Normative Multi-Agent Systems

Edited by

Giulia Andrighetto
Guido Governatori
Pablo Noriega
Leendert W. N. van der Torre

DAGSTUHL
FOLLOW-UPS

Dagstuhl Follow-Ups – Vol. 4 www.dagstuhl.de/dfu

Editors

Giulia Andrighetto
ISTC-CNR; EUI
Italy
giulia.andrighetto@istc.cnr.it

Guido Governatori
NICTA
St. Lucia, Australia
guido.governatori@nicta.com.au

Pablo Noriega
IIIA – CSIC
Barcelona, Spain
pablo@iiia.csic.es

Leendert W. N. van der Torre
University of Luxembourg
Luxembourg
leon.vandertorre@uni.lu

ACM Classification 1998
I.2.11 Distributed Artificial Intelligence

ISBN 978-3-939897-51-4

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at http://www.dagstuhl.de/dagpub/978-3-939897-51-4.

Publication date
April, 2013

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Digital Object Identifier: 10.4230/DFU.Vol4.12111.i

ISBN 978-3-939897-51-4 **ISSN 1868-8977** **http://www.dagstuhl.de/dfu**

DFU – Dagstuhl Follow-Ups

The series *Dagstuhl Follow-Ups* is a publication format which offers a frame for the publication of peer-reviewed papers based on Dagstuhl Seminars. DFU volumes are published according to the principle of Open Access, i.e., they are available online and free of charge.

ISSN 1868-8977

www.dagstuhl.de/dfu

Contents

Norms in MAS: Definitions and Related Concepts
Tina Balke, Célia da Costa Pereira, Frank Dignum, Emiliano Lorini, Antonino Rotolo, Wamberto Vasconcelos, and Serena Villata 1

Normative Reasoning and Consequence
Jan Broersen, Stephen Cranefield, Yehia Elrakaiby, Dov Gabbay, Davide Grossi, Emiliano Lorini, Xavier Parent, Leendert W. N. van der Torre, Luca Tummolini, Paolo Turrini, and François Schwarzentruber 33

Computational Models for Normative Multi-Agent Systems
Natasha Alechina, Nick Bassiliades, Mehdi Dastani, Marina De Vos, Brian Logan, Sergio Mera, Andreasa Morris-Martin, and Fernando Schapachnik 71

Regulated MAS: Social Perspective
Pablo Noriega, Amit K. Chopra, Nicoletta Fornara, Henrique Lopes Cardoso, and Munindar P. Singh 93

(Social) Norm Dynamics
Giulia Andrighetto, Cristiano Castelfranchi, Eunate Mayor, John McBreen, Maite Lopez-Sanchez, and Simon Parsons 135

Simulation and NorMAS
Tina Balke, Stephen Cranefield, Gennaro Di Tosto, Samhar Mahmoud, Mario Paolucci, Bastin Tony Roy Savarimuthu, and Harko Verhagen 171

The Uses of Norms
Munindar P. Singh, Matthew Arrott, Tina Balke, Amit Chopra, Rob Christiaanse, Stephen Cranefield, Frank Dignum, Davide Eynard, Emilia Farcas, Nicoletta Fornara, Fabien Gandon, Guido Governatori, Hoa Khanh Dam, Joris Hulstijn, Ingolf Krueger, Ho-Pun Lam, Michael Meisinger, Pablo Noriega, Bastin Tony Roy Savarimuthu, Kartik Tadanki, Harko Verhagen, and Serena Villata 191

Normative Multi-Agent Systems. *Dagstuhl Follow-Ups, Vol. 4.* ISBN 978-3-939897-51-4.
Editors: Giulia Andrighetto, Guido Governatori, Pablo Noriega, and Leendert W.N. van der Torre
DAGSTUHL FOLLOW-UPS Dagstuhl Publishing
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany

Preface

As research in Multi-Agent Systems (MAS) has been expanding its focus from from the individual, cognitive focussed, agent models to models of socially situated agents, MAS researchers have been showing rising interest in social theories. Particular attention has been given to normative concepts because it is expected that norms could play as key a role in articulating agent interactions as the one norms play in human social intelligence. Thus, the label of "normative multi-agent system" has been attached to systems where individual and collective behaviour is affected by norms. This book is not a state of the art of normative multi-agent systems, nor a systematic description of the key concepts, or a compendium of the most salient challenges. However, the reader will find in its chapters something of each of these three contents because *Normative Multi-Agent Systems* is an effort to clarify the ideas behind the label and to put in perspective the work that is being done in this area.

Normative Multi-Agent Systems is the outcome of the 2012 Schloss Dagstuhl Seminar on Normative Multi-Agent Systems[1], the third in a series of Schloss Dagstuhl seminars on Normative Multi-Agent Systems. The first seminar (07122)[2], in 2007, had the aim of identifying common definitions, ontologies, research problems and applications in the field. The second seminar (09121)[3], in 2009, had instead the aim of discussing these fundamental concepts in relation to the use of norms as a regulatory mechanism in human and artificial systems. Building on the work of these two workshops, the 2012 seminar was convened to produce a forward-looking account of current research in the area. Some forty specialists were invited to prepare short position papers along seven research topics. Prior to the seminar, these papers went under a review process, and discussed among authors contributing to the same topic. After this process, authors were encouraged to prepare new position papers that became the basis for short presentations. These presentations and the preceding work gave substance to discussion groups that were formed during the workshop around particular norm related topics. These groups reported their findings in plenary sessions, provoking a lively debate, and eventually drafted the seven chapters that make this book.

The chapters of this Dagstuhl Follow-Ups volume focus on the following topics.

Chapter 1, titled *Norms in MAS: Definitions and Related Concepts*, provides an introductory presentation of normative multi-agent systems (nMAS). The main idea of the chapter is that any definition of nMAS should preliminarily clarify meaning, scope, and function of the concept of norm. On account of this idea, the authors focus on three definitions and some related requirements for nMAS. For each of such definitions, some guidelines for developing normative MAS have been proposed. Then, it has been discussed how to relate the concept of normative MAS to different conceptions of norms and how norms can be used within the systems. Finally, some specific issues that open research questions or that exhibit interesting overlaps with other disciplines have been identified.

Chapter 2, called *Normative Reasoning and Consequence*, provides a general introduction to deontic logic and normative reasoning. Then, the authors discuss why normative reasoning is relevant for normative multi-agent systems and point out the advantages of formal methods in multi- agent systems. Finally, current research challenges are discussed.

[1] `http://www.dagstuhl.de/12111`
[2] `http://www.dagstuhl.de/07122`
[3] `http://www.dagstuhl.de/09121`

Normative Multi-Agent Systems. *Dagstuhl Follow-Ups, Vol. 4.* ISBN 978-3-939897-51-4.
Editors: Giulia Andrighetto, Guido Governatori, Pablo Noriega, and Leendert W.N. van der Torre
DAGSTUHL FOLLOW-UPS Dagstuhl Publishing
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany

Chapter 3, titled *Computational Models for Normative Multi-Agent Systems*, addresses the problem of building normative multi-agent systems. It takes a closer look at computational logic approaches for the design, verification and the implementation of normative multi-agent systems. Finally, an overview of current research challenges is provided.

Chapter 4, *Regulated MAS: Social Perspective*, addresses the problem of building normative multi-agent systems in terms of regulatory mechanisms. It describes a static conceptual model through which one can specify normative multi-agent systems along with a dynamic model to capture their operation and evolution. The chapter proposes a typology of applications and discusses some open problems.

Chapter 5, titled *(Social) Norm Dynamics*, is concerned with the *dynamics* of social norms. In particular the chapter concentrates on the lifecycle that social norms go through, focusing on the generation of norms, the way that norms spread and stabilize, and finally evolve. The cognitive mechanisms behind norm compliance, the role of culture in norm dynamics, and the way that trust affects norm dynamics have been finally discussed.

Chapter 6, *Simulation and NorMAS*, discusses state of the art and future perspective of the study of norms with simulative methodologies, in particular employing agent-based simulation. The authors discuss the research challenges that they feel more apt to be tackled by the simulative approach. Finally, indications for the realization of a NorMAS simulation platform, illustrated by selected scenario, conclude the chapter.

Chapter 7, called *The Uses of Norms*, concludes this Dagstuhl Follow-Ups volume. It presents a variety of applications of norms. These applications include governance in sociotechnical systems, data licensing and data collection, understanding software development teams, requirements engineering, assurance, natural resource allocation, wireless grids, autonomous vehicles, serious games, and virtual worlds.

List of Authors

Natasha Alechina
School of Computer Science
University of Nottingham
United Kingdom
nza@cs.nott.ac.uk

Giulia Andrighetto
Institute of Cognitive Science and Technologies, ISTC-CNR
European University Institute, EUI
Italy
giulia.andrighetto@istc.cnr.it

Matthew Arrott
University of California, San Diego
USA

Tina Balke
University of Surrey
United Kingdom
t.balke@surrey.ac.uk

Nick Bassiliades
Dept. of Informatics, Aristotle University of Thessaloniki
Greece
nbassili@csd.auth.gr

Jan Broersen
Utrecht University
The Netherlands
broersen@cs.uu.nl

Henrique Lopes Cardoso
Universidade do Porto
Portugal
hlc@fe.up.pt

Cristiano Castelfranchi
Institute of Cognitive Science and Technologies, ISTC-CNR
Italy
cristiano.castelfranchi@istc.cnr.it

Amit K. Chopra
Lancaster University
United Kingdom
a.chopra1@lancaster.ac.uk

Stephen Cranefield
University of Otago
New Zealand
stephen.cranefield@otago.ac.nz

Rob Christiaanse
Vrije Universiteit
The Netherlands

Célia da Costa Pereira
Université de Nice Sophia Antipolis
France
celia.pereira@unice.fr

Mehdi Dastani Dept of Information and Computer Sciences
University of Utrecht
The Netherlands
M.M.Dastani@uu.nl

Marina De Vos Dept. of Computer Science,
University of Bath
United Kingdom
mdv@cs.bath.ac.uk

Frank Dignum
Utrecht University
The Netherlands
F.P.M.Dignum@uu.nl

Gennaro Di Tosto
Utrecht University
The Netherlands
g.ditosto@uu.nl

Yehia Elrakaiby
University of Luxembourg
Luxembourg

Davide Eynard
Università della Svizzera italiana
Switzerland
davide.eynard@usi.ch

Emilia Farcas
University of California at San Diego
USA

Normative Multi-Agent Systems. *Dagstuhl Follow-Ups, Vol. 4.* ISBN 978-3-939897-51-4.
Editors: Giulia Andrighetto, Guido Governatori, Pablo Noriega, and Leendert W.N. van der Torre
DAGSTUHL FOLLOW-UPS Dagstuhl Publishing
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany

Nicoletta Fornara
Università della Svizzera italiana
Switzerland
nicoletta.fornara@usi.ch

Dov Gabbay
King's College London
United Kingdom
dov.gabbay@kcl.ac.uk

Fabien Gandon
INRIA Sophia Antipolis
France
Fabien.Gandon@inria.fr

Guido Governatori
NICTA
Australia
guido.governatori@nicta.com.au

Davide Grossi
University of Liverpool
England
d.grossi@liverpool.ac.uk

Joris Hulstijn
Delft University of Technology
The Netherlands
J.Hulstijn@tudelft.nl

Hoa Khanh Dam
University of Wollongong
Australia
hoa@uow.edu.au

Ingolf Krueger
University of California at San Diego
USA
ikrueger@ucsd.edu

Ho-Pun Lam
NICTA
Australia

Brian Logan
School of Computer Science
University of Nottingham
bsl@cs.nott.ac.uk

Maite Lopez-Sanchez
University of Barcelona
maite_lopez@ub.edu

Emiliano Lorini
Paul Sabatier University - Toulouse
France
Emiliano.Lorini@irit.fr

Samhar Mahmoud
Kings College - London
United Kingdom
Samhar samhar.mahmoud@kcl.ac.uk

Eunate Mayor
LMTG/GET UMR5563,
IRD-CNRS-Universite P. Sabatier Toulouse III
eunate.mayor@gmail.com

John McBreen
Wageningen University
Wageningen, The Netherlands
johnmcbreen@gmail.com

Sergio Mera
Departamento de Computación, FCEyN
Universidad de Buenos Aires
Argentina
smera@dc.uba.ar

Michael Meisinger
University of California at San Diego
USA

Andreasa Morris-Martin
Dept. of Computer Science
University of Guyana
andreasa.morris@uog.edu.gy

Pablo Noriega
IIIA-CSIC
Spain
pablo@iiia.csic.es

Mario Paolucci
LABSS, ISTC-CNR Rome
Italy
mario.paolucci@istc.cnr.it

Simon Parsons
Brooklyn College
City University of New York
USA
parsons@sci.brooklyn.cuny.edu

Xavier Parent
University of Luxembourg
Luxembourg
xavier.parent@uni.lu

Antonino Rotolo
University of Bologna
Italy
antonino.rotolo@unibo.it

Bastin Tony Roy Savarimuthu
University of Otago
New Zealand
TonyR@infoscience.otago.ac.nz

Fernando Schapachnik
Departamento de Computación, FCEyN
Universidad de Buenos Aires
Argentina
fernando@schapachnik.com.ar

François Schwarzentruber
ENS Cachan / IRISA
France
francois.schwarzentruber@bretagne.ens-cachan.fr

Munindar P. Singh
North Carolina State University
USA
singh@ncsu.edu

Kartik Tadank
Deutsche Bank
USA

Luca Tummolini
National Research Council
Italy
luca.tummolini@istc.cnr.it

Paolo Turrini
University of Luxembourg
Luxembourg
paolo.turrini@uni.lu

Leendert W. N. van der Torre
University of Luxembourg
Luxembourg
leon.vandertorre@uni.lu

Wamberto Vasconcelos
University of Aberdeen
United Kingdom
wvasconcelos@acm.org

Harko Verhagen
Stockholm University
Sweden
verhagen@dsv.su.se

Serena Villata
INRIA Sophia Antipolis
France
serena.villata@inria.fr

Norms in MAS: Definitions and Related Concepts

Tina Balke[1], Célia da Costa Pereira[2], Frank Dignum[3], Emiliano Lorini[4], Antonino Rotolo[5], Wamberto Vasconcelos[6], and Serena Villata[7]

1 University of Surrey, UK
2 Université de Nice Sophia Antipolis, France
3 Utrecht University, The Netherlands
4 Paul Sabatier University – Toulouse, France
5 University of Bologna, Italy
6 University of Aberdeen, UK
7 INRIA Sophia Antipolis, France

Abstract

In this chapter we provide an introductory presentation of normative multi-agent systems (nMAS). The key idea of the chapter is that any definition of nMAS should preliminarily clarify meaning, scope, and function of the concept of norm. On account of this idea, we focus on three definitions and some related requirements for nMAS. For each of such definitions we propose some guidelines for developing nMAS. Second, we suggest how to relate the concept of nMAS to different conceptions of norms and how norms can be used within the systems. Finally, we identify some specific issues that open research questions or that exhibit interesting overlaps with other disciplines.

1998 ACM Subject Classification I.2.11 Distributed Artificial Intelligence

Keywords and phrases Norms, MAS

Digital Object Identifier 10.4230/DFU.Vol4.12111.1

1 Introduction

Normative systems are those in which "norms play a role and which need normative concepts in order to be described or specified" [70, preface]. There has been in the last years an increasing interest in normative systems in the computer science community, due, among the other reasons, to the AgentLink RoadMap's [65] observation that norms must be introduced in agent technology in the medium term for infrastructure for open communities, reasoning in open environments and trust and reputation. Indeed, normative multi-agent systems (nMAS) revolve around the idea that, while the main objective of MAS research is to design systems of autonomous agents, it is likewise important that agent systems may exhibit global desirable properties. One possible strategy to achieve this goal is that, like in human societies, such desirable properties be ensured by normative constraints: the interaction of artificial agents, too, adopts normative models whose goal is to govern agents' behavior through normative systems in supporting coordination, cooperation and decision-making. The deontic logic and artificial intelligence and law communities, for instance, agree about the structure and properties of norms [38]. The nMAS community, too, has strong and obvious connections with the development of rule-based systems and technologies.

Layout of the Chapter

It is widely acknowledged that normative concepts can play an important role in MAS and many themes and methods have obtained a reasonable degree of consensus. However, there

Normative Multi-Agent Systems. *Dagstuhl Follow-Ups, Vol. 4.* ISBN 978-3-939897-51-4.
Editors: Giulia Andrighetto, Guido Governatori, Pablo Noriega, and Leendert W.N. van der Torre; pp. 1–31
DAGSTUHL FOLLOW-UPS Dagstuhl Publishing
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany

are still different views in regard to some fundamental research questions, such as the kind of norms to be used, or the way to use them.

The layout of the chapter is as follows:

- In Section 2 we will offer a general discussion of basic theoretical assumptions behind any meaningful usage of norms in MAS.
- In Section 3 we will discuss three definitions of nMAS, one emerging from social sciences and two that revise the ones proposed in [20, 23, 18, 19] and that emerged from past NorMAS workshops. Notice that we will not assume that the social perspective of norms be necessarily contrasted with the legal one. In fact, these two views are often taken to be symmetrically opposed: in the social paradigm norms fall within a bottom-up approach to normativity that is based on the concept of norm emergence; in the legal paradigm norms are mostly defined within a top-down, authority-based and institutionalised perspective. While it is far from obvious that norm emergence does not play any role in legal systems [28], we prefer framing the definitions of nMAS in a slightly different way: the key problem is clarifying what norms are supposed to do in MAS and what normative strategy is best suited to govern agents' social interaction.
- In Section 5 we will address some open (and somehow overlooked) research issues in MAS and show that they are relevant for nMAS and for establishing interesting links with other disciplines.

2 A General Picture: The Nature of Norms

In order to determine the possible roles of norms in MAS we will have a look at the position of norms in human society. In a very general sense norms regulate the interactions between an individual and the society (i.e., other individuals or groups of individuals). A consequence of this very simple statement is that norms would not make any sense if we consider only one individual in an environment (which possibly might contain other individuals, but which are not distinguished from other environmental elements). Of course, this individual might be influenced by the environment. It can be constrained by physical properties or enticed by them to perform or avoid certain (sequences of) actions. However, patterns of behavior that emerge from this situation would not be called norms, but rather habits.

The second aspect of norms is that they regulate the interactions between (groups of) individuals as part of and through the *social reality.* Although this is not the place to expand very much on the nature of social reality (see, e.g., [88] for some analysis) we will say a few words about it as far as it determines the nature of norms. In short social reality is a reality that is created and exists solely by a kind of joint agreement of the individuals in a society. It therefore does not exist without there being individuals (like the physical reality), but it always exists in some form when there are two or more individuals that are interdependent. Some famous examples of elements of social reality are concepts such as *money* and *ownership.* Because the elements of social reality only exist virtually (and by virtue of agreement) they need to be connected in some way to the physical reality in order for them to be observable, manipulable, changeable, etc. Thus money is represented by coins and banknotes and ownership is represented by documents. However, this relation is not completely fixed and static. For example, centuries ago only coins of a certain material (gold or copper) counted as money. Later promissary notes of certain individuals also counted as money. Nowadays most money is only represented as data in a computer.

By the mere fact that norms are part of social reality they are also created by mutual agreement and thus their nature is not entirely fixed but malleable by mutual agreement as well. However, as with most concepts of social reality the meaning and nature of norms is not completely random. There is a so-called core part which is (relatively) stable and a penumbra which might be more flexible. We will come back to this later when talking about different types of norms.

First, let's see what further consequences follow from the fact that norms are part of social reality. Because they are a concept of social reality they do not physically constrain the relations between individuals. Therefore it is possible to violate them. Most people would even argue that this is a fundamental property of norms. This does not have to be a property of the representation (or implementation) of the norm in the physical reality. If a prohibition to ride a train without a ticket is represented by a fence and a door to enter the train platform which can only be opened using a (valid) train ticket then it becomes physically impossible to violate this norm. However, not all norms can be implemented in physical reality in a way that the violation of them is impossible[1]. To argue this point, let's first come back to the main claim that norms regulate the interactions between an individual and other individuals in a society. These interactions between individuals consist of actions performed by the individuals that influence the behavior or result of actions of the others. Thus there are mainly two ways to regulate these interactions. First one can use prohibitions of actions that an individual might want to perform (because they are considered to interfere in a negative way with the actions of other individuals). Secondly, one can resort to obligations ordering an individual to perform at a certain moment an action (because it is considered necessary for the success of the interaction).

The second type of regulation of behavior cannot be implemented by a constraint in the environment, and so agents' regimentation can be accordingly difficult. Basically one can only prevent the violation of such a norm if it is possible to force the agent to perform the desired action. However, with people as well as any other autonomous agents this is not possible because their decisions for actions are precisely taken autonomously (except maybe for some extreme cases). So, most commonly the implementation of such a norm is done by giving an incentive to perform the action or a deterrent to avoid the action. E.g., when you buy a product you must pay for it. Why would a person pay? Because if you pay I will send you the product (there is a reward/incentive that you want to achieve—because it helps achieving the goal of possessing the product, which was the purpose of buying it in the first place). Or if you don't pay you will never get a product from me anymore.

So, the implementation of this norm actually is given by specifying what are the consequences of performing or avoiding the wanted action where the consequences are things that are typically under control of the other individuals or the environment. Thus, because we cannot connect norms directly to the internal mechanisms of an individual, they have to do with the physical reality by indicating what are the physical or observable consequences of individuals' behavior. However, this indirect way of influencing the behavior might not be successful. The individual still makes her own decision based on her internal motivations and the relation she wants to maintain with society. Thus if she decides that the consequences of not performing an action are less important than performing another action she will violate the norm. This leads us to a more general question on how norms influence individuals.

As said before, norms are part of social reality. As such, norms will use elements of social reality and their representations in physical reality to influence individuals. On the other

[1] This is always possible when one could create a physical reality that allows only for simple interactions: in this way, norms are not in fact needed because they can be implemented as physical constraints; this is discussed further in Section 3.2.

hand norms connect individuals to this social reality. So, how this relation works depends on how social reality connects with the attitudes of the individual. This is the research area of social psychology. Unfortunately this research is for a large part descriptive in nature or describes mechanisms for very limited, detailed phenomena. There are no grand unifying theories that could be used as basis for our assumptions on how norms would fit into this framework. Therefore we will limit ourselves to some intuitive (albeit perhaps faulty) remarks about how a plausible relation might look.

First of all we have to assume that individuals are basically social entities and thus social reality does influence in some way private attitudes and behavior. Although this seems obvious for people it is less so for agents. The BDI (Belief-Desire-Intention) models of agents do not intrinsically contain social attitudes. We claim that values are an essential element to connect private and social attitudes. Values play an important role in social reality and, e.g., determine part of the culture of society. Values can also be seen as private high level standards against which the relative importance of different attitudes is measured. Introducing values right away leads to a type of norm that is recognized in human society as *moral norms*. A moral norm is some regulation of interactions that, when it is followed, leads to behaviors that promote some shared values. The punishment following from the violation of the norm moves from the fact that the value is demoted: if we consider that the behavior of an individual is motivated ultimately by these shared values as well, a violation of the norm would also mean going against her own values. This right away shows that if the individual does not share (all) the values this might not be the case. Also there can be situations where two values would motivate contradictory behaviors. In that case a choice has to be made depending on the most important value and the consequences of the actions. For example, do you go into a burning house to safe a victim or stay out to safe your own life?

One could argue that moral norms are a special kind of *social norms*, which use the relation between the individual and social reality to influence its behavior. Intuitively social norms do not have explicit sanctions or rewards associated with them. The sanctions on violating a social norm come from a change of the relation between the individual and social reality. This can, e.g., be lowering of status in a group of friends after failing to help one of them. The effectiveness of these norms depends heavily on the importance of the affected social reality for the individual.

Depending on which concepts are available to describe social reality elements one can distinguish many more norm types. Each of these also only makes sense if the individual also has corresponding concepts available to take changes in its relation towards those elements into account. For example, *group norms* are norms pertaining to the members of a group, but also defining that group. Fulfilling the group norms is taken as a signal of belonging to the group and violating them is seen as detaching yourself from the group (unless you are the leader or contender and want to show your leadership in violating a norm). These norms only can influence an individual's behavior if group membership is a motivation for behavior (and not just a factual belief) of the agent. In a similar vein one can use the fact that most individuals are fulfilling a norm as a motivating factor to also fulfil the norm. This only works if being similar to other individuals (in a certain context) can be seen as motivation for an individual.

The above remarks show that the number of types of norm depend essentially on the richness of the social reality concepts available and whether the individual has motivational attitudes using its relation to those aspects of social reality. The more concepts that are available the more types of norms can be distinguished and the more complex the mechanism can be made for them to regulate interactions. On the other hand if no assumptions can be made about the individual's attitudes or decision mechanism we have to fall back on the most simple type of norms: the *regimented norms*. These norms use concepts directly related to physical reality and thus can be used as practical constraints on behavior. However, it is usually impossible (as said above) to implement all norms in this way. Most often this is remedied by constructing rigid interaction patterns where control resides always outside the individuals and unwanted behavior can be blocked, while desired actions can be forced or an alternative can be forced.

3 Requirements and Definitions for Normative MAS

Any attempt to offer a general picture of the role of norms in MAS should preliminarily assume a robust theory explaining what normativity means. Any theory serving this purpose can be fruitfully taken from social sciences, where the role of norms is typically explained in regard to real human societies. The previous section presented some general remarks in that preliminary perspective.

Indeed, a general picture can be given, as the concepts of "norm" and "normative system" have been investigated in many distinct disciplines, such as philosophy, sociology, law, and ethics. However, looking at the literature of these disciplines, even at a first glance it has to be noticed that these two terms as well as related terms such as "policies", "laws", "regulations", "convention", "values", "morals" will show vast differences when it comes to their definition and distinction from one another. This is even the case when looking at smaller research fields such as nMAS, where authors have used a multitude of ways to define these concepts, differentiate different types of norms, policies, laws, etc. as well as to model them. This lack of consensus reduces the scientific rigor of communication, modeling, and clarity of thought when it comes to discussing nMAS. That is why we start this chapter by given a brief outline of the above mentioned concepts and try to approach a definition by characteristics as well as some guidelines to distinguish them. We start by focusing on norms in particular as they are a cornerstone of nMAS, and by defining them also give distinctions for the other concepts mentioned above. As the large number of existing definitions of the concept of norms indicates, there is not "the one definition" which defines a standard way of thinking about norms. This is not surprising as norms are important in many disciplines and the term is also widely (but loosely) used by humans in everyday life. This makes a complete consensus not only hard but impossible and that is why we do not aim at a single definition [54]. Instead we believe that consensus is achievable by defining norms (as well as their relating terms) with the basic concepts pertained to them and by looking how these concepts are understood and picked up in the different research disciplines.

Perhaps, a minimal starting point for characterizing norms in MAS should the following points into account:

Rule structure of norms. It is widely acknowledged that norms have usually a conditional structure that captures the applicability conditions of the norm and its effects when it is triggered[2]. This very general view highlights an immediate link between the concepts of norm and rule.

[2] Norms can be also unconditioned, that is their effects may not depend upon any antecedent condition. Consider, for example, the norm "everyone has the right to express his or her opinion". Unconditioned norms can formally be reconstructed in terms of conditionals with no antecedent conditions.

Types of norm. There are many types of norms. The common sense meaning of norm refers to them having a regulatory and mainly prescriptive character. But norms express not only regulations about how to act. For example, von Wright [97] classified norms into the following main types (among others):

1. determinative rules, which define concepts or constitute activities that cannot exist without such rules. These rules are also called in the literature 'constitutive rules';
2. technical rules, which state that something has to be done in order for something else to be attained;
3. prescriptions, which regulate actions by making them obligatory, permitted, or prohibited. These norms, to be complete, should indicate
   - who (norm-subjects)
   - does what (the action theme)
   - in what circumstances (conditions of application), and
   - the nature of their guidance (the mood).

Basic features of normative systems. Herbert Hart, among other philosophers, clarified for the law under what conditions normative systems exist [49]. However, Hart's remarks can be somewhat generalized and can help us identify some very basic features of almost any other normative domain:

norm recognition and hierarchies: it is possible to state or identify criteria for normative systems to establish whether norms belong to them; also, normative systems can assign to their norms a different ranking status and organize them in hierarchies;

norm application: it is possible to state or identify criteria for normative systems to correctly apply their norms to concrete cases;

norm change: it is possible to state or identify criteria for changing normative systems.

However, these questions, rather than solving the fundamental definitional problems regarding the nature of nMAS, open new questions and research perspectives. Following earlier discussions resulting from previous NorMAS meetings [20, 23, 18, 19], we can formulate three definitions of norms and nMAS, which we discuss in the remainder of this section.

3.1 The Social Definition

In the social sciences norms are often seen as customary rules of behaviour that coordinate the interactions in groups and societies [11]. Many norms are enacted in following simple behavioural rules, like not to bump into other people on the street. Other norms are more complex, triggering sanctions against individuals departing from such rules. The former are sometimes called *conventions* and the latter *norms*.

The widely shared view across the social sciences and, in fact, also law and philosophy, is that norms are exogenous variables constraining individual behaviour. The most obvious form of enforcing norms is by threat of social disapproval or punishment in the event of norm violation. Less obviously, norms can also be internalised by socialising individuals to them. In particular, sociology and pedagogy have studied the socialisation of individuals into society, for instance in the educational system. Norms can also arise from purely voluntary coordination of interaction. Economics in particular has studied the function of norms in providing efficient market outcomes. The justification for this kind of research is that "norms coordinate expectations which reduce transaction costs in interactions that possess multiple equilibria" [98]. If a norm supports the rise of a focal solution to a coordination problem it can be considered a form of social capital.

Against this background, in the social sciences norms typically are viewed as means to define these relationships, and portraying as well as regulating the behaviour of the members of a group or society as well as the society as a whole [71].

Research on norms in the social sciences tends to focus on the **social function** of norms (see the ideas in [69] for example), the **social impact** of norms (e.g. [74]), or mechanisms leading to the **emergence and creation** of norms (e.g. [33]).

With regard to the aspect of the social function, norms are often concerned with the behaviour that members of a social group should (or should not) perform regardless of the possible consequences (also referred to as obligations). Furthermore they deal with the expectations resulting from the anticipation of the other actors in the system with regard to this behaviour.

This idea is formalized by Tuomela and Bonnevier-Tuomela [94], for example, who specify a social norm as a norm having the following form: "An [individual] of the kind F in group G ought to perform task T in situation C"[3] and thereby highlight the four major aspects of norms in social science: (i) that individuals can be different in their behaviour and might perform different roles in a society (i.e., F), (ii) the importance of the relationship of the individual to a group or society (i.e., G), (iii) the obligation to perform (and indirectly not to perform) certain tasks (i.e., T) as a result of a norm, as well as (iv) the notion of context-dependency (i.e., C) of norms, i.e., that norms might only be valid in particular situations.

Another important aspect Tuomela and Bonnevier-Tuomela [94] stress is the distinction between so-called "r-norms" and "s-norms". The former are norms that are "created by an authority or body of agents authorized to represent the group (this body can also be the entire group)" [94]. This authority might, for example, be a governing or legislative body, the operator of a platform, or a chosen leader. S-norms are norms emerging as a feature of a (social) normative context, i.e., the result of mutual beliefs about the way a particular situation should be handled, general codes of conduct or conventions. Thus, in contrast to r-norms, the concept of s-norm is not based on rules defined by any authority, but tends to highlight the social aspect of norms. The inclusion of s-norms in this distinction is a rather important feature of the social sciences. Based on the idea of an s-norm, in addition to the notion of norms being guidance on what actors are ought or expected to do, in the social sciences norms are also viewed as an information source of what is perceived to be "normal" in a group or population [93]. In the social sciences, this "normal" behaviour is explained as emerging as a general pattern of behaviour by the agents of a MAS making choices without any centralized planning [1].

With regard to research on the social impact of norms, the focus is placed on the utility provided to or taken away from the actors involved in an interaction. Utility is defined here as the relative (both positive and negative) satisfaction achieved by the actors. This utility can be either internal, such as emotion levels and energy, or external, in the case of money, etc. Social impact research analyses the effect of utilities of the different stakeholders in a society resulting from specific norms as well as on the society as a whole [73].

Besides these works on the social function, the social impact as well as the emergence and creation of norms, in the social sciences (in particular in philosophy) further important scientific contributions have been made dealing with normative positions [90]. The two normative positions we have talked about in this chapter so far are permissions and obligations.

[3] Despite the strong emphasis of this statement on obligations, i.e., what one ought to do, [94] in the course of their definition broadens it to include permission (i.e, what one may do) as well.

Although these describe norms to a large extent, within the world of legal theory several more positions can be found, including power, duty, right, liability, disability, claim and immunity for example [51]. The position of *power* is of particular interest as a restraint on the (physical) power of agents. In social science, power can have two different forms, which are both described by [67]: (i) **legal power**, and (ii) **physical power**.

Whereas the former specifies whether an actor is "empowered" to perform a certain action in a legal sense, the latter establishes whether he is physically able to carry out the actions necessary to exercise his legal power [55]. Whereas this distinction between physical and normative power is relatively easy to make, it is important not to confuse normative power and the term "permissions" explained earlier as well as to understand how the different concepts are linked in a system and to define when which agent has which power or permission well.

▶ Problem 1. How to define the relation between physical and legal power as well as permission in a system and how to define at which point of time or in which state an agent has which powers (both physical and legal) as well as permissions?

To explain the difference between normative power and permission [67, p. 409] gives an illustrative example:

> [. . .] consider the case of a priest of a certain religion who does not have the permission, according to instructions issued by the ecclesiastical authorities, to marry two people, only one of whom is of that religion, unless they both promise to bring up the children in that religion. He may [in his function as priest] nevertheless have the power to marry the couple even in the absence of such a promise, in the sense that if he goes ahead and performs the ceremony, it still counts as a valid act of marriage under the rules of the same church even though the priest may be subject to reprimand or more severe penalty for having performed it.

What is important to note besides this difference between permission and legal power, is the difference in legal and physical power. Thus, despite the priest having the legal power, physical power does not follow automatically. Thus, the priest might be incapacitated by being sick for example, and as a consequence is physically not able to perform the normative action of marrying two people, despite having the legal power to do so. In the opposite direction the two notions of power also do not have a coercive relation: having the practical possibility and power to act does not necessarily imply that a legal power is also existing. To give an example, someone might not have the legal power to conduct a marriage, despite being physically able to do so (e.g., by not being physically incapacitated and knowing the procedure, etc.).

As a result of this view on norms in the social sciences, some common features can be found, such as the idea that norms are rules that define what is considered right or wrong by the majority of a population. The definition of the majority of the population and the particular features of the population (i.e., its size or composition) therefore are domain dependent [52]. Norms furthermore spread; that implies that they are acquired and communicated through means that can be direct (e.g., communication between the actors in a system) or indirect (such as adaption and learning processes by the actors).

Taking these common basic concepts pertaining to norms, a first social science inspired definition of nMAS could be made as follows:

▶ **Definition 1** (nMAS – The Social Definition)**.** A normative multi-agent system is a MAS governed by restrictions on patterns of behaviour of the agents in the system that are actively or passively transmitted and have a social function and impact. These patterns are sometimes represented as actions to be performed [4]. In principle they dictate what actions (or outcomes) are permitted, empowered, prohibited or obligatory under a given set of conditions as well as specify effects of complying or not complying with the norms [6]. They can emerge within the nMAS or can be created a priori by the nMAS designer. Norms should be contextual, prescriptive and followable [87].

The notion of **contextuality** refers to norms only being applicable in a specific context and not in general. To give an example: despite being generally valid, rules for driving a car are only applicable in traffic settings and these traffic settings provide the context for the application of the respective norms. This notion of contextuality also sets a scope to the time a norm is in force and consequently requires the definition of the activation, as well as existence and deactivation of a norm.

▶ Problem 2. How to specify the context(s) in which norms apply and do not apply and how to ensure that agents can determine which context they are acting in?

That norms should be **prescriptive** refers to the fact that "those who are knowledgeable of a rule also know that they can be held accountable if they break it" [87, p. 41]. Thus, norms specify what actions an actor "must not" perform (prohibition), "is empowered to perform" (legal power), "must perform" (obligation) or may perform ("permission") if the actors want to avoid sanctions for non-compliance with the norms being imposed on them. What is important to note is that the **prescriptive** criterion (i.e. that agents knowledgeable of norms should know that they can be held accountable for breaking them) does not necessarily imply the opposite to be true, i.e., that one can only be held accountable if one knew about a norm. However it poses the problem that as a system designer one might want to account for this difference.

▶ Problem 3. How to deal with a lack of normative awareness and if it is being considered, how to check the lack of normative awareness if an agent's knowledge base is not accessible?

Finally, rules should be *followable* in the sense that it should be physically possible for the actors in a system to both perform and not to perform prohibited, obligatory or permitted actions, as well as to obtain the legal power to do so [5]. This poses another problem:

▶ Problem 4. How to ensure that norms are followable for agents at any state/point of time, or phrased differently, how to ensure that two (sets of) norms do not conflict with one another (e.g. by allowing and not allowing an action at the same time) [61]?

In addition to this social science inspired definition, the NorMAS community has discussed definitions emphasising further features of MAS in previous NorMAS meetings. These definitions are the focus of the next two sections.

3.2 The Norm-change Definition

The norm-change definition runs as follows:

▶ **Definition 2** (nMAS – The Norm-change Definition [20])**.** "A normative multi-agent system is a multi-agent system together with normative systems in which agents on the one hand can decide whether to follow the explicitly represented norms, and on the other the normative systems specify how and to what extent the agents can modify the norms."

Hence, under this view norm dynamics plays a crucial role, which opens a number of related questions. In [18, 19] three guidelines for developing nMAS were derived from Definition 2.

▶ Guideline 1. Motivate which definition of nMAS is used and so explain which of the following choices should be adopted:

1. norms must be explicitly represented in agents and in the system in a declarative way (the 'strong' interpretation), or
2. norms must be explicitly represented in the overall system specification (the 'weak' interpretation), or
3. none of the above interpretations should be adopted.

It was argued in [18, 19] that the strong interpretation must be preferred whenever we want to prevent a too generic notion of norm. In fact, we should avoid trivializing this notion, which is a risk when we see any specification requirement as a norm that the system has to comply with. However, the weak interpretation is sometimes more suitable to address the following problems:

▶ Problem 5 (Norm compliance). How we can check whether a system complies with relevant norms applicable to it?

▶ Problem 6 (Norm implementation). How can we design a system such that it complies with a given set of norms?

Problems 5 and 6 amount to studying the concept of compliance at runtime and by design, so they can be meaningful also when we adopt for nMAS the strong reading. Notice that both problems require, in general, the articulation of the conditions under which the relevant norms are part of the normative system at hand and can correctly be triggered and applied: these issues correspond, as we said, to the first two basic features of any normative domain.

Finally, notice that any attempt to address Definition 2 and Guideline 1 also requires the preliminary clarification of the types of norm we need to embed within nMAS. This clarification is very relevant, since different types of norms sometimes correspond to different formal (and logical) models and so distinct options may differently affect the choice between the strong and weak interpretations for the explicit representation of norms within nMAS. We have already mentioned von Wright's norm classification. In [38] an extensive analysis of requirements for representing norms has been proposed for the law. Consider the following aspects, which contribute to classifying norms and which can be extended to other normative domains besides the law[4]:

Temporal properties [42]. Norms can be qualified by temporal properties, such as:

1. the time when the norm is in force;
2. the time when the norm can produce effects; and
3. the time when the normative effects hold.

Normative effects. There are many normative effects that follow from applying norms, such as obligations, permissions, prohibitions and also more articulated effects such as those introduced for the law, for example, by Hohfeld (see [86]). Below is a rather comprehensive classification of normative effects [84]:

Evaluative, which indicate that something is good or bad, is a value to be optimised or an evil to be minimised. Consider, for example, "Human dignity is valuable", "Participation ought to be promoted";

[4] Gordon *et al.* [38] also study whether existing rule interchange languages for the legal domain are expressive enough to fully model all the features listed below (and those recalled below, Section 3.3): RuleML, SBVR, SWRL, RIF, and LKIF.

Qualificatory, which ascribe a normative quality to a person or an object. Consider, for example, "x is a citizen";

Definitional, which specify the meaning of a term. Consider, for example, "Tolling agreement means any agreement to put a specified amount of raw material per period through a particular processing facility";

Deontic, which, typically, impose the obligation or confer the permission to do a certain action. For example, "x has the obligation to do A";

Potestative, which attribute powers. For example, "A worker has the power to terminate his work contract";

Evidentiary, which establish the conclusion to be drawn from certain evidence. Consider, for example, when the sentence "It is presumed that dismissal was discriminatory" is concluded from some piece of evidence;

Existential, which indicate the beginning or the termination of the existence of a normative entity. For example, "The company ceases to exist";

Norm-concerning effects, which state the modifications of norms; for the law: abrogation, repeal, substitution, and so on.

Definition 2 raises two other fundamental research questions, which concern, respectively, whether agents in nMAS can violate norms and how and why norms can be changed in nMAS.

Hence, the second guideline follows from the fact that agents, insofar as they are supposed to be autonomous, can decide whether to follow the norms. Indeed, it would be misleading for the specification of a nMAS to disregard "the distinction between normative behavior (as it *should be*) and actual behavior (as it *is*)" [70, preface]. Avoiding making this distinction is misleading for three reasons: if any "illegal behavior is just ruled out by specification" then

- we are unable to "specify what should happen if such illegal but possible behaviors occurs!" [70, preface];
- we fail to adopt a meaningful concept of norm, since philosophers and deontic logicians mostly agree that genuine norms (and their effects) can be violated (as an extreme example, it does not make any sense to say that $A \wedge \neg A$ is forbidden); and
- agents cannot violate norms and so we do not model one important aspect of agents' autonomy in normative agent architectures and decision making [29].

Accordingly, a theoretically sound definition of nMAS would assume that agents can violate norms, so if a norm is a kind of constraint, the question immediately is raised what is special about them. While hard constraints are restricted to preventative control systems in which violations are impossible, soft constraints are used in detective control systems where violations can be detected. This justifies the following guideline:

▶ Guideline 2. Make explicit why your norms are a kind of (soft) constraint that deserve special analysis.

A typical illustration of how normative soft constraints work is the situation in which one can enter a train without a ticket, but may be checked and sanctioned. In contrast, a supposed illustration of a hard-constraint implementation of a norm is the situation in which one cannot enter a metro station without a ticket [18, 19]. However, a closer inspection of the metro example shows that, strictly speaking, this does not correspond to a genuine case where violations are made impossible, but only where they are normally and in most cases prevented to occur: indeed, one could, for instance, break the metro barriers and travel without any ticket. When violations are impossible in any conceivable way, the concept of norm does not make much sense.

On the other hand, if the norms are represented as soft constraints, then the problem is to check if the process of monitoring violations is correctly managed, since this detective control is the result of actions of agents and therefore subject to errors and influenceable by actions of other agents. For example, it may be the case that violations are not detected often enough, there are conflicting obligations in the normative system, that agents are able to block the sanction or update the normative system, etc.

More information on compliance and norm violation is given in Chapter 5.

The third guideline follows from the fact that norms can be changed by the agents or by the system. Suppose, for example, that a nMAS must be checked against some legal system. As is well-known, one of the peculiar features of the law is that it necessarily takes the form of a dynamic normative system [56]. Hence, the life-cycle of agents must be described with respect to a changing set of norms. Similar considerations can be applied to many other normative domains, as we argued that it is possible to state or identify criteria for changing many types of normative system:

▶ Guideline 3 (Norm change). Explain why and how norms can be changed at runtime.

In general, in nMAS a norm can be made by an agent, as legislators do in a legal system, or there can be an algorithm that observes agent behavior, and suggests a norm when it observes a pattern. The agents can vote on the acceptance of the norm [62]. Likewise, if the system observes that a norm is often violated, then apparently the norm does not work as desired, and it undermines the trust of the agents in the normative system, so the system can suggest that the agents can vote whether to retract or change the norm.

More on norm change can be found in Chapter 2 and 6.

3.3 The Mechanism Design Definition

The mechanism design definition of nMAS runs as follows:

▶ **Definition 3** (nMAS – The Mechanism Design Definition [23]). "A normative multi-agent system is a multi-agent system organized by means of mechanisms to represent, communicate, distribute, detect, create, modify, and enforce norms, and mechanisms to deliberate about norms and detect norm violation and fulfilment."

Norms are rules used to guide, control, or regulate desired system behavior. An nMAS system is a self-organizing system, and norms can be violated. Boella *et al.* [18, 19] derive two guidelines from this definition, which focus on the role of norms, either as a mechanism or as part of a larger institution or organization.

▶ Guideline 4. Discuss the use and role of norms always as a mechanism in a game-theoretic setting.

▶ Guideline 5. Clarify the role of norms in your system as part of an organization or institution.

Both these guidelines lead to handling more specific research problems:

▶ Problem 7 (Norms and games). A relevant problem has to do with investigating the connection between games and norms. In fact, games can explain that norms should satisfy various properties and also the role of various kinds of norms in a system. For example, Bulygin [25] explains why permissive norms are needed in normative systems using his "Rex, Minister and Subject" game. Boella and van der Torre introduce a game theoretic approach to normative systems [22] to study violation games, institutionalized games, negotiation games, norm creation games, and control games. Norms should satisfy various properties to be effective as a mechanism to obtain desirable behavior. For example, the system should not sanction without reason, and sanctions should not be too mild or too harsh.

▶ Problem 8 (Norms and their functions). Another research problem consists of providing a clarification of the different role that norms can play in agents' societies. As we mentioned in the previous section, norms may have a number of different effects, and so they do not only impose duties and establish sanctions for their violation. Hence, in a game-theoretic perspective they not only have a preventive character, but, for instance, also provide incentives. However, moral incentives are very different from financial or legal incentives. For example, the number of violations may *increase* when financial sanctions are imposed, because the moral incentive to comply with the norm is destroyed [59, 35, p. 18–20]. Moreover, norms and trust have been discussed to analyze backward induction (which is an iterative process in game theory for solving finite extensive form or sequential games) [53].

▶ Problem 9 (Norms and organizational design). How do norms contribute to design agents' organizations? Norms are addressed to roles played by agents [21] and used to model organizations as first class citizens in multi-agent systems. In particular, constitutive norms are used to assign powers to agents playing roles inside the organization. Such powers allow the issuing of commands to other agents, making formal communications and restructuring the organization itself, for example, by managing the assignment of agents to roles. Moreover, normative systems also allow modeling the structure of an organization and not only the interdependencies among the agents of an organization. Legal institutions are defined by Ruiter [85] as "systems of [regulative and constitutive] rules that provide frameworks for social action within larger rule-governed settings". They are "relatively independent institutional legal orders within the comprehensive legal orders".

Hence, Definition 3, Guideline 4 and 5 and the related research problems require, too, additional clarification on the types of norm we need to model for nMAS. Also in this second perspective, many of Gordon *et al.*'s requirements [38] for specifically representing norms in the law are directly applicable to modeling roles, organizations and institutions. Important requirements for legal rule languages from the field of AI and Law include the following:

Isomorphism [8]. To ease validation and maintenance, there should be a one-to-one correspondence between the rules in the formal model and the units of natural language text which express the rules in the original normative sources, such as sections of legislation. This entails, for example, that a general rule and separately stated exceptions, in different sections of a statute, should not be converged into a single rule in the formal model.

Rule semantics. Any language for modeling norms should be based on a precise and rigorous semantics, which allows for correctly computing the effects that should follow from a set of norms.

Defeasibility [37, 79, 86]. When the antecedent of a norm is satisfied by the facts of a case, the conclusion of the rule presumably holds, but is not necessarily true. The defeasibility of norms breaks down in the law into the following issues:

Conflicts [79]. Rules can conflict, namely, they may lead to incompatible legal effects. Conceptually, conflicts can be of different types, according to whether two conflicting rules

- are such that one is an exception of the other (i.e., one is more specific than the other);
- have a different ranking status; or
- have been enacted at different times;

Accordingly, rule conflicts can be resolved using principles about rule priorities, such as:

- *lex specialis*, which gives priority to the more specific rules (the exceptions);
- *lex superior*, which gives priority to the rule from the higher authority (see 'Authority' above); and
- *lex posterior*, which gives priority to the rule enacted later (see 'Temporal parameters' above).

Exclusionary norms [79, 86, 37]. Some norms provide one way to explicitly undercut other rules, namely, to make them inapplicable.

Contributory reasons or factors [86]. It is not always possible to formulate precise rules, even defeasible ones, for aggregating the factors relevant for resolving a normative issue. Consider, for example, "The educational value of a work needs to be taken into consideration when evaluating whether the work is covered by the copyright doctrine of fair use."

Norm validity [42]. Norms can be invalid or become invalid. Deleting invalid norms is not an option when it is necessary to reason retroactively with norms which were valid at various times over a course of events. For instance, in the law:

1. The *annulment* of a norm is usually seen as a kind of repeal which invalidates the norm and removes it from the legal system as if it had never been enacted. The effect of an annulment applies *ex tunc*: annulled norms are prevented from producing any legal effects, also for past events.
2. An *abrogation* on the other hand operates *ex nunc*: The norm continues to apply for events which occurred before the rule was abrogated.

Legal procedures. Norms not only regulate the procedures for resolving normative conflicts (see above), but also for arguing or reasoning about whether or not some action or state complies with other norms [40]. In particular, norms are required for procedures which

1. regulate how to detect violations of the law (for the law is not sufficient that a violation is detected, but how this happens: illegal detection may lead to void effects); or
2. determine the normative effects triggered by norm violations, such as reparative obligations, which are meant to repair or compensate violations (the law distinguishes different types of sanction that can be applied to the same wrongdoings).

Persistence of normative effects [43]. Some normative effects persist over time unless some other and subsequent event terminates them. For example: "If one causes damage, one has to provide compensation". Other effects hold on the condition and only while the antecedent conditions of the rules hold. For example: "If one is in a public building, one is forbidden to smoke".

Values [7]. Usually, norms promote some underlying values or goals. Modeling norms sometimes needs to support the representation of these *values* and *value preferences*, which can play also the role of meta-criteria for solving norm conflicts. (Given two conflicting norms r_1 and r_2, value v_1, promoted by r_1, is preferred to value v_2, promoted by r_2, and so r_1 overrides r_2.)

Some of these requirements, as they are formulated above (they are recalled from [38]), are peculiar of the legal domain only or, at least, of any "codified" system of norms (consider, e.g., the "Isomorphism" requirement). However, almost all can be easily adjusted to fit many other normative domains. Besides some very general requirements, such as "Defeasibility" and "Rule semantics"—which correspond to aspects widely acknowledged for most normative domains—the other requirements are also important for nMAS. Consider, for instance, the problem of the temporal persistence of norm effects, the fact that norms can be valid only under some conditions, or the role of exclusionary reasons.

4 Norms, Policies, Laws and Conventions

We previously noted that currently a multitude of views on the definition of the terms "norm" and "nMAS" exists. In that perspective, we highlighted some of these definitions by identifying generally established characteristics of norms and gave some guidelines on how to distinguish norms from other concepts. This section now has a closer look at related terms such as "policies", "laws" and "conventions" and points out differences as well as similarities with norms. In detail, this section recalls the definitions on norms given earlier (especially in Section 3.1) and gives a short overview of the meaning of the term "policy" in the research domains from which they have been borrowed, namely social science, political science, economics and law. Afterwards we relate the gained information to the concepts of "laws" and "conventions".

As pointed out in Section 3.1, in the social sciences norms are often seen as customary rules of behaviour that coordinate the interactions in groups and societies [11] as well as give advice on how individuals should behave. Norms therefore often have some social goal, such as the reduction of transaction costs in coordination and collaboration situations.

Not all norms are, however, geared towards efficiency and even if norms may fulfill important social functions they cannot be explained solely on the basis of this function. In fact, even if a means-end relationship between a norm and a social goal exists, this may not be the reason the norm came to be [30, p. 322]. In addition, many norms may persist even if they are inefficient or contested. In particular philosophy of law has studied how new norms can be justified [45, p. 631]. This is relevant because in post-traditional societies legislation has become a key feature of the integration of society and the procedures bringing about new laws themselves represent institutionalised norms. This has received attention from different disciplines because institutionalising norms departs from seeing norms purely as interactions of individuals. Legal science, and to a certain extent also political science, considers norms to be hierarchically differentiated and policies are part of this hierarchy.

In social science, policies are understood as instruments to implement norms [95], and are used by policy makers to encourage society to adopt certain norms[5]. The constitution of the state embodies basic legal norms and these constrain politics and policies. On the other hand policies also represent emerging social norms, which may over time come to transform the governing body. According to Lowi's dictum that policies determine politics [64] political actors seek to implement new norms through policies, eventually changing politics too. In doing so they draw on formal (e.g., the law) and informal (e.g., social norms) institutions. Over time even the hard-wired norms of the governing body may be changed despite the hierarchical relationship between governing body and policies. Thus, existing legal norms and new social norms emerging in the political process can be said to form a recursive cycle.

From the above statements one can also infer information about the differences between laws and conventions with respect to norms. Laws are typically forming a system of rules and guidelines which are enforced through a judicial system to govern behavior, wherever possible. What is important about laws is that they are made by some authority of the

[5] Policies are not the only way norms can be implemented, but they can also emerge as generally accepted social behaviour. Thus, policies only cover a certain part of norms and are generally understood to be preceded by them.

system (e.g., a government) and are explicitly written down and made publicly available.

In contrast to laws, conventions are a set of agreed, stipulated or generally accepted standards, social norms or criteria, often taking the form of a custom, which are not necessarily written down, but are often transmitted through other (informal) means. Although a (social) convention is a regularity widely observed by some group of agents, the reverse—that every regularity is a convention—is not always true. To give an example of this, we all eat, sleep, and breathe, yet these are not conventions [81].

In contrast, the fact that everyone in the UK drives on the left-hand side of the road rather than the right is a convention [60], which started from an information behaviour and was made formal by means of laws later on [66]. With respect to norms, conventions are often seen as simpler rules, with limited or no sanctioning attached to them, whereas norms tend to be more complex and often have some form of enforcement idea attached to them.

5 Specific Developments, Open Questions and Trades with Other Disciplines

In the previous sections we mentioned several research issues and general aspects of normativity in MAS. We will recall in the remainder some of them by suggesting research lines for nMAS that, though important, have not yet received sufficient attention in the MAS community:

- The concept of moral agency, especially in a cognitive perspective;
- The concept of group norm;
- The connection between argumentation and norms;
- Conceptual vagueness and fuzziness of legal norms.

For each of them we will indicate some open problems and research perspectives. We will finally identify possible links among them in terms of what benefits each research line can offer to the others.

5.1 Moral Agency and the Mental Side of Normativity

Although the concepts of morality and moral agency have been extensively studied in social philosophy and in the social sciences, they have been so far less studied in the areas of multi-agent systems and normative multi-agent systems. Some works have been proposed on the extension of the BDI (Belief, Desire, Intention) model with normative concepts such as the concept of obligation [24, 41], but none of them have really focused on the integration of moral aspects into the architecture of a cognitive agent.

Developing formal models of cognitive agents integrating a moral dimension is a promising research avenue for these two areas. Indeed, as shown by social scientists [34, 32], decisions of human agents are often affected by moral sentiments and moral concerns (e.g. concerns for fairness or equity). Therefore, to take the presence of moral attitudes into account becomes extremely important when developing formal and computational models of artificial agents which are expected to interact with human agents (e.g., trading agents, recommender systems, and tutoring agents).

A model of moral agency should be able to explain the two different origins of an agent's motivations. Some of them originate from the agents' desires. A desire can be conceived as an agent's attitude which consists of an anticipatory mental representation of a pleasant state of affairs (the representational dimension of a desire) that motivates the agent to achieve it (the motivational dimension of a desire). In this perspective, the motivational dimension of an agent's desire is realized through its representational dimension. For example when an agent desires to be at the Japanese restaurant eating sushi, he imagines himself eating sushi at the Japanese restaurant and this representation gives him pleasure. This pleasant representation motivates him to go to the Japanese restaurant in order to eat sushi.

Agents are motivated not only by their desires but also by their moral values. Moral values, and more generally moral attitudes (ideals, standards, etc.), originate from an agent's capability of discerning what from his point of view is (morally) good from what is (morally) bad. If an agent has a certain ideal φ, then he thinks that the realization of the state of affairs φ ought to be promoted because φ is good in itself.[6]

Morality is a composite cognitive phenomenon. Aspects of this phenomenon that deserve to be studied and to be implemented in the computational architecture of a cognitive agent are, for example:

- the concept of moral choice, i.e., how the utility of a given decision option for an agent is determined by both the agent's desires and the agent's moral values [48]; and
- moral emotions such as guilt, moral pride and reproach [46].

5.2 Group Norms

Group norms address groups of individuals, affecting their joint behaviours, which arises in many situations; consider, e.g., an obligation on the sales team to meet once a week, a prohibition on gatherings of more than x people, or a permission for a group visit to a building. This section makes a case for the importance of representing and processing such norms, raises issues which should be investigated, and sketches how research on group norms could connect coordination mechanisms and normative reasoning.

5.2.1 Description of the Topic

Group norms can be seen as those norms addressing collections of individuals and affecting their *joint behaviours*. This is a specific interpretation of group norms, as there are other kinds of regulations aimed at groups of people (as opposed to individuals), but these do not concern joint behaviours.

For instance, a norm establishing that "non-EU nationals must join queue Q", although addressing a group, does not place constraints on individuals' joint behaviours, that is, non-EU nationals do not have to agree on how to act collectively. On the other hand, norms such as "at most 5 people are allowed in room R", "procedure P can only be carried out by a team of 3 people" and "gatherings of more than 3 people are forbidden", all influence the collective behaviour of those whom the norm addresses. Individuals will need to agree on what to do and when, in order to abide by norms whilst striving to achieve their goals.

Coordination is essential for agents to adequately process such group norms. Although there are many ways in which autonomous entities may interact, ranging from a simple Contract Net protocol [92], to auctions, negotiations and argumentation, the outcome of this activity is an agreed joint plan of action, listing what each party will do and when. Research on group norms will ultimately connect coordination mechanisms (and group deliberation), norm representation and individual (normative) reasoning.

[6] A similar distinction has also been made by philosophers. See [89] for a recent philosophical analysis of how an agent may want something without desiring it and the problem of reasons for acting based on values and independent from desires.

5.2.2 Background

Work on collective agency (e.g., [26, 27, 76]) and collective obligations (e.g., [44]) have addressed similar concerns. These approaches represent norms over actions, also establishing a group of agents to whom the norms apply. Some approaches regard a group norm as a shorthand for a norm which applies to all/some members of the group (e.g., [27]), whereas other approaches (e.g., [44]) regard a group norm (more specifically, a collective obligation) as a shared complex action requiring individual contributions (i.e., simpler actions) from those individuals of the group.

Research on joint action and coalitions (e.g., [68, 83, 50]) is also relevant as it looks into individual deliberation when coordination is required. Work exploring aspects of delegation (e.g., [72, 63]) sheds light on how norms can be transfered among individuals and groups.

The concept of *roles* used in work on societies, electronic institutions and organisations, also provides means to address collections of individuals. We note that norms addressed at roles are a useful shorthand for specialised norms addressed at individuals, yet the usual definition of role norms does not aim to influence the joint behaviour of individuals.

A flexible means to specify groups of agents is needed, in order to capture the class of group norms we have in mind. Moreover, group specifications should be compact, allowing for an intensional definition of those belonging to the group. For instance, being able to represent a group with at most 3 workers (from a potentially larger universe of workers) is useful, as some of the norms we want to capture require flexibility, compactness, and precision.

5.2.3 Current Understanding

Group norms ultimately influence collective behaviour. Individuals must agree on a joint plan of action to achieve certain goals or to avoid some situations. In order to do this, individuals must be aware of *i)* the norms in place; *ii)* their membership of groups (and hence whether any group norms apply to them); *iii)* how their behaviour conforms or not to any applicable norms (and whether there are incentives to abide or not by the norms).

Any account of group norms for nMAS needs to devise an expressive, compact and precise specification language. Such an account also requires developing mechanisms to process these norms, in order to check their applicability (to individuals), and to endow individuals with means to factor these norms in when deciding on rational individual behaviours.

Individual choices also involve collective deliberation on joint courses of action. Indeed, group norms address collections of individuals; hence, even though individual action are ultimately the ones performed, in groups some interconnected actions can together "count as" group actions. For instance, when 4 people lift a square table the individual actions are to lift each of the 4 corners. The choice of which individual action(s) each member of the group chooses, taking into account any norms in place, is thus very important.

5.2.4 Questions, Challenges & Expected Lines of Research

The concept of group norms has not been adequately investigated. These norms, however, do exist in reality, and they influence individuals and, ultimately, groups. Phenomena arising from such group norms and their processing is of great importance to policy makers, and designers of agents and autonomous systems.

It seems to us that there are two strands for a promising analysis of group norms. On the one hand, group norms can be defined in suitable operational semantics, which describe the meaning of norms as influencing a collective agreement over a joint plan. On the other hand, we can investigate model-theoretic aspects of group norms as a deontic logic for coalitions.

We notice the potential for strategic reasoning, whereby individuals may form groups so as to avoid norms or indeed have norms on them. This is the case, for instance, of agents joining or forming a group because a permission is in place for the group. Likewise, individuals may avoid being part of a group because some unwanted norm is in place. Strategic formation and dissolution of groups thus allows agents to behave in a norm-compliant fashion but avoiding penalties associated with norm violation.

5.3 Argumentation and Norms

Norms and argumentation are two research areas which are becoming more and more connected over the last decade, in the legal field, in knowledge representation, ethics, or linguistics, and most recently, in agreement technologies in computer science[7]. Norms are used to set the space of legal agreements (or commitments) and argumentation is used to choose among the possible agreements [12]. Moreover, we may consider that norms set not only the scope of possible legal agreements, but also the way we can choose among these possible agreements.

