy sets and use the result for the choice of a normatively appropriate sanction. This further provides the basis for the systematic consideration of characteristics of social comparison (Festinger, 1954).

Multi-Level Clustering on Meso-Level

Immediate extensions of the approach could include the optimisation of the clustering mechanism to minimise user intervention. This could be facilitated by automated detection of sub-groups on different social levels by applying hierarchical clustering techniques. This approach is particularly useful for the analysis of intra- and inter-group relationships in complex social systems, and thus applicable to peace and conflict studies as well as research into organisational structures in social systems in the wider sense.

Fuzzy Concepts in Social Computing

Applications of IT2FS for social concepts have so far been limited. Their use can naturally extend outside the domain of normative multi-agent systems or agent-based simulation, such as the area of *social computing* in which the interpretation and application of fuzzy data (e.g. ambiguous or vague human input) is a central concern. Structuring input based on fuzzy techniques can facilitate refined context-sensitive responses. A further interesting prospect could be the refinement of software based on the continuous adaptation of derived fuzzy sets based on user behaviour.

8.5 Conclusion

This work showcases the use of agent-based modelling and simulation for the purpose of institutional analysis. Our approach was primarily driven by a historical scenario, the Maghribī Traders Coalition, which Avner Greif contrasted to the contemporary medieval Genoese traders using the approach of Analytic Narratives (see Subsection 2.2.3). We reviewed this model with respect to selected underlying assumptions. Applying ABM, we augmented the original scenario by exploring details that had either been ignored by high-level assumptions (secrecy among Genoese traders, different role conceptualisations in both societies) or had not been previously explored in the literature (Maghribīan apprenticeship system). Applying an agent-based approach facilitated a shift towards the reflection of social characteristics that had not found consideration in earlier work, such as the notion of status (jāh);[5] but more importantly, we explored the notion of experientially driven normative behaviour. As such this work emphasised the analysis of institutions and their stabilisation from a behavioural perspective, reducing the rationality focus required by Greif's original analysis. The assumption of rationality should not be dismissed entirely, but may require further contextualisation or be of secondary role when looking at specific problems (e.g. the apprenticeship system).[6] As such we believe that the application of ABM principles, along with the structural representation of institutions and analytical tools, provide suitable means to model and analyse institutions on arbitrary levels of granularity. This approach offers a valuable analytical tool for the exploration of sub-problems that cannot be captured on higher levels of abstraction.

Consequently, the tools developed in the course of this case-oriented exploration, the institution modelling constructs as well as the analytical extension based on Interval Type-2 Fuzzy Sets, are thus intentionally of general nature so as to allow reuse within and beyond the domain of institutional analysis. Applying such a toolset to a concrete case as done in this work not only advocates the movement towards an emphasis of behavioural aspects in the context of institutional analysis, but also complements the traditional comparative-static approaches, such as the IAD framework (Ostrom et al., 1994) and Analytic Narratives (Bates et al., 1998) (see Subsection 2.2.3). This approach highlights a dynamic perspective that furthers the identification and more in-depth understanding of institutions by emergence as opposed to explicit specification. We show that the use of agent-based modelling and simulation can offer promising contributions to the field of institutional analysis, leveraging a dynamic behavioural perspective on institutions.

We believe that particularly those latter technical contributions, the institution modelling constructs as well as the analytical extension, take important steps to establish ABM as a means to further our understanding of social processes. Adding to the strengths of ABM, the modelling of complex social systems, our contributions enrich its explanatory means to retrace what *did happen*, but also to suggest what *might have happened*, and potentially, to look ahead and predict what *might happen in the future*.

[5] It reflected an essential characteristic of the Maghribīs but was virtually irrelevant in its contrasted society, the Genoese. Its importance was emphasised by the fact that Maghribī trade agreements were not based on monetary remuneration but repaid by reciprocal trade obligations (Subsection 3.1.3).

[6] With respect to the apprenticeship system, we concluded that the use of dedicated mentors would not have minimised the absolute cheater numbers in all cases, but assured a comparatively cheater-free trader coalition to the benefit of all members.

References

Acemoglu, D. and Robinson, J. (2012). *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. Crown Business, New York (NY).

Acharya, A. (2004). How Ideas Spread: Whose Norms Matter? Norm Localization and Institutional Change in Asian Regionalism. *International Organization*, 58(2):239–275.

Acharya, T. and Ray, A. K. (2005). Fuzzy Set Theory in Image Processing. In *Image Processing: Principles and Applications*, pages 209–226. John Wiley & Sons, Hoboken (NJ).

Ackerman-Lieberman, P. I. (2007). *A Partnership Culture. Jewish Economic and Social Life Seen through the Legal Documents of the Cairo Geniza*. PhD thesis, Princeton University Press, Princeton (NJ).

Ackerman-Lieberman, P. I. (2012). Commercial Forms and Legal Norms in the Jewish Community of Medieval Egypt. *Law and History Review*, 30(4):1007–1052.

Ackerman-Lieberman, P. I. (2014). *The Business of Identity*. Stanford University Press, Stanford (CA).

Act of Parliament, United Kingdom (1967). Sexual Offences Act 1967. http://www.legislation.gov.uk/ukpga/1967/60/pdfs/ukpga_19670060_en.pdf. Accessed on: 26th October 2014.

ACT-R Research Group (2015). ACT-R. http://act-r.psy.cmu.edu. Accessed on: 25th August 2015.

Airaldi, G. (1969). *Le carte di Santa Maria delle Vigne di Genova (1103-1392)*. Genoa.

Airaldi, G. (1997). The Genoese Art of Warfare. In Agius, D. A. and Netton, I. R., editors, *Across the Mediterranean frontiers: Trade, Politics and Religions, 650-1450*. Brepols, Turnhout.

Airiau, S., Sen, S., and Villatoro, D. (2014). Emergence of conventions through social learning. *Autonomous Agents and Multi-Agent Systems*, 28(5):779–804.

Alam, S. J., Geller, A., Meyer, R., and Werth, B. (2010). Modelling Contextualized Reasoning in Complex Societies with "Endorsements". *Journal of Artificial Societies and Social Simulation*, 13(4):6.

Alchian, A. (1950). Uncertainty, Evolution and Economic Theory. *Journal of Political Economy*, 58(3):211–221.

Alchourròn, C. and Bulygin, E. (1971). *Normative Systems*. Springer, Heidelberg.

Aldewereld, H., Dignum, F., García-Camino, A., Noriega, P., Rodríguez-Aguilar, J., and Sierra, C. (2007). Operationalisation of Norms for Electronic Institutions. In Noriega, P., Vázquez-Salceda, J., Boella, G., Boissier, O., Dignum, V., Fornara, N., and Matson, E., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems II*, volume 4386 of *Lecture Notes in Computer Science*, pages 163–176. Springer, Berlin.

Aldewereld, H., Grossi, D., Vázquez-Salceda, J., and Dignum, F. (2006). Designing Normative Behaviour Via Landmarks. In Boissier, O., Padget, J., Dignum, V., Lindemann, G., Matson, E., Ossowski, S., Sichman, J., and Vázquez-Salceda, J., editors, *Coordination, Organizations, Institutions, and Norms in Multi-Agent Systems*, volume 3913 of *Lecture Notes in Computer Science*, pages 157–169. Springer, Berlin.

Amblard, F., Sichman, J. S., Takadama, K., Sawyer, K., Antunes, L., Gilbert, N., Davidsson, P., Conte, R., and Moss, S. (2014). Multi-Agent-Based Simulation (MABS). http://www.pcs.usp.br/ mabs/. Accessed on: 1st October 2014.

Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Annals of Mathematical Statistics*, 23(2):193–212.

Andersson, K. P. (2004). Who Talks with Whom? The Role of Repeated Interaction in Decentralized Forest Governance. *World Development*, 32(2):233–249.

Andrighetto, G., Campenni, M., Cecconi, F., and Conte, R. (2010a). The Complex Loop of Norm Emergence: A Simulation Model. In Takadama, K., Cioffi-Revilla, C., and

Deffuant, G., editors, *Simulating Interacting Agents and Social Phenomena*, pages 19–35. Springer, Berlin.

Andrighetto, G., Villatoro, D., and Conte, R. (2010b). Norm internalization in artificial societies. *AI Communications*, 23(4):325–339.

Anscombe, G. E. M. (1958). Modern Moral Philosophy. *Philosophy*, 33(124).

Aoki, M. (2001). *Toward a Comparative Institutional Analysis*. MIT Press, Cambridge (MA).

Archer, M. S. (1979). *Social Origins of Educational Systems*. Sage, London.

Archer, M. S. (1988). *Culture and Agency: The Place of Culture in Social Theory*. Cambridge University Press, Cambridge (UK).

Archer, M. S. (1995). *Realist Social Theory: The Morphogenetic Approach*. Cambridge University Press, Cambridge (UK).

Aref, A. and Tran, T. (2014). Using Fuzzy Logic and Q-Learning for Trust Modeling in Multi-agent Systems. In *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, pages 59–66.

Arredondo, T. V., Freund, W., Navarro-Guerrero, N., and Castillo, P. (2013). Fuzzy Motivations in a Multiple Agent Behaviour-Based Architecture. *International Journal of Advanced Robotic Systems*, 10:313.

Artikis, A., Kamara, L., Pitt, J., and Sergot, M. (2005). A Protocol for Resource Sharing in Norm-Governed Ad Hoc Networks. In Leite, J., Omicini, A., Torroni, P., and Yolum, P., editors, *Declarative Agent Languages and Technologies II*, volume 3476 of *Lecture Notes in Computer Science*, pages 221–238. Springer, Berlin.

Astren, F. (2012). Goitein, Medieval Jews, and the "New Mediterranean Studies". *The Jewish Quarterly Review*, 102(4):513–531.

Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books, New York (NY).

Axelrod, R. (1986). An Evolutionary Approach to Norms. *The American Political Science Review*, 80(4):1095–1111.

Axelrod, R. (1997a). Advancing the Art of Simulation in the Social Sciences. In Conte, R., Hegselmann, R., and Terna, P., editors, *Simulating Social Phenomena*, pages 21–40. Springer, Berlin.

Axelrod, R. (1997b). Advancing the Art of Simulation in the Social Sciences. *Complexity*, 3(2):16–22.

Axelrod, R. (2006). Simulation in the social sciences. In Renard, J.-P., editor, *Handbook of Research on Nature Inspired Computing for Economics and Management*, pages 90–100. Idea Group, Hershey.

Aydinonat, N. E. (2006). Institutions: Theory, History and context-specific Analysis. *History of Economic Ideas*, XIV(3):145–158.

Ayres, C. E. (1944). *The Theory of Economic Progress*. University of North Carolina Press, Chapel Hill (NC).

Babbage, C. (1835). *On the Economy of Machinery and Manufactures*, volume 4. Charles Knight, London.

Balard, M. (2001). *Croisades et orient Latin (XIème-XIVème siècles)*. Armand Colin, Paris.

Baldwin, S. (1968). *Business in the Middle Ages*. Cooper Square Publishers, New York (NY).

Balev, S., Baudry, J., Dutot, A., Pigné, Y., and Savin, G. (2014). GraphStream – A Dynamic Graph Library. http://graphstream-project.org/. Accessed on: 28th August 2014.

Balke, T. and Gilbert, N. (2014). How Do Agents Make Decisions? A Survey. *Journal of Artificial Societies and Social Simulation*, 17(4):13.

Balke, T. and Villatoro, D. (2012). Operationalization of the sanctioning process in utilitarian artificial societies. In Cranefield, S., van Riemsdijk, M., Vázquez-Salceda, J., and Noriega, P., editors, *Coordination, Organizations, Institutions, and Norms in Agent System VII*, volume 7254 of *Lecture Notes in Computer Science*, pages 167–185. Springer, Berlin.

Bandura, A., Ross, D., and Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models. *Journal of Abnormal and Social Psychology*, 63(3):575–582.

Barabási, A.-L. and Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512.

Barry, B. (1989). *A Treatise on Social Justice: Theories of Justice*. University of California Press, Berkeley (CA).

Barry, B. (1995). *Justice as Impartiality*, volume A Treatise on Social Justice, Vol. II. Clarendon Press, Oxford (UK).

Basurto, X., Kingsley, G., McQueen, K., Smith, M., and Weible, C. M. (2010). A Systematic Approach to Institutional Analysis: Applying Crawford and Ostrom's Grammatical Syntax. *Political Research Quarterly*, 63(3):523–537.

Bates, R. H., Greif, A., Levi, M., Rosenthal, J.-L., and Weingast, B. R. (1998). *Analytic Narratives*. Princeton University Press, Princeton (NJ).

Baumeister, R. F., Bratslavsky, E., and Finkenauer, C. (2001). Bad is Stronger than Good. *Review of General Psychology*, 5(4):323–370.

Bedau, M. A. (1997). Weak emergence. In Tomberlin, J., editor, *Philosophical Perspectives: Mind, Causation, and World*, pages 375–399. Blackwell, Malden (MA).

Belton, R. K. (2005). Competing Definitions of the Rule of Law. Implications for Practitioners. Rule of Law Series 55, Carnegie Endowment for International Peace, Washington, D.C.

Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press, New York (NY).

Bloomfield, L. (1933). *Language*. Holt, New York (NY).

Boella, G. and van der Torre, L. (2006). A Logical Architecture of a Normative System. In Goble, L. and Meyer, J.-J., editors, *Deontic Logic and Artificial Normative Systems*, volume 4048 of *Lecture Notes in Computer Science*, pages 24–35. Springer Berlin Heidelberg.

Boella, G., van der Torre, L., and Verhagen, H. (2007). Introduction to Normative Multi-Agent Systems. In Boella, G., van der Torre, L., and Verhagen, H., editors, *Normative Multi-agent Systems, Dagstuhl Seminar Proceedings 07122*, Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

Boissier, O., Balbo, F., and Badeig, F. (2011). Controlling Multi-party Interaction within Normative Multi-agent Organizations. In De Vos, M., Fornara, N., Pitt, J., and Vouros, G., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems VI*, volume 6541 of *Lecture Notes in Computer Science*, pages 357–376. Springer, Berlin.

Boissier, O., Bordini, R. H., Hübner, J. F., Ricci, A., and Santi, A. (2015). JaCaMo Project. http://jacamo.sourceforge.net/. Accessed on: 1st September 2015.

Bourdieu, P. (1977). *An Outline of a Theory of Practice*. Cambridge University Press, London.

Bourdieu, P. (1981). Men and Machines. In Knorr-Cetina, K. and Circourel, A. V., editors, *Advances in Social Theory and Methodology*, pages 304–317. Routledge & Kegan Paul, London.

Bourdieu, P. (1984). *Distinction: A Social Critique of Judgment and Taste*. Harvard University Press, Cambridge (MA). (Originally published: 1979).

Bourdieu, P. (1985). The Genesis of the Concepts of Habitus and Field. *Sociocriticism*, 2(2):11–24.

Bourdieu, P. (1990a). *In Other Words*. Polity Press, Cambridge (UK).

Bourdieu, P. (1990b). *The Logic of Practice*. Stanford University Press, Stanford (CA).

Bowles, S. and Gintis, H. (2004). The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations. *Theoretical Population Biology*, 65(1):17–28.

Boyd, R. (1988). The Evolution of Sizable Groups. *Journal of Theoretical Biology*, 132(3):337–356.

Braithwaite, J. and Drahos, P. (2000). *Global Business Regulation*. Cambridge University Press, Cambridge (UK).

Bratman, M. (1987). *Intentions, Plans, and Practical Reason*. Harvard University Press, Cambridge (MA).

Braudel, F. (1996). *The Mediterranean and the Mediterranean World in the Age of Philip II*, volume 2 volumes. University of California Press, Berkeley (CA). Originally published: 1949.

Bravo, G., Squazzoni, F., and Boero, R. (2012). Trust and Partner Selection in Social Networks: An Experimentally Grounded Model. *Social Networks*, 34(4):481–492.

Brinberg, D. and McGrath, J. (1985). *Validity and the Research Process*. SAGE Publishing Co., Beverly Hills (CA).

Brooks, L., Iba, W., and Sen, S. (2011). Modeling the emergence and convergence of norms. In Walsh, T., editor, *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, volume 1 of *IJCAI '11*, pages 97–102, Barcelona.

Brousseau, E. and Raynaud, E. (2006). The Economics of Private Institutions: An Introduction to the Dynamics of Institutional Frameworks and to the Analysis of Multilevel Multi-Type Governance. *Available at SSRN: http://ssrn.com/abstract=920225*.

Burt, R. S. (1982). *Toward a Structural Theory of Action: Network Models of Social Structure, Perception, and Action*. Academic Press, New York (NY).

Bushouse, B. K. (2011). Governance Structures: Using IAD to Understand Variation in Service Delivery for Club Goods with Information Asymmetry. *Policy Studies Journal*, 39(1):99–113.

Byrne, E. H. (1916). Commercial Contracts of the Genoese in the Syrian Trade of the Twelfth Century. *Quarterly Journal of Economics*, 31:128–170.

Cagan, P. (1956). The monetary dynamics of hyperinflation. In Friedman, M., editor, *Studies in the Quantity Theory of Money*. University of Chicago Press, Chicago (IL).

Caldas, J. C. and Coelho, H. (1999). The Origin of Institutions: socio-economic processes, choice, norms and conventions. *Journal of Artificial Societies and Social Simulation*, 2(2):1.

Calvert, R. L. (1995). Rational actors, equilibrium, and social institutions. In Knight, J. and Sened, I., editors, *Explaining Social Institutions*, pages 57–93. University of Michigan Press, Ann Arbor (MI).

Campos, J., López-Sánchez, M., Rodrıguez-Aguilar, J., and Esteva, M. (2009). Formalising situatedness and adaptation in electronic institutions. In Hübner, J., Matson, E., Boissier, O., and Dignum, V., editors, *Coordination, Organizations, Institutions and Norms in Agent Systems IV*, volume 5428 of *Lecture Notes in Computer Science*, pages 126–139. Springer, Berlin.

Cardoso, H. and Oliveira, E. (2009). A Context-Based Institutional Normative Environment. In Hübner, J., Matson, E., Boissier, O., and Dignum, V., editors, *Coordination, Organi-*

zations, Institutions and Norms in Agent Systems IV, volume 5428 of *Lecture Notes in Computer Science*, pages 140–155, Berlin. Springer.

Carley, K. M., Prietula, M. J., and Lin, Z. (1998). Design Versus Cognition: The interaction of agent cognition and organizational design on organizational performance. *Journal of Artificial Societies and Social Simulation*, 1(3):4.

Castelfranchi, C. (1995). Guarantees for autonomy in cognitive agent architecture. In Wooldridge, M. J. and Jennings, N. R., editors, *Proceedings of the Workshop on Agent Theories, Architectures and Languages, ATAL '94*, volume 890 of *Lecture Notes in Artificial Intelligence*, pages 56–70, Berlin. Springer.

Castelfranchi, C. (2003). Formalising the informal? dynamic social order, bottom-up social control, and spontaneous normative relations. *Journal of Applied Logic*, 1(1–2):47–92.

Castelfranchi, C. (2014). Minds as social institutions. *Phenomenology and the Cognitive Sciences*, 13(1):121–143.

Castelfranchi, C., Conte, R., and Paolucci, M. (1998). Normative reputation and the costs of compliance. *Journal of Artificial Societies and Social Simulation*, 1(3):3.

Ceccardi, C. R. and Monleone, G., editors (1923–29). *Annali Genovesi di Caffaro e dei suoi Continuatori. 1099–1240*. Municipio di Genova, Genoa. 4 Volumes.

Chevaleyre, Y., Endriss, U., Lang, J., and Maudet, N. (2007). A Short Introduction to Computational Social Choice. In van Leeuwen, J., Italiano, G., van der Hoek, W., Meinel, C., Sack, H., and Plášil, F., editors, *Proceedings of the 33rd Conference on Current Trends in Theory and Practice of Computer Science*, volume 4362 of *Lecture Notes in Computer Science*, pages 51–69, Berlin. Springer.

Choi, B.-I. and Rhee, F. C.-H. (2009). Interval type-2 fuzzy membership function generation methods for pattern recognition. *Information Sciences*, 179(13):2102–2122.

Chomsky, N. (2010). Some simple evo-devo theses: How true might they be for language? In Larson, R. K., Deprez, V. M., and Yamakido, H., editors, *Approaches to the Evolution of Language*, pages 45–62. Cambridge University Press, Cambridge (UK).

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6):1015–1026.

Cialdini, R. B. and Trost, M. R. (1998). Social influence: Social norms, conformity, and compliance. In Gilbert, D. T., Fiske, S. T., and Lindzey, G., editors, *The Handbook of Social Psychology*, volume 2, pages 151–192. Oxford University Press, New York (NY), 4th edition.

Cioffi-Revilla, C. (1981). Fuzzy sets and models of international relations. *American Journal of Political Science*, 25(1):129–159.

Citarella, A. O. (1971). A puzzling question concerning the relation between the Jewish communities of Christian Europe and those represented in the Geniza Documents. *Journal of the American Oriental Society*, 91(3):390–397.

Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data. *SIAM Rev.*, 51(4):661–703.

Coase, R. H. (1937). The Nature of the Firm. *Econometrica*, 4(16):386–405.

Coase, R. H. (1959). The Federal Communications Commission. *Journal of Law and Economics*, 2(1):1–40.

Cohen, M. (2013). A Partnership Gone Bad: Business Relationships and the Evolving Law of the Cairo Geniza Period. *Journal of the Economic and Social History of the Orient*, 56(2):218–263.

Cohen, M. R. (2006). Geniza for Islamicists, Islamic Geniza, and the "New Cairo Geniza". *Harvard Middle Eastern and Islamic Review*, 7:129–145.

Cohen, P. R. (1985). *Heuristic Reasoning about Uncertainty: An Artificial Intelligence Approach*. Pitman Publishing, Marshfield (MA).

Coleman, J. S. (1987). Norms as social capital. In Radnitzky, G. and Bernholz, P., editors, *Economic Empirialism*. Paragon, New York (NY).

Coleman, J. S. (1990). *Foundations of Social Theory*. Harvard University Press, Cambridge (MA).

Committee on Homosexual Offences and Prostitution (1957). *Report of the Committee on Homosexual Offences and Prostitution*. Her Majesty's Stationery Office, London.

Commons, J. R. (1936). Institutional Economics. *American Economic Review*, 21(4):648–657.

Commons, J. R. (1968). *Legal Foundations of Capitalism*. University of Wisconsin Press, Madison (WI).

Conte, R., Andrighetto, G., and Campennl, M., editors (2014). *Minding Norms: Mechanisms and dynamics of social order in agent societies*. Oxford University Press, Oxford (UK).

Conte, R. and Castelfranchi, C. (1995). Understanding the functions of norms in social groups through simulation. In Gilbert, N. and Conte, R., editors, *Artificial Societies. The Computer Simulation of Social Life*, pages 252–267. UCL Press, London.

Conte, R. and Castelfranchi, C. (1999). From conventions to prescriptions. towards an integrated view on norms. *Artificial Intelligence and Law*, 7(4):323–340.

Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., Loreto, V., Moat, S., Nadal, J.-P., Sanchez, A., Nowak, A., Flache, A., San Miguel, M., and Helbing, D. (2012). Manifesto of computational social science. *The European Physical Journal Special Topics*, 214(1):325–346.

Conte, R., Gilbert, N., and Sichman, J. S. (1998). MAS and Social Simulation: A suitable commitment. In Sichman, J. S., Conte, R., and Gilbert, N., editors, *Multi-Agent Systems and Agent-Based Simulation*, volume 1534 of *Lecture Notes in Artificial Intelligence*, pages 1–9, Berlin. Springer.

Conte, R. and Paolucci, M. (2002). *Reputation in Artificial Societies: Social Beliefs for Social Order*, volume 6 of *Multiagent Systems, Artificial Societies, and Simulated Organizations*. Springer, Berlin.

Cook, M. (2003). *Forbidding Wrong in Islam*. Cambridge University Press, Cambridge (UK).

Cooter, R. and Landa, J. T. (1984). Personal versus impersonal trade: The size of trading groups and contract law. *International Review of Law and Economics*, 4:15–22.

Corapi, D., De Vos, M., Padget, J., Russo, A., and Satoh, K. (2011). Norm refinement and design through inductive learning. In De Vos, M., Fornara, N., Pitt, J., and Vouros, G., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems VI*, volume 6541 of *Lecture Notes in Computer Science*, pages 77–94. Springer, Berlin.

Cortell, A. P. and Davis Jr., J. W. (2000). Understanding the Domestic Impact of International Norms: A Research Agenda. *International Studies Review*, 2(1):65–87.

Cortes-Rios, J. C., Gomez-Ramirez, E., de-la Vega, H. A. O., Castillo, and Melin, P. (2014). Optimal design of interval type 2 fuzzy controllers based on a simple tuning algorithm. *Applied Soft Computing*, 23:270–285.

Cranefield, S. (2006). A Rule Language for Modelling and Monitoring Social Expectations in Multi-agent Systems. In Boissier, O., Padget, J., Dignum, V., Lindemann, G., Matson, E., Ossowski, S., Sichman, J., and Vázquez-Salceda, J., editors, *Coordination, Organizations, Institutions, and Norms in Multi-Agent Systems*, volume 3913 of *Lecture Notes in Computer Science*, pages 246–258. Springer, Berlin.

Cranefield, S. (2007). Modelling and Monitoring Social Expectations in Multi-agent Systems. In Noriega, P., Vázquez-Salceda, J., Boella, G., Boissier, O., Dignum, V., Fornara, N., and Matson, E., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems II*, volume 4386 of *Lecture Notes in Computer Science*, pages 308–321. Springer, Berlin.

Cranefield, S. (2014). Agents and Expectations. In Balke, T., Dignum, F., van Riemsdijk, M. B., and Chopra, A. K., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems IX*, volume 8386 of *Lecture Notes in Computer Science*, pages 234–255. Springer, Berlin.

Crawford, S. E. and Ostrom, E. (1995). A Grammar of Institutions. *The American Political Science Review*, 89(3):582–600.

Crawford, S. E. and Ostrom, E. (2005). A Grammar of Institutions. In *Understanding Institutional Diversity*, chapter 5, pages 137–174. Princeton University Press, Princeton (NJ).

Criado, N., Julián, V., Botti, V., and Argente, E. (2010). A Norm-Based Organization Management System. In Padget, J., Artikis, A., Vasconcelos, W., Stathis, K., da Silva, V., Matson, E., and Polleres, A., editors, *Coordination, Organizations, Institutions and Norms in Agent Systems V*, volume 6069 of *Lecture Notes in Computer Science*, pages 19–35. Springer, Berlin.

da Silva, V. and Zahn, J. (2014). Normative Conflicts that Depend on the Domain. In Balke, T., Dignum, F., van Riemsdijk, M. B., and Chopra, A. K., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems IX*, volume 8386 of *Lecture Notes in Computer Science*, pages 311–326. Springer, Berlin.

da Silva, V. T. (2007). Implementing Norms that Govern Non-Dialogical Actions. In Boella, G., van der Torre, L., and Verhagen, H., editors, *Normative Multi-agent Systems, Dagstuhl Seminar Proceedings 07122*, Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

da Silva Figueiredo, K., Torres da Silva, V., and de Oliveira Braga, C. (2011). Modeling Norms in Multi-agent Systems with NormML. In De Vos, M., Fornara, N., Pitt, J., and Vouros, G., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems VI*, volume 6541 of *Lecture Notes in Computer Science*, pages 39–57. Springer, Berlin.

Dattathreya, M. S., Singh, H., and Meitzler, T. (2012). Detection and Elimination of a Potential Fire in Engine and Battery Compartments of Hybrid Electric Vehicles. *Advances in Fuzzy Systems*, 2012:11 pages. Article ID 687652.

de Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., and Fehr, E. (2005). The Neural Basis of Altruistic Punishment. *Science*, 305(5688):1254–1258.

de Roover, R. (1958). The Organization of Trade. In Postan, M. M., Rick, E. E., and Miltey, M., editors, *Cambridge Economic History of Europe*, volume 3, pages 42–118. Cambridge University Press, Cambridge (UK).

Department of Near Eastern Studies at Princeton University (2014). Princeton Geniza Project. http://etc.princeton.edu/genizaproject. Accessed on: 1st April 2014.

Diamond, J. M. (1997). *Guns, Germs, and Steel: The Fates of Human Societies*. W.W. Norton, New York (NY).

Diesing, P. (1966). Objectivism vs. Subjectivism in the Social Sciences. *Philosophy of Science*, 33(1/2):124–133.

Dignum, F., Morley, D., Sonenberg, E., and Cavedon, L. (2000). Towards socially sophisticated bdi agents. In *Proceedings of the Fourth International Conference on Multi-Agent Systems*, pages 111–118.

Drogoul, A., Amouroux, E., Caillou, P., Gaudou, B., Grignard, A., Marilleau, N., Taillandier, P., Vavasseur, M., Vo, D.-A., and Zucker, J.-D. (2013). Gama: Multi-level and complex environment for agent-based models and simulations. In *Proceedings of the 12th International Conference on Autonomous Agents and Multi-Agent Systems*, AAMAS '13, pages 1361–1362.

Durkheim, E. (1933). *The Division of Labour in Society*. Free Press, New York (NY). Originally published: 1893.

Durkheim, E. (1952). *Suicide*. Routledge & Kegan Paul, London. Originally published: 1897.

Durkheim, E. (1964). *The Rules of Sociological Method*. Free Press, New York (NY). Originally published: 1894.

Dybalova, D., Testerink, B., Dastani, M., and Logan, B. (2014). A framework for programming norm-aware multi-agent systems. In Balke, T., Dignum, F., van Riemsdijk, M. B., and Chopra, A. K., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems IX*, volume 8386 of *Lecture Notes in Computer Science*, pages 364–380. Springer, Berlin.

Edel, A. (1959). The concept of levels in social theory. In Gross, L., editor, *Symposium on Social Theory*, pages 167–195. Row Peterson, Evanston (IL).

Edmonds, B. and Moss, S. (2005). From KISS to KIDS – An 'Anti-simplistic' Modelling Approach. In Davidsson, P., Logan, B., and Takadama, K., editors, *Multi-Agent and Multi-Agent-Based Simulation*, volume 3415 of *Lecture Notes in Computer Science*, pages 130–144. Springer, Berlin.

Edwards, J. and Ogilvie, S. (2009). Contract Enforcement, Institutions and Social Capital: The Maghribi Traders Reappraised. Technical Report 0928, Cambridge Working Papers in Economics.

Edwards, J. and Ogilvie, S. (2012). Contract Enforcement, Institutions and Social Capital: The Maghribi Traders Reappraised. *The Economic History Review*, 65(2):421–444.

Edwards, J. and Ogilvie, S. (March 2008). Contract Enforcement, Institutions and Social Capital: the Maghribi Traders Reappraised. *CESifo Working Paper Series*, (2254). Available at SSRN: http://ssrn.com/abstract=1107801.

Elsenbroich, C. and Gilbert, N. (2014). *Modelling Norms*. Springer, Dordrecht.

Elster, J. (1989). Social Norms and Economic Theory. *The Journal of Economic Perspectives*, 3(4):99–117.

Encyclopædia Britannica (2009). Mediterranean Sea: depth contours and submarine features. http://www.britannica.com/EBchecked/media/614/The-Mediterranean-Sea. Accessed on: 20th July 2014.

Englemore, R. and Morgan, T., editors (1988). *Blackboard Systems*. Addison-Wesley, Wokingham.

Epstein, J. and Axtell, R. (1996). *Growing Artificial Societies - Social Science from the Bottom Up*. Brookings Institution, Press MIT Press, Washington D.C., Cambridge (MA).

Epstein, J. G., Möhring, M., and Troitzsch, K. (2003). Fuzzy-logical rules in a multi-agent system. In *SimSoc VI Workshop*, Groningen, The Netherlands. (Page 25).

Epstein, J. M. (2001). Learning to Be Thoughtless: Social Norms and Individual Computation. *Computational Economics*, 18(1):9–24.

Epstein, J. M. (2007). *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton University Press, Princeton (NJ).

Epstein, S. A. (1994). Secrecy and Genoese commercial practices. *Journal of Medieval History*, 20(4):313–325.

Epstein, S. A. (1996). *Genoa and the Genoese, 958–1528*. The University of North Carolina Press, Chapel Hill (NC).

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Simoudis, E., Han, J., and Fayyad, U., editors, *Second International Conference on Knowledge Discovery and Data Mining*, pages 226–231, Portland, Oregon. AAAI Press.

Esteva, M. (2003). *Electronic Institutions: from specification to development*. PhD thesis, Institut d'Investigació en Intel-ligència Artificial, Bellaterra, Spain.

Esteva, M., Rodríguez-Aguilar, J.-A., Sierra, C., Garcia, P., and Arcos, J. (2001). On the Formal Specification of Electronic Institutions. In Dignum, F. and Sierra, C., editors, *Agent Mediated Electronic Commerce*, volume 1991 of *Lecture Notes in Computer Science*, pages 126–147. Springer, Berlin.

Evertsz, R., Busetta, P., Pedrotti, M., Ritter, F., and Bittner, J. (2008). Cojack – achieving principled behaviour variation in a moderated cognitive architecture. In *Proceedings of the 16th Conference on Behavior Representation in Modeling and Simulation (BRIMS '08)*, Orlando (FL). University of Central Florida.

Ewert, U. C. and Sunder, M. (2011). Trading Networks, Monopoly and Economic Development in Medieval Northern Europe: an Agent-Based Simulation of Early Hanseatic Trade. In *European Historical Economics Society Conference, Dublin, 2011*.

Ewing, J. (1889). On hysteresis in the relation of strain to stress. *British Association Reports*, (502).

Fehr, E., Fischbacher, U., and Gächter, S. (2002). Strong Reciprocity, Human Cooperation and the Enforcement of Social Norms. *Human Nature*, 13:1–25.

Ferber, J. (1999). *Multi-Agent Systems - An Introduction to Distributed Artificial Intelligence*. Addison-Wesley, Harlow (UK).

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, 7(2):117–140.

Finnemore, M. and Sikkink, K. (1998). International norm dynamics and political change. *International Organization*, 52(4):887–917.

Flentge, F., Polani, D., and Uthmann, T. (2001). Modelling the emergence of possession norms using schemes. *Journal of Artificial Societies and Social Simulation*, 4(4):3.

Forrester, J. (1971). Counterintuitive Behavior of Social Systems. *Technology Review*, 73(3):52–68.

Foucault, M. (1977). *Discipline and Punish: The Birth of the Prison*. Pantheon Books, New York (NY).

Franklin, S. and Graesser, A. (1996). Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agent. In *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Language (ECAI)*, pages 21–25.

Franks, H., Griffiths, N., and Anand, S. S. (2014). Learning agent influence in MAS with complex social networks. *Autonomous Agents and Multi-Agent Systems*, 28:836–866.

Frantz, C. (2010). Micro-agents revisited: A Modern Reimplementation of the Micro-agent Layer of the Otago Agent Platform (OPAL). Master's thesis, University of Koblenz, Koblenz.

Frantz, C. (2014). Thesis Resources. http://thesis.christopherfrantz.org. Accessed on: 6th July 2014.

Frantz, C., Nowostawski, M., and Purvis, M. (2010). Multi-agent Platforms and Asynchronous Message Passing: Frameworks Overview. Information Science Discussion Paper 2010/07, Dept. of Information Science, University of Otago, New Zealand, Dunedin. http://hdl.handle.net/10523/892.

Frantz, C., Purvis, M. K., and Nowostawski, M. (2014a). Agent-Based Modeling of Information Transmission in Early Historic Trading. *Social Science Computer Review*, 32(3):393–416.

Frantz, C., Purvis, M. K., Nowostawski, M., and Savarimuthu, B. T. R. (2013). nADICO: A Nested Grammar of Institutions. In Boella, G., Elkind, E., Savarimuthu, B. T. R., Dignum, F., and Purvis, M. K., editors, *PRIMA 2013: Principles and Practice of Multi-Agent Systems*, volume 8291 of *Lecture Notes in Artificial Intelligence*, pages 429–436, Berlin. Springer.

Frantz, C., Purvis, M. K., Nowostawski, M., and Savarimuthu, B. T. R. (2014b). Modelling Institutions using Dynamic Deontics. In Balke, T., Dignum, F., van Riemsdijk, M. B., and Chopra, A. K., editors, *Coordination, Organizations, Institutions and Norms in Agent Systems IX*, volume 8386 of *Lecture Notes in Artificial Intelligence*, pages 211–233, Berlin. Springer.

Frantz, C., Purvis, M. K., Nowostawski, M., and Savarimuthu, B. T. R. (2015a). Analysing the Apprenticeship System in the Maghribi Traders Coalition. In Grimaldo, F. and Norling, E., editors, *Multi-Agent-Based Simulation XV*, volume 9002 of *Lecture Notes in Artificial Intelligence*, pages 180–196. Springer, Berlin.

Frantz, C., Purvis, M. K., Purvis, M. A., Nowostawski, M., and Lewis, N. D. (2014c). Fuzzy Modelling of Economic Institutional Rules. In Sadeghian, A. and Tahayori, H., editors, *Frontiers in Higher-Order Fuzzy Sets*, pages 87–129. Springer, New York.

Frantz, C., Purvis, M. K., Savarimuthu, B. T. R., and Nowostawski, M. (2014d). Analysing the Dynamics of Norm Evolution using Interval Type-2 Fuzzy Sets. In *WI-IAT '14 Pro-*

ceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), volume 3, pages 230–237.

Frantz, C., Purvis, M. K., Savarimuthu, B. T. R., and Nowostawski, M. (2014e). Modelling Dynamic Normative Understanding in Agent Societies. In Dam, H. K., Pitt, J., Xu, Y., Governatori, G., and Ito, T., editors, *Principles and Practice of Multi-Agent Systems - 17th International Conference, PRIMA 2014*, volume 8861 of *Lecture Notes in Artificial Intelligence*, pages 294–310, Berlin. Springer.

Frantz, C., Purvis, M. K., Savarimuthu, B. T. R., and Nowostawski, M. (2015b). Modelling the Impact of Role Specialisation on Cooperative Behaviour in Historic Trader Scenarios. In Ghose, A. K. and Oren, N., editors, *Coordination, Organizations, Institutions and Norms in Agent Systems X*, volume 9372 of *Lecture Notes in Artificial Intelligence*, Berlin. Springer.

Friedman, M. A. (1982). Polygyny in Jewish Tradition and Practice New Sources from the Cairo Geniza. *American Academy for Jewish Research*, 49:33–68.

Friedman, M. A. (1986). *Jewish Polygyny in the Middle Ages*. The Bialik Institute, Jerusalem.

Fuggetta, A., Picco, G., and Vigna, G. (1998). Understanding Code Mobility. *IEEE Transactions on Software Engineering*, 24(5):342–361.

Galan, J. M. and Izquierdo, L. R. (2005). Appearances can be deceiving: lessons learned re-implementing Axelrod's 'evolutionary approach to norms'. *Journal of Artificial Societies and Social Simulation*, 8(3):2.

Gambetta, D. (1988). Can we trust trust? In Gambetta, D., editor, *Trust: Making and Breaking Cooperative Relations*, pages 213–237. Basil Blackwell, Oxford.

García-Camino, A., Rodrıguez-Aguilar, J.-A., Sierra, C., and Vasconcelos, W. (2007). Norm-Oriented Programming of Electronic Institutions: A Rule-Based Approach. In Noriega, P., Vázquez-Salceda, J., Boella, G., Boissier, O., Dignum, V., Fornara, N., and Matson, E., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems II*, volume 4386 of *Lecture Notes in Computer Science*, pages 177–193. Springer, Berlin.

Gâteau, B. and Boissier, O. (2007). Normative Multi-Agent Organizations: Modeling, Support and Control. In Boella, G., van der Torre, L., and Verhagen, H., editors, *Normative*

Multi-agent Systems, Dagstuhl Seminar Proceedings 07122, Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

Ghasem-Aghaee, N. and Ören, T. (2003). Towards Fuzzy Agents with Dynamic Personality for Human Behavior Simulation. In *Proceedings of the 2003 Summer Computer Simulation Conference, Montreal, Canada, July 20-24*, pages 3–10, San Diego (CA).

Ghorbani, A., Aldewereld, H., Dignum, V., and Noriega, P. (2013a). Shared Strategies in Artificial Agent Societies. In Aldewereld, H. and Sichman, J., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems VIII*, volume 7756 of *Lecture Notes in Computer Science*, pages 71–86. Springer, Berlin.

Ghorbani, A., Bots, P., Dignum, V., and Dijkema, G. (2013b). MAIA: a Framework for Developing Agent-Based Social Simulations. *Journal of Artificial Societies and Social Simulation*, 16(2):9.

Ghose, A. and Savarimuthu, B. T. R. (2013). Norms as Objectives: Revisiting Compliance Management in Multi-agent Systems. In Aldewereld, H. and Sichman, J., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems VIII*, volume 7756 of *Lecture Notes in Computer Science*, pages 105–122. Springer, Berlin.

Giardini, F. and Conte, R. (2012). Gossip for Social Control in Natural and Artificial Societies. *Simulation*, 88(1):18–32.

Giddens, A. (1979). *Central problems in social theory: Action, structure, and contradiction in social analysis*. University of California Press, Los Angeles (CA).

Giddens, A. (1984). *The Constitution of Society*. Polity Press, Cambridge (UK).

Gil, M. (1976). *Documents of the Jewish pious foundations from the Cairo Geniza*. Brill, Leiden.

Gil, M. (1997). *A History of Palestine, 634-1099*. Cambridge University Press, Cambridge (UK). Originally published: 1983.

Gilbert, N. (1995). Emergence in social simulation. In *Artificial Societies: The Computer Simulation of Social Life*, pages 144–156. UCL Press, London.

Gilbert, N. (2004). Agent-based social simulation: dealing with complexity. In *The Complex Systems Network of Excellence*, volume 9, pages 1–14.

Gilbert, N. (2008). *Agent-Based Models*. Sage Publications, London.

Gilbert, N. (2014). The Journal of Artificial Societies and Social Simulation. http://jasss.soc.surrey.ac.uk/JASSS.html. Accessed on: 25th Sept. 2014.

Gilbert, N. and Troitzsch, K. G. (2005). *Simulation for the Social Scientist*. Open University Press, Maidenhead.

Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, 206:169–179.

Gintis, H. (2009). *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton University Press, Princeton (NJ).

Global Initiative to End All Corporal Punishment of Children (2015). States which have prohibited all corporal punishment. http://www.endcorporalpunishment.org/progress/prohibiting-states/. Accessed on: 9th August 2015.

Goffman, E. (1961). *Asylums: essays on the social situation of mental patients and other inmates*. Anchor Books, Garden City (NY).

Goitein, S. (1973). *Letters of Medieval Jewish Traders*. Princeton University Press, Princeton (NJ).

Goitein, S. D. (1960). The Unity of the Mediterranean World in the "Middle" Middle Ages. *Studia Islamica*, (12):29–42.

Goitein, S. D. (1967–2000a). *A Mediterranean Society: The Jewish Communities of the Arab World as Portrayed in the Documents of the Cairo Geniza*. University of California Press, Berkeley (CA). (6 Volumes).

Goitein, S. D. (1974). *Jews and Arabs: Their Contacts through the Ages*. Schocken, New York (NY). Originally published: 1955.

Goitein, S. D. (2000b). *A Mediterranean Society: The Jewish Communities of the Arab World as Portrayed in the Documents of the Cairo Geniza*, volume I: Economic Foundations. University of California Press, Berkeley (CA). Originally published: 1967.

Goldberg, J. L. (2005). *The Geographies of Trade and Traders in the Eleventh-Century Mediterranean: A Study Based on Documents from the Cairo Geniza*. PhD thesis, Columbia University, New York (NY).

Goldberg, J. L. (2011). On reading Goitein's A Mediterranean Society: a view from economic history. *Mediterranean Historical Review*, 26(2):171–186.

Goldberg, J. L. (2012a). Choosing and Enforcing Business Relationships in the Eleventh Century Mediterranean: Reassessing the 'Maghribī Traders'. *Past & Present*, 216(1):3–40.

Goldberg, J. L. (2012b). The use and abuse of commercial letters from the Cairo Geniza. *Journal of Medieval History*, 38(2):127–154.

Goldberg, J. L. (2012c). *Trade and Institutions in the Medieval Mediterranean: The Geniza Merchants and their Business World*. Cambridge University Press, New York (NY).

González De Lara, Y. (2008). The secret of Venetian success: a public-order, reputation-based institution. *European Review of Economic History*, 12(3):247–285.

Goode, W. J. (1978). *The Celebration of Heroes: Prestige as a Social Control System*. University of California Press, Berkeley (CA).

Goody, J. (1983). *The Development of Family and Marriage in Europe*. Cambridge University Press, Cambridge (UK).

Graham, H. (1996). Smoking prevalence among women in the European Community 1950–1990. *Social Science & Medicine*, 43(2):243–254.

Granovetter, M. S. (1985). Economic Action, Social Structure, and Embeddedness. *American Journal of Sociology*, 91(3):481–510.

Greif, A. (1989). Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders. *The Journal of Economic History*, 49(4):857–882.

Greif, A. (1992). Institutions and International Trade: Lessons from the Commercial Revolution. *The American Economic Review*, 82(2):128–133.

Greif, A. (1993). Contract Enforceability and Economic Institution in Early Trade: The Maghribi Traders' Coalition. *The American Economic Review*, 83(3):525–548.

Greif, A. (1994). Cultural Beliefs and the Organization of Society: A historical and theoretical reflection on Collectivist and Individualist Societies. *The Journal of Political Economy*, 102(5):912–950.

Greif, A. (1998). Self-enforcing political systems and economic growth: Late medieval genoa. In *Analytic Narratives*, pages 23–63. Princeton University Press, Princeton (NJ).

Greif, A. (2006). *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*. Cambridge University Press, New York (NY).

Greif, A. (2008). Contract Enforcement and Institutions among the Maghribi Traders: Refuting Edwards and Ogilvie. *CESifo Working Paper Series*, (2350). Available at SSRN: http://ssrn.com/abstract=1159681.

Greif, A. (2012). The Maghribi traders: a reappraisal? *The Economic History Review*, 65(2):445–469.

Greif, A., Iyigun, M., and Sasson, D. (2012). Social Institutions and Economic Growth: Why England and Not China Became the First Modern Economy. *Working Paper*. Available at SSRN: http://ssrn.com/abstract=1783879.

Greif, A. and Laitin, D. (2004). A Theory of Endogenous Institutional Change. *American Political Science Review*, 98(4):633–652.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S. K., Huse, G., Huth, A., Jepsen, J. U., Jørgensen, C., Mooij, W. M., Müller, B., Pe'er, G., Piou, C., Railsback, S. F., Robbins, A. M., Robbins, M. M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R. A., Vabø, R., Visser, U., and DeAngelis, D. L. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198(1–2):115–126.

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., and Railsback, S. F. (2010). The {ODD} protocol: A review and first update. *Ecological Modelling*, 221(23):2760 – 2768.

Groenewegen, J., Spithoven, A., and den Berg, A. V. (2010). *Institutional Economics*. Palgrave Macmillan, London.

Grofman, B. and Landa, J. (1983). The Development of Trading Networks among Spatially Separated Traders as a Process of Proto-Coalition Formation: The Kula Trade. *Social Networks*, 5(4):347–365.

Grossi, D., Aldewereld, H., and Dignum, F. (2006). Ubi Lex, Ibi Poena: Designing Norm Enforcement in E-Institutions. In Noriega, P., Vázquez-Salceda, J., Boella, G., Boissier, O., Dignum, V., Fornara, N., and Matson, E., editors, *Coordination, Organizations, Institutions, and Norms in Multi-Agent Systems II*, volume 4386 of *Lecture Notes in Computer Science*, pages 107–120, Berlin. Springer.

Grusec, J. E. and Kuczynski, L. (1997). *Parenting and Children's Internalization of Values: A Handbook of Contemporary Theory*. John Wiley & Sons, New York (NY).

Gurvitch, G. (1964). *The Spectrum of Social Time*. D. Reidel Publishing Company, Dordrecht.

Gutman, M. (2011). Social norms and attitudes about smoking (1991-2010). RWJF Retrospective Series, Robert Wood Johnson Foundation.

Habermas, J. (1987). *The Theory of Communicative Action*, volume II. Beacon Press, Boston (MA).

Hagras, H. (2004). A Hierarchical Type-2 Fuzzy Logic Control Architecture for Autonomous Mobile Robots. *IEEE Transactions on Fuzzy Systems*, 12(4):524–539.

Hagras, H. (2007). Type-2 FLCs: A New Generation of Fuzzy Controllers. *IEEE Computational Intelligence Magazine*, 2(1):30–43.

Harbord, D. (2006). Enforcing cooperation among medieval merchants: The Maghribi traders revisited. MPRA Paper 1889, Market Analysis Ltd.

Hardin, G. (1968). The Tragedy of the Commons. *Science*, 162(3859):1243–1248.

Harris, R. (2009). The institutional dynamics of early modern eurasian trade: The corporation and the commenda. *Journal of Economic Behavior and Organization*, 71(3):606–622.

Hassan, S., Garmendia, L., and Pavón, J. (2007). Agent-Based Social Modeling and Simulation with Fuzzy Sets. In Corchado, E., Corchado, J., and Abraham, A., editors, *Innovations in Hybrid Intelligent Systems*, volume 44 of *Advances in Intelligent and Soft Computing*, pages 40–47. Springer, Berlin.

Hassan, S., Salgado, M., and Pavón, J. (2011). Friendship Dynamics: Modelling Social Relationships through a Fuzzy Agent-Based Simulation. *Discrete Dynamics in Nature and Society*, 2011, Article ID 765640:19 pages.

Hawe, G. I., Coates, G., Wilson, D. T., and Crouch, R. S. (2012). Agent-based simulation for large-scale emergency response: A survey of usage and implementation. *ACM Computing Surveys*, 45(1):Article No. 8.

Haynes, C., Miles, S., and Luck, M. (2014). Monitoring the impact of norms upon organisational performance: A simulation approach. In Balke, T., Dignum, F., van Riemsdijk, M. B., and Chopra, A. K., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems IX*, volume 8386 of *Lecture Notes in Computer Science*, pages 103–119. Springer, Berlin.

Heath, B., Hill, R., and Ciarallo, F. (2009). A Survey of Agent-Based Modeling Practices (January 1998 to July 2008). *Journal of Artificial Societies and Social Simulation*, 12(4):9.

Heckbert, S. (2013). MayaSim: An Agent-Based Model of the Ancient Maya Social-Ecological System. *Journal of Artificial Societies and Social Simulation*, 16(4):11.

Hedström, P. and Swedberg, R. (1998). *Social Mechanisms – An Analytical Approach to Social Theory*. Cambridge University Press, Cambridge (UK).

Helmke, G. and Levitsky, S. (2004). Informal Institutions and Comparative Politics: A Research Agenda. *Perspectives on Politics*, 2(4):725–740.

Henrich, J. (2000). Does culture matter in economic behaviour? Ultimatum game bargaining among the Machiguenga of the Peruvian Amazon. *American Economic Review*, 90(4):973–979.

Himma, K. (2009). Artificial agency, consciousness, and the criteria for moral agency: what properties must an artificial agent have to be a moral agent? *Ethics and Information Technology*, 11(1):19–29.

Hinkle, D. E., Wiersma, W., and Jurs, S. G. (2003). *Applied Statistics for the Behavioral Sciences*. Houghton Mifflin, Boston (MA), 15th edition.

Hodgson, G. (1988). *Economics and Institutions: A Manifesto for A Modern Institutional Economics*. Polity Press and University of Pennsylvania Press, Philadelphia (PA).

Hodgson, G. M. (2002). The Evolution of Institutions: An Agenda for Future Theoretical Research. *Constitutional Political Economy*, 13(2):111–127.

Hodgson, G. M. (2006). What Are Institutions? *Journal of Economic Issues*, 40(1):1–25.

Hodgson, G. M. (2014). On fuzzy frontiers and fragmented foundations: some reflections on the original and new institutional economics. *Journal of Institutional Economics*, 10(4):591–611.

Hofstede, G., Hofstede, G. J., and Minkov, M. (2010). *Cultures and Organizations: Software of the Mind*. McGraw-Hill, New York (NY), 3rd edition.

Hohfeld, W. N. (1913). Some Fundamental Legal Conceptions as Applied in Judicial Reasoning. *Yale Law Journal*, 23:16–59.

Hollander, C. and Wu, A. (2011a). Using the process of norm emergence to model consensus formation. In *Fifth IEEE International Conference on Self-Adaptive and Self-Organizing Systems (SASO)*, pages 148–157.

Hollander, C. D. and Wu, A. S. (2011b). The Current State of Normative Agent-Based Systems. *Journal of Artificial Societies and Social Simulation*, 14(2):6.

Hollingsworth, J. R. (2000). Doing institutional analysis: implications for the study of innovations. *Review of International Political Economy*, 7(4):595–644.

Honig, W. K. and Staddon, J. E. R., editors (1977). *The handbook of operant behavior*. Prentice Hall, Englewood Cliffs (NJ).

Höppner, F., Klawonn, F., Kruse, R., and Runkler, T. (1999). *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*. Wiley, Chichester.

Horden, P. and Purcell, N. (2000). *The Corrupting Sea: A Study of Mediterranean History*. Blackwell, Oxford (UK).

Huang, H.-P., Liu, Y.-H., Liu, L.-W., and Wong, C.-S. (2010). Applications of Advanced Fuzzy Logic Techniques in Fuzzy Image Processing Scheme. *Advances in Fuzzy Mathematics*, 5(1):71–76.

Hübner, J., Boissier, O., and Bordini, R. (2010). A Normative Organisation Programming Language for Organisation Management Infrastructures. In Padget, J., Artikis, A., Vasconcelos, W., Stathis, K., da Silva, V., Matson, E., and Polleres, A., editors, *Coordination, Organizations, Institutions and Norms in Agent Systems V*, volume 6069 of *Lecture Notes in Computer Science*, pages 114–129. Springer, Berlin.

Hungwe, B. (2014). Zimbabwe's multi-currency confusion. http://www.bbc.com/news/world-africa-26034078. Accessed on: 2nd October 2014.

Hwang, C. and Rhee, F. (2007). Uncertain fuzzy clustering: interval type-2 fuzzy approach to C-means. *IEEE Transactions on Fuzzy Systems*, 15(1):107–120.

Ibn Khaldun (1967). *The Muqaddimah: An Introduction to History*. Princeton University Press, Princeton (NJ). Originally published: around 1377 AD.

International Encyclopedia of the Social Sciences (2008). Corporal punishment. http://www.encyclopedia.com/topic/corporal_punishment.aspx. Accessed on: 26th October 2014.

ISO (1996). ISO/IEC 14977 – Information technology - Syntactic metalanguage - Extended BNF. Accessed on: 1st July 2013.

Janowitz, M. (1975). Sociological Theory and Social Control. *American Journal of Sociology*, 81(1):82–108.

Jehel, G. (1995). Expéditions navales ou croisade? l'activité militaro-diplomatique de gênes dans l'occident méditerranéen (xème-xivème siècle). In Balard, M. and Ducellier, A., editors, *Coloniser au Moyen Age*. Colin, Paris.

Jennings, N. R. and Wooldridge, M. (2000). Agent-Oriented Software Engineering. *Artificial Intelligence*, 117:277–296.

Jensen, K. (1992). *Coloured Petri-Nets – Basic concepts, Analysis Models and Practical Use*, volume 1. Springer, Heidelberg.

Jiang, H., Vidal, J. M., and Huhns, M. N. (2007). Ebdi: An architecture for emotional agents. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multi-Agent Systems*, AAMAS '07, pages 11:1–11:3, New York (NY). ACM.

Jiang, J., Aldewereld, H., Dignum, V., and Tan, Y.-H. (2013). Norm contextualization. In Aldewereld, H. and Sichman, J., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems VIII*, volume 7756 of *Lecture Notes in Computer Science*, pages 141–157. Springer, Berlin.

John, R. and Coupland, S. (2007). Type-2 fuzzy logic: a historical view. *IEEE Computational Intelligence Magazine*, 2(1):57–62.

John, R., Innocent, P., and Barnes, M. (2000). Neuro-fuzzy clustering of radiographic tiba image data using type-2 fuzzy sets. *Information Sciences*, 125(1–4):65–82.

Johnson, D. and Bering, J. (2006). Hand of God, Mind of Man: Punishment and Cognition in the Evolution of Cooperation. *Evolutionary Psychology*, 4:219–233.

Joskow, P. L. (2004). New institutional economics: A report card. Technical report, Presidential Address of the Annual Conference of the International Society of New Institutional Economics, Budapest, Hungary. (Essay based on talk given in Sept. 2003).

Kaelbling, L. P., Littman, M. L., and Moore, A. W. (1996). Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, 4:237–285.

Kaplan, H. and Gurven, M. (2005). The natural history of human food sharing and cooperation: A review and a new multi-individual approach to the negotiation of norms. In Gintis, H., Bowles, S., Boyd, R., and Fehr, E., editors, *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*, pages 75–114. MIT Press, Cambridge (MA).

Karnik, N. N. and Mendel, J. M. (1999). Applications of type-2 fuzzy logic systems to forecasting of time-series. *Information Sciences*, 120(1–4):89–111.

Kaser, M. (1965). *Roman Private Law*. Butterworths, Durban.

Katele, I. B. (1988). Piracy and the Venetian State: The Dilemma of Maritime Defense in the Fourteenth Century. *Speculum*, 63:865–889.

Kedar, B. Z. (1976). *Merchants in Crisis: Genoese and Venetian men of affairs and the fourteenth-century depression*. Yale University Press, New Haven (CT).

King, T., Dignum, V., and van Riemsdijk, M. (2014). Re-checking normative system coherence. In Balke, T., Dignum, F., van Riemsdijk, M. B., and Chopra, A. K., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems IX*, volume 8386 of *Lecture Notes in Computer Science*, pages 275–290. Springer, Berlin.

Kingston, C. and Caballero, G. (2009). Comparing theories of institutional change. *Journal of Institutional Economics*, 5(2):151–180.

Kinzig, A. P., Ehrlich, P. R., Alston, L. J., Arrow, K., Barrett, S., Buchman, T. G., Daily, G. C., Levin, B., Levin, S., Oppenheimer, M., Ostrom, E., and Saari, D. (2013). Social Norms and Global Environmental Challenges: The Complex Interaction of Behaviors, Values, and Policy. *BioScience*, 63(3):164–175.

Klir, G. and Folger, T. (1988). *Fuzzy Sets, Uncertainty and Information*. Prentice Hall, Englewood Cliffs (NJ).

Klügl, F. (2008). A validation methodology for agent-based simulations. In *Proceedings of the 2008 ACM Symposium on Applied Computing*, SAC '08, pages 39–43, New York (NY). ACM.

Knight, J. (1992). *Institutions and Social Conflict*. Cambridge University Press, Cambridge (UK).

Kollingbaum, M. and Norman, T. (2006). Informed Deliberation During Norm-Governed Practical Reasoning. In Boissier, O., Padget, J., Dignum, V., Lindemann, G., Matson, E., Ossowski, S., Sichman, J., and Vázquez-Salceda, J., editors, *Coordination, Organizations, Institutions, and Norms in Multi-Agent Systems*, volume 3913 of *Lecture Notes in Computer Science*, pages 183–197. Springer, Berlin.

Kollingbaum, M., Norman, T., Preece, A., and Sleeman, D. (2007). Norm conflicts and inconsistencies in virtual organisations. In Noriega, P., Vázquez-Salceda, J., Boella, G., Boissier, O., Dignum, V., Fornara, N., and Matson, E., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems II*, volume 4386 of *Lecture Notes in Computer Science*, pages 245–258. Springer, Berlin.

Kreps, D. M. (1990). *Game Theory and Economic Modelling*. Clarendon Press, Oxford.

Lacroix, B., Mathieu, P., and Kemeny, A. (2009). The Use of Norms Violations to Model Agents Behavioral Variety. In Hübner, J., Matson, E., Boissier, O., and Dignum, V., editors, *Coordination, Organizations, Institutions and Norms in Agent Systems IV*, volume 5428 of *Lecture Notes in Computer Science*, pages 220–234. Springer, Berlin.

Laird, J. E., Newell, A., and Rosenbloom, P. S. (1987). SOAR: An Architecture for General Intelligence. *Artificial Intelligence*, 33(1):1–64.

Lambeck, K., Anzidei, M., Antonioli, F., Benini, A., and Esposito, A. (2004). Sea level in Roman time in the Central Mediterranean and implications for recent change. *Earth and Planetary Science Letters*, (224):563–575.

Lehmann, E. L. (2006). *Nonparametrics: Statistical Methods Based on Ranks*. Springer, Berlin.

Lesani, M. and Montazeri, N. (2009). Fuzzy Trust Aggregation and Personalized Trust Inference in Virtual Social Networks. *Computational Intelligence*, 25(2):51–83.

Lessig, L. (1999). *Code and Other Laws of Cyberspace*. Basic Books, New York (NY).

Lewis, D. K. (1969). *Convention: A Philosophical Study*. Harvard University Press, Cambridge (MA).