5.3.1 Background

In law, Bench-Capon et al. [9] present how argumentation theory has been used in legal reasoning. For instance, legal disputes arise out of a disagreement between two parties and may be resolved by presenting arguments in favor of each party's position. These arguments are proposed to a judging entity, who will justify the choice of the arguments he accepts with an argument of his own, with the aim to convince the public. The common conclusion shared by such works is that argumentation has the potential to become a useful tool for people working in the legal field. Even if a common answer from lawyers when they are asked about what argumentation theory can do for them is that it can be used to deduce the consequences from a set of facts and legal rules, and to detect possible conflicts, there is much more to argumentation. Following the example proposed by Bench-Capon et al. [9], a case is not a mere set of facts, but it can be seen as a story told by a client to his lawyer. The first thing the lawyer does is to interpret this story in a particular legal context. The lawyer can interpret the story in several different ways, and each interpretation will require further facts to be obtained. Then the lawyer has to select one of the possible interpretations, she has to provide arguments to persuade the judging entity of the client's position, and to rebut any further objection. The major topics that emerge as relevant in norms and argumentation include, among others, case based reasoning [3, 82], arguing about conflicts and defeasibility in rule based systems [91, 77, 80], dialogues and dialectics [36], argument schemes [39, 10], and arguing about the successfulness of the attacks [31, 78].

5.3.2 Current Understanding

Existing works (see Section 5.3.1) on norms and argumentation can be categorized into two different classes, namely (i) arguing about norms, and (ii) norms about argumentation (for a review of the literature, see [75]). The former includes the greater part of existing works in

[7] Agreement technologies refer to computer systems in which autonomous software agents negotiate with one another in order to come to mutually acceptable agreements.

the area of norms and argumentation, such as approaches which aim at resolving conflicts and dilemmas, looking in particular at how norms interact with other norms, arguing about norm interpretation and dynamics, arguing about norm adoption, acceptance and generation, representing norm negotiation, and arguing about contracts. In spite of all the existing literature on these topics, several challenges have still to be addressed and resolved. For instance, the introduction of frameworks where the individuals can discuss the merits and the effects of the norms to be adopted in the society, and the proposal of preference models allowing the detection and reasoning about norm interactions are fundamental steps to approaching the two research areas. The latter class of research includes a smaller set of existing works, and it aims at addressing the challenges about dialogue and debate protocols, reasoning about epistemic norms, and enforcement models of the burden of proof. For instance, the introduction of new techniques to verify whether a virtual agent complies with an epistemic norm, and the development of tools able to support judging entities and lawyers to enforce the burden of proof are further challenges for agreement technologies. Finally, besides norms about argumentation and arguing about norms, direct formal relations between deontic logic – in particular input/output logic – and abstract argumentation have been considered [14, 15, 16], leading to a number of additional challenges.

5.3.3 Questions, Challenges & Expected Lines of Research

Open challenges in this field are connected with the following research topics:

- Arguing about norms:
  1. *societal modelling and control*: where individuals debate about the merits of norms and their effects;
  2. *societal modelling and control*: where individuals persuade others about the utility of norm adoption;
  3. *constitutive norms*: more than two agents performing ontology alignment;
  4. *constitutive norms*: avoiding the need for the central ontology mapping repository;
  5. *regulative norms*: considering norms in practical reasoning;
  6. *normative constraints*: complex normative reasoning for deadlines, norm violation, norm fulfillment;
  7. *normative constraints*: using argument schemes to reason about norms being or not in force;
  8. *normative conflict*: developing reacher preference models and logics for reasoning about norm interaction;
  9. *practical reasoning*: integration of domain specific knowledge and inference using argument schemes;
  10. *practical reasoning*: new reasoning heuristics;
  11. *monitoring norms*: identifying argument schemes (in the sense of [39, 10]), which reason about uncertainty;
  12. *monitoring norms*: weighting up conflicting uncertain evidence.

- Norms about argumentation:
  1. *dialogue*: interplay between dialectical norms (those norms that specifically govern dialogues and the exchange of arguments) and procedural norms (see Section 3.3);
  2. *dialogue*: modelling dialogues where several norms regulate a dialogue;
  3. *burden of proof*: tools for supporting people in the legal field to verify proof standards.

We claim that these future challenges have to be addressed both from the theoretical and from the design point of view. We need to define new innovative models using argumentation theory in legal reasoning or applying norms in the argumentation, and tools that really implement the proposed models and theories in order to not leave such theories at the pure abstract level. Consider, for instance, the application of norms to the argumentation process. Proof standards and burden of proof are key examples of norms applied to argumentation. While several burden of proof have been theoretically defined in the literature, such as *burden of claiming* and *burden of questioning*, a challenge to be addressed consists in the development of tools to support the humans operating in the legal field. The idea is to start from systems like Carneades[8], which already provide a tool for modeling legal dialogues, and improve them to support the interaction with humans. For instance, a judge can use such a tool to look at the argumentation framework which models the trial, and she will be able to detect the possible "irregularities" with respect to the burden of proof. Moreover, the tool should provide the judge with a summary of the argumentation framework representing the trial's arguments. A possible way for formalizing such a summary may be to use argumentation patterns [96], where meaningful sets of arguments together with their attack relations are identified and treated as a unique piece of information with a precise meaning. The same tool can be used by lawyers to detect the possible weak points of a deliberation. In this way, the lawyer will know exactly the weak points to appeal. The development of such a tool based on burden of proof is a big challenge in norms and argumentation.

5.4 Applying Norms in a Flexible and Adaptive Way: Fuzziness in Legal Interpretation

Legal interpretation is a mechanism allowing legal norms to be adapted to unforeseen situations. This section outlines promising research lines in nMAS on the role of interpretation in legal reasoning. As recalled in Section 3, norms have typically a conditional structure and may be thus represented as a rule $b_1, \ldots, b_n \Rightarrow l$ such that l is the legal effect linked to the norm. The degree associated to l depends on the degrees of truth of conditions for each b_i. These degrees depend in turn on the goal associated with the norm. An interesting approach is to define the fuzzy set $b'_i = f(b_i, g_j)$ where the value of each b'_i increases or decreases according to the match between b_i and the goal associated with the norm j. The degree of match depends on how concepts relevant to the norm are defined in a domain ontology.

5.4.1 Description of the Topic and Background

5.4.1.1 Legal Interpretation

Since norms have a conditional structure such as $b_1, \ldots, b_n \Rightarrow l$ (if $b_1, \ldots, b_n$ hold, then l should be the case), if l states that, e.g., some p is obligatory, then an agent is compliant with respect to this norm if p is obtained whenever $b_1, \ldots, b_n$ are derived. Many logical models of legal reasoning assume that conditions of norms give a complete description of their applicability (for a discussion, see [86]). However, this assumption is too strong, due to the complexity and dynamics of the world. Norms cannot take into account all the possible conditions in which they should or should not be applied, giving rise to the so called "penumbra": while we can often identify a core of cases which can clearly be classified as belonging to the concept, there is a penumbra of hard cases, in which the membership

[8] https://github.com/carneades/carneades

of the concept can be disputed [49]. Moreover, not only does the world change, giving rise to circumstances unexpected by the legislator who introduced the norm, but even the ontology of reality can change with respect to the one constructed by the law to describe the applicability conditions of norms. Consider, e.g., the problems concerning the application of existing laws to privacy, intellectual property or technological innovations in healthcare. To cope with unforeseen circumstances, the judicial system, at the moment in which a case concerning a violation is discussed in court, is empowered to interpret, i.e., to change norms, under some restrictions not to go beyond the purpose from which the norms stem.

5.4.1.2 Categories and Metaphors

Legal systems are the product of human mind and are then written in a natural language. This implies two facts :

- the basic processes of human cognition have to be taken into account when interpreting norms;
- as natural languages are inherently vague and imprecise, so are norms.

The application of laws to a new situation is a metaphorical process: the new situation is mapped on to a situation in which applying law is obvious, by analogy. Here, by “metaphor” we mean using a well understood, prototypical situation to represent and reason about a less understood, novel situation. Metaphors are one of the basic building blocks of human cognition [58].

Norms are written with references to categories. Take, for instance, Section 2 of the US Marihuana Tax Act of 1937:

> Every person who imports, manufactures, . . . , administers, or gives away marihuana shall . . . pay the following special taxes . . .

This norm makes reference to concepts such as person, import, manufacture, administer, and marihuana which may be described as categories of entities or actions. Applying this norm to a particular case means recognising that a particular individual may be categorised as a person, that what he/she does may be categorised as giving away something, and that what he/she gives away may be categorised as marihuana.

As pointed out by Lakoff [57], “Categorisation is not a matter to be taken lightly. There is nothing more basic than categorization to our thought, perception, action, and speech”. The folk theory that categories are defined by common properties is not entirely wrong, but it is only a small part of the story. It is now clear that categories may be based on prototypes. Some categories are vague or imprecise; some do not have gradation of membership, while others do. The category US Senator is well defined, but categories like rich people or tall men are graded, simply because there are different degrees of richness and tallness. However, it is important to notice that these degrees of membership depend both on the the context in which the norm will be applied and on the goal associated with the norm. To be considered tall in the Netherlands is not the same as to be considered tall in Portugal, for example. We have than first to consider the context and than to consider the goal associated with the norm. If the goal is context-dependent, both aspects can be considered at same time.

An effective and natural tool for formally modeling these phenomena is Fuzzy Logic, which is indeed suitable to capture all the above issues related to categories. More precisely, a category may be represented as a fuzzy set: the membership of an element to a category is a graded concept. As a result, we get that a norm may apply to a given situation only to a certain extent and different norms may apply to different extents to the same situation.

5.4.1.3 Fuzzy Logic

As is well known, fuzzy logic was initiated by Lotfi Zadeh with his seminal work on fuzzy sets [99]. Fuzzy set theory provides a mathematical framework for representing and treating vagueness, imprecision, lack of information, and partial truth.

Very often, we lack complete information in solving real world problems. This can be due to several causes. First of all, human expertise is of a qualitative type, hard to translate into exact numbers and formulas. Our understanding of any process is largely based on imprecise, "approximate" reasoning. However, imprecision does not prevent us from performing successfully very hard tasks, such as driving cars, improvising on a chord progression, or trading financial instruments. Furthermore, the main vehicle of human expertise is natural language, which is in its own right ambiguous and vague, while at the same time being the most powerful communication tool ever invented.

Fuzzy sets are a generalization of standard sets obtained by replacing the characteristic function of a set A, χ_A which takes values in $\{0,1\}$ ($\chi_A(x) = 1$ iff $x \in A$, $\chi_A(x) = 0$ otherwise) with a *membership function* μ_A, which can take any value in $[0,1]$. The value $\mu_A(x)$ is the membership degree of element x in A, i.e., the degree to which x belongs in A. A fuzzy set is completely defined by its membership function. Therefore, we can use it to define the core and penumbra of normative concepts. Indeed, given a fuzzy set A, its *core* is the (conventional) set of all elements x such that $\mu_A(x) = 1$ while its *support* is the set of all x such that $\mu_A(x) > 0$.

Since a fuzzy set is completely defined by its membership function, the question arises of how the shape of this function is determined. From an engineering point of view, the definition of the ranges, quantities, and entities relevant to a system is a crucial design step. In fuzzy systems all entities that come into play are defined in terms of fuzzy sets, that is, of their membership functions. The determination of membership functions is then correctly viewed as a problem of design. As such, it can be left to the sensibility of a human expert or more objective techniques can be employed. Alternatively, optimal membership function assignment, of course relative to a number of design goals that have to be clearly stated, such as robustness, system performance, etc., can be estimated by means of a machine learning or optimization method. In particular, evolutionary algorithms have been employed with success to this aim.

The usual set-theoretic operations of union, intersection, and complement can be defined for fuzzy sets as a generalization of their counterparts on standard sets. Likewise, in fuzzy logic the set of true proposition and its complement, the set of false propositions, are fuzzy. The degree to which a given proposition P belongs to the set of true propositions is its degree of truth. Much in the same way, a one-to-one mapping can be established as well between fuzzy sets and fuzzy predicates. In classical logic a predicate of an element of the universe of discourse defines the set of elements for which that predicate is true and its complement, the set of elements for which that predicate is not true. In fuzzy logic, by contrast, these sets are fuzzy and the degree of truth of a predicate of an element is given by the degree to which that element is in the set associated with that predicate.

5.4.2 Current Understanding and Open Challenges

Again, let $b_1, \ldots, b_n \Rightarrow l$ be a norm. We can represent each b_i as a proposition of the form 'x is A', where x is a variable and A is a category represented as a fuzzy set.

As in [17], we can assume that goals are assigned to norms. The role of such goals, which can be context-dependent or not, is to pose the limits within which the interpretation process of the judicial systems must stay when interpreting norms. As a consequence, the definition of category changes: we have to consider the degree to which x belongs to a category A with respect to the reason or goal behind the norm. Thus, a category can have as many membership functions as the number of goals behind the norm. For example, let us consider the following example in [17]: if x is a vehicle and y is a park, then it is forbidden for any x to enter y. If the x is a bicycle and the goal of the norm is to limit air pollution, the degree to which x belongs to the category "vehicle" should be different than in the case in which the goal of the norm would be the "safety of kids walking in the park".

5.5 Connections among the Research Topics

In this section we outline some connections among the topics presented in the previous parts of Section 5. These links show what benefits each research line can offer to the others.

5.5.1 Moral Agency: Relations with the Other Topics

- *Group norms* (Section 5.2) – There are different ways to explain the origin of moral attitudes such as ideals, standards and values. Social scientists (see, e.g., [13]) have defended the idea that there exist innate moral principles in human agents such as the principle of fairness which are the product of biological evolution. Other ideals are the product of the internalization of some external norm. A possible explanation is based on the hypothesis that moral judgments are true or false only in relation to and with reference to some agreement between people forming a group or a community. More precisely, an agent's ideals are simply norms of the group or community to which the agent belongs that have been internalized by the agent.[9] This is the essence of the philosophical doctrine of moral relativism [47].
  An interesting research direction for the NorMAS community is to provide models clarifying the relationships between group norms and moral values.
- *Argumentation and norms* (Section 5.3) – Morality also has interesting connections with argumentation theory. For instance, it would be interesting to develop models of argumentation in which agents may advance arguments supported by moral values, standards or ideals (e.g. "Do not speak too loudly. It is unpolite.").

5.5.2 Group Norms: Relations with the Other Topics

- *Argumentation and norms* (Section 5.3) – When a group norm is violated or complied with, sanctions or rewards associated with the norm should be distributed among group members. A promising way to facilitate this is via argumentation: members would propose a distribution (or just their own share of the reward or sanction) with its justification, and engage in an argument. Typical justifications would be "I deserve a higher share of the reward because I performed the most expensive individual action" or "You deserve more blame because you failed to perform the cheapest/simplest of the actions". Additionally, it might be useful for an agent to belong to a group or not, so as to take advantage, for instance, of permissions (which the agent would not otherwise have), or because other members of the group will act in an attractive way (say, performing actions which the agent would have to do otherwise). If a group specification is vague or may have different interpretations (see also Section Section 5.4), then agents may engage in arguments

[9] See [2] for a cognitive model of norm internalization.

making a case towards or against their membership to a group. The two issues (blame or praise apportioning and membership) merge when an agent is trying to make a case against being blamed (and fined) because they do not belong to the group.

- *Moral agency* (Section 5.1) – Group norms precisely define reference communities (namely, the groups to which the norm applies) against which agents' moral emotions can be gauged. When a group norm is violated or complied with, then each agent should provide her own assessment of her degree of blameworthiness or praiseworthiness with respect to the group and her own contribution(s). The "other" parties from which admiration or reproach stems are members of the group, expressing their opinions on other members of the group. These two sources of opinion (namely, "self" and "others") will inform the blame/praise apportioning for sanctions/rewards.
- *Norms vs. policies* (Section 4) – The differentiation between norms and policies is useful in many ways. If research can provide a formal distinction between these concepts, with precisely defined points of contact, then policies provide the *design rationale* of norms[10]. For instance, a policy of "protecting minorities and their culture" may influence or explain the design of the norm "a member of a minority may speak her own language in court". Policies addressing groups will give rise to group norms; if a group norm is contested—agents may challenge the norm, asking why it should be complied with—-then its policy is presented as the rationale.
- *Fuzziness of legal norms* (Section 5.4) – In realistic settings, checking the membership of a group may not be a clear-cut issue. For instance, an agent checking its membership of a group such as "tall people" or "heavy equipment" could have *degrees of truth*, which can be captured and studied with fuzzy interpretation techniques. Interestingly, the perceived degree of membership to a group might inform the agent as to how much importance the agent should give to the norm. Blame and praise apportioning among members of a group, when a norm is violated or complied with, could make use of the agent's perceived degree of membership to a group.

5.5.3 Argumentation and Norms: Relations with the Other Topics

- *Moral agency* (Section 5.1) – Moral agency may lead to an improvement of preference-based argumentation, where the preferences are not simply assigned in the abstract way, but they are grounded on the moral norms regulating the behaviour, and thus the way of running argumentations, of the single agents.
- *Fuzziness of legal norms* (Section 5.4) – Norm interpretation leads to new challenges in argumentation theory, in particular the challenge is to go behind legal disputes by seeing arguments as, for instance, alternative legal theories exchanged that define a same concept and its scope, instead of single arguments, to assess their goodness.

5.5.4 Fuzziness of Legal Norms: Relations with the Other Topics

- *Moral agency* (Section 5.1) – The moral aspects of an agent could be integrated in the process of norm interpretation as a new and more specific component besides the context and the goal associated to the norm. The fact that a same action receives different moral evaluations may depend on a different ways in which this action is classified.

[10] These terms are used here following their meaning within the multi-agent systems research community.

- *Group norms* (Section 5.2) – The notion of belonging to a group of agents with respect to a norm is introduced. In the presence of a norm, the agent has to decide if it is a member of a group addressed/affected by the norm, that is, it has to interpret the legal meaning of the norm.
- *Argumentation and norms* (Section 5.3) – Argumentation-based persuasion could be used in order to change the interpretation of a norm by an agent. For example, arguing about different goals of norm categories means arguing about different membership functions for those categories.

6 Summary

This chapter set the scene for nMAS by giving some definitions of norms and nMAS based on common characteristics as well as pointing out requirements for nMAS. We presented three views on nMAS: a social, a norm change and a mechanisms design one. We furthermore discussed several guidelines for the development of nMAS proposed in [18, 19] and compared them with some of the requirements for legal knowledge representation outlined in [38].

We assumed that norms are used to coordinate, organize, guide, regulate or control interaction among distributed autonomous systems or entities; and that nMAS use these norms to govern these systems using restrictions on patterns of behaviour of the agents in the system.

The so-called “social science” definition looks at norms that are actively or passively transmitted and have a social function and impact.

The so-called “norm-change” definition supports the derivation of those guidelines that require to motivate which definition of normative multi-agent system is used; also, this definition is meant to to make explicit why norms are a kind of soft constraints deserving special analysis, and to explain why and how norms can be changed at runtime. The so-called “mechanism design” definition entails the guidelines recommending to discuss the use and role of norms as a mechanism in a game-theoretic setting and to clarify the role of norms in the multi-agent system. The formal requirements of Gordon et al. [38] offer a complementary analysis to the ones in [18, 19], as they provide a fine-grained account of the notions of norm and normative system.

Finally, we considered in some detail four potential research lines concerning the nature of nMAS: (i) the concept of moral agency, especially in a cognitive perspective, (ii) the concept of group norm, (iii) the connection between argumentation and norms, and (iv) the role of conceptual vagueness and fuzziness in norm interpretation and application.

References

1 Giulia Andrighetto, Rosaria Conte, Paolo Turrini, and Mario Paolucci. Emergence in the loop: Simulating the two way dynamics of norm innovation. In Guido Boella, Leon van der Torre, and Harko Verhagen, editors, *Normative Multi-agent Systems*, number 07122 in Dagstuhl Seminar Proceedings, 2007.

2 Giulia Andrighetto, Daniel Villatoro, and Rosaria Conte. Norm internalization in artificial societies. *AI Communication*, 23(4):325–339, 2010.

3 Kevin D. Ashley. *Modeling legal argument - reasoning with cases and hypotheticals*. Artificial Intelligence and Legal Reasoning. MIT Press, 1990.

4 Robert Axelrod. An evolutionary approach to norms. *The American Political Science Review*, 80(4):1095–1111, December 1986.

5 Tina Balke. *Towards the Governance of Open Distributed Grids – A Case Study in Wireless Mobile Grids.* PhD thesis, University of Bayreuth, Chair for Information Systems Management, 2011.

6 Tina Balke and Daniel Villatoro. Operationalization of the sanctioning process in hedonic artificial societies. In *Workshop on Coordination, Organization, Institutions and Norms in Multiagent Systems @ AAMAS 2011, Taiwan*, 2011.

7 Trevor Bench-Capon. The missing link revisited: The role of teleology in representing legal argument. *Artificial Intelligence and Law*, 10(2-3):79–94, 2002.

8 Trevor Bench-Capon and Frans Coenen. Isomorphism and legal knowledge based systems. *Artificial Intelligence and Law*, 1(1):65–86, 1992.

9 Trevor Bench-Capon, Henry Prakken, and Giovanni Sartor. *Argumentation in Legal Reasoning.* Argumentation in Artificial Intelligence. Springer, 2010.

10 Floris Bex, Henry Prakken, Chris Reed, and Douglas Walton. Towards a formal account of reasoning about evidence: Argumentation schemes and generalisations. *Artif. Intell. Law*, 11(2-3):125–165, 2003.

11 Cristina Bicchieri and Ryan Muldoon. Social norms. The Stanford Encyclopedia of Philosophy, 2011. Stanford University Press.

12 Holger Billhardt, Roberto Centeno, Carlos E. Cuesta, Alberto Fernández, Ramón Hermoso, Rubén Ortiz, Sascha Ossowski, J. Santiago Pérez-Sotelo, and Matteo Vasirani. Organisational structures in next-generation distributed systems: Towards a technology of agreement. *Multiagent and Grid Systems*, 7(2-3):109–125, 2011.

13 Kenneth Binmore. *Natural justice.* Oxford University Press, New York, 2005.

14 Alexander Bochman. Collective argumentation and disjunctive logic programming. *J. Log. Comput.*, 13(3):405–428, 2003.

15 Alexander Bochman. Production inference, nonmonotonicity and abduction. In *Proceeding of the Eighth International Symposium on Artificial Intelligence and Mathematics (AMAI 2004)*, 2004.

16 Alexander Bochman. Propositional argumentation and causal reasoning. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI 2005)*, pages 388–393, 2005.

17 Guido Boella, Guido Governatori, Antonino Rotolo, and Leendert van der Torre. A logical understanding of legal interpretation. In *KR'2010*, 2010.

18 Guido Boella, Gabriella Pigozzi, and Leendert van der Torre. Five guidelines for normative multiagent systems. In *JURIX*, pages 21–30, 2009.

19 Guido Boella, Gabriella Pigozzi, and Leendert van der Torre. Normative systems in computer science - ten guidelines for normative multiagent systems. In Guido Boella, Pablo Noriega, Gabriella Pigozzi, and Harko Verhagen, editors, *Normative Multi-Agent Systems*, number 09121 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2009. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany.

20 Guido Boella and Leendert Van Der Torre. Introduction to normative multiagent systems. *Computational and Mathematical Organization Theory*, 12:71–79, 2006.

21 Guido Boella and Leendert van der Torre. The ontological properties of social roles in multi-agent systems: Definitional dependence, powers and roles playing roles. *Artificial Intelligence and Law Journal (AILaw)*, 2007.

22 Guido Boella, Leendert van der Torre, and Harko Verhagen. Normative multi-agent systems. In *Internationales Begegnungs und Porschungszentrum fur Informatik (IBFI)*, 2007.

23 Guido Boella, Leendert van der Torre, and Harko Verhagen. Introduction to the special issue on normative multiagent systems. *Autonomous Agents and Multi-Agent Systems*, 17(1):1–10, 2008.

24 Jan Broersen, Mehdi Dastani, Joris Hulstijn, and Leendert van der Torre. Goal generation in the BOID architecture. *Cognitive Science Quarterly*, 2(3-4):428–447, 2002.

25 E. Bulygin. Permissive norms and normative systems. In A. Martino and F. Socci Natali, editors, *Automated Analysis of Legal Texts*, pages 211–218. Publishing Company, Amsterdam, 1986.

26 José Carmo. Collective agency, direct action and dynamic operators. *Logic Journal of the IGPL*, 18(1):66–98, 2010.

27 José Carmo and Olga Pacheco. Deontic and action logics for organized collective agency, modeled through institutionalized agents and roles. *Fundam. Inform.*, 48(2-3):129–163, 2001.

28 James Coleman. *Foundations of Social Theory.* Belknap Press, 1990.

29 R. Conte, C. Castelfranchi, and F. Dignum. Autonomous norm-acceptance. In *Intelligent Agents V (ATAL98)*, LNAI 1555, pages 319–333. Springer, 1999.

30 Jon Elster. *The cement of society: a study of social order.* Cambridge University Press, 1989.

31 Arthur M. Farley and Kathleen Freeman. Burden of proof in legal argumentation. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law (ICAIL 1995)*, pages 156–164, 1995.

32 Ernst Fehr and Klaus M. Schmidt. Theories of fairness and reciprocity: Evidence and economic applications. In *Advances in Economics and Econometrics.* Cambridge University Press, 2003.

33 Nigel Gilbert. Emergence in social simulation. In *Artificial Societies: The Computer Simulation of Social Life*, pages 144–156. UCL Press, 1995.

34 Herbert Gintis, Samuel Bowles, Robert Boyd, and Ernst Fehr, editors. *Moral sentiments and material interests.* MIT Press, Cambridge, 2005.

35 Uri Gneezy and Aldo Rustichini. A fine is a price. *The Journal of Legal Studies*, 29(1):1–18, 2000.

36 Thomas F. Gordon. The pleadings game: Formalizing procedural justice. In *Proceedings of the Fourth International Conference on Artificial intelligence and Law (ICAIL 1993)*, pages 10–19, 1993.

37 Thomas F. Gordon. *The Pleadings Game: An Artificial Intelligence Model of Procedural Justice.* Kluwer, Dordrecht, 1995.

38 Thomas F. Gordon, Guido Governatori, and Antonino Rotolo. Rules and norms: Requirements for rule interchange languages in the legal domain. In *RuleML*, pages 282–296, 2009.

39 Thomas F. Gordon and Douglas Walton. Legal reasoning with argumentation schemes. In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Law (ICAIL 2009)*, pages 137–146, 2009.

40 Guido Governatori. Representing business contracts in RuleML. *International Journal of Cooperative Information Systems*, 14(2-3):181–216, 2005.

41 Guido Governatori and Antonino Rotolo. BIO logical agents: Norms, beliefs, intentions in defeasible logic. *Journal of Autonomous Agents and Multi Agent Systems*, 17(1):36–69, 2008.

42 Guido Governatori and Antonino Rotolo. Changing legal systems: Legal abrogations and annulments in defeasible logic. *The Logic Journal of IGPL*, 18(1):157–194, 2010.

43 Guido Governatori, Antonino Rotolo, and Giovanni Sartor. Temporalised normative positions in defeasible logic. In *10th International Conference on Artificial Intelligence and Law (ICAIL05)*, pages 25–34. ACM Press, 2005.

44 Davide Grossi, Frank Dignum, Lambèr Royakkers, and Jean-Jules Meyer. Collective obligations and agents: Who gets the blame? In *Procs. 7th Int'l Workshop on Deontic Logic*

in Computer Science (DEON 2004), volume 3065 of *Lecture Notes in Computer Science*, pages 129—-145. Springer, 2004.

45 Juergen Habermas. *Between facts and norms: contributions to a discourse theory of law and democracy.* Polity Press, 1999.

46 Jonathan Haidt. The moral emotions. In R. J. Davidson, K. R. Scherer, and H. H. Goldsmith, editors, *Handbook of Affective Sciences*, pages 852—870. Oxford University Press, 2003.

47 Gilbert H. Harman. *Explaining value and other essays in moral philosophy.* Clarendon Press, Oxford, 2000.

48 John Harsanyi. Morality and the theory of rational behaviour. In A. K. Sen and B. Williams, editors, *Utilitarianism and beyond.* Cambridge University Press, Cambridge, 1982.

49 H.L.A. Hart. *The concept of law.* Clarendon, Oxford, 1994.

50 Andreas Herzig and Emiliano Lorini. A dynamic logic of agency I: STIT, abilities and powers. *Journal of Logic, Language and Information*, 19:89–121, 2010.

51 Wesley Newcomb Hohfeld. Some fundamental legal conceptions as applied in judicial reasoning. *The Yale Law Journal*, 23(1):16–59, 1913.

52 Christopher D. Hollander and Annie S. Wu. The current state of normative agent-based systems. *Journal of Artificial Societies and Social Simulation*, 14(2), 2011.

53 Martin Hollis. *Trust within reason.* Cambridge University Press, Cambridge, 1998.

54 Matthew Interis. On norms: A typology with discussion. *American Journal of Economics and Sociology*, 70(2):424–438, April 2011.

55 Andrew J. I. Jones and Marek J. Sergot. A formal characterisation of institutionalised power. *Logic Journal of the IGPL*, 4(3):427–443, 1996.

56 Hans Kelsen. *General theory of norms.* Clarendon, Oxford, 1991.

57 G. Lakoff. *Women, Fire, and Dangerous Things.* University of Chicago Press, Chicago, 1987.

58 G. Lakoff and M. Jonhson. *Metaphors We Live By.* University of Chicago Press, Chicago, 1980.

59 Steven D. Levitt and Stephen J. Dubner. *Freakonomics : A Rogue Economist Explores the Hidden Side of Everything.* William Morrow, New York, May 2005.

60 David Lewis. *Convention: a philosophical study.* Harvard University Press, 1969.

61 Tingting Li, Tina Balke, Marina De Vos, and Julian Padget. Conflict detection in composite institutions. In *Proceedings of the International Workshop on Agent-based Modeling for Policy Engineering*, 2012.

62 Emiliano Lorini, Dominique Longin, Benoit Gaudou, and Andreas Herzig. The logic of acceptance: grounding institutions on agents' attitudes. *Journal of Logic and Computation*, 19(6):901–940, 2009.

63 Emiliano Lorini, Nicolas Troquard, Andreas Herzig, and Cristiano Castelfranchi. Delegation and mental states. In *Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2007)*, pages 610–612. ACM Press, 2007.

64 Theodore J. Lowi. Four systems of policy, politics, and choice. *Public Administration Review*, 32(4):298–310, 1972.

65 Michael Luck, Peter McBurney, and Chris Preist. *Agent Technology: Enabling Next Generation Computing.* AgentLink, 2003. Electronically available, `http://www.agentlink.org/roadmap/`.

66 Neil MacCormick. *Legal reasoning and legal theory.* Clarendon law series. Clarendon Press, 1978.

67 David Makinson. On the formal representation of rights relations: Remarks on the work of stig kanger and lars lindahl. *Journal of Philosophical Logic*, 15(4):403–425, 1986.

68 Thomas Ågotnes and Natasha Alechina. Reasoning about joint action and coalitional ability in kn with intersection. In *Procs. 12th int'l conf. on Computational logic in multi-agent systems (CLIMA'11)*, volume 6814 of *Lecture Notes in Computer Science*, pages 139–156, Berlin, Heidelberg, 2011. Springer-Verlag.

69 Robert K. Merton. *Social Theory and Social Structure.* Free Press, 1968.

70 John.-Jules Meyer and Roel Wieringa. *Deontic Logic in Computer Science: Normative System Specification.* John Wiley & Sons, Chichester, England, 1993.

71 Martin Neumann. Homo socionicus: a case study of simulation models of norms. *Journal of Artificial Societies and Social Simulation*, 11(4):6, 2008.

72 Timothy J. Norman and Chris Reed. A logic of delegation. *Artif. Intell.*, 174:51–71, January 2010.

73 Douglass C. North. *Institutions, Institutional Change and Economic Performance (Political Economy of Institutions and Decisions).* Cambridge University Press, October 1990.

74 Douglass C. North. Institutions, transaction costs and productivity in the long run. Economic History 9309004, EconWPA, September 1993.

75 Nir Oren, Antonino Rotolo, Leon van der Torre, and Serena Villata. Norms and argumentation. In *Handbook on Agreement Technologies.* Springer, Dordrecht, 2013.

76 Olga Pacheco and José Carmo. A role based model for the normative specification of organized collective agency and agents interaction. *Autonomous Agents and Multi-Agent Systems*, 6:145–184, March 2003.

77 Henry Prakken. A logical framework for modelling legal argument. In *Proceedings of the Fourth International Conference on Artificial intelligence and Law (ICAIL 1993)*, pages 1–9, 1993.

78 Henry Prakken, Chris Reed, and Douglas Walton. Dialogues about the burden of proof. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law (ICAIL 2005)*, pages 115–124, 2005.

79 Henry Prakken and Giovanni Sartor. A dialectical model of assessing conflicting argument in legal reasoning. *Artificial Intelligence and Law*, 4(3-4):331–368, 1996.

80 Henry Prakken and Giovanni Sartor. A dialectical model of assessing conflicting arguments in legal reasoning. *Artif. Intell. Law*, 4(3-4):331–368, 1996.

81 Michael Rescorla. Convention. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy.* Spring 2011 edition, 2011.

82 Edwina L. Rissland, David B. Skalak, and M. Timur Friedman. BankXX: A Program to Generate Argument Through Case-Base Research. In *Proceedings of the Fourth International Conference on Artificial intelligence and Law (ICAIL 1993)*, pages 117–124, 1993.

83 Lambèr Royakkers. Combining deontic and action logics for collective agency. In *13th annual conf. on Legal knowledge and information systems (JURIX 2000)*, volume 64 of *Front. Artif. Intell. Appl.*, pages 135–146. IOS Press, December 2000.

84 Rossella Rubino, Antonino Rotolo, and Giovanni Sartor. An OWL ontology of fundamental legal concepts. In *Proceedings of JURIX 2006*, pages 101–110, 2006.

85 D.W.P. Ruiter. A basic classification of legal institutions. *Ratio Juris*, 10(4):357–371, 1997.

86 Giovanni Sartor. *Legal Reasoning: A Cognitive Approach to the Law.* Springer, Dordrecht, 2005.

87 Susan B. Schimanoff. *Communication rules: Theory and Research.* Sage Publications, 1980.

88 John R. Searle. *The Construction of Social Reality.* The Free Press, New York, 1995.

89 John R. Searle. *Rationality in Action.* MIT Press, Cambridge, 2001.

90 Marek J. Sergot. A computational theory of normative positions. *ACM Transactions on Computational Logic*, 2(4):581–622, October 2001.

91 Marek J. Sergot, Fariba Sadri, Robert A. Kowalski, F. Kriwaczek, Peter Hammond, and H. T. Cory. The British Nationality Act as a logic program. *Commun. ACM*, 29(5):370–386, 1986.

92 Reid G. Smith. The contract net protocol: High-level communication and control in a distributed problem solver. *IEEE Trans. Comput.*, 29(12):1104–1113, December 1980.

93 Goran Therborn. Back to norms! on the scope and dynamics of norms and normative action. *Current Sociology*, 50:863–880, 2002.

94 Raimo Tuomela and Maj Bonnevier-Tuomela. Norms and agreements. *European Journal of Law, Philosophy and Computer Science*, 5:41–46, 1995.

95 Geoffrey Vickers. Values, norms and policies. *Policy Science*, 4:103–111, 1973.

96 Serena Villata, Guido Boella, and Leendert van der Torre. Argumentation patterns. In *Proceedings of the 8th International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2011)*, pages 133–150, 2011.

97 Georg Henrik von Wright. *Norm and Action.* Routledge, London, 1963.

98 H. Peyton Young. Social norms. The New Palgrave Dictionary of Economics, 2008. Macmillan.

99 Lofti A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

Normative Reasoning and Consequence

Jan Broersen[1], Stephen Cranefield[2], Yehia Elrakaiby[3], Dov Gabbay[4], Davide Grossi[5], Emiliano Lorini[6], Xavier Parent[3], Leendert W. N. van der Torre[3], Luca Tummolini[7], Paolo Turrini[3], and François Schwarzentruber[8]

1 Utrecht University, The Netherlands
2 University of Otago, New Zealand
3 University of Luxembourg, Luxembourg
4 King's College London, United Kingdom
5 University of Liverpool, England
6 Paul Sabatier University – Toulouse, France
7 National Research Council (CNR) / ISIC – Rome, Italy
8 ENS Cachan / IRISA, France

Abstract

In this chapter we first provide a general introduction to the research area methodology and relevance, then we discuss normative reasoning for multi-agent systems, and finally we discuss current research challenges. We cover the main issues in modern deontic logic, which is much broader than the traditional modal logic framework of deontic logic, with an emphasis to our intended audience. To emphasize this broadness, we typically refer to "deontic logic and normative systems" rather than deontic logic only.

1998 ACM Subject Classification I.2.11 Distributed Artificial Intelligence

Keywords and phrases Norms, MAS

Digital Object Identifier 10.4230/DFU.Vol4.12111.33

1 Introduction

We first give a general introduction to the research area methodology and relevance, then we discuss normative reasoning for multi-agent systems, and finally we discuss current research challenges.

The intended audience is a general multi-agent systems audience, and no previous knowledge of deontic logic is presupposed. For a more detailed and in depth discussion on deontic logic, we refer to the handbook of deontic logic and normative systems.

We cover the main issues in modern deontic logic, which is much broader than the traditional modal logic framework of deontic logic, with an emphasis to our intended audience. To emphasize this broadness, we typically refer to "deontic logic and normative systems" rather than deontic logic only.

The paper is based on discussions during Dagstuhl seminar "Normative Multi-Agent Systems"[1], and has subsequently been written by participants of the workshop, extended with a few additional authors.

[1] `http://www.dagstuhl.de/12111`

Normative Multi-Agent Systems. *Dagstuhl Follow-Ups, Vol. 4.* ISBN 978-3-939897-51-4.
Editors: Giulia Andrighetto, Guido Governatori, Pablo Noriega, and Leendert W.N. van der Torre; pp. 33–70
DAGSTUHL FOLLOW-UPS Dagstuhl Publishing
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany

2 Methodology and relevance

We explain what deontic logic and normative reasoning are, why normative reasoning is relevant for normative multi-agent systems, what the advantages of formal methods in multi-agent systems are, and whether normative reasoning is special in comparison to other kinds of reasoning.

2.1 What are deontic logic and normative reasoning?

Succinctly put, deontic logic can be defined as the formal study of normative reasoning. In this section, we explain what this definition means.

Generally speaking, one might define logic as the study of the principles of correct reasoning. It tells us whether certain conclusions follow from some given assumptions. The truth of propositions and the validity of reasoning are quite distinct. Empirical scientists, private detectives *etc.* are concerned with the first. Logicians are concerned with the second. For instance, the logic principle $\neg(A \wedge \neg A)$, where $\neg$ stands for negation and $\wedge$ stands for conjunction, is known as the principle of non-contradiction. It says nothing about the truth-value of A and $\neg A$, but only that they cannot be true at the same time.

Sentential logic (propositional logic, first-order logic, *etc*) looks at the logical relationships amongst utterances that assert that something can be judged true or false, like "the cat is on the mat." Such a sentence is usually called a declarative one. By contrast, deontic logic is the study of the logical relationships among propositions that assert that certain actions or states of affairs are obligatory, forbidden or permitted. Deontic comes from the Greek *deon* meaning "that which is binding, right." The latter sentences are usually called directives. A typical example is "you should not eat with your fingers." Such a sentence can be formalized as $\bigcirc\neg p$, where $\bigcirc$ is read as "it is obligatory that", and p is read as "eating with your fingers." One might add various levels of granularity using first-order logic, or using other modal operators (here, a modal operator for agency).

It is traditional to take the view that logic is topic neutral. The laws of biology might be true only of living creatures, and the laws of economics are only applicable to groups of agents that engage in financial transactions. But the principles of logic are universal principles which are more general than biology and economics. Thus, the principle of non-contradiction mentioned above applies to both biology and economics. The same point can be made about the principles of deontic logic. It does not matter where the directives come from. These can be part of a moral code, or a legal system, or a set of traffic regulations, *etc.*

There are two main types of directive sentences that are studied in deontic logic: regulative norms and constitutive norms. The first are obligations, prohibitions and permissions part of a normative system. The second say what counts as what within a normative system. In the example below, the first premise is a constitutive norm, while the second premise and the conclusion are regulative norms:

- Bikes count as vehicles.
- Vehicles are not allowed to access public parks.
- Therefore, bikes are not allowed to access public parks.

Deontic logic studies the two, and their interaction.

2.2 Why is normative reasoning relevant for NorMAS?

In recent years, the study of Multi-Agent Systems (MAS) has undergone what can be called a normative turn, shifting the emphasis on normative issues in agent organizations. The

AgentLink Roadmap, published in 2005, considers norms as key for the development of MAS. This shift in focus has given rise to a new interdisciplinary research area called "NorMAS" (Normative Multi-Agent Systems). It can be defined as the intersection of normative systems and multi-agent systems.

One motivation comes from systems where artificial and human agents interact. Since the use of norms is a key element of human social intelligence, norms may be essential too for artificial agents that collaborate with humans, or that are to display behavior comparable to human intelligent behavior. By integrating norms and individual intelligence normative multi-agent systems provide a promising model for human and artificial agent cooperation and coordination, group decision making, multi-agent organizations, regulated societies, electronic institutions, secure multi-agent systems, and so on.

Another key point for using normative reasoning is that it is an independent tool for building interoperability standards and for reasoning about them, a sort of lingua franca that preserves the heterogeneity of the agents and their autonomy.

Multi-agent research has been driven by the need to find a substantially more realistic model than those used thus far. It is a common assumption in many multi-agent systems that agents will behave as they are intended to behave. There are many circumstances in which such an assumption must be abandoned. In particular, in open agent societies, where agents are programmed by different parties, where there is no direct access to an agent's internal state, and where agents do not necessarily share a common goal, it cannot be assumed that all agents will behave according to the system norms that govern their behavior. Agents must be assumed to be untrustworthy because they act on behalf of parties with competing interests, and so they may fail to, or even choose not to, comply with the society's norms in order to achieve their individual goals. It is then usual to impose sanctions to discourage norm violating behavior and to provide some form of reparation when it does occur.

An alternative use of normative reasoning comes from so-called logic based agents. Agents need to represent their environment in order to be able to act intelligently. The assumption (already made by McCarthy in 1958 – see [97]) is that an efficient representation can be found using logic. The idea is that agents only need to store "basic" facts about the external world, and derive the rest using logic. This hints at what Russell and Norvig call "logic-based agents" (see also Wooldridge [148]). An agent is considered having amongst others two components:

- a knowledge base;
- a set of deduction rules (a "program").

The knowledge base stores information about the world. By applying the deduction rules to the knowledge base, the agent gets more information about the world, and can interact with it more intelligently. In the case of a normative multi-agent system, the knowledge base contains roles played by the agent, obligations and permissions associated with these roles, constitutive rules, and the like. The goal of deontic logic is to identify the relevant deduction rules used to extract information from the knowledge base, thus conceived.

2.3 Advantages of formal methods

Deontic logic and other formalisms for normative reasoning are examples of formal methods with formal semantics. Formal methods can be used as a modeling language when designing multi-agent systems, for explaining their structure to other designers, or for reasoning about the system.

In general, the advantages of using formal methods over informal ones may be seen by contrasting them with other methods, like the Unified Modeling Language (UML) – see *e.g.*

[59]. This graphical language is a *de facto* standard for object-oriented modeling. Its success within the industry has resulted in a number of attempts to adapt the UML notation for modeling agent systems. Odell and colleagues [107] have discussed several ways in which the UML notation might usefully be extended to enable the modellng of agent systems. The proposed modification include:

- support for expressing concurrent threats of interaction, thus enabling UML to model such well-known agent protocols as the Contract-Net
- a notion of "role" that extends that provided in UML, and in particular, allows the modeling of an agent playing many roles roles, which are usually associated with obligations, permissions and powers.

In spite of being urgently required, little work has been done on extending UML with such normative notions (though it has been studied in the context of so-called business rules in that community).

There are many papers that explain that UML has indeed a semantics even if it is not based on formal methods and why OMG decided this way. UML and formal methods have two completely different purposes. Many papers also proposed formal methods for UML (mainly based on Petri Nets), which is used in specific domains and applications. This it is not generally adopted, because all the proposed formal methods are too specific for the general purposes UML aims at.

A graphical language like UML is useful. For it mainly facilitates communication amongst researchers with different backgrounds. But it has a number of pitfalls, among which is the fact that it is error-prone. This is where formal methods come into the picture. They provide a mathematically rigorous framework for modeling normative multi-agent systems, so advanced tool support can be given. The modeling language is given a formal semantics, which constrains the intuitive characterization of the normative notions being used. The language also comes equipped with a complete axiomatic characterization. On the one hand, the meaning of the deontic concepts is given by the axioms governing their use. On the other hand, a corollary to completeness is consistency. There is a guarantee that the framework is consistent. Without such a guarantee, the move to the implementation level would be pointless: an inconsistent framework could be as easily implemented as a consistent one, but it would be useless. It should be remembered that in the early decades of the 20th century the advent of logic was mainly motivated by such foundational questions of mathematics as the question of how to establish the consistency of arithmetics. Logic is the only tool available for this task.

2.4 Is normative reasoning special?

Logic is a very broad field. There are many different logics around, all differing in language, ontological commitments, epistemological commitments, *etc.* One of these logics, or classes of logics, is deontic logic.

Some people make a distinction between logics that study the notion of inference itself, and logics that use logical inference to model reasoning about a phenomena. Examples of the latter are temporal logic and epistemic logic, and examples of the former are (non-classical) logics like intuitionistic logic, relevance logic and linear logic. For example, intuitionistic reasoning prescribes an *alternative way* to come from arbitrary premisses to entailed conclusions. The same holds for relevance logic, and other alternatives to classical logic. The question we raise here is whether or not deontic reasoning is special in this very same sense, that is, does (or should) deontic logic aim at systems that give an alternative way to come from *arbitrary* premises to their entailed conclusions? In this sense, are deontic logics non-classical?

Though some philosophers seem to pursue the first position where deontic logic is special, many computer scientists see deontic logic as aimed at designing formal systems for coming from *deontic* premises to entailed *deontic* conclusions. But, so the pragmatic argument goes, the logic governing these entailment relations can itself be quite classical. If that is true, then the burden of the deontic logician can shift from designing alternative deontic inference mechanisms to designing rich enough, but classical languages for specifying deontic conditions.

Let us take as an example the famous Chisholm 'paradox' [45], consisting of the sentences (1) "you should help", (2) "if you help you should tell you will", (3) "if you do not help you should not tell you will", (4) "you do not help". We can have at least two views on the way to approach it. The first is that we somehow need to find the right general deontic inference procedure to come to the right conclusion (you should not tell you will help) starting from the propositions representing the sentences. We might for instance claim that the conditionals in the example (sentences 2 and 3) have to be dealt with in a special, deontic way. The second view is that the propositions that make up the example hide a lot of structure (temporal structure, agency, intention) that should be made explicit. After this structure has been made explicit, in classical logics of time, agency or intention, then the deontically correct conclusion follows by classical reasoning.

This relates directly to another interesting issue. Sometimes, in discussions with other logicians, deontic logicians have to defend deontic logic against the claim that there is not a single principle of deontic logic that is non-disputed. Indeed, if one aims at designing a 'core' logic of deontic reasoning, one may end up with a very weak system, since for every suggested principle, some deontic logician might raise his hand and come with a concrete scenario and the claim that this is a counterexample. It may be that such counterexamples introduce context that interferes with the deontic reasoning. The solution to such situations would then not be to leave the logical language as it is and adapt (weaken) the logic, but to leave the logic intact and enrich the language to include the formerly hidden concepts.

So, deontic reasoning is special either way. For some it is special, because they are convinced deontic consequence cannot be classical. For others, especially for those adhering to the pragmatic computer scientist point of view, deontic logic is special because of the many different contexts that influence correct deontic reasoning in concrete examples: there is influence from all kinds of modalities, like belief, intentions, action, time, trust, ethical systems, and different views on the phenomenon of agency. Many researchers seem to believe that in order to be practically useful for computer science, deontic logic should incorporate more than just deontic modalities.

3 Background

We discuss research issues in deontic logic and NorMAS. We focus on two actual topics: norm change and proof methods. Subsection 3.3, titled *Norm Change*, shows the need to apply the tools of logic to reason about the dynamics of normative systems. On the one hand this part discusses the changes of a normative system in time, analyzing legal phenomena such as *ex tunc* and *ex nunc* regulations, on the other hand it presents a classical AGM like approach to norm change, treated as a special case of theory change. Subsection 3.4, titled *Proof Methods*, surveys the proof methods employed in deontic logic so far. they have been given less attention in philosophical logic, but they play a central role in computer science.

3.1 Current research trends in deontic logic

This section describes the different problems that are addressed in deontic logic, and gives a short literature survey. A more comprehensive overview of the state of the art can be found in the forthcoming *Handbook of Deontic Logic* [61].

3.1.1 Norm without truth

A first problem is to reconstruct deontic logic in accord with the idea that norms are neither true nor false. There are two approaches.

The mainstream approach is to reconstruct deontic logic as a logic of normative propositions. The idea is that, though norms are neither true nor false, one may state that (according to the norms), something ought to be done: the statement "John ought to leave the room" is, then, a true or false description of a normative situation. Such a statement is usually called a normative proposition, as distinguished from a norm. The Input/Output (I/O) framework of Makinson and van der Torre [91], and the bi-modal system NOBL due to Åqvist [11], are two different reconstructions of deontic logic as a logic of normative propositions, thus conceived.

The other approach consists in reconstructing deontic logic as a logic of imperatives. This approach is documented in Hansen [70, 71], to which the reader is referred for further details.

3.1.2 Reasoning about norm violation

The system SDL has one modality OA and one accessibility relation R, where xRy means that y is an ideal world relative to x. Unfortunately such a system is not adequate for representing contrary to duty obligations. We also note that some obligations have a temporal aspect to them and that even in the case where there is no real temporal aspect, there is nevertheless a progression along the axis of violations. The Chisholm example has both a temporal progression $\pm y_0 < ... < \pm y_n$ and a violation progression $\pm x_0 < ... < \pm x_n$. We can add the "temporal" relation R with the modalities Nec and Y (yesterday), to stand and represent any type of progression. Nec is flexible enough to do that. We now have models of the form (S,R, R') where the R' ideal worlds are dispersed among the other worlds. In such semantics and language, we can express the general Chisholm set and more. The intuitive concept is that when we have a set of obligations involving both real temporal progression and violation progression, we try to move along a path which will satisfy the obligations, by trying to pass through various ideal worlds in the correct way. Such a logic will have no paradoxes, because the facts correspond to families of paths and the contrary to duty obligations are wffs of the language.

3.1.3 Normative conflicts

There are two main questions here. The first one is: how can deontic logic accommodate possible conflicts between norms? The first systems of deontic logic precluded the possibility of any such conflict. This makes them unsuitable as a tool for analyzing normative reasoning. Different ways to accommodate normative conflicts have been studied over the last fifteen years. A comparative study of them can be found in Goble [65].

The second question is: how can the resolution of conflicts amongst norms be semantically modeled? An intuitively appealing modeling approach consists in using a priority relation defined on norms. There have been several proposals to this effect, and the reader is referred to the discussions in Boella and van der Torre [31], Hansen [70, 71], Horty [78] and Parent

[111]. An open question is whether tools developed for so-called non-monotonic reasoning are suitable for obligations and permissions.

3.1.4 Time

Most formalisms do not have temporal operators in the object language, nor do they have, in their standard formulation, an interpretation in temporal models. Yet for several scenarios and decisions involving deontic reasoning, the temporal aspect of the reasoning seems crucial, and several researchers have sought to study logics for the interactions between temporal and deontic modalities. The research question is: what is the relation between deontic conditionals and temporal deontic modalities?

Two natural concepts to be considered are 'validity time' and 'reference time' of an obligation, prohibition or permission. The validity time is the point in time where a deontic modality is true (surpassing the issue of section 3.1.1 here we simply assume normative modalities have truth values relative to some coherent body of norms that is left implicit) and the reference time is the point in time the obligation, prohibition or permission applies to. For instance, we can have the obligation now (validity time) to show up at the dentist's tomorrow (reference time).

Systems dealing with these temporal differences have been studied, for instance, in [12, 135]. Subtleties in expressing deontic temporal statements involving deontic deadlines have been studied in [40, 37].

3.1.5 Action

We often think of deontic modalities as applying to actions instead of states of affairs. The problems arising in this area are the following: how do we combine deontic modalities with logics of action? How do deontic and action modalities interact. Which action formalisms are best suited for a deontic extension?

Two approaches to deontic action logic prominent in the literature are dynamic deontic logic [100] and deontic *stit* logic [77]. In dynamic deontic logic normative modalities are reduced to dynamic logic action modalities by using violation constants. Prohibition, for instance, is modeled as the dynamic logic conditional assertion that if the action is executed, a violation will occur. In deontic *stit* logic, the perspective on action is different. Where in dynamic logic actions are objects that are given proper names in the object language, in *stit* logic actions derive their identity from the agent(s) executing them and the effect they achieve. This allows for a proper theory of agency, ability and joint ability. In [77] normativity is introduced in *stit* theory by means of a deontic ideality ordering. But the alternative of violation constants has also been used in the *stit* context [22, 38].

3.1.6 Permissive norms

For a long time, it was naively assumed that permission can simply be taken as the dual of obligation, just as possibility is the dual of necessity in modal logic. Something is permitted if its negation is not forbidden. Nowadays in deontic logic a more fine-grained notion of permission is used. The notions of explicit permission, dynamic permission, and permission as exception to a pre-existing obligation are also used. (A dynamic permission is forward-looking, and is like a constitutional right – it sets limits on what can be forbidden). One main finding is that these normative concepts can all be given a well-defined semantics in terms of Input/Output logic [93, 33, 134, 133]. The main open problem concerns their proof-theory, which is still lacking.

3.1.7 Constitutive norms

So-called regulative norms describe obligations, prohibitions and permissions. So-called constitutive norms make possible basic 'institutional' actions such as the making of contracts, the issuing of fines, the decreeing of divorces. Basically they tell us what counts as what for a given institution. As pointed out in [32], constitutive norms have been identified as the key mechanism to normative reasoning in dynamic and uncertain environments, for example to realize agent communication and electronic contracting.

The paper [82] by Jones and Sergot is often credited for having launched the area. There the counts-as relation is viewed as expressing the fact that a given action "is a sufficient condition to guarantee that the institution creates some (usually normative) state of affairs." A conditional connective $\Rightarrow_s$ is used to express the "counts-as" connection holding in the context of an institution s.

When defining constitutive norms, the main issue is in defining their relation with regulative norms. To this end, Boella and van der Torre [30] use the notion of a logical architecture combining several logics into a more complex logical system, also called logical input/output nets (or lions).

3.2 Current research trends in NorMAS

3.2.1 New standard for deontic reasoning

A normative system is used to guide, control, or regulate desired system behavior. We can distinguish four traditional ways to look at normative reasoning in the deontic logic literature. Von Wright's system KD [145] distinguishes good and bad, or right and wrong. Anderson's reduction represents norms by their violation conditions. Hansson's preference-based semantics [72] makes it possible to represent tradeoffs among norms. Makinson [90] criticizes the hegemony of modal logic and proposes an alternative iterative approach. His iterative detachment approach and alternative candidates for a new standard represent the norms explicitly. In the *Handbook of deontic logic*, which is currently in preparation, the classical modal logic framework is mainly confined to the historical chapter. A chapter presents the alternatives to the modal framework, and three concrete approaches, input/output logic, the imperativist approach, and the algebraic conceptual implication structures or cis approach. There are also other candidates for a new standard, such as nonmonotonic logic or deontic update semantics.

Deontic formalisms should be able to capture more applied scenarios. For example, input/output logic is relatively abstract, so can it solve the problems left open by SDL? Just like SDL has been extended with all kinds of things, also input/output logic or abstract normative systems can be extended with all kinds of things. For several classical problems it has been shown that the input/output logic framework can give new insights. In my order of preference. It has been shown that the input/output logic approach to permission, the most classical problem of normative reasoning, has been a big step forward. Not only conceptually distinguishing kinds of norms, but also providing proof systems. Second, it has been shown that the existing semantic solutions to contrary-to-duty reasoning and dilemmas can be reproduced, leaving to a better framework for their formal analysis and comparison (such as proof rewriting techniques). Third, it has been shown that priorities among rules can be studied more systematically. Fourth, it has been shown that reasoning about obligations and time can be done more systematically, including proof systems. Input/output logic completely ignores agent interactions (which are fundamental for – say – social norms). The same holds for all other approaches. There is a very important challenge here, but also here it is crucial to have norms explicitly in the language. There are several contributions on games and deontic logic, but they do not make norms explicit.

3.2.2 The internal and the external

When an individual or a group of individuals is confronted with a number of possible choices, often the question arises of what that individual should do. In the history of deontic logic two perspective have been taken in modeling these type of concepts:

- In the first, norms assume an internal or utilitarian character: actions that are obligatory for a player (or a group of players) are those that are best for the player itself (or, in a general sense, meet the preferences of some players).
- In the second, norms assume an external or systemic character: choices are judged against predetermined interests, specified from outside the system. This is the classical view of deontic logic, which has its roots in Anderson's work, and has been explicitly connected to agency by Meyer's Dynamic Deontic Logic framework.

3.2.3 Expectations

Much MAS research has investigated the use of *commitments* and *norms* to provide social semantics and control to interactions within societies of agents [35, 55, 142, 29]. These constructs both represent socially contextualised constraints on the future behavior of agents, with the fulfilment or violation of these constraints having significance within some formal context (a specific formalised interaction protocol or the code of conduct of a society). Stripping away the social context, we are left with a less formal type of constraint: the expectations that an agent has about the future.

Expectations represent a potential future state of affairs that an agent has an interest in tracking over time (which may be represented by an explicit goal or some computational mechanism that implicitly embodies that goal). While they can be seen as the core aspect of commitments or instantiated norms, they can also arise for less formal reasons. For example, short-term team tactics in sport are based around expectations about the behaviours of other team members. Also, an agent may plan its practical reasoning around expectations that are justified by its own experience, but which need to be tracked in case they turn out to be violated in some situations.

3.2.4 Norms and argumentation

In law, Bench-Capon *et al.* present how argumentation theory has been used in legal reasoning. For instance, legal disputes arise out of a disagreement between two parties and may be resolved by presenting arguments in favor of each party's position. These arguments are proposed to a judging entity, who will justify the choice of the arguments he accepts with an argument of his own, with the aim to convince the public. The common conclusion shared by such works is that argumentation has the potential to become a useful tool for people working in the legal field. Even if a common answer from lawyers when they are asked about what argumentation theory can do for them is that it can be used to deduce the consequences from a set of facts and legal rules, and to detect possible conflicts, there is much more in argumentation. Following the example proposed by Bench-Capon et al., a case is not a mere set of facts, but it can be seen as a story told by a client to his lawyer. The first thing the lawyer does is to interpret this story in a particular legal context. The lawyer can interpret the story in several different ways, and each interpretation will require further facts to be obtained. Then the lawyer has to select one of the possible interpretations, she has to provide arguments to persuade the judging entity of the client's position, and to rebut any further objection. The major topics that emerge as relevant in norms and argumentation include, among others, case based reasoning, arguing about conflicts and defeasibility in rule based systems, dialogues and dialectics, argument schemes, and arguing about the successfulness of the attacks.

3.2.5 Logics for MAS and NorMAS

Several logical systems have been proposed in the last twenty years to model the properties of agents, multi-agent systems (MAS) and normative multi-agent systems (NorMAS). Among them we should mention Propositional Dynamic Logic PDL, Computational Tree Logic CTL , Coalition Logic CL and Alternating-time Temporal Logic ATL, STIT logic (the logic of "seeing to it that") by Belnap, Horty and coll., Dynamic Logic of Agency DLA. Some relationships between these different logical systems have been studied. For instance, it has been shown that both CL and CTL are fragments of ATL, that the 'strategic' variant of STIT logic embeds ATL and that DLA embeds both CL and STIT.

3.2.6 Visualizing normative reasoning

Tosatto et al. [136] promote the use of visual reasoning formalisms for normative reasoning, and insist that the formalism should have a clear and unambiguous semantics. Moreover, they believe that a visual formalism is best accompanied by a logical one, and they therefore refer to visualization of normative reasoning rather than a visual reasoning formalism. There are some related approaches. For example, for business processes there is Declare by van der Aalst's research group [113] (see also Marco Montali's book [101]), and Baldoni et al.'s proposal for commitments [21].

3.3 Norm change

There are two competing theories of norm change, developed as branch of theory change, and as theory of legal dynamics.

3.3.1 Revision of a set of norms

In general, a code G of regulations is not static, but changes over time. For example, a legislative body may want to introduce new norms or to eliminate some existing ones. A different (but related) type of change is the one induced by the fusion of two (or more) codes as it is addressed in the next section.

Little work exists on the logic of the revision of a set of norms. To the best of our knowledge, Alchourrón and Makinson were the first to study the changes of a legal code [8, 9]. The addition of a new norm n causes an enlargement of the code, consisting of the new norm plus all the regulations that can be derived from n. Alchourrón and Makinson distinguish two other types of change. When the new norm is incoherent with the existing ones, we have an *amendment* of the code: in order to coherently add the new regulation, we need to reject those norms that conflict with n. Finally, *derogation* is the elimination of a norm n together with whatever part of G implies n.

In [8] a "hierarchy of regulations" is assumed. Few years earlier, Alchourrón and Bulygin [6] already considered the *Normenordnung* and the consequences of gaps in this ordering. For example, in jurisprudence the existence of precedents is an established method to determine the ordering among norms.

However, although Alchourrón and Makinson aim at defining change operators for a set of norms of some legal system, the only condition they impose on G is that it is a non-empty and finite set of propositions. In other words, a norm x is taken to be simply a formula in propositional logic. Thus, they suggest that "the same concepts and techniques may be taken up in other areas, wherever problems akin to inconsistency and derogation arise" ([8], p. 147).

This explains how their work (together with Gärdenfors' analysis of counterfactuals) could ground that research area that is now known as *belief revision*. Belief revision is the formal studies of how a set of propositions changes in view of a new information that may cause an inconsistency with the existing beliefs. Expansion, revision and contraction are the three belief change operations that Alchourrón, Gärdenfors and Makinson identified in their approach (called AGM) and that have a clear correspondence with the changes on a system of norms we mentioned above. Hence, the following question needs to be addressed:

> How to revise a set of regulations or obligations? Does belief revision offer a satisfactory framework for norms revision?

Some of the AGM axioms seem to be rational requirements in a legal context, whereas they have been criticized when imposed on belief change operators. An example is the *success* postulate, requiring that a new input must always be accepted in the belief set. It is reasonable to impose such a requirement when we wish to enforce a new norm or obligation. However, it gives rise to irrational behaviors when imposed to a belief set, as observed for instance in [62].