Li, T., Jiang, J., Aldewereld, H., De Vos, M., Dignum, V., and Padget, J. (2014). Contextualized institutions in virtual organizations. In Balke, T., Dignum, F., van Riemsdijk, M. B., and Chopra, A. K., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems IX*, volume 8386 of *Lecture Notes in Computer Science*, pages 136–154. Springer, Berlin.

Liang, Q. and Mendel, J. (2001). MPEG VBR video traffic modeling and classification using fuzzy techniques. *IEEE Transactions on Fuzzy Systems*, 9(1):183–193.

Libecap, G. D. (1989). *Contracting for Property Rights*. Cambridge University Press, Cambridge (UK).

Linder, A., editor (1997). *Jews in the Legal Sources of the Early Middle Ages*. Wayne State University Press, Detroit (MI).

Liu, C.-F., Yeh, C.-Y., and Lee, S.-J. (2012). Application of Type-2 Neuro-Fuzzy Modeling in Stock Price Prediction. *Applied Soft Computing*, 12(4):1348 – 1358.

Liu, F. and Mendel, J. M. (2008a). Aggregation using the fuzzy weighted average as computed by the Karnik-Mendel algorithms. *IEEE Transactions on Fuzzy Systems*, 16(1):1–12.

Liu, F. and Mendel, J. M. (2008b). Encoding Words Into Interval Type-2 Fuzzy Sets Using an Interval Approach. *IEEE Transactions on Fuzzy Systems*, 16(6):1503–1521.

Loasby, B. J. (2000). Market institutions and economic evolution. *Journal of Evolutionary Economics*, 10(3):297–309.

Lombard, M. (1948). Mahomet et Charlemagne, le probleme économique. *Annales Économies, Sociétés, Civilisation*, 3(2):188–199.

Long, Z., Yuanc, Y., and Long, W. (2014). Designing fuzzy controllers with variable universes of discourse using input-output data. *Engineering Applications of Artificial Intelligence*, 36:215–221.

Lopez, R. S. (1975). *Su e giù per la storia di Genova*. Università di Genova, Istituto di paleografia e storia medievale, Genoa.

Lopez, R. S. (1976). *The Commercial Revolution of the Middle Ages, 950-1350*. Cambridge University Press, Cambridge (UK).

Lopez, R. S. (1982). Genoa. In Strayer, J. R., editor, *Dictionary of the Middle Ages*, pages 383–387. Scribner, New York (NY).

López y López, F. and Luck, M. (2004). Towards a Model of the Dynamics of Normative Multi-Agent Systems. In Lindemann, G., Moldt, D., and Paolucci, M., editors, *RASTA 2002*, volume 2934 of *Lecture Notes in Artificial Intelligence*, pages 175–194, Heidelberg. Springer.

López y López, F., Luck, M., and Puebla, A. (2003). Modelling Norms for Autonomous Agents. In *Proceedings of Fourth Mexican International Conference on Computer Science*, pages 238–245.

Lotzmann, U. and Möhring, M. (2009). Simulating norm formation: An operational approach. In *Proceedings of the 8th International Conference on Autonomous Agents and Multi-Agent Systems - Volume 2*, AAMAS '09, pages 1323–1324.

Luke, S. (2015). ECJ – A Java-based Evolutionary Computation Research System. https://cs.gmu.edu/ eclab/projects/ecj/. Accessed on: 28th August 2015.

Luke, S., Cioffi-Revilla, C., Panait, L., Sullivan, K., and Balan, G. (2005). MASON: A Multiagent Simulation Environment. *Simulation*, 81(7):517–527.

Macal, C. and North, M. (2010). Tutorial on agent-based modelling and simulation. *Journal of Simulation*, 4(3):151–162.

Macy, M. W. and Willer, R. (2002). From Factors to Actors: Computational Sociology and Agent-Based Modeling. *Annual Review of Sociology*, 28:143–166.

Mahmoud, M. A., Ahmad, M. S., Yusoff, M. Z. M., and Mustapha, A. (2014). A Review of Norms and Normative Multiagent Systems. *The Scientific World Journal*, 2014:23 pages. Article ID 684587.

Mann, H. B. and Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics*, 18(1):50–60.

Márquez, B., Castanon-Puga, M., Castro, J., Suarez, E., and Magdaleno-Palencia, S. (2011). Fuzzy Models for Complex Social Systems Using Distributed Agencies in Poverty Studies. In Mohamad Zain, J., Wan Mohd, W., and El-Qawasmeh, E., editors, *Software Engi-*

neering and Computer Systems, volume 179 of *Communications in Computer and Information Science*, pages 391–400. Springer, Berlin.

Marx, K. (1977). *Capital: A Critique of Political Economy*, volume 1. Vintage Books, New York (NY). Originally published: 1867.

Matsumoto, M. and Nishimura, T. (1998). Mersenne Twister: A 623-dimensionally Equidistributed Uniform Pseudo-random Number Generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1):3–30.

Matthews, R. B., Gilbert, N. G., Roach, A., Polhill, J., and Gotts, N. M. (2007). Agent-based land-use models: a review of applications. *Landscape Ecology*, 22(10):1447–1459.

McKenna, D. (2013). Towards a Grammar of Local Political Institutions. In *The Party's Over? 63rd Political Studies Association Annual International Conference, 25–27 March 2013, City Hall Cardiff.*

McNamara, P. (1996). Making Room for Going Beyond the Call. *Mind*, 105(419):415–450.

Ménard, C. and Shirley, M. M. (2014). The Future of new Institutional Economics: From Early Intuitions to a New Paradigm? *Journal of Institutional Economics*, 10(4):541–565.

Mendel, J. (2001). *Uncertain Rule-based Fuzzy Logic Systems: Introduction and New Directions*. Prentice Hall, Upper Saddle River (NJ).

Mendel, J. (2003). Fuzzy sets for words: a new beginning. In *The 12th IEEE International Conference on Fuzzy Systems, 2003 (FUZZ '03)*, volume 1, pages 37–42.

Mendel, J. (2007a). Computing with Words: Zadeh, Turing, Popper and Occam. *Computational Intelligence Magazine, IEEE*, 2(4):10–17.

Mendel, J., John, R., and Liu, F. (2006a). Interval Type-2 Fuzzy Logic Systems Made Simple. *IEEE Transactions on Fuzzy Systems*, 14(6):808–821.

Mendel, J. M. (2007b). Advances in type-2 fuzzy sets and systems. *Information Sciences*, 177(1):84–110.

Mendel, J. M. (2007c). Type-2 fuzzy sets and systems: an overview. *IEEE Computational Intelligence Magazine*, 2(1):20–29.

Mendel, J. M., Hagras, H., and John, R. I. (2006b). Standard background material about interval type-2 fuzzy logic systems that can be used by all authors. http://cis.ieee.org/images/files/Documents/standards.t2.win.pdf. Accessed on: 1st October 2014.

Menger, C. (1892). On the Origin of Money. *Economic Journal*, 2:239–255.

Merton, R. K. (1994). Durkheim's Division of Labor in Society. *Sociological Forum*, 9(1):17–25.

Meyer, J.-J. C. (1988). A Different Approach to Deontic Logic: Deontic Logic Viewed as a Variant of Dynamic Logic. *Notre Dame Journal of Formal Logic*, 29(1):109–136.

Meyer, J.-J. C. and Wieringa, R. J. (1993). *Deontic Logic in Computer Science: Normative System Specification*. John Wiley & Sons, Chichester (UK).

Michel, F., Ferber, J., and Drogoul, A. (2009). Multi-agent systems and simulation: A survey from the agent community's perspective. In Uhrmacher, A. and Weyns, D., editors, *Multi-Agent Systems: Simulation and Applications*, pages 3–52. CRC Press, Boca Raton (FL).

Milgram, S. (1967). The small world problem. *Psychology Today*, 1(1):61–67.

Milgrom, P. R., North, D. C., and Weingast, B. R. (1990). The Role of Institutions in the Revival of the Trade: The Law Merchant, Private Judges, and the Champagne Fairs. *Economics and Politics*, 2(1):1–23.

Miller, S., Gongora, M., and John, R. (2011). Interval Type-2 Fuzzy Modelling and Simulated Annealing for Real-World Inventory Management. In Corchado, E., KurzyÅĎski, M., and Woźniak, M., editors, *Hybrid Artificial Intelligent Systems*, volume 6678 of *Lecture Notes in Computer Science*, pages 231–238. Springer, Berlin.

Miller, S., Góngora, M. A., Garibaldi, J. M., and John, R. (2012). Interval Type-2 Fuzzy Modelling and Stochastic Search for Real-World Inventory Management. *Soft Computing*, 16(8):1447–1459.

Montealegre Vázquez, L. and López y López, F. (2007). An Agent-Based Model for Hierarchical Organizations. In Noriega, P., Vázquez-Salceda, J., Boella, G., Boissier, O., Dignum, V., Fornara, N., and Matson, E., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems II*, volume 4386 of *Lecture Notes in Computer Science*, pages 194–211. Springer, Berlin.

Montesquieu (1949). *The Spirit of the Laws*. Macmillan, New York (NY). Originally published: 1748.

Morales, J., López-Sánchez, M., Rodríguez-Aguilar, J. A., Wooldridge, M., and Vasconcelos, W. (2013). Automated synthesis of normative systems. In *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-Agent Systems*, AAMAS '13, pages 483–490.

Morales, J., López-Sánchez, M., Rodríguez-Aguilar, J. A., Wooldridge, M., and Vasconcelos, W. (2014). Minimality and Simplicity in the On-line Automated Synthesis of Normative Systems. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems*, AAMAS '14, pages 109–116.

Morris, C. (1972). *The Discovery of the Individual, 1050-1200*. Harper and Row, New York (NY).

Moss, S. (2008). Alternative Approaches to the Empirical Validation of Agent-Based Models. *Journal of Artificial Societies and Social Simulation*, 11(1):5.

Mouzelis, N. (1991). *Back to Sociological Theory*. Macmillan, Basingstone (UK).

Mukherjee, P., Sen, S., and Airiau, S. (2007). Emergence of Norms with Biased Interaction in Heterogeneous Agent Societies. In *Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, pages 512–515.

Mulaa, J., Polera, R., and Garcia-Sabater, J. P. (2007). Material Requirement Planning with fuzzy constraints and fuzzy coefficients. *Fuzzy Sets and Systems*, 158(7):783–793.

Myerson, R. (2004). Justice, Institutions, and Multiple Equilibria. *Chicago Journal of International Law*, 5:91–107.

Myerson, R. B. (1999). Nash Equilibrium and the History of Economic Theory. *Journal of Economic Literature*, 37(3):1067–1082.

Nakamaru, M. and Levin, S. A. (2004). Spread of two linked social norms on complex interaction networks. *Journal of Theoretical Biology*, 230(1):57–64.

Nakamoto, S. (2009). Bitcoin: A Peer-to-Peer Electronic Cash System. https://bitcoin.org/bitcoin.pdf. Accessed on: 1st August 2014.

Nash, J. F. (1950). Equilibrium Points in N-Person Games. *Proceedings of the National Academy of Sciences of the United States of America*, 36(1):48–49.

Navarro, Z. (2006). In Search of a Cultural Interpretation of Power: The Contribution of Pierre Bourdieu. *IDS Bulletin*, 37(6):11–22.

Nelson, R. R. (1994). The Co-evolution of Technology, Industrial Structure, and Supporting Institutions. *Industrial and Corporate Change*, 3(1):47–63.

Neumann, M. (2008). Homo Socionicus: a Case Study of Simulation Models of Norms. *Journal of Artificial Societies and Social Simulation*, 11(4):6.

Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press, Oxford (UK).

Noriega, P. (1997). *Agent-Mediated Auctions: The Fishmarket Metaphor*. PhD thesis, Universitat Autonoma de Barcelona, Barcelona.

North, D. C. (1981). *Structure and Change in Economic History*. Norton, New York (NY).

North, D. C. (1990). *Institutions, Institutional Change, and Economic Performance*. Cambridge University Press, New York (NY).

North, D. C. (1991). Institutions. *Journal of Economic Perspectives*, 5(1):97–112.

North, D. C. (2005). *Understanding the Process of Economic Change*. Princeton University Press, Princeton (NJ).

North, D. C., Wallis, J. J., and Weingast, B. R. (2009). *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*. Cambridge University Press, New York (NY).

North, M. J., Collier, N. T., Ozik, J., Tatara, E., Altaweel, M., Macal, C. M., Bragen, M., and Sydelko, P. (2013). Complex Adaptive Systems Modeling with Repast Simphony. *Complex Adaptive Systems Modeling*, 1(1):3.

Nowak, M. A. and Sigmund, K. (1998). The Dynamics of Indirect Reciprocity. *Journal of Theoretical Biology*, 194(4):561–574.

Nowak, M. A. and Sigmund, K. (2005). Evolution of Indirect Reciprocity. *Nature*, 437:1291–1298.

Oakerson, R. J. (1992). Analyzing the commons: A framework. In Bromley, D. W. and Feeny, D., editors, *Making the Commons Work: Theory, Practice, and Policy*, pages 41–59. ICS Press, San Francisco (CA).

Object Management Group (2014). Unified Modeling Language (UML). http://www.omg.org/spec/UML/. Accessed on: 22th October 2014.

Odelstad, J. (2007). Agents, Norms and Forest Cleaning. In Boella, G., van der Torre, L., and Verhagen, H., editors, *Normative Multi-agent Systems, Dagstuhl Seminar Proceedings 07122*, Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

Oliver, P. (1980). Rewards and Punishments as Selective Incentives for Collective Action: Theoretical Investigations. *American Journal of Sociology*, 85(6):1356–1375.

Oren, N., Luck, M., and Miles, S. (2010). A Model of Normative Power. In *Proceedings of the 9th International Conference on Autonomous Agents and Multi-Agent Systems*, AAMAS '10, pages 815–822.

Oren, N., Panagiotidi, S., Vázquez-Salceda, J., Modgil, S., Luck, M., and Miles, S. (2009). Towards a Formalisation of Electronic Contracting Environments. In Hübner, J., Matson, E., Boissier, O., and Dignum, V., editors, *Coordination, Organizations, Institutions and Norms in Agent Systems IV*, volume 5428 of *Lecture Notes in Computer Science*, pages 156–171. Springer, Berlin.

Ören, T. and Ghasem-Aghaee, N. (2003). Personality Representation Processable in Fuzzy Logic for Human Behavior Simulation. In *Proceedings of the 2003 Summer Computer Simulation Conference, Montreal, Canada, July 20-24*, pages 11–18, San Diego (CA). SCS.

Ortony, A., Clore, G., and Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge University Press, New York (NY).

Ostrom, E. (1986). An Agenda for the Study of Institutions. *Public Choice*, 48(1):3–25.

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, New York (NY).

Ostrom, E. (2005a). Animating Instiutional Analysis. In *Understanding Institutional Diversity*, chapter 4, pages 99–134. Princeton University Press, Princeton (NJ).

Ostrom, E. (2005b). *Understanding Institutional Diversity*. Princeton University Press, Princeton (NJ).

Ostrom, E. (2005c). Why Classify Generic Rules? In *Understanding Institutional Diversity*, chapter 6, pages 175–185. Princeton University Press, Princeton (NJ).

Ostrom, E. (2008). Developing a method for analyzing institutional change. In Batie, S. and Mercuro, N., editors, *Alternative Institutional Structures: Evolution and Impact*, pages 48–77. Routledge, London.

Ostrom, E. (2011). Background on the Institutional Analysis and Development Framework. *Policy Studies Journal*, 39(1):7–27.

Ostrom, E. and Crawford, S. E. (2005). Classifying Rules. In *Understanding Institutional Diversity*, chapter 7, pages 186–215. Princeton University Press, Princeton (NJ).

Ostrom, E., Gardner, R., and Walker, J. (1994). *Rules, Games, and Common-Pool Resources*. University of Michigan Press, Ann Arbor (MI).

Panagiotidi, S., Alvarez-Napagao, S., and Vázquez-Salceda, J. (2014). Towards the Norm-Aware Agent: Bridging the Gap Between Deontic Specifications and Practical Mechanisms for Norm Monitoring and Norm-Aware Planning. In Balke, T., Dignum, F., van Riemsdijk, M. B., and Chopra, A. K., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems IX*, volume 8386 of *Lecture Notes in Computer Science*, pages 346–363. Springer, Berlin.

Panagiotidi, S., Vázquez-Salceda, J., and Dignum, F. (2013). Reasoning over Norm Compliance via Planning. In Aldewereld, H. and Sichman, J., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems VIII*, volume 7756 of *Lecture Notes in Computer Science*, pages 35–52. Springer, Berlin.

Parker, J. (2000). *Structuration*. Open University Press, Buckingham (UK).

Parsons, T. (1951). *The Social System*. Routledge, New York (NY).

Parsons, T. (1961). *Theories of Society: Foundations of Modern Sociological Theory*. Free Press, New York (NY).

Parsons, T. (1967). *Sociological Theory and Modern Society*. Free Press, New York (NY).

Parsons, T. (1976). A Few Considerations on the Place of Rationality in Modern Culture and Society. *Revue européenne des sciences sociales*, Tome XIV(38–39).

Patrone, A. M. N. D. and Airaldi, G. (1986). Genova e la liguria nel medioevo. In Airaldi, G., editor, *Communi e Signorie nell'Italia settentrionale: il Piemonte e la Liguria*. Utet, Genoa.

Pavlov, I. P. (1927). *Conditional Reflexes*. Dover Publications, New York (NY).

Pereira, D., Oliveira, E., Moreira, N., and Sarmento, L. (2005). Towards an architecture for emotional bdi agents. In *Proceedings of 12th Portuguese Conference on Artificial Intelligence (EPIA '05)*, pages 40–47. Springer.

Peristiany, J. G. (1966). Introduction. In Peristiany, J. G., editor, *Honour and Shame: The Values of Mediterranean Society*, pages 9–18. University of Chicago Press, Chicago (IL).

Perreau de Pinninck, A., Sierra, C., and Schorlemmer, M. (2008). Distributed Norm Enforcement Via Ostracism. In Sichman, J., Padget, J., Ossowski, S., and Noriega, P., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems III*, volume 4870 of *Lecture Notes in Computer Science*, pages 301–315. Springer, Berlin.

Picault, S. and Mathieu, P. (2011). An Interaction-oriented Model for Multi-scale Simulation. In Walsh, T., editor, *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, volume 1 of *IJCAI'11*, pages 332–337. AAAI Press.

Pirenne, H. (1956). *Mohammed and Charlemagne*. Barnes and Noble, New York (NY). Originally published: 1935.

Podolny, J. M. (1994). Market Uncertainty and the Social Character of Economic Exchange. *Administrative Science Quarterly*, 39(3):458–483.

Popper, K. R. (1945). *The Open Society and Its Enemies*. Routledge and Kegan Paul, London. 2 volumes.

Posner, E. (2000). *Law and Social Norms*. Harvard University Press, Cambridge (MA).

Posner, R. A. and Rasmusen, E. B. (1999). Creating and Enforcing Norms, with Special Reference to Sanctions. *International Review of Law and Economics*, 19(3):369–382.

Poundstone, W. (1992). *Prisoner's Dilemma*. Doubleday, New York (NY).

Pryor, J. (1977). The origins of the commanda contract. *Speculum*, 51(4):5–37.

Puterman, M. L. (1994). *Markov Decision Processes*. Wiley, Hoboken (NJ).

Radford, R. A. (1945). The Economic Organization in a P.O.W. Camp. *Economica*, 12(48):189–201.

Ramanath, A. M. and Gilbert, N. (2004). The Design of Participatory Agent-Based Social Simulations. *Journal of Artificial Societies and Social Simulation*, 7(4):1.

Ramírez, C. L., Castillo, O., Melin, P., and Díaz, A. R. (2011). Simulation of the Bird Age-Structured Population Growth Based on an Interval Type-2 Fuzzy Cellular Structure. *Information Sciences*, 181(3):519–535.

Rao, A. S. and Georgeff, M. P. (1995). BDI Agents: From Theory to Practice. *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS-95)*, pages 312–319.

Rawls, J. (2005). *A Theory of Justice*. Harvard University Press, Cambridge (MA). Originally published: 1971.

Raymond, I. and Lopez, R., editors (2001). *Medieval Trade In the Mediterranean World: Illustrative Documents*. Columbia University Press, New York (NY).

Raytheon BBN Technologies (2015). Cougaar. http://cougaar.org/. Accessed on: 1st September 2015.

Ren, F., Zhang, M., and Bai, Q. (2007). A Fuzzy-Based Approach for Partner Selection in Multi-Agent Systems. In *6th IEEE/ACIS International Conference on Computer and Information Science (ICIS)*, pages 457–462.

Rhee, F. (2007). Uncertain Fuzzy Clustering: Insights and Recommendations. *IEEE Computational Intelligence Magazine*, 2(1):44–56.

Ritzer, G. (1979). Toward an Integrated Sociological Paradigm. In Snizek, W., Fahrman, W. E., and Miller, M. K., editors, *Contemporary Issues in Theory and Research*, pages 25–46. Greenwood Press, Westport (CT).

Ritzer, G. (1981). *Towards an Integrated Sociological Paradigm: The Search for an Exemplar and an Image of the Subject Matter*. Allyn & Bacon, Boston (MA).

Ritzer, G. and Gindoff, P. (1994). Agency-structure, micro-macro, individualism-holism-relationalism: A metatheoretical explanation of theoretical convergence between the united states and europe. In Sztompka, P., editor, *Agency and Structure: Reorienting Social Theory*, pages 3–23. Gordon and Breach, Yverdon.

Ritzer, G. and Goodman, D. J. (2003). *Sociological Theory*. McGraw-Hill, New York (NY), 6th edition.

Ritzer, G. and Goodman, D. J. (2008). *Modern Sociological Theory*. McGraw-Hill, New York (NY), 7th edition.

Riveret, R., Rotolo, A., and Governatori, G. (2007). Interaction between Normative Systems and Cognitive Agents in Temporal Modal Defeasible Logic. In Boella, G., van der Torre, L., and Verhagen, H., editors, *Normative Multi-agent Systems, Dagstuhl Seminar Proceedings 07122*, Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

Rodríguez-Aguilar, J. A. (2001). *On the Design and Construction of Agent-mediated Electronic Institutions*. PhD thesis, Universitat Autonoma de Barcelona, Barcelona.

Roland, G. (2004). Understanding Institutional Change: Fast-moving and Slow-moving Institutions. *Studies in Comparative International Development*, 38(4):109–131.

Rostovtzeff, M. I. (1957). *The Social & Economic History of the Roman Empire*. Oxford University Press, Oxford (UK). (Revised edition by J. M. Fraser, Originally published: 1926).

Rostovtzeff, M. I. (1986). *The Social & Economic History of the Hellenistic World*. Oxford University Press, Oxford (UK). Originally published: 1941.

Ruffle, B. and Sosis, R. (2007). Does it pay to pray? Costly ritual and cooperation. *The B.E. Journal of Economic Analysis & Policy*, 7(1):1–37.

Ruffle, B. J. and Sosis, R. H. (2010). Do Religious Contexts Elicit More Trust and Altruism? An Experiment on Facebook. Discussion Paper 10-02, Ben-Gurion University of the Negev, Department of Economics.

Rutherford, M. (2001). Institutional Economics: Then and Now. *Journal of Economic Perspectives*, 15(3):173–194.

Ruttan, V. W. (2006). Social Science Knowledge and Induced Institutional Innovation: An Institutional Design Perspective. *Journal of Institutional Economics*, 2(3):249–272.

Saam, N. J. and Harrer, A. (1999). Simulating Norms, Social Inequality, and Functional Change in Artificial Societies. *Journal of Artificial Societies and Social Simulation*, 2(1):2.

Sabeur, E. and Denis, G. (2007). Human behavior and social network simulation: Fuzzy sets/logic and agents-based approach. In *Proceedings of the 2007 Spring Simulation Multiconference - Volume 2*, SpringSim '07, pages 102–109, San Diego (CA). Society for Computer Simulation International.

Sadeghian, A. and Tahayori, H., editors (2015). *Frontiers of Higher Order Fuzzy Sets*. Springer, Berlin.

Sakellariou, I., Kefalas, P., and Stamatopoulou, I. (2008). Enhancing NetLogo to Simulate BDI Communicating Agents. In Darzentas, J., Vouros, G., Vosinakis, S., and Arnellos, A., editors, *Artificial Intelligence: Theories, Models and Applications*, volume 5138 of *Lecture Notes in Computer Science*, pages 263–275, Berlin. Springer.

Santos, M. and Pitt, J. (2014). Emotions and Norms in Shared Spaces. In Balke, T., Dignum, F., van Riemsdijk, M. B., and Chopra, A. K., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems IX*, volume 8386 of *Lecture Notes in Computer Science*, pages 157–176. Springer, Berlin.

Sargent, R. G. (2013). Verification and validation of simulation models. *Journal of Simulation*, 7(1):12–24.

Savarimuthu, B., Cranefield, S., Purvis, M., and Purvis, M. (2010). Internal Agent Architecture for Norm Identification. In Padget, J., Artikis, A., Vasconcelos, W., Stathis, K., da Silva, V., Matson, E., and Polleres, A., editors, *Coordination, Organizations, Institutions and Norms in Agent Systems V*, volume 6069 of *Lecture Notes in Computer Science*, pages 241–256. Springer, Berlin.

Savarimuthu, B. T. R. (2011). *Mechanisms for norm emergence and norm identification in multi-agent societies*. PhD thesis, University of Otago, Dunedin, New Zealand.

Savarimuthu, B. T. R. and Cranefield, S. (2011). Norm creation, spreading and emergence: A survey of simulation models of norms in multi-agent systems. *Multiagent and Grid Systems*, 7(1):21–54.

Savarimuthu, B. T. R., Cranefield, S., Purvis, M., and Purvis, M. (2007). Norm emergence in agent societies formed by dynamically changing networks. In *2007 IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, pages 464–470.

Savarimuthu, B. T. R., Cranefield, S., Purvis, M., and Purvis, M. (2008). Role Model Based Mechanism for Norm Emergence in Artificial Agent Societies. In Sichman, J., Padget, J., Ossowski, S., and Noriega, P., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems III*, volume 4870 of *Lecture Notes in Computer Science*, pages 203–217. Springer, Berlin.

Savarimuthu, B. T. R., Cranefield, S., Purvis, M. A., and Purvis, M. K. (2013a). Identifying prohibition norms in agent societies. *Artificial Intelligence and Law*, 21:1–46.

Savarimuthu, B. T. R., Padget, J., and Purvis, M. A. (2013b). Social Norm Recommendation for Virtual Agent Societies. In Boella, G., Elkind, E., Savarimuthu, B. T. R., Dignum, F., and Purvis, M. K., editors, *PRIMA 2013: Principles and Practice of Multi-Agent Systems*, volume 8291 of *Lecture Notes in Computer Science*, pages 308–323. Springer, Berlin.

Schacht, J. (1983). *An Introduction to Islamic Law*. Oxford University Press, Oxford (UK).

Schelling, T. (1960). *The Strategy of Conflict*. Harvard University Press, Cambridge (MA).

Schelling, T. (1978). *Micromotives and Macrobehaviour*. W. W. Norton and Company, New York (NY).

Schlüter, A. and Theesfeld, I. (2010). The Grammar of Institutions: The challenge of distinguishing between strategies, norms, and rules. *Rationality and Society*, 22(4):445–475.

Schotter, A. (1981). *The Economic Theory of Social Institutions*. Cambridge University Press, Cambridge (MA).

Schumpeter, J. A. (1909). On the Concept of Social Value. *Quarterly Journal of Economics*, 23(2):213–232.

Schwartz, S. (2010). *Were the Jews a Mediterranean Society? Reciprocity and Solidarity in Ancient Judaism*. Princeton University Press, Princeton (NJ).

Scott, W. R. (2008). Approaching Adulthood: The Maturing of Institutional Theory. *Theory and Society*, 37(5):427–442.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, London.

Searle, J. R. (1995). *The Construction of Social Reality*. Allen Lane, London.

Searle, J. R. (2005). What Is an Institution? *Journal of Institutional Economics*, 1(1):1–22.

Selten, R. (2001). What is Bounded Rationality? In *Bounded Rationality: The Adaptive Toolbox*, pages 13–36. MIT Press, Cambridge (MA).

Sen, O. and Sen, S. (2010). Effects of social network topology and options on norm emergence. In Padget, J., Artikis, A., Vasconcelos, W., Stathis, K., da Silva, V., Matson, E., and Polleres, A., editors, *Coordination, Organizations, Institutions and Norms in Agent Systems V*, volume 6069 of *Lecture Notes in Computer Science*, pages 211–222. Springer, Berlin.

Sen, S. and Airiau, S. (2007). Emergence of Norms Through Social Learning. In Veloso, M., editor, *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, IJCAI'07, pages 1507–1512, San Francisco (CA). Morgan Kaufmann Publishers Inc.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4):591–611.

Shepsle, K. A. (1989). Studying Institutions: Some Lessons from the Rational Choice Approach. *Journal of Theoretical Politics*, 1(2):131–147.

Shepsle, K. A. and Weingast, B. R. (1984). When Do Rules of Procedure Matter? *Journal of Politics*, 46(1):206–221.

Shoham, Y. and Tennenholtz, M. (1992). On the synthesis of useful social laws for artificial agent societies. In *Proceedings of the 10th National Conference on Artificial Intelligence (AAAI '92)*, pages 276–281, San Jose (CA).

Shoham, Y. and Tennenholtz, M. (1997). On the emergence of social conventions: modeling, analysis, and simulations. *Artificial Intelligence*, 94(1–2):139–166.

Siddiki, S., Weible, C. M., Basurto, X., and Calanni, J. (2011). Dissecting Policy Designs: An Application of the Institutional Grammar Tool. *The Policy Studies Journal*, 39(1):79–103.

Simmel, G. (1902). The Number of Members as Determining the Sociological form of the Group II. *American Journal of Sociology*, 8(2):158–196.

Simmel, G. (1964). *The Sociology of Georg Simmel*. Free Press, New York (NY). Originally published: 1908.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1):99–118.

Singh, H., Gupta, M. M., Meitzler, T., Hou, Z.-G., Garg, K. K., Solo, A. M. G., and Zadeh, L. A. (2013). Real-Life Applications of Fuzzy Logic (Special Issue).

Singh, S. and Kearns, M. (2002). Near-Optimal Reinforcement Learning in Polynomial Time. *Machine Learning*, 49(2):209–232.

Situngkir, H. (2007). The Ribbon of Love: Fuzzy-Ruled Agents in Artificial Societies. BFI Working Paper Series, Department of Computational Sociology, Bandung Fe Institute, Bandung.

Skyrms, B. and Pemantle, R. (2000). A Dynamic Model of Social Network Formation. *Proceedings of the National Academy of Sciences of the United States of America*, 97(16):9340–9346.