On the other hand, when we turn to a proper representation of norms, like in the input/output logic framework, the AGM principles prove to be too general to deal with the revision of a normative system. For example, one difference between revising a set of propositions and revising a set of regulations is the following: when a new norm is added, coherence may be restored modifying some of the existing norms, not necessarily retracting some of them. The following example will clarify this point:

▶ **Example 1.** If we have $\{(\top, a), (a, b)\}$ and we have that c is an exception to the obligation to do b, then we need to retract (c, b). Two possible solutions are $\{(\neg c, a), (a, b)\}$ or $\{(\top, a), (a \wedge \neg c, b)\}$.

Future research must investigate whether general patterns in the revision of norms exist and how to formalize them.

3.3.2 Legal dynamics

One peculiar feature of the law is that it necessarily takes the form of a dynamic normative system [83, 73]. Despite the importance of norm-change mechanisms, the logical investigation of legal dynamics is still much underdeveloped.

As is well-known, the AGM framework distinguishes three types of change operation over theories. Contraction is an operation that removes a specified sentence ϕ from a given theory Γ (a logically closed set of sentences) in such a way as Γ is set aside in favor of another theory Γ^-_ϕ which is a subset of Γ not containing ϕ. Expansion operation adds a given sentence ϕ to Γ so that the resulting theory Γ^+_ϕ is the smallest logically closed set that contains both Γ and ϕ. Revision operation adds ϕ to Γ but it is ensured that the resulting theory Γ^*_ϕ be

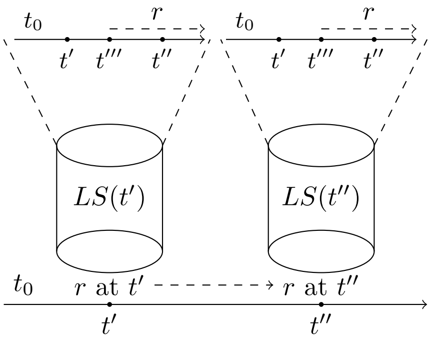


Figure 1 Legal System at t' and t''.

consistent [7]. Alchourrón, Gärdenfors and Makinson argued that, when Γ is a code of legal norms, contraction corresponds to norm derogation (norm removal) and revision to norm amendment.

AGM framework has the advantage of being very abstract but works with theories consisting of simple logical assertions. For this reason, it is perhaps suitable to capture the dynamics of obligations and permissions, not of legal norms. In fact, it is essential to distinguish norms from obligations and permissions [60, 68]: the latter ones are just possible effects of the application of norms and their dynamics do not necessarily require to remove or revise norms, but correspond in most cases to instances of the notion of *norm defeasibility* [68]. Very recently, some research has been carried out to reframe AGM ideas within rule-based logical systems, which take this distinction into account [132, 117]. However, also these attempts suffer from some drawbacks, as they fail to handle the following aspects of legal norm change:

1. the law usually regulate its own changes by setting specific norms whose peculiar objective is to change the system by stating what and how other existing norms should be modified;
2. since legal modifications are derived from these peculiar norms, they can be in conflict and so are defeasible;
3. legal norms are qualified by temporal properties, such as the time when the norm comes into existence and belongs to the legal system, the time when the norm is in force, the time when the norm produces legal effects, and the time when the normative effects hold.

Hence, legal dynamics can be hardly modeled without considering defeasibility and temporal reasoning. Some recent works (see, e.g., [68]) have attempted to address these research issues. All norms are qualified by the above mentioned different temporal parameters and the modifying norms are represented as defeasible meta-rules, *i.e.*, rules where the conclusions are temporalized rules.

If $t_0, t_1, \ldots, t_j$ are points in time, the dynamics of a legal system LS are captured by a time-series $LS(t_0), LS(t_1), \ldots, LS(t_j)$ of its versions. Each version of LS is called a *norm repository*. The passage from one repository to another is effected by legal modifications or simply by temporal persistence. This model is suitable for modeling complex modifications such as retroactive changes, *i.e.*, changes that affect the legal system with respect to legal effects which were also obtained before the legal change was done. The dynamics of norm change and retroactivity need to introduce another time-line within each version of LS (the time-line placed on top of each repository in Figure 1). Clearly, retroactivity does not imply that we can really change the past: this is "physically" impossible. Rather, we need to set a mechanism through which we are able to reason on the legal system from the viewpoint of its current version but *as if* it were revised in the past: when we change some $LS(i)$ retroactively, this does not mean that we modify some $LS(k)$, $k < i$, but that we move back from the perspective of $LS(i)$. Hence, we can "travel" to the past along this inner time-line, *i.e.*, from the viewpoint of the current version of LS where we modify norms. Figure 1 shows a case where the legal system LS and its norm r persist from time t' to time t''; however, such a norm r is in force in LS (it can potentially have effects) from time t''' (which is between t' and t'') onwards.

3.3.3 Dynamic logic approaches

Inspired by recent theoretical and technical developments in the logical study of dynamics—especially the dynamics of informational attitudes such as knowledge and belief[2]—some scholars have proposed models for the 'dynamification' of several kinds of deontic logics.

At the heart of these approaches lies the notion of structure transformation. Let us consider for instance a semantic analysis of obligations based on an ideality ordering among worlds. This semantics lends itself easily to a view of obligation dynamics based on ways of manipulating that ideality ordering. To make a simple example, following [140], the enactment of a command that ϕ be the case could be rendered by the modification of that ideality ordering in such a way that all ϕ-states are ranked as more ideal than all $\neg\phi$-states. The upshot is the modeling of different forms of norm dynamics in terms of different operations on their semantic structures. Other recent contributions along these lines, although based on different structures, are for instance [17, 15].

The advantage of this approach is to maintain a clear link with the underlying logical semantics of deontic notions. How the two perspectives can be technically bridged is very much an open issue.

3.4 Proof systems for deontic logic

This section is devoted to present the different proof systems available for different deontic logic formalisms. The first attempt to proof systems for deontic logic is probably one of Mally [94]. His formal system is based on the classical propositional calculus.

3.4.1 Standard deontic logic KD

Standard deontic logic is the monadic modal logic KD defined as the valid formulas on the class on serial frames. Sahlqvist theorem [119] gives an Hilbert style axiomatization made up with:

- all tautologies of classical propositional logic
- $O(\phi \to \psi) \to (O\phi \to O\psi)$
- $O\phi \to \neg O \neg \phi$
- modus ponens rule
- necessitation rule: from $\vdash \phi$ infer $\vdash O\phi$

A tableau system for KD exists: you extend the tableau system for K by the following rule:

$$\frac{(w\ O\phi)}{(w\ R\ v)(v\ \phi)}$$

[2] See [139] for a recent comprehensive overview

meaning that you add a successor in all labels containing a formula of the form $O\phi$. There exists an implementation of KD and variants ($\mathsf{D4}$, *etc.*) in KED [14]. Such a tableau system is implemented in generic tableau provers [54], [123].

3.4.2 Dyadic deontic logic

Dyadic deontic logic has the dyadic modality $O(\psi \mid \phi)$ as primitive syntactical construct. It is read as "ψ is obligatory conditional upon ϕ being the case". Expressed in BNF notation, the syntax may be defined by

$$\phi ::= p \mid \neg\phi \mid \phi \wedge \psi \mid O(\psi \mid \phi)$$

The semantics is given in terms of Kripke models equipped with a binary relation $\geq$ defined on the universe of the model. The latter relation is used to rank possible worlds in terms of betterness. Compared to Kripke models based on a usual accessibility relation, the main novelty is that the semantics distinguishes various grades of ideality. This is needed to model the notion of CTD obligation, the antecedent of which refers to a sub-optimal situation where a primary obligation is violated. A Kripke model equipped with an accessibility relation uses a binary classification of worlds as good/bad. This binary classification is too rigid to reason about norm violation.

Formally a model becomes a triple $\mathcal{M} = (W, \geq, V)$ where:

- W is a non-empty set of worlds w, w', ...;
- $\geq$ is a binary relation on W; intuitively, $w \geq w'$ may be read as "w is at least as good as w' ";
- V is a valuation, which associates with each possible world a set of propositional formulae (intuitively, the set of those that are true at that world).

Intuitively the evaluation rule for the dyadic obligation operator puts $\bigcirc(\psi \mid \phi)$ true at a world in a model whenever ψ holds in all the best (according to ranking $\geq$) worlds where ϕ is true. This may be expressed as follows:

$w \models \bigcirc(\psi/\phi)$ iff $w' \models \psi$ for all w' such that

$$w' \models \phi \;\&\; \forall w'' \; (w'' \models \phi \Rightarrow w' \succeq w'' \;)$$

There are different systems of dyadic deontic logic depending on the constraints the relation $\geq$ satisfies. In the paper that launched the area, Hansson [72] distinguished three main systems, which he called DSDL1, DSDL2, DSDL3. They are defined as shown in the following table:

	constraints on $\geq$
DSDL1	reflexivity
DSDL2	reflexivity, and limitedness
DSDL3	reflexivity, transitivity, totalness, and limitedness

Roughly speaking, limitedness is the condition that any chain of strictly better worlds is finite. The role of the limitedness condition is, thus, to rule out infinite sequences of strictly better worlds.

It is possible to supplement the logic with additional operators. In particular, given limitedness, the universal modality $\Box$ may be introduced by means of the definition $\Box\phi \equiv$

$O(\bot \mid \neg\phi)$, where $\bot$ denotes a contradiction. It is also possible to add the unary modal operator $Q\phi$. Intuitively, $Q\phi$ says that we are in an ideal situation where ϕ holds. Note that, in a language with $\Box$ and Q has primitive syntactical constructs, $O(\psi \mid \phi)$ is equivalent to $\Box(Q\phi \to \psi)$. The latter just encodes into the syntax the truth-conditions used for the former.

Proof method for these logics is still an active research area. Known results are for DSDL3 mostly. Spohn [131] gave a weakly complete axiomatization for a fragment of DSDL3, with no iterated deontic modalities, and no truthfunctional compound propositions. Åqvist [10] extended Spohn's weak completeness result to the full language of DSDL3. Parent [109] strengthens Åqvist's result into a strong completeness one. A corollary of Spohn's weak completeness result is decidability of the system.

For DSDL2, a partial axiomatization result is available only. Parent [110] provides a strongly complete axiomatization using a language with Q and $\Box$ as primitive syntactical constructs. Åqvist [10] conjectured an alternative axiomatization using the dyadic obligation operator as primitive syntactical construct. This is his system F. His conjecture has not been settled yet.

The axiomatization problem for DSDL1 is still an open one. Åqvist [10] conjectured an axiomatization, which he called E. It is obtained from F by leaving out a suitable axiom. Åqvist's conjecture has not been settled yet.

Rönnedal gave 16 different tableau proof systems for different variants of dyadic deontic logic [116] where he changes the semantics: he works with a usual accessibility relation indexed by sentences. It remains to be seen what a tableau construction would be like in the original setting. Furthermore, the termination problem is not discussed.

3.4.3 See-to-it-that logic

See-to-it-that logic (*stit*) is a framework to deal with agency. Although it can be (and often is) studied completely independent of any normative context, the connection with normative issues is never far away. Agency and normativity are so closely entangled that this justifies looking at proof systems for *stit* in the context of this chapter.

See-to-it-that logic typically provides modal constructions $[J : stit]\phi$ meaning that the group of agents J sees to it that ϕ is true. The construction $[\emptyset : stit]\phi$ for empty set coalition stands for 'ϕ is necessarily true' and the operator $[\emptyset : stit]$ is called historical necessity. The semantics is given in terms of branching time and choice structures and the reader may refer to [23] for details. We give here the semantics of *stit* without time operator given in terms of Kripke structure [74].

Let AGT be a finite set of agents. A Kripke model for the logic *stit* is a tuple $\mathcal{M} = (W, R, V)$ where:

- W is a set of points;
- R is a mapping associating to every agent $i \in AGT$ an equivalence relation R_i on W such that for all $(w_1, w_2, \ldots, w_n) \in W^n$, $\bigcap_{i \in AGT} R_i(w_i) \neq \emptyset$;
- V is a valuation function.

Intuitively, R_i is nothing more than the equivalence relation corresponding to the choice of agent i. When $u \in R_i(w)$ then agent i's current choice at w cannot distinguish between w and u. The truth conditions are given by:

- $\mathcal{M}, w \models [J : stit]\phi$ iff for all $u \in \bigcap_{i \in J} R_i(w)$, we have $\mathcal{M}, u \models \phi$.

There exists a variant of *stit* called deliberate *stit* providing only the historical necessity and constructions $[J : dstit]\phi$ standing for 'J deliberatively sees to it that ϕ is true'. The construction $[J : dstit]\phi$ is defined as a macro of $[J : stit]\phi \wedge \neg[\emptyset : stit]\phi$.

A Hilbert axiomatization is said to be orthodox if the axiomatization is defined by a finite set of axioms schemas (then all instances of the axioms where we substitute any proposition by a formula are theorems) and the two following rules: modus ponens and necessitation rule. Unfortunately there is no so called orthodox finite Hilbert axiomatization for Chellas' *stit* if there are more than 3 agents in the system [74].

Nevertheless when we consider syntactic fragments of Chellas' *stit*, there may exist some Hilbert axiomatizations. For instance, there exists an orthodox axiomatization for the individual *stit* that is the fragment where we only allow construction $[\emptyset : stit]\phi$ and constructions $[\{i\} : stit]\phi$ for individual agents. For individual deliberative *stit* ($[\emptyset : stit]\phi$ and $[\{i\} : dstit]\phi$ for all agents i), Xu's gave an orthodox finite Hilbert axiomatization in [150]. A finite axiomatization for individual Chellas' *stit* and an alternative axiomatization for deliberative individual *stit* may be found in [20]. The axiomatization of individual Chellas' *stit* is given by:

- $S5$ axioms for each modality $[\emptyset : stit]$ and $[\{i\} : stit]$;
- $[\emptyset : stit]\phi \rightarrow [\{i\} : stit]\phi$;
- $\Diamond[\{1\} : stit]\phi_1 \wedge \ldots \Diamond[\{n\} : stit]\phi_n \rightarrow \Diamond([\{1\} : stit]\phi_1 \wedge \ldots [\{n\} : stit]\phi_n)$.

where $\Diamond$ is the dual operator of $[\emptyset : stit]$. The last axiom is interesting and states the independence of agents.

Then, if the set of allowed coalitions in the language are $\emptyset$, $\{i\}$ then there is an axiomatization. Schwarzentruber [122] exhibits a bigger set of allowed coalitions in the language such that there exists a finite Hilbert axiomatization. Lorini and Schwarzentruber [89] exhibit a fragment of Chellas' *stit* logic, with a restriction on the modal depth and an axiomatization for it is given. An axiomatization of *stit* plus the linear temporal logic operators is given by Lorini [88]. Non-terminating tableaux for deliberative individual *stit* are given by Wansing [147].

3.4.4 Other formalisms

In general methods are imported from other areas, in particular from conditional logic and the logic of counterfactuals [86], and van der Torre and Tan [143] uses Boutilier's axiomatization in modal logic [36] to represent DSDL3 and several other logics such as Prohairetic Deontic Logic.

Meyer [100] proposes an approach based on Propositional Dynamic Logic PDL with an atom to designate a violation situation. He gives the Hilbert style axiomatization of PDL. Balbiani [19] also proposes an alternative deontic logic based on PDL.

Goble [66] gave strongly complete axiomatizations for logics incorporating multiple accessibility relations and multiple betterness ranking on alternative worlds to represent distinct normative standards The language uses two monadic modal operators: $O_e\phi$ and $O_a\phi$. The first says that there exists a normative standard in which ϕ is obligatory, and the second says that it is so according to all the normative standards. There are open axiomatization problems for the dyadic counterparts to these modalities.

Alternative kinds of proof systems are developed in input/output logic [91, 92].

4 Research challenges

This section presents a number of fundamental challenges for deontic logic and normative systems. They are represented by research areas with contiguous interests – such as the study of norms – that we believe constitute a frontier where the machinery of deontic logic and normative systems can show its added value.

Subsection 4.1, titled *Norms and Games*, presents a summary of game-theoretical approaches to norms. In particular it distinguishes between two understanding of norms in the field: norms as rules of the game, the so-called mechanism design perspective; and norms as equilibria, the so-called stable state perspective. We believe that deontic logic and normative systems, applying their reasoning tools, can make explicit several foundational issues in norms and games.

Subsection 4.2, titled *Norms and Responsibility*, introduces two dimensions of responsibility in multi-agent systems, namely: (i) responsibility as in 'who did it?', which refers to agents performing an action violating some prescription, (ii) responsibility as in 'who is to blame?' referring to the agents who are instead to be accountable for the damage brought about. The two dimensions do not necessarily coincide, but only a precise modelling of their properties and their consequences can help establishing responsibility in a normative system.

Subsection 4.3, titled *Abstract and Concrete*, shows that normative concepts do not share the same level of detail. We go from extremely general laws, such as constitutional rights, to extremely concrete ones, such as civil law regulations. This part presents an interesting connections between the level of concreteness of regulations and the type of actions that they recommend and sketches an interesting similarity with the logics for ability needed to reason about them.

Subsection 4.5, titled *Visualization*, argues that unlike several important areas of multi-agent systems – such as argumentation – deontic logic and normative systems still lack a representation that is easy to visualize and work with – such as Dung graphs for the case of argumentation.

Subsection 4.6, titled *Proof Methods*, argues about the need to have more general results in the field, such as systematic tableau methods and correspondence results. For instance several important formalism employed in the area of deontic logic, such as dyadic deontic logic, are still not well understood in terms of better-behaved normal modal logics. Bridging the gap would allow to transfer the variety of results available for the latter – such as Salqvist completeness – to the former.

Subsection 4.7, titled *From Deontic Logic to Norms and Policies*, discusses the relation between the formalisms employed to reason about computer systems and their actual implementation. The chapter touches upon subjects such as policies, that are widespread in the practice of disciplines such as security and software engineering.

Subsection 4.8, titled *Expectations*, starts from the observation that norms can have an impact on the practical reasoning of individual agents. Then it goes on arguing that a norm-aware agent must consider the constraints that the prevailing norms impose on its own future behaviour, but it can also benefit by considering how those norms constrain the behaviour of other agents.

Subsection 4.9, titled *Agreement Technologies* presents challenges for Agreement Technologies. In particular it revisits the metaphor of sandbox, which sees methods and mechanisms from the fields of semantic alignment, norms, organization, argumentation and negotiation, as well as trust and reputation are part of a "sandbox" to build software systems based on a technology of agreement. This parts presents an input/output perspective on it.

	C	D
C	2, 2	0, 3
D	3, 0	1, 1

	L	R
L	1, 1	0, 0
R	0, 0	1, 1

Figure 2 Prisoner's dilemma (with C=cooperate and D=defect) and Coordination game (with L=left and R=right).

4.1 Norm and games

Generally speaking, the contributions in the literature in the intersection between games and norms[3] can be divided into two main branches: the first, mostly originating from economics and game theory [47, 79, 80], exploits normative concepts, such as institutions or laws, as *mechanisms* that enforce desirable properties of strategic interactions; the second, that has its roots in the social sciences and evolutionary game theory [138, 48] views norms as *equilibria* that result from the interaction of rational individuals.

4.1.1 Norms as mechanisms

This section presents the view of norms as constraints that, once imposed on players' behaviour, enforce desirable social outcomes in games. In this view, norms are conceived as the rules of the game[4], and it is the most common approach to norms within the so-called New Institutional Economics[5]. An interpretation of this view from the standpoint of game theory is developed in [79], which models the rules of the game in terms of the theory of mechanism design.

In brief, institutions are seen as collective procedures geared towards the achievement of some desirable social outcomes[79]. An example of them are auctions, *viz.* mechanisms to allocate resources among self-interested players. In many auctions goods are not assigned to the bidder valuing them most as bidders might find it convenient to misrepresent their preferences. In such situations mechanism design can be used to enforce the desirable property of truth telling. For instance, when the bidders submit independently and anonymously and the winner pays an amount equivalent to the bid of the runner-up, truth telling is a dominant strategy.[6] In other words, in a second-price sealed bid auction, independently of the way bidders value the auctioned good, they cannot profitably deviate from stating their preferences truthfully.

Viewing norms as mechanisms assigns to norms the same role as auctions. Just like in auctions, norms are supposed to make no assumptions on the preferences of the participating agents. They merely define the possible actions that participants can take, and their consequences. Slightly more technically, they are *game forms* (or mechanisms), *viz.* games without preferences.

Two aspects of this view are worth stressing. First, it clearly explains the rationale for norms and institutions: they exist to guarantee that socially desirable outcomes are realized as equilibria of the possible games that they support (*implementation*). Second, it

[3] For a more detailed discussion on these topics see [69].

[4] As far as we know, this location has been introduced in [106].

[5] New Institutional Economics has brought institutions and norms to the agenda of modern economics, viewing them as the social and legal frameworks of economic behaviour. See [47] for a representative paper.

[6] This is the so-called Vickrey auction. See [125, Ch. 11] for a neat exposition.

presupposes some sort of infallible enforcement: implementation can be obtained only by assuming that players play within the space defined by the rules, which represents a strong idealization of how institutions really work.[7]

The view of norms as mechanisms is by no means limited to economic analysis of interaction, but it has been also successfully applied in computer science to regulate the behaviour of computer systems. A game-transformation approach has been pioneered by [126] in order to engineer laws which guarantee the successful coexistence of multiple programs and programmers. It has been further explored in the multi-agent systems community in [141], to study temporal structures obeying systemic requirements, and [41] , which has made the role of norms explicit in leading players' behaviour to a desirable outcome.

4.1.2 Norms as equilibria

Starting from the classical problem of the spontaneous emergence of social order, the game-theoretic analysis of norms has focused in particular on informal norms enforced by a community of agents, *i.e. social* norms. From this perspective, the view of norms as Nash equilibria has been first suggested by Schelling [121], Lewis [85] and Ullmann-Margalit [138]. A Nash equilibrium is a combination of strategies, one for each individual, such that each player's strategy is a best reply to the strategies of the other players. Since each player's beliefs about the opponent's strategy are correct when part of an equilibrium, this view of norms highlights the facts that a norm is supported by self-fulling expectations.

However, not every Nash equilibrium seems like a plausible candidate for a norm. In the Prisoner's Dilemma (see Figure 2) mutual defection is a Nash equilibrium of the game without being plausibly considered a norm-based behavior. In fact, the view of norms as Nash equilibria has been refined by several scholars. Bicchieri [24], for instance, has suggested that, in the case of norms conformity is always *conditional* upon expectations of what other players will do. Moreover, in this model, norms are different from mere conventions, in that norms are peculiar of mixed-motives games (e.g. the Prisoner's Dilemma) and operate by transforming the original games into coordination ones.

Another influential view of norms characterized them as devices that solve equilibrium selection problems. A comprehensive and concise articulation of this view can be found in [26] which emphasizes two key features of norms. First, as equilibria, they determine self-enforcing patterns of collective behavior[8], e.g., making cooperation an equilibrium of the (infinitely iterated) Prisoner's Dilemma. Second, since repeated interaction can create a large number of efficient and inefficient equilibria, a norm is viewed as a device to select among them—a paradigmatic example of a game with multiple equilibria is the game on the right in Figure 2, known as the coordination game.

Finally, it has been recently suggested that a norm is best captured as a correlating device that implements a correlated equilibrium of an original game in which all agents play strictly pure strategies [64]. A correlated equilibrium is a generalization of the Nash equilibrium concept in which the assumption that the players' strategies are probabilistically independent is dropped. When playing their part on a correlated equilibrium the players condition their choice on the same randomizing device [18]. Since the conditions under which a correlated

[7] This problematic assumption has been put under discussion extensively in [80].

[8] Self-enforcement is the type of phenomenon captured by the so-called *folk theorem*. The theorem roughly says that, given a game, any outcome which guarantees to each player a payoff at least as good as the one guaranteed by her minimax strategy is a Nash equilibrium in the infinite iteration of the initial game (cf. [108, Ch. 8]).

equilibrium is played are less demanding than those characterizing Nash equilibria, the view of norms as a correlating device seems more plausible. Moreover, the correlating device is seen as a device that suggests separately to each player what she is supposed to do and thus seems to better characterize the prescriptive nature of norms [49]. On the other hand, since such correlating devices are viewed as an emergent property of a complex social system, their origins is left unclear.

Although an equilibrium-based analysis of norms might provide a rationale for compliance, it does not explain how such norms can possibly arise in strictly competitive situations—like the Prisoner's Dilemma. Such explanation can be obtained, on the other hand, by adding an evolutionary dimension to the standard game-theoretic framework, as studied for instance in [129].

4.2 Norms and responsibility

The notion of responsibility has two connotations. On the one hand, there is responsibility as in 'who did it?'. But another use of the term responsibility identifies it with 'who is to blame?'. These two forms of responsibility are closely related, but do not coincide; an employee can be responsible for something that happened in the sense that he/she intentionally conducted the activity, without the employee being responsible for what happened in the legal or moral sense; it might be, for instance, that according to the regulations his/her superior is to blame.

Normative systems define who is to blame for what circumstances under which conditions. A standard example is formed by our systems of law. Other examples are systems of moral values, religious commandments, social conventions, *etc.* A normative system defines (either explicitly, as in the law, or implicitly through the collective beliefs of some society, as in social conventions) if an agent that is responsible for something that occurred or might occur (maybe due to some other agent) is in violation from the point of view of that system. Deontic logic models the reasoning of agents having to make decisions and draw normative conclusions in the context of a normative system (What do I have to do? What am I allowed to do? What am I forbidden to do?). Although deontic logic has now been studied for over 60 years, only fairly recently a connection with agency and different forms of responsibility was made [77][23]. Exploring the logical connections between agency, responsibility and normative systems is one of the challenges for the theory of normative systems in the near future.

Responsibilities can be the result of commitments agents made to other agents (or themselves). Baldoni et al. [96] face the problem of defining control, safety, and responsibility in a chain of commitments.

Part of the challenge is the search for logical theories about *degrees* of responsibility and their associated *degrees* of blame relative to a normative system. In this context it is natural to look at probabilities. When we think of responsibilities, probabilistic action plays a central role [149]. For instance, the responsibility for an action may be related to the (subjective) chance of success for that action; if an agent does not have full control over the outcome of an action it can only be partly responsible for bad outcomes. Also, it is very natural to think of having responsibilities relative to a normative system as having to optimize the chance to obey obligations and having to avoid the risk of violations. With the exception of [39] very little is known about logical models relating probability, agency and normative systems.

4.3 Abstract versus concrete norms

In computer science, abstraction is one of the techniques that can be found in almost any sub-area. For instance, in model checking, abstraction is used to get a handle on the complexity of search spaces. In software design abstraction is used for the traceability of requirements. In planning theory, abstraction is used in HTN planning [57]. In logic, abstraction is used in generalized logic.

It makes sense then to assume that one of the new challenges for normative systems in computer science is to integrate them with abstraction techniques. We briefly discuss three sub-areas where abstraction already plays a modest role and that are promising candidates for coming to a more general view on abstraction and norms.

Abstraction of actions. First, in the area of deontic action logic, two views emerged. The first is Meyer's work on dynamic deontic logic [100]. Here the idea is that deontic action logic can be reduced to dynamic logic plus a violation constant. Dynamic logic is a formalism designed for reasoning about pre- and postconditions of basic and complex programs. So the central element of Meyer's contribution is the claim that reasoning about agentive action can be modeled as reasoning about programs. The second view is the *stit* (seeing to it that) view, and it has its background in philosophy [23]. Here the view on action itself is more abstract; an act is a relation between an agent and what this agent achieves. Now we can see *stit* actions as abstractions of dynamic logic actions. The *stit* view on action is close to the view in HTN planning: using *stit* operators we can reason about abstract action and about how they can be refined into more concrete action. In dynamic logic, this is exactly the other way around: we can reason about concrete basic action and how they relate to more complex action expressed in terms of them.

Abstraction of normative systems. Second, also the difference between, for instance, constitutional law and normal law can be seen as a matter of abstraction. Constitutional law might be seen as setting the general, more abstract stage for normal law to take effect. This is a different form of abstraction, that applies to the normative systems themselves and not to the actions that are regulated by such systems. Whether abstraction of normative systems takes the form of orderings over such systems [8], of meta-level descriptions, or of any other relation between them, is one of the challenges.

Abstraction of norms and normative contexts. Third, the difference between general norms, independent of a particular moment, a particular agent, and a particular choice situation, and specific obligations relative to a point in time, an agent and the action it chooses, can also be easily viewed as a form of abstraction. Here the difference is between concrete norms (obligations) and abstract, more general norms. This form of abstraction does not concern normative systems as a whole and also does also not concern regulated actions. What the relation with these other forms of abstraction is, is again one of the challenges.

4.4 Visualization of normative reasoning

Successful reasoning formalisms in Artificial Intelligence such as Bayesian networks, causal networks, belief revision, dependence networks, CP-nets, Dung's abstract argumentation systems, come with intuitive and simple visualizations. Traditionally deontic logic has been associated with preference orders [72], which have an intuitive visualization too. However, it is less clear how to extend this visualization of pre-orders to other aspects of normative reasoning.

In general, we see two approaches to visualization, depending on the audience for which the visualization is developed. On the one hand, we may aim to illustrate a derivation in all its details, and on the other hand, we may look for an abstract approach that visualizes the rough structure of normative reasoning, hiding the more detailed structure. Such an abstract approach may also be used to summarize a more complex derivation. In this paper we follow the latter approach. We thus aim at a visualization that can be understood by non-experts in normative reasoning.

An intuitive and simple visualization for abstract normative systems is important to make them adopted in real applications. The idea shares the motivation with Dung's argumentation networks for non-monotonic reasoning [56], with visual languages such as UML for object-oriented software engineering[9] [118], and TROPOS-like visual representation of early and late requirements[10] [102, 103], *etc.*

Though we promote the use of visual reasoning formalisms for normative reasoning, we insist that the formalism should have a clear and unambiguous semantics. Moreover, we believe that a visual formalism is best accompanied by a logical one, and we therefore refer to visualization of normative reasoning rather than a visual reasoning formalism.

It would be beneficial if the notation were suited both for printed documents and hand-written notes, or machine-processable and paper-and-pencil, which might mean that shading and dashed circles are best avoided. However, we expect that we need more advanced techniques than just diagrams that can be printed, for example using interactive visualizations.

4.5 Proof methods

4.5.1 Dyadic deontic logic

One first important issue is to resolve the axiomatization problem and the decidability problem for DSDL1 and DSDL2 proposed by Hansson.

In dyadic deontic logic there is no systematic way to obtain an axiomatization from a specific class of frames. This is because of the form the evaluation rule for the dyadic obligation operator has.

Another important issue would be to obtain a general correspondence between axiomatization and constraints on the semantics. The idea is not only to focus on DSDL1, DSDL2 and DSDL3 proposed by Hansson but also new systems. One may wonder what the axiomatization of dyadic deontic modal would be like when the relation $\leq$ is reflexive and euclidian. The aim is to obtain a general result in the same flavour that the Sahlqvist theorem gives such a correspondence in normal modal logic [119].

Another main concern is automated reasoning and especially addressing the satisfiability problem. Among all the existing type of proof systems, tableaux are good candidate for providing algorithms. This is especially true for non-classical logics as modal logic, intuitionistic logic and description logic. Indeed, in tableaux contrary to Hilbert or sequent calculus, most of the rules are deterministic in such that a way that the calculus is guided. Furthermore rules are often designed so that terms that are generated are getting strictly smaller and smaller so that it guarantees that the rewriting process of the proof system is terminating. In that case, we say than tableaux rules are strictly analytic according to Fitting's terminology [58]. If some rules are not strictly analytic the termination is no more guaranteed and there are some loop-check techniques to enforce termination without loosing

[9] http://www.uml.org/

[10] http://www.troposproject.org/node/120

soundness and completeness of the algorithm. This is classical for modal logic S4 of reflexive and transitive frames [75].

Here an important problem is to devise proof systems – for instance tableaux – that can be turned into an algorithm. For this, it would also be interesting to have a generic result: for instance we may treat decidability at once for DSDL1, DSDL2, DSDL3 by providing a set of common tableaux rules extended with specific rules for each logic DSDL1, DSDL2, DSDL3.

4.5.2 *stit* challenges

No proof-theory for *stit* logics integrating the deontic modalities is available. A first challenge would be to devise one such.

In Chellas' *stit* logic, we conjecture that a syntactic fragment of it does not distinguish *stit* models from super additive ones if, and only if there exists an orthodox Hilbert finite axiomatization that generates all the validities of the fragment.

There is also a long avenue concerning the proof system – such as tableaux – that may be transformed into an algorithm for providing effective procedures to deal with the satisfiability problem of some fragments of *stit* logic.

4.6 From deontic logic to norms and policies

The deontic concepts of permission, prohibition and obligation have received attention in disciplines other than deontic logic. This is primarily due to the practicality of the notions studied by deontic logic as these notions regulate and coordinate our lives together, making deontic logic valuable for the study of topics of considerable practical significance such as morality, law, social and business organizations (their norms, as well as their normative constitution), and security systems [98].

Among disciplines where deontic logic is relevant is research on security policy languages and models. Security policy languages aim at the practical specification of security requirements in information systems, whereas security models identify the basic concepts and elements necessary to study and analyze these requirements. Security policies and models are not separated concepts as policy languages are often underpinned by a security model. In security policy languages, permissions and prohibitions are used to specify access control requirements [120, 81, 2] whereas obligations have been used to specify requirements such as availability [53], privacy [104] and usage control [112, 152].

Although studying the same concepts, deontic logic and research on security policy languages and models have different research objectives. As a branch of symbolic logic, deontic logic is primarily interested in the study of *valid* inferences when deontic concepts are considered. More specifically, it is often the case that modal logics and Kripke's possible worlds semantics are used to provide a formal view of the semantics of deontic concepts. This allows the proof of the completeness and soundness of the axiomatic with respect to the possible world semantics and the analysis of various deontic puzzles.

On the other hand, security policy languages are more concerned with the clear specification, analysis and enforcement of a system's security requirements, rather than studying valid inferences from deontic concepts. For this reason, norms in policy languages are generally expressed using less abstract constructs than those used in deontic logics for a clear and intuitive representation of requirements. Policy languages also consider the computational aspects of the language to allow efficient analysis and enforcement of system requirements. For this reason, policy languages are generally formalized using tractable fragments of first-order predicate logic such as Datalog [137, 1].

Another possible way to compare deontic logic and security policies is to consider them as models of deontic concepts. A *model* is typically a symbolic or physical representation of a concept or an object that is intended to simplify the understanding, validation or analysis of the modeled concept or object. A model should be simple with great explanatory power. Since these two requirements are typically contradictory, a trade-off often occurs between the simplicity and efficiency of the model on one hand and its explanatory power on the other hand. Since security models (associated with policy languages) and deontic logics are symbolic representations of the deontic concepts of permission, prohibition and obligation, they may be thought of as models which provide the necessary elements and tools to understand, analyze, and/or enforce deontic concepts. In this context, deontic logics are general models for reasoning about deontic concepts whereas security models are more application-oriented but also more efficient and allow clearer specification of norms.

One research challenge is therefore bringing together research on deontic logic and security policy languages and models. This should be beneficial to both communities as it would add to the practicality of deontic logic and provide a more rigorous and formal foundation to research on security policies and languages. Representative research work that considers the use of deontic logic reasoning style and the specification of practical security requirements are [63, 16].

4.7 Norms and expectations

Norms can have an impact on the practical reasoning of individual agents. A norm-aware agent must consider the constraints that the prevailing norms impose on its own future behaviour, but it can also benefit by considering how those norms constrain the behaviour of other agents. If the agent has reason to be confident that other agents will follow the norms, then its own planning can be simplified by adopting this assumption. However, the agent has an interest in knowing whether the norms are indeed followed by the other agents that it interacts with. If its assumption of their compliance turns out to be unwarranted, then its intentions and plans must be reconsidered. Thus, norms induce *expectations*: anticipations of the future course of events or states of affairs, together with a goal to know whether the anticipated future eventuates. This goal may be represented explicitly [42] or by some monitoring mechanism that implicitly embodies it [51].

An activated norm gives rise to an expectation that represents the core temporal content of an activated norm, where the key concern is under what conditions the expectation becomes fulfilled or violated, and in what form the expectation persists from one state to the next if it has not been fulfilled or violated, and other issues such as the source of the norm and the consequences of violation are left to a more specialised and contextual layer of social reasoning. While the definition of norm fulfilment, violation and persistence is often trivial in many normative languages, especially those in which norms express predicates that should hold in some ideal state, these concepts are more subtle when expectations can have a complex temporal structure, e.g. after paying a magazine subscription I expect to receive an issue each month for a year [51].

While expectations can arise from norms, they can also arise from contracts, from commitments resulting from agent interaction, from joint plans (e.g. in team plays in sport), or simply from agents' own observations of the regularities in their environments. Thus, studying techniques for formally modelling and reasoning about expectations promises to lead to a unified treatment of commitments, norms and other constructs, with the notions of future expectation, fulfilment and violation as a core module. Furthermore, while much work on deontic logic has focused on norms with a propositional content, sometimes with the additional of a deadline, expectations arising for non-normative reasons (e.g. joint plans) may lead to requirements for greater temporal expressiveness. The development of expectation languages and reasoning techniques that meet these requirements can, in a modular account of norms and expectations, also lead to richer representation of the temporal aspects of norms.

4.7.1 Current understanding

Expectations have been studied in a number of different settings, as outlined below.

Castelfranchi and colleagues have studied the role that expectation plays as a form of mental anticipation and its relationship with conventions, commitments, obligations, emotions and trust [43, 42]. In their approach, an expectation combines a belief, representing a mental anticipation of a future state or event, with a goal to know whether the anticipated future occurs as predicted. This can be seen as a precursor to the development of social norms, conventions, and commitments. This work has been applied in the development of a computational model for agents that combines a BDI engine with expectation-driven deliberation and an affective control mechanism based on expectations [114].

Alberti et al. have proposed modelling agent interaction protocols using an explicit representation of expectations [4]. In this approach, protocols are expressed using logical rules defining how future expectations on agents' communicative acts arise from observations of current and past communicative acts. Time is treated using explicit time variables and comparisons between the times of different events. The abductive proof procedure $\mathcal{S}$CIFF [3], which includes positive and negative expectation predicates, is used to verify agents' compliance with protocols. An extension of the event calculus based on $\mathcal{S}$CIFF has been proposed for the specification of the social semantics of agent interactions in terms of commitments, and their run-time verification [44]. It has also been shown how the deontic logic concepts of obligation, prohibition and permission can be mapped to the notion of expectation used in this line of research [5].

Cranefield and colleagues have focused on the temporal aspect of expectations, and developed a logic for modelling the activation, fulfilment and violation of rules of expectations with a rich temporal structure, as well as an associated model checking technique for monitoring expectations [50, 51, 52]. Initial work used a first order metric interval temporal logic (with guarded quantification) [50], but more recent work has focused on a propositional linear temporal logic with future and past time operators and some hybrid logic features. The work has been applied in monitoring agents' expectations in the Second Life virtual world[11] [51, Section 6.2], and the expectation checker has been integrated with the Jason BDI interpreter [115].

Nickles et al. [105] introduced the notion of *expectation-oriented modelling*, in which explicit representations of agent expectations are used both as part of the agent design process and agents' run-time execution. In this approach, agent interactions are specified using a graph-based formalism called *expectation networks* in which nodes represent event occurrences and annotated edges encode information about how the occurrence of events result in expectations of other subsequent events.

Wallace and Rovatsos [146] defined an approach for specifying and executing an agent's

[11] http://secondlife.com

practical social reasoning in terms of expectations. They consider an expectation to be a conditional belief that is associated with a specific test condition that will (eventually) confirm or refute the expectation. The agent's pre-existing knowledge of its social context is encoded by defining, for each expectation, how positive and negative test results will result in the activation and deactivation of expectations. From this, an "expectation graph" is derived, representing the possible transitions between sets of active expectations. Another set of rules is defined to specify when agent actions should be generated based on the agents' current beliefs and queries on the expectation graph.

4.7.2 Open research questions

Some open research questions related to expectations are listed below. Some of these questions apply equally well to norms, but may be usefully addressed by focusing on expectations in the first instance, without the added complications of social context and the questions of how the expectations are initiated and what the consequences of fulfilment and violation are.

- What is the relationship between expectations, commitments and norms? Can existing approaches to modelling commitments and norms be expressed in terms of expectations plus additional social context? Is there a common formal model of expectation that can be seen to underlie a range of approaches?
- How expressive can formal models of expectation be while still allowing tractable run-time monitoring? For example, can an appropriate guarded fragment of first order predicate logic be combined with a temporal logic for convenience of encoding complex expectations? What monitoring techniques can be used with different restrictions on expectation expressiveness?
- What techniques can be used to answer questions such as the following, for different levels of expectation expressiveness? (i) Is a set of expectations consistent? (ii) What outcomes are possible for an agent assuming that a given a set of expectations will be fulfilled? (iii) What plans are consistent with a set of expectations? (iv) How can an agent plan to meet its goals and fulfil any expectations applying to its own behaviour, given a set of expectations?
- What are desirable properties for the semantics of expectation, for example, what closure conditions should apply to a set of expectations? If an agent expects ϕ and also expects ψ, should it also expect $\phi \wedge \psi$? While this might seem plausible, if expectations consist of expected beliefs and goals to track them, as proposed by Castelfranchi, then should the agent have a goal to track the truth of $\phi \wedge \psi$?
- Can existing approaches for defining the model-theoretic semantics for commitments (such as the work of Singh [127]), or normative concepts such as obligation, be adapted to include expectations as a core module?
- How can the connection between expectations and practical reasoning architectures, such as the BDI approach, be formalised?

Below, we sketch out one possible way that the first question above might be addressed.

4.7.3 Towards a formalism linking expectation, commitments and norms

This section briefly (and somewhat informally) illustrates how a logic of expectations can be used to provide common semantics for the fulfilment and violation of commitments and norms with rich temporal content. The presentation is based on the logic of Cranefield and

Winikoff [51]. A shorter overview of the key aspects of the logic is presented by Cranefield et al. [52, Section 3].

Commitments and norms have been formalised and operationalised using a wide range of formalisms and computational mechanisms.[12] This section adopts a combination of the event calculus [84, 124] and the logic of expectations.[13]

The event calculus is a formalism for specifying the effects of actions. Amongst many other uses, it has been used directly or in a modified form to specify and reason about agent interaction protocols in terms of commitments [151] and to define norm-governed multi-agent systems [13, 46].

4.7.3.1 Expectations and commitments

We begin by considering the specification of agent communication acts in terms of commitments, following the ideas (but not the formalism) of Verdicchio and Colombetti [144]. A partial specification in terms of the event calculus might look like this:

$$\begin{aligned}
&\mathit{initiates}(\mathit{inform}(x,y,\phi),\ \mathit{comm}(t,x,y,\phi),\ t) && (1)\\
&\mathit{initiates}(\mathit{request}(x,y,\phi),\ \mathit{precomm}(t,y,x,\phi),\ t) && (2)\\
&\mathit{terminates}(\mathit{accept}(x,y,\phi,t_1),\ \mathit{precomm}(t_1,x,y,\phi),\ t_2) && (3)\\
&\mathit{initiates}(\mathit{accept}(x,y,\phi,t_1),\ \mathit{comm}(t_1,x,y,\phi),\ t_2)\\
&\qquad \leftarrow \mathit{holds\_at}(\mathit{precomm}(t_1,x,y,\phi),\ t_2) && (4)\\
&\mathit{terminates}(\mathit{refuse}(x,y,\phi,t_1),\ \mathit{precomm}(t_1,x,y,\phi),\ t_2) && (5)
\end{aligned}$$

This states that the sending of an inform message from x to y with content ϕ establishes a fluent (dynamic predicate) expressing that a commitment holds from x to y that ϕ is true. A request initiates a precommitment, which is terminated when the request is accepted or refused (the *request* and *refuse* communicative acts include the time that the precommitment was established in order to disambiguate different requests with the same content). If the request is accepted, a commitment is established. The time at which the commitment (or precommitment, if applicable) was established is recorded in the *comm* fluent. This is important for linking commitments (with their additional social context, x and y) to expectations.

We now model the relationship between commitments and expectations:

$$\begin{aligned}
&\mathit{holds\_at}(\boldsymbol{exp}(\mathit{true},\phi,t,\phi),\ t) \leftarrow \mathit{holds\_at}(\mathit{comm}(t,x,y,\phi),\ t) && (6)\\
&\mathit{holds\_at}(\mathit{fulf\_comm}(t_1,x,y,\phi),\ t_2)\\
&\qquad \leftarrow \mathit{holds\_at}(\mathit{comm}(t_1,x,y,\phi),\ t_2) \wedge \mathit{holds\_at}(\boldsymbol{fulf}(\mathit{true},\phi,t_1,\_),\ t_2) && (7)\\
&\mathit{holds\_at}(\mathit{viol\_comm}(t_1,x,y,\phi,\psi),\ t_2)\\
&\qquad \leftarrow \mathit{holds\_at}(\mathit{comm}(t_1,x,y,\phi),\ t_2) \wedge \mathit{holds\_at}(\boldsymbol{viol}(\mathit{true},\phi,t_1,\psi),\ t_2) && (8)
\end{aligned}$$

Here, constructs from the logic of expectations are written in bold. $\boldsymbol{exp}(\lambda,\rho,t,\phi)$ states that ϕ *would be* expected to hold currently if there were a rule of expectation with condition λ and content ρ, due to that rule having fired in the (possibly prior) state at time t. Similarly, ***fulf*** and ***viol*** express the fulfilment or violation of a current expectation. Note that these formulae define expectation, fulfilment and violation *relative* to the rule represented by the first two

[12] An partial survey is given by Cranefield et al. [52].

[13] This combination has not yet been formalised or implemented.

arguments. In any state, a countably infinite number of instances of these formulae will hold (e.g. all possible unconditional rules would result in a current expectation). It is therefore up to an implemented system to only track the expectations that are of relevance.[14] The formulae λ, ρ and ϕ are expressed using a form of linear temporal logic with past operators, but any future states available in the model are ignored when evaluating the condition λ. Once a rule has fired, the formula that is expected to hold (the last argument of ***exp***) is initially ρ. However, this may refer to the future, and thus as long as it is not fufilled or violated in any state, it is partially evaluated and 'progressed' to the following state (e.g. "In the next state, ϕ should hold" progresses to "ϕ should hold"). This means that the expected formula is always expressed from the viewpoint of the present—not the time at which the rule was initially triggered.

Clause 6 states that an unconditional rule of expectation is triggered at the time at which a commitment is created (note that t appears twice in the right hand side). Clauses 7 and 8 state that a commitment is fulfilled (respectively violated) if it currently exists (having been established at some time t) and the corresponding unconditional rule of expectation would have resulted in a current fulfilment (respectively violation) if it had fired at time t. The predicate *viol_comm* has an additional final argument (compared to *fulf_comm*) that encodes the residual formula ψ that was violated in the current state. This is likely to differ from the original commitment after partial evaluation and progression across a number of states occurring between t_1 and the present.

We assume the use of an extended version of the event calculus that incorporates the semantics of ***exp***, including the progression of expectations that are not fulfilled or violated, and can perform on-line and/or off-line determination of fulfilment and violation, e.g. by using the technique of Cranefield and Winikoff [51]. This means we need not explicitly use *holds_at* or *initiates* to define how expectations are progressed. However, we use *holds_at* to define the *first* time at which a rule of expectation becomes relevant to the system, to indicate that it should be tracked starting at that time.

4.7.3.2 Expectations and norms

The approach sketched out above can also be used to express protocol-based norms of institutional power, permission and obligation in terms of expectations. For example, the event calculus based approach of Artikis and Sergot [13] and the related work by Cliffe et al. [46] could be adapted to make use of the ***exp***, ***fulf*** and ***viol*** operators.

In this section we show how the logic of expectations can also be used to define the fulfilment and violation of conditional rule-based norms. We assume that norms are encoded by propositions of the form $\mathit{norm}(\lambda, \rho, \mathit{sanction})$, where λ is the condition under which the norm holds, ρ is a linear temporal logic formula encoding the norm as a constraint on the present and future states of the world, and *sanction* encodes a sanction to be applied if the norm is violated. The sanction is an example of the additional contextual information that might be associated with a norm in contrast to the strictly temporal focus of an expectation. This example uses an additional operator from the logic of expectations: ***truncs***. A formula ***truncs***(ϕ) ("truncate model and evaluate with strong finite model semantics") is true when ϕ can be determined to hold without the use of any future information that might be available in the trace under consideration. This is necessary when checking for fulfilment or violation

[14] In the context of the event calculus, this could be done implicitly by a hypothetical combination of the event calculus with the logic of expectations. Alternatively, explicit fluents could be used in the clauses above to record the currently relevant rules of expectation.

of expectations that might involve future-oriented temporal operators (which may occur nested inside past-oriented operators).[15]

$$
\begin{aligned}
&\mathit{holds\_at}(\boldsymbol{exp}(\lambda, \rho, t, \rho),\ t)\\
&\qquad \leftarrow \mathit{norm}(\lambda, \rho, \mathit{sanction}) \wedge \mathit{holds\_at}(\boldsymbol{truncs}(\lambda),\ t) && (9)\\
&\mathit{holds\_at}(\mathit{fulf\_norm}(\lambda, \rho, t_1,\ \mathit{sanction}),\ t_2)\\
&\qquad \leftarrow \mathit{norm}(\lambda, \rho, \mathit{sanction}) \wedge \mathit{holds\_at}(\boldsymbol{fulf}(\lambda, \rho, t_1, \_),\ t_2) && (10)\\
&\mathit{holds\_at}(\mathit{viol\_norm}(\lambda, \rho, t_1, \phi, \mathit{sanction}),\ t_2)\\
&\qquad \leftarrow \mathit{norm}(\lambda, \rho, \mathit{sanction}) \wedge \mathit{holds\_at}(\boldsymbol{viol}(\lambda, \rho, t_1, \phi),\ t_2) && (11)
\end{aligned}
$$

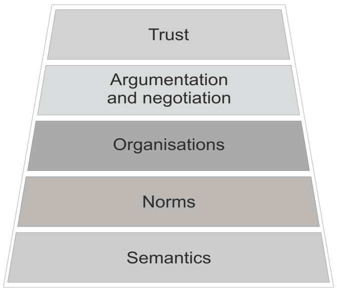
These clauses state that when a conditional norm is triggered, a corresponding expectation is triggered. The fulfilment or violation of an expectation that corresponds to a triggered norm results in the fulfilment or violation of the norm. We assume that norms are static, so do not use *holds_at* to check for their existence. Note that the time the norm was triggered serves to distinguish different fulfilments or violations of the same norm—it appears as the second argument in *fulf_norm* and *viol_norm*. As in the commitments case, we add an additional argument (ϕ) to the predicate *viol_norm* to record the residual violated formula derived from the right hand side of the norm after partial evaluation and progression since the norm was triggered.

4.8 Agreement technologies

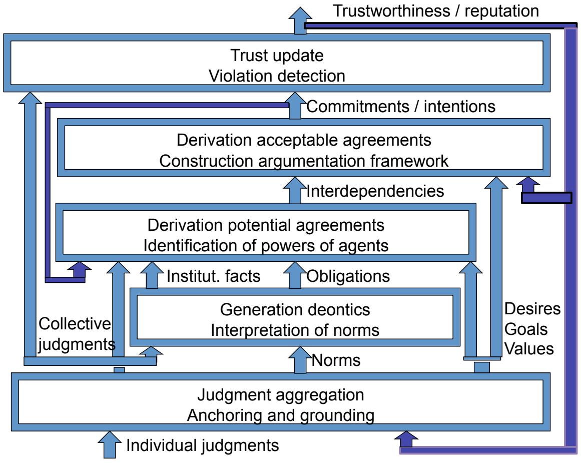
Billhardt *et al.* [25] envision that methods and mechanisms from the fields of semantic alignment, norms, organization, argumentation and negotiation, as well as trust and reputation are part of a "sandbox" to build software systems based on technologies of agreement. Based on a well known definition of coordination as management of dependencies between organisational activities [95], they distinguish the detection of dependencies from taking a decision on which coordination action to apply. Their call-by-agreement interaction method first establishes an agreement for action, and the actual enactment of the action is requested thereafter. The normative context determines rules of the game, *i.e.* interaction patterns and additional restrictions. The so-called agreement technologies "tower" or stack of semantic alignment, norms, organization, argumentation, negotiation, trust and reputation is visualized in Figure 3.

Semantic technologies form the basis to deal with semantic mismatches and alignment of ontologies to give a common understanding of norms or agreements, defining the set of possible agreements. Norms and organizations determine constraints that the agreements, and the processes to reach them, have to satisfy. Organisational structures define the capabilities of the roles and the power and authority relationships among them. Argumentation and negotiation methods are used to make agents reach agreements. The agents use trust mechanisms that summarise the history of agreements and subsequent agreement executions in order to build long-term relationships between the agents. Billhardt *et al.* emphasize that these methods should not be seen in isolation, as they may well benefit from each other.

[15] Formulae of this type might be unlikely to appear as norm conditions, but for completeness we allow for this possibility rather than imposing syntactic restrictions on λ.



■ **Figure 3** Agreement Technologies Tower [25].



■ **Figure 4** Architecture of Agreement Process.

4.8.1 Agreement process

Instead of combining the technologies in a sandbox, Boella and van der Torre [34] introduce a combined agreement process, whose architecture is visualized in Figure 4.

The individual judgments and preferences are grounded in observations and opinions, and aggregated into collective judgments, norms, desires, values and goals. The collective judgments and the norms in force are interpreted [28], and used to generate institutional facts, obligations and permissions. The collective judgments, institutional facts and obligations are used to identify the actions the agents can perform, and their power to satisfy the desires and goals of themselves as well as of other agents. This creates a network of dependencies among the agents. The dependencies among the agents can be used to construct an argumentation framework. Based on the desires and goals of the agents, they negotiate and commit to acceptable agreements. The resulting intentions are fed back into the argumentation and negotiation component, when new agreements are negotiated. The behavior of agents and their commitments is monitored, and in case of detection of violations of agreements the trustworthiness and reputation of the involved agents is updated. The trustworthiness of agents is fed back into the judgment aggregation operator, as well as in the argumentation and negotiation component.

4.8.2 Normative reasoning

The agreement technologies sandbox suggests a bottom-up approach, in the sense that each reasoning technique is studied in its own community, with its own conferences and its own journals. There is a semantic web conference and journal, a deontic logic in computer science and normative multi-agent systems conference, an argumentation conference and journal, and so on. The challenge of reasoning for agreement technologies is to define the relations among them, such that a coherent framework arises. In that sense, the architecture of the agreement process introduced in this paper is more top down. We now discuss how these reasoning techniques can be combined.

4.8.3 Input/output perspective

The input/output perspective on the architecture of the agreement process considers each individual reasoning method as a black box, defined by its input/output behavior, and studies their interaction. Makinson and van der Torre [91] introduce input/output logic for norms to generate institutional facts, obligations and permissions. Bochman [27] uses it to define argumentation in a causal framework. The normative theory can be used to formalize Castelfranchi's theory of dependence networks and social commitments, which has to be extended with a theory or roles. Singh [128] specializes the general way to treat conditionalization in input/output logic for the setting of trust with inferences for completion, commitments, and teamwork that do not arise with conditionals in general, but are important for an understanding of trust.

Missing is an input/output perspective on semantic alignment. Fragments of classical logic such as description logics are used to reason about ontologies, but have less to say about aggregation and alignment. We propose to adopt a judgment aggregation perspective for this step [87].

5 Concluding remarks

For further information, consult the upcoming deontic logic handbook, and the proceedings of the DEON conference series.

References

1 S. Abiteboul, R. Hull, and V. Vianu, editors. *Foundations of Databases: The Logical Level*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1995. ISBN 0201537710.

2 A. Abou El Kalam, S. Benferhat, A. Miège, R. El Baida, F. Cuppens, C. Saurel, P. Balbiani, Y. Deswarte, and G. Trouessin. Organization based access control. *Policies for Distributed Systems and Networks, IEEE International Workshop on*, 0:120, 2003. ISBN 0-7695-1933-4.

3 M. Alberti, F. Chesani, M. Gavanelli, E. Lamma, P. Mello, and P. Torroni. Verifiable agent interaction in abductive logic programming: the sciff framework. *ACM Transactions on Computational Logic*, 9(4), 2008.

4 M. Alberti, M. Gavanelli, E. Lamma, F. Chesani, P. Mello, and P. Torroni. Compliance verification of agent interaction: a logic-based software tool. *Applied Artificial Intelligence*, 20(2):133–157, 2006.

5 M. Alberti, M. Gavanelli, E. Lamma, P. Mello, P. Torroni, and G. Sartor. Mapping deontic operators to abductive expectations. *Computational & Mathematical Organization Theory*, 12:205–225, 2006.

6 C. Alchourrón and E. Bulygin. The expressive conception of norms. in [76] 95–124.

7 C. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.

8 C. Alchourron and D. Makinson. Hierarchies of regulations and their logic. In R. Hilpinen, editor, *New Studies in Deontic Logic*, pages 125–148. Reidel, Dordrecht., 1981.

9 C. Alchourrón and D. Makinson. On the logic of theory change: Contraction functions and their associated revision functions. *Theoria*, 48:14–37, 1982.

10 L. Åqvist. *Introduction to deontic logic and the theory of normative systems.* Bibliopolis Napoli,, Italy, 1987. This paper gives an alternative proof of weak completeness of DSDL3 given by Spohn in 1975.

11 L. Åqvist. Alchourrón and bulygin on deontic logic and the logic of norm-propositions: axiomatization and representability results. *Logique et Analyse*, 51(203):225–261, 2008.

12 L. Åqvist and J. Hoepelman. Some theorems about a tree system of deontic tense logic. In R. Hilpinen, editor, *New Studies in Deontic Logic*, pages 187–221. Reidel, 1981.

13 A. Artikis and M. Sergot. Executable specification of open multi-agent systems. *Logic Journal of the IGPL*, 18(1):31–65, 2010.

14 A. Artosi, P. Cattabriga, and G. Governatori. Ked: A deontic theorem prover. In *Workshop on Legal Application of Logic Programming*, pages 60–76, 1994. This is a prover for "deontic" SDL but in fact for several modal logics.

15 G. Aucher, G. Boella, and L. van der Torre. Prescriptive and descriptive obligations in dynamic epistemic deontic logic. In G. Governatori and G. Sartor, editors, *Proceedings of the 10th International Conference on Deontic Logic in Computer Science (DEON 2010)*, volume 6181 of *LNAI*, pages 150–161, 2010.

16 G. Aucher, G. Boella, and L. van der Torre. A dynamic logic for privacy compliance. *Artif. Intell. Law*, 19(2-3):187–231, 2011.

17 G. Aucher, D. Grossi, A. Herzig, and E. Lorini. Dynamic context logic. In X. He, J. Horty, and E. Pacuit, editors, *Proceedings of LORI 2009*, volume 5834 of *LNAI*. Springer, 2009.

18 R. Aumann. Correlated equilibrium as an expression of bayesian rationality. *Econometrica*, 55:1–18, 1987.

19 P. Balbiani. Logical approaches to deontic reasoning: From basic questions to dynamic solutions. *International Journal of Intelligent Systems*, 23(10):1021–1045, 2008. This article contains a deontic logic made up with PDL.

20 P. Balbiani, A. Herzig, and N. Troquard. Alternative axiomatics and complexity of deliberative stit theories. *Journal of Philosophical Logic*, 37(4):387–406, 2008. This paper gives an axiomatization for both individual Chellas' STIT and individual deliberative STIT.

21 M. Baldoni, C. Baroglio, and E. Marengo. Behavior-oriented commitment-based protocols. In H. Coelho, R. Studer, and M. Wooldridge, editors, *ECAI*, volume 215 of *Frontiers in Artificial Intelligence and Applications*, pages 137–142. IOS Press, 2010. ISBN 978-1-60750-605-8.

22 P. Bartha. Conditional obligation, deontic paradoxes, and the logic of agency. *Annals of Mathematics and Artificial Intelligence*, 9(1-2):1–23, 1993.

23 N. Belnap, M. Perloff, and M. Xu. *Facing the future: agents and choices in our indeterminist world.* Oxford, 2001.

24 C. Bicchieri. *The Grammar of Society: The Nature and Dynamics of Social Norms.* Cambridge University Press, 2006.

25 H. Billhardt, R. Centeno, C. E. Cuesta, A. Fernández, R. Hermoso, R. Ortiz, S. Ossowski, J. S. Pérez-Sotelo, and M. Vasirani. Organisational structures in next-generation distributed systems: Towards a technology of agreement. *Multiagent and Grid Systems*, 7(2-3): 109–125, 2011.

26 K. Binmore. The origins of fair play. *Proceedings of the British Academy*, 151:151–193, 2007.

27 A. Bochman. *Explanatory Nonmonotonic Reasoning.* World Scientific Publishing Company, New York, London, 2005.

28 G. Boella, G. Governatori, A. Rotolo, and L. van der Torre. A logical understanding of legal interpretation. In F. Lin, U. Sattler, and M. Truszczynski, editors, *KR*. AAAI Press, 2010.

29 G. Boella, G. Pigozzi, M. Singh, and H. Verhagen, editors. Special issue on normative multiagent systems. *Logic Journal of the IGPL*, 18(1), 2010.

30 G. Boella and L. van der Torre. A logical architecture of a normative system. In [67] 24–35.

31 G. Boella and L. van der Torre. Permissions and obligations in hierarchical normative systems. In *Proceedings of the 9th International Conference on Artificial Intelligence and Law, ICAIL 2003, June 24-28, Edinburgh, Scotland, UK*. ACM, 2003. Revised version to appear in *Artificial Intelligence and Law*.

32 G. Boella and L. van der Torre. Constitutive norms in the design of normative multiagent systems. In *Computational Logic in Multi-Agent Systems, 6th International Workshop, CLIMA VI*, LNCS 3900, pages 303–319. Springer, 2006.

33 G. Boella and L. van der Torre. Institutions with a hierarchy of authorities in distributed dynamic environments. *Artificial Intelligence and Law*, 16(1):53–71, 2008.

34 G. Boella and L. van der Torre. Reasoning for agreement technologies. Submitted.

35 G. Boella, L. van der Torre, and H. Verhagen, editors. Special issue on normative multiagent systems. *Computational & Mathematical Organization Theory*, 12(2–3), 2006.

36 C. Boutilier. Conditional logics of normality: A modal approach. *Artif. Intell.*, 68(1): 87–154, 1994.

37 J. Broersen. Strategic deontic temporal logic as a reduction to ATL, with an application to Chisholm's scenario. In L. Goble and J.-J. Meyer, editors, *Proceedings 8th International Workshop on Deontic Logic in Computer Science (DEON'06)*, volume 4048 of *Lecture Notes in Computer Science*, pages 53–68. Springer, 2006.

38 J. Broersen. Deontic epistemic *stit* logic distinguishing modes oof *Mens Rea*. *Journal of Applied Logic*, 9(2):127–152, 2011.

39 J. Broersen. Probabilistic action and deontic logic. In *Proceedings of the 12th International Workshop on Computational Logic in Multi-Agent Systems*, volume 6814 of *Lecture Notes in Artificial Intelligence*, pages 293–294. Springer, 2011.

40 J. Broersen, F. Dignum, V. Dignum, and J.-J. Meyer. Designing a deontic logic of deadlines. In A. Lomuscio and D. Nute, editors, *Proceedings 7th International Workshop on Deontic Logic in Computer Science (DEON'04)*, volume 3065 of *Lecture Notes in Computer Science*, pages 43–56. Springer, 2004.

41 N. Bulling and M. Dastani. Verifying normative behaviour via normative mechanism design. In T. Walsh, editor, *IJCAI*, pages 103–108. IJCAI/AAAI, 2011. ISBN 978-1-57735-516-8.

42 C. Castelfranchi. For a systematic theory of expectations. In *Proceedings of the European Cognitive Science Conference*, 2007.

43 C. Castelfranchi, F. Giardini, E. Lorini, and L. Tummolini. The prescriptive destiny of predictive attitudes: From expectations to norms via conventions. In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, pages 222–227, 2003. URL `http://csjarchive.cogsci.rpi.edu/proceedings/2003/pdfs/61.pdf`.

44 F. Chesani, P. Mello, M. Montali, and P. Torroni. Commitment tracking via the reactive event calculus. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI)*, pages 91–96. Morgan Kaufmann, 2009.

45 R. Chisholm. Contrary-to-duty imperatives and deontic logic. *Analysis*, 24:33–36, 1963.

46 O. Cliffe, M. De Vos, and J. Padget. Modelling normative frameworks using answer set programing. In *Proceedings of the 10th International Conference on Logic Programming*

and Nonmonotonic Reasoning, volume 5753 of *Lecture Notes in Computer Science*, pages 548–553. Springer, 2009. URL `http://dx.doi.org/10.1007/978-3-642-04238-6_56`.

47 R. Coase. The problem of social cost. *Journal of Law and Economics*, 1, 1960.

48 J. Coleman. *Foundations of Social Theory.* Belknap Harvard, 1990.

49 R. Conte and C. Castelfranchi. *Cognitive and Social Action.* UCL Press, 1995.

50 S. Cranefield. A rule language for modelling and monitoring social expectations in multi-agent systems. In *Coordination, Organizations, Institutions, and Norms in Multi-Agent Systems*, volume 3913 of *Lecture Notes in Artificial Intelligence*, pages 246–258. Springer, 2006.

51 S. Cranefield and M. Winikoff. Verifying social expectations by model checking truncated paths. *Journal of Logic and Computation*, 21(6):1217–1256, 2011.

52 S. Cranefield, M. Winikoff, and W. Vasconcelos. Modelling and monitoring interdependent expectations. In S. Cranefield, M. B. van Riemsdijk, J. Vázquez-Salceda, and P. Noriega, editors, *Coordination, Organizations, Institutions, and Norms in Agent System VII*, volume 7254 of *Lecture Notes in Artificial Intelligence*, pages 149–166. Springer, 2012.

53 F. Cuppens, N. Cuppens-Boulahia, and T. Ramard. Availability enforcement by obligations and aspects identification. In *ARES '06: Proceedings of the First International Conference on Availability, Reliability and Security*, pages 229–239. IEEE Computer Society, Washington, DC, USA, 2006. ISBN 0-7695-2567-9.

54 L. del Cerro, D. Fauthoux, O. Gasquet, A. Herzig, D. Longin, and F. Massacci. Lotrec: the generic tableau prover for modal and description logics. *Automated Reasoning*, pages 453–458, 2001. This is the first paper about the generic tableau prover Lotrec developped in Toulouse. SDL is implemented.

55 F. Dignum and R. van Eijk, editors. Special issue on agent communication. *Autonomous Agents and Multi-Agent Systems*, 14(2), 2007.

56 P. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.*, 77(2):321–358, 1995.

57 K. Erol, J. Hendler, and D. S. Nau. Htn planning: Complexity and expressivity. In *In Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94)*, pages 1123–1128. AAAI Press, 1994.

58 M. Fitting. *Proof methods for modal and intuitionistic logics.* D. Reidel, Dordrecht, 1983. This book provides proof methods and define "strictly analytic".

59 M. Fowler. *UML Distilled: A Brief Guide to the Standard Object Modeling Language.* Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 3 edition, 2003. ISBN 0321193687.

60 G. P. G. Boella and L. van der Torre. A normative framework for norm change. In *Proc. AAMAS 2009.* ACM, 2009.

61 D. Gabbay, J. Horty, R. van der Meyden, X. Parent, and L. van der Torre (eds). *Handbook of Deontic Logic and Normative Systems.* To appear with College Publications, London.

62 D. Gabbay, G. Pigozzi, and J. Woods. Controlled revision — an algorithmic approach for belief revision. *Journal of Logic and Computation*, 13:3–22, 2003.

63 V. Genovese, D. Garg, and D. Rispoli. Labeled sequent calculi for access control logics: Countermodels, saturation and abduction. In *In Proceedins of the 25th IEEE Computer Security Foundations Symposium*, 2012.

64 H. Gintis. *The Bounds of Reason.* Princenton University Press, 2010.

65 L. Goble. Prima facie norms, normative conflicts and dilemmas. Forthcoming in [61].

66 L. Goble. Preference semantics for deontic logics. part i – Simple models. *Logique et Analyse*, 46:383–418, 2008.

67 L. Goble and J.-J. Meyer, editors. *Deontic Logic and Artificial Normative Systems. 8th International Workshop on Deontic Logic in Computer Scicence, DEON 2006, Utrecht, July 2006, Proceedings.* Springer, Berlin, 2006.

68 G. Governatori and A. Rotolo. Changing legal systems: Legal abrogations and annulments in defeasible logic. *The Logic Journal of IGPL*, 18(1):157–194, 2010.

69 D. Grossi, L. Tummolini, and P. Turrini. Norms in game theory. In *Handbook of Agreement Technologies.* 2012.

70 J. Hansen. Deontic logics for prioritized imperatives. *Artificial Intelligence and Law, forthcoming*, 2005.

71 J. Hansen. Prioritized conditional imperatives: problems and a new proposal. *Journal of Autonomous Agents and Multi-Agent Systems*, 17(1):11–35, 2008.

72 B. Hansson. An analysis of some deontic logics. *Nous*, 3(4):373–398, 1969. This paper presents variant of dyadic deontic logic differing by the constraints in the model, especially DSDL1, DSDL2, DSDL3.

73 H. Hart. *The Concept of Law.* Clarendon, Oxford, 1994.

74 A. Herzig and F. Schwarzentruber. Properties of logics of individual and group agency. *Advances in modal logic*, 7:133–149, 2008. This paper recalls the result for individual Chellas' STIT (axiomatization and complexity results) and states that group Chellas' STIT is undecidable and not finitely axiomatizable.

75 A. Heuerding, M. Seyfried, and H. Zimmermann. Efficient loop-check for backward proof search in some non-classical propositional logics. In *TABLEAUX '96: Proceedings of the 5th International Workshop on Theorem Proving with Analytic Tableaux and Related Methods*, pages 210–225. Springer-Verlag, London, UK, 1996. This article presents the technique of loop-check.

76 R. Hilpinen, editor. *New Studies in Deontic Logic.* Reidel, Dordrecht, 1981.

77 J. Horty. *Agency and Deontic Logic.* 2001.

78 J. Horty. Defaults with priorities. *Journal of Philosophical Logic*, 36:367–413, 2007.

79 L. Hurwicz. Institutions as families of game forms. *Japanese Economic Review*, 47(2):113–132, 1996.

80 L. Hurwicz. But who will guard the guardians? *American Economic Review*, 98(3):577–585, 2008.

81 S. Jajodia, P. Samarati, and V. S. Subrahmanian. A logical language for expressing authorizations. *Security and Privacy, IEEE Symposium on*, 0:0031, 1997. ISSN 1540-7993.

82 A. Jones and M. Sergot. A formal characterisation of institutionalised power. *Journal of IGPL*, 3:427–443, 1996.

83 H. Kelsen. *General Theory of Norms.* Clarendon, Oxford, 1991.

84 R. Kowalski and M. Sergot. A logic-based calculus of events. *New Generation Computing*, 4:67–69, 1986.

85 D. Lewis. *Convention: A Philosophical Study.* Cambridge University Press, 1969.

86 D. Lewis. Semantic analyses for dyadic deontic logic. In S. Stenlund, editor, *Logical Theory and Semantic Analysis*, pages 1 – 14. Reidel, Dordrecht, 1974.

87 C. List and C. Puppe. Judgment aggregation: A survey. In P. Anand, C. Puppe, and P. Pattanaik, editors, *Oxford handbook of rational and social choice.* Oxford University Press, New York, 2009.

88 E. Lorini. A stit logic analysis of commitment and its dynamics (to appear). *Journal of Applied Logic*, 2012. This article contains the axiomatization of STIT + the future operator.

89 E. Lorini and F. Schwarzentruber. A logic for reasoning about counterfactual emotions. *Artificial Intelligence*, 175(3):814–847, 2011. This paper presents a STIT logic for coutnerfactual emotions. It presents a decidable fragment of group Chellas' STIT and a Hilbert axiomatization.

90 D. Makinson. On a fundamental problem of deontic logic. In [99], 29–53.

91 D. Makinson and L. van der Torre. Input-output logics. *Journal of Philosophical Logic*, 29 (4):383–408, 2000.

92 D. Makinson and L. van der Torre. Constraints for input-output logics. *Journal of Philosophical Logic*, 30(2):155–185, 2001.

93 D. Makinson and L. van der Torre. Permissions from an input-output perspective. *Journal of Philosophical Logic*, 32(4):391–416, 2003.

94 E. Mally. Grundgesetze des sollens. *Mind*, 36(141):124–b, 1927. This paper is the first approach to deontic logic. It is based on classical propositional logic.

95 T. Malone and K. Crowston. The interdisciplinary study of coordination. *Computing Surveys*, 26 (1):87–119, 1994.

96 E. Marengo, M. Baldoni, C. Baroglio, A. Chopra, V. Patti, and M. Singh. Commitments with regulations: reasoning about safety and control in regula. In Sonenberg et al. [130], pages 467–474.

97 J. McCarthy. Programs with common sense. In *Proceedings of the Teddington Conference on the Mechanization of Thought Processe*, pages 756–91. Her Majesty's Stationery Office, London: Her Majesty's Stationery Office, 1959.

98 P. McNamara. Deontic logic. In D. Gabbay and J. Woods, editors, *Handbook of the History of Logic*, volume 7, pages 197–289. North-Holland Publishing, Amsterdam, 2006.

99 P. McNamara and H. Prakken, editors. *Norms, Logics and Information Systems.* IOS, Amsterdam, 1999.

100 J.-J. Meyer. A different approach to deontic logic: Deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic*, 29:109–136, 1988.

101 M. Montali. *Specification and Verification of Declarative Open Interaction Models - A Logic-Based Approach*, volume 56 of *Lecture Notes in Business Information Processing.* Springer, 2010. ISBN 978-3-642-14537-7. 1-383 pp.

102 D. Moody. The physics of notations: Toward a scientific basis for constructing visual notations in software engineering. *IEEE Transactions on Software Engineering*, 35(6):756–779, 2009.

103 D. Moody and J. van Hillegersberg. Evaluating the visual syntax of uml :an analysis of the cognitive effectiveness of the uml family of diagrams. In *Software Language Engineering*, volume 5452 of *Lecture Notes in Computer Science*, pages 16–34. . Springer, 2009.

104 Q. Ni, E. Bertino, and J. Lobo. An obligation model bridging access control policies and privacy policies. In *SACMAT '08: Proceedings of the 13th ACM symposium on Access control models and technologies*, pages 133–142. ACM, New York, NY, USA, 2008. ISBN 978-1-60558-129-3.

105 M. Nickles, M. Rovatsos, and G. Weiss. Expectation-oriented modeling. *Engineering Applications of Artificial Intelligence*, 18:891–918, 2005. URL `http://dx.doi.org/10.1016/j.engappai.2005.05.002`.

106 D. North. *Institutions, Institutional Change and Economic Performance.* Cambridge University Press, Cambridge, 1990.

107 J. Odell, H. Van Dyke Parunak, and B. Bauer. Representing agent interaction protocols in UML. In *IN OMG DOCUMENT AD/99-12-01. INTELLICORP INC*, pages 121–140. Springer-Verlag, 2001.

108 M. Osborne and A. Rubinstein. *A Course in Game Theory.* MIT Press, 1994.

109 X. Parent. On the strong completeness of Åqvist's dyadic deontic logic G. *Deontic Logic in Computer Science*, pages 189–202, 2008. This article contains the strong completeness of an axiomatization of dyadic deontic logic DSDL3.

110 X. Parent. A complete axiom set for Hansson's deontic logic DSDL2. *Logic Journal of IGPL*, 18(3):422–429, 2010. Axiomatization for DSDL2.

111 X. Parent. Moral particularism in the light of deontic logic. *Artif. Intell. Law*, 19(2-3): 75–98, 2011.

112 J. Park and R. Sandhu. The $UCON_{ABC}$ Usage Control Model. *ACM Transactions on Information and System Security (TISSEC)*, 7(1):128 – 174, February 2004.

113 M. Pesic, H. Schonenberg, and W. M. P. van der Aalst. The declare service. In A. H. M. ter Hofstede, W. M. P. van der Aalst, M. Adams, and N. Russell, editors, *Modern Business Process Automation*, pages 327–343. Springer, 2010. ISBN 978-3-642-03120-5.

114 M. Piunti, C. Castelfranchi, and R. Falcone. Expectations driven approach for situated, goal-directed agents. In *Proceedings of the 8th AI*IA/TABOO Joint Workshop "From Objects to Agents": Agents and Industry: Technological Applications of Software Agents*, pages 104–111. Seneca Edizioni Torino, 2007.

115 S. Ranathunga, S. Cranefield, and M. Purvis. Integrating expectation handling into BDI agents. In *Programming Multi-Agent Systems*, volume 7217 of *Lecture Notes in Artificial Intelligence*, pages 74–91. Springer, 2012.

116 D. Rönnedal. Dyadic deontic logic and semantic tableaux. *Logic and Logical Philosophy*, 18(3-4):221–252, 2009. TODO.

117 A. Rotolo. Retroactive legal changes and revision theory in defeasible logic. In G. Governatori and G. Sartor, editors, *Proceedings of the 10th International Conference on Deontic Logic in Computer Science (DEON 2010)*, volume 6181 of *LNAI*, pages 116–131. Springer, 2010. ISBN 978-3-642-14182-9.

118 J. Rumbaugh. Notation notes: Principles for choosing notation. *Journal of Object-Oriented Programming*, 8(10):11–14, 1996.

119 H. Sahlqvist. Completeness and correspondence in the first and second order semantics for modal logic. *Studies in Logic and the Foundations of Mathematics*, 82:110–143, 1975. This paper explains a correspondance between constraints on Kripke models and axioms.

120 R. Sandhu, E. Coyne, H. Feinstein, and C. Youman. Role-based access control models. *IEEE Computer*, 29(2):38–47, 1996.

121 T. Schelling. *The Strategy of Conflict.* Oxford University Press, 1966.

122 F. Schwarzentruber. Complexity results of stit fragments (to appear). *Studia logica*, 2011. This paper exhibits synctatic fragments of Chellas' group STIT by restricting the coalitions allowed in the language. Those fragments are finitely axiomatizable.

123 F. Schwarzentruber. Lotrecscheme. *Electronic Notes in Theoretical Computer Science*, 278: 187–199, 2011. This is a paper about some improvements of the tableau prover Lotrec and presents a prototype for it. SDL is implemented.

124 M. Shanahan. The event calculus explained. In M. Wooldridge and M. Veloso, editors, *Artificial Intelligence Today: Recent Trends and Developments*, volume 1600 of *Lecture Notes in Artificial Intelligence*, pages 409–430. Springer, 1999.

125 Y. Shoham and K. Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic and Logical Foundations.* Cambridge University Press, 2008.

126 Y. Shoham and M. Tennenholtz. Social laws for artificial agent societies: Off-line design. *Artificial Intelligence*, 73(12):231–252, 1995.

127 M. Singh. Semantical considerations on dialectical and practical commitments. In *Proceedings of the 23rd National Conference on Artificial Intelligence (AAAI)*, pages 176–181. AAAI Press, 2008.

128 M. Singh. Trust as dependence: a logical approach. In Sonenberg et al. [130], pages 863–870.

129 B. Skyrms. *Evolution of the Social Contract.* Cambridge University Press, 1996.

130 L. Sonenberg, P. Stone, K. Tumer, and P. Yolum, editors. *10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011), Taipei, Taiwan, May 2-6, 2011, Volume 1-3.* IFAAMAS, 2011.

131 W. Spohn. An analysis of Hansson's dyadic deontic logic. *Journal of Philosophical Logic*, 4(2):237–252, 1975. This paper contains a weak completeness result for DSDL3.

132 A. Stolpe. Norm-system revision: theory and application. *Artif. Intell. Law*, 18(3):247–283, 2010.

133 A. Stolpe. Relevance, derogation and permission. In G. Governatori and G. Sartor, editors, *DEON*, volume 6181 of *Lecture Notes in Computer Science*, pages 98–115. Springer, 2010.

134 A. Stolpe. A theory of permission based on the notion of derogation. *J. Applied Logic*, 8 (1):97–113, 2010.

135 R. Thomason. Deontic logic as founded on tense logic. In R. Hilpinen, editor, *New Studies in Deontic Logic*, pages 165–176. Reidel, 1981.

136 S. Tosatto, G. Boella, L. van der Torre, and S. Villata. Visualizing normative systems. In *Procs. of DEON 2012*, LNCS. Springer, 2012.

137 D. Ullman. *Principles of database and knowledge-base systems, Vol. I.* Computer Science Press, Inc., New York, NY, USA, 1988. ISBN 0-88175-188-X.

138 E. Ulmann-Margalit. *The Emergence of Norms.* Oxford: Clarendon Press, 1977.

139 J. van Benthem. *Logical Dynamics of Information and Interaction.* Cambridge University Press, 2011.

140 J. van Benthem, D. Grossi, and F. Liu. Deontics = betterness + priority. In G. Governatori and G. Sartor, editors, *Proceedings of the 10th International Conference on Deontic Logic in Computer Science (DEON 2010)*, volume 6181 of *LNAI*, pages 50–65. Springer, 2010.

141 W. van der Hoek, M. Roberts, and M. Wooldridge. Social laws in alternating time: Effectiveness, feasibility, and synthesis. *Synthese*, 156:1:1 – 19, 2007.

142 L. van der Torre, G. Boella, and H. Verhagen, editors. Special issue on normative multiagent systems. *Autonomous Agents and Multi-Agent Systems*, 17(1), 2008.

143 L. van der Torre and Y.-H. Tan. Contrary-to-duty reasoning with preference-based dyadic obligations. *Ann. Math. Artif. Intell.*, 27(1-4):49–78, 1999.

144 M. Verdicchio and M. Colombetti. A commitment-based communicative act library. In *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 755–761. ACM, 2005.

145 G. von Wright. Deontic logic. *Mind*, 60:1–15, 1951.

146 I. Wallace and M. Rovatsos. Bounded practical social reasoning in the ESB framework. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems*, pages 1097–1104. IFAAMAS, 2009. URL `http://dl.acm.org/citation.cfm?id=1558109.1558166`.

147 H. Wansing. Tableaux for multi-agent deliberative-stit logic. *Advances in modal logic*, 6: 503–520, 2006. This paper presents a non-terminating tableau for individual deliberative-STIT logic.

148 M. Wooldridge. *An Introduction to MultiAgent Systems (2. ed.).* Wiley, 2009. 1-461 pp.

149 R. Wright. Causation, responsibility, risk, probability, naked statistics, and proof: Pruning the bramble bush by clarifying the concepts. *Iowa Law Review*, 73:1001–1077, 1988.

150 M. Xu. Axioms for deliberative stit. *Journal of Philosophical Logic*, 27(5):505–552, 1998. This paper gives an axiomatization for individual deliberative STIT.

151 P. Yolum and M. Singh. Reasoning about commitments in the event calculus: An approach for specifying and executing protocols. *Annals of Mathematics and Artificial Intelligence*, 42:227–253, 2004. URL `http://dx.doi.org/10.1023/B:AMAI.0000034528.55456.d9`.

152 X. Zhang, F. Parisi-Presicce, R. Sandhu, and J. Park. Formal model and policy specification of usage control. *ACM Trans. Inf. Syst. Secur.*, 8(4):351–387, 2005. ISSN 1094-9224.

Computational Models for Normative Multi-Agent Systems

Natasha Alechina[1], Nick Bassiliades[2], Mehdi Dastani[3],
Marina De Vos[4], Brian Logan[1], Sergio Mera[5],
Andreasa Morris-Martin[6], and Fernando Schapachnik[5]

1 School of Computer Science, University of Nottingham
Nottingham, NG8 1BB, UK
{nza,bsl}@cs.nott.ac.uk

2 Dept. of Informatics, Aristotle University of Thessaloniki
54124 Thessaloniki, Greece
nbassili@csd.auth.gr

3 Dept of Information and Computer Sciences, University of Utrecht
Princetonplein 5, 3584CC Utrecht, The Netherlands
M.M.Dastani@uu.nl

4 Dept. of Computer Science, University of Bath
Bath, BA2 7AY, UK
mdv@cs.bath.ac.uk

5 Departamento de Computación, FCEyN, Universidad de Buenos Aires,
Intendente Güiraldes 2160 – Ciudad Universitaria – C1428EGA, Buenos Aires,
Argentina
{smera,fschapachnik}@dc.uba.ar

6 Dept. of Computer Science, University of Guyana
Turkeyen, Greater Georgetown, Guyana
andreasa.morris@uog.edu.gy

Abstract

This chapter takes a closer look at computational logic approaches for the design, verification and the implementation of normative multi-agent systems. After a short overview of existing formalisms, architectures and implementation languages, an overview of current research challenges is provided.

1998 ACM Subject Classification I.2.11 Distributed Artificial Intelligence, I.2.4 Knowledge Representation Formalisms and Methods

Keywords and phrases Norm verification, Computational Architectures for Normative MAS, Programming Normative Systems

Digital Object Identifier 10.4230/DFU.Vol4.12111.71

1 Introduction

Multi-agent systems consist of interacting autonomous agents. While individual agents aim at achieving their own design objectives, the overall objectives of multi-agent systems can be guaranteed by regulating and organising the behaviour of individual agents and their interactions. There have been various proposals for regulating and organising the behaviours of individual agents. Some of these proposals advocate the use of coordination artifacts that are specified in terms of low-level coordination concepts such as synchronization [5, 28]. Other approaches are motivated by organisational models, normative systems, or electronic

Normative Multi-Agent Systems. *Dagstuhl Follow-Ups, Vol. 4.* ISBN 978-3-939897-51-4.
Editors: Giulia Andrighetto, Guido Governatori, Pablo Noriega, and Leendert W.N. van der Torre; pp. 71–92
Dagstuhl Publishing
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany

institutions [35, 36, 49, 57, 62, 31]. These approaches have resulted in what is often called normative multi-agent systems. In such systems, the behaviours of individual agents are regulated by means of norms and organisational rules that are either used by individual agents to decide how to behave, or are enforced or regimented through monitoring and sanctioning mechanisms. In these systems, a social and normative perspective is conceived as a way to make the development and maintenance of multi-agent systems easier to manage. A plethora of social concepts (e.g., roles, groups, social structures, organisations, institutions, norms) has been introduced in multi-agent system methodologies (e.g. Gaia [88]), models (e.g. OperA [34]), specification and modelling languages (e.g. S-MOISE+ [58], INSTAL[25] and ISLANDER [36]) and computational frameworks (e.g. AMELI [35]).

With the significant advances in the area of autonomous agents and multi-agent systems in the last decade, promising technologies for the development and engineering of normative multi-agent systems have emerged. The result is a variety of programming languages, execution platforms, and computational tools and techniques that facilitate the development and engineering of normative multi-agent systems. This chapter provides an overview of computational tools that can be used to develop and build normative multi-agent systems. The main themes of this chapter are the languages for representing deontic notions (such as norms themselves, conflicts of norms, violations of norms, etc.) and the algorithms to implement tools capable of analyzing and verifying norms, and to implement normative system platforms capable of e.g. monitoring norm violations, and implementing agents capable of deliberating about norms.

2 Background

Computational models for norms are essential for the development of normative multi-agent systems. Such models can be used to specify and verify normative multi-agent systems, to implement normative multi-agent systems, or to allow software agents, electronic institutions and organisations to reason about norms at run-time. In this section, we provide an overview of the existing computational models of norms and how they are applied in multi-agent systems.

2.1 Norm Specification and Verification

This section focuses on the use of logic languages and automated reasoning to describe and understand the behavior of normative systems and analyze their properties. Coherence, in the sense of absense of conflicts, is one of the key properties of interest. Under this approach the object under study is typically the normative system itself. This means that the type of questions we can aim at answering using logic machinery are less related to particular executions of a set of normative agents (which is the focus of sections 2.2 and 2.3) and they are closer to understanding the behavior of universal properties or interactions within the system. It also involves being able to provide a mathematical proof of such properties (a proof whose extent, depending on the language and automated methods used, can cover a fragment of the models under consideration or the whole class of them).

Let's provide an example of the type of properties we mentioned and consider a legislator that is drafting a normative system that regulates the traffic of a city. During this process she can pose a series of very natural questions related to the quality of the normative system. For example, are these rules going to force a driver to, at the same time, grant and deny the right-of-way to another vehicle? Can a pedestrian and a vehicle be allowed to cross an intersection simultaneously, and therefore cause an accident? Is the system applicable to all type of vehicles, or is there a “gap” that makes, for example, the speed limit not to be enforced on motorcycles?

The type of questions mentioned above involve dealing with the concept of obligation, permission and prohibition. *Deontic logic* is the field of logic concerned about these concepts and the formalisms we are going to talk about in this section naturally fall under this category. Other related concepts such as right, liberty, power, immunity, etc. can also be included in this definition. Legal theories, at least since Hohfeld in 1913 [54], logic-based approaches like [69] and more recently [80, 81, 75] have attempted to clarify these concepts.

The primal property that is subject of verification is what is known as *coherence*: whenever the set of rules is “contradictory” in any sense, sometimes referred to as “absence of conflicts”. As stated in the literature (e.g., [52]), the problem cannot be simply reduced to logical consistency. To exemplify this, suppose you have a norm stating that it is forbidden to cross red lights, and another that says that it is forbidden but paying a fine is way of “fixing” it. Is there a conflict or not? Another example is a rule that imposes an obligation and a certain reparation as a way of “fixing” things if the obligation is not fulfilled. If another rule prohibits this same action, then there is another potential conflict.

The importance of absence of conflicts in a normative system depends of course on the context, given that different legal instruments operate at different levels of abstraction. Constitutions and parliamentary bills are often very general and their ambiguity is a necessary price to get enough consensus for their approval. However, as we descend the legal hierarchy we find everyday regulations having other requirements: they tend to be operational, describing how a set of actors (or agents) should behave in a daily manner. Ambiguities and gaps are usually a problem. More important, they are meant to be interpreted by common individuals, not specifically trained in Law, and it is seldom the case where conflicts that arise from them are resolved in a Court of Law. This does not mean that they are conflict-free. On the contrary, an important amount of time and effort is wasted in trying to overcome the deficiencies of everyday regulations. This is one place where doing logical verification of norms can add real value.

Much work has been done in exploring deontic-type logics and trying to cope with conflict detection. Many different languages have been proposed with motley theoretical and computational properties. There is no “one-size-fits-all” formalism, and there are many aspects (like complexity, expressiveness, conciseness, a language natural to the context of use, etc.) that live in opposite sides of the scales. Considering the strong computational flavor of this work, what follows aims at presenting the state-of-the-art formalisms currently being developed by the community, focusing on the ones that incorporated such concepts into working tools. There is much valuable work on the theoretical or philosophical side of deontic logic, but this falls outside the scope of this section. Besides a first generation of efforts that were aimed at specific normative sets, somehow hard-coded into the tools (e.g., [50, 20]), we describe what is considered as the current generation of practical formalisms.

The authors in [51] propose a language that deals with automated conflict detection in norms by using a tool that supports ontologies and translates normative propositions into a Prolog program. However, the analysis is restricted to logical contradiction, which has the limitations we already mentioned. In order to describe a real system, where rule interactions are naturally more involved, being able to deal with a more general concept such as coherence is a must.

BCL [46, 47] is a contract specification language based on defeasible logic that is meant for monitoring, allows to build executable versions, can detect conflicts among rules off-line and provides features like clause normalisation. Recently it incorporated support for some temporal reasoning. On the other hand, it does not offer support for specifying background theories, which are very useful for restricting the class of models under consideration and greatly contribute to improve reasoning performance.

$\mathcal{CL}$ [78, 37] is another logical language based on dynamic logic that treats deontic operators as first-class citizens and has support for conditional obligations, permissions and prohibitions, as well as for (nested) contrary-to-duties and contrary-to-prohibitions. It provides its own coherence checker, CLAN [38]. Contextual information has to be encoded as deontic rules. Being based on dynamic logic, certain predicates are hard to express: for instance, it is easy to say that if a book is borrowed, a book should eventually be returned, but it would be difficult to formalize the borrowing of multiple books and their corresponding returning; the correct version being pretty involved.

FL [44, 45] is also a language and companion tool. The language is a wrapper for LTL with some features such as variables and fluents (called *intervals*) to ease expressions. Its main characteristic is that it uses existing LTL model checkers like SPIN [56], DiViNE [14] or NuSMV [23] to do the verification. Authors claim that as there are strong similarities between temporal software specifications and vast amounts of norms, it makes sense to reutilise tools that are being developed and optimised since at least a couple of decades instead of building new ones. On the other hand, it doesn't support deontic operators as first class citizens and somehow entails a notion of total consistency.

All of these tools face the challenge of scale: they have been tested with somehow realistic albeit small case studies. What is needed to cope with work load from real users?

For instance, there is the theoretical question of which complexity class are they in. Experience from other fields suggests this is not enough, as sometimes many problems become practically tractable although their theoretical worst case is not. What can be said about the analysis of norms? In other words, what is their "practical complexity"?

If it is about computing real cases, expressivity also comes into play. Can we say everything that is needed? Is there a simple way or contrived expressions cannot be avoided?

More importantly, a right balance between expressivity and complexity has to be achieved. What other fields have done to achieve the same goal seems promising in this case too. Build usable tools, apply them to real cases (also let others use them), understand what is lacking and what is overflowing, revise the underlying theory, repeat the cycle.

2.2 Computational Architectures for Normative Multi-Agent Systems

In the literature on multi-agent systems, there have been many proposals for specification languages and logics to specify and reason about normative multi-agent systems, virtual organisations, and electronic institutions (e.g., [62, 77, 17, 1]). How to specify, develop, program, and execute such normative systems was one of the central themes that were discussed and promoted during the AgentLink technical fora on programming multi-agent systems (see [30, 29] for the general report of these technical fora). In this section we focus on normative architectures; formalisms that were initially created to support a designer in specifying and verifying normative systems. The next section discusses the middleware and the programming languages needed for running normative systems. We only briefly discuss the ISLANDER and $\mathcal{M}$OISE$^+$ architectures in this section as they have a stronger emphasis on the programming part. They will be discussed in more detail in the next section.

All architectures in this section start with the assumption that the norms of the systems are centrally maintained. The agents themselves can be norm-aware or not. We will start with the more formal and logic-based frameworks before proceeding to the more pragmatic models.

OPERA [34, 72] is founded on the assumption that multi-agent systems imitate human societies. It therefore adopts an organisational structure for its norm modelling. It originates from the idea that agent societies cannot depend on individual agents, all with their own goals, to bring about the society's goals without a framework that specifically enforces this. The goals and norms of the society must be specified, independently of the participating agents since there is no guarantee that the agents will have a desire to achieve the society's goals. This also assures the autonomy of individual participating agents. Each of them can decide to ignore this organisation goal or even rebel against it. Agents are associated with roles or groups of roles within the organisation. The norms are defined on the roles rather than on the agents themselves. With each role an agent assumes, the agent will receive a set of responsibilities and capabilities. The norms are defined in order to allow the organisation to achieve its goals, irrespective of the agents' individual goals. The achievement of the organisation goals is directed by means of landmarks, i.e. states to be achieved, and scenes, a roadmap landmarks . These concepts are formalised using first order logic, allowing for verification of various properties, like for example "is it possible for the organisation to achieve its goals despite violation of the norms by some agents".

OperettA is an open-source tool that provides an organisation-oriented development environment. It was developed in order to support developers in designing and maintaining organization models based mainly on the OPERA framework. It provides specialised editors for all the main components of OPERA's operational model. The model once completed can be viewed as XML data. Additionally, OperettA supports the verification and simulation of the OPERA models.

The INSTAL framework [24] is a normative framework architecture with a formal mathematical model to specify, verify and reason about the norms that govern an open distributed system. The INSTAL approach has opted for an event-driven institutional approach: the norms are expressed on the events/actions of the participants rather than the normative state. Deviation from a norm results in a violation. The premise of the model is that events trigger the creation of institutional fluents. Inspired by Jones and Sergot's [63] account of institutional power and the notion of 'counts-as', the generation relation is used to explain the connection between actions and their interpretation in the context of the institution. The effects of events, actions or institutional events – in terms of the initiation or termination of brute facts [61] and institutional fluents – is described by the consequence relation. Thus, given an event and a state of the institutional model, represented as a set of (institutional) fluents, the next state can be determined by the transitive closure of the generation relation and the consequence relation. The system was extended to be able to deal with interacting normative systems [25].

The formal model and semantics is translated in an equivalent logic program under the answer set semantics [13]. Answer set programming is a declarative programming paradigm with an operational semantics. To support the designer, an associated action language, INSTAL, is provided. Designing a set of norms can be an erroneous process. To support the designer, an inductive logic programming system [26] was developed. On the of a set desired outcomes, also called use-cases, it revises the current specification to take these outcomes into account.

OCeAN (Ontology CommitmEnts Authorizations Norms) [39] is a metamodel that can be used to specify electronic institutions. Those institutions, thanks to a process of contextualization in a specific application domain, can be used in the design of different open systems by enabling the interaction of autonomous agents.

The fundamental concepts of this model, which need to be specified when designing electronic institutions are: (i) an ontology for the definition of the concepts used in the communication and in the specification of the rules of the interaction; (ii) the definition of the events, the actions, and the institutional actions and events that may happen or can be used in the interaction among agents; (iii) the definition of the roles that the agents may play; (iv) an agent communication language (ACL) for the exchange of messages; (v) the definition of the institutional powers for the actual performance of institutional actions; (vi) a set of norms for the specification of obligations, prohibitions, and permissions.

This model has been formalized using the Discrete Event Calculus (DEC) [39], which is a version of the Event Calculus. Recently some parts of this model have been formalized using Semantic Web technologies [41, 40].

One of the early specification tools of multi-agent systems in terms of institutional rules and norms is ISLANDER [36] based on the concepts of electronic institutions as proposed in [74]. The same concepts are the foundation for the INSTAL and OCEAN architectures discussed earlier. In ISLANDER institutions are constructed by means of a graphical user interface. The framework associates roles to agents. The roles offer standardised patterns of behaviour and agents can assume various roles when interacting with the electronic institutions. Scenes describe the normative paths that agents in their respective roles need to follow; the scene and the role describe what an agent can do, is permitted or obligated to do at any given time. Agent's behaviour is further restricted by normative rules, dictating which actions are permitted and which ones are not. The key aspect of ISLANDER approach is that norms can never be violated by the agents. A simulation tool SIMDEI [6] is provided for to animate and verify the institution before deployment.

$\mathcal{M}$OISE$^+$ [60] is a framework for specification by means of social and organisational concepts. This modelling language can be used to specify multi-agent systems through three organisational dimensions: structural, functional, and deontic.

The systems presented in this section fall into two categories. The first group consists of OPERA and $\mathcal{M}$OISE$^+$. Both systems take an organisational approach and express high-level norms concerning a state of the system/organisation. The other architectures, INSTAL, ISLANDER, OCEAN and [42, 82], take their inspiration from institutions where the norms are expressed at level of the actions of the participants. These are referred to as low-level norms. High-level norms can be used to represent *what* the agents should establish — in terms of a declarative description of a system state — rather than specifying *how* they should establish it. So far, to our knowledge, there is no system that allows for the specification of both high and low level norms.

All systems offer a variety of social, normative and organisational constructs to make the specification of the normative system easier. INSTAL offers the least of them, providing a more concise model at the expense of the designer having to hard-code these constructs.

Of all the systems described above $\mathcal{M}$OISE$^+$ is the only architecture that is not grounded in a logical/mathematical formalism. This implies that the soundness and properties of the normative system cannot be analysed through formal analyses and verification mechanisms.

2.3 Programming Normative Multi-Agent Systems

The development of normative multi-agent systems requires programming languages that provide constructs to implement social and organisational concepts and abstractions in order to regulate the behaviour of individual agents. There are two ways to regulate the behavior of individual agents by means of such concepts. First, the programming languages for individual agents can be extended with constructs that allow the representation and reasoning about social and organisational concepts. Such constructs should allow multi-agent programmers to implement agents that make their decisions not only based on their individual goals and beliefs, but also based on the existing social and organisational rules. The idea is that individual agents can be implemented in terms of cognitive and social abstractions such that their behaviours are determined upon reasoning about such abstractions.

Second, one can regulate the behaviour of individual agents exogenously by means of a program external to individual agent programs. The implementation of exogenous mechanisms requires abilities to monitor and control the behaviours of individual agents. The idea is to have an external organisation component that is able to monitor, evaluate and control the behaviour of individual agents. The question is what should be monitored and how the agents' behaviours can be evaluated and influenced. As the internal structure of individual agents cannot be assumed in general, their external behaviours (i.e., communication and interaction with the environment) are the only controllable entities. The organisation software entity can thus observe agents' external behaviour, evaluate them based on social and organisational rules, and determine how to respond to such behaviour. For example, if the organisation specification disallows certain agents to interact, then the organisation component should be able to block such interactions. This suggests that the agents' actions (e.g., communication, environment actions including sense actions) should be processed and managed through the external organisation component, i.e., the organisation component intermediates the interaction between agents as well as the interaction between agents and the environment.

In the previous section, we looked at different approaches to specify and verify normative systems. In this section, we focus on the implementation and the deployment of normative systems through the architectures or programming languages previously mentioned. We start with systems were norms are externalised before looking at programming normative agents.

2.3.1 Programming Normative Organisations

One of the early modelling languages for specifying institutions in terms of institutional rules and norms is ISLANDER [36]. In order to interpret institution specifications and execute them, a computational platform, called AMELI [35], has been developed. This platform implements an infrastructure that, on the one hand, facilitates agent participation within the institutional environment and supports their communication and, on the other hand, enforces the institutional rules and norms as specified in the institutional specification. The key aspect of ISLANDER/AMELI is that norms can never be violated by the agents. In other words, systems programmed via ISLANDER/AMELI make only use of regimentation in order to guarantee the norms to be actually followed. Another characteristic of this proposal is that norms concern communication actions and express which communication actions are permitted, obliged or forbidden.

A related modelling language is $\mathcal{M}$OISE$^+$ [60]. In contrast to ISLANDER [36], $\mathcal{M}$OISE$^+$ focuses more on organisational structures and specifies multi-agent systems through three organisational dimensions: structural, functional, and deontic. Another difference between $\mathcal{M}$OISE$^+$ and ISLANDER/AMELI is that the first is concerned with more high-level norms pertaining to declarative descriptions of the system while the latter considers norms as low-level procedures that directly refer to actions. Various programming frameworks have been proposed to implement and execute $\mathcal{M}$OISE$^+$ specifications. For example, $\mathcal{S}$-$\mathcal{M}$OISE$^+$ [58] is an organisational middleware that provides agents access to the communication layer and the current state of the specified organisation. Another characteristic feature of this middleware is that it allows agents to change the organisation and its specification, as long as such changes do not violate organisational constraints. In the artifact version of this framework, ORG4MAS, various organisational artifacts are used to implement specific components of an organisation such as group and goal schema. In this framework, a special artifact, called reputation artifact, is introduced to manage the enforcement of the norms.

Like ISLANDER/AMELI, $\mathcal{S}$-$\mathcal{M}$OISE$^+$ does not allow agents to violate organisational rules and norms, i.e., norms are regimented rather than being enforced by sanctions. In the artifact version of this framework, ORG4MAS, the enforcement of norms is assumed to be managed indirectly through a reputation mechanism, but it remains unclear how such a reputation system realizes sanctioning. Moreover, both AMELI and $\mathcal{S}$-$\mathcal{M}$OISE$^+$ lack a complete operational semantics that captures all aspects of normative systems, including the enforcement of norms. An explicit formal and operational treatment of norm enforcement is essential for a thorough understanding and analysis of computational frameworks of normative multi-agent systems. A great contribution of ISLANDER/AMELI and $\mathcal{M}$OISE$^+$/$\mathcal{S}$-$\mathcal{M}$OISE$^+$ is the repertoire of social and organisational concepts, which support high-level specification of multi-agent organisations.

In addition to the above approaches, there are some proposals that advocate the use of standard technologies for programming organisations and institutions, e.g., `powerJava` [11] and `powerJade` [10]. These proposals are designed to implement institutions in terms of roles. While `powerJava` extends Java with programming constructs to implement institutions, `powerJade` proposes similar extensions to the Jade framework. Like AMELI, $\mathcal{S}$-$\mathcal{M}$OISE$^+$, and ORG4MAS, these programming frameworks consider an institution as an exogenous coordination mechanism that manages the interactions between participating computational entities (objects in `powerJava` and agents in `powerJade`). However, unlike AMELI, $\mathcal{S}$-$\mathcal{M}$OISE$^+$, and ORG4MAS, the proposed programming frameworks provide only a limited set of social concepts such as role and power. A role is defined in the context of an institution (e.g., a student role is defined in the context of a school) and encapsulates capabilities, also called powers, that its players can use to interact with the institution and with other roles in the institution (e.g., a student can participate in an exam). For an object or an agent to play a role in an institution in order to gain specific abilities, they should satisfy specific requirements as well. In `powerJava` roles and organisations are implemented as Java classes. In particular, a role within an institution is implemented as an inner class of the class that implements the organisation. Moreover, the powers that a player of a role gains and the requirements that the player of the role should satisfy are implemented as methods of the class that implements the role. In `powerJade`, organisations, roles and players are implemented as subclasses of the Jade agent class. The powers that the player of a role gains and the requirements that a player of a role should satisfy are implemented as Jade behaviours (associated to the role).

A dedicated programming language that supports the implementation of normative multi-agent organisations is 2OPL (Organisation Oriented Programming) [31, 86]. Similar to the abovementioned computational frameworks, this programming framework considers an organisation as a software entity that exogenously coordinates the interaction between agents and their shared environment. In particular, the organisation is a software entity that manages the interaction between the agents themselves and between agents and the shared environment. 2OPL provides programming constructs to specify 1) the initial state of an organisation, 2) the effects of agents' actions in the shared environment, and 3) the applicable norms and sanctions. In contrast to AMELI, $\mathcal{S}$-$\mathcal{M}$OISE$^+$ and ORG4MAS, norms in 2OPL can be either enforced by means of sanctions or regimented. In the first case, agents are allowed to violate norms after which sanctions are imposed. In the second case, norms are considered as constraints that cannot be violated. The enforcement of norms by sanctions is a way to guarantee higher autonomy for the agents and higher flexibility for the multi-agent system. The interpreter of 2OPL is based on a cyclic control process. At each cycle, the observable actions of the individual agents (i.e., communication and environment actions) are monitored, the effects of the actions are determined, and norms and sanction are imposed if necessary. Comparing to other approaches, 2OPL has the advantage that it comes with a complete operational semantics such that normative organisation programs can be formally analysed by means of verification techniques [9]. This organisation oriented programming language is extended with programming constructs that support the implementation of concepts such as obligation, permission, prohibition, deadline, norm change, and conditional norm [86, 84, 85].

Rule Responder [76] is an open source framework for creating virtual organizations as multi-agent systems that support collaborative rule-based agent networks on the Semantic Web, where independent agents engage in conversations by exchanging event messages and cooperate to achieve (collaborative) goals. Rule Responder agents communicate in conversations that allow implementing different agent coordination and negotiation protocols. It utilizes messaging reaction rules from Reaction RuleML for communication between the distributed agent inference services, employing various semantically annotated primitives, such as speech acts, or message content information, such as deontic organizational norms, purposes or goals and values etc., which are represented as ontological concepts, using RDF Schema or OWL. The (semi-)autonomous agents use rule engines and Semantic Web rules to describe and execute derivation and reaction logic which declaratively implements the organizational semiotics and the different distributed system/agent topologies with their negotiation/coordination mechanisms. The supported reasoning engines, with their languages, in Rule Responder are, among others, Prova [66] with a Prolog-like language, OO jDREW [12] with RuleML/POSL, and DR-Device [15] with a defeasible logic rule language. Compared to the systems and frameworks presented above, Rule Responder provides the necessary flexible infrastructure to build normative organizations, being agnostic to the modelling/programming language that is used to implement social and organisational concepts. Any of the reasoning engines supported by Rule Responder can be used in an adhoc manner for this purpose. One such example, presented in the next subsection, is DR-DEVICE-M [65]. Furthermore, Rule Responder is extensible enough to allow the incorporation of more specific-to-the task normative reasoning engines, such as the ones presented throughout this section.

2.3.2 Programming Norm Aware Agents

Normative programming frameworks and middleware to support the development of normative multi-agent organisations are often designed to inter-operate with existing BDI-based agent programming languages, and there has been considerable recent work on approaches to programming agents which operate as part of a normative system.

For example, $\mathcal{J}$-$\mathcal{M}$OISE$^+$ [60] is designed to inter-operate with the $\mathcal{S}$-$\mathcal{M}$OISE$^+$ [58] middleware and allows Jason [18] agents to access and update the state of an $\mathcal{S}$-$\mathcal{M}$OISE$^+$ organization. Similarly, the JaCaMo programming framework combines the Jason, Cartago [79], and $\mathcal{S}$-$\mathcal{M}$OISE$^+$ platforms. In JaCaMo, the organisational infrastructure of a multi-agent system consists of organisational artefacts and agents that together are responsible for the management and enactment of the organisation. An organisational artefact employs a normative program which in turn implements a MOISE$^+$ specification. A programming language for the implementation of normative programs as well as a translation of MOISE$^+$ specifications into normative programs is described in [59]. JaCaMo provides similar functionality to $\mathcal{J}$-$\mathcal{M}$OISE$^+$ in allowing Jason agents to interact with organisational artefacts, e.g., to take on a certain role. However, while these approaches allow a developer to program e.g., when an agent should adopt a role, the Jason agents have no explicit mechanisms to reason about norms and their deadlines and sanctions in order to adapt their behaviour at run time.

In [2], the the notion of *norm-aware* agents is introduced. Norm-aware agents deliberate on their goals, norms and sanctions before deciding which plan to select and execute. A norm-aware agent will violate a norm (accepting the resulting sanction) if it is in the agent's overall interests to do so, e.g., if meeting an obligation would result in an important goal of the agent becoming unachievable. Norm-aware agents are related to the notion of *deliberate normative agents* in [22], and are capable of *behaving according to a role specification in a normative organization* and *reasoning about violations* in the sense of [87]. Programming norm-aware agents in conventional BDI-based agent programming languages is difficult, as they lack support for deliberating about goals, norms, sanctions and deadlines. To address this, Alechina et al. [2] introduce the norm-aware agent programming language N-2APL. N-2APL is based on 2APL [27] and provides support for beliefs, goals, plans, norms, sanctions and deadlines. Alechina et al. show that N-2APL agents are rational in the sense of committing to a set of plans that will achieve the agent's most important goals and obligations by their deadlines while respecting its most important prohibitions. N-2APL agents are designed to interoperate with the 2OPL architecture for normative systems [31]. However, as the idea of organisational artefacts based on normative programs is closely related to the 2OPL architecture, Alechina et al. speculate that N-2APL agents could also be used in the JaCaMo framework.

Another approach that integrates norms in a BDI-based agent programming architecture is proposed in [70]. This extends the AgentSpeak(L) architecture with a mechanism that allows agents to behave in accordance with a set of non-conflicting norms. As in N-2APL, the agents can adopt obligations and prohibitions with deadlines, after which plans are selected to fulfil the obligations or existing plans are suppressed to avoid violating prohibitions. However, unlike N-2APL, [70] does not consider scheduling of plans with respect to their deadlines or possible sanctions.

Defeasible deontic logic is a representation and reasoning formalism for normative agent knowledge [65, 4, 48, 71]. It combines the non-monotonicity of defeasible logic with the normative modalities of deontic logic; with the former offering a computational environment of the latter. Defeasible and deontic logic provide the means for modeling multi-agent systems, where each agent is characterized by its own cognitive profile and normative system, as well as policies, which define privacy requirements for a user, access permissions for a resource, individual rights etc.

There are a few systems that implement defeasible deontic logics. One of them is DR-DEVICE-M [65], an extension of the DR-DEVICE Semantic Web-aware defeasible reasoning engine [15], with a mechanism for expressing modal and deontic logic operators. DR-DEVICE-M is based on work presented in [48], the main difference being that it adopts a theory transformation for turning a modal defeasible theory into a non-modal one, while [48] is based on a meta-program formalization of defeasible logic [68]. Furthermore, DR-DEVICE-M supports an extensible agent type definition scheme, where the agent is declared in the rule base, while modality interactions are dealt with parametrically, via external agent-type definition files. Compared to the programming frameworks discussed above, DR-DEVICE-M, as well as other similar defeasible deontic logic reasoning engines [3, 67, 71], cannot fully implement a programming environment for norm aware agents, but can play the role of the reasoning mechanism about norms.

3 Challenges

This section provides a list of issues that are currently challenging the development of normative multi-agent systems. These challenges concern computational systems for norms that can be used to specify, verify, and implement norms in multi-agent systems.

3.1 Verification of Norms

In its most general definition norm verification systems take a set of norms and some properties that they are suppose to comply with, and verify whether they do or do not. Ideally, they should provide counter examples if they do not comply. The term *offline* is sometimes used to differentiate this procedure from a similar one that some systems need to do at run-time (see Section 3.7). Many times the property of interest is absence of conflicts, and the set of norms represents a contract, a law, or some other legal artefacts.

Verification of norms face the community with the need to advance in at least two foundational issues:

- Finding a common family of formalisms where different interpretations of legal concepts, and the consequences of such interpretations, could be represented and discussed over a shared background. Even when formal semantics are provided, it is not always clear how to establish a fair comparison between existing proposals, or to understand what these formalisms have in common.
- Starting to consider not only expressiveness but also complexity and performance, so we can also build actual tools that help us produce better real-world norms. Significant research has been done in the direction of capturing the intended meaning of deontic modalities and fulfilling the expressiveness requirements for modest-sized normative systems. But the question of how computationally tractable the proposed systems are is usually overlooked.

These concepts should be discussed both in general and in specific contexts. This is particularly important because even if the general problem is hard, there could be "niche markets" where practical contributions to state-of-the-art and society could be made, like two-party contracts.

3.1.1 Need for Common Playground

Clarifying the meaning of legal concepts is a worthwhile endeavor. It's important to know if, for example, empowering someone to marry two individuals that have immunity to marriage is equivalent to having the liberty of bringing about that two individuals are married when they have the freedom of being married.

Although the community has not yet reach final consensus over the meaning of prohibition, obligation and permission, there is at least some level of agreement over their basic properties. Introduced in 1913 by Hofheld [54], other modalities such as *claim right*, *power*, *freedom* and *immunity* are still less clear. Many attempts have been made at formalizing the Hofhelian concepts (e.g., [64, 69, 81, 80]). However, semantic-wise, most of them went no further than structured language, leaving many questions unresolved. For instance, [80] introduces the concept of directed modalities to express sentences like "It is obligatory that Tom pays Mary $1000 in order to advance Mary's interests". If Mary is using the money to pay a blackmailer or to buy cancer-causing cigarettes, is she advancing her interests? According to whom? Can Tom deny the payment claiming that she would not use the money "to advance her interests"? [81] is more precise about which operator combinations are consistent given a few assumptions about the underlying logic, but because it only considers some basic modalities, and because the logic is not fixed, we still don't know if, for example, being empowered but forbidden makes any sense.

In order to fully exploit the benefits of a formal language, formal semantics are also needed. A legal concept is fully understood if, given a legal corpora that uses it in combination with others, we can set apart legal from illegal behaviors. Such behaviors are usually shaped as narratives of events, i.e., traces. Given a state of affairs and a legal corpora, only some combinations of actions and events are legal: (future) traces again. So if we want to focus our efforts in practical approaches for improving the quality of real-life normative systems, what's needed are trace-friendly semantics.

There are currently many deontic proposals tailored for different situations (e.g., monitoring, conflict analysis, judging, etc). As mentioned in Section 2.1, the dream of a one-size-fits-all, experience suggest, is only Utopian. However, we could aim for founding logics on a family of related mathematical structures, forming a common base where comparisons become easier. Much as when giving Kripke semantics to modal logics allowed to easily compare axiomatic systems whose logic relation at that point was unknown [43].

Comparing again with Software Engineering, there are Architecture Description Languages and methodologies tailored for specific purposes such as avionics, web systems, finances, etc. In the same vein, maybe we need different deontic languages for trade contracts (e.g., supporting some numeric capabilities) than for finding conflicts in criminal law, where different ways of expressing intention become important. This is one of the challenges that need to be addressed: for each domain of interest (e.g., business contracts, civil law, SOAP, etc.) determine the expressivity requirements, which types of conflicts need to be detected and find tools that handle them properly.

3.1.2 Dealing with Complexity

Computational complexity is a topic on its own right, and we discuss it in the next section, but there is also the complexity that humans face when trying to understand large and complex pieces of information. Deontic formalisms willing to deal with real-world sized corpora should allow to model different levels of abstraction and to apply compositional reasoning [16]. These are powerful tools to cope with large, interrelated, complex and sometimes difficult to attain sets of concepts.

At the micro level, doing detailed intra-norm conflict analysis is definitively a must, but also to be able to understand complex inter-norm interactions. To exemplify, we would like to know if sending an email *counts as* placing an order in much the same way as sending a signed order form does, or if a given contract presents some contradiction surrounding those two practices. Zooming out, we would also like to know if placing an order surpassing one's paying capacity is considered illegal by some national law, abstracting away the particulars of how the order was placed. Compositional reasoning seems like the right tool, allowing to concentrate into details, checking their local correctness and then analyzing global correctness from clearly abstracted local properties.

However, care should be exercised when reasoning compositionally as this small but representative example shows: if some higher-ranked norm states that email orders are considered void for amounts over a limit, the particulars of how the order was placed cannot be abstracted away. The topic of compositional reasoning and abstraction in the deontic world, we believe, deserves further analysis.

3.1.3 The Quest for Tools

The predominant efforts carried out so far on normative systems, and deontic systems in general, share a strong philosophical bias. Identifying legal concepts and structuring them into operators, analyzing their interactions, expressiveness and derived paradoxes drove the main research agenda of the last decades. Even though there was valuable work in the direction of building real tools ([78, 46, 83, 51], etc.), this effort is far from being completed. Computer Science as a discipline, and in particular Computational Logic, has evolved significantly and today is much closer to provide usable tools that can cope with scales that match real-life cases. The area of modal logic, and specifically model checking, has proven to be very successful at this respect (e.g, [53]). It is a very good opportunity to make the most of this momentum and apply it for building usable tools for legal drafting and norm verification.

There is a very well-known usability problem in tools that deal with formal systems. Reconciling a smooth and natural user experience and the right level of abstraction with a rigorous formal language is a very difficult issue to overcome. The most frequent result are tools that demand very specialized technical skills to the end user. There are many similarities between software specification and legislative sciences [45], but there is a difference here: even though the final user for both communities is probably a non-technical person, people that *build* software, the developers, are inherently closer to formal verification. On the contrary, people that *build* laws, the legislators, are probably not familiar with the concept. For the time being, we see real tools aiding the legislative process only in the presence of a strong side-by-side collaboration between specialized technicians and legislators. The first major difficulty that has to be solved, we claim, is not really offering a pleasant experience for the non-technical user, but designing tools that can beat real-world sized problems.

We mentioned scalability, but the fundamental aspect we think has been largely overlooked for formal legal systems is computational tractability. Although some work has been done where practical considerations are analyzed (e.g., [8, 78]), many proposed systems analyze only the philosophical aspects, and no practical short-term role for them in the legislative process seems to be devised. And going farther from theoretical complexity, the ability to count on mature, highly tuned, inference tools is key. It is well-known that a full-fledged inference tool can many times provide better empirical results working with a reasonably complex logic than a young, made-from-scratch tool developed for a theoretically less complex logic. The work done in SMT-solvers, theoretically intractable, but successful in many real-life cases (e.g., [33]) is a good example. This point nicely connects to the considerations we made in the previous section: the ability to compare logics, and therefore know how much they overlap, will allow us to reuse off-the-shelf inference tools that have decades of development effort on their backs.

3.2 Operational Semantics for Norms

Operational semantics are used to formally specify the meaning of programming languages. They allow the specification of program executions and their formal verifications. It is therefore essential that programming languages for normative multi-agent systems (both for normative organizations and norm-aware agents) come with their operational semantics. Operational semantics require an operationalization of the normative concepts that are involved in the programming languages. An operationalization of normative concepts assumes a computational representation and a reasoning engine.

The expressivity of norms in most agent-based programming languages (e.g., 2OPL and N-2APL) is limited to atomic cases. In some of these programming languages, norms are represented by atomic expressions such as $O(i, \phi)$ (agent i has an obligation to bring about ϕ, where ϕ is a conjunction of literals) such that reasoning with these concepts is limited to reasoning about ϕ. In some programming languages the deontic concepts are evaluated on states and operationalized by ensuring that each state either satisfies the deontic expressions (e.g., for $O(i, \phi)$ ensuring that each state satisfies ϕ) or otherwise be considered as a state in which agent i is in violation. In these programming languages, one can specify sanctions that should be imposed when specific violation occurs. A sanction ψ (often a conjunction of literals) is imposed by updating the state with ψ. In some other agent-based programming languages, the deontic expressions are extended with deadlines, e.g., the extended expression $\langle O(i, \phi), d, \psi \rangle$ indicates that agent i is obliged to bring about ϕ before deadline d otherwise sanction ψ should be imposed. The deadline in such an expression determines the temporal scope of the deontic element. Such extended deontic expressions are evaluated on paths (a sequence of states) and operationalized by ensuring that each execution path ending with a state satisfying deadline d has a state that satisfies the deontic element, or otherwise the state satisfying d should be updated with sanction ψ.

An important challenge in this respect is to have more expressive but computationally tractable fragments of normative concepts that can be operationalized and incorporated in the compilers or interpreters of agent-based programming languages. Also, the set of deontic concepts can be extended to include a variety of social and normative concepts such as responsibility, delegation of responsibility, trust, commitments, roles and their enactments. The incorporation of these concepts requires their operationalization which in turn requires computational representation and reasoning engines. A particular difficulty in this regard is the possibility of conflicts that can emerge between various normative concepts (e.g., agent i is both obliged and prohibited to bring about ϕ or agent i is responsible for ϕ but has no permission to control or change it) such that any operationalization of these concepts should be able to detect and resolve such conflicts. Another challenge is the operationalization of deontic concepts applied to groups of individuals, e.g., group obligation and collective responsibility. In particular, when and how such concepts can be created, maintained, released, and how to cope with violations and sanctions towards individual agents. Finally, normative concepts can be applied to both states and actions, including communication actions (e.g., agent i is obliged to bring about ϕ or to perform action α). State- and action-based norms can interact and it is a challenge to propose programming constructs to implement both kinds of normative concepts while governing their interactions.

3.3 Programming Norm Change

Norm change is an interesting but complex phenomenon. There are many aspects related to norm change: the entity/authority that can issue a norm change (e.g., can every agent issue a norm change?), the types of norms that can be changed (e.g., can a concrete norm applied to a specific agent change or can a general norm applied to a set of agents change too? do we allow constitutive and regulative norms to change?), the constraints that norm change mechanism should satisfy (e.g., under which conditions a norm can be adopted and what are the consequences?), and how to handle norm conflicts as a consequence of norm change. Other dynamic aspects of norms are known as legal annulment and legal abrogation. Legal abrogation refers to cancellation of a norm except for effects that were obtained already before the cancellation, and legal annulment refers to the total cancellation of a norm including all (earlier) effects of that norm.

Most aspects have been extensively investigated in the literature. For example, Artikis [7] has presented an infrastructure allowing agents to modify a protocol (a set of laws) at runtime, Bou et al. [19] and Campos et al. [21] have investigated how norms in electronic institutions can change at runtime, and Oren et al. [73] have considered how powers can be used to create, delete and modify norms. Despite these research works on norm change not much attention has been given to programming languages that support the development and implementation of normative multi-agent systems in which norms can change at runtime. Such programming languages require constructs to create, charge, and discharge norms. It should be noted that it is often difficult, if not impossible, to operationalize phenomena such as legal annulment.