Smajgl, A., Izquierdo, L., and Huigen, M. G. A. (2010). Rules, Knowledge and Complexity: How Agents Shape their Institutional Environment. *Journal of Modelling and Simulation of Systems*, 1(2):98–107.

Smajgl, A., Izquierdo, L. R., and Huigen, M. (2008). Modeling endogenous rule changes in an institutional context: The ADICO Sequence. *Advances in Complex Systems*, 2(11):199–215.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. W. Strahan and T. Cadell, London.

Sosis, R. (2005). Does Religion Promote Trust? The Role of Signaling, Reputation, and Punishment. *Interdisciplinary Journal of Research on Religion*, 1(7):1–30.

Squazzoni, F. (2008). The Micro-Macro Link in Social Simulation. *Sociologica*, 1:1–25.

Squazzoni, F. (2012). *Agent-Based Computational Sociology*. Wiley, Chicester (UK).

Stanford Encyclopedia of Philosophy (2011). Homosexuality. http://plato.stanford.edu/entries/homosexuality. Accessed on: 26th October 2014.

Stein, P. (1999). *Roman Law in European History*. Cambridge University Press, Cambridge (UK).

Sugawara, T. (2011). Emergence and Stability of Social Conventions in Conflict Situations. In Walsh, T., editor, *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, volume 1 of *IJCAI'11*, pages 371–378. AAAI Press.

Sumner, W. G. (1906). *Folkways: A Study of the Sociological Importance of Usages, Manners, Customs, Mores, and Morals*. Ginn and Co., Boston (MA).

Sun, R. (2003). A tutorial on clarion 5.0. http://www.cogsci.rpi.edu/ rsun/sun.tutorial.pdf. Accessed on: 20th August 2015.

Sun, R. (2007). Cognitive Social Simulation Incorporating Cognitive Architectures. *IEEE Intelligent Systems*, 22(5):33–39.

Sun, R. (2009). Cognitive architectures and multi-agent social simulation. In Lukose, D. and Shi, Z., editors, *Multi-Agent Systems for Society*, volume 4078 of *Lecture Notes in Computer Science*, pages 7–21, Berlin. Springer.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press, Cambridge (MA).

Szalay, L. B. (1981). Intercultural communication – a process model. *Journal of Intercultural Relations*, 5(2):133–146.

Szalay, L. B. and Deese, J. (1978). *Subjective Meaning and Culture: An Assessment Through Word Associations*. Lawrence Erlbaum Associates, Mahwah (NJ).

Tadelis, S. (2013). *Game Theory: An Introduction*. Princeton University Press, Princeton (NJ).

Telecom Italia (2015). JADE - Java Agent DEvelopment Framework. http://jade.tilab.com. Accessed on: 1st September 2015.

Tesfatsion, L. and Judd, K. L., editors (2006). *Handbook of Computational Economics - Agent-Based Computational Economics*, volume 2. North Holland, Amsterdam.

Testerink, B., Dastani, M., and Meyer, J.-J. (2014). Norms in Distributed Organizations. In Balke, T., Dignum, F., van Riemsdijk, M. B., and Chopra, A. K., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems IX*, volume 8386 of *Lecture Notes in Computer Science*, pages 120–135. Springer, Berlin.

Thakur, R. (2001). Global Norms and international humanitarian law: an Asian perspective. *International Review of the Red Cross*, 83(841):19–44.

Thomas, W. and Znaniecki, F. (1920). *The Polish Peasant in Europe and America*, volume 2. Octagon, New York (NY).

Thorndike, E. L. (1911). *Animal Intelligence*. Macmillan, New York (NY).

Tisue, S. and Wilensky, U. (2004). NetLogo: Design and Implementation of a Multi-Agent Modeling Environment. In *SwarmFest*, Ann Arbor (MI).

Trencansky, I. and Cervenka, R. (2005). Agent Modeling Language (AML): A Comprehensive Approach to Modeling MAS. *Informatica*, 29(4):391–400.

Troitzsch, K. G. (2009). Perspectives and challenges of agent-based simulation as a tool for economics and other social sciences. In *Proceedings of the 8th International Conference on Autonomous Agents and Multi-Agent Systems - Volume 1*, AAMAS '09, pages 35–42.

Tuomela, R. (1995). *The Importance of Us: A Philosophical Study of Basic Social Notions*. Stanford University Press, Stanford (CA).

Tuomela, R. (2002). *The philosophy of social practices: A collective acceptance view*. Cambridge University Press, Cambridge (UK).

Tuomela, R. (2003). Collective Acceptance, Social Institutions, and Social Reality. *American Journal of Economics and Sociology*, 62(1):123–165.

Turner, J. (1991). *Social Influence*. Open University Press, Buckingham (UK).

Udéhn, L. (2001). *Methodological Individualism: Background, History and Meaning*. Routledge, London.

Udovitch, A. L. (1962). At the origin of the Western commanda: Muslim, Israel, Byzantium. *Speculum*, 37(2):198–207.

Udovitch, A. L. (1967). Labor Partnerships in Early Islamic Law. *Journal of the Economic and Social History of the Orient*, 10(1):64–80.

Udovitch, A. L. (1970). *Partnership and Profit in Medieval Islam*. Princeton University Press, Princeton (NJ).

Udovitch, A. L. (1977). Formalism and informalism in the social and economic institutions of the medieval islamic world. In Banani, A. and Jr., S. V., editors, *Individualism and Conformity in Classical Islam*, pages 61–81. Harrassowitz, Wiesbaden.

Ullmann-Margalit, E. (1977). *The Emergence of Norms*. Clarendon Library of Logic and Philosophy. Clarendon Press, Oxford (UK).

Urbano, P., Balsa, J., Antunes, L., and Moniz, L. (2009). Force versus majority: A comparison in convention emergence efficiency. In Hübner, J., Matson, E., Boissier, O., and Dignum, V., editors, *Coordination, Organizations, Institutions and Norms in Agent Systems IV*, volume 5428 of *Lecture Notes in Computer Science*, pages 48–63. Springer, Berlin.

van Doosselaere, Q. (2009). *Commercial Agreements and Social Dynamics in Medieval Genoa*. Cambridge University Press, Cambridge (UK).

Vanberg, V. J. (2002). Rational Choice versus Program-Based Behaviour: Alternative Theoretical Approaches and Their Relevance for the Study of Institutions. *Rationality and Society*, 14(1):7–53.

Vázquez-Salceda, J., Aldewereld, H., and Dignum, F. (2004). Implementing Norms in Multiagent Systems. In Lindemann, G., Denzinger, J., Timm, I., and Unland, R., editors, *Multiagent System Technologies*, volume 3187 of *Lecture Notes in Computer Science*, pages 313–327. Springer, Berlin.

Veblen, T. (1899). *The Theory of the Leisure Class: An Economic Study of Institutions*. Macmillan, New York (NY).

Veblen, T. (1904). *The Theory of Business Enterprise*. Transaction Books, New Brunswick (NJ).

Verhagen, H. J. (2000). *Norm Autonomous Agents*. PhD thesis, The Royal Institute of Technology and Stockholm University, Stockholm, Sweden.

Viganò, F. and Colombetti, M. (2008). Model checking norms and sanctions in institutions. In Sichman, J., Padget, J., Ossowski, S., and Noriega, P., editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems III*, volume 4870 of *Lecture Notes in Computer Science*, pages 316–329. Springer, Berlin.

Villatoro, D., Andrighetto, G., Sabater-Mir, J., and Conte, R. (2011a). Dynamic Sanctioning for Robust and Cost-efficient Norm Compliance. In Walsh, T., editor, *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, volume 1 of *IJCAI'11*, pages 414–419. AAAI Press.

Villatoro, D. and Sabater-Mir, J. (2009). Categorizing Social Norms in a Simulated Resource Gathering Society. In Hübner, J., Matson, E., Boissier, O., and Dignum, V., editors, *Coordination, Organizations, Institutions and Norms in Agent Systems IV*, volume 5428 of *Lecture Notes in Computer Science*, pages 235–249. Springer, Berlin.

Villatoro, D., Sabater-Mir, J., and Sen, S. (2011b). Social Instruments for Robust Convention Emergence. In Walsh, T., editor, *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, volume 1 of *IJCAI'11*, pages 420–425. AAAI Press.

Villatoro, D., Sen, S., and Sabater-Mir, J. (2009). Topology and memory effect on convention emergence. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Volume 02*, WI-IAT '09, pages 233–240.

Vinogradoff, P. (1929). *Roman Law in Medieval Europe*. Oxford University Press, London, 2nd edition.

von Hayek, F. A. (1942). Scientism and the Study of Society - Part I. *Economica*, 9(35):267–291.

von Hayek, F. A. (1973). *Law Legislation and Liberty*, volume 1. University of Chicago Press, Chicago (IL).

von Neumann, J. and Morgenstern, O. (1947). *Theory of Games and Economic Behavior*. Princeton University Press, Princeton (NJ).

von Wright, G. H. (1951a). *An Essay in Modal Logic*. North Holland, Amsterdam.

von Wright, G. H. (1951b). Deontic logic. *Mind*, 60(237):1–15.

von Wright, G. H. (1963). *Norm and Action: A Logical Enquiry*. Routledge & Kegan Paul, London.

Vu, T. M., Siebers, P.-O., and Wagner, C. (2013). Comparison of Crisp Systems and Fuzzy Systems in Agent-Based Simulation. In *13th UK Workshop on Computational Intelligence (UKCI)*, pages 54–61.

Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and non-nested Hypotheses. *Econometrica*, 57(2):307–333.

Walker, S. T. (1998). *Both Pretense and Promise: The Political Economy of Privatization in Africa*. PhD thesis, Indiana University, Bloomington (IN).

Wallach, W. and Allen, C. (2010). *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press, New York (NY).

Waller, W. T. (1988). The concept of habit in economic analysis. *Journal of Economic Issues*, XXII(1):113–126.

Walpole, R. E., Myers, R. H., Myers, S. L., and Ye, K. (2012). *Probabilities & Statistics for Engineers & Scientists*. Prentice Hall, Boston (MA), 9th edition.

Watkins, C. (1989). *Learning from Delayed Rewards*. PhD thesis, Cambridge University, Cambridge (UK).

Watkins, C. and Dayan, P. (1992). Technical Note: Q-Learning. *Machine Learning*, 8(3-4):279–292.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442.

Weber, M. (1978a). *Economy and Society*. University of California Press, Berkeley (CA). Originally published: 1922.

Weber, M. (1978b). *Selection in Translation*. Cambridge University Press, Cambridge (UK).

Wellman, M. (2015). Putting the Agent in Agent-Based Modeling (Edited transcript of a talk presented at the 13th International Conference on Autonomous Agents and Multi-agent Systems (Paris, France), May 2014, on receipt of the ACM/SIGAI Autonomous Agents Research Award). http://web.eecs.umich.edu/srg/?page_id=1579. Accessed on: 1st September 2015.

Westerhoff, F. and Franke, R. (2012). Agent-based models for economic policy design. BERG Working Paper 88, University of Bamberg, Bamberg.

White, J. (1997). Mobile agents. In Bradshaw, J., editor, *Software Agents*, pages 437–472. AAAI Press.

Wieringa, R. and Meyer, J.-J. (1993). Applications of deontic logic in computer science: A concise overview. In *Deontic Logic in Computer Science: Normative System Specification*, pages 17–40, New York (NY). John Wiley & Sons.

Williamson, O. E. (1975a). *Markets and Hierarchies, Analysis and Antitrust Implications: A Study in the Economics of Internal Organization*. Free Press, New York (NY).

Williamson, O. E. (1975b). *The Economic Institutions of Capitalism*. Free Press, New York (NY).

Williamson, O. E. (1998). Transaction Cost Economics: How it works; Where it is headed. *De Economist*, 146(1):23–58.

Williamson, O. E. (2000). The New Institutional Economics: Taking Stock, Looking Ahead. *Journal of Economic Literature*, 38(3):595–613.

Windrum, P., Fagiolo, G., and Moneta, A. (2007). Empirical Validation of Agent-Based Models: Alternatives and Prospects. *Journal of Artificial Societies and Social Simulation*, 10(2):8.

Wu, D. and Mendel, J. (2014). Designing practical interval type-2 fuzzy logic systems made simple. In *2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 800–807.

Wu, D. and Mendel, J. M. (2009). A Comparative Study of Ranking Methods, Similarity Measures and Uncertainty Measures for Interval Type-2 Fuzzy Sets. *Information Sciences*, 179(8):1169–1192.

Wu, D. and Tan, W. (2006). Genetic Learning and Performation Evaluation of Type-2 Fuzzy Logic Controllers. *Engineering Applications of Artificial Intelligence*, 19(8):829–841.

Xenophon (1914). *Cyropaedia: The Education of Cyrus*. William Heinemann Ltd., London. Originally published: between 431 and 355 BC.

Yager, R. R. (1993). Families of OWA operators. *Fuzzy Sets and Systems*, 59(2):125–148.

Young, H. P. (1993). The Evolution of Conventions. *Econometrica*, 61(1):57–84.

Young, H. P. (1998). *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press, Princeton (NJ).

Yu, C., Zhang, M., Ren, F., and Luo, X. (2013). Emergence of Social Norms Through Collective Learning in Networked Agent Societies. In *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems*, AAMAS '13, pages 475–482.

Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control*, 8(3):338–353.

Zadeh, L. A. (1973). Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(1):28–44.

Zadeh, L. A. (1975a). The Concept of a Linguistic Variable and its Application to Approximate Reasoning - I. *Information Sciences*, 8(3):199–249.

Zadeh, L. A. (1975b). The Concept of a Linguistic Variable and its Application to Approximate Reasoning - II. *Information Sciences*, 8(4):301–357.

Zadeh, L. A. (1996). Fuzzy Logic = Computing with Words. *IEEE Transactions on Fuzzy Systems*, 4(2):103–111.

Zadeh, L. A. (1999). From Computing with Numbers to Computing with Words – From Manipulation of Measurements to Manipulation of Perceptions. *IEEE Transactions on Circuits and Systems - I: Fundamental Theory and Applications*, 45(1):105–119.

Zeigler, B. (1976). *Theory of Modeling and Simulation*. Wiley Interscience, New York (NY).

Zeigler, B. P., Kim, T. G., and Praehofer, H. (2000). *Theory of Modeling and Simulation*. Academic Press, Orlando (FL), 2nd edition.

Zerubavel, E. (1980). If Simmel Were A Fieldworker: On Formal Sociological Theory And Analytical Field Research. *Symbolic Interaction*, 3(2):25–34.

Zschache, J. (2012). Producing public goods in networks: Some effects of social comparison and endogenous network change. *Social Networks*, 34(4):539–548.

Appendices

A

Mediterranean Free Trade Debate

In this section we explore a further aspect in which Greif's theses appear polarised and require clarification, if not revision. For this purpose we will look at the circumstances that entailed trade across the Mediterranean, addressing the question of constraints that affected the interaction between the different societies that shared their connection to the Mediterranean sea as their commonality.

A.1 The 'Mediterranean Studies'

The thesis of an inclusive view on the Mediterranean region has been championed by Goitein and Braudel, despite their different research foci. Goitein's understanding as "sociographer" (Astren, 2012)[1] made the *individual* subject of his analysis, while Braudel (1996) emphasised the *longue durée*. Braudel put an emphasis on the long-term historical development as opposed to the concentration on individual events, let alone the anecdotal individuals that were at the centre of Goitein's analysis. We will not enter an extended debate on the associated issue of the 'Mediterranean identity' – the question whether something like a common archetypical 'Mediterranean type' existed. We leave this highly disputed concern to histo-

[1]Goldberg: "... Goitein was a compiler, ..." (Goldberg, 2011).

rians and social anthropologists[2] with Goitein (2000a) and Braudel (1996) advocating this view, and critical voices from central figures in the field of Mediterranean history, such as Rostovtzeff (1957, 1986) and in particular Pirenne (1956), who formulated what is famously known as the 'Pirenne thesis'. Pirenne claimed that the European Medieval civilisation was fundamentally influenced by the Arab expansion along the North African coast and what is now known as the Middle East as well as Spain and Sicily from the 7th century onwards. In Pirenne's view this expansion ultimately led to the demise of the devolving remains of the Roman empire, which in consequence, cut off Western Europe from the previous vivid trade interactions with the Eastern Mediterranean, and left Europe as an agrarian society at subsistence level. According to Pirenne, this isolation from cultural and technological influences based on interaction enabled Charlemagne to establish the Carolingian Empire that defined a new, Western-rooted civilisation and governance, controlling the chaos that had dominated Europe after the decline of the Roman empire. To return to the discussion, this theory challenges the concept of Mediterranean unity, according to which such identity could have hardly existed, given that the Geniza material Goitein used for his analysis set in well after the beginning of the Islamic expansion. However, the Pirenne thesis itself is not undisputed, generally suggesting the overemphasis of the influence of Arabic trading. However, Goitein did not discuss those inconsistencies (Horden and Purcell, 2000). A last noteworthy source for a comprehensive historical overview of Mediterranean history, with extensive consideration of geographical characteristics (similar to Braudel), is the work by Horden and Purcell (2000), who integrate the different theories by providing an updated overview of geographical, economic, political and religious aspects of the Mediterranean societies and offering information that challenges the previous 'Mediterranean studies'.

A.2 Goitein's Mediterranean Unity Thesis

Although the previous excursus appears as a deviation, it is closely related to the investigation of some of the assumptions building the foundation for Greif's analysis, foremostly the absence of institutional and environmental boundaries that could have prevented free trade among all Mediterranean communities.[3] Given the absence of trade limitations, Greif (1994) suggests that the Maghribīs' unwillingness to extend their trade relationships beyond

[2]The possibly most cynical approach to summarise commonalities is offered by Peristiany (Peristiany, 1966) who suggested that the moustache was a essential commonality one would observe with males around the Mediterranean, implicitly challenging the lacking discrimination of correlation and causation (here: an actual common history) by advocates of the existence of 'the Mediterranean'.

[3]"The *geniza* documents indicate that eleventh-century Mediterranean trade was free, private, and competitive. ..." (Greif, 1989, 2006)

coalition boundaries could only be explained by incompatible cultural beliefs (as discussed in Subsection 3.2.2), which supports his bifurcated concentration on contrasting cultural factors. The presumed absence of legal constraints in combination with the lucrative trade opportunities Maghribī traders could expect when trading with al-Rūm (Arabic: 'the land of the Romans'),[4] a culture-rooted deterrent for interactions between Southern Europeans and Maghribīs must have been significant.

In this context it is important to realise that Greif's work strongly relies on Goitein's authoritative analysis of the Geniza documents. A central aspect of Goitein's studies is the understanding of the Mediterranean as a culture of unity[5] – a state he considered to persist until around 1050 AD (Goitein, 1960),[6] despite the clear religious, ethnic and political boundaries that separated the Mediterranean. He saw this unity grounded in a set of complementary legal, socio-economic and historical factors (Goitein, 1960):

- the *Enforcement of Law as a Individualised Matter* – Individuals were not judged based on the territorial rules, but rather based on the litigant's origin or religious affiliation. In Goitein's view the cross-fertilising effect of this was to build a conception of commonality across the vast area surrounding the Mediterranean sea (or, to emphasise its binding natural feature simply referred to as "the sea" (Astren, 2012)).

- the *Trading Activity as Social Foundation* – The communities around the Mediterranean shared a spirit of trade, a remainder that Goitein considers to be rooted in the "bourgeois revolution" (Goitein, 1960) in the eight and ninth century that, in his view, had effectuated the transformation into stronger integrated societies. This was greatly facilitated by the connecting nature of trade, enabling (and likewise enabled by) what Goitein saw as a "free trade community" (Goitein, 2000b).

- the *Common History of the Mediterranean Societies* – Goitein recognises the shared roots of the different societies, which he associates with Sumer and Akkad, along with roots reaching into what today is Iran. Surprisingly he acknowledges but disregards the fact that the Mediterranean in its entirety had only been 'unified' by the Roman Empire.

[4]The Geniza offers anecdotes of the possibility to sell pepper of inferior quality in Europe (Goitein, 2000b). In another case a Maghribī trader managed to sell Indian brazilwood to a European trader at a profit of 150 percent (Greif, 1989).

[5]This theme is well-reflected in the title of his five volume seminal work 'A Mediterranean Society ...'.

[6]For Goitein (1960) the crusades disrupted the culture of unity which he considered to persist until around 1050 AD. This is in stark contrast to Pirenne (1956), who saw the Arab conquests of Northern Africa, Middle East, and the conquering of Spain in the seventh century onwards as fundamental for the breach of that unity, with the crusades and the onset of the Spanish reconquista as corresponding reactions.

Those aspects, that are underlying Goitein's work and support the idea of Mediterranean unity bear issues of concern with respect to the inclusion of the European societies into his conception of 'unity'.

Law as Individualised Matter Inasmuch as law was an individualised affair in the Islamic world, with the coexistence of Jewish jurisprudence with an overarching legal framework based on the Ḥanafī school (Udovitch, 1970; Ackerman-Lieberman, 2014) that spread across and beyond the Fatimid Empire, the legal systems of Western Europe highlighted the personal principle and coexisted with local customary laws.[7] However, from the sixth century onward, private law in the remains of what consisted of the former Western Roman Empire was increasingly replaced by local variants promoted by the respective conquerors, known as *vulgarization of Roman law* (Kaser, 1965).[8] Instead of relying on a system of courts, freemen were to judge individuals' disputes, with success often relying of outnumbering the oaths given in favour of a litigant (Stein, 1999).

Only the rediscovery of the Justinian code by 11th century Italian scholars, drove the forming nation states but also the demand for regulating trade, marked its return in a modified form into the legal systems of the forming Western nation states, replacing locally varying customary rules with stronger legal unification, marking a shift from 'personality' to the concept of 'territoriality' as a fundamental legal principle (Braithwaite and Drahos, 2000). More significant with respect to its ancestry of international law is the unification of rulings based on occupational needs. The demand for a unification of occupation-specific law, in particular serving the emerging trader class, was fundamental for the reduction of uncertainty in long-distance trade, such as the provision of security guarantees to traders in foreign territories. The *lex mercatoria*, or *law merchant* (Milgrom et al., 1990) was originally based on trade customs that were increasingly adopted (e.g. contractual obligations) and enforced by local rulers (e.g. settlement of disputes, legal protection for foreign traders), with the Champagne fairs (Milgrom et al., 1990; Braithwaite and Drahos, 2000; Baldwin, 1968) being the prototypical example. Legal principles based on personality would not have served the diverse international trader community that sought for lowered transaction costs as part of their increasing professionalism.

[7] Vinogradoff (1929) highlights an interesting example: "[The situation was] very intricate when members of different nationalities, living under different laws, were brought together to transact business with each other. As Bishop Agobard of Lyons tells us about 850, it happened constantly that of five people meeting in one room, each followed a law of his own." (Vinogradoff, 1929).

[8] Braithwaite and Drahos (2000) suggest the onset of vulgarization after the publication of the main parts of the actual *Justinian Corpus Iuris Civilis* at around 533/4 AD, manifested by the capturing of vast territory by the Lombards in 568 AD, who were of Germanic origin.

This leaves us with the observation that European law in the time frame observed by Goitein was – as he suggests – still largely based on the personality principle, with a shift towards territoriality in Italy in the eleventh century, which coincides with the essential core of trade-related Geniza material.[9]

Trading Activity as Social Foundation Goitein's second characteristic, the notion of trade as social foundation for the 'unity' thesis is undoubted. Using the Genoese society as a representative for Southern European societies, Goitein's second thesis finds support in the simple phrase "Ianuensis ergo mercator" (Genoese therefore merchant) (Airaldi, 1969), although being predated and paired with piracy activities as a major source of income (Greif, 2006; van Doosselaere, 2009).

The importance of trading, but also its cross-fertilising effect is seen in the most important institutional instrument that Italian traders had adopted from the Eastern Mediterranean, the *commenda* (as discussed in Section 3.1.3). Lombard (1948) suggests that trade interactions between Europeans and Muslims continued well after the trade interruptions that Pirenne had associated with the Arab conquests (in the 7th century onwards). Instead of ceasing trade altogether, Lombard noted the replacement of Syrian Muslim traders with other, often Jewish, Eastern traders that served the Southern European Mediterranean ports (Citarella, 1971). Inasmuch as Europeans demanded the influx of luxury goods such as silk, the Fatimid Empire had a constant demand in natural resources, such as wood, which was scarcely available along the North African coast, but, at that time, abundant in the European forests (Lombard, 1948).

While Arabic vessels regularly frequented European ports, Venetian ships ventured to the Eastern Mediterranean Arabic ports from around 850 AD onwards, and Genoese traders only actively participating in the eleventh century, with the First Crusade marking the onset (Patrone and Airaldi, 1986). Independent of the comparatively late participation in long-distance trade the Genoese' self-understanding is reflected by their leaders, being "governments of the merchants, by the merchants, for the merchants" (Lopez, 1976).

Common History of the Mediterranean Societies The final aspect, the suggestion of a common cultural background based on the ancient societies of the Sumer and Akkad indicates a focal point of Goitein's analysis.

Goitein's perspective is centred in the Middle East, with particular focus on the symbiotic Islamic-Jewish cultural relationship (Goitein, 1974) fostered by the Maghribīs. The indirect

[9]Goldberg (2012c) allocates the highest number of documents between 1040 and 1080 AD.

influences onto European society (numeric system; institutional instruments as explored in Subsection 3.1.3) are undisputed, but the hostile nature in which Southern Europeans and Islamic rulers coexisted challenges the concept of Mediterranean unity. Examples include Genoa's Saracen raids in the tenth century (van Doosselaere, 2009) followed by constant maritime warfare with Western Islamic territories (Airaldi, 1997), and Genoa's active maritime participation in the crusades in the eleventh century as an example of offensive military action (Jehel, 1995; Balard, 2001; van Doosselaere, 2009). It thus appears unclear if Goitein's conceptualisation of the Mediterranean society entailed Southern Europeans. This understanding is supported by the fact that Goitein's time frame of analysis is defined by the trade-related Geniza writing, starting in the 990 AD and ceding in the early 13th century, well after the Arab conquest of Northern Africa - and thus the Maghreb, the place of origin of the Maghribīs. An ethnocentric view on the Fatimid Empire, a view that extended to the East (up to what is nowadays Iran), well beyond the immediate Mediterranean (see e.g. Goitein (1960); Astren (2012)) might have led Goitein to suggest an understanding of free-trade that was mostly concerned with the trade across the Islamic part of the Mediterranean, including al-Andalus, Sicily and the North African coast up to the Eastern Mediterranean, which captured roughly half of the Mediterranean Basin.

While Greif largely builds on Goitein's work, he grounds his views on Fatimid Empire and Southern Europeans on cultural traits, the collectiveness but also individualism. Assuming this differing cultural background indirectly challenges the thesis of common history Goitein entertains as the last pillar of Mediterranean unity.

A.3 Alternative Influence Factors

Instead of fully committing to a particular thesis, such as pure focus on cultural differences as trade deterrent for Maghribīs and Southern Europeans (as advocated by Greif (2006)), let us assume a more factual position and widen the range of possible factors unveiled by recent literature.

We can broadly structure those into *political/institutional environment* and *technological environmental* factors.

A.3.1 Institutional Factors

Laying out a refined picture of the *institutional environment* the Maghribīs faced, Goldberg (2011, 2012a) highlights some of the regulatory constraints that limited the Maghribīs' ability to act even *within* the Fatimid Empire, such as the higher tax charge for non-locals, and double the customs charge for non-Muslims in particular. Trade *outside* the Fatimid Empire, however, underlay even stronger discrimination, with the prohibition to trade holy relics or slaves, and attempts to exclude Jews from sea trade in general.[10] Looking at the regulatory constraints within and outside the empire, the Mediterranean Basin could hardly be seen as a free-trade zone, but as Goldberg puts it, "... a world in which special privileges are acquired by mixed claims of locality, economic power, and personal relationships, and in which privilege and power can be shared, and indeed formed part of a merchant's worth and currency ..." (Goldberg, 2011).

The Maghribīs sought empowerment in mutual associates that facilitated trade, and offered concrete benefits such as guaranteed free storage opportunities to fill gaps in contract enforcement in long-distance trade, in short: their power lay in the their relationship network, the aṣḥābunā. An environment of inconsistent location-dependent sovereignties[11] and privileges could have challenged the expansion of the Maghribī network, not only for establishing presence in the first place but also to do so to a significant extent. To assure a functioning monitoring and private-order enforcement system, a critical amount of traders would have needed to settle in Southern European trade locations in order to witness goods receipt and processing. Extending the Maghribīan monitoring system to Genoa would have required a multilateral decision to establish multiple homesteads at the same time, an endeavour that would have implied considerable risk, especially given the less stable institutional environment[12] and limited reach of individual sovereignties, such as the Genoese or Venetian city states.

From an institutional perspective, the incentives for Maghribīs to engage in long-running engagement with Genoese were thus limited, despite the promising margins realised when trading with remote traders in the local vicinity (see e.g. Goitein (2000b)).

A.3.2 Technological Factors

Horden and Purcell (2000) derived another *technological* aspect considered neither by Goitein, Greif nor Goldberg. They find that the naval technology available to the Maghribīs was ill-fitted for long-distance journeys, with travels taking the character of *cabotage*, thus travelling

[10] Astren (2012) refers to an imperial decree documented by Linder (1997) that granted Venetians trade privileges in Constantinople, under the condition that no Jews, Lombards, Amalfitans and merchants from Bari were onboard of vessels.

[11] The above-mentioned decree with its refined discrimination based on trader origin serves as a good example.

[12] The central government was considered dysfunctional and could only provide poor protection against outside threats (Greif, 2006).

close to the coast and stopping at nearly any harbour along the route to replenish water; the heavy weight of the prevalent amphorae – wood caskets had yet to be introduced – did not permit sufficiently large volumes that would have allowed crossing the open sea (Horden and Purcell, 2000; Astren, 2012). Likewise larger ships capable of crossing the Mediterranean, though existing, were only introduced in the Mediterranean after the Maghribī traders' activity had ended. These technological limitations were complemented by sailors' reluctance to travel into the open sea (Astren, 2012).

Thus, considering the bathymetric map shown in Figure A.1,[13] cabotage made the travel from Alexandria to al-Andalus via Sicily retraceable.[14]

Source: Encyclopædia Britannica (2009)
By courtesy of Encyclopaedia Britannica, Inc., Copyright 2009; used with permission.