A proposal to facilitate norm change by means of dedicated programming constructed is presented in [32]. In this work, norm change is considered at the level of both norm instances and norm schemes. Norm instances are detached norms, i.e., norms that are already issued and directed to specific agents (e.g., agent i is obliged to fulfill its payment before next Wednesday) whereas norm schemes are general norms that can be applied and instantiated to specific agents (e.g., payments should be fulfilled within one week after signing the contract). Both norm instances and norm schemes are specified by means of rules which can be applied at run-time to modify norms. In this framework, changing a norm scheme may have consequences for the existing norm instances. For example, a norm scheme that states that a payment should be fulfilled before one week after signing the contract can be changed to a norm where the deadline is extended with one more week. The question is what needs to be done with the norm instances that are already issued before a norm change takes place. A related and more general problem in this respect is the problem of organisation change. Multi-agent organisations aim at controlling and coordinating the behaviour of individual agents by means of a plethora of concepts including normative concepts. Besides normative concepts, multi-agent organisations uses concepts such as roles, scenes, capabilities, services and databases that can be changed at runtime. A full-fledge programming language that supports the implementation of multi-agent organisation should therefore provide constructs and mechanisms that can be operationalized to manage the organisational changes.

3.4 Norm Adaptation and Emergence

Norm emergence and adaptation in open normative systems is related to norm change discussed issues discussed previously. Like society, norms continuously change. Some norms get discarded while new ones get introduced. Designers of these systems cannot foresee all possible changes that might occur to the system. The evolution of the system should not be prescribed nor regimented. In some systems, for instance, participating agents should be able to collectively decide which norms can abandoned or adapted. This could be the case for norms that are constantly violated by a majority of the agent population. While agents can bring up the adoption of a norm to a vote based on the behaviour of the participants.

3.5 Scalable Architectures for Normative Systems

A key challenge for large, distributed normative systems is scalability. In a normative multi-agent system, the interaction between agents and the environment is governed by a normative exogenous organisation, which is specified by a set of conditional norms with their associated deadlines and sanctions. The normative organisation monitors environmental changes resulting from the activities of agents to: determine any obligations to be fulfilled or prohibitions that should not be violated by the agents; determine any violations based on the deadlines of the detached obligations and prohibitions; and impose any corresponding sanctions. The role of the exogenous organisation is thus to continuously 1) monitor the behaviour of agents, 2) evaluate the effects they bring about in the environment, 3) determine norms that should be respected, 4) check if any norm is violated and 5) take necessary measures (i.e., imposing sanctions) when norms are violated. This continuous process is often implemented by a so-called control cycle [31]. For large, open, distributed normative MAS, implementing the control cycle is non-trivial. State information may be distributed, and updated asynchronously by many agents, which may join or leave the system at any time. Such systems demand scalable architectures for behaviour and norm monitoring. One approach is to distribute such functions, however this entails additional complexity, e.g., transactional approaches to state update. For very large scale systems, 100% monitoring may be infeasible, and probabilistic guarantees of the detection of norm violations may be more appropriate. Such approaches, in turn, may impact the design of sanctions, and, when applied to behaviour monitoring, the design of the conditional norms which specify the normative exogenous organisation itself.

3.6 Algorithms for Compliance Checking

By this, we mean algorithms for checking whether external behaviour of agents conforms to the system's norms.

Compliance checking will require more or less computationally demanding algorithms depending on the structure of norms (e.g. do we need to monitor for bad states/bad events or do we need to analyse histories in order to detect a violation?). For some systems, existing algorithms can be re-used. For example algorithms for checking database integrity, or methods for checking business process compliance, or on-the-fly LTL model-checking [55]. For more complex norms we may need to design new algorithms or come up with non-trivial modifications of existing ones.

3.7 Normative Conflict Detection and Resolution

We addressed conflict detection at Section 3.1. The types of systems mentioned there usually need only to point out conflicts and leave resolution in the hands of a human being. There are some systems, however, that need to both detect and resolve conflicts at run-time, without human intervention, autonomous normative agents being an example.

It is seldom the case that offline mechanisms, which are usually more resource consuming, are fit for the purpose of run-time detection of conflicts. Lightweight detection mechanisms are needed, and those probably require changes in normative languages (see Section 3.2) to relax some assumptions or provide more information for the automatic resolution of conflicts. The whole topic of how a norm-aware agent should resolve conflicts at runtime requires further investigation, and probably domain-specific solutions instead of a general one.

3.8 Programming Languages for Norm-Aware Agents

Current approaches to normative agent programming such as N-2APL [2] have significant limitations, and many basic questions remain unanswered. For example, if the priorities of norms are commensurable, norm-aware deliberation is intractable — are there contexts where commensurable priorities are essential or contexts where 'best effort' (i.e., potentially intractable) deliberation is acceptable? What sorts of deadlines should be associated with norms — a single point in time or an interval of time? Should deadlines be interpreted as 'soft' or 'hard', or are both required? How can an agent developer estimate the time required to execute a plan in order to find out if an obligation can be discharged by its deadline? Should plan execution times rather be expressed as probabilistic profiles?

Additional challenges arise from overlaps between programming languages for implementing norm-aware agents and the operational semantics assumed for norms (see Section 3.2). For example, current approaches to normative agent programming have focussed on state-based norms, and would need to be extended to support action-based norms and their interactions with state-based norms. Similarly, existing approaches lack support for norm-aware deliberation about norms that apply to groups of agents, or normative interactions between agents, such as delegation of responsibility.

4 Conclusions

Looking at other fields that reached maturity, it is common to see an iterative pattern, each cycle consisting of an expansion phase, where new concepts are discovered, new entities are named and new methods are proposed, followed by a period of synthesis, where some of those are pruned or at least classified and people start to agree on which problems are best suited to be worked with which tools.

As was previously mentioned, the field of Computational Models for Normative Multi-Agent Systems involves a plethora of social concepts like roles, groups, social structures, organisations, institutions, norms, etc. that have been introduced in multi-agent system methodologies, models, specification and modelling languages, and computational frameworks.

If our Computational Models are to advance, it might be time to start classifying the various models and tools. Which tools are best for designing/implementing on-line monitoring? And for off-line checking? For programming action-based semantics? And for society-imposed regulations? How different tools and methodologies compare? Any of them solves the open issues?

This chapter leans toward providing some information to answer such type of questions, hopping to serve the practitioner by presenting a clear view of what we have and what we still need.

References

1 T. Ågotnes, W. van der Hoek, and M. Wooldridge. Robust normative systems. In Padgham, Parkes, Muller, and Parsons, editors, *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, pages 747–754, Estoril, Portugal, May 2008. IFAMAAS/ACM DL.

2 Natasha Alechina, Mehdi Dastani, and Brian Logan. Programming norm-aware agents. In Vincent Conitzer, Michael Winikoff, Lin Padgham, and Wiebe van der Hoek, editors, *Proc. of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, volume 2, pages 1057–1064, Valencia, Spain, June 2012. IFAAMAS.

3 Grigoris Antoniou, Nikos Dimaresis, and Guido Governatori. A modal and deontic defeasible reasoning system for modelling policies and multi-agent systems. *Expert Systems with Applications*, 36(2, Part 2):4125 – 4134, 2009.

4 Grigoris Antoniou, Thomas Skylogiannis, Antonis Bikakis, Martin Doerr, and Nick Bassiliades. Dr-brokering: A semantic brokering system. *Knowledge-Based Systems*, 20:61–72, 2007.

5 F. Arbab, L. Astefanoaei, F. de Boer, M. Dastani, J.-J.Ch. Meyer, and N. Tinnermeier. Reo connectors as coordination artifacts in 2APL systems. In *Proceedings of the 11th Pacific Rim International Conference on Multi-Agents (PRIMA 2008)*, volume LNCS 5357, pages 42–53. Springer, 2009.

6 Josep Lluís Arcos, Marc Esteva, Pablo Noriega, Juan A. Rodríguez-Aguilar, and Carles Sierra. Engineering open environments with electronic institutions. *Eng. Appl. of AI*, 18(2):191–204, 2005.

7 Alexander Artikis. Dynamic protocols for open agent systems. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 97–104, 2009.

8 Alexander Artikis, Marek Sergot, and Jeremy Pitt. Specifying norm-governed computational societies. *ACM Trans. Comput. Logic*, 10:1:1–1:42, January 2009.

9 L. Astefanoaei, M. Dastani, J.-J. Ch. Meyer, and F. Boer. On the semantics and verification of normative multi-agent systems. *International Journal of Universal Computer Science*, 15(13):2629–2652, 2009.

10 M. Baldoni, G. Boella, M. Dorni, R. Grenna, and A. Mugnaini. powerJADE: Organizations and roles as primitives in the JADE framework. In *In of WOA 2008: Dagli oggetti agli agenti, Evoluzione dell'agent development: metodologie, tool, piattaforme e linguaggi*, pages 84–92, 2008.

11 M. Baldoni, G. Boella, and L. Van Der Torre. Roles as a coordination construct: Introducing powerJava. In *In Proceedings of 1st International Workshop on Methods and Tools for Coordinating Concurrent, Distributed and Mobile Systems*, volume 150 (1), pages 9–29. Electronic Notes in Theoretical Computer Science, 2005.

12 Marcel Ball, Harold Boley, David Hirtle, Jing Mei, and Bruce Spencer. The oo jdrew reference implementation of ruleml. In Asaf Adi, Suzette Stoutenburg, and Said Tabet, editors, *Rules and Rule Markup Languages for the Semantic Web*, volume 3791 of *Lecture Notes in Computer Science*, pages 218–223. Springer Berlin Heidelberg, 2005.

13 Chitta Baral. *Knowledge Representation, Reasoning and Declarative Problem Solving.* Cambridge Press, 2003.

14 J. Barnat, L. Brim, I. Černá, P. Moravec, P. Ročkai, and P. Šimeček. DiVinE – A Tool for Distributed Verification (Tool Paper). In *Computer Aided Verification*, volume 4144/2006 of *LNCS*, pages 278–281. Springer Berlin / Heidelberg, 2006.

15 Nick Bassiliades, Grigoris Antoniou, and Ioannis P. Vlahavas. A defeasible logic reasoner for the semantic web. *Int. J. Semantic Web Inf. Syst.*, pages 1–41, 2006.

16 S. Berezin, S. Campos, and E. Clarke. Compositional reasoning in model checking. *Compositionality: The Significant Difference*, pages 81–102, 1998.

17 G. Boella and L. van der Torre. Substantive and procedural norms in normative multiagent systems. *Journal of Applied Logic*, 6:152–171, 2008.

18 Rafael H. Bordini, Michael Wooldridge, and Jomi Fred Hübner. *Programming Multi-Agent Systems in AgentSpeak using Jason.* Wiley Series in Agent Technology. John Wiley & Sons, 2007.

19 Eva Bou, Maite López-Sánchez, and Juan A. Rodríguez-Aguilar. Adaptation of autonomic electronic institutions through norms and institutional agents. In *Proc. of the 7th International Conf. on Engineering Societies in the Agents World (ESAW)*, pages 300–319, 2006.

20 Joost Breuker, Emil Petkov, and Radboud Winkels. Drafting and validating regulations: The inevitable use of intelligent tools. In *AIMSA '00: Proceedings of the 9th International Conference on Artificial Intelligence*, pages 21–33, London, UK, 2000. Springer-Verlag.

21 Jordi Campos, Maite López-Sánchez, Juan Antonia Rodríguez-Aguilar, and Marc Esteva. Formalising situatedness and adaptation in electronic institutions. In *Proceedings of Coordination, Organizations, Institutions and Norms in Agent Systems (COIN@AAAI)*, pages 126–139, Berlin, Germany, 2009. Springer.

22 C. Castelfranchi, F. Dignum, Catholijn M. Jonker, and Jan Treur. Deliberative normative agents: Principles and architecture. In N. R. Jennings and Y. Lesperance, editors, *Intel-*

ligent Agents VI. Proceedings of the 6th International Workshop on Agent Theories and Architectures and Languages and ATAL'99., LNAI, pages 364–378. Springer Verlag, 2000.

23 A. Cimatti, E. Clarke, F. Giunchiglia, and M. Roveri. NuSMV: a new symbolic model checker. *International Journal on Software Tools for Technology Transfer (STTT)*, 2(4):410–425, 2000.

24 Owen Cliffe, Marina De Vos, and Julian Padget. Answer set programming for representing and reasoning about virtual institutions. In Katsumi Inoue, Satoh Ken, and Francesca Toni, editors, *Computational Logic for Multi-Agents (CLIMA VII)*, volume 4371 of *LNAI*, pages 60–79, Hakodate, Japan, May 2006. Springer.

25 Owen Cliffe, Marina De Vos, and Julian Padget. Specifying and reasoning about multiple institutions. In Pablo Noriega, Javier Vázquez-Salceda, Guido Boella, Olivier Boissier, Virginia Dignum, Nicoletta Fornara, and Eric Matson, editors, *Coordination, Organization, Institutions and Norms in Agent Systems II – AAMAS 2006 and ECAI 2006 Int'l Workshops, COIN 2006 Hakodate, Japan, May 9, 2006 Riva del Garda, Italy, August 28, 2006*, volume 4386 of *LNCS*, pages 67–85. Springer Berlin / Heidelberg, 2007.

26 Domenico Corapi, Marina De Vos, Julian Padget, Alessandra Russo, and Ken Satoh. Normative design using inductive learning. *Theory and Practice of Logic Programming*, 11:783–799, 2011.

27 M. Dastani. 2APL: a practical agent programming language. *International Journal of Autonomous Agents and Multi-Agent Systems*, 16(3):214–248, 2008.

28 M. Dastani, F. Arbab, and F.S. de Boer. Coordination and composition in multi-agent systems. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'05)*, pages 439–446. 2005.

29 M. Dastani and J. Gomez-Sanz. Programming multi-agent systems. *The Knowledge Engineering Review*, 20(2):151–164, 2006.

30 M. Dastani and J.J. Gomez-Sanz. Agentlink iii technical forum group, programming multiagent systems. `http://people.cs.uu.nl/mehdi/al3promas.html`.

31 M. Dastani, D. Grossi, J.-J. Ch. Meyer, and N. Tinnemeier. Normative multi-agent programs and their logics. In *Proc. Workshop on Knowledge Representation for Agents and Multi-Agent Systems*, LNCS 5605, pages 16–31, 2009.

32 Mehdi Dastani, John-Jules Meyer, and Nick Tinnemeier. Programming norm change. *Journal of Applied Non-classical Logics*, Published online, 2012.

33 Leonardo De Moura and Nikolaj Bjørner. Satisfiability modulo theories: introduction and applications. *Commun. ACM*, 54:69–77, September 2011.

34 V. Dignum. *A Model for Organizational Interaction.* PhD thesis, Utrecht University, SIKS, 2004.

35 M. Esteva, J.A. Rodríguez-Aguilar, B. Rosell, and J.L. Arcos. AMELI: An agent-based middleware for electronic institutions. In *Proceedings of the Third International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2004)*, pages 236–243, New York, US, July 2004.

36 Marc Esteva, David de la Cruz, and Carles Sierra. ISLANDER: an electronic institutions editor. In *Proceedings of the First International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2002)*, pages 1045–1052, Bologna, Italy, 2002.

37 Stephen Fenech, Gordon J. Pace, and Gerardo Schneider. Automatic conflict detection on contracts. In *ICTAC '09: Proceedings of the 6th International Colloquium on Theoretical Aspects of Computing*, pages 200–214, Berlin, Heidelberg, 2009. Springer-Verlag.

38 Stephen Fenech, Gordon J. Pace, and Gerardo Schneider. Clan: A tool for contract analysis and conflict discovery. In Zhiming Liu and Anders P. Ravn, editors, *ATVA*, volume 5799 of *Lecture Notes in Computer Science*, pages 90–96. Springer, 2009.

39 N. Fornara and M. Colombetti. Specifying artificial institutions in the event calculus. In V. Dignum, editor, *Handbook of Research on Multi-Agent Systems: Semantics and Dynamics of Organizational Models*, Information science reference, chapter XIV, pages 335–366. IGI Global, 2009.

40 Nicoletta Fornara. Specifying and monitoring obligations in open multiagent systems using semantic web technology. In A. Elçi, M. Tadiou Kone, and M. A. Orgun, editors, *Semantic Agent Systems: Foundations and Applications*, volume 344 of *Studies in Computational Intelligence*, chapter 2, pages 25–46. Springer-Verlag, 2011.

41 Nicoletta Fornara and Marco Colombetti. Representation and monitoring of commitments and norms using owl. *AI Commun.*, 23(4):341–356, 2010.

42 A. Garcia-Camino, P. Noriega, and J. A. Rodriguez-Aguilar. Implementing norms in electronic institutions. In *Proc. of the 4th Int'l Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2005)*, pages 667–673, New York, NY, USA, 2005.

43 R. Goldblatt. Mathematical modal logic: a view of its evolution. *Journal of Applied Logic*, 1(5-6):309–392, 2003.

44 Daniel Gorín, Sergio Mera, and Fernando Schapachnik. Model Checking Legal Documents. In *Proceedings of the 2010 conference on Legal Knowledge and Information Systems: JURIX 2010*, pages 111–115, December 2010.

45 Daniel Gorín, Sergio Mera, and Fernando Schapachnik. A Software Tool for Legal Drafting. In *FLACOS 2011: Fifth Workshop on Formal Languages and Analysis of Contract-Oriented Software*, pages 1–15. Elsevier, 2011.

46 G. Governatori and D.H. Pham. Dr-contract: An architecture for e-contracts in defeasible logic. *International Journal of Business Process Integration and Management*, 4(3):187–199, 2009.

47 Guido Governatori, Joris Hulstijn, Régis Riveret, and Antonino Rotolo. Characterising deadlines in temporal modal defeasible logic. In *Proceedings of the 20th Australian joint conference on Advances in artificial intelligence*, AI'07, pages 486–496, Berlin, Heidelberg, 2007. Springer-Verlag.

48 Guido Governatori and Antonino Rotolo. Bio logical agents: Norms, beliefs, intentions in defeasible logic. *Autonomous Agents and Multi-Agent Systems*, 17:36–69, 2008. 10.1007/s10458-008-9030-4.

49 D. Grossi. *Designing Invisible Handcuffs.* PhD thesis, Utrecht University, SIKS, 2007.

50 N. Den Haan. Tracs: A support tool for drafting and testing law. In *Jurix 92: Information Technology and Law*, pages 63–70, 1992.

51 S. Hagiwara and S. Tojo. Discordance Detection in Regional Ordinance: Ontology-based Validation. In *Proc. of the 2006 Conference on Legal Knowledge and Information Systems: JURIX 2006: The Nineteenth Annual Conference*, pages 111–120. IOS Press, 2006.

52 Jörg Hansen, Gabriella Pigozzi, and Leendert W. N. van der Torre. Ten philosophical problems in deontic logic. In Guido Boella, Leendert W. N. van der Torre, and Harko Verhagen, editors, *Normative Multi-agent Systems*, volume 07122 of *Dagstuhl Seminar Proceedings.* Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, 2007.

53 K. Havelund, M. Lowry, and J. Penix. Formal analysis of a space-craft controller using SPIN. *IEEE Transactions on Software Engineering*, pages 749–765, 2001.

54 W.N. Hohfeld. Some fundamental legal conceptions as applied in judicial reasoning. *Yale Lj*, 23:16, 1913.

55 Gerard J. Holzmann. On-the-fly model checking. *ACM Comput. Surv.*, 28(4es):120, 1996.

56 Gerard J. Holzmann. *The Spin Model Checker: Primer and Reference Manual.* Addison-Wesley, 2003.

57 J.F. Hübner, O. Boissier, R. Kitio, and A. Ricci. Instrumenting multi-agent organisations with organisational artifacts and agents: Giving the organisational power back to the agents. *International Journal of Autonomous Agents and Multi-Agent Systems*, 20:369–400, 2010.

58 J.F. Hübner, J.S. Sichman, and O. Boissier. $\mathcal{S}-\mathcal{M}\mathrm{OISE}^{+}$: A middleware for developing organised multi-agent systems. In *Proceedings of the international workshop on Coordination, Organizations, Institutions, and Norms in Multi-Agent Systems*, volume 3913 of *LNCS*, pages 64–78. Springer, 2006.

59 Jomi Fred Hübner, Olivier Boissier, and Rafael H. Bordini. From organisation specification to normative programming in multi-agent organisations. In Jürgen Dix, João Leite, Guido Governatori, and Wojtek Jamroga, editors, *Computational Logic in Multi-Agent Systems, 11th International Workshop, CLIMA XI, Lisbon, Portugal, August 16-17, 2010. Proceedings*, volume 6245 of *Lecture Notes in Computer Science*, pages 117–134. Springer, 2010.

60 Jomi Fred Hübner, Jaime Simão Sichman, and Olivier Boissier. Developing organised multiagent systems using the $\mathcal{M}\mathrm{OISE}^{+}$ model: programming issues at the system and agent levels. *International Journal of Agent-Oriented Software Engineering*, 1(3/4):370–395, 2007.

61 John R. Searle. *The Construction of Social Reality.* Allen Lane, The Penguin Press, 1995.

62 A. J. I. Jones and M. Sergot. On the characterization of law and computer systems. In J.-J. Ch. Meyer and R.J. Wieringa, editors, *Deontic Logic in Computer Science: Normative System Specification*, pages 275–307. John Wiley & Sons, 1993.

63 Andrew J.I. Jones and Marek Sergot. A Formal Characterisation of Institutionalised Power. *ACM Computing Surveys*, 28(4):121, 1996.

64 S. Kanger and H. Kanger. Rights and parliamentarism. *Theoria*, 32(2):85–115, 1966.

65 Efstratios Kontopoulos, Nick Bassiliades, Guido Governatori, and Grigoris Antoniou. A modal defeasible reasoner of deontic logic for the semantic web. *Int. J. Semantic Web Inf. Syst.*, 7(1):18–43, 2011.

66 Alexander Kozlenkov, Rafael Penaloza, Vivek Nigam, Loic Royer, Gihan Dawelbait, and Michael Schroeder. Prova: Rule-based java scripting for distributed web applications: A case study in bioinformatics. In *EDBT Workshops'06*, pages 899–908, 2006.

67 Ho-Pun Lam and Guido Governatori. The making of spindle. In *Proceedings of the 2009 International Symposium on Rule Interchange and Applications*, RuleML '09, pages 315–322, Berlin, Heidelberg, 2009. Springer-Verlag.

68 M. J. Maher and G. Governatori. A semantic decomposition of defeasible logics. In *Proceedings of the sixteenth national conference on Artificial intelligence and the eleventh Innovative applications of artificial intelligence conference innovative applications of artificial intelligence*, AAAI '99/IAAI '99, pages 299–305, Menlo Park, CA, USA, 1999. American Association for Artificial Intelligence.

69 D. Makinson. On the formal representation of rights relations. *Journal of philosophical Logic*, 15(4):403–425, 1986.

70 Felipe Rech Meneguzzi and Michael Luck. Norm-based behaviour modification in BDI agents. In Carles Sierra, Cristiano Castelfranchi, Keith S. Decker, and Jaime Simão Sichman, editors, *8th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, pages 177–184. IFAAMAS, 2009.

71 Donald Nute, editor. *Defeasible Deontic Logic: Essays in Nonmonotonic Normative Reasoning*, volume 263 of *Synthese Library.* Kluwer Academic Publishers, Dordrecht, Holland, 1997.

72 Daniel Okouya and Virginia Dignum. Operetta: a prototype tool for the design, analysis and development of multi-agent organizations. In *AAMAS (Demos)*, pages 1677–1678. IFAAMAS, 2008.

73 N. Oren, M. Luck, and S. Miles. A model of normative power. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 815–822, 2010.

74 Pablo Noriega. *Agent mediated auctions: The Fishmarket Metaphor.* PhD thesis, Universitat Autonoma de Barcelona, 1997.

75 Gordon J. Pace and Fernando Schapachnik. Types of rights in two-party systems: A formal analysis. In Burkhard Schäfer, editor, *JURIX*, volume 250 of *Frontiers in Artificial Intelligence and Applications*, pages 105–114. IOS Press, 2012.

76 Adrian Paschke and Harold Boley. Rule responder: Rule-based agents for the semantic-pragmatic web. *International Journal on Artificial Intelligence Tools*, 20(06):1043–1081, 2011.

77 H. Prakken and M. Sergot. Contrary-to-duty obligations. *Studia Logica*, 57:91–115, 1996.

78 Cristian Prisacariu and Gerardo Schneider. $\mathcal{CL}$: An action-based logic for reasoning about contracts. In *WoLLIC '09: Proceedings of the 16th International Workshop on Logic, Language, Information and Computation*, pages 335–349, Berlin, Heidelberg, 2009. Springer-Verlag.

79 Alessandro Ricci, Mirko Viroli, and Andrea Omicini. Give agents their artifacts: the A&A approach for engineering working environments in MAS. In Edmund H. Durfee, Makoto Yokoo, Michael N. Huhns, and Onn Shehory, editors, *6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2007)*. IFAAMAS, 2007.

80 G. Sartor. Fundamental legal concepts: A formal and teleological characterisation*. *Artificial Intelligence and Law*, 14(1):101–142, 2006.

81 M. Sergot. A computational theory of normative positions. *ACM Transactions on Computational Logic (TOCL)*, 2(4):581–622, 2001.

82 V. Torres Silva. From the specification to the implementation of norms: an automatic approach to generate rules from norms to govern the behavior of agents. *JAAMAS*, 17(1):113–155, 2008.

83 Ellis Solaiman, Carlos Molina-jimenez, and Santosh Shrivastava. Model checking correctness properties of electronic contracts. In *Proceedings of the International conference on Service Oriented Computing (ICSOC03)*, pages 303–318. Springer-Verlag, 2003.

84 N. Tinnemeier, M. Dastani, and J.-J. Ch. Meyer. Roles and norms for programming agent organizations. In Decker, Sichman, Sierra, and Castelfranchi, editors, *Proceedings of the Eight International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, pages 121–128. IFAMAAS/ACM DL, 2009.

85 N. Tinnemeier, M. Dastani, and J.-J. Ch. Meyer. Programming norm change. In van der Hoek, Kaminka, Lespérance, Luck, and Sen, editors, *Proceedings of the Ninth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, pages 957–964. IFAMAAS/ACM DL, 2010.

86 N. Tinnemeier, M. Dastani, J.-J. Ch. Meyer, and L. van der Torre. Programming normative artifacts with declarative obligations and prohibitions. In *Proceedings of IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, pages 145–152. IEEE Computer Society, 2009.

87 M. Birna van Riemsdijk, Koen V. Hindriks, and Catholijn M. Jonker. Programming organization-aware agents: A research agenda. In *Proceedings of the Tenth International Workshop on Engineering Societies in the Agents' World (ESAW'09)*, volume 5881 of *LNAI*, pages 98–112. Springer, 2009.

88 F. Zambonelli, N.R. Jennings, and M. Wooldridge. Developing multiagent systems: The Gaia methodology. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 12(3):317–370, 2003.

Regulated MAS: Social Perspective

Pablo Noriega[1], Amit K. Chopra[2], Nicoletta Fornara[3], Henrique Lopes Cardoso[4], and Munindar P. Singh[5]

1 IIIA-CSIC, Spain
2 Lancaster University, UK
3 Università della Svizzera italiana, Switzerland
4 Universidade do Porto, Portugal
5 North Carolina State University, USA

Abstract

This chapter addresses the problem of building normative multi-agent systems in terms of regulatory mechanisms. It describes a static conceptual model through which one can specify normative multi-agent systems along with a dynamic model to capture their operation and evolution. The chapter proposes a typology of applications and presents some open problems. In the last section, the authors express their individual views on these matters.

1998 ACM Subject Classification I.2.11 Distributed Artificial Intelligence

Keywords and phrases NormMAS, Norms, Open Interaction

Digital Object Identifier 10.4230/DFU.Vol4.12111.93

1 Introduction

Central to the idea of a normative multi-agent systems is the important distinction between regimentation and regulation, first drawn in the present technical setting by Jones and Sergot [42]. The key distinction is that regimentation arises in a system that forces or precludes certain actions whereas regulation arises in a system that neither forces nor precludes the relevant actions but merely regulates the participants so that those actions respectively occur or fail to occur. Specifically, when we think of multi-agent systems—which by definition consist of multiple agents—subject to various constraints, regimentation becomes ensuring the agents are simply (i.e., "physically") unable to violate some constraints whereas regulation becomes ensuring the agents choose not to violate the constraints, despite being able to violate them.

In other words, the notion of regulation is central to the idea of autonomy and thus legitimises the general idea of norms as we understand them. For if an agent were simply unable to violate a constraint, the constraint would appear less like a norm and more like a physical law, such as that of the conservation of energy.

A note on terminology: for our present purposes, we treat a regulation as a norm that is of social provenance and applies on the interactions of the participants. In this manner, we would not include within regulations the following kinds of norms: personal norms (never mislead my friends) and social conventions (greet everyone at the start of a meeting).

Regulation as the idea of control despite autonomy involves the obvious idea of a life cycle of regulations being promulgated, obeyed (or disobeyed), enforced, updated, and revoked. The notion of regulations presumes what we term a *normative architecture* or a social backdrop. The notion of regulations, however, is more general than any such normative architecture in which they may exist—regulations can exist in any setting where multi-agent systems make sense. Specifically, we can see regulations being applied in a setting involving a designated

Normative Multi-Agent Systems. *Dagstuhl Follow-Ups, Vol. 4.* ISBN 978-3-939897-51-4.
Editors: Giulia Andrighetto, Guido Governatori, Pablo Noriega, and Leendert W.N. van der Torre; pp. 93–133
DAGSTUHL FOLLOW-UPS Dagstuhl Publishing
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany

governor [54] for each agent, just as much as in a setting involving social control. And, regulations can be promulgated by a central authority just as much as being democratically decided. Our emphasis, as multi-agent systems researchers, falls on normative architectures that respect the autonomy of the participants and carry a conceptually decentralised flavour. Even such architectures may be realised in systems wherein the participants elect a governor who either enforces the regulations they jointly promulgate or promulgates and enforces regulations on its own initiative. It is not surprising that many computational architectures reflect one of the cases where a governor is either elected or appointed, even if the remaining participants might be notionally peers of each other.

1.1 Example

To motivate this discussion further, let us consider the example setting of traditional commerce. Commerce clearly involves autonomous parties: buyers and sellers at the very least and often members of an extensive ecosystem of suppliers, shippers, payment processors, and ratings agencies. Since the parties are autonomous, regimentation is out of the question: you cannot prevent a seller from selling illegal goods or from failing to ship goods that the buyer has paid for nor the buyer from refusing to pay for goods ordered. But regulation is essential. The buyer and seller can regulate their transaction to some extent by entering into a contract that specifies the transaction. Or, they may adopt the regulations in place in their social environment, such as the city or marketplace where they operate. In either case, in general, each party would rely upon another entity to help enforce a regulation when its interests are at stake in the satisfaction of the regulation. This entity could be a government agency, an industry board, or a nominated third-party that the participants agree upon. Although the parties may also function without such an entity to back the regulations, such a situation becomes rarer as the stakes go up.

Now what would change if we move to electronic commerce? Clearly, the same or similar roles are still involved. It is obvious that there is an equal need for regulation in virtual settings as well. E-commerce is usually facilitated through marketplaces such as eBay wherein the buyers and sellers can meet to conduct business. The marketplace serves as a promulgator and enforcer of regulations. Thus the most common form of e-commerce employs a centralised architecture. In the days of Usenet prior to the Web, it was common for users to find each other and conduct transactions without a formal marketplace. This is analogous to transactions where the parties find each other through Craigslist today. Such transactions might involve one party sending a payment to another to purchase the specified goods. In such settings, too, regulations apply though there is no ready means to enforce them. Potentially, in some countries such as in the US and Europe, the government can get involved if broader regulations against fraud in commerce appear to be violated.

Let us imagine that we have a virtual marketplace. Clearly the entities that live in such a marketplace are not people but their software agents. This leads to the question of to whom do the regulations apply: the agents or the people? An important idea here is one of responsibility and accountability, for example, as delineated by Mamdani and Pitt [50]. One can imagine a virtual world populated by software agents where each agent acts on its own behalf and is therefore subject to the virtual world's regulations. But in the more common uses of agents, especially in settings such as commerce that involve an external reality, the agents are merely representatives of people. The agents could be intelligent and function without minor guidance but insofar as they are representatives of people, the regulations apply to people, who must bear the consequences of obeying or disobeying them. The situation is analogous to a business owner using an accountant to prepare a government filing. It is the business owner who is subject to government regulations and would bear the consequences of complying or not, unless the accountant has violated regulations pertaining to the field of accountancy.

Restricted in this manner, we view each agent as representing a principal. The computational system, such as the marketplace above, facilitates interactions among the agents by supporting the necessary bookkeeping but does not have a life of its own. That is, the computational system is merely an instrument. The life belongs to the participants, including the principal whom the marketplace viewed as an agent represents. Actions in the computational systems *count as* actions in the real world where they are subject to appropriate regulations.

1.2 Layout of the Chapter

The main objective of this chapter is to formulate the research challenges of regulated MAS. The remainder of this chapter is organised as follows. Section 2 takes a closer look at some typical applications of regulated MAS. Section 3 provides a deeper motivation of normative architectures wherein regulations are feasible, discussing the common traits of normative architectures. Section 4 describes the main components of a conceptual model for regulations. Section 5 introduces how such a conceptual model can be operationalised in the above architectures. Section 6 discusses some dynamic aspects of regulations, especially how they evolve over time. Section 7 relates regulated MAS with other perspectives in computer science, with an emphasis on the software engineering of sociotechnical systems. Section 8 illustrates a real-life example of a regulated MAS drawn from the domain of open innovation. Section 9 summarises some open research problems pertaining to regulated MAS. Section 10 provides a soapbox for each of the authors to describe their personal views.

2 Normative Applications

Regulated MAS serve one main function, to set a “level playing ground” (as D. North [55] postulates for institutions). Putting it bluntly, regulated MAS create an *institutional reality* that is different from the physical reality. In the institutional reality, only *institutional facts and actions* exist. As we discuss in the next section, these two realities have some correspondence thanks to the “constitutive” norms. Those norms produce, on one hand, the legitimacy of the regulated MAS to create an institutional reality that is governed by regulations and to enforce these on participants and, on the other hand, the entitlements needed by agents to act within that institutional reality and consequently held liable in the actual world. In addition, those constitutive norms also determine the ontology that will exist in the institutional reality and the *counts-as* relationship that establishes a correspondence between facts and actions in the physical world, and institutional facts and actions (see Searle [62]). The ultimate purpose of a regulated MAS is to articulate interactions where several *autonomous* agents may be involved. By fixing ontology and regulating admissible and legal actions, the regulated MAS reduces uncertainty and simplifies decision-making “surrounding agents with reliable and perceivable patterns of events that allow them to make reasonable and stable calculations about behaviour” [65, p. 78]. Furthermore, governance, by providing some degree of control over undesired behaviour, serves to allocate risk and limit the exposure and liability of agents.

A regulated MAS is supposed to be implemented *properly* (with respect to the three integrity challenges discussed on Page 99), and to make the aforementioned institutional reality work. The implementation also needs to support five components that are essential for that *normative architecture* we mentioned above:

- A "virtual space" where agents may interact.
- A "shared ontology" to which all agents may univocally refer to.
- An "interaction model" that determines what the primitive or atomic agent actions are and how they may be interlaced into activities involving many agents.
- A "set of regulations" that affect agent interactions. In addition a collection of norms of different sorts, this set may contain an explicit mention to regimented constraints and may also include (*meta*)norms that regulate how existing norms may be modified or revoked and how new norms may be introduced in the set.
- A "governance model" which consists of two complementary elements. First, some principles about compliance (i.e., constraints—within the regulated MAS—whose enforcement is regimented, norms whose enforcement is discretional—in the sense that the sanctions may be imposed or not depending on the judgement of the enforcers—and yet another type of norms whose application is not discretional). Second, enforcement mechanisms that contend with violations; that is, how infractions are ascertained and then dealt with (e.g., a centralised mechanism, some law enforcement agents, self-regulated peer-to-peer control).

What are the conditions that make a regulated MAS useful in practice? There is a threefold answer:

First, when the agents whose participation is regulated have the following features: (i) they are autonomous towards observance of norms (ii) their internal decision-making and motivation is beyond the control of the regulated MAS, (iii) the agents may be malicious or incompetent and thus are likely to infringe norms, (iv) may "enter or leave" (be active in) the virtual space at will, (v) are independent from each other or do not represent the same principal and thus may have different and perhaps conflicting individual goals, and (vi) there is some liability in their actions.

Second, when the situations where these agents meet have the following features: (i) regulated activities are repetitive (no need to regulate one-shot situations), (ii) activities are performed in a shared social context (at any point in time several individuals share the same state of the world, the same regulations apply to all and any institutional action by any of them affects the shared state of the world) (iii) regulated interactions are perceivable and applicable conventions are ostensible.

Third, the regulated MAS is backed by an organisation and supported by technological artifacts that resolve the "integrity challenges" that are discussed in Section 5. Namely: (i) create the stable interaction environment and maintain it in proper operation, (ii) manage access and "identity" of participating agents, (iii) implement the governance model, (iv) assure the persistence and enforcement of regimented constraints as well as the sound management of regulations.

2.1 Towards a typology of regulated MAS applications

Since the notion of a normative MAS is recent and few multi-agent system approaches include normative aspects explicitly as part of their conceptual framework there are not many examples of actual regulated MAS applications. Nevertheless, the examples reported in the last chapter of this book provide a suggestive indication of the types of applications of regulated MAS that are being or will be developed. Other examples of applications, which may qualify as regulated MAS in the sense just described, may be drawn from a number of "standard" MAS that have been developed from an organisational or an institutional standpoint. Additionally, some conventional systems and even some multi-agent systems that organise social or collective activities may be reified as regulated MAS, even if the actual normative component in some of these maybe flimsy. Note, however, that for all these examples, the provisos stated in Section 7 should be kept in mind.

A look at those three sources of examples allows us to venture four distinct types of applications that involve agents, situations and organisational functionalities for which regulated MAS are appropriate.

Hard-wired sociotechnical systems. Systems where the conventions that regulate agent interactions are established at design time and are issued and maintained by the owner of the system [69]. Although many of these conventions are hard-wired into mostly static workflows in a regimented way, there may also be some norms that might be violated and need enforcement and regulations may evolve over time. The balance between regimentation and enforceability (and the corresponding enforcement mechanisms as well as the dynamics of the regulations) respond to pragmatic aspects like efficiency, ease of use, accountability, trust-worthiness, risk, and liability. Typical examples are e-markets (for on-line sales, e.g., eBay and PayPal), enterprise information systems (for hospital management, hotels) and web-based conventional social activities (for example, e-government transactions, e-learning or some forms of mobile health care).

Agent-reified conventional regulated environments. This type includes social conventional or traditional activities that are subject to norms, but are now web-enabled in some way and involve agents that perform regulatory functions; for example, collective writing in Wikipedia. Some of these sociotechnical systems may be properly labeled regulated MAS. Examples of these type in Chapter 7 of this book are the one about norms in open source software repositories by Savarimuthu and Dam; the one by Villata and Gandon, on data licensing in the web; and the one by Fornara and Eynard on data collection.

Artificial social systems. Two distinct breeds: (i) Those systems used for modelling and simulation in fields like experimental economics, sociology or policy-making; for example, the UAV (unmanned autonomous vehicles) example by Governatori and Lam, and the water management policy simulator example by Noriega (both in Ch. 7) and (ii) Virtual worlds, for immersive remote interactions, like the ones discussed by Cranefield and Verhagen (in Ch. 7); and virtual worlds as those used in games, like those discussed by Dignum (also in Ch. 7). In both types of applications, the advantage of a regulated MAS is that design concerns about agents and the environment are clearly separated; the use of explicit formal norms allows for a more abstract and flexible specification of conventions, scenarios and performance indicators, and in addition these may support to some extent off-line and on-line reasoning by designers and agents.

Open sociotechnical ecosystems. In these systems a regulated MAS furnishes the normative architecture that enables the inception and runtime support of new organisations, agreements or collective activities that are subject to their own regulations and may be created by centralised off-line design or by on-line peer-to-peer collaboration of participating agents. Examples of these systems are the virtual enterprises scenario proposed in [45], the Ocean Observatory Initiative (OOI) reported by Singh et al. in Chapter 7 of this book and the Green Open Innovation Platform described below in Section 8.

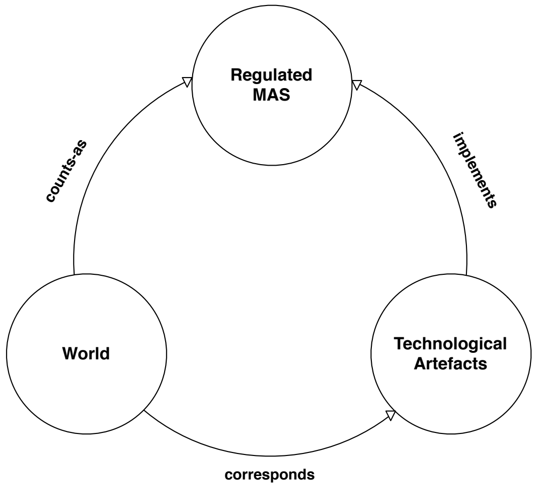


Figure 1 The relationships between the parts of a sociotechnical system.

3 Normative Architecture

The foregoing sections made the case for regulated MAS as involving agents who represent principals. Let us now consider the architecture of a regulated MAS, which we dub the *normative architecture*, in conceptual terms. The agents function as autonomous with respect to other agents but they derive their autonomy from the principals they represent, thereby making the principals accountable for their actions. In this manner, regulations make the most sense in settings that combine the social and the technical spheres. In this regard, we can treat regulated MAS as *sociotechnical systems* [69]. Sometimes this term is used for systems that merely involve human interaction with a computer; here, we specifically mean interactions between socially autonomous entities, such as people and organisations, though the interactions involve and are mediated through technical artefacts such as computers.

Figure 1 describes how a regulated MAS may be understood. At one side is the world, which we can think of approximately as the “physical” world. More carefully, it represents the objective reality for the purposes of modelling. For example, in commerce, the world provides a home for the goods being bought and sold as well as the infrastructure such as for shipping and paying that commerce needs. In a multi-user virtual environment, the world may simply be the virtual reality in which avatars of the users exist and function.

At the other side lie the technical artefacts. For our purposes, these are computational representations of the world as well as means (programs) to manipulate them. We can resolve the distinction between representation and reality simply in terms of how we choose to model. Whatever is endogenous falls into the technical artefacts and whatever is exogenous falls into the world. For example, although physical goods, vehicles, currency are clearly the province of the world, we might treat the services that deal with such objects as internal or endogenous. A key distinction is that whatever is endogenous can be manipulated computationally from the model and vice versa. Referring back to Section 1, the endogenous parts can be regimented (the bank account is never overdrawn) but what is exogenous can only be regulated (the customer shouldn’t write a check for more money than he has in his account) [69].

At the top, we place the regulated MAS itself, which exists in the world and is realised through the technical artefacts. The whole idea of commerce, for example, resides at this level. People may pass objects back and forth but only in some suitable settings can such

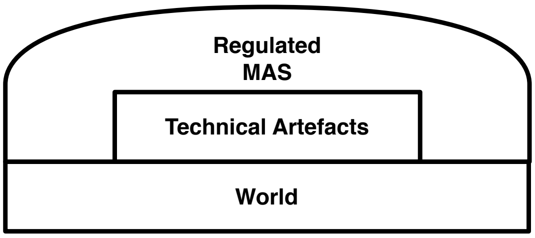


Figure 2 A sociotechnical system schematically.

actions *count as* [62] a commercial transaction. In particular, these settings involve the satisfaction of various norms, and impose regulations upon the transaction—for example, that goods can only be sold by someone who has ownership of them (not just possession) and a successful transaction results in a change of ownership.

The above characterises a regulated MAS in conceptual terms. How can we realise such a system? Figure 2 describes a sociotechnical system in schematic terms. We view such a system as having three main parts. The sociotechnical system is grounded in the physical world, inhabited by exogenous resources such as vehicles (autonomous or otherwise). Overlaid on the physical world is the technical (i.e., computational) structure, inhabited by information resources such as databases. Normative reality, home to the regulations, exists over the physical world and technical artifacts. In this manner, social interactions—that is, those subject to regulations—may be realised in the physical world through the mediation as needed of the technical artifacts.

The main conceptual challenge in regulated MAS is to maintain their *integrity* if only to ensure that regulations are being obeyed or to find out whether they are being disobeyed and by whom. Ensuring integrity in a sociotechnical system is nontrivial. Since a sociotechnical system brings together three concerns in its modelling, it is important to recognise that each of them may potentially be the cause of an integrity violation. First is the challenge in ensuring that the regulated MAS itself is sound: that is, the regulations do not conflict with each other and are potentially jointly achievable in principle. Chapter 2 addresses these formal challenges.

Second is the challenge of ensuring that the technical artefacts are correct. Let us put aside implementation-level concerns, such as that the sorting routine being used is correct, which we can capture in terms of the underlying infrastructure. The main conceptual aspect of integrity of the artefacts comes down to what is termed *identity* in computing parlance [20] but better corresponds to an ability to identify relevant objects. Identification is a crucial function of sociotechnical systems. Not surprisingly, the ability to identify principals is crucial for accountability: if we don't know who's who, we have no legitimate basis for determining if a regulation is being obeyed or disobeyed. The ability to identify technical artefacts is equally as crucial because it enables tracing actions to their principals and thus potentially determining who was responsible for obeying or disobeying a regulation. In traditional human societies, for example, a small village, each party would be known to every other party. In larger human societies, identification becomes difficult. In modern settings, the government, which plays the "governor" role alluded to in Section 1, additionally provides a means to identify the participants. Likewise, in virtual settings such as eBay, the resident authority (think of it as the governor of the eBay marketplace) provides the identification.

The provisioning of identification is a common feature of regulated settings. An ability to identify a participant is a prerequisite for regulating its interactions. Identifiability can be achieved definitionally at the regulated MAS level. It can be trivially implemented at the technical level: simply give everyone an account with which they must login before participating in the sociotechnical system. Achieving identifiability at the physical level is nontrivial and the challenge segues into the one below.

Third is the challenge in ensuring that the physical world, which recall provides the infrastructure upon which actions and events of relevance to the regulations arise, is not corrupted. For example, if the underlying messaging system is corrupted and delivers bogus messages or fails to deliver correct messages, the integrity of the regulated MAS would be violated. In general, such threats from the infrastructure are a major security threat to a regulated MAS. Figure 2 captures the intuition that the regulated MAS ought to control the relevant aspects of the physical world. For example, eBay viewed as a regulated MAS controls the infrastructure through which auctions are created by sellers and bid submitted by buyers. Such control is essential for eBay to determine which bids were in time, who if anyone won an auction, what item did they buy, and how much they committed to pay for it. In essence, the regulated MAS takes responsibility for the physical world as a way to ensure its own integrity. The above assumption of control, however, may not hold in practice—and arguably is never met in practice. For example, underlying eBay's infrastructure and its execution by users lie computers and networks that eBay does not control. Thus we would need techniques for regulated MAS to function correctly despite threats from the infrastructure. Indeed, human societies are not paralysed because of the existence of such threats and norms can provide a way of dealing with them.

4 Conceptual Model

This section describes the fundamental application-independent components that all open interaction systems have in common. We base the foregoing claim, on our experience [19, 25, 45]. Those application-independent parts should be integrated with application-dependent concepts and concrete values of some parameters in order to realise an actual interaction system. The main advantage of this model is that, in principle, it may be used for the specification of a different type of systems used in different domains, from the definition of marketplaces for the improvement of e-commerce to the definition of collaborative/coordinating/social ecosystems for supporting collaborative work and social coordination. In the following we will describe what we consider are the main components for the design of those systems, how those components should be used for the realisation of a system, and the main functionalities that should be implemented in those systems for their correct execution. A fundamental assumption behind the definition of this model is that the interacting parties are *autonomous entities*; that is. these agents may be human or software, each agent has its own goals, each one belongs to a specific principal and, furthermore, each of these agents may violate the norms that regulate its interactions with others.

4.1 Design Components

We may distinguish between those components whose main goal is to *enable social interaction* among the participants agents, and those components whose main goal is to *regulate such interactions.* Regarding the first type of components, we need to specify a set of *conventions* for realising conversations or interactions. Those conventions regard:

- The definition of the common *ontologies* that the agents need to share in order to interact. They consists of concepts and properties used for describing the objects on which the interaction is focused and the shared knowledge of the interacting parties that evolves during the interaction. Those ontologies have an application-independent part that defines for instance the concept of action, event, obligation, and so on, and an application-dependent part focused on the domain of the application. Some of the concepts represented in those ontologies have a direct mapping to objects in the real world, some other exists only in the institutional reality of the regulated MAS.

- The definition of those *actions* that agents may perform. We assume that within the regulated MAS, agents use only *communicative acts* for interacting. Those acts can be defined in terms of the *preconditions* for their successful performance and in terms of their *effects* on the state of the interaction, that is, on the state of a set of application independent and dependent components whose existence is jointly accepted by the interacting agents. For example, the effects of some communicative acts may be to create or modify the *normative relationship* among the agents: an agent that bids for a product in an auction commits to pay the amount of the bid. The fact that by performing a communicative act an agent can change normative relationships at runtime, can be modelled using constitutive rules [62] (X counts-as Y in C) for associating an appropriate semantics to the communicative acts that agents perform within a specific protocol enactment, except for declarations [27].
- The definition of the relevant *events* that may happen during an interaction, (again in term of preconditions and effects), the most common of which is the passing of time. Events may be referred to in the content of communicative acts, for example, as a promise to perform an action before a given deadline.

For example, in the definition of a marketplace the domain language might contain a definition of the different types of products that may be exchanged, their properties, some constrains that specify the conditions under which the values of their properties are correct—for example rules for changing the price of a product—and some terminological properties that are valid for the domain in question—like the fact that a bank transfer is a particular type of payment. The actions can be the act of buying, selling, paying, and delivering a product.

Regarding the components of the conceptual model devoted to regulate the interaction or conversation of the agents, it is important to observe that very often in human social life, interactions happen in a specific *context*; for example in a school, in a bar, in a market, in a university. The context is useful for disambiguating the meaning of certain terms and their properties, and for the fact that it introduces some predefined normative relationship among the agents playing certain roles. This is usually done with the aim of bringing the interaction to a certain final goal, for example an auction can be used to reach an agreement on the price of a given product, an exam is used to grade a student. Moreover, this is useful for avoiding the complex task of starting every negotiation from scratch, where agents need to reach agreements on the rules of the interaction.

As the preceding remark suggests, it is fundamental to clearly define the *context* where the interaction will take place and the *norms* that will regulate the interaction in that context. It is also important to regulate agent interactions in such a way that agents may benefit from their autonomy. In order to make it possible for the interacting agents to profit their autonomy, it is important to regulate their interaction in a way that allows them as much freedom as possible to choose what communicative action to perform among the set of available ones. Therefore it is necessary to be able to express a list of normative constrains that can specify:

- What actions are permitted. One possible default for the system is that if an action is not permitted, it is prohibited, and the prohibition can be violated. Another default may be that every action by default is permitted, unless it is explicitly prohibited, and the prohibition can be violated.

- What actions are prohibited in a regimented fashion (prohibitions that cannot be violated). In the OCeAN model [25] in order to avoid confusion between this notion of prohibition and the previous one, this notion of prohibition is formalised using the notion of *institutional power* [43]: if an agent does not have the institutional power to perform an institutional action (an action that changes the value of a property whose value is shared by all the agents involved in the interaction, the effects of the attempt to perform the action are void;
- What actions are obligatory, and obligations can be violated. Very often an obligation has an associated deadline that specifies the instant of time when the obliged action has to be performed.

Given that those context-dependent norms are often defined at design time, they are expressed in terms of the *roles* that the agents will be able to play at run time. Those roles very often are related by subsumption and incompatibility relationships that create a social model for the agents.

Usually those normative constraints are in force in a specific and delimited context of interaction that should be clearly represented in the model and should be distinguished from other contexts where usually other norms, other roles, and other ontologies apply. Those contexts of interaction may be called *scenes* [54], *spaces* [73], *orgs* [69], or groups (when their distinguishing characteristic is represented by the agents involved in the interaction). The introduction of the construct of contexts of interaction as an application independent component of the model requires to define the rules for regulating the creation at runtime of new contexts of interaction, for letting agents enter and exit from contexts, and for destroying a given context.

4.2 Design Functionalities

In order to actually enable an open interaction among autonomous-heterogenous agents belonging to different principals, the previously described components should be properly formalised for the specific application domain where the interaction system will be used. For example, in the definition of an e-marketplace the different types of auctions (English, Dutch, double, . . .) and contracts should be specified (these are the contexts of interaction), in terms of domain ontologies, roles (seller, buyer, auctioneer, participant, . . .), actions (buying, selling, paying, bidding, . . .), and norms for regulating the performance of the actions available in a given space.

Once a system is defined, it is necessary, before its execution, to test if the formalisation has some specific properties. For example, to test that the definitions of all the ontologies are consistent, that the defined roles are all used, that the preconditions for the performance of the intended actions may be satisfied and that that the applicable norms are not in contradiction—in spite of the fact that when norms refer to intervals and instants of time, deadlocks are difficult to check at design time and hence usually need to be dealt with at runtime. Another important functionality is the one for proving that all the possible evolutions of the interaction system are constrained within a given set of boundaries that will allow the interacting agents to reach certain predefined goals.

Usually the process of realising those functionalities depends on the language used for formalising the system at design time (see 9.1.1 below). For example, if logical languages like decidable description logics (DL) (that are the basis of Semantic Web Languages like OWL) are used, then some of those properties may be checked thanks to the use of available DL reasoners (like Pellet or HermiT[1]).

Once a formal specification is finalised and some of its properties are checked, it will be used at runtime for actually executing an open dynamic interaction system whose components, i.e., agents and norms, may dynamically change during the interaction. This process will be described in detail in the following section.

5 Operational Model

Given the conceptual model introduced in the previous section, here we describe the challenges that are important to take into account from an operational perspective. That is, we identify the specific concerns that need to be addressed when developing a normative multi-agent system.

5.1 Operational Components

The first challenge one needs to address when designing a normative multi-agent system is *promulgation*: when and how are norms created into the normative environment within which norm-regulated interactions take place. There are (at least) two possible approaches.

Design-time norms. The most common approach regarding norm promulgation consists of relying on the system designer to anticipate the possible encounters that are to be governed using a normative approach, and to define the norms most suitable to those encounters. While keeping the possibility of addressing an open scenario in which interacting agents are concerned, this approach assumes that norm promulgation is a design challenge. Such perspective fits many real-world applications identified in Section 2.

Runtime norms. Some applications of normative multi-agent systems inherently require that the norms applicable to a certain interaction are at least adopted at runtime. A typical case is electronic contracting [45]: when establishing a contract, two (or more) agents negotiating on behalf of their principals choose the normative setting that will regulate the enactment of such contract (e.g., by specifying the type of the contract they are establishing). More sophisticated norm specification mechanisms include, on one side, the negotiation (or configuration) of norms, as well as their assembly and, on the other side, the emergence of social norms as discussed in Chapter 5.

Another operational concern in a normative multi-agent system is related to the *observability* of agent actions. This concern is directly related with the functional purpose of a norm, which is to be able to distinguish compliant from noncompliant behaviours. It is therefore a crucial aspect of norm *monitoring* to be able to assess the relevant agent actions that enable the normative environment to determine whether norms are being complied with or not. Observability is particularly tricky when dealing with prohibition norms. In some applications (e.g., auditing), when dealing with obligations an assumption of self-motivated compliance demonstration can be in place, i.e., it is in the best interest of agents to publicise fulfilment. Detecting prohibition violations, on the other hand, is much harder because it demands for a pervasive character of the monitoring system; in practice, different verifiability levels exist [76] regarding the actions agents can perform.

[1] http://clarkparsia.com/pellet/, http://hermit-reasoner.com/

Closely related with this challenge is the choice of implementation of norms [38], which is tightly coupled with the freedom agents are allowed when playing in the normative environment. Two approaches have been pursued: (i) *regimentation*, which prevents unwanted outcomes by imposing constraints; (ii) *enforcement* [37], which consists of using mechanisms to influence the decisions agents make, letting them choose whether to fulfil or violate norms.

Regimented norms are only viable in totally controlled scenarios, in the sense that agent autonomy is reduced by making sure that disallowed behaviours do not have any effect in the system (this is the original approach to engineering electronic institutions [54, 21]). On the other hand, enforcement means that external (as far as the agents themselves are concerned) mechanisms must be put in place so as to influence agent behaviour. Examples of such mechanisms include, among others, sanctions (either normative or utilitarian), incentives, or reputation.

Sanctions, in particular, may be employed following two different (nonexclusive) policies: retribution aims at compensating the addressee of a violation, while deterrence puts an emphasis on punishing the violator so as to discourage future violations. Taking into account this distinction, it may be desirable to incorporate both sanctions that concern actions to be performed by the violator and sanctions that materialise as actions that the norm enforcer is authorised to perform [24].

As stated above, monitoring is a central operational component in normative multi-agent systems. In rich social interaction scenarios there may also be a need to address *blame assignment*, in the sense that norm infraction may not entail guilt of the agent that is the subject of the violation taking place.

Summarising, when designing a normative multi-agent system a number of core operational components have to be engineered. This section has identified norm promulgation, observability, monitoring and enforcement as the key elements to be addressed. Additional concerns are raised in Section 9.

5.2 Operational Functionalities

From a functional perspective, and from the point of view of interacting agents, when designing a normative multi-agent system, a number of concerns need to be addressed. These are related to the operational components identified above.

The first functionality is important in open systems, and concerns the possibility for agents to enter and leave a normative multi-agent system. When entering the system, an agent adheres to the set of norms that regulate agents behaviour. Admission may encompass specific constraints that must be met. Leaving the system may demand for checking certain conditions (e.g., the agent must not have any pending obligations).

Another important functional aspect is the establishment and updating of norms at runtime. When enabling this possibility, it is necessary to engineer appropriate mechanisms that allow agents to negotiate the norms that guide their further interactions and to feed those norms into the normative environment. Accommodating specific infrastructures may facilitate this task, e.g., through coordination artefacts [40] or normative frameworks [46].

Naturally, the next functionality concerns enactment, i.e., designing the means through which agent interactions are assessed by the system. This is intrinsically related with the challenge of observability, and is crucial to allow for monitoring to take place.

6 Evolution

In general terms, when considering the evolution of a normative system, one needs to consider two kinds of changes: norm *promulgation*, which consists on establishing a new norm; and norm *derogation*, which removes a norm. Modifying a norm, e.g., by changing its applicability conditions, may be seen as a derogation followed by a promulgation.

A first step towards encompassing evolution in the normative multi-agent system has already been identified in the previous section, and concerns normative dynamism from the agents' point of view. When norms are to be created at runtime [45], the normative environment denotes an evolving facet. It may also be the case that within a particular institution, different organisations are established at runtime. Those organisations may have their own normative structure [75].

Typical in these approaches is to embed into the normative system some kind of infrastructure determining the normative changes that can be introduced. This infrastructure, specified at design-time, dictates the changes at runtime that agents may introduce in the normative system (their *degrees of freedom* [1]).

Dynamism may also be an important and desirable property from the normative system's point of view. In this case, two questions need to be addressed. The first question relates to how to evaluate the performance of the system as a whole. The second one relates to the changes that can be introduced in the normative multi-agent system so as to improve its overall performance.

Performance evaluation demands for an analysis of how effective a normative system is in terms of regulating the multi-agent environment. One may approach this issue by observing the behaviour of agents and assessing if the population as a whole complies with the norms: if it does not, some changes in the normative system may need to be introduced, e.g., by adjusting enforcement levels (as in [49]). In some cases, however, excessive or inadequate regulation may hinder the system by preventing some better overall performance from taking place. In such situations, although agents may be complying with norms, it might be in the best interest of the system for them to do otherwise.

Noncompliance can therefore be seen as an opportunity to change the normative system, by considering violations as alternative enactments that must be reacted-upon through changes in the normative structure. Those changes are not targeted to deviating agents, but instead towards the normative system as a whole.

Another aspect of change to take into account are the propagation effects of norm introduction and updates (see Ch. 2).

A possible approach to enable a runtime evolution of the normative system is to allow the designer to specify at compile-time a *normative transition function* that specifies the feasible norm changes and the conditions for their realisation [7, 74]. This approach however assumes it is possible to anticipate each normative update that the system may need.

7 Mapping to Other Perspectives

In this section, we discuss some approaches and conceptions of multi-agent systems that bear some similarity to regulated multi-agent systems (MAS), but turn out to be fundamentally different. In all of these, a regulated MAS-based modelling approach would potentially be a better fit.

7.1 Sociotechnical Systems

Traditionally, the field of sociotechnical systems has been concerned with the interplay between humans and technology in an organisational setting: how technology affects humans and vice versa. It developed with the recognition that the efficiency-focused design of work removed from the concerns of the workers or end-users (the social components) and the culture of the organisation tends to end up being self-defeating. This theme was picked up in software engineering in two major ways: (1) how to understand, elicit, and model the requirements for sociotechnical systems and how to manage change [53, 33, 2], and (2) how to model sociotechnical systems themselves [83]. Often, there was substantial overlap between the two, for example, as in i* [83]. Closer to the multi-agent systems literature, Flores et al.'s well-known work on the design of work [23] is best seen in the light of sociotechnical systems.

However, unlike regulated MAS, models of sociotechnical systems developed in software engineering are invariably conceptually centralised. This means that conceptually there would be a single thread of control in the system. Baxter and Sommerville [2], for instance, ascribe goals to the system; the system would potentially adapt in pursuit of its goals. Further, in their conception, systems are hierarchically decomposed into subsystems, each of which may either be a social component or a technical component. Yu's [83] models are peer to peer; however, his approach is mentalist and therefore implies conceptually centralised systems.

Models of sociotechnical system may well support physical distribution via the notion of components. One could then talk about the "interactions" among the components. These interactions, however, are merely *technical* in the sense that they are merely the means to an end—the system goals. Thus even though a system may be physically distributed, it remains conceptually centralised. In fact, it turns out that all of traditional software engineering has a conceptually centralised perspective on systems [14]. The reason is that even the most fundamental conception of systems in software engineering is *machine-oriented*. In other words, the primary objective of software engineering is to design machines that transform inputs to suitable outputs. The conceptual centralisation in software engineering is not surprising given that sociotechnical systems research, even outside of software engineering, has largely confined itself to organisational settings.

Regulated MAS represent a more general model for sociotechnical systems. With regulated MAS, one can model inter-organisational settings, which cannot be modelled with current approaches. Further, one can potentially argue that even for intra-organisational applications, regulated MAS-based models would be superior to the top-down models of traditional software engineering, because, after all, even *within* an organisation there would be multiple autonomous agents. Many of the approaches advocated by those in sociotechnical systems research and software engineering (for example, ethnomethodology and other methods from the social sciences) could potentially be employed toward the design of regulated MAS just the same as a centralised sociotechnical systems.

7.2 Agent-Oriented Software Engineering

Can agent-oriented software engineering (AOSE) be employed for designing regulated MAS? The answer is no. Although, many AOSE approaches give prominence to interaction, they take a conceptually centralised perspective. Some AOSE approaches are logically centralised approaches geared toward efficient problem-solving. Jennings [41], for example, describes a scheduling problem that is addressed by *distributing* it across intelligent agents. Both Zambonelli et al. [84] and Vázquez-Salceda et al. [77] acknowledge the distinction between open and closed systems and emphasise interactions. However they falter in important details, betraying if not conceptually centralised mindsets, at least considerable conceptual difficulties. For example, Zambonelli et al. (p. 328) identify the "resources that the MAS can exploit, control or consume when it is working toward the achievement of the organisational goal." Vàzquez-Salceda et al. model the objectives of social systems as goals; further, the social systems are themselves controllers (p. 338): "Facilitation roles are usually provided by agents controlled by the society, and follow a trivial contract." Tropos [4], another prominent AOSE methodology, is goal-oriented and advocates a top-down system design process starting with stakeholder goals and ending with the coded "agents".

7.3 Design Norms

A regulated MAS displays the following characteristics.

- The norms are social in that they are relations among agents.
- The norms are social in that the state of a norm would progress only due to explicit communication among agents. Logically, the social state of a regulated MAS is a conjunction over the states of individual norms. The specification of how communication affects the norms is referred to as a protocol.
- The social state of a MAS may be computed by observing the communications in the MAS. The social state is distinct from the internal state of any of the agents in the MAS; in general, there may be no overlap between the two (although in distributed systems, there would be no unique social state because of asynchrony; instead, there would be a local social state corresponding to the messages each agent observes [15]).

Commitments, for example, are created, discharged, delegated, and so on *only* by explicit communication—in distributed systems, via asynchronous messaging—among the agents. Commitment protocols specify how commitments among agents progress with interaction [82]. Analogously to commitments, Singh extends the treatment to other kinds of norms such authorisation, power, and so on [69].

The above characteristics set regulated MAS apart from approaches where norms are inserted into agent designs by *fiat*. This includes the social laws-based approach [63], which is essentially a distributed artificial intelligence approach. In the design-by-fiat approaches, the agents are not autonomous (they do not represent real world principals); they are instead agents in the technical sense. Social laws, which are sometimes referred to as norms, are essentially constraints on agent designs (specified with respect to a state space common to all agents). By contrast, norms in regulated MAS are not constraints on agent design; in fact, they make no reference to agent designs whatsoever.

Regulated MAS is a more general approach than social laws. One can potentially design agents to conform with the norms established during their interaction; however, one cannot apply the social laws approach to models systems of autonomous agents.

7.4 Compliance

Norms in regulated MAS are intimately tied with the idea of *compliance*. Broadly speaking, an agent is compliant with a norm if it does not violate it. Hence, compliance is fundamentally a runtime correctness criterion. Further, it can be determined by observing solely the communications of the agent (both to and from) [78, 70, 60] (this is because of the notion of social state discussed above). This is a crucial point: that compliance would be determined from observations implies that one does not have to look into agent designs to determine compliance (which would be impossible anyway in systems of autonomous agents). This should not be taken to mean that one cannot design agents for compliance. Given a set of norms, one can design agents to comply with the norms. Some refer to this design problem as that of compliance; we, however, reserve the term *compliance* for the runtime sense described above.

If an agent is noncompliant, a compensating norm may kick in, and if an agents violates that as well, then another compensating norm, and so on. At some point though, some norm for which no compensatory norm is specified may potentially be violated. At that point, we say that the violation must be handled outside the regulated MAS; in other words, it must be escalated to the surrounding organisational context [75]. Hence, it seems useful to frame a broader notion of compliance that would take into account the relations among norms (for example, via compensation) and its potential escalation outside the regulated MAS. Singh's idea of explicitly representing the context of commitments in a MAS as an agent [66] could be useful in capturing this broader notion of compliance. In [71], Singh et al. present several patterns of commitments that involve the use of the context agent.

Norms in regulated MAS serve as logical bases for compliance. They not only specify the conditions for compliance, but also who is responsible to whom. Compliance has received much attention in the business process community; however, this community mostly approaches it as a design problem, for example, as in [35, 64]. These approaches resemble the design-by-fiat approach discussed above: they lack the observational perspective, having no notion of communication and social state.

8 Demonstration: GOI a Sociotechnical System for Open Innovation

The *Green Open Innovation platform* (GOI) is a sociotechnical system to support "open innovation" [12] in the realm of sustainable economy. It enables business interactions among a community of firms and individuals who are interested in potential collaboration.[2]

GOI was originally conceived as some sort of social network portal with the standard "Facebook" functionalities—group definition, private mail, forums and "like" / "dislike" markings—on top of which there would be one interaction context for inscribing a "challenge" (either as an offer or as a request) and another interaction context for a "prediction market" (to show support to challenges).

In practice, though, the design evolved so that the platform is articulated around an "agreement space" that consists of four specialised contexts of interaction (agreement preparation, agreement negotiation, agreement monitoring, and quality assurance) that make use of multiple on-line services and repositories to support social coordination (not unlike what is advocated below in Section 10.2). Additionally, the platform is designed to support *APIs* for some standardised activities like crowd-sourcing and prediction markets, plus mash-ups and partnering for ancillary services like an employment market and some environmental certification services. Figure 3 sketches that setup. The system is designed to allow humans and software agents to participate.

The system may be recognised as a regulated MAS in as much as the platform provides the first three components of a normative architecture described in section 2 in P. 96. Namely, a virtual space, a common ontology and an interaction model. In fact, the platform provides an institutional infrastructure in the sense that it regiments ontology and means

[2] GOI is a proof of concept prototype for the development of a commercial sociotechnical system. It is an on-going project involving several private companies, NGO's and universities. It is partially funded by grants from the Spanish and regional governments.

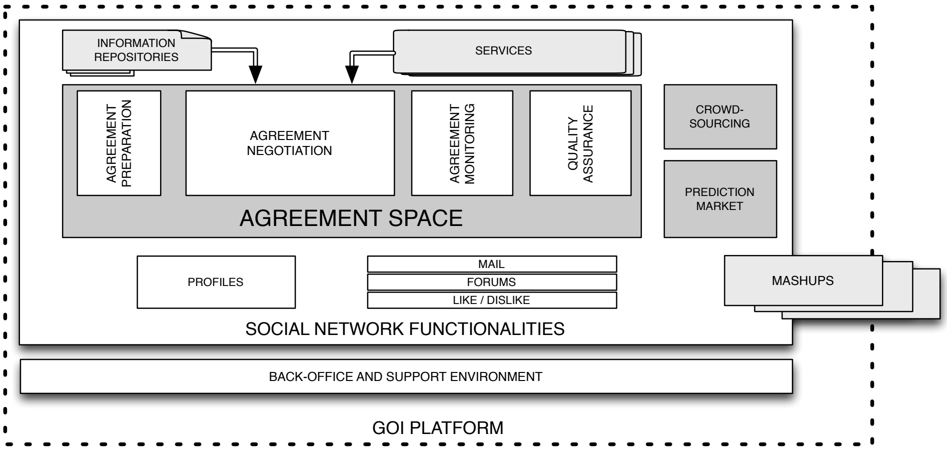


Figure 3 The GOI platform.

of communication, as well as the procedures that govern how to pass from one context of interaction to other and the procedures that certain processes must follow. For example, the platform regiments how challenges are issued, taken-up and monitored, or how to contest an active agreement. However, the agreement space also includes norms that may be violated as well as devices to contend with violations; thus containing the last two components of a normative architecture.

The agreement space, in broad terms, hosts a population of agents that are entitled to enter into agreements and are “active”—i.e., ready to be invited to an agreement, searching for partners for establishing an agreement, or being involved in the negotiation or in the enactment of an agreement—and are “ubiquitous”—i.e., agents may be involved in several agreements at a given time. The space is “open”—in the sense that agents may enter and leave at will—and it has some regimented ground rules on what are the primitive and atomic actions, the procedures that govern the basic agreement process cycle from start to end and the procedures to access and update GOI repositories and invoke services, as shown in Figure 4.

Perhaps the most interesting element of the GOI platform is that some agreements that are reached and executed inside the agreement space are in fact contracts whose clauses are negotiated among parties and their execution is monitored by the platform. The platform provides different means to facilitate this contracting. For instance, it has a repository of model contracts, whose clauses are standard and therefore negotiation is reduced to agreeing on parameters. Another resource is the availability of negotiation procedures that take the form of virtual “negotiation rooms” where a GOI member requests the platform a “room” to negotiate a contract. This member may want a room to hold an open call for tenders, hence may also request to have GOI staff run the tender for him (with software agents that perform the duties of, say, gate-keeper and auctioneer). Another member may, likewise, wish to negotiate one single model contract with N different potential partners simultaneously and would therefore require N rooms for one-to-one-negotiation, each with a software mediator with some explicit criteria for rejecting and admitting counteroffers, and a single arbiter to close the deal. Since not all contracts are honoured and some may be contested by other GOI members, the platform also provides means to resolve disputes, both in the form of automated ODR facilities or by making expert (human) mediators available. Finally, the platform keeps track of all actions where it is involved and therefore keeps information about participants

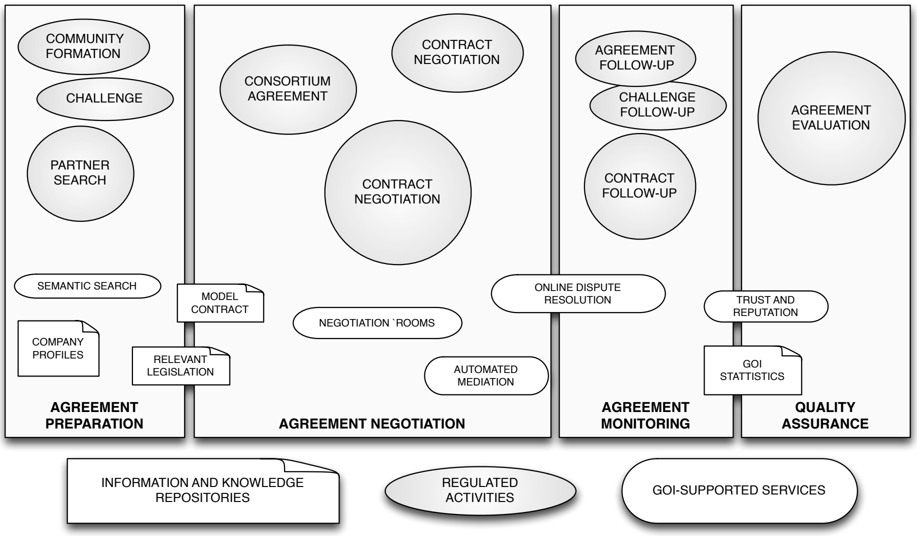


Figure 4 Main functionalities of the agreement space.

and transactions and uses it to provide different forms of quality assurance ranging from "blue book" rankings of members to specific trust and reputation measures that may be associated to particular agreements, types of agreements, communities and subcommunities, and so on. In other words, the GOI platform involves a regulated multi-agent system where each agreement includes its own normative content, is subject to some procedural and functional norms that regulate how it may be signed and how it should be executed, is also subject to the norms that govern malpractice and defaulting and, finally, some forms of quality assurance rules apply to it.

The GOI platform has a centralised design and its evolution is for the moment limited to the changes brought about by the addition of new services and interaction contexts. From an implementation perspective, the platform environment is also centralised, although agreements are distributed processes. Actually, there is also a central governance model for the basic operation, however each agreement spawns a (sub)context of interaction that is governed by its own norms and whose effects are, in principle, not propagated to the contexts of other agreements.[3]

9 A Map of Open Problems

This section outlines some open problems around the core notions of regulated MAS. The outline follows the three main phases of the regulated MAS lifecycle (design, enactment and evolution) and then focusses within those phases on some activities that are characteristic of normative environments in general. For each of those activities we mention those topics we find particularly relevant for norMAS and where we consider open problems abound.

[3] Each agreement generates a local institutional state that is part of the GOI institutional state (see Page 114) but the platform ought to guarantee the integrity and identity of agents and their institutional facts, thus some regimented control is imposed on the validation of agreements and their monitoring. Implementation follows the ideas discussed in [22, 19, 32]; hence, in essence, the agreement space is a large institution where each agreement is a new sub-institution that is specified and run peer-to-peer on demand.

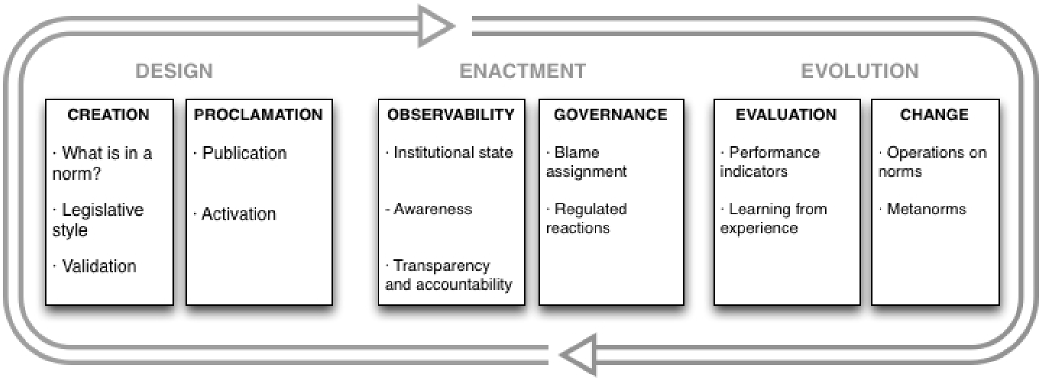


Figure 5 Some challenging topics within the lifecycle of regulated multi-agent systems.

Figure 5 summarises this outline. We elaborate on concerns introduced in Sections 4, 5 and 6 but note that while in this section we simply allude to some salient topics that we believe deserve a more thorough treatment, in the next section the authors of this chapter delineate some open problems that they find alluring.

9.1 Design Phase

Several design challenges emerge around the methodology for developing regulated MAS and the need for good enough metamodels to describe and specify them. In the case of methodology, the space for innovation is in dealing with those aspects that pertain to normative notions and how these are merged with the more conventional aspects for which methodologies have been proposed (see Section 10.2). In what corresponds to metamodels, the goal is to specify in a cohesive way all the components of the regulated MAS (see Section 5 and 10.1 below). The following subsections discuss several issues that need to be taken into account at design, in this section, however, we limit the discussion to design choices in two areas. On one hand, the *creation* of a boot-strapping nucleus of a regulated system including the particular norms and other regulatory devices (like normative roles, enforcement mechanisms or validation methods) that constitutes the original regulated MAS; and, on the other hand, a *proclamation* process that makes that original regulated MAS ready to be enacted and used by participants. Each of these areas include topics that are rich in concepts and complex in operationalisation, hence open for innovation.

9.1.1 Creation

Assuming proper methodology and metamodels are at hand, one still has to deal with the particular choices for each component of the regulated MAS in order to specify the system and make it operational. Let's review three areas of opportunity associated to the process of creation itself:

What is in a norm? The normative elements of a regulated MAS may have implicit requirements whose representation should properly take into account. These are a few that may be worth elucidating.

First, there are three key questions worth posing with respect to the intended *purpose* of the norms: (1) *What is the role that the norm is intended to play?* One may want norms to serve as a way of restricting unwanted actions and promoting desirable outcomes (hence appropriate descriptions of obligations and prohibitions are essential); likewise one could design them as a means to reducing the number of possible actions or outcomes in order to simplify the decision-making of participants (hence the representation of procedural norms needs to be paid special attention); moreover, norms may also be understood as a manner of creating a space for interaction (thus constitutive norms are paramount), or eventually the promotion of some collective objectives (in this case, then, such objectives should be properly captured in the form of the norms and their compliance incentives). (2) *What are the values that norms promote?* Values like fairness, trustworthiness, accountability are usually associated with the normative system as a whole but particular norms and their combination bias the system in one direction or another. Thus for example, full observability and strict governance may contribute to transparency, regimentation towards trustworthiness and unobtrusive governance towards flexibility. (3) *What are the pragmatic benefits of a norm?* The question here is to be able to establish a proper trade-off between the costs of observance and enforcement of norms and those of compliance and noncompliance. Some norms may have little effect on the reliability of the system but still impose agents an undue cost in deciding whether or not to comply with them; some norms may be practical for some population profiles and not for others and different sanctions may adapt better to some situations than others.

Next, there are three considerations with respect to the formal features of the *structure of the norm* to keep in mind to make sure that such structure is appropriate with respect to the desired expressiveness of the norm, the conditions for adoption of the norm and their compliance by agents and the governance mechanisms of the system: (1) The choice of the *syntax*, the *constitutive elements* involved (label, conditions of activation, deontic features, beneficiary, subject, ...) and the *ancillary elements* associated with the norm (linked norms, contrary to duty actions, ...). (2) The *crispness* of the statement of the norm and its applicability. In other words: Is the way that the norm is expressed and enacted precise enough for the fulfilment of its intended purpose? Is there an objective way of determining when an agent is complying or not with a norm? Does the expression of the norm achieve some desirable degree of flexibility (for discretional enforcement, to adapt properly to the evolution of population or changes of the application context)? (3) *Coherence of the normative corpus.* This has to do with the different formal properties that the system as a whole should exhibit. In some cases the preferred notion is that of logical consistency but in some cases it may suffice with a narrower notion of conflict-free sets of norms or, in more general terms, in choosing some particular notion of “consequence” that is appropriate for the system. Depending on those preferences, the designer would have to answer questions like: May conflicts be dealt with during runtime? Are there possibilities of deadlocks? Are there any norms that are impossible to comply with? How to enforce norms when the system has not proven to be conflict-free? Are norms conducive to the overall purpose of the regulated MAS? To what extent?

Finally, the design of the system should take into account *what an agent is intended to do with norms.* There are three capabilities that are assumed of agents in this respect: *Reasoning* about norms: What are the decision-making capabilities needed by agents in order to comply with the promulgated and active normative elements? To what extent are agents presumed to infer the state of the system in order to know whether a norm applies, has been violated and what the consequences of a violation might be? *Adoption* of norms: Is the ontology of the regulated MAS compatible with the norm (i.e., with respect to the objects, actions, roles, and such involved in the norm)? Is the governance structure appropriate for enforcing the norm (detection, blame assignment, sanctioning and reparation)? Is pertinent information about entitlements, co-dependence of normative elements, effects of noncompliance, and so on properly represented and communicated to implicated parties? *Compliance* with norms: What behaviour is perceived and by whom? What information about a norm should an agent be informed of: social values associated, whether or not it is active, conditions of application, regulated reactions, subjects of the norm, consequences of not-compliance, liability, enforceability conditions?

Legislative style. The designer needs to make choices about the normative features that have to be functional when the sociotechnical system is originally enacted and how those features should evolve once the system becomes active. Three matters of concern may affect those choices and all three have plenty of open problems.

Locus of control: The issue is to determine who controls the changes that take place in the normative system. One extreme is a "demiurgic" style: the designer that creates the regulated system has control of it as a whole and introduces changes as needed. The other extreme is where a minimal set of regulations are instituted and participants are able to introduce new norms and appropriate forms of observability and enforcement once the system is in operation. Because the reasons for introducing a change and the choice of the type of change that best applies, are manifold, the usual solution is an intermediate position. However, as discussed below in 9.3, finding out how to determine what those intermediate positions are, and how to implement them is almost unexplored territory.

Other design choices determine the *balance between regimentation and the different degrees of enforceability*. What are the matters that should be taken into account? For instance, considerations about accountability, robustness, and transparency may tilt the balance towards regimentation, while on the other hand addressing or allocating risk through constitutive conventions (e.g., requiring bonds and guarantees from principals, or relying on a conventional legal system to deal with severe transgressions), or the need for flexibility together with the presumption of reliable autonomous agents tilt the balance towards self-governance.

A third closely related matter is the *regulation attitude*. In this case, again, in most cases the final choice is bounded by two opposite styles. On one side is what we may call "preemptive enforcement" where—as in a typical Napoleonic legal tradition—presuming any agent would break the law if given a chance, the system designer anticipates all possible infractions, makes these and their ad-hoc corrective reactions explicit, and commits to a strict enforcement of this list. On the other side there is a "laissez faire" attitude where it s presumed that most individuals abide by the law most of the time. In this case—mimicking a Common Law tradition—undesired behaviour is expressed in general terms and only when someone is caught cheating and proven guilty after a due process, a harsh exemplary punishment is applied.

Validation of the design. With the remarks in Section 3 in mind, there is room for innovation with respect to testing and validation of the system from the formal as well as the engineering perspectives. For instance, the *off-line validation* of the sets of norms under different criteria and techniques, from model-checking to coherence theory. *Computational complexity* of norm-abidance and *scalability* of the system with respect to the increase of norms and agents. *Expressiveness* of metamodels with respect to the intended performance of the regulated MAS. *Reliability* of the operationalisation processes. Although we are referring here to validation at design-time, there are limitations to what may be tested and proven off-line, hence some *runtime validation* may be needed, should be considered as part of the design and then implemented. In these matters the key is reaching an appropriate balance between what may be proved and what is satisfactory from an engineering perspective.

9.1.2 Proclamation

This process involves two aspects: on the one hand, how participating agents need to be informed about the system so that they will be able to play accordingly, and on the other hand, what needs to be operational so that the system may be enacted. Mirroring standard legal practice we may distinguish two sources of design choices where interesting problems arise: publication of the conventions and their activation.

Publication. The designer will have to commit towards those elements that need to be working from the start of and will also need to decide for each one of these elements how to make them known, when, and to whom. A mere enumeration of the elements involved in publication is enough to show the richness of this topic: constitutive conventions (ontology, primitive and nonprimitive actions, interaction model, access requirements, entitlements), architecture of the system (governance model, dynamics), normative content (different types of norms and metanorms) and eventually, the operational semantics of the system. The "how" part has two dimensions: the degree of formality (logical or otherwise) of those components and the process by which the components are made part of the system. Another aspect where innovation is needed is in the ergonomic side of communicating those elements: what type of expression, syntax, interface are appropriate, when, and for whom.

Activation. Some conventions—including most regimented constraints—are established at design time to become active the moment the system is enacted and hence are applicable from the start to any agent that intends to participate. However, while the system is being enacted, norms may be added, modified or revoked and these situations start to apply to participants at some point. Generally speaking, such activation may be triggered either by time (e.g., so many days after it is published) or by an event (e.g., once a commitment is made). The challenges in this topic come from different sources: from the regulated MAS perspective, the immediate ones are how to validate that an agent that intends to participate complies with the conventions that regiment its admission, and then how the system deals with latent norms. From an individual agent's perspective, how is an agent made aware of those norms that may be applicable to it.

9.2 Enactment Phase

We do not touch upon the computational and implementation aspects of enactment, only on some topics that may be interesting from a regulatory perspective while the system is active. Furthermore, for sake of brevity, most of our comments refer to noncompliance with norms (and the corresponding negative sanctions), although they apply *mutatis mutandi* to the compliance of actions that have a positive reward associated.

9.2.1 Observability

Institutional State. The essential feature of regulated systems is that there are norms that individuals may or may not infringe. Hence, at some points, although individuals decide whether to comply with *applicable* norms, the system as a whole, or its enforcement devices—and usually some participant agents as well—may have to assess that a compliance or noncompliance took place. Such assessments involve the difficult technical (operational) problem of representing and updating the "institutional state" of the regulated MAS (see Section 7.3). In other words, what are the values of those variables that represent the institutional facts at a given moment, how atomic actions are filtered into the regulated MAS and how those actions that are deemed "institutional" modify the value of those variables.

Assuming that the institutional state is properly represented and implemented, the designer still has to choose how this state is accessed by participants while actions are taking place. In other words, the regulated MAS needs to address, the *ex-ante* aspects of "awareness"; and the two complementary *ex-post* aspects of "transparency" and "accountability". These three types of aspects are solved by regulating what part of the institutional state is revealed, to whom, when and how.

Awareness. The challenges in the *ex-ante* phase of compliance assessment reside in what is revealed before an action takes place.

From an individual agent's perspective, that revealed information is needed to support two decision-making tasks: On the one hand, the individual agent needs to be aware of that state of the system to realise what active norms apply to it in order to decide whether or not to perform an action that may infringe those norms. On the other hand, an individual needs to be aware of the state of the system in order to form expectations about what may happen then: who may act, what actions may be attempted and what effects these may have.

From the system's perspective, the challenge is again twofold: to determine what actions are feasible (nonregimented), and to determine which agents are subject to active (nonregimented) norms and therefore have the possibility of complying or not with it.

Transparency and accountability. The first type of *ex-post* aspects, "transparency" refers to what is revealed (to individuals and to the system) about actions that take place and their institutional effects. "Accountability", in turn, refers to information about who performs an action and who is affected by that action. The two of them together become input for determining awareness in subsequent institutional states. Both of them are, evidently, key for enforcement and should consequently be in line with the enforcement model of the system and the enforcement style we mentioned above. Both are particularly challenging when the regulated mulitagent system involves nested or concurrent regulated MAS where commitments established in one regulated MAS may have effects in other regulated MAS. Transparency and accountability also need to address the complementary aims of "need to know" and "need to share".

The opacity of some actions may be appropriate. For instance, in some mediated negotiations, only the mediator is informed of offers and counteroffers and each party is unaware of what the other part is actually proposing. In some cases, while opacity is required for some purpose, it needs to be compensated by some means. For instance, even if a noncompliant action is itself opaque to law enforcer agents, an infraction may be inferred by these agents if they perceive some effects of that action, or they are informed of the infraction by other agents that witnessed or inferred it on their own; likewise, even if a punishable action is perceived by an enforcer agent, the agent may decide—or be compelled—to ignore it. Opacity may also have adverse effects. For instance, if an infraction is not immediately dealt with, its indirect consequences may be difficult to foresee and contend with.

9.2.2 Governance

This area includes those actions that follow the assessment of compliance. Namely the processes of determining whether or not an infraction took place—analogously, a reward-deserving action—and then react accordingly. In the judiciary tradition, governance involves three main processes: prosecution, trial, and punishment. Although some of these issues have already received attention from different perspectives, the topics are rich and still largely open for regulated multi-agent systems. We will next describe the most salient ones in two areas, blame assignment (involving prosecution and trial to some extent) and reactions (including punishment and other). As with the previous paragraphs, our comments are biased towards noncompliance and sanctions but similar ones would apply to rewards and desired actions.

Blame assignment. Given that a punishable (or reward deserving) action has occurred, the challenges reside in determining that the action took place and should be punished (or rewarded), determining who is involved and who is responsible for the infraction (or reward). The assessment of infractions and culprits will depend on the observability of actions and ultimately on the enforcement model of the system and the enforcement style. For instance, if the enforcement of some norms is delegated to enforcer agents, when these agents observe or infer the infraction, they would need to be able to identify the beneficiaries along with the casualties associated with the infraction. Provided a good level of accountability is available, these enforcers should then be able to assign blame. Once the infraction is acknowledged and culprits identified, reparatory actions and sanctions are enacted according to the norms that regulate these processes. The process of determining who the culprits are is not necessarily straightforward because, depending on the observability or the infraction, the identity of perpetrators may not be revealed, and even if the agents who perform the invalid action are properly identified, it may still be necessary to prove who the actual guilty parts are. In this case some sort of *due process* needs to be activated. When a regulated multi-agent system contemplates the existence of a due process, several components need to be in place. Namely, some notion of proof that an infraction has occurred, the resources to elaborate and validate that proof and the procedures that correspond to (i) bringing charges and evidence against suspects, (ii) defending innocence against charges (iii) evaluating evidence and applicable norms and (iv) formulating a resolution about the innocence or guilt of the accused.

Regulated reactions. The types of reactions that are worth studying may be organised in two large blocks: punishment and reward on one hand and damage control.

Punishment and rewards. How the punishment is expressed (threat, argument ad baculum, and others), Grounds for punishment (what needs to be proved to deserve that punishment or reward). Purpose of punishment (to teach, to encourage, to retaliate), types of punishment (direct or indirect, private or public, with rhetorical information associated, with social or monetary cost). Management of punishment (whether it is applicable on the spot, delegated to another regulated MAS, and so on).

Damage control. Identify direct and indirect effects of an infraction. Measure costs of transgression, evaluate costs of detection, blame assignment, and punishment. Identify and implement reparatory actions (fix damages, compensate victims). Here belongs the challenging world of *contrary-to-duty obligations* that has received ample formal treatment and whose implementation is nontrivial (see [13, 9, 34]).

9.3 Evolution Phase

Several remarks on this phase were made above in Section 6 and more will be made in the last section of this chapter. Here we simply mention a few more open topics. and for the sake of presentation we mention topics about only two aspects: performance evaluation and change.

Note however, that from the design perspective (even when evolution is postulated as a bottom-up process) it is advantageous to identify the several conditions that might make the regulated system evolve and, consequently, include devices to handle that evolution. One may argue that the type and protagonism of those devices depend largely on the expected evolution of the system, and different types of devices will need proper grounds to determine their application. In general, one would need to identify (in the metamodel and the methodology) (i) a reasonable list of devices, (ii) for each device, the type of situations that justify its use, (iii) the elements of the system that are involved in those situations and (iv) the interplay of those elements in a given situation.

As an example to illustrate the richness of the problem, note that one of the many devices to make a regulated system evolve is to change some of its norms; moreover, one of the ways that norms may be changed is directly by the system (not by participants), and those changes might be advisable, for instance, because performance of the system has decreased due to, for example, changes in the population profile or the environmental conditions. Now realise that, even in this rather simple case where we assume that *one* evolution device of the system is to change norms, in order to decide what norms to change we would then need to foresee, at design time, that performance of the system can be measured, that changes in population and environmental conditions may be assessed, that the relationship of those changes with performance are made explicit, that the set of norms that may be modified (added, deleted, changed) in order to achieve the desired performance levels can be determined and, finally, that the actual modifications may be accomplished.

9.3.1 Evaluation

Performance indicators. It is not unusual to have sociotechnical systems involving stakeholders with competitive goals and thus having regulations intended to achieve equilibria of different sorts. However it is not always clear how these equilibria might be identified without an explicit reference to some variables involved in the operation of the system and their combination as indicators that are meaningful in terms of the objectives of the stakeholders or the system as a whole. In general, variables and indicators are useful to assess the quality of the system (or parts of it) and to choose between alternative implementations of regulations at design time and to guide evolution at runtime. It is a challenging task to deal with performance indicators that evaluate how a system performs or improves in important but elusive qualities such as fairness, trustworthiness, and accountability.

Learning from experience. As suggested above, the choice and combination of performance indicators depends not only on the objectives but also on the devices that are available to make the system evolve and on the elements that are involved in the application of such devices: from changing some parameters of a norm, to changing the set of norms. Part of the tuning—not only of the original normative system but notably also of the evolution mechanisms themselves—may be achieved at design-time by stress-testing and running simulations; and there is ample opportunity for development of the current practices and tools.

9.3.2 Change

Operations on norms. A full typology of operations on the set of norms (promulgation, amendment, suspension, abrogation, annulment, ...), as well as the operations on their governance aspects and their implementation is still to be attempted from the regulated MAS perspective. Some works in this community suggest [5, 74, 36] that a systematic treatment of the different operations is far from trivial.

Metanorms. Another line of work that is largely open for research is to include all those aspects that determine the dynamics of the system as a distinct core component of a regulated MAS as postulated in Section 6 (p. 105). Several approaches may be taken and a few have already deserved attention like, for example, the use of metanorms to choose among predefined sets of norms [1, 7], or the use of case-based reasoning to introduce new norms when conflict among norms are detected [52] but a holistic proposal towards a general normative transition function, as suggested before in 10.4.1, is still to appear.

10 Authors' Perspectives

This section is meant to complement the rest of the chapter by including personal views of the authors. Each author has chosen topics, length and structure.

10.1 Noriega: In Light of Applications

I would like to make the following remarks under the light of regulated multi-agent systems that are going to be used in open sociotechnical systems that are intended to work in the real world.

In this context, what I believe to be the most fundamental task is to build a general framework on top of which actual normative multi-agent systems may be specified and implemented. By a framework I mean three main components: a normative architecture or "metamodel" for the specification of normative multi-agent systems, the computational counterpart of this metamodel that would allow to implement and run such specifications, and the methodologies to guide the actual implementation of such norMAS.

Following the experiences reported in [19], I believe that such a metamodel can be built along the lines described above in sections 3, 4 and 5, in order to support the five components enumerated in Section 2. The computational counterpart should produce an "institutional environment" where particular regulated MAS are enacted, all regimented constraints are enforced and several regulated MAS may concurrently exist. That implementation may opt for a centralised governing of the full environment (including the possibly several regulated systems embedded in the shared institutional environment) or choose to have a distributed architecture to handle particular norMAS [59]. The methodologies should facilitate four activities: the proper specification, implementation and testing of particular norMAS—including whatever agents are needed to perform those regulatory roles entailed by the specification—and, in the fourth place, assess the adequacy of the implementation of the norMAS with respect to the intended functionality in the world.

If taken to heart, this task involves several challenges that being already in this community's agenda, as suggested in the previous section, are far from being won. The ones I see as the most significant are the following:

10.1.1 Achieving Good-Enough Expressiveness and Automation of the Normative Languages

Let's assume that the framework includes a rich enough meta-modelling ontology to capture several interaction models, normative corpora and governance models. Then, for the specification of particular regulated MAS, it will be necessary to use several normative languages (and many "types of norms", e.g., Crawford and Ostrom [17]) and possibly non standard notions of "consequence", that should be compatible with the features discussed in 7.3 and 7.4. But, in addition, those norms will need to be accompanied with other linguistic and formal construct so that the *declarative* and *inferential* features of such norms carry the intended pragmatic load within the normative MAS. So, for instance, in order to deal with *procedural norms*, it may be advantageous to use, say, commitment-based protocols, hence one has to introduce a proper language [27, 82] as part of the regimented constraints. Likewise, to implement *contrary to duty functionalities*, one needs declarative languages that accommodate the subtleties of features like the set of linked norms, including sanctions and reparations, as well as some automated means to support the complex inferential processes involved in blame assignment and proper reparation.

The implementation of these features should take into account, not only what the regulated systems themselves should be able to accomplish at run time, but also how agents (software as well as humans) become informed of all those features; in order to take them into account both at design and at running time.

10.1.2 Designing for Noncompliance

The framework, as I see it, serves a general-purpose boot-strapping function, in the sense that it should support particular normative MAS and provide the environment where actual normative MAS are embedded. Therefore, the metamodel should support the modelling of different governance models and provide the *expressive features* to specify those different governance models, and the framework should support their *automation*. Moreover, the framework should support the interoperation of several of these regulated MAS in a common environment.

In these conditions, regimented governance is unavoidable for certain aspects of the environment (e.g., common ontology, atomic interactions, social semantics, how new regulated MAS are embedded in the environment) and maybe also within the particular regulated MAS. However, in addition, in most regulated MAS, one would need to have nonregimented conventions that deal with the forms of enforcement described in page 104 and in Section 4.

In practical terms it is unlikely that there may be a general treatment of governance modelling and implementation. Nevertheless, it is still hard but worth attempting to develop some sort of "standard" devices to deal with the several aspects of governance (detecting, ascertaining, evaluating, assigning blame, applying sanctions) and assemble such devices as the enforcement model for a particular regulated MAS. Assuming regimentation is properly handled in the framework (like suggested in [19]), there are two ways to approach the enforcement of nonregimented regulation. The first one is to rely on an omniscient filtering device—like the one needed for regimentation—that automatically produces all consequences of a new fact, thus updating the institutional state [74, 40, 31]. The second one is to rely on some form of law enforcement roles that are capable of perceiving some actions and react with or without discretion in accordance with the relevant regulations; this applies also to informal sanctions, applied by individuals who may not be playing proper law-enforcement roles at all.

Another path worth exploring is that given that transgressions will take place, the governance model may in some instances grant the offender the benefit of doubt and allow the wrongdoer to argue the case. This would serve two purposes, allow for some useful forms of ambiguity and to improve regulations.

10.1.3 Designing with Ambiguity

Conventional legal institutions incorporate ambiguity in ordinary legislative practice for two main reasons: to prevent overregulation and to allow practitioners and enforcers some leeway for interpretation. These reasons also suggest the need and advantages of including ambiguity as a design assumption for regulated MAS.

In general terms regulated MAS come across “opaque” contexts where some undesirable situations are difficult to properly identify at design time. Some obvious situations are: (i) Regulating the wrong problem. For example in the *mWater* system described in the next chapter, it is easy to focus only on pricing conventions when the real problems reside in the lack of supply and in the conflicts provoked by the use of traded rights. (ii) Regulating the wrong population; for instance, unwittingly ruling out human agents, by setting the bidding-clock pace in an auction too fast. (iii) Changing population: In the open innovation scenario of Section 8, it is rather likely that the client base becomes more international and more sophisticated as the level of activity increases. But then standard contracts and quality assurance criteria would have to be tuned to the new situation [7]. (iv) Unforeseen undesirable outcomes; In automated trading, software trading agents may have equivalent bidding algorithms and enter into unending ties. (v) Volatile contexts where the grounds for choosing particular actions may change depending on circumstances that, because of their unpredictability or variety, are not worth hardwiring in stable norms, but rather be localised to those contexts.

Mirroring the points just made, the following are obvious heuristics to start bringing ambiguity into design: identifying volatile contexts in key business processes; building malleability in constitutive norms (for instance, committing to a good-enough ontology while including norms to update it); separating stable norms (that may be hardwired as regimented or fully specified enforced regulations) from norms that are better represented in the system as norms that govern decision-making of law enforcers; using flexible-enough governance models and choosing performance indicators that measure different types of “fitness” of the regulations.

At any rate, whatever heuristics are devised, the components where ambiguity will need to be instrumented are the norms that govern the evolution of the system. And there is where ultimate design with ambiguity will be achieved.

10.2 Chopra: Social Computing, A Software Engineering Perspective

The nature of applications is changing. Earlier they were logically-centralised; now they are becoming increasingly interaction-oriented. Social networks, social cloud, healthcare information systems, virtual organisations, and so on are evidence of the shift. In such applications, *autonomous* social actors (could be individuals or organisations) interact in order to exchange services and information. I refer to applications involving multiple autonomous actors as *social applications.*

Unfortunately, software engineering hasn’t kept up with social applications. It remains rooted in a logically centralised perspective of systems dating back to its earliest days and continues to emphasise low-level control and data flow abstractions. In requirements engineering, for instance, the idea that specifications are of *machines*, that is, controllers, is firmly entrenched. Software architecture applies at the level of the internal decomposition of a machine into message-passing components. In other words, it helps us *realise* a machine as a physically distributed system. However, the machine-oriented worldview cannot account for social applications in a natural manner.

I understand *social computing* as the joint computation by multiple autonomous actors. By "joint", I refer simply to their interactions and the *social relationships* that come about from the interaction, not necessarily cooperation or any other form of logical centralisation. In fact, each actor will maintain its own local view of the social relationships—there is no centralised computer or knowledge base. The relationships themselves may take the form of commitments, trust, or some other suitable social norm. The purpose of the computation may be to loan a bicycle or a couch to a peer, to schedule a meeting or a party, to carry out a multiparty business transaction, to provide healthcare services, to schedule traffic in smart cities, to manage the distribution of electricity in smart grids, to build consensus on an issue via argumentation, or globally distributed software development itself—*anything* that would involve interaction among autonomous actors.

Clearly, we are already building social applications, even with current software engineering approaches. For example, online banking is a social application in which a customer interacts with one or more banks to carry out payments, deposits, and transfers. Social networks such as Facebook and LinkedIn facilitate interactions among their users. However, just because we can build social applications, it does not mean we are building them the right way. Right now, all these applications are built in a heavily centralised manner: banks provide all the computational infrastructure; so does Facebook. Users of these infrastructures are just that—*users*, no different from those of an elevator or an operating system. In other words, current software engineering produces only low-level technical solutions.

My vision of social computing instead embraces the social. It recognises the autonomy of actors. Instead of control flow or message flow, it talks about the meanings of messages in terms of social relationships. Computation refers to the progression of social relationships as actors exchange messages, not to any actor's internal computations (although these too could be accounted for). The different aspects of my vision constitute a challenging research program. What form would specifications of social applications take? What would be the principles, abstractions, and methodologies for specifying social applications? On what basis would we say that an actor is behaving correctly in a social application? How would we help an actor reason about specifications of social applications with respect to its own goals and internal information systems? What kind of infrastructure would we need to run social applications? The answers to these questions and the realisation of my vision will lead to a software engineering vastly more suited to social applications.

10.3 Fornara: Formalising and Executing Open Normative Systems

The process of formalising at design time and executing at run time open interaction systems where the behaviour of agents is regulated by norms, presents interesting open issues.

A first one regards how to choose and use existing formal and programming languages and architectural solutions for defining and implementing in efficient and effective way open interaction systems. Those type of systems are composed by many concepts that evolve in time on the basis of the events that happen and the actions performed by the agents, therefore one problem is efficiently represents the evolution in time of their state. A crucial requirement in the specification of norms is that their content should be a description of the action that should, should not, or can be performed by the agents: that is a description of a *template* of the action. In order to let the autonomous agents to exploit their capability to dynamically plan their action by taking into account the norms that constrain their behaviour, it would be better that those template should not describe the actions in all details. This by making it possible that different concrete actions will match with the described template. For example a norm may express the obligation to pay a certain amount of money to Robert without specifying the actor of the action and making it possible to perform the payment in many different ways, like for example by bank transfer or in cash or by cheque.

In those type of systems it is necessary to implement specific components for realising the following required functionalities. First of all in order to *enforce* those norms it is necessary to be able to efficiently *compare and match* the real actions that are performed by the agents at runtime with their description formalised in the template of the action that is used as content of the norm. Practically it is also necessary to define a component able to *monitor* and to *regulate* (in the case of power) the evolution of the state of the system.

At least two more languages turn out to be necessary, one for the formalisation of the concepts the other for the implementation of the functionalities. In this case fundamental problems are how to combine them, and how to decide what to represent in a given language and what in another.

Reusability is a strong requirement in the development of software systems, therefore the definition of design mechanisms that make it possible to *re-use* at least part of a given specification of a system is another important challenge. The process of designing and developing open systems has to take into account also the design and development of the *agents* able to interact through those systems. The openness of those systems and the fact that they are running in an open network like Internet, implies that one agent may decide to interact simultaneously with different systems and that one system can be modularised in different components and distributed on different platforms, therefore a critical issue is the possibility to *combine* different specifications with few or none added re-design or re-programming work.

An important issue in open systems is related to how agents can perceive the shared state of the system and in some cases, when declarative communicative acts have to be performed (for example declaring an auction open) to use them for interacting with other agents. Regarding this aspect there are interesting existing approaches and existing frameworks coming from studies on distributed event-based systems and *environments* [81]. Agent environments can be considered an interesting architectural component that can be used for *mediating* agents action and for enabling agents to perceive the state of the interaction. The functionalities of the environment can also be extended to realise the *monitoring* of agent actions and the realisation of concrete mechanisms for *norm enforcement.* In this context, an interesting challenge would be the study of how to integrate and extend existing formal models and frameworks for the realisation of agent environments [6, 58] with the concepts and functionalities required by open normative systems.

One possible approach for trying to tackle some of the previously described challenges consists in formalising norms, and the corresponding obligations, prohibitions, permissions, and institutional powers, using *Semantic Web* technologies. In particular by using OWL 2 DL, a *description logic* language recommended by W3C for Semantic Web applications, for the specification of the main concepts of the model.

The main advantage of the choice of these languages is first the possibility to re-use ontologies and combine them thanks to the fact that two or more ontologies can be simply merged by taking the union of their axioms [39]. Another advantage is the possibility of using the semantics of the concepts used to describe the template of the actions contained in the content of the norms to effectively match them with real actions that happen in the system; for example being able to deduce that a bank transfer or a cash payment are both payment. From the point of view of the technologies and tools available for supporting the use of Semantic Web Technologies, important advantages are: (i) the availability of well studied and optimised reasoners (like Fact++, Pellet, Racer Pro, HermiT) for deducing knowledge or for checking the consistency of certain set of norms; (ii) the possibility to use tools for ontology editing (like Protégé) and library for automatic ontology management (like OWL-API or JENA).

Given that OWL 2 DL is mainly a language for expressing knowledge bases it may not be expressive enough for implementing running dynamic interaction systems, therefore a further crucial challenge is to study how to effectively and efficiently combine OWL 2 DL ontologies with *rule languages*—like Datalog, SWRL (Semantic Web Rule Language), or RIF (Rule Interchange Format) that is a W3C Recommendation from 22 June 2010—and with programming languages—like Java—for representing the dynamic evolution of the state of the system and for monitoring and enforcing the norms. Some preliminary studies on how to use Semantic Web Languages for realising open systems are presented in [28, 26].

The approach of defining different artificial institutions and then to combine them to realise different concrete open system is one of the possible approach that tries to tackle the problem of re-usability of artificial institutions. An open and crucial challenge is to study the process of transforming a conceptual model of an artificial institutions in a concrete running software where the interaction among autonomous agents can actually take place.

Concerning the challenge of situating open systems in agent environments a possible approach could be the introduction of a formal description of spaces of interaction and objects as first-class entities that shape the environment and describe all its structural components, including norms. A crucial challenge in this approach is managing AI and spaces interdependencies in terms of events and norms that regulate those events. Initial studies in this direction are presented in [73, 29].

10.4 Lopes Cardoso: Achieving Open Normative Environments

Most approaches to modelling normative multi-agent systems are based on statically defined normative scenarios. Even when assuming an open stance in terms of the nature of interacting agents (which are seen as heterogeneous self-interested entities that cannot be assumed to be benevolent), such approaches do not accommodate a normative-level autonomy [79] of interacting agents. As such, when entering the system agents adhere to the normative scenario that has been predefined. This is the approach we see in organization-oriented models (such as [18]) or in dialogical electronic institutions [54].

Looking at *open* normative environments from an extended perspective, norm-autonomous agents should be able to choose the norms they wish to govern their relationships. Therefore, it is not only the origin or benevolence of agents that is not certain (which is a must in open multi-agent systems), but more importantly the norms that are to be handled by the environment cannot be totally predicted in advance. It is thus important to think of the normative environment as providing infrastructures for fulfilling two distinct and complementary roles:

- *Normative setup.* How can agents be assisted in their effort to specify the norms that better fit their intended interaction scenario?
- *Norm monitoring and enforcement.* What kinds of mechanisms can the environment employ with the aim of monitoring and enforcing norm-regulated relationships?

While the latter has been widely addressed by several researchers working both on monitoring (e.g., [51, 48]) and enforcement (e.g., [57, 24, 37, 49]), the issue of runtime normative setup has been mostly neglected—as mentioned before, designed normative environments are typically closed as far as the normative space is concerned. Nevertheless, this issue is important in application domains configuring regulated coalition settings, such as agent-based electronic contracting or Virtual Enterprise formation.

10.4.1 Environment Design and Normative Evolution

The study of the *environment* as a first-class entity in multi-agent systems (MAS) is quite recent [81]. A role that the environment may fulfil is that of normative state maintenance, by employing appropriate norm monitoring policies. This has been identified as a viable alternative to deploying "institutional agents" responsible for this task: the governing responsibility is transferred to the environment itself [61]. Typically, a rule-based infrastructure that defines reactions to events is the approach taken to realise this kind of environment, allowing not only to regulate agent interaction but any action that is taken within the environment.

Along this line, *coordination artefacts* [56] have been proposed as abstractions encapsulating and providing a coordination service that agents can exploit in a social context. Coordination artifacts aim at enabling the creation/composition of social activities and at ruling those activities, from a normative perspective. Several proposals have been made to conceive artefacts of several types. This includes the use of organizational mechanisms to address the normative aspects of a MAS [40], mostly from a normative state monitoring perspective.

The structuring of a normative environment into different social contexts has been targeted by a number of researchers. *Institutional spaces* [72] are an approach to segment the environment into different institutions; these benefit from a common environment infrastructure in terms of perception and evolution models. Institutions can also be empowered to govern other institutions [16]. A slightly different perspective is to consider an institutional environment as being composed of several *normative contexts* [47] governing different but sometimes interdependent social relationships. A hierarchical organisation of contexts (or institutions, for that matter) allows designing the environment with norm inheritance models in place.

A few researchers have tackled the problem of changing the norms of the environment at runtime. Theoretical approaches assign to *constitutive norms* the role of defining the possible changes to the normative system [3]. More practically-oriented efforts consider defining appropriate constructs that allow the system designer to define at compile time a *normative transition function* that specifies the possible norm changes and the conditions for their realisation [7, 74]. Letting the agents choose at runtime the changes to introduce in a normative system (defined in terms of a dynamic protocol) specified at design-time has also been suggested in [1]. Again, the possible options are application-specific: a set of possible values for specific fluents are the *degrees of freedom* agents have when proposing protocol modifications.

10.4.2 Challenges and Perspectives

The notion of an *open normative environment* tries to look at the environment from outside any preexisting organisation, putting the emphasis instead on infrastructural components that allow *normative interactions* to take place. By normative interactions we mean not only interactions governed by norms, but also norms that are established by a deliberative process based on interaction.

Given the current approaches to address normative environments (whether disguised as artificial/electronic institutions or referred to as environments in their own right), the openness of multi-agent systems is not fully addressed. We argue in favor of the need to develop appropriate infrastructures for enabling runtime creation of normative specifications. Although, as mentioned above, there are some approaches in the literature that deal in some limited way with dynamic adaptation of norms, this issue has not yet been addressed from a domain-independent perspective.

Having software agents that are able to deliberatively establish on their own some normative specification of a norm-regulated relationship is a challenging task. Although there are several relevant research contributions regarding normative reasoning (mostly from a norm compliance perspective), the specification of appropriate infrastructural components facilitating normative setup is lacking.

Approaches such as coordination artifacts have been exploited as mediated interaction mechanisms. A similar mechanism may be explored for assisting the setup of normative relationships. The use of normative frameworks is another promising approach to ease this task. Norm inheritance or adaptation mechanisms can be explored. To this end, insights from legal theory are certainly pertinent sources of inspiration to develop environments that include such facilities.

10.5 Singh: A Normative Basis for Trust

Open settings, where norms apply, inevitable bring out the problem of decision-making: How can each party decide on how it should engage the others? Trust is a key ingredient in such decision making, leading us to another question: How can each party determine how much trust to place in another autonomous party? To be an effective basis for decision making, the estimation of trust must incorporate (1) the interaction—task or transaction—being considered by the decision maker, (2) the social or organisational relationships, and (3) the relevant context.

The following are the main approaches to trust today.

Today's distributed computing approaches hard-code some patently naïve assumptions, e.g., about the effectiveness of certificate chains [8]. Thus, they provide little or no rational basis for decision making in realistic settings.

Today's analytics-based approaches seek to estimate the trustworthiness of a party based on an analysis of its attributes, behaviour, and relationships [30, 44, 80]. However, existing approaches largely hide the deep structure that one would informally associate with trust in natural human and organisational settings. In particular, these approaches provide simple calculi that associate numeric measures or qualitative descriptions of trust that seek to summarise the trustworthiness of a party as a single real number or nominal value.

Today's cognitive approaches build ontologies and knowledge models of contexts and situations but in a way that gets into the particulars of a domain of application. Further, they provide complex definitions seeking to formalise how humans subjectively understand trust, but such definitions call upon concepts for which we cannot easily obtain any data to use as a basis for prediction or analysis [11, 10]. Thus, although the cognitive themes and domain models are useful, such approaches are not easy to apply computationally.

I propose a research direction that seeks to address the above gap in the modelling of trust in a manner that exploits the inherently normative nature of multi-agent systems to address how we can analyse and engender trust. If our collective research efforts succeed, our main contribution will be a principled approach to trust that provides rich abstractions for modelling tasks and social or organisational relationships that apply in distributed systems and can be grounded in analytics-based decision making.

10.5.1 Norms for Trust

The foremost idea underlying trust is that the extent to which one party, Alice, may justifiably trust another party, Bob, depends on how Alice and Bob interact, including what Alice observes regarding Bob. I begin from two key points about trust.

Autonomy. Trust carries a connotation of choice: trust is meaningful only when the trusting party or *truster* can choose to proceed or not with the given interaction (with its counterparty, the *trustee*). And, the trustee can choose whether to carry out the interaction in question with the requisite quality. The latter element is diminished in cases of instrumental trust, wherein the truster treats the trustee as an instrument—as part of the infrastructure—and thus lacking in autonomy. In such cases, the question of trust reduces to the truster's trust in the reliability of the trustee.

Exposure. Trust carries a connotation of vulnerability: it is meaningful only when the trusting party has some skin in the game. If we were to somehow guarantee full protection to the trusting party, it could decide independently of its trust in any counterparty.

Given the foregoing points, I preserve the basic intuition that Alice's trust in Bob is strengthened when Bob meets or exceeds her expectations and weakened otherwise. Alice could observe Bob from a distance, as he interacts with others, but her learning of him would be the most relevant in circumstances where Alice personally faces a certain vulnerability to Bob's potential malice (or, in the case of an instrument, incompetence)—that is, when Alice has placed trust in Bob.

In departure from traditional research on trust, I express the relevant expectations formally in terms of *normative relationships* between Alice, Bob, and the other parties in the given system. From efforts in modelling large-scale systems from the standpoints of contractual relationships and decentralised administration, I have identified a small set of norm types [69]. Specifically, is Bob authorised to do something, committed to doing it, or prohibited from doing it? To what extent are Alice and Bob encounters regimented (so as to prevent violation) by their environment? To what extent is Alice directly protected? To what extent is Alice indirectly protected: would Bob be sanctioned (punished) for a violation? The lower the regimentation or protection the greater is Alice's vulnerability. Based on these, we can analyse Bob's actions from Alice's viewpoint—and Alice's actions from Bob's. Thus for each party we can provide a basis for determining how much trust to place in the other.

The norms I propose can all be expressed in conditional logic, e.g., [67]. Their logical basis offers a clear way to assign semantics to trust based on sets of possible computation paths [68]. We can use the semantics as a standard of correctness for any formal trust calculus, including as a basis for analytics. For example, Alice ought to trust Bob to no greater an extent for keeping his commitment to do P and Q than for keeping his commitment to do P alone. Such semantics can provide a basis for a rich variety of trust calculi that may be specialised for particular kinds of distributed software systems. Thus, if Alice determines (for example, via data mining) that Bob is trustworthy (to a specific extent) for P and Q, she should infer that Bob is at least as trustworthy for P. A potential practical benefit of the proposed approach is dealing with heterogeneous observations from which trust needs to be determined in the field. Such observations often do not respect simple patterns (such as being clear-cut positive or negative about the same P) from which one can compute a probability. I conjecture that a normative approach can provide a basis for extracting the most information from the observations that arise in practice.

Initially, we might pursue a Bayesian approach, which relates well both to analytics and to a semantics based on computation paths. In subsequent studies, it would be worthwhile to consider richer representations such as those based on utility theory.

Any approach to trust can be difficult to evaluate, and especially so if we bring in sophisticated concepts such as norms. Existing public datasets, e.g., for social networks, are limited and do not specify interactions. Indeed, the limited nature of available datasets is one of the important reasons for the popularity of the simplistic analytics-based approaches to trust. The following are two potential evaluation approaches.

- Obtain a text-rich dataset of user comments on each other and carry out text mining to infer assessments of the implicit norms involved and the felicity of the interactions involved.
- Develop a new dataset based on one or more games that we can have users play against each other. Such a dataset would likely be small but would point the way toward richer modelling.

10.5.2 A Call to Arms

To summarise, we see today a significant gap between trust theory and practice. Analytical and distributed systems approaches to trust involve shallow models not representative of real applications; even when these approaches talk of social networks, they do so in a manner that disregards any meaningful characterisation of the underlying social relationships. The cognitive models are richer but incomplete, yet difficult to characterise empirically from practically obtainable data.

The proposed norm-based research program will contribute a principled approach to trust that provides includes semantically rich abstractions for tasks and social or organisational relationships, yet which can be grounded in analytics-based decision making. This program is ambitious: it doesn't seek incremental improvements to current approaches, but to introduce a sea change in how trust is approached in research and practice, beginning from a foundation in norms.

Acknowledgments

Munindar Singh's effort was partially supported by the U.S. Army Research Office under grant W911NF-08-1-0105. The content of this paper does not necessarily reflect the position or policy of the U.S. Government; no official endorsement should be inferred or implied.

Nicoletta Fornara's effort is supported by the Hasler Foundation project nr. 11115-KG and by the SER project nr. C08.0114 within the COST Action IC0801 Agreement Technologies.

Henrique Lopes Cardoso's effort is supported by Fundação para a Ciência e a Tecnologia (FCT), under project PTDC/EIA-EIA/104420/2008.

Pablo Noriega's effort has been partially supported by the Spanish Ministry of Science and Technology through the Agreement Technologies CONSOLIDER project under contract CSD2007-0022, and the Generalitat of Catalunya grant 2009-SGR-1434.

References

1 Alexander Artikis. Dynamic protocols for open agent systems. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 1*, AA-

MAS, pages 97–104, Richland, SC, 2009. International Foundation for Autonomous Agents and Multiagent Systems.

2 Gordon Baxter and Ian Sommerville. Socio-technical systems: From design methods to systems engineering. *Interacting with Computers*, 23(1):4–17, 2011.

3 G. Boella and L. van der Torre. Regulative and Constitutive Norms in Normative Multiagent Systems. In D. Dubois, C.A. Welty, and M.-A. Williams, editors, *Ninth International Conference on Principles of Knowledge Representation and Reasoning*, pages 255–266. AAAI Press, Whistler, Canada, 2004.

4 Paolo Bresciani, Anna Perini, Paolo Giorgini, Fausto Giunchiglia, and John Mylopoulos. Tropos: An agent-oriented software development methodology. *Autonomous Agents and Multi-Agent Systems*, 8(3):203–236, 2004.

5 Jan Broersen. Issues in designing logical models for norm change. In George Vouros, Alexander Artikis, Kostas Stathis, and Jeremy Pitt, editors, *Organized Adaption in Multi-Agent Systems*, volume 5368 of *Lecture Notes in Computer Science*, pages 1–17. Springer Berlin Heidelberg, 2009.

6 Stefano Bromuri and Kostas Stathis. Distributed agent environments in the ambient event calculus. In *Proceedings of the Third ACM International Conference on Distributed Event-Based Systems*, DEBS '09, pages 1–12, New York, NY, USA, 2009. ACM.

7 J. Campos, M. López-Sánchez, J. A. Rodríguez-Aguilar, and M. Esteva. Formalising situatedness and adaptation in electronic institutions. In J. Hubner, E. Matson, O. Boissier, and V. Dignum, editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems IV*, LNAI 5428, pages 126–139. Springer, 2009.

8 Marco Carbone, Mogens Nielsen, and Vladimiro Sassone. Formal model for trust in dynamic networks. In *Proceedings of the 1st International Conference on Software Engineering and Formal Methods (SEFM)*, pages 54–63. IEEE Computer Society, 2003.

9 J. Carmo and A. Jones. Deontic logic and contrary-to-duties. *Handbook of philosophical logic*, 8:265–343, 2002.

10 Cristiano Castelfranchi and Rino Falcone. *Trust Theory: A Socio-Cognitive and Computational Model.* Agent Technology. John Wiley & Sons, Chichester, UK, 2010.

11 Cristiano Castelfranchi, Rino Falcone, and Francesca Marzo. Being trusted in a social network: Trust as relational capital. In *Trust Management: Proceedings of the iTrust Workshop*, volume 3986 of *LNCS*, pages 19–32, Berlin, 2006. Springer.

12 Henry W. Chesbrough. *Open Innovation: The New Imperative for Creating and Profiting from Technology.* Harvard Business School Press, Boston, MA, 2003.

13 R.M. Chisholm. Contrary-to-duty imperatives and deontic logic. *Analysis*, pages 33–36, 1963.

14 Amit K. Chopra and Paolo Giorgini. Requirements engineering for social applications. In *Proceedings of the 5th International i* Workshop*, volume 766 of *CEUR Workshop Proceedings.* CEUR-WS.org, 2011. 138–143.

15 Amit K. Chopra and Munindar P. Singh. Multiagent commitment alignment. In *Proceedings of the Eighth International Conference on Autonomous Agents and MultiAgent Systems*, pages 937–944, 2009.

16 O. Cliffe, M. De Vos, and J. Padget. Specifying and reasoning about multiple institutions. In P. Noriega, J. Vázquez-Salceda, G. Boella, O. Boissier, V. Dignum, N. Fornara, and E. Matson, editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems II*, LNAI 4386, pages 67–85. Springer, 2007.

17 S.E.S. Crawford and E. Ostrom. A grammar of institutions. *American Political Science Review*, pages 582–600, 1995.

18 V. Dignum, J. Vazquez-Salceda, and F. Dignum. Omni: Introducing social structure, norms and ontologies into agent organizations. In *Programming Multi-Agent Systems, Second International Workshop ProMAS 2004*, volume 3346 of *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, pages 181–198, New York, NY, United States, 2005. Springer Verlag, Heidelberg, D-69121, Germany.

19 Mark d'Inverno, Michael Luck, Pablo Noriega, Juan A. Rodriguez-Aguilar, and Carles Sierra. Communicating open systems. *Artificial Intelligence*, 186(0):38 – 94, 2012.

20 Carl M. Ellison. Establishing identity without certificate authorities. In *Proceedings of the 6th USENIX Security Symposium*, pages 67–76, 1996.

21 M. Esteva, B. Rosell, J. A. Rodríguez-Aguilar, and J. L. Arcos. Ameli: An agent-based middleware for electronic institutions. In *Third International Joint Conference on Autonomous Agents and Multiagent Systems*, volume 1, pages 236–243, New York, USA, 2004. IEEE Computer Society.

22 Marc Esteva, Juan A. Rodriguez-Aguilar, Josep Lluis Arcos, and Carles Sierra. Socially-aware lightweight coordination infrastructures. In *AAMAS'11 12th International Workshop on Agent-Oriented Software Engineering*, pages 117–128, 2011.