Figure A.1: Bathymetric Overview of the Mediterranean

However, from Sicily Maghribīs could have likewise ventured along the Western Italian coastline up to Genoa. However, doing so, they would have crossed the territorial boundaries that existed between the Fatimid Empire and Southern Europe, not only risking to be engaged in military threats, but also remain under constant threat of piracy in those waters, a trade-deterring aspect acknowledged by Goitein (1960). From a Maghribī perspective seabound travel, be it to Genoa, Pisa, or Venice, would have thus incurred significant travel along a rugged coastline through foreign territory with frequently changing sovereignties in

[13]Although the map reflects the current depth levels (2009), eventual deviations from the historical bathymetry can be neglected, given that sea-level rise since Roman times (2000 years ago), i.e. around one millennium before the Maghribīs' operations, amounts to around 1.35 meter (Lambeck et al., 2004).

[14]Note the relatively shallow Strait of Sicily in Figure A.1.

the form of city states and warlords, hardly integrated by a weak central government (Greif, 2006). In fact, travelling to Southern Europe meant to transgress into an institutional environment whose boundaries were by no means as coherent and unified as it was the case for the Fatimid Empire; the particularised remnants of the Roman Empire did not offer a unified institutional backing,[15] leading traders to be more aggressive (which is exemplified in Genoa's history in piracy (Epstein, 1996; Schwartz, 2010)), and drove the fierce economic competition among the Italian city states, such as Venice, Pisa and Genoa (see e.g. González De Lara (2008)). This shifted the institutional trajectory towards increasing militarisation, both for political reasons (such as defence) as well as economic reasons (e.g. guarding convoys of trade vessels) (González De Lara, 2008). Development of advanced naval technology was essential for a community that inherently built on maritime trade, and, given the varying political environments but also more challenging coastline,[16] could not rely on cabotage travel. Furthermore, the competition between the Italian city states for the domination of foreign trade posts intensified piracy in the Mediterranean waters (Katele, 1988). The Maghribīs, as a minority in their own society, did not have such resources, and given their distributed operation, such investment may have hardly been useful, especially since they could rely on their own culturally and institutionally homogeneous stable environment.

A.3.3 Concluding Thoughts

Looking at the amounting evidence, we see a fuzzy picture that interleaves cultural, political (including institutions in the wider sense, such as legalistic mechanisms to maintain societal organisation), environmental and institutional differences (institutions in the narrow sense, such as specific institutional instruments to control economic cooperation) – aspects that extend beyond Greif's separation based on cultural differences. This suggests the existence of much more tangible concerns Maghribīs would have needed to take into account when interacting with Southern European traders, an aspect that motivates us to relax Greif's high-level assumption of cultural incompatibility as part of our analysis.

[15] At this stage recall the vulgarization of Roman law with Pagan law.

[16] Note the rocky coastline of the Peleponnes and the climatically exposed Kythirian Straits that connect the deep Ionian sea with the Aegaen Sea.

B
Conceptual Overview of the Simulation Environment

The simulations developed as part of this model share a uniform conceptual basis that has been adapted to accommodate the specifics of the presented simulation scenarios.

B.1 Simulation Environment

As its foundation the environment uses the Java-based simulation toolkit MASON (Luke et al., 2005) (Version 17), developed at the George Mason University that provides the core scheduling functionality. It guarantees fair random scheduling[1] of *Steppables* as the basic execution unit.

[1]MASON relies on the Mersenne Twister (Matsumoto and Nishimura, 1998) implementation of MASON's sister project, the evolutionary computation toolkit ECJ (Luke, 2015).

B.2 Agent Implementation

Agents are Java classes that implement the Steppable interface without constraining access to JVM and third party libraries. The individual agent execution cycles are implemented as described in the Algorithms provided in Subsections 4.2.1, 4.2.2, and Section 6.3.

Beyond the execution semantics, the implementation of communication, network links and memory are of relevance for the reconstruction of scenarios.

B.3 Communication

Agents are uniquely identified by IDs, which are used to deliver messages. We further maintain a directory service that holds all registered agents, and can be used to randomly pick fellow agents.

The communication between individual agents is modelled using direct method calls on respective other agents, which are resolved using the directory service mentioned above. To ensure deterministic and fair execution, communication (between agents and directory services as well as between agents) occurs synchronously.

B.4 Network Links

The initial two scenarios depend on explicit inter-agent relationships.

In the Investor Communication Model (Subsection 4.2.1) those are generated during initialisation and remain static over runtime. We use the GraphStream (Balev et al., 2014) library (Version 1.1) to generate those (using the provided network generation algorithm implementations). Internally, agents maintain lists of their respective partners, which they draw upon for randomised interaction.

In the Apprenticeship Model (Subsection 4.2.2) relationships dynamically change at runtime and are created whenever an apprenticeship is established or new ṣuḥba partners are found. Dissolution of links occurs whenever a partner is identified as a cheater, or upon any partner's death.

B.5 Memory

The implementation of memory modules varies in a simulation-dependent manner. The first two simulation models in Subsections 4.2.1 and 4.2.2 rely on primitive memory representations.

The Investor Communication Model (Subsection 4.2.1) maintains information about cheating operators and fellow investors using a fixed length queue with FIFO semantics, and is parameterised as specified in the context of the model description. Repeated entries (i.e. information about the same individual) are aggregated and treated as the most recent (i.e. last added) entry.

The Apprenticeship Model (Subsection 4.2.2) maintains information about aṣḥāb memberships and announced cheaters as hash maps. Traders do not lose information about identified cheaters throughout their lifetime.

The later experiments (Chapter 6 onwards) rely on a Java-based Q-Learning implementation (see Subsection 2.3.3) that captures action-reaction pairs as state-action sequences. The parameterisation is described in the context of the experiment in Section 6.3.

B.6 Deriving nADICO Statements

B.6.1 From Q-values to nADICO Statements

To derive the ascribed normative understanding as described in Subsection 6.2.5, individual `action-reaction` pairs are extracted along with the associated Q-values.

To achieve this, all `action-reaction` pairs are decomposed and grouped by `action`, with Q-values attached to the respective `reaction`.

The individual reactions (read: sanctions) are represented in a data structure representing institutional statements of type `ADIC` (sanction-less norm; see Section 5.3).

The individual components are mapped as follows:

- `A` – Since no specific information about sanctioners is available, the `A` is represented as a wildcard (*), reflecting the unknown nature of the sanctioner (see Subsection 5.3.4).
- `D` – The deontic value holds the Q-value associated with the `reaction`.
- `I` – The aim holds the string representation of the reaction (e.g. `reward`).
- `C` – Since our model does not explicitly capture context, the value of this component is not constrained to a specific condition. Recall that the specification of this component is optional (see Section 5.2).

All generated `ADIC` statements are then combined using *or* operators and attached as consequences to a leading `ADIC` statement that captures the `action` (`I`) and the actor's role in the `A` component (e.g. Maghribi, Investor, or Operator). Similar to the nested statements, the conditions (`C`) are not explicitly specified.

Following this the deontic value that is to be attached to the Deontic component (`D`) of the monitored institutional statement is derived from the individual Q-values using one of the strategies introduced in Subsection 6.2.5.

B.6.2 Deriving the Deontic Range

The deontic range is used to resolve deontic values to compartments. It is generated based on the mean value across a sliding window of past maximum and minimum Q-values. Deontic compartments are then dynamically calculated based on the deontic range configuration as described in Subsection 6.2.5, i.e. either by deriving equally sized symmetrically allocated compartments, or by enforcing a zero-based range centre and separately distributing the positive and negative range equally across the respective compartments.

B.6.3 Representing Stability

Since the Dynamic Deontics concept captures the institutional characteristic of stability, generated nADICO statements whose deontic values penetrate the tolerance zones around the deontic extremes (i.e. are relatively large or relatively small) are cached. This allows to establish the number of simulation rounds throughout which those statements continuously penetrated the tolerance zone in order to designate those statements as either stable obligation or prohibition norms.

The same approach is used to monitor those statements' deontic values for deviation from the deontic extremes, before disestablishing the respective statements' obligating or prohibiting natures, and readmitting them to the fluid permissive part of the deontic range.

B.7 Fuzzy Sets

The core Java-based IT2FLS Module implementation described in Chapter 7 largely follows the MatLab reference implementation provided by Liu and Mendel (2008b), apart from introducing richer object representations, fixing bugs, and extending the configurability of the original implementation. However, we extended it with visualisation capabilities (using JFreeChart) that can capture individual input intervals and the resulting Interval Type-2 Fuzzy Sets individually as well as in a comprehensive dashboard overview.

The module interface is accessible to individual agents, who generate their input intervals from the individual deontic compartment boundaries and submit those for fuzzy set generation.

B.8 Data Collection

The data collection is facilitated by a Statistics module that is scheduled independently from the individual agents and collects the dependent variable values at the end of each simulation round (i.e. after randomly scheduling all agents). The output is written to CSV files that are analysed using the statistics toolkit R (Version 3.0.3).

The actual simulation runs were performed using Java 8 (Update 25) running on Microsoft Windows 7.

Figure B.1 schematically integrates the discussed architectural components, with the memory component abstractly representing the specific memory types for the different simulations, and, where applicable, the derivation of nADICO statements and deontic range.

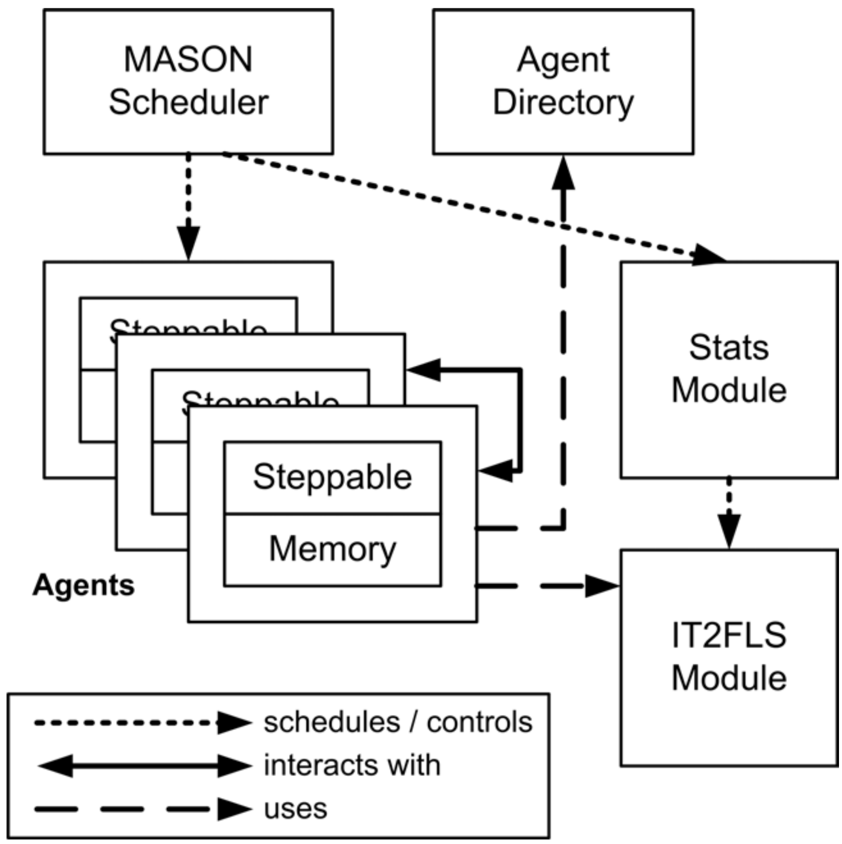


Figure B.1: Architectural Overview of the Simulation Environment

C
Genoese Communication Model

C.1 Sensitivity Analysis

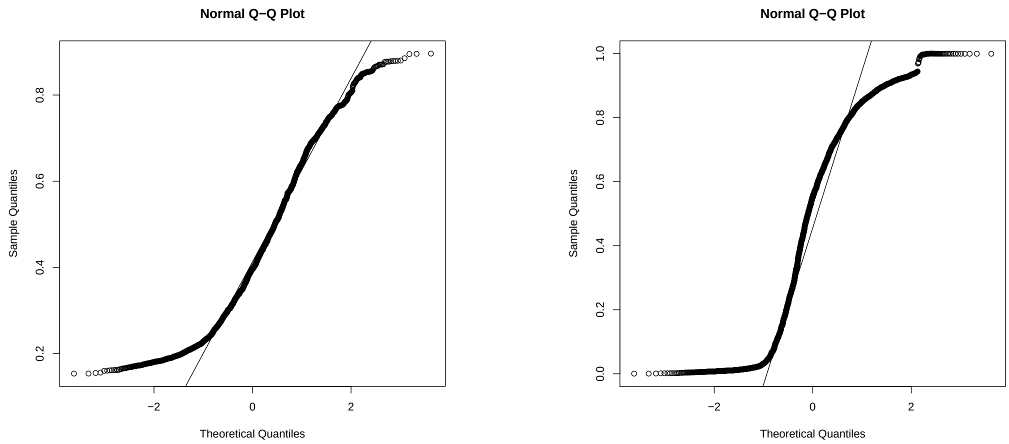
To advise the selection of values for parameters that are central to our investigation, we explore the impact of individual inputs on the dependent variables *effectivity* and *efficiency*. We thus perform parameter sweeping with focus on the parameters *memory size* (*memoryEntries*), *max. number of advice requests* (*maxRequests*) and *initial probability of truthful reporting* ($p_{truthful}$) across all simulation scenarios (different network types, Genoese and Maghribī communication patterns as introduced in Subsection 4.2.1). The ranges for the individual parameters are specified in Table C.1.

Table C.1: Parameter Ranges for Sensitivity Analysis

Parameter	Range
rq	5-20 (step size: 5)
tth	0.5-0.7 (step size: 0.1)
mem	10-50 (step size: 10)

Since the results do not follow a normal distribution as established based on the Shapiro-

Wilk (Shapiro and Wilk, 1965) and Anderson-Darling (Anderson and Darling, 1952) tests[1] (for visual confirmation refer to the respective Q-Q-Plots in Figure C.1), we rely on non-parametric tests to determine the relationship of input variables to outputs.



(a) Q-Q-Plot for Effectivity (b) Q-Q-Plot for Efficiency

Figure C.1: Q-Q-Plots for Effectivity and Efficiency

Table C.2 shows the correlations (based on Spearman's ρ) between the individual input variables *memory size* (abbreviated as 'mem' in the following tables and figures), *max. number of advice requests* ('rq'), and *initial probability of truthful reporting* ('tth') with respect to effectivity ('effectivity') and efficiency ('efficiency'). A visual representation is provided in the correlogram in Figure C.2. Observing the results, all independent variables have negligible (< 0.3)[2] or low positive correlation (between 0.3 and 0.5) with effectivity. For efficiency we can observe a moderate positive correlation (between 0.5 and 0.7) for the initial probability of truthful reporting ('tth'). Furthermore, all correlations – apart from the relationship of memory size ('mem') and efficiency – are significant.[3] At this stage, the central determinant for efficiency, and to a lesser extent effectivity, is the initial probability of truthful reporting.

Table C.2: Correlation across all Network Types

Parameter	effectivity	p value	efficiency	p value
rq	0.311	0	0.224	0
tth	0.204	0	0.609	0
mem	0.313	0	0.02	0.236

[1] In all cases the p values approximate 0.

[2] We base our interpretation of correlation strength on Hinkle et al. (2003).

[3] The correlation between effectivity and efficiency as observable in the correlogram in Figure C.2 can be ignored, since both are interrelated dependent variables as introduced in Subsection 4.2.1.

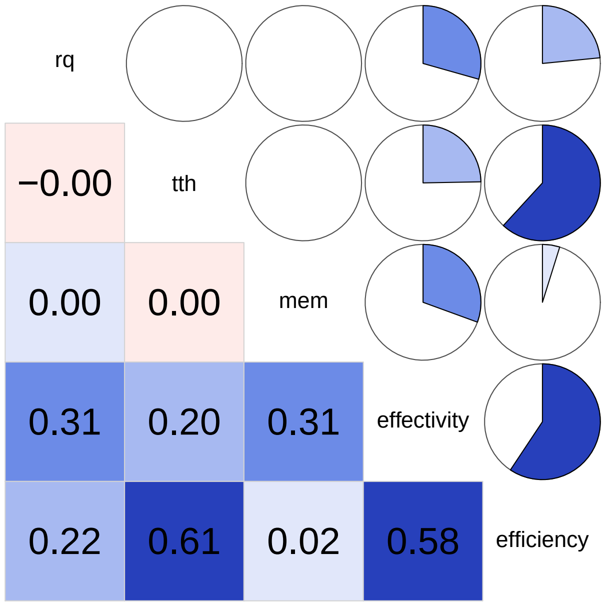


Figure C.2: Correlogram across all Network Types

Since we perform and evaluate simulation runs across different network topologies, it is important to isolate the sensitivity for individual network configurations to identify variables that we would otherwise mistakenly dismiss as irrelevant. We thus provide differentiated analyses for individual network topologies below, with FXD indicating fixed random directed networks (Table C.3 and Figure C.3a), FXDM referring to fixed random undirected networks (Table C.4 and Figure C.3b), WS to small-world networks (Table C.5 and Figure C.4a), and BA to scale-free network configurations (Table C.6 and Figure C.4b).

Table C.3: Correlations in Directed Random Networks (FXD)

Parameter	effectivity	p value	efficiency	p value
rq	0.342	0	0.246	0
tth	0.262	0	0.709	0
mem	0.326	0	0.05	0.146

However, the differentiated results offer consistent support for the initial probability of truthful advising as the essential independent variable, with strongest support for the fixed directed network (FXD) (high positive correlation with efficiency) and lowest correlation for the scale-free network (moderate positive correlation with efficiency). All other independent variables show low correlation levels (i.e. below 0.5) at best. For the purpose of our simulation, the parameterisation of memory size ('mem') and number of advice requests ('rq') is

Table C.4: Correlations in Undirected Random Networks (FXDM)

Parameter	effectivity	p value	efficiency	p value
rq	0.297	0	0.264	0
tth	0.192	0	0.642	0
mem	0.301	0	0.026	0.437

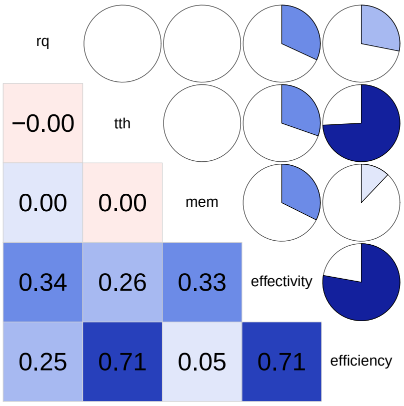
Table C.5: Correlations in Small-World Networks (WS)

Parameter	effectivity	p value	efficiency	p value
rq	0.302	0	0.27	0
tth	0.185	0	0.588	0
mem	0.265	0	-0.05	0.145

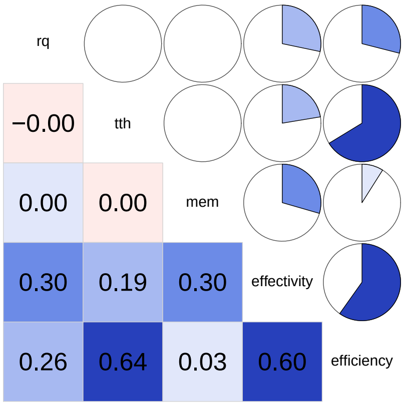
Table C.6: Correlations in Scale-Free Networks (BA)

Parameter	effectivity	p value	efficiency	p value
rq	0.319	0	0.127	0
tth	0.188	0	0.555	0
mem	0.381	0	0.064	0.059

thus secondary. We fix the number of requests at a maximum of 10. Since investors only draw on additional advice requests if they have not been able to receive information about a potential future employee, this number appears to be rarely exhausted. The memory length is likewise of secondary nature, but bears mildly stronger relevance than the number of requests. Assuming that agents can maintain a memory about around on fifth to one quarter of conduct of fellow investors, we fix this value at 40, which implies a coverage of 22.5 percent. The central parameter for refined exploration is thus the initial probability of truthfulness, an aspect we thus use as the essential independent input variable for the remainder of the experimental evaluation.

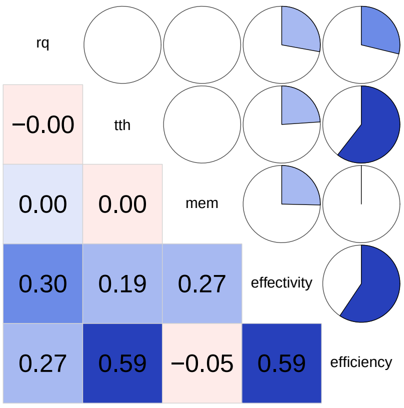


(a) Correlogram for FXD Networks

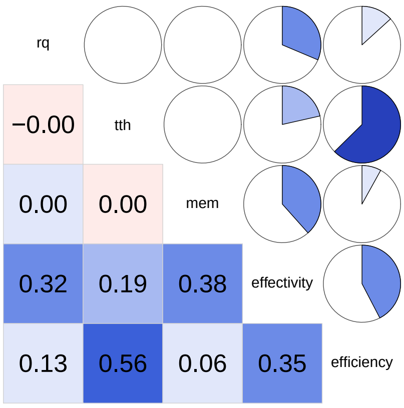


(b) Correlogram for FXDM Networks

Figure C.3: Correlograms for FXD and FXDM Networks



(a) Correlogram for WS Networks



(b) Correlogram for BA Networks

Figure C.4: Correlograms for WS and BA Networks

C.2 Rewiring Probabilities

Table C.7: Effectivity Statistics for different Rewiring Probabilities for Watts-Strogatz Networks across different Levels of Initial Truthfulness (Genoese Scenario)

Initial	**Rewiring Probabilities**						**Statistical Measures**			
Truthfulness	0.05	0.1	0.2	0.3	0.4	0.5	mean	σ	min	max
0.5	0.332	0.352	0.368	0.372	0.371	0.372	0.361	0.016	0.332	0.372
0.51	0.356	0.371	0.383	0.394	0.389	0.391	0.381	0.015	0.356	0.394
0.55	0.398	0.414	0.422	0.428	0.428	0.427	0.419	0.012	0.398	0.428
0.6	0.435	0.445	0.457	0.46	0.458	0.459	0.452	0.01	0.435	0.46
0.7	0.498	0.507	0.509	0.512	0.512	0.512	0.508	0.006	0.498	0.512
0.8	0.556	0.556	0.562	0.562	0.566	0.561	0.56	0.004	0.556	0.566
0.9	0.59	0.592	0.593	0.595	0.595	0.594	0.593	0.002	0.59	0.595

Table C.8: Efficiency Statistics for different Rewiring Probabilities for Watts-Strogatz Networks across different Levels of Initial Truthfulness (Genoese Scenario)

Initial	**Rewiring Probabilities**						**Statistical Measures**			
Truthfulness	0.05	0.1	0.2	0.3	0.4	0.5	mean	σ	min	max
0.5	0.013	0.015	0.017	0.017	0.017	0.016	0.016	0.002	0.013	0.017
0.51	0.02	0.023	0.023	0.025	0.022	0.025	0.023	0.002	0.02	0.025
0.55	0.147	0.16	0.172	0.169	0.159	0.158	0.161	0.009	0.147	0.172
0.6	0.566	0.577	0.569	0.565	0.569	0.564	0.568	0.005	0.564	0.577
0.7	0.876	0.876	0.87	0.876	0.872	0.876	0.874	0.003	0.87	0.876
0.8	0.951	0.952	0.952	0.952	0.95	0.95	0.951	0.001	0.95	0.952
0.9	0.976	0.976	0.976	0.976	0.976	0.976	0.976	0	0.976	0.976

Table C.9: Effectivity Statistics for different Rewiring Probabilities for Watts-Strogatz Networks across different Levels of Initial Truthfulness (Maghribī Scenario)

Initial	**Rewiring Probabilities**						**Statistical Measures**			
Truthfulness	0.05	0.1	0.2	0.3	0.4	0.5	mean	σ	min	max
0.5	0.574	0.599	0.607	0.647	0.584	0.628	0.607	0.027	0.574	0.647
0.51	0.619	0.662	0.71	0.706	0.734	0.743	0.696	0.047	0.619	0.743
0.55	0.623	0.669	0.726	0.755	0.773	0.779	0.721	0.063	0.623	0.779
0.6	0.622	0.669	0.725	0.754	0.772	0.779	0.72	0.063	0.622	0.779
0.7	0.618	0.666	0.723	0.753	0.769	0.778	0.718	0.063	0.618	0.778
0.8	0.62	0.664	0.721	0.75	0.768	0.778	0.717	0.063	0.62	0.778
0.9	0.613	0.661	0.719	0.749	0.766	0.773	0.714	0.064	0.613	0.773

Table C.10: Efficiency Statistics for different Rewiring Probabilities for Watts-Strogatz Networks across different Levels of Initial Truthfulness (Maghribī Scenario)

Initial	**Rewiring Probabilities**						**Statistical Measures**			
Truthfulness	0.05	0.1	0.2	0.3	0.4	0.5	mean	σ	min	max
0.5	0.35	0.469	0.61	0.626	0.73	0.699	0.58	0.145	0.35	0.73
0.51	0.401	0.45	0.545	0.641	0.643	0.639	0.553	0.107	0.401	0.643
0.55	0.589	0.631	0.68	0.705	0.719	0.725	0.675	0.054	0.589	0.725
0.6	0.686	0.725	0.76	0.777	0.788	0.791	0.755	0.042	0.686	0.791
0.7	0.829	0.848	0.867	0.874	0.88	0.884	0.864	0.021	0.829	0.884
0.8	0.914	0.923	0.933	0.936	0.939	0.938	0.931	0.01	0.914	0.939
0.9	0.953	0.957	0.963	0.964	0.965	0.966	0.961	0.005	0.953	0.966

C.3 Results for Genoese Scenario

Table C.11: Simulation Results for Informal Communication among Genoese Traders (1/2)

Initial Truthfulness	Network Type	Effectivity			Efficiency		
		mean	σ	p value	mean	σ	p value
0.5	NoComm	0.195	0.019		1	0	
0.5	FXD	0.25	0.035	9.28e-09	0.02	0.019	2.904e-12
0.5	FXDM	0.241	0.034	1.285e-07	0.013	0.007	2.906e-12
0.5	WS (0.05)	0.332	0.035	2.661e-16	0.013	0.005	2.904e-12
0.5	WS (0.1)	0.352	0.038	6.652e-17	0.015	0.003	2.901e-12
0.5	WS (0.2)	0.368	0.038	6.652e-17	0.017	0.006	2.901e-12
0.5	WS (0.3)	0.372	0.039	6.652e-17	0.017	0.008	2.901e-12
0.5	WS (0.4)	0.371	0.035	6.652e-17	0.017	0.006	2.906e-12
0.5	WS (0.5)	0.372	0.035	6.652e-17	0.016	0.006	2.904e-12
0.5	BA	0.334	0.069	2.526e-08	0.161	0.341	3.967e-11
0.51	FXD	0.276	0.038	3.833e-13	0.068	0.063	2.909e-12
0.51	FXDM	0.26	0.037	5.067e-11	0.027	0.018	2.904e-12
0.51	WS (0.05)	0.356	0.033	6.652e-17	0.02	0.007	2.904e-12
0.51	WS (0.1)	0.371	0.04	6.652e-17	0.023	0.009	2.906e-12
0.51	WS (0.2)	0.383	0.04	6.652e-17	0.023	0.012	2.904e-12
0.51	WS (0.3)	0.394	0.03	6.652e-17	0.025	0.01	2.904e-12
0.51	WS (0.4)	0.389	0.03	6.652e-17	0.022	0.011	2.906e-12
0.51	WS (0.5)	0.391	0.026	6.652e-17	0.025	0.011	2.909e-12
0.51	BA	0.36	0.043	4.881e-13	0.068	0.18	1.097e-11
0.55	FXD	0.364	0.033	6.652e-17	0.531	0.105	2.909e-12
0.55	FXDM	0.334	0.033	6.652e-17	0.411	0.143	2.906e-12
0.55	WS (0.05)	0.398	0.025	6.652e-17	0.147	0.086	2.909e-12
0.55	WS (0.1)	0.414	0.028	6.652e-17	0.16	0.091	2.909e-12
0.55	WS (0.2)	0.422	0.028	6.652e-17	0.172	0.09	2.909e-12
0.55	WS (0.3)	0.428	0.023	6.652e-17	0.169	0.087	2.909e-12
0.55	WS (0.4)	0.428	0.03	6.652e-17	0.159	0.072	2.909e-12
0.55	WS (0.5)	0.427	0.027	6.652e-17	0.158	0.088	2.909e-12
0.55	BA	0.395	0.028	6.652e-17	0.112	0.058	2.909e-12
0.6	FXD	0.457	0.028	6.652e-17	0.777	0.029	2.909e-12
0.6	FXDM	0.414	0.03	6.652e-17	0.748	0.03	2.909e-12
0.6	WS (0.05)	0.435	0.03	6.652e-17	0.566	0.093	2.909e-12
0.6	WS (0.1)	0.445	0.029	6.652e-17	0.577	0.094	2.909e-12

Table C.12: Simulation Results for Informal Communication among Genoese Traders (2/2)

Initial Truthfulness	Network Type	Effectivity			Efficiency		
		mean	σ	p value	mean	σ	p value
0.6	WS (0.2)	0.457	0.031	6.652e-17	0.569	0.1	2.906e-12
0.6	WS (0.3)	0.46	0.024	6.652e-17	0.565	0.091	2.906e-12
0.6	WS (0.4)	0.458	0.028	6.652e-17	0.569	0.095	2.909e-12
0.6	WS (0.5)	0.459	0.024	6.652e-17	0.564	0.097	2.909e-12
0.6	BA	0.424	0.029	6.652e-17	0.379	0.119	2.906e-12
0.7	FXD	0.558	0.028	6.652e-17	0.909	0.015	2.909e-12
0.7	FXDM	0.517	0.028	6.652e-17	0.906	0.012	2.909e-12
0.7	WS (0.05)	0.498	0.029	6.652e-17	0.876	0.018	2.906e-12
0.7	WS (0.1)	0.507	0.028	6.652e-17	0.876	0.016	2.904e-12
0.7	WS (0.2)	0.509	0.031	6.652e-17	0.87	0.03	2.909e-12
0.7	WS (0.3)	0.512	0.026	6.652e-17	0.876	0.022	2.904e-12
0.7	WS (0.4)	0.512	0.028	6.652e-17	0.872	0.023	2.909e-12
0.7	WS (0.5)	0.512	0.026	6.652e-17	0.876	0.024	2.906e-12
0.7	BA	0.476	0.025	6.652e-17	0.78	0.044	2.909e-12
0.8	FXD	0.605	0.025	6.652e-17	0.959	0.006	2.898e-12
0.8	FXDM	0.581	0.025	6.652e-17	0.957	0.007	2.904e-12
0.8	WS (0.05)	0.556	0.03	6.652e-17	0.951	0.008	2.896e-12
0.8	WS (0.1)	0.556	0.027	6.652e-17	0.952	0.006	2.909e-12
0.8	WS (0.2)	0.562	0.029	6.652e-17	0.952	0.007	2.909e-12
0.8	WS (0.3)	0.562	0.028	6.652e-17	0.952	0.009	2.904e-12
0.8	WS (0.4)	0.566	0.025	6.652e-17	0.95	0.007	2.904e-12
0.8	WS (0.5)	0.561	0.03	6.652e-17	0.95	0.007	2.904e-12
0.8	BA	0.52	0.031	6.652e-17	0.909	0.026	2.909e-12
0.9	FXD	0.614	0.025	6.652e-17	0.977	0.003	2.876e-12
0.9	FXDM	0.606	0.026	6.652e-17	0.977	0.003	2.901e-12
0.9	WS (0.05)	0.59	0.029	6.652e-17	0.976	0.003	2.896e-12
0.9	WS (0.1)	0.592	0.028	6.652e-17	0.976	0.003	2.901e-12
0.9	WS (0.2)	0.593	0.029	6.652e-17	0.976	0.003	2.904e-12
0.9	WS (0.3)	0.595	0.024	6.652e-17	0.976	0.002	2.898e-12
0.9	WS (0.4)	0.595	0.027	6.652e-17	0.976	0.003	2.898e-12
0.9	WS (0.5)	0.594	0.026	6.652e-17	0.976	0.003	2.906e-12
0.9	BA	0.563	0.028	6.652e-17	0.962	0.013	2.909e-12

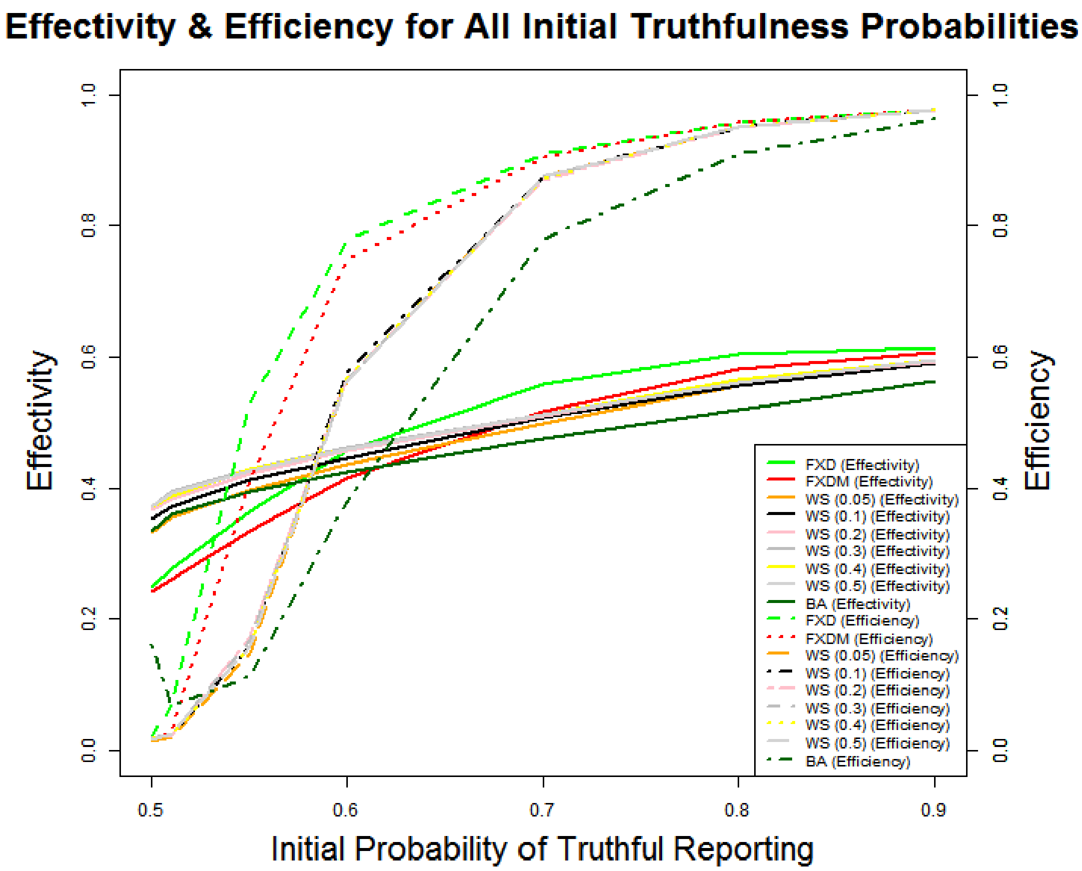


Figure C.5: Effectivity/Efficiency Values across Range of $p_{truthful}$

C.4 Results for Maghribī Scenario

Table C.13: Simulation Results for Informal Communication using Maghribī Communication Patterns (1/2)

Initial Truthfulness	**Network Type**	**Effectivity**			**Efficiency**		
		mean	**σ**	**p value**	**mean**	**σ**	**p value**
0.5	NoComm	0.195	0.019		1	0	
0.5	FXDM	0.635	0.264	3.679e-09	0.699	0.191	1.475e-09
0.5	WS (0.05)	0.574	0.122	1.264e-15	0.35	0.254	1.097e-11
0.5	WS (0.1)	0.599	0.168	5.067e-11	0.469	0.277	3.967e-11
0.5	WS (0.2)	0.607	0.22	1.532e-12	0.61	0.248	4.596e-10
0.5	WS (0.3)	0.647	0.228	3.564e-11	0.626	0.212	1.476e-09
0.5	WS (0.4)	0.584	0.272	1.842e-09	0.73	0.201	3.933e-08
0.5	WS (0.5)	0.628	0.263	1.643e-10	0.699	0.192	1.476e-09
0.5	BA	0.588	0.197	4.233e-10	0.497	0.279	4.596e-10
0.51	FXDM	0.715	0.203	1.062e-13	0.69	0.13	4.596e-10
0.51	WS (0.05)	0.619	0.048	6.652e-17	0.401	0.148	2.909e-12
0.51	WS (0.1)	0.662	0.062	6.652e-17	0.45	0.17	2.909e-12
0.51	WS (0.2)	0.71	0.102	7.982e-16	0.545	0.159	1.097e-11
0.51	WS (0.3)	0.706	0.168	9.806e-13	0.641	0.129	1.097e-11
0.51	WS (0.4)	0.734	0.144	4.457e-15	0.643	0.105	1.096e-11
0.51	WS (0.5)	0.743	0.149	1.388e-13	0.639	0.116	1.097e-11
0.51	BA	0.654	0.128	3.379e-14	0.44	0.183	2.909e-12
0.55	FXDM	0.79	0.031	6.652e-17	0.736	0.018	2.906e-12
0.55	WS (0.05)	0.623	0.043	6.652e-17	0.589	0.036	2.909e-12
0.55	WS (0.1)	0.669	0.038	6.652e-17	0.631	0.036	2.909e-12
0.55	WS (0.2)	0.726	0.035	6.652e-17	0.68	0.03	2.909e-12
0.55	WS (0.3)	0.755	0.032	6.652e-17	0.705	0.021	2.909e-12
0.55	WS (0.4)	0.773	0.032	6.652e-17	0.719	0.019	2.904e-12
0.55	WS (0.5)	0.779	0.032	6.652e-17	0.725	0.02	2.906e-12
0.55	BA	0.694	0.03	6.652e-17	0.416	0.134	2.909e-12
0.6	FXDM	0.788	0.031	6.652e-17	0.802	0.014	2.906e-12
0.6	WS (0.05)	0.622	0.04	6.652e-17	0.686	0.035	2.904e-12
0.6	WS (0.1)	0.669	0.04	6.652e-17	0.725	0.027	2.906e-12

Table C.14: Simulation Results for Informal Communication using Maghribī Communication Patterns (2/2)

Initial Truthfulness	Network Type	Effectivity			Efficiency		
		mean	σ	p value	mean	σ	p value
0.6	WS (0.2)	0.725	0.037	6.652e-17	0.76	0.022	2.909e-12
0.6	WS (0.3)	0.754	0.033	6.652e-17	0.777	0.019	2.909e-12
0.6	WS (0.4)	0.772	0.033	6.652e-17	0.788	0.019	2.909e-12
0.6	WS (0.5)	0.779	0.031	6.652e-17	0.791	0.017	2.906e-12
0.6	BA	0.693	0.029	6.652e-17	0.439	0.109	2.909e-12
0.7	FXDM	0.787	0.032	6.652e-17	0.892	0.013	2.906e-12
0.7	WS (0.05)	0.618	0.044	6.652e-17	0.829	0.027	2.909e-12
0.7	WS (0.1)	0.666	0.039	6.652e-17	0.848	0.021	2.909e-12
0.7	WS (0.2)	0.723	0.038	6.652e-17	0.867	0.018	2.909e-12
0.7	WS (0.3)	0.753	0.033	6.652e-17	0.874	0.016	2.906e-12
0.7	WS (0.4)	0.769	0.033	6.652e-17	0.88	0.015	2.909e-12
0.7	WS (0.5)	0.778	0.033	6.652e-17	0.884	0.014	2.906e-12
0.7	BA	0.69	0.03	6.652e-17	0.747	0.092	2.909e-12
0.8	FXDM	0.785	0.032	6.652e-17	0.946	0.007	2.904e-12
0.8	WS (0.05)	0.62	0.042	6.652e-17	0.914	0.014	2.906e-12
0.8	WS (0.1)	0.664	0.039	6.652e-17	0.923	0.013	2.906e-12
0.8	WS (0.2)	0.721	0.038	6.652e-17	0.933	0.008	2.904e-12
0.8	WS (0.3)	0.75	0.033	6.652e-17	0.936	0.009	2.904e-12
0.8	WS (0.4)	0.768	0.033	6.652e-17	0.939	0.007	2.906e-12
0.8	WS (0.5)	0.778	0.033	6.652e-17	0.938	0.008	2.904e-12
0.8	BA	0.693	0.031	6.652e-17	0.89	0.038	2.906e-12
0.9	FXDM	0.784	0.032	6.652e-17	0.971	0.003	2.898e-12
0.9	WS (0.05)	0.613	0.044	6.652e-17	0.953	0.008	2.901e-12
0.9	WS (0.1)	0.661	0.04	6.652e-17	0.957	0.006	2.906e-12
0.9	WS (0.2)	0.719	0.038	6.652e-17	0.963	0.005	2.904e-12
0.9	WS (0.3)	0.749	0.033	6.652e-17	0.964	0.005	2.906e-12
0.9	WS (0.4)	0.766	0.034	6.652e-17	0.965	0.004	2.904e-12
0.9	WS (0.5)	0.773	0.033	6.652e-17	0.966	0.004	2.904e-12
0.9	BA	0.713	0.032	6.652e-17	0.951	0.016	2.906e-12

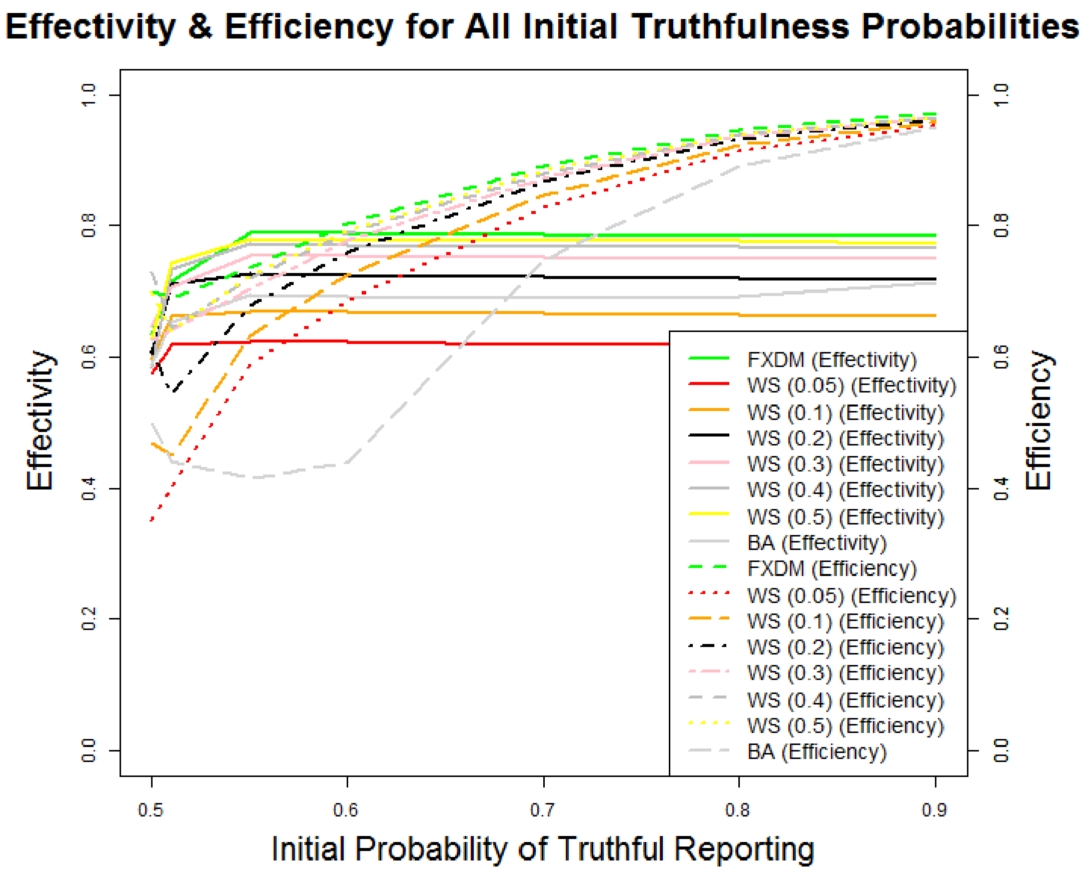


Figure C.6: Effectivity/Efficiency Values across Range of $p_{truthful}$

C.5 Investor Network Distributions according to van Doosselaere

Based on the statistical aggregates of the commenda node degree distributions by van Doosselaere (2009), we attempted an approximation of candidate power law function parameters. However, this approach is challenged by the coarse binning of network degrees. Table C.15 shows the original categories and the derived category mean we use for the approximation process. The corresponding distribution is shown in Figure C.7. Plotting the cumulative distribution function (CDF) on a log-log scale, we can observe the characteristic near linear signature. Based on Clauset et al. (2009)'s method for truncating power law function (to accommodate the approximation of finite distributions), we determine candidate model parameters for an approximated power law distribution ($p(x) \propto x^{-\alpha}$, $x > x_{min}$), with x_{min} describing the starting point of the distribution, and α indicating its cut-off. Comparing the synthetic estimates and the real distribution visually in Figure C.8, the unsupervised result (in black colour) offers an optimistic scaling parameter α (which controls the tail of the distribution) of around 1.7 and a minimal value (x_{min}) of 4. The supervised approximation across the value range of the original distributions shows stronger alignment with the original distribution based on an α value of 1.5 (in red colour). However, compared to the power law distributions observed by Clauset et al. (2009) an α value of 1.5 reflects an exceptionally low value; the lowest α values for their collection of empirical cases lie at around 1.7, with the majority of values lying between 2 and 3 (Clauset et al., 2009).

Table C.15: Commenda Network Distribution (1154-1164 AD) according to van Doosselaere

Node Degree Category (van Doosselaere (2009))	**Mean value**	**Count**
1	1	241
2	2	77
3	3	21
4	4	13
5	5	8
6-7	6.5	5
8-9	8.5	7
10-13	11.5	4
14-19	16.5	5
20+	20	2

To support our visual analysis, we perform a goodness-of-fit test. We do so by bootstrap-

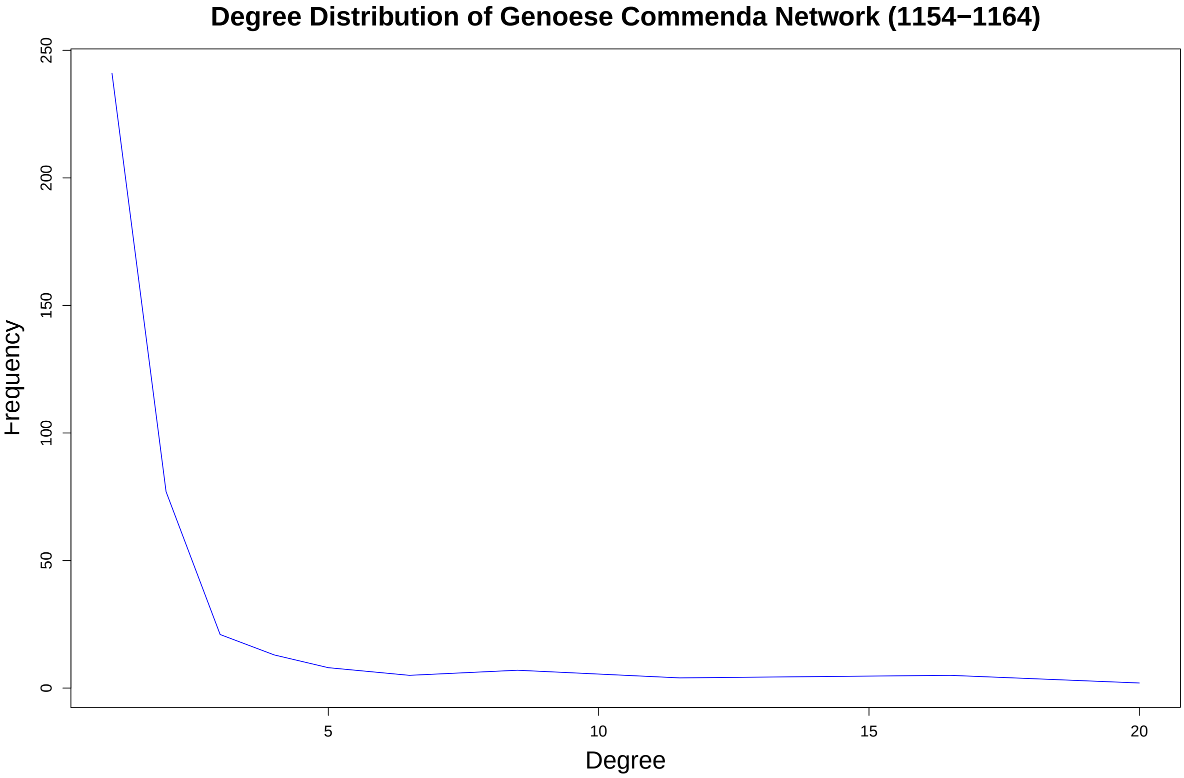


Figure C.7: Distribution of Commenda Relationships (1154-1164 AD) according to van Doosselaere

ping power law distributions from the inferred parameters and calculating the Kolmogorov-Smirnov statistic. Clauset et al. (2009) use the fraction of times the value for the generated distribution is larger than the originally inferred power law model, which is expressed as a p value, with p values $>$ a chosen significance level.[4] Performing 10000 runs,[5] we arrive at a value of 0.8642 which supports, i.e. does not rule out, the existence of a power law distribution.

Refining the analysis by excluding alternative distributions, we compare the synthetic power law distribution with exponential and log-normal distributions based on parameters inferred from the originally generated distribution. We do so using Vuong (1989)'s closeness test that indicates the likelihood of one of two compared models to be closer to the true distribution. It does so based on two measures, one being the ratio of log-likelihoods between two distributions' data ($\mathscr{R}$), whose sign indicates the closeness of the original data

[4]Clauset et al. (2009) advocate a level of 0.1 to avoid overly rigid dismissal of candidate distributions from further exploration.

[5]Clauset et al. (2009) suggest the determination of bootstrapping runs based on the rule $\frac{1}{4}\varepsilon^{-2}$, with ε indicating the desired level of accuracy, thus $\varepsilon = 0.005$ for our case.

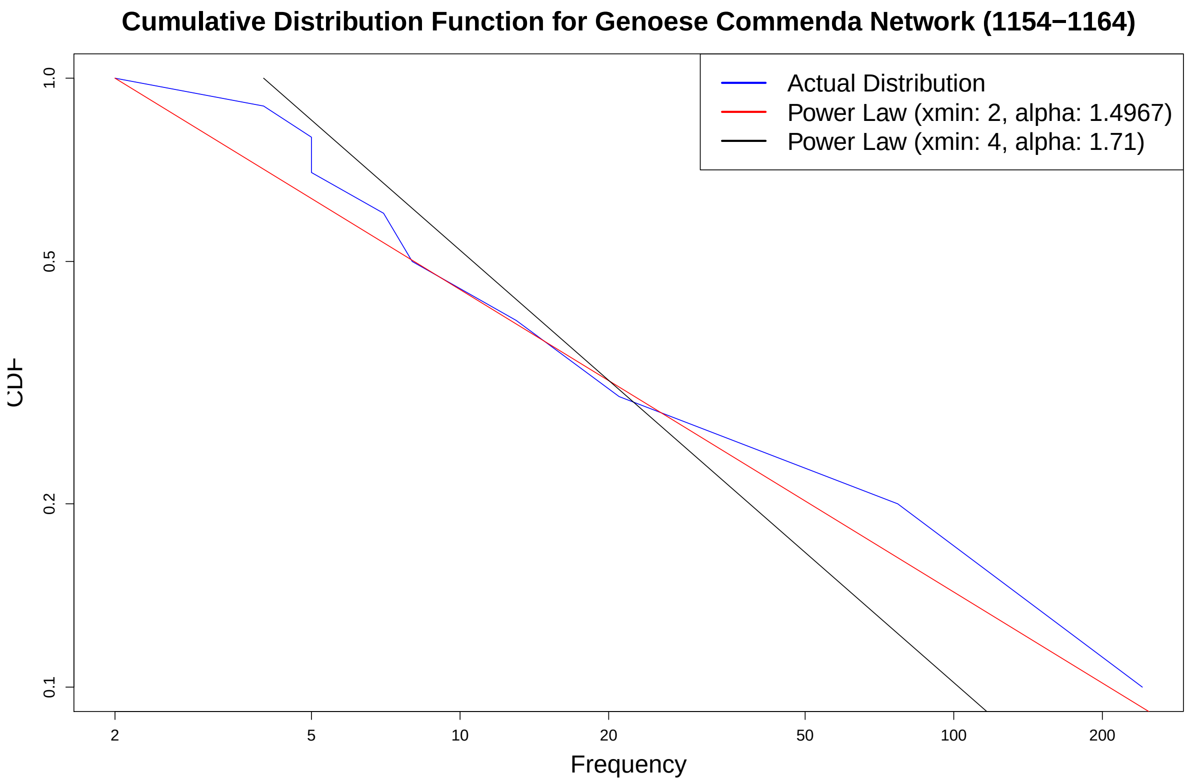


Figure C.8: Cumulative Frequency Distribution for Commenda Relationships (1154-1164 AD) according to van Doosselaere

to a given model. Here a *low* associated p-value[6] indicates statistical significance of $\mathscr{R}$'s sign, with a positive result supporting the first model, and negative sign favouring the second model. Table C.16 shows the results for 30 comparisons against either alternative distribution. The results favour the power law distribution over the exponential distribution (positive $\mathscr{R}$, p < 0.1). However, the comparison to the log-normal distribution, though in favour of the power law distribution (positive $\mathscr{R}$), is not significant (p > 0.1). Clauset et al. (2009) suggest to gain stronger support by sampling data subsets from the generated distributions with varying levels of x_{min} (to extract partitions of the distribution's long tail). However, in the light of the coarse categorisation of higher degree levels of the original distribution (see Table C.15), we see limited value in a refined exploration of the distribution's long tail. The commenda relationships thus show power law characteristics (albeit highlighting a low scaling parameter (α) of around 1.5) without being fully conclusive about the nature of the distribution.

[6] Clauset et al. (2009) suggest a value < 0.1.

Table C.16: 'Goodness of Fit' Results

Compared distribution type	p value		$\mathscr{R}$	
	mean	σ	**mean**	σ
log-normal	0.6094	0.2659	0.2745	0.6963
exponential	0.1006	0.1011	25.5980	88.2569

D

Apprenticeship Model

D.1 Upper Jāh vs. Lower Jāh

Table D.1: Relative Cheater Reduction of Apprenticeship System (compared to Apprenticeship-Free Variant) for Parameters Upper Jāh vs. Lower Jāh Difference

Upper Jāh \ Lower Jāh	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.1	0.136	0.155	0.15	0.14	0.157	0.16	0.179	0.172	0.216	0.178
0.2	0.141	0.098	0.127	0.13	0.143	0.135	0.141	0.163	0.117	0.017
0.3	0.153	0.114	0.129	0.128	0.128	0.148	0.117	0.124	0.128	0.015
0.4	0.151	0.142	0.132	0.113	0.136	0.145	0.136	0.133	0.116	−0.014
0.5	0.144	0.114	0.143	0.147	0.169	0.152	0.144	0.122	0.137	−0.027
0.6	0.144	0.126	0.122	0.153	0.16	0.166	0.138	0.149	0.144	−0.054
0.7	0.149	0.134	0.135	0.172	0.158	0.159	0.185	0.161	0.157	−0.044
0.8	0.142	0.135	0.141	0.149	0.183	0.162	0.154	0.156	0.132	−0.057
0.9	0.141	0.077	0.136	0.15	0.15	0.162	0.141	0.161	0.11	−0.055
1	0.153	0.136	0.134	0.147	0.136	0.153	0.143	0.126	0.116	−0.044

Table D.2: Absolute Cheater Numbers in Apprenticeship System for Parameters Upper Jāh vs. Lower Jāh Difference

		Lower Jāh									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Upper Jāh	0.1	127.321	124.821	125.866	125.924	126.02	124.16	122.979	122.432	118.77	123.067
	0.2	126.682	121.657	120.358	119.052	118.989	118.802	118.193	116.022	118.87	119.001
	0.3	125.362	116.899	112.456	112.485	111.045	109.907	111.775	110.501	108.28	108.76
	0.4	125.891	115.074	108.19	108.337	105.733	104.941	105.986	105.973	104.217	102.005
	0.5	125.207	117.23	106.08	103.48	101.952	100.465	101.161	104.041	99.956	100.672
	0.6	128.505	116.003	106.882	101.609	101.257	100.44	101.376	99.743	99.036	100.262
	0.7	125.728	116.808	106.744	100.821	100.936	100.288	96.603	98.46	97.885	98.373
	0.8	127.09	117.751	106.09	102.7	98.071	98.435	99.062	95.628	98.142	98.329
	0.9	126.034	117.312	106.358	102.081	99.515	97.246	98.467	97.189	97.887	97.075
	1	124.628	117.08	106.849	102.731	99.094	97.491	96.606	97.665	97.179	95.933

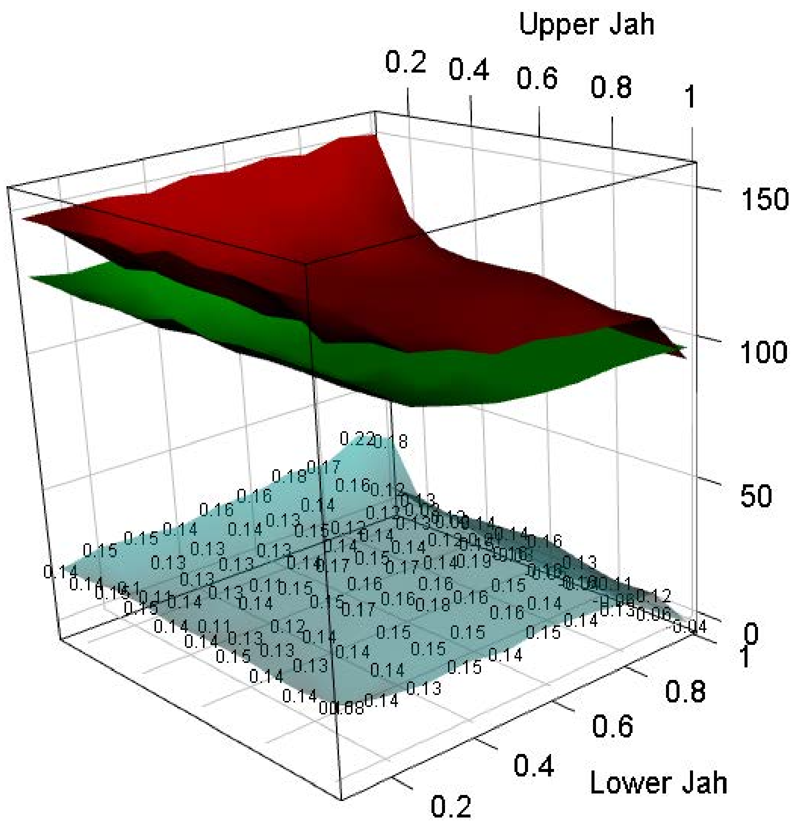


Figure D.1: Number of Cheaters with Relationship to Non-Cheater(s) for Varied Upper and Lower Jāh Differences

D.2 Cheating Probability vs. Lower Jāh

Table D.3: Relative Cheater Reduction of Apprenticeship System (compared to Apprenticeship-Free Variant) for Parameters Cheating Probability vs. Lower Jāh Difference

Cheating Probability \ Lower Jāh	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.1	0.157	0.133	0.107	0.088	0.13	0.12	0.11	0.119	0.105	−0.055
0.2	0.137	0.128	0.119	0.119	0.152	0.143	0.124	0.142	0.117	−0.043
0.3	0.152	0.126	0.129	0.156	0.137	0.143	0.133	0.111	0.114	−0.06
0.4	0.149	0.121	0.125	0.142	0.155	0.142	0.151	0.131	0.121	−0.048
0.5	0.153	0.136	0.144	0.148	0.145	0.159	0.144	0.114	0.115	−0.059
0.6	0.154	0.125	0.147	0.166	0.155	0.143	0.145	0.15	0.107	−0.052
0.7	0.139	0.127	0.141	0.162	0.154	0.155	0.137	0.144	0.108	−0.073
0.8	0.166	0.14	0.145	0.178	0.153	0.164	0.146	0.13	0.121	−0.082
0.9	0.133	0.129	0.126	0.172	0.162	0.144	0.15	0.134	0.115	−0.057
1	0.157	0.138	0.157	0.186	0.17	0.162	0.16	0.145	0.129	−0.048

Table D.4: Absolute Cheater Numbers in Apprenticeship System for Parameters Cheating Probability vs. Lower Jāh Difference