23 Fernando Flores, Michael Graves, Brad Hartfield, and Terry Winograd. Computer systems and the design of organizational interaction. *ACM Transactions on Information Systems*, 6:153–172, 1988.

24 N. Fornara and M. Colombetti. Specifying and enforcing norms in artificial institutions. In G. Boella, L. van der Torre, and H. Verhagen, editors, *Normative Multi-agent Systems*, volume 07122 of *Dagstuhl Seminar Proceedings*. Schloss Dagstuhl, 2007.

25 N. Fornara and M. Colombetti. Specifying artificial institutions in the event calculus. In V. Dignum, editor, *Handbook of Research on Multi-Agent Systems: Semantics and Dynamics of Organizational Models*, Information science reference, chapter XIV, pages 335–366. IGI Global, 2009.

26 Nicoletta Fornara. Specifying and monitoring obligations in open multiagent systems using semantic web technology. In A. Elçi, M. Tadiou Kone, and M. A. Orgun, editors, *Semantic Agent Systems: Foundations and Applications*, volume 344 of *Studies in Computational Intelligence*, chapter 2, pages 25–46. Springer-Verlag, 2011.

27 Nicoletta Fornara and Marco Colombetti. Operational specification of a commitment-based agent communication language. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2*, AAMAS '02, pages 536–542, New York, NY, USA, 2002. ACM.

28 Nicoletta Fornara and Marco Colombetti. Representation and monitoring of commitments and norms using owl. *AI Commun.*, 23(4):341–356, 2010.

29 Nicoletta Fornara and Charalampos Tampitsikas. Using OWL Artificial Institutions for dynamically creating Open Spaces of Interaction. In *Proceedings of the AT 2012 First International Conference on Agreement Technologies, October 15 - 16, 2012 in Dubrovnik, Croatia*, volume 918 of *CEUR Workshop Proceedings*, pages 281–295, 2012.

30 Karen Fullam and K. Suzanne Barber. Dynamically learning sources of trust information: Experience vs. reputation. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, pages 1062–1069, Honolulu, May 2007. IFAAMAS.

31 Andres Garca-Camino, Pablo Noriega, and Juan A. Rodrguez-Aguilar. Implementing norms in electronic institutions. In Simon Thompson Michal Pechoucek, Donald Steiner, editor, *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '05)*, pages 667–673, Utrecht, NL, 2005. ACM Press.

32 Andres Garcia-Camino, Juan A. Rodriguez-Aguilar, and Wamberto Vasconcelos. A distributed architecture for norm management in multi-agent systems. In Jaime Sichman,

Julian Padget, Sascha Ossowski, and Pablo Noriega, editors, *Coordination, Organization, Institutions and Norms in agent systems III*, volume 4870 of *Lecture Notes in Computer Science*, pages 275–286. Springer Berlin / Heidelberg, 2008.

33 Joseph Goguen. Requirements engineering as the reconciliation of technical and social issues. In M. Jirotka and J. Goguen, editors, *Requirements Engineering: Social and Technical Issues*, pages 165–200. Academic Press, 1994.

34 G. Governatori and A. Rotolo. Logic of violations: A gentzen system for reasoning with contrary-to-duty obligations. *Australasian Journal of Logic*, 4:193–215, 2006.

35 Guido Governatori, Zoran Milosevic, and Shazia Sadiq. Compliance checking between business processes and business contracts. In *Proceedings of the Tenth International Distributed Object Computing Conference*, pages 221–232, 2006.

36 Guido Governatori and Antonino Rotolo. Changing legal systems: Abrogation and annulment. part ii: Temporalised defeasible logic. In Guido Boella, Gabriella Pigozzi, Munindar P. Singh, and Harko Verhagen, editors, *NORMAS*, pages 112–127, 2008.

37 D. Grossi, H. Aldewereld, and F. Dignum. *Ubi Lex, Ibi Poena*: Designing norm enforcement in e-institutions. In P. Noriega, J. Vázquez-Salceda, G. Boella, O. Boissier, V. Dignum, N. Fornara, and E. Matson, editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems II*, volume LNAI 4386, pages 101–114. Springer, 2007.

38 D. Grossi, D. Gabbay, and L. Torre. The norm implementation problem in normative multi-agent systems. In Mehdi Dastani, Koen V. Hindriks, and John-Jules Charles Meyer, editors, *Specification and Verification of Multi-agent Systems*, pages 195–224. Springer US, 2010.

39 Pascal Hitzler, Markus Krötzsch, and Sebastian Rudolph. *Foundations of Semantic Web Technologies*. Chapman & Hall/CRC, 2009.

40 J. Hübner, O. Boissier, R. Kitio, and A. Ricci. Instrumenting multi-agent organisations with organisational artifacts and agents. *Autonomous Agents and Multi-Agent Systems*, 20:369–400, 2010. 10.1007/s10458-009-9084-y.

41 Nicholas R. Jennings. On agent-based software engineering. *Artificial intelligence*, 117(2):277–296, 2000.

42 Andrew J. I. Jones and Marek J. Sergot. On the characterisation of law and computer systems: The normative systems perspective. In John-Jules Ch. Meyer and Roel J. Wieringa, editors, *Deontic Logic in Computer Science: Normative System Specification*, Wiley Professional Computing, chapter 12, pages 275–307. John Wiley and Sons, Chichester, UK, 1993.

43 Andrew J. I. Jones and Marek J. Sergot. A Formal Characterisation of Institutionalised Power. *Logic Journal of the IGPL / Bulletin of the IGPL*, 4:427–443, 1996.

44 Audun Jøsang. A subjective metric of authentication. In *Proceedings of the 5th European Symposium on Research in Computer Security (ESORICS)*, volume 1485 of *LNCS*, pages 329–344, Louvain-la-Neuve, Belgium, 1998. Springer.

45 H. Lopes Cardoso and E. Oliveira. Electronic institutions for B2B: Dynamic normative environments. *Artificial Intelligence and Law*, 16(1):107–128, 2008.

46 H. Lopes Cardoso and E. Oliveira. Norm defeasibility in an institutional normative framework. In M. Ghallab, C.D. Spyropoulos, N. Fakotakis, and N. Avouris, editors, *Proceedings of the 18th European Conference on Artificial Intelligence (ECAI 2008)*, pages 468–472, Patras, Greece, 2008. IOS Press.

47 H. Lopes Cardoso and E. Oliveira. A context-based institutional normative environment. In J. Hubner, E. Matson, O. Boissier, and V. Dignum, editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems IV*, LNAI 5428, pages 140–155. Springer, 2009.

48 H. Lopes Cardoso and E. Oliveira. Monitoring directed obligations with flexible deadlines: a rule-based approach. In M. Baldoni, J. Bentahar, J. Lloyd, and M. B. Van Riemsdijk,

editors, *Declarative Agent Languages and Technologies VII*, LNAI, pages 51–67. Springer, 2010.

49 H. Lopes Cardoso and E. Oliveira. Social control in a normative framework: An adaptive deterrence approach. *Web Intelligence and Agent Systems*, 9:363–375, December 2011.

50 Ebrahim (Abe) Mamdani and Jeremy Pitt. Responsible agent behavior: A distributed computing perspective. *IEEE Internet Computing*, 4(5):27–31, September 2000.

51 S. Modgil, N. Faci, F. Meneguzzi, N. Oren, S. Miles, and M. Luck. A framework for monitoring agent-based normative systems. In Decker, Sichman, Sierra, and Castelfranchi, editors, *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems*, pages 153–160. IFAAMAS, Budapest, Hungary, 2009.

52 Javier Morales, Maite López-Sánchez, and Marc Esteva. Using experience to generate new regulations. In *Proceedings of the twenty-second International Joint Conference on Artificial Intelligence IJCAI'11*, pages 307–312, Barcelona, Spain, 16/07/2011 2011. AAAI Press, USA.

53 John Mylopoulos, Lawrence Chung, and Brian Nixon. Representing and using nonfunctional requirements: a process-oriented approach. *IEEE Transactions on Software Engineering*, 18(6):483 –497, 1992.

54 Pablo Noriega. *Agent-Mediated Auctions: The Fishmarket Metaphor. PhD thesis Universitat Autònoma de Barcelona, 1997.* Number 8 in IIIA Monograph Series. IIIA, 1999.

55 Douglass C. North. *Institutions, Institutional change and economic performance.* Cambridge University Press, 1990.

56 Andrea Omicini, Alessandro Ricci, Mirko Viroli, Cristiano Castelfranchi, and Luca Tummolini. Coordination artifacts: Environment-based coordination for intelligent agents. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 1*, AAMAS '04, pages 286–293, Washington, DC, USA, 2004. IEEE Computer Society.

57 P. Pasquier, R. A. Flores, and B. Chaib-Draa. Modelling flexible social commitments and their enforcement. In M.-P. Gleizes, A. Omicini, and F. Zambonelli, editors, *Engineering Societies in the Agents World V*, volume 3451 of *Lecture Notes in Artificial Intelligence*, pages 139–151. Springer, Toulouse, France, 2005.

58 Alessandro Ricci, Michele Piunti, and Mirko Viroli. Environment programming in multiagent systems: an artifact-based perspective. *Autonomous Agents and Multi-Agent Systems*, 23(2):158–192, September 2011.

59 David Robertson. A lightweight coordination calculus for agent systems. In *Declarative Agent Languages and Technologies. DALT 2004*, volume 3476, pages 183–197. Springer, 2005.

60 William N. Robinson and Sandeep Purao. Specifying and monitoring interactions and commitments in open business processes. *IEEE Software*, 26(2):72–79, 2009.

61 Michael Schumacher and Sascha Ossowski. The governing environment. In Danny Weyns, H. Van Dyke Parunak, and Fabien Michel, editors, *Environments for Multi-Agent Systems II*, volume 3830 of *Lecture Notes in Computer Science*, pages 88–104. Springer Berlin / Heidelberg, 2006.

62 John R. Searle. *The Construction of Social Reality.* Free Press, New York, 1995.

63 Yoav Shoham and Moshe Tennenholtz. On social laws for artificial agent societies: Off-line design. *Artificial Intelligence*, 73(1-2):231–252, 1995.

64 Alberto Siena, Giampaolo Armellin, Gianluca Mameli, John Mylopoulos, Anna Perini, and Angelo Susi. Establishing regulatory compliance for information system requirements: An experience report from the health care domain. In *Proceedings of the 29th International Conference on Conceptual Modeling*, volume 6412 of *LNCS*, pages 90–103. Springer, 2010.

65 Herbert A. Simon. *Reason in Human Affairs.* Stanford University Press, 1983.

66 Munindar P. Singh. An ontology for commitments in multiagent systems: Toward a unification of normative concepts. *Artificial Intelligence and Law*, 7:97–113, 1999.

67 Munindar P. Singh. Semantical considerations on dialectical and practical commitments. In *Proceedings of the 23rd Conference on Artificial Intelligence (AAAI)*, pages 176–181, Chicago, July 2008. AAAI Press.

68 Munindar P. Singh. Trust as dependence: A logical approach. In *Proceedings of the 10th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, pages 863–870, Taipei, May 2011. IFAAMAS.

69 Munindar P. Singh. Norms as a basis for governing sociotechnical systems. *ACM Transactions on Intelligent Systems and Technology (TIST)*, pages 1–21, 2013. To appear; available at `http://www.csc.ncsu.edu/faculty/mpsingh/papers`.

70 Munindar P. Singh and Amit K. Chopra. Correctness properties for multiagent systems. In *Proceedings of the Sixth Workshop on Declarative Agent Languages and Technologies*, volume 5948 of *LNCS*, pages 192–207. Springer, 2009.

71 Munindar P. Singh, Amit K. Chopra, and Nirmit Desai. Commitment-based service-oriented architecture. *IEEE Computer*, 42(11):72–79, 2009.

72 C. Tampitsikas, S. Bromuri, and M. Schumacher. Manet: A model for first-class electronic institutions. In S. Cranefield and P. Noriega, editors, *12th International Workshop on Coordination, Organization, Institutions and Norms in Agent Systems (COIN)*, pages 105–119, Taipei, Taiwan, 2011.

73 Charalampos Tampitsikas, Stefano Bromuri, Nicoletta Fornara, and Michael Ignaz Schumacher. Interdependent Artificial Institutions In Agent Environments. *Applied Artificial Intelligence*, 26(4):398–427, 2012.

74 Nick Tinnemeier, Mehdi Dastani, and John-Jules Meyer. Programming norm change. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1*, AAMAS, pages 957–964, Richland, SC, 2010. International Foundation for Autonomous Agents and Multiagent Systems.

75 Y. B. Udupi and M. P. Singh. Dynamics of contracts-based organizations: A formal approach based on institutions. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, AAMAS, pages 19:1–19:3, New York, NY, USA, 2007. ACM.

76 J. Vázquez-Salceda, H. Aldewereld, and F. Dignum. Implementing norms in multiagent systems. In G. Lindemann, J. Denzinger, I. J. Timm, and R. Unland, editors, *Multiagent System Technologies*, volume 3187 of *Lecture Notes in Artificial Intelligence*, pages 313–327, Erfurt, Germany, 2004. Springer Verlag, Heidelberg, D-69121, Germany.

77 Javier Vázquez-Salceda, Virginia Dignum, and Frank Dignum. Organizing multiagent systems. *Autonomous Agents and Multi-Agent Systems*, 11:307–360, 2005.

78 Mahadevan Venkatraman and Munindar P. Singh. Verifying compliance with commitment protocols: Enabling open Web-based multiagent systems. *Autonomous Agents and Multi-Agent Systems*, 2(3):217–236, September 1999.

79 H. J. E. Verhagen. *Norm Autonomous Agents*. PhD thesis, The Royal Institute of Technology and Stockholm University, 2000.

80 Yonghong Wang and Munindar P. Singh. Formal trust model for multiagent systems. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1551–1556, Hyderabad, 2007. IJCAI.

81 Danny Weyns, Andrea Omicini, and James Odell. Environment as a first class abstraction in multiagent systems. *Autonomous Agents and Multi-Agent Systems*, 14(1):5–30, 2007.

82 Pınar Yolum and Munindar P. Singh. Flexible protocol specification and execution: Applying event calculus planning using commitments. In *Proceedings of the 1st International*

Joint Conference on Autonomous Agents and MultiAgent Systems, pages 527–534. ACM Press, 2002.

83 Eric S.K. Yu. Towards modelling and reasoning support for early-phase requirements engineering. In *Proceedings of the Third IEEE International Symposium on Requirements Engineering*, pages 226–235, 1997.

84 Franco Zambonelli, Nicholas R. Jennings, and Michael Wooldridge. Developing multiagent systems: The Gaia methodology. *ACM Transactions on Software Engineering Methodology*, 12(3):317–370, 2003.

(Social) Norm Dynamics

Giulia Andrighetto[1], Cristiano Castelfranchi[2], Eunate Mayor[3], John McBreen[4], Maite Lopez-Sanchez[5], and Simon Parsons[6]

1 Institute of Cognitive Science and Technologies, ISTC-CNR
Rome, Italy and European University Institute, EUI, Florence, Italy
giulia.andrighetto@istc.cnr.it

2 Institute of Cognitive Science and Technologies, ISTC-CNR
Rome, Italy
cristiano.castelfranchi@istc.cnr.it

3 LMTG/GET UMR5563, IRD-CNRS-Universite P. Sabatier Toulouse III
Toulouse, F-31400, France
eunate.mayor@gmail.com

4 Wageningen University
Wageningen, The Netherlands
johnmcbreen@gmail.com

5 University of Barcelona
Gran Via de les Corts Catalanes 585, 08007 Barcelona, Spain
maite_lopez@ub.edu

6 Brooklyn College, City University of New York
2900 Bedford Avenue, Brooklyn, NY 11210, USA
parsons@sci.brooklyn.cuny.edu

Abstract

This chapter is concerned with the *dynamics* of social norms. In particular the chapter concentrates on the lifecycle that social norms go through, focusing on the generation of norms, the way that norms spread and stabilize, and finally evolve. We also discuss the cognitive mechanisms behind norm compliance, the role of culture in norm dynamics, and the way that trust affects norm dynamics.

1998 ACM Subject Classification I.2.11 Intelligent agents, Multi-agent systems

Keywords and phrases Social norms, Norm generation, Norm spreading, Norm evolution, Trust, Culture

Digital Object Identifier 10.4230/DFU.Vol4.12111.135

1 Introduction

This chapter aims to identify the major aspects of norm dynamics, by which we mean the way that norms come into being and change through their life, as well as some of the relevant factors or determinants of the process that underlies this change. The need for a deep understanding of these dynamics is becoming a compelling task for the Normative Multi-Agent Systems (NorMAS) community because of the systems that the community wants to develop. Now, the members of the NorMAS community have a wide range of interests with respect to multi-agent systems so it behooves us to explain, before we get much further, what perspective the authors take on the dynamics of norms. We focus here on two view of multi-agent systems, views that we distinguish by referring to them as the 'engineering' perspective and the 'sociological' perspective.

Normative Multi-Agent Systems. *Dagstuhl Follow-Ups, Vol. 4.* ISBN 978-3-939897-51-4.
Editors: Giulia Andrighetto, Guido Governatori, Pablo Noriega, and Leendert W.N. van der Torre; pp. 135–170
DAGSTUHL FOLLOW-UPS Dagstuhl Publishing
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany

From the *engineering* perspective, we are interested in building multi-agent systems that are flexible and open. In other words, rather than rigid systems in which all possible behaviors are determined at design-time, we are interested in developing systems that are open—in the sense that agents can enter or leave the system at any time—and systems that are able to evolve over time. This evolution will include the evolution of the mechanisms that govern the system. This desire for flexibility points to the use of norms to regulate the behavior of these systems, and so we are interested in the properties of norms in human societies, seeing this as a rich source of evidence for the development of norms in multi-agent systems. From the *sociological* perspective, we are interested in understanding human societies, in particular the mechanisms that allow these systems to self-regulate, developing and modifying rules of behavior. From this perspective, multi-agent systems are a useful research tool, one that allows models of societies to be constructed and experimented with in a way that is not possible with human subjects.

Of course, these two views are neither exhaustive (there are other perspectives within the NorMAS community) nor are they exclusive (research can contribute to both perspectives), but the reader should bear both these perspectives in mind when reading this chapter.

Now, we also need to stress that *norm* is a polysemic term with possible multiple meaning, and is used in a number of different ways. Let us distinguish three basic and fundamental meanings of the term 'norm' that we use in this chapter.

1. Norm meant as *normal* behavior or feature. In this sense 'norm' is understood as a reference to the normal (or Gaussian) distribution and we use the term in a statistical sense. What is normal? What corresponds to the norm in this sense? That which is around (standard deviation) the median value (mean). More extreme, and thus rare, values (behaviors, features) that are not *regular* are considered as *deviant*. This is just a statistical and descriptive notion of norm. It does not necessarily have a prescriptive nature nor it is assumed as ideal and model. Of course, usually very deviant features are not assumed as ideal, good. For example very small feet or very big ones. However, for example for intelligence an abnormal measure can be assumed as an ideal model, the best.
2. Norm meant as *model* or *standard*, *ideal-typical.* An example is provided by the normative theory of rationality in economics that indicates what would be the perfect rational (economically efficient) reasoning and decision. 'Normative' in this sense does not mean *prescriptive*; it is not an imperative or obligation to reason in such a way. It just is a useful parameter, a model, for comparing different decision processes and evaluate them. It is well known that human beings do not reason and decide in such a way [1] (consider for example, Herbert Simon's *bounded rationality*; cognitive heuristics and biases; affective intelligence, etc.). Also in other domains, it is clear that the ideal-typical characterization is not statistically dominant; for example only 15% of population has eyesight as ideally described in the textbooks of ophthalmology; but it is a precious comparative standard for evaluating the various deviations from it.
3. Finally, norm meant as an (explicit or implicit) imperative regarding the behavior of social actors, i.e. specifying how they should behave. Norms that are explicit are often those formal norms introduced by organizations or societies through some kind of legislative or regulatory process. Such norms are often handed down to societies of agents from above. Other norms, and this is what we mean here by the term *social norm*, are those that are

[1] Though, of course, mainstream economics continues to pretend that they do.

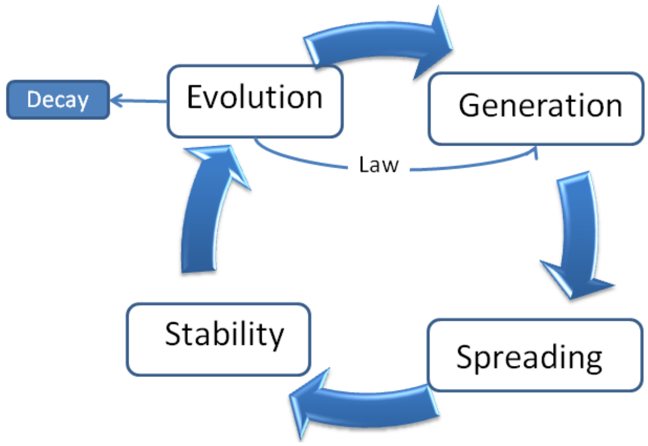


Figure 1 The lifecycle for norms.

implicit and emerge from some form of evolution and self-organization by the agents that are regulated by them.

We view the lifecycle for norms as following the cycle in Figure 1, and this cycle forms the basis for the discussion in this chapter. The cycle starts at the top right. Norms are generated, and after generation may spread through a population and become widely adopted. If the norm spreads, it will eventually reach stability, having been adopted by some portion of the population and failing to grow any further within the population (though, of course, it may stabilize at a point where it has been adopted by the whole population). With stability norms will persist. Some will then decay, being superceded by new norms, some will evolve and some may be codified into law. The norms that evolve or become law can be considered as generating new candidate norms that will then spread or not.

This cycle is not identical for every type of norms. Some norms, such as social norms, emerge spontaneously from self-organizing groups of agents, and are usually implicitly (but sometimes explicitly) *negotiated* by the agents. Other norms, for example the legal and the organizational ones, are *decided* by some institution (like the parliament or the boss) and are officially *proclaimed* by an authority[2]. So the process of norm generation and the process of spreading are not identical for all kinds of norms. The spreading of a custom or social norm is based on behavioral messages (signals) and on conformity, imitation, etc.; while the spreading of a given official instruction or rule is realized through other specific channels, e.g., publication, issuing, etc. We should also note that norm lifecycle as we have sketched it mainly applies to the second and third class of norms we distinguished, but can also be seen to apply to statistical norms.

The consumption of tobacco illustrates a number of aspects of the lifecycle and it well applies to norms of different kinds. Tobacco consumption in Europe (and because of the availability of data we will consider consumption by women) through the last hundred years or so can be considered as a statistical norm, a social norm, and has ended up regulated as a legal norm (in the form of a prohibition). In Europe [43], it seems that smoking was not widespread among women until the twentieth century. It subsequently became a norm that spread, with close to 50% of the population in the UK, Denmark and the Netherlands

[2] This is the original etymology of *lex*, that is, officially declared.

smoking by 1970. In the UK this number then stabilized, with somewhere between 40% and 50% of women smoking by 1990 (when the data in [43] ends). At the same time, the smoking norm for men was being discarded — in 1950 70–90% of the male population in Northern Europe was smoking, and this fell to 30-40% by 1990. Looking more closely (again using data from [43]), it seems that among smokers, what was consumed changed in the early years of the twentieth century. Driven by the tobacco industry, consumption among (predominately male) smokers shifted from dark tobacco—cured in traditional ways and consumed in pipes, cigars and hand-rolled cigarettes—to manufactured cigarettes that used a lighter tobacco. In the lifecycle, this can be seen as the evolution of a norm, with the consumption of dark tobacco evolving into the consumption of blond tobacco, and the means of consumption evolving also. The new norm stabilized, and the old norm decayed.

Given the data in [43], what we are primarily talking about here is a statistical norm—the number of people smoking from which we can infer what 'typical' behavior would be. However, given that smoking is an activity under the control of the agents that indulge in it (unlike, for example, their height), it is not much of a stretch to imagine that the statistical norm is accompanied by a social norm, that is the figures we quote above are a reflection of how socially acceptable it was for European women to smoke.

Tobacco consumption also illustrates how laws can lead to new norms. Various countries have brought in smoking bans (i.e., prescriptive norms of the third kind) in the last ten years or so. To some extent these bans have had the desired effect of helping the smoking norm to be discarded by some members of the population. However, they have also led to a new norm—that of people standing outside a bar or restaurant in order to smoke while still occupying space inside the building (leaving food and drinks at the table).

Having briefly sketched the norm lifecycle in this section, we will examine it in more detail in the rest of this chapter. The principal steps of the dynamics of norms—generation, spreading, stabilization and evolution—are discussed in some detail in Sections 3.1, 3.2, 3.3. We suggest that for a well-founded and innovative study of norms, it is necessary to look at the cognitive mechanisms underlying the dynamics of norms, and we describe these in Section 2. In addition, we believe that it is also important to consider the role played by trust (see Section 4) and cultural variation (see Section 5).

2 The Minds of the Agents

If one aims to understand how social norms are able to regulate the conduct of intelligent and autonomous agents, it is necessary to analyse the cognitive mediators that make it possible to transform normative requests into actions [29, 30, 31]. This section aims to present a model of the cognitive mechanisms underlying norm compliance.

With *social norms*, we refer here to informal rules which prescribe what we ought or ought not to do. Social norms are behaviours spreading over a society on condition that the corresponding prescriptions and mental representations (namely, sets of beliefs and goals concerning the norm) spread as well. We will then refer to the third definition of norms provided in the introduction, the one that focuses on their *prescriptive* nature.

Norms are influencing mechanisms, aimed to direct the future conduct of the individuals subject to the norm. To influence the behaviour of autonomous systems, a complex mental dynamics must be activated. In order for autonomous agents to undertake (or refrain from undertaking) a certain course of action, it is not sufficient that they know (i.e., have the belief) that such a course of action is desired by someone [30]. It is necessary for them also to have the goal of performing such an action. A goal is here meant in the very general sense derived from cybernetics, i.e., a desired state of the world triggering and driving actions [30, 68]; for an analysis of the differences between goals and other motivational states, see [23].

Since social norms are intended to guide the behaviour of agents, they are a powerful means for the generation of new goals in the minds of the individuals subject to the norm. The repeated exposure to the normative conduct of others and to their normative requests and expectations (be they implicit or explicit) generates in the minds of individuals the belief that there is a norm and there is a widespread will which prescribes that he or she should (is asked to) observe that impersonal and widespread will (normative expectation). This normative knowledge has then the power to activate or to generate the goal to observe the norm in question.

But what is the mental path or process through which the normative requests and expectations generate these goals, goals which are, in themselves, normative as well?

In order to understand the motivational power of normative requests and expectations, it is useful to sketch a cognitive anatomy of the latter type of representations. As suggested by [67], an expectation is a hybrid mental representation which consists of an epistemic component and a motivational one. Expectations are different from simple predictions (epistemic representations about the future), in that the individual not only has the belief (a prediction with a certain degree of certainty) that a certain event or state will occur, but also has the goal (with a specific value of importance) that it will (or will not) occur, and, to this end, actively monitors the process.

Expectations are not just beliefs, they imply a personal concern about the anticipated event. Having an expectation means that the agent is waiting for something (not) to happen. There is a goal involved, that is why there are positive and negative expectations. Specifically, a normative expectation simultaneously consists of (a) the belief about the existence of the norm and (b) the goal that the norm is (not) complied with (by those who are subject to it).

To reinforce the probability that a normative expectation will be fulfilled, the holder must act in such a way as to influence others' behaviour and possibly even their minds. They should provide the addressee with the relevant information, thus he must communicate the existence of the norm and that there is a (widespread and impersonal) will that it should be observed, and, in the opposite case, that there is possibility of punishment or sanction. This prescriptive nature is a crucial aspect of the normative mechanism. The normative expectation contains, in itself, a request, which individuals must decide whether to adopt or not, and in order to do this, they must recognize the normative goal which it transmits.

How is it possible that the goal (need, desire, objective, request, order, ...) of somebody else succeeds in regulating behavior of an autonomous agent? Autonomous agents have the ability to choose whether or not to have a given goal and this choice is conditioned to the belief that such a goal is a means by which they can achieve other goals that they already possess. An autonomous agent acts to achieve her own goals and must have reasons for choosing whether to act as she does. In particular, if she accepts another's request, she must have good reasons for doing so. The general mechanism by which an autonomous agent adopts external requests, called goal adoption, has been described at some length in [30, 68] . Here, suffice it to say that an agent (the adopter) will adopt another agent's (i.e., the adoptee's) goal as hers, on condition that she, the adopter, comes to believe that the achievement of the adoptee's goal will increase the chances that she will in turn achieve one of her previous goals. For example, I will accept your request to lend you my laptop, if in return you allow me to borrow your car tonight. When the external request is a prescription, a special application of this process occurs, i.e., norm adoption. I will adopt the norm if, say, I think that by doing so I avoid getting a fine, obtain others' approval, build a good reputation, etc. General adoption leads to the formation of social goals (achieve somebody else's goals). Norm adoption leads to the formation of normative goals (comply with the norms), i.e., a goal that an agent happens to have because and to the extent she has the correspondent normative belief.

An agent can decide to adopt a norm, and form a normative goal for several reasons (for a detailed analysis, [30, 68]):

- Instrumental reason: the subject adopts the normative goal if she believes she can get something in return (avoid punishment, obtain approval, praise, reward, etc.).
- Cooperative reason: the subject adopts the normative goal to achieve a common goal. Norm-adoption is cooperative when it is value-driven, that is, when the subject shares both the end of the norm and the belief that the norm achieves it. For example, an agent may decide to conform to the refuse-recycling norm because she believes that, by doing so, she helps to reduce the negative impact of not doing so on the environment.
- Terminal reason: The subject wants to observe the whole set of norms she is subject to, as ends in themselves. She has the terminal goal or value that "norms be respected" (Kantian morality).

It is interesting to notice, that normative goals can be formed for different reasons, also for self-regarding reasons, as in instrumental norm adoption. However, this does not mean that the generated goal is not normative in its nature, it is. A normative goal is a goal relativised to a normative belief, i.e., held because and to the extent that it is believed to be exacted by a norm. For a goal to be normative, all that is necessary is that it is based upon internal representations of a normative nature [27].

In this section, we have described the standard path of norms in agents' minds. However, this path is rarely followed in its totality and it is more plausible to consider that shortcuts take place during the computation of the normative input. Norm conformity in humans is usually due to the automatization and simplification of a such a rich cognitive decision process. After a repeated normative learning and especially for simple behavioral rules (like stopping at the red light) the subject just reacts automatically to the normative stimulus. For example, when you are driving and you see a red traffic light you will automatically stop. In this situation, it is not necessary that input follows the complete mental path.

The normative beliefs and reasoning remain inactive, but they remain present in the subject's mind and they will be re-activated if needed. For example, a given routine must be blocked when given conditions activate inconsistent prescriptions. A car driver stopping at a red traffic light might see a policeman asking her to move on. In such a case, the car driver needs to be able to retrieve control of her action, block the automatism, and decide which normative input should be given priority. Here the norm is explicitly taken into account and we reason about violating it or not [5, 29].

3 The Lifecycle of Social Norms

This section goes more deeply into the *Norm Lifecycle* introduced in Section 1. Lifecycle models of norms have been also proposed by [85] and [92] (see also the chapters "Normas in MAS: Definitions and Related Concepts" and "Simulation and NorMAS" in this volume). [85] takes a simulation perspective, while [92] examines norm dynamics from a lower-level operational point of view, i.e., in terms of its key states and transitions. The lifecycle model we present in this chapter is not identical for all kinds of norms, we believe that it is widely applicable. As already described, the lifecycle starts with a *generation* phase. Generated norms may then *spread* through the population and eventually reach *stability* (i.e., persistence) if they are adopted by enough individuals in the population. Afterwards, different evolution stages may occur: norms may *decay*, being superceded by new norms; norms may evolve; or alternatively, they may be *codified into law*. We discuss how the last two cases close the lifecycle.

3.1 Norm Generation

The life cycle in Figure 1 starts with the norm generation process. This process may be triggered when a brand new organisation lacks norms to structure individual interactions or when an existing organisation requires the addition of new norms. The intrinsically dynamic nature of Internet and its commonalities with organisational MAS approaches offer many opportunities for the emergence of organizations with specific objectives. Some examples can be found in economic coalitions, working teams, social communities, etc. Defining the rules of the game for brand new organizations may not be a straightforward process, so we claim that the automatic generation of organisational rules will become an important topic in the Normative MAS area if we understand the term "norm" in its broad sense of social conventions.

Norm generation can be tackled from different approaches. In this section we provide a brief overview of the main approaches that have been taken so far by the research community, their assumptions and the open issues that still require further consideration.

Formal approaches to norm generation have been generally referred to as *norm synthesis*. These approaches exhaustively enumerate the full state space in order to define norms that ensure access to goal states whilst disallowing access to undesirable states in the state space. Therefore, exhaustive formal approaches explicitly define at design time which actions are allowed and which actions are forbidden within a specific scenario. By definition, their results are proven to be complete. Nevertheless, the intrinsic complexity of these formal methods causes them to be intractable for real world systems. In fact, Shoham and Tennenholtz [90] have shown this complexity to be NP-Complete by performing a reduction to 3-SAT. Furthermore, a typical assumption of these approaches is that the system will be static (or at least that the system dynamics will be known and included at design time in the state space being explored).

An attempt to reduce complexity is the work by Christelis and Rovatsos [25], which proposes an automated method for synthesising prohibitive norms in planning-based domains at design time. This method includes norm generalisation and performs a local search around the state space, disallowing access to undesirable states but ensuring accessibility to goal states. Despite its improvements, this approach remains intractable for real world systems. Regarding system dynamics, it is also important to consider the agents that participate in the organisation, since it may be the case that agents decide not to comply with the generated norms. Ågotnes et al. [1] have formally (logically) studied the extent to which a normative system is robust, i.e., its effectiveness when populated by non-compliant agents. Still at design time, they investigate different robustness notions such as identifying those key agents that are necessary and/or sufficient to guarantee the effectiveness of a normative system or the proportion of an agent population that must comply in order to ensure success.

On the other hand, some research approaches define systems that require participant agents to be involved in *agreement* processes on the norms to follow. These systems constitute a democratic mechanism that reaches agreements (i.e., converge) eventually. Nevertheless, they may require participating agents to be enriched with abilities that go beyond its social basic capabilities (that is, those required for performing specific tasks or for participating in the community to obtain an individual or common goal). An example is the work of

Artikis [6], where agents are able to update norms by using an argumentation protocol to support discussion. Furthermore, they can even discuss about the argumentation protocol itself in order to update it. Design-time requirements for these systems are thus related to the argumentative capabilities of the agents, the specification of argumentation protocols to follow, and the knowledge agents need in order to be able to successfully complete the generation process. Nevertheless actual norm generation occurs at run-time, and therefore, the time and effort (in terms of reasoning resources) required for proposing and agreeing upon a new norm needs to be taken into consideration. Generation time may also be influenced by specific protocol parameters such as the degree of consensus required for a norm to be generated.

Nowadays, *norm* (or convention) *emergence* provides the approach to norm generation that is attracting most attention from the research community. This mechanism does not require any oversight or centralized control and thus it is defined in settings where agents lack of an explicit organisation. Norm emergence is based on the autonomy of agents for choosing individually a solution (i.e., a specific behaviour convention) from a space of alternative solutions. Afterwards, they consider a convention has emerged when a majority of agents actually choose the same actions (that is, the society converges towards a convention when agents perform the same actions). For example, in a traffic scenario, agents may decide whether to drive on the left or on the right, and we will say that a norm has emerged when most agents drive on the same side of the street.

Agents are generally considered to be self-interested, like in [44], where Griffiths et al. study how to establish a suitable set of norms in a decentralized population and the problem of group recognition by using observable tags as markings, traits or social cues attached to individuals. These approaches usually consider a number of repeated interactions involving pairs of agents (repeated two-player games). Differences lie in the assumptions about agent interaction: some consider general interaction patterns such as uniform random one-on-one interaction probabilities [89, 91] whereas others study the impact of societal topology that constrains the interactions of individuals [33, 60, 99]. This norm generation process is therefore intermingled with the subsequent steps of spreading and stability in the norm life cycle. Since these steps are further detailed in the remaining of this section, rather than covering all approaches exhaustively, we will just mention that it is also possible to consider some kind of observation as a requirement or instrument to adopt norms [33, 81, 99] and that internalisation is also proposed as a mechanism to guarantee norm acquisition [28].

Overall, it is worth mentioning that norm emergence approaches include some parameters that may be decided at design time. For example, one that is common to all of them is the threshold of compliance in the society that will be used to consider a norm to have emerged. Additionally, although most works on norm emergence are empirical, there are exceptions such as the one by Brooks et al. [18] that presents a mathematical model of the emergence of norms based on utilities. Specifically, they model the emergence of norms in societies of agents who adapt their likelihood of choosing one from a finite set of options based on their experience from repeated one-on-one interactions with other members in the society. Their goal is to study both the process of emergence of norms as well as to predict the likely final convention that will emerge if agents had preconceived biases for certain options. All these approaches to norm emergence pose very interesting research questions, nevertheless, it is important to notice: i) how much they rely on the fact that the initial set of alternative norms have to be known (by the agents) at design time; as well as that ii) convergence often depends on initial conditions. From our point of view, the research community can transform these limitations into research opportunities when trying to overcome them. As for any other research area, the proposed approaches should be as general (and domain-independent) as possible. And this should apply also to the empirical approaches by taking advantage of those general MAS characteristics that are common to most systems.

An alternative approach to norm generation is that from Morales et al. [73] which learns norms based on run-time experience. This solution follows a 'division of concerns' paradigm where the majority of agents participating in the organization can act by simply conducting their domain specific activities without enduring the process of establishing new norms, which is left for specialized regulatory agents. These regulatory agents are staff (i.e., organisational) agents devoted to: first, proposing (a set of) rules for the system; and second, to monitoring to what extent these rules are effective in regulating the organization. This empowerment distribution is not meant to generate a centralized totalitarian system though. Instead, their aim is that these regulatory agents propose norms that smooth organizational activity and agent interactions. This is done by detecting when conflicts arise and by including new norms with the aim of preventing those conflicts from happening in the future. Once these regulations have been established to the system, these regulatory agents will be also in charge of evaluating their adequacy based on the organizational run-time experience. This is measured in terms of agent compliance and the resulting conflicts: agents may decide whether to observe norms or not and these decisions may have different consequences in the system. Therefore, the system does not focus on an enforcement mechanism that invalidates agent actions whenever they do not conform with the established rules. On the contrary, regulatory agents observe the agents and "listen" to them, since agents may not comply with norms if they consider they are not necessary (see 3.2.2 for a more detailed discussion on violation). The underlying assumption here is that it is possible to identify an unnecessary norm whenever agents violate it and no bad consequences (i.e., conflicts) arise. In fact, if this is the case in a number of experiences along the system execution, regulatory agents can then consider they have gathered enough evidence against the norm so that they can discard the norm from the set of currently established norms.

Overall, it is worth noticing that this empirical approach does not explore the complete state space but that it is flexible enough to deal with system dynamics. It can be seen as a non-intrusive, autonomy preserving, norm generation mechanism. Moreover, the amount of knowledge it requires at design time is rather limited. The basic requisites for the regulatory agents are conflict identification, and monitoring of agents' norm compliances and violations. Nevertheless, some parameters still need to be defined at design time. Some examples are the quality threshold that is required for a norm not to be discarded or the amount of evidence that each norm violation/compliance provides. Convergence time in this case depends on the conflict frequency.

As in Savarimuthu et al. [84], previously introduced approaches for including norms in a multi-agent system can be classified into two different categories. The first one is the top-down approach, where an institutional mechanism specifies (prescribes) norms that regulate agents' behaviour. Formal approaches in which norms are specified at design time could also be included in this category. The second one is the bottom-up approach, where agents locally interact in order to spread the adoption of suitable norms within the society. Both emergence and agreement approaches fit in this category. Nevertheless, a balanced mixture between top-down and bottom-up approaches may lead to the best results. This is the reason Morales et al. [73] advocate for methods where norms are proposed by regulatory (top) agents and validated against the experience of domain (bottom) agents.

3.2 Norm Spreading and stabilization

Once norms have been generated, transmission is the process by which norms spread from one agent to another. As shown by a great number of works on the classification of norms [101, 79], there is no single modality through which a norm can be transmitted. A norm can, for example, be communicated through explicit commands, orders, or requests, both written and oral, “do this”, “don’t do that”, or by means of declarations which mention or imply deontic predicates, such as “x must/must not do y”, or “you must/must not do y”. Evaluations are also powerful means through which norms are transmitted, in phrases such as “paying your taxes is right/tax evasion is wrong”, an evaluation relative to the state of things which is derived from the execution (or not) of normative action is expressed (for a more detailed analysis of the communication of norms, see [27]). [54] refers to this type of transmission techniques as *active transmission*. “Active transmission occurs when one agent purposefully broadcasts a set of norms to neighboring agents” [54]. The authors suggest that this type of transmission is usually accompanied by social enforcement, such as social sanctions aimed to deter others from violation. In multi-agent systems, examples of the use of active transmission is the use of norm entrepreneurs [51] and that of sanction [3, 20, 61, 63, 98]. Over time this process favours the diffusion of norms throughout the entire group.

3.2.1 Implicit normative communication

In this work, we aim to stress how the spreading of norms does not occur exclusively through active transmission and how a privileged role in this process of norm spreading is played by communication realised through actions or behavioral implicit communication ([54] refers to this process as *passive transmission*). With Behavioral Implicit Communication (BIC), we refer to a specific type of communication in which there is no specialized signal (neither arbitrary acoustic signal nor codified gesture), but the message is conveyed through a practical action. In BIC, the subject performs a usual practical action (e.g., walking, drinking, etc.), knowing that somebody is observing her and is able to understand the behavior (or the result of the behavior) she performed, and this is one of the goal for performing the practical action (for a more complete taxonomy of the many forms of implicit behavioral communication, [24]). In other words, “X performs the behavior b in order Y perceives it and on such a basis believes that p” . Consider, for example, a friend who, during a dinner party at her own house, places the ashtrays only on the balcony. By this action, she is communicating to the guests that they can only smoke outside and that smoking is not permitted inside her apartment. This form of communication is *behavioral* because it exploits usual practical behavior that is not conventionalized (e.g., an arbitrary acoustic signal or a codified gesture); and it is *implicit* because, not being codified and specialized, its communicative character is unmarked, undisclosed, non-manifest and thus deniable. Behavioral implicit communication plays a key role in the spreading of norms. In particular, in the next subsection, we will show how the acts of obeying, violating and defending norms in terms of BIC provide interesting insights for understanding the dynamics of norm spreading.

3.2.2 Obedience and Violation

The action of obeying a norm is an act which can convey important information. It can convey, for example, to whoever observes (and monitors), the information that the conduct prescribed by the norm has been performed. From this information, the observer can infer new details regarding the normative actor, such as that the actor is an obedient individual and is therefore trust-worthy. When individuals realise the expressive (i.e., demostrative) power of their own actions, they can intentionally decide to perform those actions in order to influence the minds of others in a normative way. They can, for example, decide to observe a norm (also) in order to communicate the fact that they have obeyed it, thereby avoiding being punished and in the hope that they will be considered as obedient and trust-worthy citizens. In the same way, that they can decide to observe the norm simply to *set an example* and thereby influence the others to do the same. As the action of observing a social norm involves a cost, whoever obeys does not want to be the only one to uphold it, and wants the normative costs to be distributed equally [27, 30]. Thus, behavioral implicit (normative) communication simultaneously has both an informative and a prescriptive nature. On the one hand, it transmits information about (1) who performed the normative act; (2) the existence of the norm; and finally (3) the fact that the norm should be respected. On the other hand, the behavioral implicit communication goes beyond mere informing and also prescribes the correct conduct to follow, asking the addressees themselves to comply with the same norm and indicating the consequences if not. Through behavioral implicit communication normative expectations and requests are transmitted and spread within the population.

The violation of a norm can also be a communicative act. For example, the desire to communicate to others (especially to those in authority, be their parents or the state) that one has violated a norm is characteristic of the provocative behavior of adolescents, and of revolutionary movements. One recalls Gandhi publically burning his South African pass[3], a document which all Indians had to have in their possession at all times, but which no white South African did. On this occasion, Gandhi communicated by means of a gesture, not with words, his indignation at the injustice, racism and exploitation to which Indians were subjected.

A norm can be perceived by an agent as more or less salient. With salience, we refer to the perceived degree of importance and strength of a norm [5, 13, 26, 61, 98, 102]. The actions of others, e.g., their compliance or violation of a norm, are important cues from which an individual can infer how salient a norm is. This is particularly true when no punishment follows the violation. Even if it is not the actor's intention, the act of violation signals that a norm is poorly salient, and the lack of punishment reinforces even more this perception. Conversely, by observing a large number of acts of compliance with a norm, agents can infer that the norm is highly salient and deserves respect. Psychological evidence suggests that the more a norm is perceived as salient, the more likely it will be complied with [26]. The perception of a lack of salience or of the weakening of salience can cause a reduction of the motivation (be it instrumental or terminal) to conform to the norm. If an individual, for example, observes a norm for reasons which are purely instrumental (or to avoid punishment), the perception that both the norm and the normative will have lost their strength allows him to infer that the motivation to defend the norm has also diminished, and, as a consequence, the probability of being punished has decreased too. If, however, an individual obeys a norm for terminal reasons, because a norm, as such, must be respected, then the belief that the social norm is slowly loosing its salience reduces the motivation to comply with it. When norm transmission is not explicit it is possible that the observer misunderstands the norm thus activating a process of norm change or norm innovation [4, 48]. A simplified approach

[3] Under the 1906 Transvaal Asiatic Registration Law all Asians who were eight years old or more and resident in the Transvaal were required to register with the Transvaal government and carry a registration document. This was an extension of the exiting 'pass laws' which already restricted the movement of the black population.

to the implicit transmission of norms has been implemented in simulations in which agents copy the norms of the more successful agents in their group [17, 37].

3.2.3 Punishment

As with norm obedience and violation, the reactions to these actions can be communicative acts through which normative requests are transmitted within a group. Punishment, if properly modelled, communicates to the offender (and also to observers) that through his conduct, he has violated a norm and that such violation is not approved of. In the large part of existing work, punishment is looked at from the classical economic perspective as a way of changing wrongdoers' conduct through the infliction of material costs [9]. As suggested by [61, 98], this way of looking at punishment is incomplete and a more insightful understanding of this practice is available once its norm-signaling nature is identified and properly exploited. The norm communicative power of punishment has been supported largely by legal theorists, who claim that a well designed punishment mechanism should explicitly express disapproval for norm violations and should provide cues for appropriate conduct [76, 93].

As claimed in [61, 98], if properly designed, punishment not only imposes a cost for the wrongdoing, but also informs violators (and the public) that the targeted behaviour is not approved of because it violates a social norm. [42] have referred to this mechanism as *sanction*, thus distinguishing it from material punishment. Since sanction communicates the presence of norms and asks people not to violate them, it allows agents to learn of the existence of norms and their relative salience and that their violation is not condoned. In particular, it will generate the belief that the violation of the norm in question will result in a sanction (this way making explicit the causal link between violation and sanction "you are being sanctioned because you violated that norm") that can be more or less severe depending on the salience of the violated norm. Thus, the information conveyed by sanction is twofold: on the one hand it communicates to the offender (and possibly to the observers) the *existence* of a norm, the consequences resulting from this violation and the *legitimacy* of this reaction; on the other hand, it indicates the specific salience of the norm and the *seriousness* of the violation. This normative information has the effect of creating in the mind of the sanctioned agent a set of normative beliefs and possibly of generating the normative goal to comply with (and possibly enforce) the norm in the future. The severity, legitimacy and frequency of sanctions are important cues from which to infer the salience of a specific norm and the seriousness of the violation, information that directly affects the cogency of the normative goal. An agent endowed with normative goals will compare them with other goals of hers and to some extent autonomously choose which one will be executed. The more cogent the normative goal is, the more likely it will outcompete other goals of the agent.

The norm-signaling component of sanction allows social norms to be activated, to increase their salience and to spread more quickly in the population than if they were enforced only by mere punishment. [61, 98] show by a simulation experiment that the use of sanction, the enforcing mechanism that supplements material punishment with normative information, promotes a higher level of norm compliance in a group, makes it more stable and reduces promotion and maintenance costs compared to the use of material punishment alone. [86] shows by simulation experiments the signalling power of punishment and its effect in favoring norm learning, while several mathematical investigations and agent based models have explicitly studied the role of punishment in the transmission and evolution of norms.

Often sanctions are accompanied by explicit messages, often oral, such as "you don't behave in this way" or "you shouldn't have done it". These messages do not necessarily have to be transmitted through explicit communication (oral or written). The fact that a form of behavior or conduct has violated a norm and that such a transgression is not approved of can also be communicated by practical actions, which can be more or less violent. Consider, for example, a pedestrian who decides to communicate his rage and indignation at the owner of a car which is illegally parked in a space reserved for disabled motorists by deliberately damaging the car's windscreen wipers. In addition, there are contextual factors that facilitate and, in some cases, amplify the behavioral implicit communication of the normative request thereby increasing its motivational power [102]. For example, when punishment is not meted out by a single individual, but by a group of people (or part of it) who co-ordinate themselves to do so, it is easier for the person sanctioned to interpret distributed punishment as aimed to defend a norm rather than driven by a personal interest [16, 97]. A cross-methodological study by [97] shows that a constant punishment level has a stronger effect when it comes from more punishers than when it comes from fewer punishers.

3.3 Norm Evolution

In his introduction to *Law and Revolution*, H. J. Berman refers to law as 'law in action', a "living process of allocating rights and duties and thereby resolving conflicts and creating channels of cooperation." [11, page 5]. This 'ongoing character' of law and institutions, its self-conscious continuity in time, appears to be built upon a conscious process of continuous development, conceived as a process not merely of change but of organic growth.

This concept of 'conscious organic development' of law has been largely applied to eleventh- and twelfth-century institutions, which "were expected gradually to adapt to new situations, to reform themselves, and to grow over long periods of time." [11, page 6]. However, neither can every change be seen as growth, nor does growth necessarily mean the expression of a deliberate will to achieve particular goals.

In Western legal tradition, law is conceived as an integrated system, a 'body' (of law), a *corpus iuris*, which continuously develops over time and generations. Unfortunately, this dynamic character is not self-sustaining. The body of law, as a coherent whole, only survives because it contains 'a built-in mechanism for organic change', that is, an internal logic. In that sense, changes are not only adaptations of the old to the new, but they are also part of a pattern of changes that is coherent with the system over time. Thus, those changes do not occur at random, but are the subject of a development process that relies on the existence of 'certain regularities and, at least in hindsight, reflects an inner necessity." [11, page 9].

We are, however, far from substantially understanding the 'mechanical' machinery lying behind the emergence and stability of social and customary norms. Although the essence of law lacks, by nature, the verifiable characteristics that are inherent to the subjects of study of exact, or 'natural' sciences, in the following sections, we examine the essential elements to configure this 'bootstrapping formula' about the stability and evolution of codified norms.

3.3.1 Stability and codification of norms

Some legal rules are not enacted by a legislator but grow instead from informal social practices[4]. Legal scholarship is divided about the relevance and consistency of such type

[4] When referring to customary law, it is often said that courts do not 'create' law, but they rather 'find' it. An example of this can be found in Cooter [32, page 216], where he relates how English merchants in the medieval trade fairs developed their own rules. In some cases, they even had their own courts. However, when those courts were unable to deal with the increasing amount of disputes generated as commerce grew, English judges where responsible for assuming jurisdiction. Since judges where outsiders, with limited knowledge of the special issues concerning trade, "instead of imposing rules,

of norms. Thus, unlike positive or natural law, customary law has received little scholarly analysis and its dynamics remain highly unknown. Part of the doctrine seems to be reluctant to accept its significance, considering customary law a secondary object of study, a 'minor' source of law that grows only when it is necessary to fill legislative gaps left by the legislator. On the contrary, however, some jurists and philosophers consider that all manifestation of law is based on a pre-existing custom; that is, that custom provides the indispensable frame of shared moral and legal reasoning in which law is embedded.

The question is how to fill this gap? The lack of a proper analysis of the dynamics that determine the normativity of customary norms constitutes an important *lacuna* in our understanding of the normative conformity of individuals, both from a social and a cultural perspective.

3.3.2 The normativity of 'informal' sources of law: open questions

When dealing with norm acquisition, the role of cognition is generally poorly explored and mostly refers to the individual motives that lead to compliance with norms. That is, the common individual traits that lead the addressees of a norm to recognize and internalize the normativity embedded in the latter. This may lead us to conclude that, from an individual point of view, human nature presents essential characteristics that predispose us for normative conformity; however, such elements are, albeit crucial pieces of the great puzzle of norm compliance, not the only ones.

Constant behavior and belief in its obligatoriness

The characterization of human beings as 'legalis homo' is not only a product of the individual spirit of men, but it is also in great measure a part of their sense of belonging to a group. Strangely enough, however, until the mid-1990s[5] the attention paid to the 'social' source[6] of law's normativity has been scarce.

Undeniably, group-awareness and group-membership play a key role in normative conformity. Many aspects of our daily lives witness behaviors which are motivated not just by internal psychological drives, but that may also be influenced by the environment in which the individual resides[7], in the form of social or peer pressure.

This 'social' feature that emerges from culture is even clearer in the case of customary law. The idea of the necessity of some sort of inherent in-group homogeneity in order to make it easy to overcome coordination issues —or any other kind of problems resulting from the coexistence of different individuals— seems intuitively appealing. However, a crucial distinction should be made before we continue, regarding the conceptual difference between regularity of behavior and rule-following. Certainly, the existence of a social rule

[English judges] tried to find out what practices already existed among the merchants and enforce them. Thus the judges dictated conformity to merchant practices, not the practices to which merchants should conform." We can find the same argument in W. Mitchell [69].

[5] It is in the 1990s when, especially due to scholars influenced by the law-and-economics approach, there was a boomlet of interest in the topic.

[6] Even if we talk about the 'social' aspect of law, in the present text we focus on the relationship between norms and single individuals, when the latter belong to a social group. No attention is given to norms that bear on the conduct of organizations of individuals, such as customary international law. The dynamics of organizational behavior may differ, of course, from the dynamics of individual behavior and those of individual behavior influenced by the feeling of membership to a social group.

[7] See, for example, the definition provided by M. Hechter and K. D. Opp in their volume entirely dedicated to the study of social norms: "[n]orms are cultural phenomena that prescribe and proscribe behavior in specific circumstances." [47, page xi].

does presuppose a regularity of conduct in the relevant group, but the latter element is not sufficient to constitute the former. Just as certain is the fact that custom can be a source of law.

Not all our habits and customs are based on rule-guided deliberations; not all habits become customary rules, and not all customary rules come from habits[8]. The justification of the normativity of customs is the outcome of a process in which individuals are aware of having an obligation to act in that particular way and act accordingly, either out of a belief in the importance of the rule or as the result of social pressure [56, page 157]. In addition to this, a rule is a social norm of a group only if non-conformity to the rule is met by a degree of adverse reaction from other members of the group. Therefore, it requires conscious following of the rule and, furthermore, critical reaction to departure from it. This is what gives it normative force in that society.

Enforceability and internal point of view

Legal norms are defined as normative rules of conduct prescribed by an authority invested with legal legitimacy. Their connection with social norms comes into sight when we study the emergence, adoption and compliance of norms not from the external point of view of the authority, but from the internal point of view of the agent[9], whose choice constitutes the last word in the implementation of normative standards.

In both cases, the adoption or rejection of those standards depends on the individual choice of the agent, based on his own expectations, beliefs, priors and preferences[10]. However, social norms do not necessarily imply enforceability in the sense in which legal norms do. They "allow for a variety of *social* mechanisms which induce norm-following behaviour and promote autonomous acceptance" [83, page 11], such as spread of reputation, social monitoring, normative influencing and rights and entitlements.

Furthermore, in order to be able to adopt a flexible approach towards normative standards, individuals must be endowed with "mechanisms for recognizing, representing, accepting norms and for solving possible conflicts among them." [83]. However, the observance by the subject of a norm as a mental object with clear, determined features is only possible if the norm remains in use for a minimum amount of time. This prevalence does not necessarily imply, though, an anchylosis of the norm, but rather something akin to -at least temporal-stability.

Stability and evolution of norms

Traditionally, at least two reasons have been considered essential for the stability of a norm. *Firstly,* a large part of the socialization process is the definition, for the individual, of prevailing social norms; since these habits or customs generally lead to a relatively seamless integration into a society or social group, few refuse to adopt them. *Secondly,* even if a particular instance of normative behavior goes against an individual's instincts or previous

[8] As Ibbetson explains, '[i]t is easy [...] to say that habit represented a factual regularity while custom was normative. This is true, but it does not explain how habit became custom, nor does it explain why custom is normative." [56, page 156].

[9] For a broader discussion on this issue, we refer the reader once more to Sartor et al. [83, page 11]. We must also note that the emphasis on this 'internal' point of view in the legal scholarship gained its force with Hart's *The Concept of Law* [45].

[10] [96] made one of the first attempts at explaining norms and norm compliance in a rational choice framework.

habits, the weight of social pressure may induce that person, nevertheless, to follow the norm. Many issues of law are about whether the addresses actions "have conformed to expectations which other parties had reasonably formed because they corresponded to the practices on which the everyday conduct of the members of the group was based." [46, page 96].

This leads us to an open question: the study of social norms and customs has led to deeper insights into the ultimate intrinsic forces behind human normative behavior and yet, "[t]he existence of social norms is one of the big unsolved problems in social cognitive science." [35, page 185]. We still know very little about how social norms are formed, the forces determining their content, and the cognitive and emotional requirements that enable a species to establish and enforce those social norms.

Issues regarding the emergence of social norms, their stability over time and their evolution are indeed questions that remain open in legal literature. In addition, the outcomes of evolutionary models of social norms are extremely sensitive to the postulated rules of transmission[11]. How is culture (and with it the social norms that evolve around culture) 'transmitted'? There are contrasting opinions in the scholarship about this topic. Here, we will not dig deeper into the subject, but rather just remember the reader what Goethe once said, "a tradition cannot be inherited —it has to be earned."

A positive 'double-feedback loop'

Unquestionably, the intrinsic nature of social norms involves change. Its link to the life of the community makes it the product of a continuously changing process, consisting in the aggregation of in-group decisions over time. However, in this accumulation of responses to everyday situations and conflicts, are we more likely to accumulate right or wrong answers? Is there a way to change the path, in cases where the wrong one is taken, or are previous choices inevitably binding?

The mechanisms by which some customs are selected and others disappear are rather obscure. The choice might not necessarily or consciously be outcome-oriented and, indeed, the 'socially optimal' solution is not always the one to survive over time[12]. Many factors affect the selection process, and understanding the causes and the course of action of that complicated selection procedure might indeed be more important than the actual outcome.

Customary law as such may be a function of institutional influence on individual and group preferences and behavioral patterns; it may as well be the end product of background facts and social norms, or the solution to coordination problems as discussed by [87, page 22]. Some scholars treat the development of customary rules as the fruit of some sort of 'social evolution', dealing with "populations of [replicating] entities, including customs and social institutions that compete for scarce resources." — see the arguments exposed by Hodgson and Kundsen [50, pages 477–486]. Other scholars consider that customs emerge as solutions to adaptive problems.

Furthermore, customary law, as opposed to civil law, is not equipped with mechanisms of systematization or codification[13], or any order which may assist in eliminating internal sources of inconsistency or conflict[14]. Despite this unsystematization, social and customary

[11] "[S]ince there is no firm basis on which to choose the rules, almost anything is possible." [38, page 73].

[12] For further discussion on path-dependance and persistence of bad choices recall, for example, [40, page 217].

[13] See Berman [11, page 328], who points out that manorial law "lacked the high degree of logical coherence and the consciously principled character of canon law."

[14] An argument in the same direction is proposed by F. [87, page 30]: "the existing array of entrenched

norms must be incorporated into the bigger picture of an integrated, coherent 'body' of law; with the additional difficulty that, although the *corpus iuris* is also flexible and relatively receptive to changes in the 'environment'[15], customary and social norms are driven by a different dynamic. This leads to an on-going tension between the different dynamics and sources of legitimacy of positive and customary law and the necessity for a sustainable and stable equilibrium.

However, this apparent 'chaos' that characterizes social and customary norms has a positive side: a higher grade of flexibility and adaptability[16]. Furthermore, insofar social and customary norms are 'closer' to what their addressees consider legitimate, to their 'internal point of view', norm-abidance is usually easier to achieve[17]. The existence of this 'double-feedback loop' between custom and culture reinforces the positive dynamics between them: "[f]ar from being a minor adjunct to the law properly so called, custom is [...] one essential component of any legal system, sufficient to sustain a rule of law under some circumstances, and one essential component of the rule of law under any and every circumstance." [75, page 100].

Notably, when the deviant behavior is 'naturally' agreed upon[18], according to the prevailing distribution of values in the community, the social forces will find it necessary to severely discourage any undesired conduct. In the case of social norms, we can consider that these attitudes of endorsement or rejection towards a particular action form part of the cognitive map through which the members of the group interpret their environment. An interpretation that is consensually constructed and commonly shared; consequently, the implementation of the enforcement mechanisms' is smoother than it would otherwise be.

An example: smoking ban in pubs

As we have explained in previous sections, when dealing with norm compliance, we can talk about two different layers or perspectives. In the first place, those addressed by the law shall consider their potential legal liability in case they don't abide the norm. Monetary sanctions, incentives and other deterrence measures are essential. However, there is also a second aspect, the morally persuasive role that is inherent to law. People obey the law because 'it's the right thing to do'. Sometimes the 'threat' of punishment is not the main reason for conformity, and non-pecuniary considerations must be taken into account.

Individuals belong to communities, which share a general respect for a system of values and norms. Norms are thus seen as a reflection of the set of behaviors that are seen as desirable of legitimate in the shared view of societal member, and whose violation elicits at least informal disapproval. This socio-cultural aspect of law has deep implications in terms of policy: instilling a norm in countries where it currently does not prevail and is culturally accepted may be a daunting task. People in such countries may find its content incompatible with the social environment in which they live, and therefore unfeasible to act accordingly.

normative customs is so complex, so large and so unsystematized that mutually exclusive customs abound."