Cheating Probability \ Lower Jāh	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.1	126.454	124.139	116.507	113.478	106.689	108.488	108.605	106.029	108.356	107.993
0.2	127.303	120.573	112.23	107.928	102.572	103.33	104.243	102.449	101.418	100.549
0.3	125.948	119.88	109.675	102.659	101.196	101.36	100.556	102.186	99.383	98.938
0.4	125.902	118.187	108.919	103.142	99.253	99.802	99.573	99.284	98.554	97.019
0.5	124.628	117.08	105.935	102.31	98.132	96.722	96.966	98.529	97.225	97.203
0.6	124.948	115.723	105.403	100.022	96.058	97.423	96.217	94.552	97.118	95.309
0.7	127.533	116.74	104.437	99.402	95.566	95.416	96.791	94.66	96.01	95.664
0.8	123.904	114.743	103.077	98.428	95.184	93.596	94.444	95.356	94.509	94.783
0.9	125.911	113.205	103.701	97.654	93.966	95.382	94.007	94.247	93.63	93.483
1	124.381	111.363	101.784	96.261	92.754	92.547	92.816	92.459	92.381	92.256

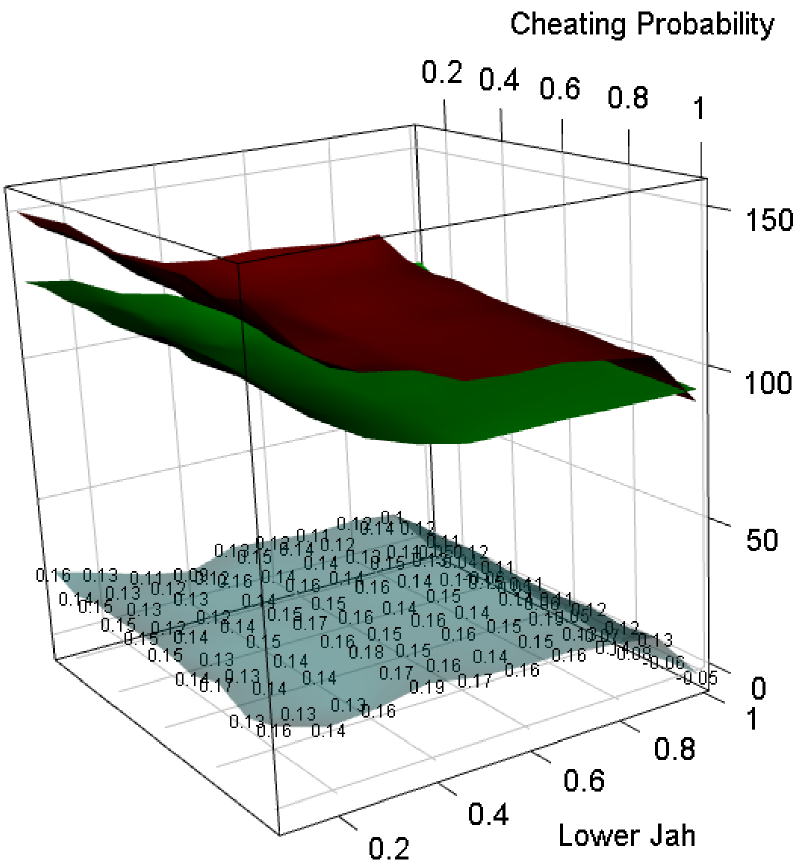


Figure D.2: Number of Cheaters with Relationship to Non-Cheater(s) for Varied Cheating Probability and Lower Jāh Differences

D.3 Observation Quota vs. Lower Jāh

Table D.5: Relative Cheater Reduction of Apprenticeship System (compared to Apprenticeship-Free Variant) for Parameters Observation Quota vs. Lower Jāh Difference

Observation Quota \ Lower Jāh	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.025	0.153	0.136	0.144	0.148	0.145	0.159	0.144	0.126	0.116	−0.044
0.05	0.19	0.196	0.146	0.138	0.098	0.091	0.091	0.057	0.086	−0.096
0.075	0.204	0.144	0.124	0.075	0.045	0.057	0.046	0.048	0.023	−0.11
0.1	0.183	0.123	0.091	0.066	0.037	0.023	0.022	0.014	−0.008	−0.17
0.125	0.17	0.106	0.085	0.033	0.018	0.013	0.013	−0.004	−0.02	−0.195
0.15	0.136	0.112	0.066	0.035	0.018	0.002	0.007	−0.01	−0.039	−0.175
0.175	0.131	0.086	0.065	0.022	−0.003	0.015	−0.009	−0.02	−0.039	−0.204
0.2	0.109	0.078	0.06	0.028	−0.001	−0.014	−0.022	−0.01	−0.041	−0.218
0.225	0.136	0.056	0.045	0.036	−0.025	−0.008	−0.013	−0.052	−0.071	−0.205
0.25	0.096	0.074	0.023	0.01	−0.018	−0.017	−0.041	−0.068	−0.069	−0.216
0.275	0.096	0.055	0.026	0.008	−0.026	−0.046	−0.025	−0.05	−0.06	−0.257
0.3	0.107	0.062	0.024	−0.011	−0.032	−0.029	−0.046	−0.05	−0.103	−0.237

Table D.6: Absolute Cheater Numbers in Apprenticeship System for Parameters Observation Quota vs. Lower Jāh Difference

Observation Quota \ Lower Jāh	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.025	124.628	117.08	105.935	102.31	98.132	96.722	96.966	97.665	97.179	95.933
0.05	97.689	89.359	87.925	84.623	83.757	83.39	83.659	85.49	82.059	84.906
0.075	87.895	83.225	79.887	79.857	78.785	77.275	77.538	76.878	78.847	78.588
0.1	81.953	77.681	76.091	73.889	73.387	73.624	73.509	74.083	73.528	75.823
0.125	77.349	73.962	71.116	71.56	69.656	70.249	70.485	70.399	70.8	73.489
0.15	76.11	70.307	68.49	68.115	67.122	67.405	66.872	67.67	69.183	70.796
0.175	72.92	68.676	66.305	66.208	65.794	65.021	65.82	65.83	66.548	69.411
0.2	71.38	66.677	64.653	63.437	63.519	63.883	63.84	63.416	63.782	68.417
0.225	68.146	65.818	62.574	61.155	62.688	61.347	61.524	62.446	63.24	66.931
0.25	67.3	62.006	61.08	59.961	60.328	59.856	60.387	61.479	61.784	65.164
0.275	67.703	62.818	60.449	59.594	59.508	59.606	59.13	59.301	60.927	65.652
0.3	65.575	60.804	58.865	58.319	57.697	57.39	58.016	57.671	59.894	63.872

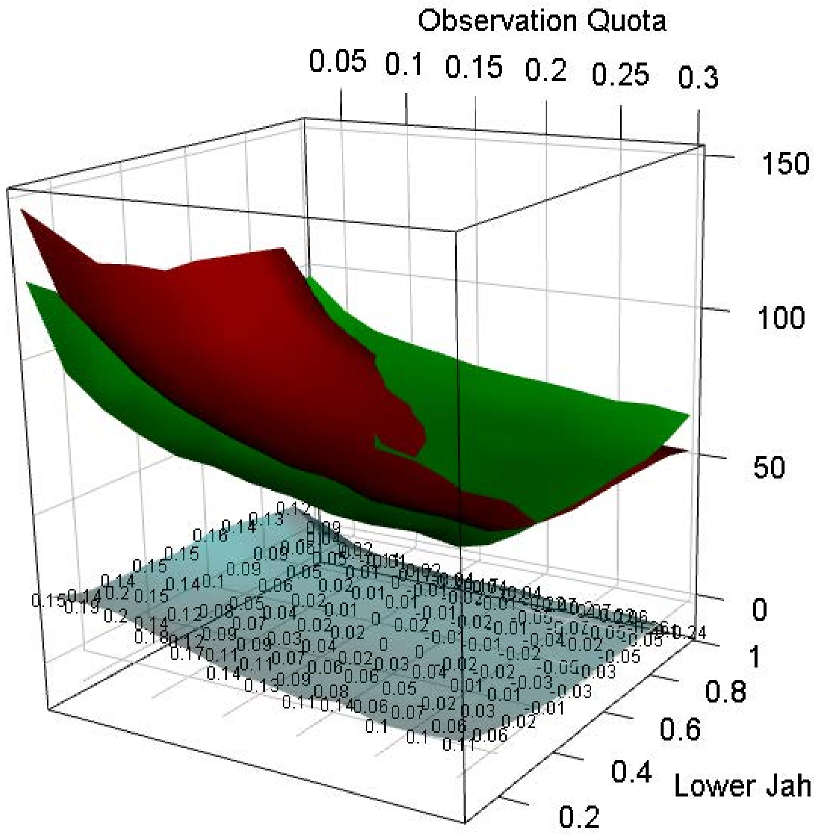


Figure D.3: Number of Cheaters with Relationship to Non-Cheater(s) for Varied Observation Quota and Lower Jāh Difference

D.4 Reduced Interaction Frequency

Table D.7: Relative Cheater Reduction of Apprenticeship System (compared to Apprenticeship-Free Variant) for Parameters Observation Quota vs. Lower Jāh Difference with Reduced Interaction Frequency

Observation Quota \ Lower Jāh	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.025	0.113	0.096	0.124	0.177	0.155	0.148	0.147	0.128	0.115	−0.065
0.05	0.17	0.189	0.171	0.152	0.111	0.118	0.109	0.094	0.081	−0.062
0.075	0.19	0.172	0.159	0.116	0.08	0.096	0.085	0.077	0.076	−0.075
0.1	0.195	0.161	0.123	0.094	0.083	0.076	0.074	0.061	0.049	−0.057
0.125	0.194	0.147	0.119	0.087	0.064	0.077	0.072	0.055	0.033	−0.083
0.15	0.176	0.142	0.102	0.085	0.068	0.066	0.069	0.06	0.033	−0.07
0.175	0.168	0.119	0.108	0.089	0.072	0.055	0.058	0.047	0.018	−0.086
0.2	0.163	0.127	0.103	0.092	0.068	0.068	0.061	0.043	0.036	−0.097
0.225	0.169	0.137	0.096	0.087	0.058	0.056	0.053	0.055	0.018	−0.081
0.25	0.175	0.127	0.108	0.094	0.064	0.061	0.044	0.042	0.023	−0.086
0.275	0.167	0.132	0.117	0.079	0.063	0.072	0.056	0.045	0.039	−0.08
0.3	0.172	0.135	0.104	0.095	0.079	0.054	0.066	0.023	0.019	−0.111

Table D.8: Absolute Cheater Numbers in Apprenticeship System for Parameters Observation Quota vs. Lower Jāh Difference with Reduced Interaction Frequency

Observation Quota \ Lower Jāh	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.025	131.38	118.924	108.128	100.762	99.337	99.121	99.161	99.378	97.636	98.954
0.05	103.254	93.566	89.641	87.977	87.502	87.2	86.834	87.526	86.89	87.755
0.075	90.849	85.982	82.34	81.977	81.917	80.847	81.359	80.816	81.425	83.094
0.1	83.478	79.611	78.783	77.948	76.96	77.178	76.952	77.188	77.636	78.495
0.125	78.341	75.916	74.568	74.56	74.71	73.651	73.435	73.584	75.168	75.943
0.15	74.577	72.12	71.75	70.587	71.138	71.551	70.678	70.371	71.908	73.157
0.175	72.886	71.169	68.602	68.936	69.264	69.517	68.762	69.432	70.209	72.134
0.2	71.486	68.301	67.446	66.252	67.1	66.618	66.958	66.905	67.799	70.882
0.225	68.848	65.653	65.193	64.455	65.653	65.423	65.103	64.699	66.544	68.599
0.25	65.747	63.57	62.823	62.496	63.342	63.079	63.352	62.948	64.055	66.951
0.275	65.316	62.4	61.189	61.497	62.151	61.578	62.262	62.345	62.418	66.108
0.3	63.965	60.524	60.183	59.31	60.144	60.705	59.92	61.903	62.245	66.121

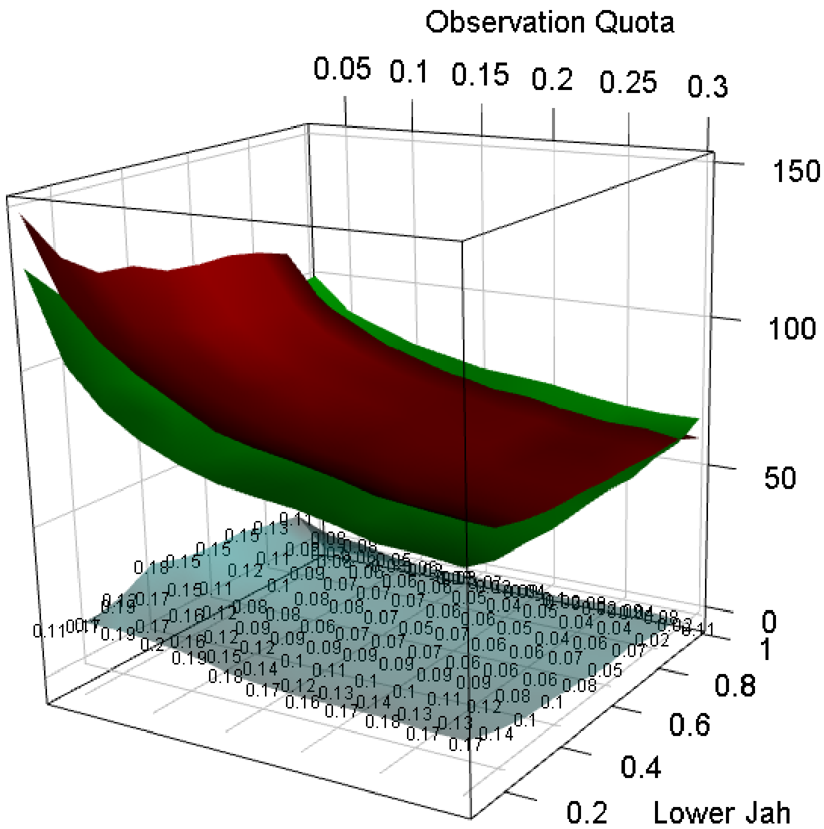


Figure D.4: Number of Cheaters with Relationship to Non-Cheater(s) for Varied Observation Quota and Lower Jāh Difference with Reduced Interaction Frequency

D.5 Reduced Number of Traders

Table D.9: Relative Cheater Reduction of Apprenticeship System (compared to Apprenticeship-Free Variant) for Parameters Observation Quota vs. Lower Jāh Difference (Reduced Interaction Frequency, 200 Agents)

	Lower Jāh									
Observation Quota	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.025	0.102	0.085	0.1	0.087	0.077	0.056	0.043	0.03	0.036	−0.164
0.05	0.152	0.162	0.174	0.159	0.137	0.13	0.138	0.118	0.114	−0.041
0.075	0.211	0.191	0.142	0.126	0.127	0.125	0.083	0.092	0.074	−0.055
0.1	0.213	0.158	0.144	0.137	0.1	0.087	0.116	0.079	0.046	−0.085
0.125	0.217	0.159	0.107	0.124	0.114	0.079	0.084	0.048	0.067	−0.079
0.15	0.16	0.139	0.12	0.1	0.087	0.098	0.08	0.094	0.045	−0.099
0.175	0.184	0.173	0.126	0.098	0.096	0.079	0.07	0.07	0.032	−0.071
0.2	0.198	0.136	0.117	0.075	0.086	0.119	0.077	0.061	0.032	−0.131
0.225	0.144	0.145	0.105	0.087	0.084	0.078	0.061	0.054	0.027	−0.084
0.25	0.201	0.135	0.127	0.127	0.096	0.088	0.079	0.067	0.02	−0.105
0.275	0.191	0.132	0.127	0.078	0.083	0.093	0.093	0.074	0.023	−0.127
0.3	0.185	0.151	0.121	0.102	0.112	0.086	0.059	0.047	0.022	−0.113

Table D.10: Absolute Cheater Numbers in Apprenticeship System for Parameters Observation Quota vs. Lower Jāh Difference (Reduced Interaction Frequency, 200 Agents)

	Lower Jāh									
Observation Quota	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.025	66.22	67.823	67.587	63.012	56.97	56.574	57.572	57.501	57.081	57.319
0.05	54.586	50.041	48.235	46.402	45.411	46.257	45.749	44.914	45.544	45.965
0.075	46.365	44.17	43.586	42.681	41.377	41.256	42.781	42.114	42.321	42.036
0.1	42.192	41.74	39.901	38.658	38.823	39.512	38.445	39.242	39.92	40.264
0.125	39.594	38.059	38.128	36.723	36.519	37.064	36.975	38.068	37.058	38.497
0.15	39.334	36.323	35.518	35.283	35.419	35.296	35.618	34.734	36.185	37.248
0.175	36.931	34.688	34.12	34.237	33.146	34.427	34.266	34.282	35.318	35.238
0.2	35.558	33.841	32.939	33.668	32.777	31.864	33.212	32.706	33.531	35.317
0.225	35.452	32.376	32.187	31.736	31.291	31.979	31.982	31.984	32.265	33.791
0.25	32.996	31.404	30.676	29.317	30.403	30.17	29.991	31.044	31.437	33.007
0.275	33.16	31.155	30.098	30.459	29.929	29.626	29.996	30.111	31.38	33.333
0.3	32.642	30.714	29.549	28.794	28.519	29.091	29.637	29.861	30.48	32.02

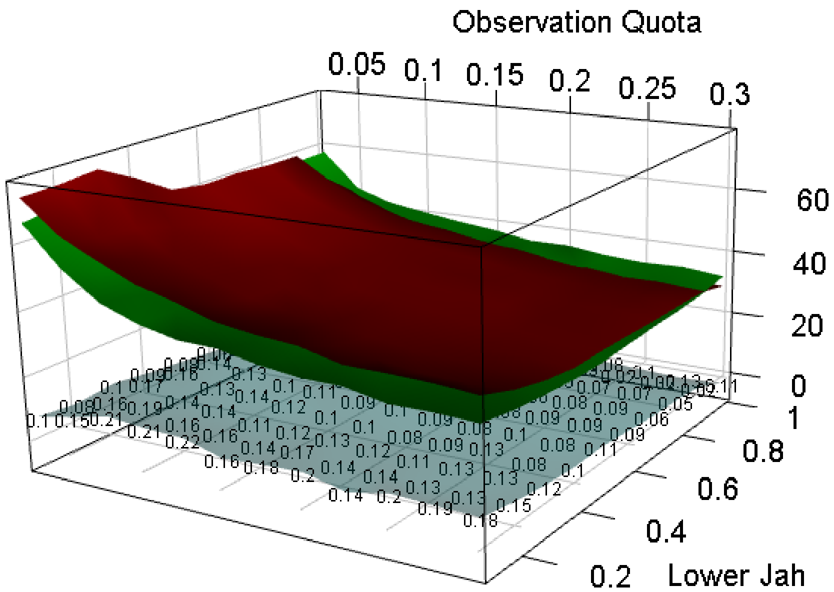


Figure D.5: Number of Cheaters with Relationship to Non-Cheater(s) for Varied Observation Quota and Lower Jāh Difference (Reduced Interaction Frequency, 200 Agents)

E

Role Differentiation Model

The following sections contain the statistical discussion for all four scenarios introduced in Section 6.3:

- Scenario 1 – Role Integration without Norm Enforcement
- Scenario 2 – Role Integration with Norm Enforcement
- Scenario 3 – Role Specialisation without Norm Enforcement
- Scenario 4 – Role Specialisation with Norm Enforcement

All scenarios are further evaluated for the opportunistic and rational deontic determination strategy (see Subsection 6.2.5). The Tables E.1 and E.2 in Section E.1 provide a respective statistical summary (mean and standard deviation) over 30 simulation runs for each configuration. Apart from describing the configuration, the table columns show the fractions of agents that associate the normative understanding for a given action with the given deontic compartments. The column headers are coded as follows: 'Number' indicates the number of agents for the respective scenario; 'Role Diff.' indicates a role-unified (Value 'F') or role-differentiated setup (Value 'T'); 'Enf.' indicates whether third-party enforcement is activated; 'Strat.' specifies the deontic determination strategy, with 'RAT' representing the rational setup, and 'OPP' the opportunistic one. Columns holding the mean value of deontic compartments are marked as follows:

- — implies *must not*
- – implies *should not*
- - implies *may not*
- o implies *indifferent*
- + implies *may*
- ++ implies *should*
- +++ implies *must*

Each column holds the mean value of agents that associate the given action with the respective deontic compartment, and is accompanied by an additional column (σ) that shows the corresponding standard deviation. The distributions effectively reflect the interpretation established in Subsection 6.3.3 and can be further supported by significance tests as discussed in Sections E.2 and E.3.

E.1 Distributions of Normative Understanding across Deontic Compartments

Table E.1: Simulation Results across 30 Simulation Runs for Action ‘trade fair’

Number	Role Diff.	Enf.	Strat.	—	σ	–	σ	-	σ	○	σ	+	σ	++	σ	+++	σ
100	F	F	RAT	0	0	0	0	0	0	0.3	0.435	0.7	0.435	0	0	0	0
100	F	T	RAT	0	0	0	0	0	0	0	0	1	0	0	0	0	0
100	F	F	OPP	0	0	0	0	0.011	0.018	0.087	0.138	0.233	0.34	0.075	0.06	0.592	0.426
100	F	T	OPP	0	0	0	0	0	0	0	0	0	0	0	0	1	0
200	T	F	RAT	0	0	0	0	0.658	0.207	0.278	0.161	0.063	0.076	0	0	0	0
200	T	T	RAT	0	0	0	0	0.55	0.219	0.418	0.2	0.032	0.026	0	0	0	0
200	T	F	OPP	0.637	0.111	0.192	0.065	0	0	0	0	0	0	0.015	0.021	0.156	0.13
200	T	T	OPP	0.443	0.1	0.094	0.039	0.265	0.048	0.013	0.015	0.12	0.052	0.008	0.012	0.057	0.103
200	F	F	RAT	0	0	0	0	0	0.001	0.241	0.409	0.758	0.409	0	0.001	0	0
200	F	T	RAT	0	0	0	0	0	0	0	0	1	0	0	0	0	0
200	F	F	OPP	0	0	0	0	0.01	0.019	0.075	0.134	0.181	0.308	0.08	0.051	0.654	0.401
200	F	T	OPP	0	0	0	0	0	0	0	0	0	0	0	0	1	0
400	T	F	RAT	0	0	0	0	0.579	0.185	0.34	0.146	0.081	0.081	0	0	0	0
400	T	T	RAT	0	0	0	0	0.565	0.164	0.399	0.151	0.036	0.042	0	0	0	0
400	T	F	OPP	0.593	0.102	0.185	0.061	0	0	0	0	0	0	0.024	0.025	0.199	0.128
400	T	T	OPP	0.444	0.062	0.095	0.023	0.278	0.048	0.011	0.011	0.125	0.04	0.003	0.004	0.043	0.083

Table E.2: Simulation Results across 30 Simulation Runs for Action ‘withhold profit’

Number	Role Diff.	Enf.	Strat.	—	σ	–	σ	-	σ	○	σ	+	σ	++	σ	+++	σ
100	F	F	RAT	0	0	0	0.002	0.091	0.077	0.659	0.314	0.25	0.366	0	0	0	0
100	F	T	RAT	0	0	0	0.002	0.382	0.046	0.618	0.047	0	0	0	0	0	0
100	F	F	OPP	0.015	0.035	0.037	0.03	0.218	0.161	0.186	0.168	0.226	0.183	0.066	0.096	0.003	0.004
100	F	T	OPP	0.035	0.019	0.185	0.04	0.281	0.052	0.483	0.043	0.016	0.024	0	0	0	0
200	T	F	RAT	0	0	0	0	0.421	0.156	0.579	0.157	0	0.002	0	0	0	0
200	T	T	RAT	0	0	0	0	0.348	0.087	0.629	0.086	0.023	0.019	0	0	0	0
200	T	F	OPP	0	0	0	0	0.705	0.121	0.275	0.121	0.02	0.014	0	0	0	0
200	T	T	OPP	0.238	0.046	0.12	0.039	0.247	0.1	0.255	0.122	0.094	0.034	0.01	0.011	0	0
200	F	F	RAT	0	0	0	0	0.102	0.078	0.69	0.286	0.208	0.338	0	0	0	0
200	F	T	RAT	0	0	0	0.001	0.383	0.042	0.616	0.042	0	0	0	0	0	0
200	F	F	OPP	0.012	0.023	0.035	0.024	0.241	0.152	0.205	0.169	0.248	0.178	0.056	0.095	0.001	0.002
200	F	T	OPP	0.034	0.012	0.197	0.028	0.286	0.036	0.464	0.037	0.019	0.029	0	0	0	0
400	T	F	RAT	0	0	0	0	0.446	0.169	0.554	0.169	0	0.001	0	0	0	0
400	T	T	RAT	0	0	0	0.001	0.358	0.088	0.615	0.082	0.026	0.024	0	0	0	0
400	T	F	OPP	0	0	0	0	0.73	0.107	0.253	0.107	0.017	0.01	0	0	0	0
400	T	T	OPP	0.242	0.031	0.125	0.03	0.248	0.092	0.243	0.104	0.091	0.025	0.008	0.008	0	0

E.2 Significance Tests

The significance tests for the respective matching Maghribī and Genoese scenarios (i.e. configuration variants of integrated role understanding vs. differentiated role understanding) are shown in Tables E.3 and E.4. Since the results are not normally distributed (see black series in Figures E.1 and E.2 at the end of this section), we rely on non-parametric tests such as the Mann-Whitney-Wilcoxon test (MWW) (Mann and Whitney, 1947) and the Kolmogorov-Smirnov test (KS) (see e.g. Lehmann (2006)) to establish the significance of the results (significance level: 0.05; 30 simulation runs for each configuration; distribution across 7 deontic compartments). Observing the results for action `trade fair` in Table E.3, for all scenarios the difference in distributions across deontic compartments is significant (i.e. p value < 0.05). The significance results for the action `withhold profit` in Table E.4 do not offer such clarity, and distribution variation is insignificant for rational scenarios and the opportunistic scenario without enforcement. Reasons for this are the lower reinforcement levels for the action `withhold profit` as well as the conservative nature of the rational configuration; as established in the Subsection 6.3.3 and observable in the tabular overview of the results in Section E.1, the rational deontic configuration drives the clustering of normative understanding around the deontic centre.[1] This becomes clearer if using the KS test (columns 'KS' and corresponding 'p value' in Tables E.3 and E.4) that puts the emphasis on changes in distribution shape as opposed to median shifts (which is characteristic for MWW

[1]Recall that the rational setup calculates means across individual deontic values, while the opportunistic setup focuses on extremal values (see Subsection 6.2.5).

tests) (Lehmann, 2006). The results for the KS test reflect the subtle distribution changes and reduce p values considerably, with all opportunistic configurations showing significant differences. However, statistical differentiation in the case of the rational configuration is too limited to permit conclusive differentiation. In addition to the scenario implications, this finding can advise the selection of deontic configurations such as the opportunistic strategy that amplify nuances of normative understanding. Moving beyond the analysis of significance, in the following section we explore the model's reaction to changing population size.

Table E.3: Significance Tests for Integrated vs. Differentiated Role Understanding (Action 'trade fair')

Enf.	Strat.	MWW (U)	p value	KS (D)	p value
F	RAT	17 507	0.001	0.192	0.001
T	RAT	16 240	0	0.271	0
F	OPP	17 917.5	0.011	0.172	0.005
T	OPP	9541	0	0.65	0

Table E.4: Significance Tests for Integrated vs. Differentiated Role Understanding (Action 'withhold profit')

Enf.	Strat.	MWW (U)	p value	KS (D)	p value
F	RAT	21 607	0.308	0.113	0.148
T	RAT	18 996.5	0.11	0.113	0.148
F	OPP	22 419	0.099	0.158	0.013
T	OPP	17 658	0.012	0.3	0

E.3 Sensitivity to Population Size

To test the model's sensitivity to changing population size, we further ran each experiment configuration with twofold agent numbers, thus comparing resulting deontic compartment distributions for 100 vs. 200 agents for the Maghribī variant, and 200 vs. 400 agents for the Genoese counterpart. As before, we draw on the Mann-Whitney-Wilcoxon test (MWW) (Mann and Whitney, 1947) and Kolmogorov-Smirnov test (KS) (see e.g. Lehmann (2006)), since non-normally distributed data prevent the use of Student's T-test.

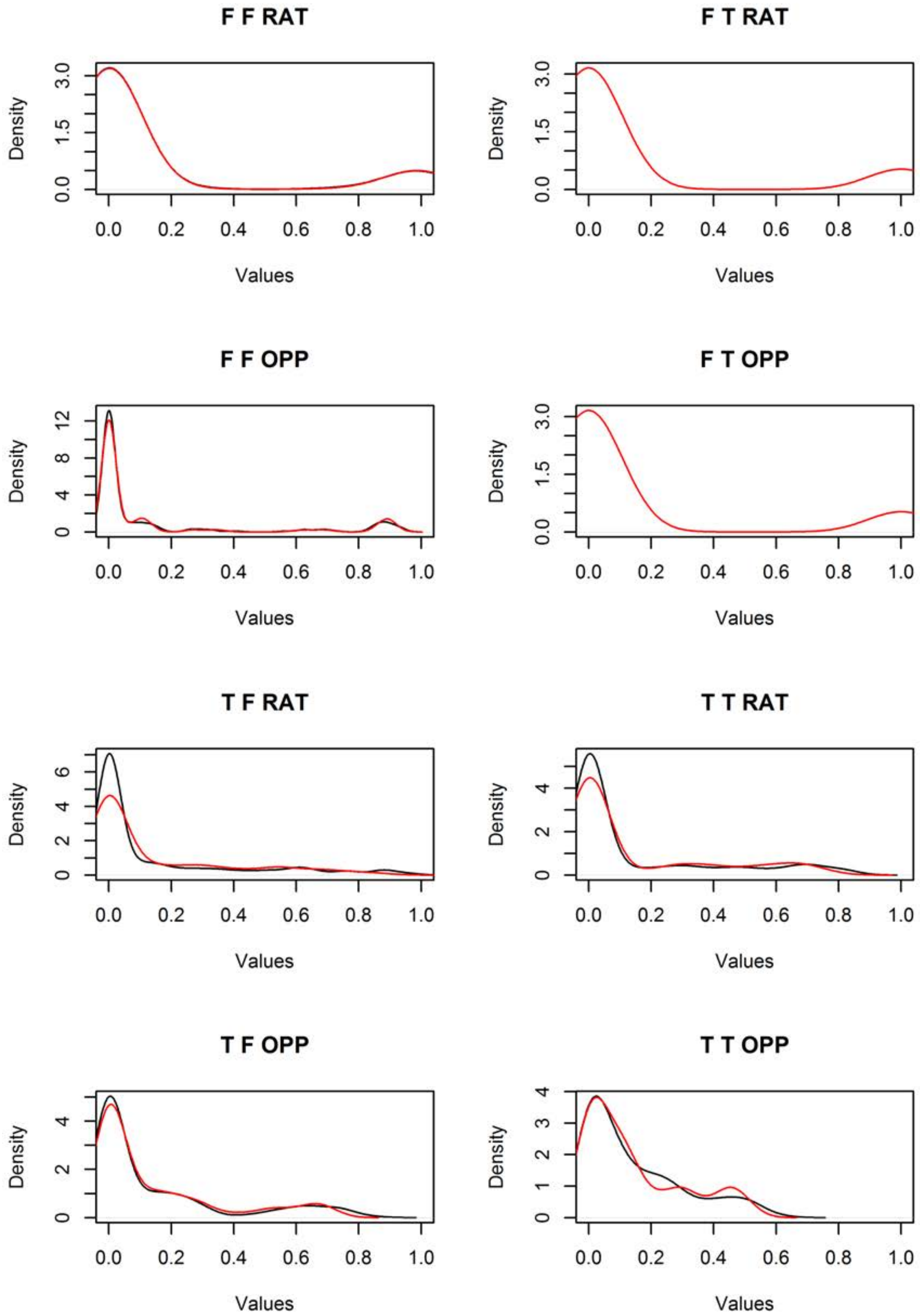
The results, separated by action, are shown in Tables E.5 and E.6. Although effectively all compared values differ (indicated by high U values), the shifts are well above the significance level, with the lowest p-value being 0.503 and most values > 0.9. This suggests that the variation of agent numbers has insignificant impact on the distribution of normative understanding across deontic compartments. Neither the MMW nor KS test could establish a significant difference in the respective distributions, thus indicating the model's robustness against changing agent numbers (within the explored range).

Table E.5: Significance Tests for Scenarios comparing Base Numbers and Twofold Numbers of Agents (Action 'trade fair')

	Role Diff.	Enf.	Strat.	MWW (U)	p value	KS (D)	p value
Maghribīs	F	F	RAT	20 595.5	0.992	0.015	1
	F	T	RAT	20 604.5	1	0	1
	F	F	OPP	20 659	0.957	0.025	1
	F	T	OPP	20 604.5	1	0	1
Genoese	T	F	RAT	20 166	0.675	0.044	0.988
	T	T	RAT	20 436	0.874	0.03	1
	T	F	OPP	19 857	0.503	0.064	0.799
	T	T	OPP	20 222	0.746	0.049	0.966

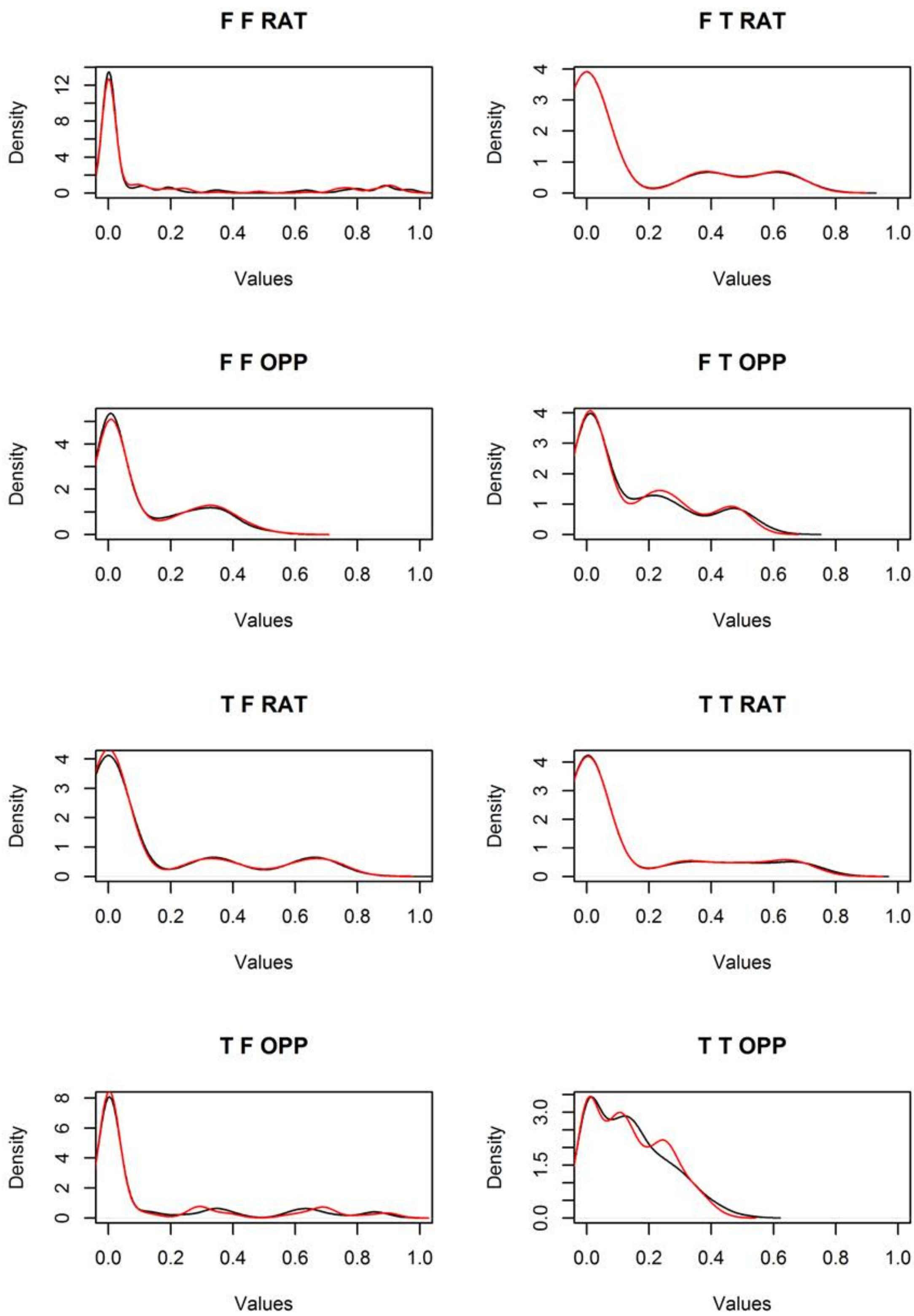
Table E.6: Significance Tests for Scenarios comparing Base Numbers and Twofold Numbers of Agents (Action ‘withhold profit’)

	Role Diff.	Enf.	Strat.	MWW (U)	p value	KS (D)	p value
Maghribīs	F	F	RAT	20657	0.959	0.025	1
	F	T	RAT	20532	0.94	0.025	1
	F	F	OPP	19862.5	0.516	0.064	0.799
	F	T	OPP	20401	0.86	0.044	0.988
Genoese	T	F	RAT	20604	1	0.02	1
	T	T	RAT	20313.5	0.784	0.03	1
	T	F	OPP	20551.5	0.96	0.039	0.997
	T	T	OPP	20569	0.976	0.049	0.966



The distribution of the base configuration is represented in black colour; the distribution of results for twofold agent numbers are shown in red colour.

Figure E.1: Distributions for Action 'trade fair'



The distribution of the base configuration is represented in black colour; the distribution of results for twofold agent numbers are shown in red colour.

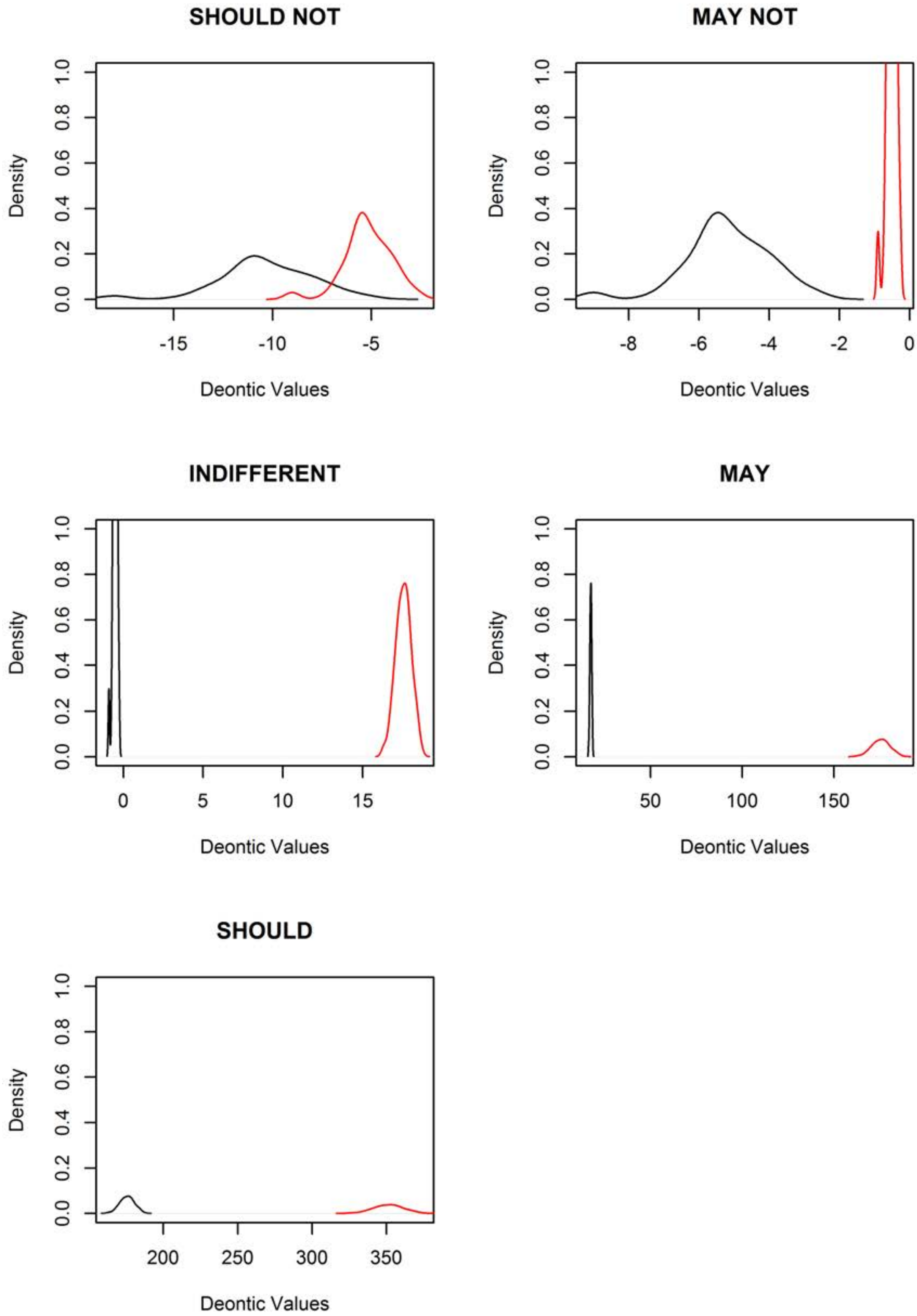
Figure E.2: Distributions for Action 'withhold profit'

F

Analysing Normative Understanding with Interval Type-2 Fuzzy Sets

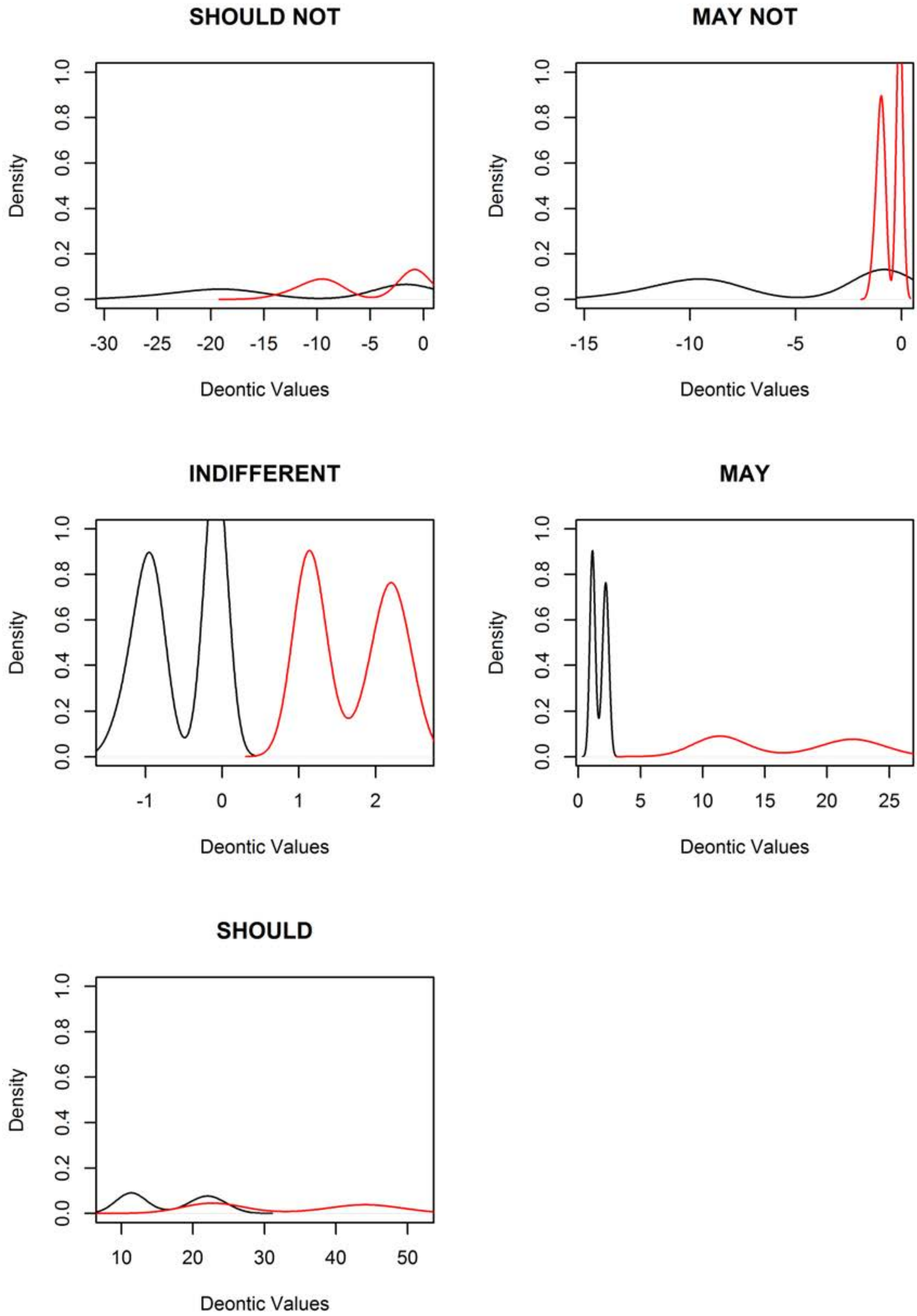
F.1 Input Interval Distributions for IT2FLS Membership Function Generation

The following figures show the distribution of input intervals for different normative compartments of the Maghribī scenario (Figure F.1) and the role-differentiated Genoese scenario (Figure F.2). Since interval sizes vary between individual agents, we plotted the distributions for the left interval boundaries and the right boundaries separately. Doing so we can observe the varying alignment across left and right boundaries, such as for the term *may not* in the Maghribī scenario, with a near unified alignment of the right boundary and more widely distributed left boundary. However, only in few cases an approximation of a normal distribution could be suggested, thus limiting the applicability of parametric tests.



The distributions of the input intervals' left boundaries are represented in black colour.
The distributions of the right boundaries are shown in red colour.

Figure F.1: Distribution of Input Intervals for Role-Integrated Maghribī Scenario



The distributions of the input intervals' left boundaries are represented in black colour.
The distributions of the right boundaries are shown in red colour.

Figure F.2: Distribution of Input Intervals for Role-Differentiated Genoese Scenario

F.2 Macro-Level Alignment of Maghribī Society without Interval Filtering

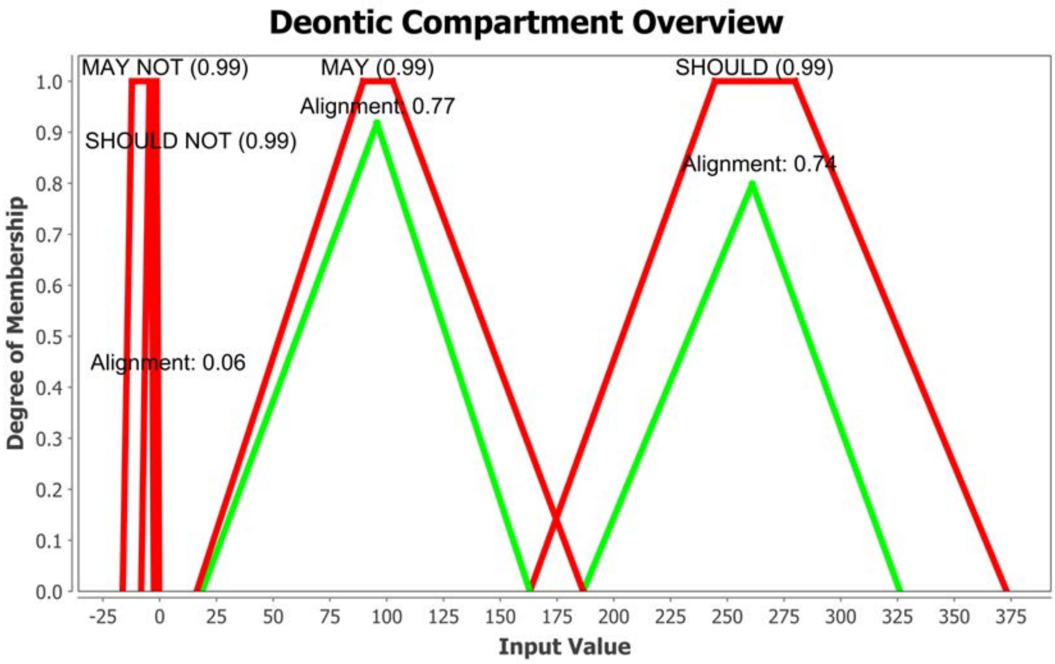


Figure F.3: IT2FS Representation of Deontic Compartments for Maghribī Scenario (Macro-Level)

Table F.1: Macro-Level Compartment Measures (Maghribī)

Compartment	Representation	Alignment	Min. Value	Max. Value
should not	0.99	0.0	−16.32	−2.57
may not	0.99	0.06	−8.16	−0.26
may	0.99	0.77	16.3	186.53
should	0.99	0.74	163.04	373.07

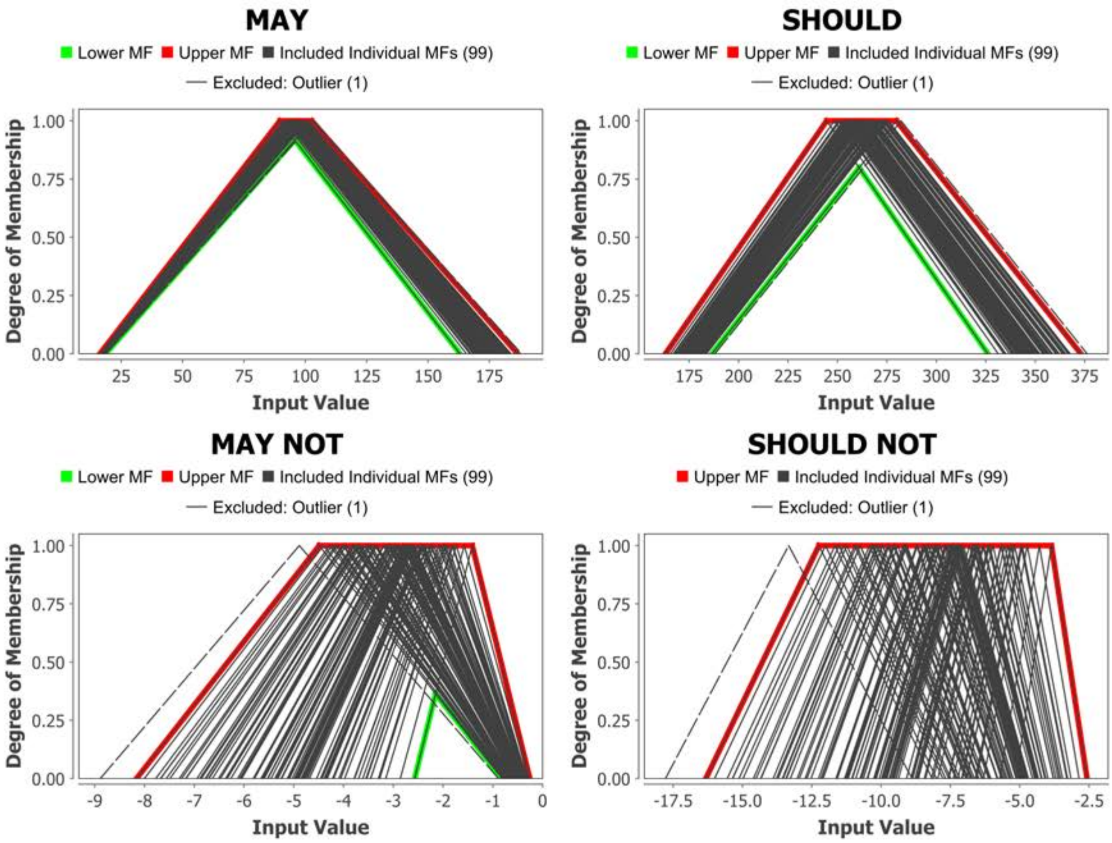


Figure F.4: Dashboard View for Maghribī Scenario (Macro-Level)

F.3 Statistical Overview of Deontic Compartments for Maghribī and Genoese Scenarios

In this section we provide a representative statistical overview of the membership function generation process for the Maghribī and Genoese societies, capturing results for the macro-level (see Section 7.4) as well as for the meso-level (as described in Subsection 7.5.2). In addition to the individual results, in Subsection F.3.3 we provide a significance test that supports the significantly differing results for both societies. All results are based on 30 simulation runs for each configuration.

F.3.1 Maghribīs

Macro-Level In Table F.2 we provide statistical measures for the macro-level variant for the Maghribī case, which are compatible with the results in Subsection 7.4.1.

Table F.2: Macro-Level Compartment Measures (Maghribī)

Compartment	Representation		Alignment		Min. Value		Max. Value	
	mean	σ	**mean**	σ	**mean**	σ	**mean**	σ
should not	0.98	0.02	0	0	−16.704	0.941	−2.554	0.338
may not	0.98	0.02	0.098	0.026	−8.352	0.47	−0.255	0.034
may	0.992	0.011	0.934	0.006	16.402	0.14	186.721	1.571
should	0.992	0.011	0.852	0.015	164.022	1.402	373.439	3.143

Meso-Level In addition, we provide a meso-level overview based on left- and rightmost clustering strategies (as introduced in Subsection 7.5.1) in Tables F.3 and F.4. The meso-level results show only minimal deviations from the entirety of all traders (macro-level) as well as between the left and right cluster, which supports the relative uniformity of the Maghribīs' normative understanding suggested in Subsection 7.4.1. Only exception is the deontic compartment *should not* that displays comparatively high standard deviation for all measures (and particularly for the compartment boundaries). The reason for this lies in the comparatively narrow range (compared to *may* and *should*) caused by poor reinforcements for negative values, which in turn gives individual reinforcements greater impact. In combination with the zero-centred deontic range configuration (which enforces zero as the centre of all compartments as shown in Subsection 6.2.5), this leads to relatively extreme shifts (given the proscriptive compartments' narrow sizes) along the left (negative) part of zero-centred deontic range if reinforcements occur. This, in consequence, drives the greater variation of compartment boundaries. The significance test results of left and right cluster measures support the otherwise strong homogeneity of the Maghribī scenario as shown in Table F.5 for p values based on Mann-Whitney-Wilcoxon test (Mann and Whitney, 1947), significance level: 0.05), suggesting that the clustering mechanism (introduced in Subsection 7.5.1) identifies the *same* cluster both for 'leftmost' and 'rightmost' cluster selection strategy. Only the comparatively narrow compartment *should not* has significantly differing interval boundaries (but only insignificant shifts in *Representation* or *Alignment* levels). As indicated above, this significance is caused by the extremely narrow intervals enforced by the zero-centred deontic compartment configuration (see Figure 7.5 for clarification) in combination with poor reinforcement levels.

Table F.3: Meso-Level Compartment Measures (Maghribī) – Left Cluster

Compartment	**Representation**		**Alignment**		**Min. Value**		**Max. Value**	
	mean	σ	**mean**	σ	**mean**	σ	**mean**	σ
should not	0.881	0.282	0.071	0.223	−16.84	1.207	−3.208	1.695
may not	0.98	0.02	0.059	0.023	−8.352	0.47	−0.255	0.034
may	0.992	0.011	0.782	0.018	16.402	0.14	186.721	1.571
should	0.992	0.011	0.749	0.022	164.022	1.402	373.439	3.143

Table F.4: Meso-Level Compartment Measures (Maghribī) – Right Cluster

Compartment	**Representation**		**Alignment**		**Min. Value**		**Max. Value**	
	mean	σ	**mean**	σ	**mean**	σ	**mean**	σ
should not	0.82	0.357	0.134	0.305	−14.57	4.324	−2.554	0.338
may not	0.98	0.02	0.059	0.023	−8.352	0.47	−0.255	0.034
may	0.992	0.011	0.782	0.018	16.402	0.14	186.721	1.571
should	0.992	0.011	0.749	0.022	164.022	1.402	373.439	3.143

Table F.5: p values for Left and Right Clusters in Maghribī Scenario (Meso-Level)

Compartment	**Representation**	**Alignment**	**Min. Value**	**Max. Value**
should not	0.751	0.102	0	0
may not	1	1	1	1
may	1	1	1	1
should	1	1	1	1

F.3.2 Genoese

Macro-Level The results for the Genoese scenario are compatible with representative runs shown in Subsections 7.4.2 and 7.5.2 respectively. The macro-level results (Table F.6) emphasise high levels of inclusiveness (*Representation*), but hardly display any alignment. As argued in Subsection 7.4.2, this suggests the existence of a heterogeneous normative landscape with sub-groups of aligned normative understanding.

Table F.6: Macro-Level Compartment Measures (Genoese)

Compartment	Representation		Alignment		Min. Value		Max. Value	
	mean	σ	**mean**	σ	**mean**	σ	**mean**	σ
should not	1	0	0	0	-30.841	2.191	-0.194	0.076
may not	1	0	0	0	-15.42	1.096	-0.019	0.008
may	1	0	0.046	0.014	0.735	0.08	27.053	0.842
should	1	0	0	0	7.351	0.8	54.104	1.684

Meso-Level *Representativeness* and *Alignment* levels for identified left and right clusters are shown in Tables F.7 and F.8. In contrast to the Maghribī case (Subsection F.3.1) the results show a *significantly differing* nature of the meso-level clusters (see Table F.9 for p values of Mann-Whitney-Wilcoxon test (Mann and Whitney, 1947), significance level: 0.05), an aspect that is retraceable from the observation of interval boundary values ('Min. Value' and 'Max. Value' in Tables F.7 and F.8). Here the only exceptions are the representation levels for the compartments *may* and *should*. This is consistent with the understanding of a bifurcated normative understanding, in which both compartments represent a *similar* fraction of *different* individuals. However, this does not carry any implications with respect to the alignment or positioning of those compartments along the deontic range.

Compared to the macro-level results, the meso-level clusters underlie more frequent changes with respect to *Representativeness* and *Alignment*, which is reflected in the comparatively high standard deviations for multiple simulation runs. Specifically, the right cluster (Table F.8) shows strong variations for the extreme narrow deontic compartments *should not* and *may not* (compared to the wide value ranges for *may* and *should*), which is caused by their comparatively poor reinforcement (see Subsection 7.5.2). As discussed for the Maghribī case (Subsection F.3.1), for the proscriptive compartments occasional reinforcements thus have strong situational impact on the cluster boundaries, leading to stronger alignment for a subset of cluster members that experienced such reinforcement (particularly visible in the compartment *should not* of the right cluster (Table F.8)). However, this comes at the price of excluding intervals that did not experience this situational reinforcement, which consequentially leads to reduced levels of *Representativeness*.

As mentioned before in Subsection F.3.1, this effect is an artefact of the zero-centred deontic compartment configuration and operationalisation based on reinforcement learning (introduced in Subsection 6.2.5). Other than pointing to poor reinforcement levels for those compartments, the results have limited relevance for the interpretation of the scenario. This aspect becomes clearer when observing higher alignment levels for strongly reinforced (and thus wider) compartments such as *may* and *should*.

Table F.7: Meso-Level Compartment Measures (Genoese) – Left Cluster

Compartment	**Representation**		**Alignment**		**Min. Value**		**Max. Value**	
	mean	**σ**	**mean**	**σ**	**mean**	**σ**	**mean**	**σ**
should not	0.498	0.004	0.029	0.036	−29.975	2.774	−7.6	0.916
may not	0.499	0.003	0.218	0.043	−14.988	1.387	−0.747	0.089
may	0.799	0.25	0.144	0.119	0.746	0.07	21.358	7.05
should	0.581	0.19	0.066	0.062	7.567	0.693	31.734	10.033

Table F.8: Meso-Level Compartment Measures (Genoese) – Right Cluster

Compartment	**Representation**		**Alignment**		**Min. Value**		**Max. Value**	
	mean	**σ**	**mean**	**σ**	**mean**	**σ**	**mean**	**σ**
should not	0.263	0.227	0.27	0.315	−1.831	1.108	−0.338	0.052
may not	0.471	0.088	0.018	0.044	−1.466	0.25	−0.029	0.005
may	0.799	0.251	0.204	0.193	1.157	0.496	26.708	0.725
should	0.581	0.191	0.234	0.123	15.848	3.606	53.193	1.228

Table F.9: p values for Left and Right Clusters in Genoese Scenario (Meso-Level)

Compartment	**Representation**	**Alignment**	**Min. Value**	**Max. Value**
should not	0	0.038	0	0
may not	0	0	0	0
may	0.839	0.039	0	0
should	0.918	0	0	0

F.3.3 Significance Test of Maghribī vs. Genoese Macro-Level Results

Supporting our earlier observations, the difference between Maghribī and Genoese scenario is significant. In all configurations for all compartments p values approximate 0, as determined using the Mann-Whitney-Wilcoxon test (Mann and Whitney, 1947) (significance level: 0.05). For all cases we can thus reject the null hypothesis that the result distributions for the Maghribī and Genoese scenarios are identical. This finding is compatible with the vastly differing distributions across deontic compartments for both scenarios observed in the previous Subsections F.3.1 and F.3.2.

Table F.10: p values for Maghribī vs. Genoese Macro-Level Results

Compartment	**Representation**	**Alignment**	**Min. Value**	**Max. Value**
should not	0	0	0	0
may not	0	0	0	0
may	0	0	0	0
should	0	0	0	0