[15] Here 'environment' is understood as the aggregate of social and cultural conditions that influence the life of individual in the community

[16] For the opposite argument, cf. Hart [45, page 89], who considers that custom tends to be static and inefficient and that, given that customary law is not under anyone's rational control, it cannot serve policy making ends.

[17] "What is more, written law will have no purchase on a community, unless it reflects the practices of that community in some way; even a law that sets out to correct custom will necessarily reflect other aspects of the customary practices of a community, or it will lack purchase in the community for which it is intended." [75, page 100].

[18] 'Naturally' as opposed to 'artificially' imposed by a legal authority.

In the case of smoking ban in pubs, the latter seems of extraordinary impact. The enforcement of such a ban is highly ineffective, and its compliance depends therefore on the underlying systems of norms and values that is embedded in that group's culture. Indeed, smoking habits are deeply rooted in greater cultural traits, such as social structures, gender roles, autonomy, authority considerations or distribution of power, to mention some of them.

Concern for risks related to smoking has raised considerably during the last decade. Not only for smokers, but also for the ones surrounding them, due to 'environmental tobacco smoke' (ETS, also called 'second-hand' smoke or 'passive smoke'). Hence, it does not seem implausible to believe that a majority of EU citizens support smoke-free public places, such as offices, restaurants and bars. This fact would be consistent with broader findings in sociological literature, that individuals exhibit similar value preference rankings across cultures. But even if this is generally true, there are also cultural differences that lead to cross-country specificities in citizen's reactions to the smoking ban in pubs. Let us take for example, three sample countries: Italy, Portugal and The Netherlands.

(a) *Italy*: is considered *par excellence* the European country in which informal ties and social relations are fundamental. This would lead to a higher compliance with the ban, due to informal enforcement mechanisms and social control. Stigmatization of smokers is higher and has further consequences. Furthermore, we believe, though restaurants and bars are an important part of cultural life, they are easily replaced. The culture of 'open-air' events is also large, due obviously to the warm Mediterranean weather, and many social meetings take place around close family and friends at private locations, making the displacement of smoking at home easier, and other adverse social consequences such as the increase in the noise in the streets not so relevant.

(b) *The Netherlands*: should present radically different reactions to the ban. Dutch citizens are considered more independent from their social ties. The informal enforcement mechanisms of the ban are hence deemed to be effective in a lesser extent. A second aspect that could be considered relevant, is that a great part of the socialization process takes place in pubs and bars. Citizens should be highly opposed to the ban, which would clearly diminish social and cultural exchange, as well as harming values such as individual freedom or autonomy.

(c) *Portugal*: represents somehow an intermediate example. In addition, the ban in Portugal was implemented in different, less radical terms. Bars were allowed to create separate spaces for smoking and non smoking in case they were bigger than 70 square meters. For bars smaller than that, it was up to the owner to decide whether to constitute the place as either smoking or non-smoking environment. This should have facilitated the transition, but it could have lead, however, to a majority of the bars becoming 'smoking allowed'.

As yet there is little published evidence that we can use to verify these predictions, though what there is suggests that they agree with what has happened. [14], for example, suggests that in Italy the ban has led to a big reduction in smoking in bars and restaurants without the need for much enforcement. There is less direct evidence for the effect of the ban in Portugal, but [64] describes a small-scale study on Portuguese restaurant workers that shows a reduction in exposure to ETS following the ban, suggesting an appreciable degree of compliance. In the case of the Netherlands, the only data we are aware of is from [71], which suggests that there is a smaller increase in support for smoking bans than in other countries.

4 Trust and the Dynamics of Norms

In this section we discuss the relationship between trust among a group of agents and the dynamics of norms within that group. We will argue that trust is essential for the adoption of a norm, and describe how we believe that trust fits into each stage of the norm lifecycle. We start by offering an overview of definitions of trust, before advancing our own position.

4.1 What is trust?

Trust is a concept that is both complex and rather difficult to pin down precisely. As a result of this slipperiness, there are a number of different definitions of trust in the literature. Sztompka [94], for example, suggests that "Trust is a bet about the future contingent actions of others", while Mcknight and Chervany [65], drawing on a range of existing definitions, offer the definition "Trust is the extent to which one party is willing to depend on something or somebody in a given situation with a feeling of relative security, even though negative consequences are possible." Gambetta [41] states that "Trust is the subjective probability by which an individual, A, expects that another individual, B, performs a given action on which its welfare depends", while Mui *et al.* [74] define trust as " a subjective expectation an agent has about another's future behavior based on the history of their encounters."

As the alert reader will have noticed, these definitions, though different, do overlap somewhat. All four definitions given here focus on trust as a mechanism for making predictions about the future actions of individuals. That is if one individual trusts another, then that first individual can make a (more or less) accurate prediction about what the other will do in the future. One might, as [41] does explicitly and [94], does implicitly[19], decide that trust can be quantified as a probability. Or one might, as Castelfranchi and Falcone [22] argue, decide that trust is more complex (and in the case of [22], decide that trust has a rational basis in reasons for beliefs about the future actions of others). This distinction need not concern us here — for our purposes it suffices to think of trust as an abstraction, a summary perhaps of some complex pattern of reasoning, for some model of how others will behave.

4.2 Trust in the norm lifecyle

The fact that trust allows agents to predict the behavior of other agents is exactly why trust, or the lack of it, plays an important part in the norm lifecycle in Figure 1. To see why this is the case, consider the spread of norms. Once a tentative norm has been generated, it will either spread through a population and eventually stabilize, or it will be discarded. Which happens depends on individuals' attitudes to the norms and their ability to predict each other's behavior, and this latter depends on the trust between individuals. As Bicchieri [12] argues, "a social norm depends for its existence on a cluster of expectations. Expectations, . . . , play a crucial role in sustaining a norm". Trust is important because of the way it helps agents to handle those expectations[20].

[19] Subjective probability having a natural interpretation as a propensity to make bets at particular odds [57, 655].

[20] One can also argue the reverse, that because norms exist, it is possible for agents to develop expectations based on their beliefs that others will follow the norms, and hence agents will trust others. While this is clearly the case, we believe that the cycle that takes observed behavior, infers trust from it, and then develops norms as a result of the trust is more general, allowing the establishment of trust and then norms in the absence of any existing norms. In other words it plays a role in norm generation as well as

Consider the norms that govern how people wait for and then board a bus. There are at least three distinct ways in which this is done. In some populations, people who arrive at a bus-stop form a queue. When the bus arrives, people board the bus in the order of the queue. In other populations, people who arrive at the bus-stop do not form a queue, but each remembers the people who were there before them, and when the bus arrives, people do not board until everyone who was at the stop before them has boarded. In yet a third population, people arriving at the bus-stop don't form a queue, and when the bus arrives, everyone boards the bus as quickly as they can with no regard for the order in which people arrived at the bus-stop. (Of course, there are populations in which some mixture of these behaviors co-exist, but we will consider simpler cases for now).

Now, consider how someone reacts when they arrive at a bus-stop where they don't know what the norm is. A typical reaction is to try to infer the norm by observation. For example, if there is a queue, take this as an indication that people will board in the order of the queue, so the right thing to do is to join the queue and let the people in front board first. In doing so, the new arrival at the bus-stop is trusting other bus riders to do the same. There is risk in this. By waiting when the bus arrives, our rider is both allowing later arrivals who don't believe in queuing to push ahead, and denying themself the opportunity to push ahead of others. However, if the queue operates as it should, our bus rider is ensuring that while they can't exploit others by pushing ahead, they will board before later arrivals, and they will not suffer whatever sanctions the bus queue imposes on the violators of norms (which in the authors' experience ranges from sharp intakes of breath and disapproving looks to loud lectures on appropriate behavior).

What we have here is an example of how trust allows norms to operate. If everyone trusts that everyone else will follow a norm[21], and in the face of a population that observes a norm and enforces there is no net benefit in violating the norm, then the norm will be stable. Of course, if there is no bad consequence of violating the norm (for example, the rest of the queue just ignores the violation), or there is no advantage in following the norm (somehow waiting in the queue means that our rider always ends up being the last person to get on the bus), then there is no incentive to following the norm and it will be discarded.

To take this example a little further, we can also see how trust is involved in the spread of a norm. Say we have a population of bus-queuers in a country with small bus shelters and a high likelihood of rain (any similarity with the situation in the United Kingdom is entirely coincidental). In such a situation, there is a disadvantage to queuing if one is not one of the first in the queue — one has to stand in the rain on rainy days because the queue extends beyond the shelter. Imagine that on such a rainy day, at one bus stop, someone who would, under the prevailing norm, have to take their place at the back of the queue, outside the shelter, decides instead to stand somewhere dry (under the awning of a nearby building), but when the bus arrives insists on boarding in order of their arrival. If they are not sanctioned (and why should they be — they may have violated the standing-in-line norm, but they haven't violated the first-in first-out spirit of the queue), then it is possible that their action will be remembered by those damp members of the queue who get to board later, and recognized as superior. These damp queuers may then copy this behavior at other bus-stops, so spreading this new norm through the population. Of course, this happening

norm spreading.

[21] Or, more correctly as we shall see below, as long as a population trusts that a sufficiently large fraction of the population will follow the norm. For the perspective that trust is based on a normative expectation see also [58].

requires trust on the part of the old-norm violators that they will not be sanctioned, and that they will still be able to board the bus in the order in which they arrived at the bus-stop. If that trust is not possible, then the new norm will never become established.

4.3 How trust develops, how trust decays

Given the essential role of trust in ensuring that norms spread and stabilize, it is worth thinking a little about how trust develops between agents. As pointed out above in the quotation from [74], trust between individuals develops as a result of the history of interactions between them. When those individuals are representative members of a population, then the trust develops as a result of all the interactions between members of those populations. The development of trust is therefore dependent on the ability of the individuals to learn, and on the repeated nature of their interactions.

One well-cited example of the development of trust even between agents that might be thought to have opposing interests is the story of the unofficial truce[22] that took place around Christmas 1914 between British and German troops (and, to a lesser extent between German and French troops) stationed in the trenches along the Western front. The truce was sufficiently widespread that a number of football games took place between groups of British and German soldiers. As Axelrod [7] points out in his description of this incident, at the time that this truce developed, units were stationed opposite each other for relatively long periods. This meant that these soldiers had a chance to establish some form of relationship with their opponents and to learn that these opponents would not take advantage of friendly actions. For example, Weintraub [100] reports that both sides had been taking a ‘live and let live’ attitude for some period before Christmas, not firing on each other during mealtimes, and observing ceasefires when both sides could retrieve the dead and injured from the no-mans land between the trenches. Here we see the importance of repetition in the development of trust.

Axelrod also gives a nice example [7] of how individuals can be representative of populations, reporting a case in which one group of Saxon soldiers from the German army apologized to the British unit they were facing for the fact that the British were shelled by a different unit:

> . . . a brave German got up on to his parapet and shouted out “We are very sorry about that; we hope noone was hurt. It is not our fault, it is that damned Prussian artilary.” (the words of an unnamed British officer quoted in [80, page 34] which in turn is cited by [7, page 84].)

Clearly the Saxons wished to distinguish themselves from the Prussians in order not to undermine the trust the Saxons had established with the British.

It is not clear how much this case represents the development of trust, and from it the adoption of a new norm, in the few months between the start of trench warfare on the Western front in September 1914 and Christmas, and how much it represents a continuation of an existing norm for opposing sides in a conflict to observe ceasefires on humanitarian grounds. Figes [36], for example, describes regular armistices at the siege of Sevastopol in the Crimean War, while Weintraub [100] points out that such incidents go back at least as far as the siege of Troy. However, it is clear that the official response was to outlaw further

[22] So unofficial that commanders actively tried to stop it happening.

unofficial truces and that this and mandatory raids on enemy lines[23] led to the steady decay of this norm — while there seem to have been further incidents during World War I, they were much less widespread. In this we see the importance of learning. Trust was adjusted as a result of experiences in the repeated interactions.

In many situations, individuals do not have the chance to develop trust in others, either at the individual level, or the population level. In Axelrod's [7] term, there is no 'shadow of the future' in such scenarios — if two individuals are going to interact just once, there is no way for one to repay the other for a kindness (not shooting during a mealtime to extend our military example), and so no motivation for the other to act kindly.

As Resnick *et al.* [78] argue, it is exactly in such situations that reputation systems come to the fore. By providing a mechanism for recording the outcome of one-off interactions, reputation mechanisms create a shadow of the future. Now the kindness or otherwise of two individuals who interact just once will be recorded, and when one of them interacts with a third, that third individual can respond to the outcome of the previous interaction. (In the trench warfare case, the reputation of the local enemy units was passed along as troops rotated in and out of a sector.)

Of course, reputation systems are not infallible. [77] describes a problem with the reputation system on eBay (a system that in general performs rather well), in that in the empirical analysis in [77], there is little negative feedback of sellers. This in turn leads to a rather over-optimistic estimate, and means that reputation isn't the best predictor. Furthermore, as [62] shows empirically, reputation systems can be manipulated, and [82] backs this up with an theoretical analysis of how reputation systems can be manipulated.

Finally, we should note that while the reputation of an individual, and hence trust in that individual, is most strongly affected by the behavior of that individual, it can be affected by others as famously argued by Akerloff [2]. Akerloff studies the 'market for lemons'[24], a market in which sellers offer indistinguishable goods some of which are of poor quality. In such a market, trust in all sellers is decreased by the presence of these poor quality goods — buyers translate the risk of loss through purchasing a lemon into a loss of trust in every seller, even though they have no personal experience of lemon purchase.

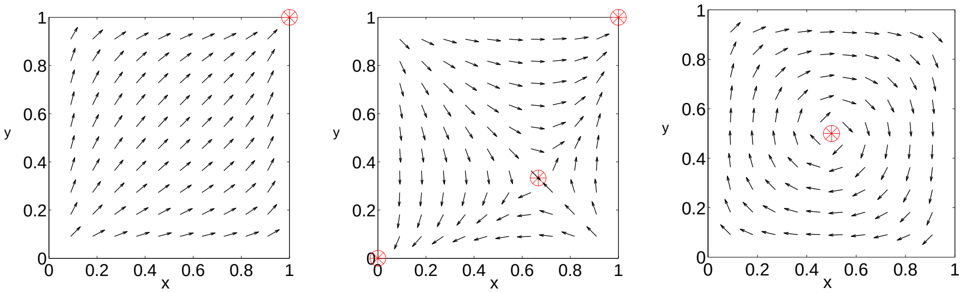
4.4 Implicit trust

Several of the authors (for example in [21, 95]) have developed formal models that can be used to explicitly represent the trust that one individual has in another. Indeed, the model from [21] is widely adopted in the multi-agent systems community and has been adapted [34] and extended [15] by a number of authors. However, while it is possible, and we believe useful, to build explicit models of trust, it is not necessary. Trust can be implicit.

Indeed we can consider several levels of implicit trust. When it uses an explicit model of trust, agent A considers the trustworthiness of agent B, and makes decisions based on that trustworthiness. One form of implicit trust is when A, rather than modelling B, just reasons about what action B will take, perhaps on the basis of B's past behavior. If A learns to predict B's behavior in interactions using fictitious play [19, 39], in which A takes B's past behavior to be a representative sample of its full range of behaviors, then the resulting probability distribution over actions can be used to help A determine what to do (which is

[23] Which, according to Axelrod [7], were not intended to reduce 'live and let live' but ended up doing so.

[24] In the US, the term 'lemon' is used to refer to a badly constructed car with many faults, and many states have a 'Lemon Law' which provides financial assistance to car buyers who have been sold a such car.



(a) Prisoner's Dilemma **(b)** Battle of the Sexes **(c)** Matching pennies

Figure 2 Replicator dynamics for three two player, two strategy games under frequency-adjusted Q-learning. Figures by kind permission of Michael Kaisers.

functionally what trust does) without A having to invoke any notion of trusting B. In this kind of scenario, as long as the agents have converged to equilibrium — so that their notion of what the other will do is no longer changing — it seems reasonable to consider that they trust one another since they have accurate predictions of what the other will do[25]. Fictitious play allows this to happen for some scenarios, for example , Miyazawa [70] showed that fictitious play converges in two-player, two strategy general sum games such as the iterated Prisoner's Dilemma, the scenario studied by Axelrod in the classic work cited above [7].

In fictitious play, A is still conscious of the choices made by B — that is what A models. A could also just concentrate on its own actions and learn what works best for it. If A does this, then since the outcomes of its actions depend to what B does, it is implicit that A learns to predict B's actions, and so very implicitly can establish trust in B. Kaisers discusses how Q-learning and its variants [59] can be used to do this learning, and shows how learning converges (and hence how A establishes trust in B). Figure 2 shows this convergence. The axes in each plot give the probability of the agents playing one of its strategies (and since there are only two strategies, by extension the probability of playing both strategies). Figure 2a shows the Prisoner's Dilemma, where the agents converge to the Nash equilibrium at $(1, 1)$[26], Figure 2b shows the Battle of the Sexes, where the agents converge to the Nash Equilibria at $(0, 0)$, $(1, 1)$ (the fixed point at $(\frac{2}{3}, \frac{1}{3})$ is not stable). Figure 2c shows the Matching Pennies game which has a fixed point at $(\frac{1}{2}, \frac{1}{2})$ to which agents do not necessarily converge[27].

4.5 Trust as a normative phenomenon

Finally, given that trust can develop and decay just like the norms in the norm lifecycle, it is worth pointing out that trust itself follows parts of the norm lifecycle. Again Figure 2 shows us how this might happen.

Above we interpreted the plots in Figure 2 as giving the probability of a specific pair of

[25] To quote BIcchieri [12] again, "norms of cooperation ... emerge as equilibria of learning dynamics in small-group interactions"

[26] One might, of course, argue that given the details of the scenario, this outcome does not represent one in which the agents trust each other, but in accurately predicting the action of the other agent, it meets our definition of trust.

[27] The fixed point only has Lyapunov stability in this particular case, so while points that are close to the fixed points will not diverge, they will not necessarily converge to the fixed point.

strategies being played. We can also interpret each point in one of the graphs as indicating the proportion of a population of agents that will pick a specific strategy. Under such an interpretation, the point $(\frac{1}{5}, \frac{3}{5})$ in Figure 2b captures what happens when, in a population of agents playing the Battle of the Sexes, 20% of the agents picking the row in the game matrix play B and 60% of the agents picking the column play B. The arrows on the plots show how the population changes its strategy from this mixture — in this case the arrow indicates that a population at $(\frac{1}{5}, \frac{3}{5})$ will change such that more agents picking the row will tend to play B, and fewer agents picking the column will tend to play B.

Under this interpretation, we can see the plots as showing trust, in terms of the ability of one agent to predict the behavior of another, spreading through the population. Look at Figure 2a, and consider a point near $(0, 0)$. Here, any given agent can be pretty sure that any other agent it encounters will play C, but it will also know that this will change over time — the choice of playing D will tend to appeal to any agent in the population so this ability to predict is not stable. The direction field tell us that the choice of playing D will tend to spread until the population reaches $(1, 1)$ and every agent knows for sure that any agent it interacts with will play D. Thus the Prisoner's Dilemma is an example where trust can spread through the population and become stable. In contrast, the fixed point at $(\frac{2}{3}, \frac{1}{3})$ in the Battle of the Sexes (Figure 2b) illustrates that it is possible for equilibria to be unstable. Without the equilibria at $(0, 0)$ and $(1, 1)$, the population playing this game would not be able to develop trust in one another, and the lack of an attracting fixed point in the Matching Pennies game is a further illustration of trust not becoming stable.

5 Culture for modifying norms' dynamics

Cultural differences in normative dynamics are considerable, as argued above. Group dynamics are an essential part of social interaction and are critical to the evolution of social norms, as already argued in Section 3.2. Relationships to others in a group context usually affect one's willingness to emulate, provoke, forgive, reproach, oppose, admire etc. the adoption of new social norms by other group members. [55] argues that recognition of the role of social relations will improve our understanding and our predictions, with regard to social norms. There is no universal mechanism for the transmission of social norms across all human cultures. It follows that at some stage in the development of NorMAS, we should tackle how cultural differences affect the dynamics of social norms. Indeed, the approaches of researchers to modelling NorMAS are undoubtedly influenced by their own cultures, and we should try to be aware of these biases when postulating mechanisms of norm transmission. An example of this comes from [8], who investigate beliefs about what it means to be human in various cultures. The dangers of ignoring culture in research into social processes has been emphasised by [49]. Members of different cultures are socialised to obey very different social norms and integrate very different value structures in their ways of living. They also have different propensities to intepret behaviour as normative.

We shall focus in this section on the differences in the processes of norm dynamics across cultures. Comparative quantitative studies of culture have usually been at the level of the nation state, due primarily to issues of data availability. While such analyses may miss many of the subtler nuances of culture, we believe that they are at a manageable level of abstraction for incorporation into the next generation of NorMAS. The empirically derived Hofstede Dimensions of Culture [52, 53], shall form the basis of our discussion of culture in NorMAS. Another cross national analysis of culture has been conducted by [88], who analysed self-reported values to elicit dimensions of values. This approach contrast with the

Table 1 Dimensions of Culture.

IND	INDividualism
PDI	Power Distance Index
MAS	MASculinity
UAI	Uncertainty Avoidance Index
LTO	Long Term Orientation
IVR	Indulgence Versus Restraint

Hofstede theory, which was derived from questions about everyday practices.

We begin with a brief introduction to the Hofstede Dimensions of Culture. We then discuss the preceding sections of the chapter from the point of view of the culture theory outlined. In the final part of this section we discuss how some of these differences can be included in the next generation of NorMAS.

5.1 A Brief Introduction to Hofstede Culture Theory

Hofstede and co-workers conceptualize culture as a limited number of major societal issues, to each of which a society finds a shared solution. These issues are conceptualized as continua, as scales with a lower and an upper end. [52, 53] call these dimensions and they describe six of them that vary across nationalities, see Table 1. These dimensions have been shown to correlate with a wide range of empirical data in the social sciences [52], notably marketing data [72].

The Hofstede model is based on questions about everyday work practices; the dimensions of values were a serendipitous finding. They refer not to convictions or beliefs but to broad tendencies to perceive the social world in a certain way. The model has grown over time, as more sources of data were consulted. The dimensions were derived using factor analysis of survey data.The latest model [52, 53] consists of six dimensions. Each of them is modelled as a continuum running along a scale from 0 to 100. This means that almost all actual values will share characteristics of both extremes.

Here they are, together with some explanation and also an impression of the perceptual capacities that agents need to have in order to accommodate the dimension in a model:

Identity: individualism versus collectivism. (IND) Essentially this is the extent to which members of a society feel responsible for themselves, or for the larger group they belong to. In the first case, rights and obligations should be the same for all people, while in the second, the boundary of the in-group is also a moral boundary beyond which rights and obligations do not hold. Agents in collectivistic cultures will act differently depending on whether other agents are in group members or not.

Hierarchy: large power distance versus small power distance. (PDI) That is the extent to which the less powerful members of a society expect and accept that power and rights are distributed unequally. Large PDI divides a society depending on positions within that society, i.e. there is stratification of social groups based on status, those with low and high positions do not mix. Agents in cultures of large power distance will respond differently to others depending on how they perceive their status to be relative to their own.

Aggression and gender: masculinity versus femininity. (MAS) This dimension is about assertive dominance and emotional gender roles. It contrasts a strong-handed, competitive orientation in 'masculine' cultures, in which people in general cannot be assumed to be trustworthy, men are supposed to be tough and women subservient and tender, versus a consensus-seeking and care-taking orientation for both women and men in 'feminine' cultures. In masculine cultures, it is desirable that agents be gendered and recognise gender.

Otherness and truth: uncertainty avoidance versus uncertainty tolerance. (UAV) In uncertainty avoiding societies, anxiety levels are high. In defence against it, strict rules, rituals, and taboos govern life. Distinctions between categories should be sharp and the unknown is considered dangerous. Out-group members and institutions will not be trusted by agents from such cultures. In uncertainty tolerant cultures, relaxation is the rule and actions are results-based rather than anxiety based. The level of fixity of all kinds of rituals goes up with uncertainty avoidance.

Immutability vs. pragmatism: short-term versus long-term orientation. (LTO) In short-term oriented societies immediate gratification of needs and keeping up social appearance, behaving well and respecting tradition are seen as virtues. In long-term oriented societies, reasoning is pragmatic and planning, foresight and perseverance are valued. Persons from such societies are more likely to learn from experience and to change their norms accordingly.

Gratification of drives: Indulgence vs. Restraint. (IVR) Indulgence stands for a society where people feel happy and healthy, and like to enjoy life and to spend time with friends, but could also slip into violence if they feel like it. Restraint stands for a society in which people feel the burden of duties heavily, and keep both positive and negative impulses in check. This dimension will heavily impact the exuberance of greetings.

Note that the six dimensions are not personality traits, but societal patterns! This means that, unlike personality traits, they will be shared by the persons with the same cultural background. Yet culture, as an unconscious set of basic values, should not be confused with conscious group affiliation.

5.2 A Discussion of Culture in the Topics Raised in Previous Sections

Three definitions of the polysemic term 'norm' are given in the introduction to this chapter, Section 1. The first, which is simply that which is most normal or most usual need not detain us here. The second is norm as an ideal. The ideals of how to behave as social actors vary significantly across cultures. In some, the more long term oriented, being pragmatic and reasonable is considered ideal, in others, the more short-term oriented to be strong, resolute and unyielding in one's convictions is more admired. The ideals of a culture, as exemplified by the heros of that culture, vary with each of the dimensions of culture, in ways that can be straightforwardly inferred from the brief description of each dimension above. The third definition of norm that was discussed is that of an "imperative regarding the behavior of social actors". These imperatives, or the resultant tendencies to behave in certain ways are, along with the manner of perceiving the world, strongly moulded by culture. However, culture is more than an exhaustive list of norms and practices, it is the deep structure embedded within these norms and practices, part of which is reflected in the six Hofstede dimensions of culture.

Section 2 discusses the cognitive architecture required to model normative processes in multi-agent systems. It forms part of the universal basis for social behaviour on the part of agents, which in the real world is modified by culture. How best to get a norm adopted in the mind of another is discussed in Section 2, and how culture can change this is discussed as the end of Section 3.3.2, with regard to real experiences of the smoking ban. It is postulated quite reasonably that the adoption of a new goal (such as a goal to observe a norm) is conditioned on the new goal being as means, in some manner, of achieving a pre-existing goal. These pre-existing goals will almost always be influenced by the culturally derived disposition to value certain states of the world over others. Individuals can reject the overt values of their society, but the traces of their culture of upbringing always remain in their way of perceiving and interpreting their social world.

Among the reasons given for motivating the adoption of a norm, in Section 2, are instrumental, cooperative and terminal reasons. Cooperative reasons apply when the norm adopter holds the values that are supported by the norm. Another more social reason to adopt the norm is simply as it is the done thing in the social group one belongs to. This comes close to envisaging the desire for compliance to the group, or being well respected by certain individuals, as a terminal reason to adopt certain norms. We may have the goal that the norms of a certain group should be respected, but which group, and why? The subtle processes of group dynamics can be influential in these decisions, decision that are not always fully conscious. To refer this to the dimensions of culture, the desire to simply fit in is strongly influenced by collectivism (or low individualism) and Restraint.

Section 4 discusses the role of trust in norm dynamics, and it is clearly a critical element for any functioning human society. Indeed, trust is so basic that it cuts across cultures. It's an essential part of social cooperation and no culture could survive without it, though how cultures manage, preserve, enforce and nurture trust, vary greatly. In low power distance, feminine cultures people are expect to be good without the presence of tough rules and supervision. In high power distance, masculine cultures that are also restrained it is usually thought very important to enforce norms strictly or else they shall not be obeyed.

Section 3.1 concerns norm generation, and shows a number of processes proposed to emulate the real world emergence of new norms. This section highlights the admittedly large gap that separates the complexity of the evolution of norm in human societies and current methods for the run-time evolution of norms. This raises the question of whether the time is right to start considering real world cultural variations in NorMAS, and for some of the models discussed this is clearly not appropriate. However, as mentioned in the introduction to this section researchers beliefs about what norms are, and how they work, is in part a product of their culture, and a broader view of these processes is likely to be helpful in the development of successful NorMAS. Also, when the adoption of a new norm is considered in different cultures, as discussed above for the smoking ban in Europe, consistencies within (and differences between) cultures related to the interpretation of behaviour as normative, the manner of enforcing norms and the readiness to adopt certain types of norms can be incorporated in currently feasible NorMAS.

Section 3.2 makes the case that observed behaviour can be interpreted normatively. If an agent sees many people perform an action that agent can interpret that this is more than a simple empirical fact, but something that can carry normative message. Here are some ways in which this process is likely to vary with national culture. In societies strongly stratified by hierarchy (large power distance), the behaviour of others of differing status is rarely a reliable guide to the social norms prevailing for those of one's own status. For example, a student may think they should use another staircase to that used by a professor. The dimension of Indulgence versus Restraint is linked to how ready people are to infer social norms from behaviour. Restraint cultures have many strict norms and are primed to interpret new behaviours in that light. Indulgent cultures are much freer in their behaviour and are more likely to see new behaviours as the result of personal preferences than normative obligation. Finally the level of individualism, as opposed to collectivism, is likely to affect the interpretation of observed behaviour. In individualist cultures personal preferences are more freely followed, while in collectivist cultures conforming the one group's behaviour and outlook is a much stronger force.

Section 3.2 also discusses obedience and violation. Notably, masculine societies feature more punishment for transgressions, while feminine societies have higher levels of forgiveness for transgressions. Restrained and collectivist cultures are likely to see far high base levels of obedience to a larger number of norms while in indulgent and individualist cultures people are much less constrained in appropriate social behaviours they may adopt. Indeed, in such cultures 'following your own star is admired', while in others obedience to social demands is admired [10].

The two-way relationship between formal laws and informal social norms is examined in Section 3.3, where the role of culture and the importance of modeling humans as actors within groups are referred to on a number of occasions. One word of caution, which is appropriate here and also applies to the examples and statements in this Section, is the possibility to read too much into the determining role of culture in any given example. The confirmation bias is an inherent part of human reasoning, and seeing a plausible explanation does not mean that it is *the* explanation[28]. This also applies to different interpretations of the role of culture in a particular phenomena, such as the varied success of the smoking ban in bars in different European countries, and well as the many institutional influences of the evolution of this particular norm. One of the major problems in getting a handle on culture in a model is its sheer pervasiveness. It's even present in how pervasive we consider culture to be. Culture is present in the smoking bar example in the institutions that created the laws and regulated their implementation. It is also present in the subtle interactions by which individuals test the the prevailing norms, and accept, reject or attempt to influence them.

5.3 Culture in the Next Generation of NorMAS

We outline some essential elements of group dynamics before discussing how these may affect norm transmission, and then how well-understood cultural variations could influence this process.

Intra-group Behaviour is critical to the transmission of social norms, and it varies considerably across cultures. The idea that people learn differently from others depending on who they are is called selective attention by [48]. When explaining why norms are enforced in some situations and not others, social context matters [55]. Some elements of human social behaviour can reasonably be assumed to be universal, such as the role of trust, see Section 4. Most social behaviour is anchored in a group context, where the relationships between individuals, and their relationships to the wider group, matter. Section 3.1 already mentions the importance of societal typology. Social norms are first and foremost a mechanism for regulating behaviour in groups. We now outline how culture can begin to be implemented in NorMAS to modify these behaviours.

5.3.1 Suggested Group Level Primitives

Each agent in the NorMAS we envisage would have a perception of each group it belongs to, including the membership of that group, the rules of behaviour appropriate to that group, and both its own and others position in the group. What do we mean by position? In order

[28] [66] argue that the fruitful clash of arguments which results from this universal bias is the main evolutionary reason for what would otherwise be a straightforward flaw in human reasoning.

to create implementable computational models this needs to be broken down into a limited set of primitives that agents can understand and reason over when making decisions. In the following we introduce one possible set of such primitives.

One of the first elements agents should look to as a guide to social behaviour is hierarchical status. This usually means age, often role (e.g. employer/employee), and sometimes gender. Hierarchical status is important in determining who can take initiative, who should look to whom for leadership, whose social cues should be followed etc..

Another important element is reputation. Who is a good member of the group, and whose behaviour is a reliable guide to the right thing to do in social situations. This is a critical element for imitation behaviour, and hence the spread of social norms. Agents will be more likely to emulate those they admire, i.e. those with a higher reputation, and they may change their criteria for judging behaviour (their social norms) if those with a currently high reputation change their behaviour. Reputation within a group is hence a critical element for altering the social norms of that group.

In addition to hierarchical status and reputation, another element that can be central to social behaviour is the relative importance of other group members. We shall refer to this as centrality, though a related idea has been named interdependence in the literature [55]. It echoes the idea of a central node in network theory, but here mean how much another individual matters in a more general sense. The importance of close family members is not necessarily derived from their position in a network. This primitive of centrality can influence the weight given to sanctions or messages from other group members. A sanction from a very central person is likely to be more harder to ignore than one from a peripheral group member.

These three elements of internal group structure can all be important in the spread of social norms, and can all operate differently across cultures.

5.3.2 Culture and Intra-group Behaviour

We hold that the empirically derived Hofstede Dimensions of Culture are at the correct level of abstraction for incorporation into the next generation of NorMAS. The Hofstede theory was derived from questions about everyday practices rather than self-reported values as in the Schwarz system of values [88].

We shall here discuss how three of the Hofstede dimensions may affect the operation of the three group primitives introduced above in the transmission of social norms. The three dimensions will shall focus on are Individualism, Masculinity and Power Distance. The full list of dimensions can be seen in Table 1, and are described in detail in [52, 53].

Power distance determines the degree of importance of hierarchical status in social behaviour. It influences the extent to which the less powerful members of a society expect and accept that power and rights are distributed unequally. In large power distance societies, high status members can disobey social norms with little resistance from those of lower status. Also, the behaviour of others of different status is not a good guide to appropriate behaviour for oneself. In low power distance societies social norms apply much more equally across the population.

Individualism is the degree to which members of a society feel responsible for themselves, or for the larger group they belong to. This dimension relates to the centrality primitive introduced above. Individualistic agents would grant more equal levels of centrality to other agents, while at the other end of the dimension, collectivist agents would be more discriminating in attributing higher centrality to others. It is harder to change one's centrality in collectivist societies where group boundaries are both more relevant and harder to pass.

The dimension of Masculinity represents, in part, the distinction between a strong-handed, competitive orientation, and a consensus-seeking and care-taking orientation in ‘feminine’ cultures. This would relate most directly to the reputation primitive, with masculine cultures having more unequal distributions of reputation, and social norm infringing acts having significant affects on reputation. In ‘feminine’ cultures reputation is more evenly assigned and not so easily lost, as such societies are more forgiving than their ‘masculine’ counterparts. Masculine societies feature more punishment for transgressions, while feminine societies have higher levels of forgiveness for transgressions.

We have argued that one important challenge for NorMAS is to incorporate social context into models of norm transmission. As a further extension, and as a warning against excessive generalisation of culture specific social contexts, we argue that cultural differences in social behaviour are important and that a first representation of some of these differences is possible in the next generation of NorMAS. As stated by Hollander and Wu [54] in their review of normative multi-agent systems, the degree to which relationships influence norm adoption is still under investigation with the social sciences. However, implementations of NorMAS that address this issue can help clarify the questions to be addressed by both social scientists and the NorMAS community.

6 Conclusions

A clear definition of a norm, or definitions of the types of norms that exist, is a necessary first step for a fruitful exploration of the question of norm dynamics. We discussed the definitions of three types of norms and applied these in the later sections of the chapter. The essential cognitive constructs required for agents to reason about norms have been introduced, along with references to more detailed discussions of this issue. The major point to note is that norms can only spread in a population if the relevant mental constructs (i.e., normative beliefs, goals and intentions) propagate as well through the minds of the individuals in that population. The reasons agents may adopt new norms are also surveyed.

This chapter presents a *Norm Lifecycle* that, although it is not identical for the different kinds of norms, starts with a *generation* phase. Briefly, norm generation corresponds to a process where new norms are proposed (either prescribed or emerged). Generated norms may then *spread* through a population and eventually reach *stability* (i.e., persistence) if they are adopted by enough individuals in the population. Afterwards, different evolution stages may occur: norms may *decay* being superceded by new norms; they may evolve; or alternatively, they may be *codified into law*. We interpret last two cases close the lifecycle, since resulting norms can be considered as norm candidates that are being generated. Thus, norm generation is a phase that can be instantiated through different methods. In this chapter we have reviewed a number of alternative approaches such as norm synthesis, agreement, emergence, or empirical learning. Strengths and limitations are discussed in terms of their complexity, required domain knowledge, or time to convergence.

The individual interactions by which norms can spread in a population have been discussed at some length. These include the norms implicitly communicated by practical actions, and more direct signals such as reproaches for the neglect of norms. The complex relationship of social norms with the law has also been elucidated. This is an unduly neglected issue in the legal domain, and one where the multi-timescale, simultaneously micro and macro approach possible with NorMAS is very promising.

We have analyzed the relationship between trust among a group of agents and the dynamics of norms within that group. An overview of definitions of trust was presented, before we advanced our own position. We hold that for the purpose of analyzing norm dynamics, it suffices to think of trust as an abstraction, a summary of some (perhaps complex) pattern of reasoning, for some model of how others will behave. Trust enables individuals to form expectations regarding the behaviour of others. We have argued that trust is essential for the adoption of a norm, and described how we believe that trust fits into each stage of the norm lifecycle.

We have finally examined how culture might be expected to effect the normative processes presented in this chapter and finally we sketched a method by which variations in human cultures can be incorporated in the next generation of NorMAS.

Acknowledgements

The authors would like to thank the reviewers for their helpful suggestions.

S. Parsons was partially funded by the National Science Foundation, under grant CNS 1117761 and Army Research Laboratory and Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document should not be interpreted as representing the official policies of those organisations or the U.S. Government.

E. Mayor benefits from a postdoctoral grant funded by the Thematic Network for Advanced Research "Sciences & Technologies for Aeronautics and Space"(RTRA STAE Foundation) in Toulouse, France, under the MAELIA project.

References

1 T. Ågotnes, W. van der Hoek, and M. Wooldridge. Robust normative systems. In Padgham, Parkes, Muller, and Parsons, editors, *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems*, pages 747–754, Estoril, Portugal, May 2008. IFAMAAS/ACM DL.

2 G. A. Akerlof. The market for 'lemons': Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500, August 1970.

3 G. Andrighetto, M. Campennì, F. Cecconi, and R. Conte. The complex loop of norm emergence: A simulation model. In Hiroshi Deguchi and et al., editors, *Simulating Interacting Agents and Social Phenomena*, volume 7 of *Agent-Based Social Systems*, pages 19–35. Springer Japan, 2010.

4 G. Andrighetto, M. Campennì, R. Conte, and M. Paolucci. On the immergence of norms: a normative agent architecture. In *Proceedings of AAAI Symposium, Social and Organizational Aspects of Intelligence Washington DC*, 2007.

5 G. Andrighetto, D. Villatoro, and R. Conte. Norm internalization in artificial societies. *AI Communications*, 23(4):325–339, 2010.

6 A. Artikis, D. Kaponis, and J. Pitt. *Multi-Agent Systems: Semantics and Dynamics of Organisational Models*, chapter Dynamic Specifications of Norm-Governed Systems. 2009.

7 R. Axelrod. *The Evolution of Cooperation.* Basic Books, 1984.

8 P. Bain, J. Vaes, Y. Kashima, N. Haslam, and Y. Guan. Folk Conceptions of Humanness: Beliefs About Distinctive and Core Human Characteristics in Australia, Italy, and China. *Journal of Cross-Cultural Psychology*, 43(1):53–58, August 2011.

9 G. S. Becker. Crime and punishment: An economic approach. *The Journal of Political Economy*, 76(2):169–217, 1968.

10 R. Benedict. *The Chrysanthemum and the Sword : Patterns of Japanese Culture.* Houghton Mifflin, Boston, 1989.

11 H. J. Berman. *Law and Revolution. The Formation of the Western Legal Tradition.* Harvard University Press, Cambridge, Massachusetts, 1983.

12 C. Bicchieri. Learning to cooperate. In C. Bicchieri, R. Jeffrey, and B. Skyms, editors, *The Dynamics of Norms*, pages 17–46. Cambridge University Press, 1997.

13 C Bicchieri. *The Grammar of Society: The Nature and Dynamics of Social Norms.* Cambridge University Press, New York, 2006.

14 N. Binkin, A. Perra, V. Aprile, A. D'Argenzio, S. Lpresti, O. Mingozzi, and S. Scondotto. Effects of a generalised ban on smoking in bars and restaurants, italy. *The International Journal of Tuberculosis and Lung Disease*, 11(5):522–527, May 2007.

15 J. Bourdon, G. Feuillade, A. Herzig, and E. Lorini. Trust in complex actions. In D. M. Gabbay and L. van der Torre, editors, *Logics in Security*, Copenhagen, Denmark, 2010.

16 R. Boyd, H. Gintis, and S. Bowles. Coordinated Punishment of Defectors Sustains Cooperation and Can Proliferate When Rare. *Science*, 328(5978):617–620, April 2010.

17 R. Boyd and P. J. Richerson. Group beneficial norms can spread rapidly in a structured population. *Journal of theoretical biology*, 215(3):287–296, April 2002.

18 L. Brooks, W. Iba, and S. Sen. Modeling the emergence and convergence of norms. In Toby Walsh, editor, *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 97–102. IJCAI/AAAI, 2011.

19 G. W. Brown. Iterative solution of games by fictitious play. In T. Koopmans, editor, *Activity Analysis of Production and Allocation*, pages 347–376. Wiley, New York, 1951.

20 M. Campennì, G. Andrighetto, F. Cecconi, and R. Conte. Normal = normative? The role of intelligent agents in norm innovation. *Mind & Society*, 8:153–172, 2009.

21 C. Castelfranchi and R. Falcone. The dynamics of trust: from beliefs to action. In *Proceedings of the Workshop on Deception, Fraud and Trust in Agent Societies*, Seattle, WA, 1999.

22 C. Castelfranchi and R. Falcone. Trust is much more than subjective probability: Mental components and sources of trust. In *Proceedings of the 33rd Hawaii International Conference on System Science*, Maui, Hawai'i, January 2000. IEEE Computer Society.

23 C. Castelfranchi and F. Paglieri. The role of beliefs in goal dynamics: Prolegomena to a constructive theory of intention. *Synthese*, pages 237–63, 2007.

24 C. Castelfranchi, G. Pezzulo, and L. Tummolini. Behavioral implicit communication (BIC): Communicating with smart environments. *International Journal of Ambient Computing and Intelligence (IJACI)*, 2(1):1–12, 2010.

25 G. Christelis, M. Rovatsos, and R. Petrick. Exploiting Domain Knowledge to Improve Norm Synthesis. In *Proceedings of he Ninth International Conference on Autonomous Agents and Multiagent Systems*, pages 831–838, 2010.

26 R. B. Cialdini, R. R. Reno, and C. A. Kallgren. 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, June 1990.

27 R. Conte. *L'obbedienza Intelligente.* Laterza, 1998.

28 R. Conte, G. Andrighetto, and M. Campenní. Internalizing norms: A cognitive model of (social) norms' internalization. *International Journal of Agent Technologies and Systems*, 2(1):63–73, 2010.

29 R. Conte, G. Andrighetto, and M. Campennì, editors. *Minding Norms. Mechanisms and dynamics of social order in agent societies.* Oxford University Press, Forthcoming.

30 R. Conte and C. Castelfranchi. *Cognitive and Social Action.* UCL Press, 1995.

31 R. Conte and C. Castelfranchi. The mental path of norms. *Ratio Juris*, 19(4):501–517, 2006.

32 R. Cooter. Structural adjudication and the new law merchant: A model of decentralized law. *International Review of Law and Economics*, 14:215–231, 1994.

33 J. Delgado, J. M. Pujol, and R. Sangüesa. Emergence of coordination in scale-free networks. *Web Intelligence and Agent Systems*, 1(2):131–138, 2003.

34 R. Demolombe and E. Lorini. A logical account of trust in information sources. In *Proceedings of the 11th International Workshop on Trust in Agent Societies*, Estoril, Portugal, may 2008.

35 E. Fehr and U. Fischbacher. Social norms and human cooperation. *Trends in Cognitive Sciences*, 8(4):185–190, 2004.

36 O. Figes. *Crimea.* Penguin, London, 2010.

37 F. Flentge, D. Polani, and T. Uthmann. Modelling the emergence of possession norms using memes. *Journal of Artificial Societies and Social Simulation*, 2001.

38 J. Fracchia and R. C. Lewontin. Does culture evolve? *History and Theory*, 38:52–78, 1999.

39 D. Fudenberg and D. K. Levine. *The Theory of Learning in Games.* MIT Press, Cambridge, MA, 1998.

40 F. Fukuyama. *The Great Disruption. Human nature and the reconstitution of social order.* The Free Press, New York, 1999.

41 D. Gambetta. Can we trust them? In D. Gambetta, editor, *Trust: Making and breaking cooperative relations*, pages 213–238. Blackwell, Oxford, UK, 1990.

42 F. Giardini, G. Andrighetto, and R. Conte. A cognitive model of punishment. In *COGSCI 2010, Annual Meeting of the Cognitive Science Society 11-14 August 2010,*. Portland, Oregon, 2010.

43 H. Graham. Smoking prevalance among women in the European Community 1950–1990. *Social Science & Medicine*, 43(2):243–254, 1996.

44 N. Griffiths and M. Luck. Norm Emergence in Tag-Based Cooperation. In *9th International Workshop on Coordination, Organization, Institutions and Norms in Multi-Agent Systems. 79-86*, 2010.

45 H. L. A. Hart. *The Concept of Law.* Clarendon Press, Oxford, 1961.

46 F. A. Hayek. *Law, Legislation and Liberty. Volume I. Rules and Order.* University of Chicago Press, Chicago, 1973.

47 M. Hechter and K. D. Opp. *Social Norms.* Russell Sage Foundation, New York, New York, 2001.

48 J. Henrich, R. Boyd, and P. Richerson. Five Misunderstandings About Cultural Evolution. *Human Nature*, 19(2):119–137, June 2008.

49 J. Henrich, S. J. Heine, and A. Norenzayan. The weirdest people in the world? *The Behavioral and brain sciences*, 33(2-3):61–83; discussion 83–135, June 2010.

50 G. M. Hodgson and T. Knudsen. The nature and units of social selection. *Journal of Evolutionary Economics*, 16:477–489, 2006.

51 M. Hoffmann. Self-organized criticality and norm avalanches. In *In Proceedings of the Symposium on Normative Multi-Agent Systems. Hatfield, UK: AISB.*, 2005.

52 G. Hofstede. *Culture's Consequences.* Sage Publication, 2nd edition, 2001.

53 G. Hofstede, G. J. Hofstede, and M. Minkov. *Cultures and Organizations: Software for the Mind, Third Edition.* McGraw-Hill, 2010.

54 C. D. Hollander and A. S. Wu. The current state of normative agent-based systems. *Journal of Artificial Societies and Social Simulation*, 14(2):6, 2011.

55 C. Horne. Explaining Norm Enforcement. *Rationality And Society*, 19:2, 2007.

56 D. Ibbetson. Custom in medieval law. In A. Perreau-Saussine and J. B. Murphy, editors, *The Nature of Customary Law. Legal, Historical and Philosophical Perspectives*, pages 151–176. Cambridge University Press, 2007.

57 E. T. Jaynes. *Probability Theory: The Logic of Science.* Cambridge University Press, Cambridge, UK, 2003. (Edited by G. L. Bretthorst).

58 A. J. I. Jones. On the concept of trust. *Decision Support Systems*, 33(3):225–232, 2002.

59 M. Kaisers and K. Tuyls. FAQ-learning in matrix games: Demonstrating convergence near nash equilibria, and bifurcation of attractors in the battle of sexes. In *Workshop on Interactive Decision Theory and Game Theory*, 2011.

60 J. E. Kittock. The impact of locality and authority on emergent conventions: Initial observations. In B. Hayes-Roth and R. E. Korf, editors, *Proceedings of the 12th National Conference on Artificial Intelligence*, pages 420–425. AAAI Press / The MIT Press, 1994.

61 B. Kokinov, A. Karmiloff-Smith, and N.J. Nersessian, editors. *Beyond the Carrot and Stick Approach to Enforcement: An Agent-Based Model.* European Conference on Cognitive Science, New Bulgarian University Press, 2010.

62 J. Lang, M. Spear, and S. F. Wu. Social manipulation of online recommender systems. In *Proceedings of the 2nd International Conference on Social Informatics*, Laxenburg, Austria, 2010.

63 U. Lotzmann, M. Mohring, and K. Troitzsch. Simulating the emergence of norms in different scenarios. *Artificial Intelligence and Law*, forthcoming.

64 J. Madureira, A. Mendes, S. Almeida, and J. P. Teixeira. Positive impact of the portuguese smoking law on respiratory health of restaurant workers. *Journal of Toxicology and Environmental Health, Part A: Current Issues*, 75(13–15):776–787, 2012.

65 D. H. McKnight and N. L. Chervany. The meanings of trust. Working Paper 96-04, Carlson School of Management, University of Minnesota, 1996.

66 H. Mercier and D. Sperber. Why do humans reason? Arguments for an argumentative theory. *The Behavioral and brain sciences*, 34(2):57–74; discussion 74–111, April 2011.

67 M. Miceli and C. Castelfranchi. The mind and the future the (negative) power of expectations the mind and the future the (negative) power of expectations the mind and the future. the (negative) power of expectations. *Theory and Psychology*, 12:335–366, 2002.

68 G. Miller, E. Galanter, and K.H. Pribram. *Plans and the structure of behavior.* New York: Holt, Rinehart and Winston., 1960.

69 W. Mitchell. *An Essay on the Early History of the Law Merchant.* Cambridge University Press, Cambridge, 1904.

70 K. Miyazawa. On the convergence of the learning process in a 2 × 2 non-zero-sum two-person game. Econometric Research Program, Research Memorandum 33, Princeton University, Princeton, 1961.

71 U. Mons, G. E. Nagelhout, R. Guignard, A. McNeill, B. van den utte, M. C. Willemsen, H. Brenner, M. Pötschke-Langer, and L. P. Breitling. Comprehensive smoke-free policies attract more support from smokers in europe than partial policies. *European Journal of Public Health*, 22, 2012.

72 M.d. Mooij. *Consumer Behavior and Culture: Consequences for Global Marketing and Advertising.* Sage, Thousand Oaks, California, 2004.

73 J. Morales, M. Lopez-Sanchez, and M. Esteva. Using experience to generate new regulations. In Toby Walsh, editor, *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, pages 307–312. IJCAI/AAAI, 2011.

74 L. Mui, M. Moteashemi, and A. Halberstadt. A computational model of trust and reputation. In *Proceedings of the 35th Hawai'i International Conference on System Sciences*, 2002.

75 J. Porter. Custom, ordinance and natural right in grantian's decretum. In A & J B Murphy Perreau-Saussine, editor, *The Nature of Customary Law. Legal, Historical and Philosophical Perspectives*, pages 79–101. Cambridge University Press, 2007.

76 E. A. Posner. *Law and Social Norms.* Cambridge MA: Harvard University Press, 2000.

77 P. Resnick and R. Zeckhauser. Trust among strangers in internet transactions: Empirical analysis of eBay's reputation system. In M. R. Baye, editor, *The Economics of the Internet and E-Commerce*, pages 127–157. Elsevier Science, Amsterdam, 2002.

78 P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara. Reputation systems: Facilitating trust in internet interactions. *Communications of the ACM*, 43:45–48, 2000.

79 A Ross. *Directives and Norms.* Routledge and Kegan Paul, London, 1969.

80 O. Rutter. *The History of the Seventh (Service) Battalion, The Royal Sussex Regiment, 1914-1919.* Times Publishing Company, London, 1934.

81 N. Salazar, J. A. Rodríguez-Aguilar, and J. L. Arcos. Convention emergence through spreading mechanisms. In W. van der Hoek, G. A. Kaminka, Y. Lespérance, M. Luck, and S. Sen, editors, *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, pages 1431–1432. IFAAMAS, 2010.

82 A. Salehi-Abari and T. White. Trust models and con-man agents: From mathematical to empirical analysis. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, Atlanta, Georgia, 2010.

83 G. Sartor, R. Conte, and R. Falcone. Introduction. agents and norms: How to fill the gap? *Artificial Intelligence and Law*, 7:1–15, 1999.

84 B. T. R. Savarimuthu and S. Cranefield. A categorization of simulation works on norms. In G. Boella, P. Noriega, G. Pigozzi, and H. Verhagen, editors, *Normative Multi-Agent Systems*, number 09121 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2009. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany.

85 B. T. R. Savarimuthu and S. Cranefield. Norm creation, spreading and emergence: A survey of simulation models of norms in multi-agent systems. *Multiagent and Grid Systems*, 7(1):21–54, 2011.

86 B. T. R. Savarimuthu, S. Cranefield, M. A. Purvis, and M. K. Purvis. Obligation norm identification in agent societies. *Journal of Artificial Societies and Social Simulation*, 13(4):3, 2010.

87 F. Schauer. Pitfalls in the interpretation of customary law. In A & J B Murphy Perreau-Saussine, editor, *The Nature of Customary Law. Legal, Historical and Philosophical Perspectives*, pages 13–35. Cambridge University Press, 2007.

88 S. H. Schwartz. A Theory of Cultural Value Orientations: Explication and Applications. *Comparative Sociology*, 5(2-3):137–182, 2006.

89 S. Sen and S. Airiau. Emergence of norms through social learning. In M. M. Veloso, editor, *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 1507–1512, 2007.

90 Y. Shoham and M. Tennenholtz. On social laws for artificial agent societies: off-line design. *Journal of Artificial Intelligence*, 73(1-2):231–252, February 1995.

91 Y. Shoham and M. Tennenholtz. On the emergence of social conventions: Modeling, analysis, and simulations. *Artificial Intelligence*, 94(1-2):139–166, 1997.

92 M. P. Singh. Norms as a basis for governing sociotechnical systems. *ACM Transactions on Intelligent Systems and Technology*, (to appear), 2013.

93 C. R. Sunstein. Social norms and social roles. *Columbia Law Review*, 96(4):903–968, 1996.

94 P. Sztompka. *Trust: A Sociological Theory.* Cambridge University Press, Cambridge, UK, 1999.

95 Y. Tang, K. Cai, P. McBurney, E. Sklar, and S. Parsons. Using argumentation to reason about trust and belief. *Journal of Logic and Computation*, (to appear), 2012.

96 E. Ullmann-Margalit. *The Emergence of Norms.* Oxford University Press, Oxford, 1977.

97 D. Villatoro, G. Andrighetto, J. Brandts, J. Sabater-Mir, and R. Conte. Distributed punishment as a norm-signalling tool. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, Valencia, Spain, 2012.

98 D. Villatoro, G. Andrighetto, R. Conte, and J. Sabater-Mir. Dynamic sanctioning for robust and cost-efficient norm compliance. In Toby Walsh, editor, *Proceedings of the 22nd*

International Joint Conference on Artificial Intelligence, pages 414–419, Barcelona, 2011. IJCAI/AAAI.

99 D. Villatoro, J. Sabater-Mir, and S. Sen. Social instruments for robust convention emergence. In Toby Walsh, editor, *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, pages 420–425. IJCAI/AAAI, 2011.

100 S. Weintraub. *Silent Night: The Story of the World War I Christmas Truce*. Simon and Schuster, New York, 2001.

101 Von G.H. Wright. *Norms and action*. Routledge and Kegan Paul, London, 1963.

102 E. Xiao and D. Houser. Emotion expression in human punishment behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 102(20):7398–7401, May 2005.

Simulation and NorMAS

Tina Balke[1], Stephen Cranefield[2], Gennaro Di Tosto[3], Samhar Mahmoud[4], Mario Paolucci[5], Bastin Tony Roy Savarimuthu[2], and Harko Verhagen[6]

1 **University of Surrey, UK**
2 **University of Otago, New Zealand**
3 **Utrecht University, The Netherlands**
4 **Kings College – London, UK**
5 **LABSS, ISTC-CNR Rome, Italy**
6 **Stockholm University, Sweden**

Abstract

In this chapter, we discuss state of the art and future perspective of the study of norms with simulative methodologies, in particular employing agent-based simulation. After presenting the state of the art and framing the simulative research on norms in a norm life-cycle schema, we list those research challenges that we feel more apt to be tackled by the simulative approach. We conclude the chapter with the indications for the realization of a NorMAS simulation platform, illustrated by selected scenarios.

1998 ACM Subject Classification I.2.11 Distributed Artificial Intelligence

Keywords and phrases Simulation, Norms, MAS

Digital Object Identifier 10.4230/DFU.Vol4.12111.171

1 Introduction

In this chapter, we present the state of the art in the agent-based simulation of norms. Simulation is emerging as one of the most important tools in the stock of the social science researcher. Indeed, simulation provides an unique way to advance understanding in theory, by building conceptual models, and at the same time to apply ideas to specific scenarios, by allowing accurate descriptions of the real world mechanisms in the models of the agents and of their interaction. Starting from the definition of the essential components of a simulation-based approach to norms, we propose a selection of conceptual challenges of relevance for the NorMAS community, and we suggest simulative approaches to deal with them. After stating basic definitions of simulation, norms and their interaction, we present an overview of simulation research on norms, then we move on to research challenges, on the basis of which we conclude the chapter with the outline of a NorMAS simulative platform.

1.1 What is simulation?

In the attempt to understand the world around us, predict its future, and build this future by policy choices, we use a modelling approach; that is, the creation of a model as a new object, which is simple to investigate, and at the same time carries some of the properties exhibited by the object of investigation. Simulation is but one of many modelling techniques, using automated computation to represent behaviours and properties of the target.

While several flavours of simulation exist, in this chapter we will focus on agent-based modelling and simulation, and on its application to the study of the norm lifecycle (see

Normative Multi-Agent Systems. *Dagstuhl Follow-Ups, Vol. 4.* ISBN 978-3-939897-51-4.
Editors: Giulia Andrighetto, Guido Governatori, Pablo Noriega, and Leendert W.N. van der Torre; pp. 171–189
DAGSTUHL FOLLOW-UPS Dagstuhl Publishing
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany

[36] and also Chapter 5 in this volume) – creation, identification, spreading, enforcement and emergence. For several reasons, that we will consider briefly now, we believe that the agent-based approach is especially suited to model the norm lifecycle.

The agent-based approach, inheriting tools and objectives of distributed artificial intelligence, is a natural way to describe complex systems: focusing on the individual and the mechanisms that control its behaviour, phenomena can be constructed generatively and puzzling, unexpected, and often enlightening results can emerge. Agent based models can also represent transient phenomena in a natural way, and do not get attracted towards equilibria that are not reachable in practice (because of sensibility issues, or of interminable reaching times; for a famous example, see [22]).

The focus on the individual distinguishes the agent-based approach from the traditional ones based on mathematical and stochastic methods where the characteristics of a population are averaged together [17]; the more complex individuals and their interactions are, the less this averaging is expected to work. Agents, instead, allow maintaining a plausible description of phenomena both at the micro and at the macro level.

Norms have an elaborate lifecycle, in which important steps – recognition, adoption, decisions to punish are examples – happen in the agent's mind. Thus, even with minimally complex norms, issues of autonomy and heterogeneity become crucial. Agents are the only existing approach capable of modelling autonomy [49] in heterogeneous agents, and are a natural paradigm for modelling adaptation, evolution, and learning.

Finally, modelling with agents reduces, with respect to equation based models, the pressure towards oversimplification that the "hard" sciences inherit from more than a century of success in understanding the physical world. This success, however, is not mirrored in the sciences that have applied the same tools to the social world, namely, the economic science, that (rather infamously) has shown to be not yet suited to deal with phase changes or crises. Simulation, and agent-based modelling and simulation in particular, is definitively a different approach to modelling social behaviours; agent modelling, in contrast to equation-based models, enables the description of agent internals in an accurate way – including mental constructs like norms.

1.2 What are norms?

While the research community does not converge on a single definition of norm, for the purposes of this paper we will consider norms as behavioural regularities that guide agents' decisions in a social environment. There are a number of features associated with norms, depending on the focus and the role they play in a multi-agent system. The two most important are:

1. **Expectations**. Norms presuppose the belief that agents around a given agent will abide to the prescribed behaviour, and they expect that agent to conform to it as well [11].
2. **Punishment**. Compliant and deviant behaviours are usually associated with more or less explicit rewards and sanctions. Although authors distinguish between *injunctive* and *descriptive* norms, associating with the latter a lesser importance for sanctions [27], depending on the social capabilities and/or the cognitive plausibility of the artificial agents it is always possible to postulate a form of reward, either positive or negative: be it in the light of the agent's own morals or emotions, social-image or reputation, the blame of his peers or the preservation of his own social-image or reputation, and – in the case of an institutional framework – where explicit fines can be imposed.

These normative aspects find an implementation in the literature in the following entities:

- conventions
- social norms
- legal norms

depending on whether the focus is on the coordination effect achieved through the conditional compliance of individual agents and their reciprocal expectation, the role of obligations and the effects of punishment on agents' decisions, the presence of an institutional authority or the implementation of normative roles in the agent population. However these three categories should not be considered mutually exclusive.

Norm dynamics is an important aspect captured by social simulation. It involves the possibility for norm *emergence* and *transition* from one category to the next (and possibly back to a previous one). For example, something that emerges as a convention can later be explicitly prescribed and enforced, and eventually become part of the legal system. It also involves the possibility of representing scenarios of *normative conflict*, where an agent's behaviour is subject to multiple normative inputs, e.g. a social and a legal norm.

Regardless of the emphasis on the correlated social and psychological phenomena ascribed to norms, they are considered the principal means to achieve social order in a population of autonomous agents, as they offer a solution to situations that pose a social dilemma.

1.3 Social Science Background

A central concern in social science in general and sociology in particular is the relation between the individual level and the societal level. The micro-macro link debate has been at the core of sociology since its creation as an area of scientific research. Mechanisms linking indiduals to societal effects and society to individual behaviour are many, norms being one of them. One of the main researchers within the social mechanisms "school" is Jon Elster. In [20] he describes a whole range of social mechanisms. Among them is the concept of social norms. A social norm is defined as an injunction to act or abstain from acting. The working mechanism is the use of informal sanctions aimed at norm violators. Sanctions may affect the material situation of the violator via direct punishment or social ostracism. An open question is the costs of sanctioning. Apart from social norms, which are followed due to the possibility of violations being observed and sanctioned, Elster describes moral norms, which are followed unconditionally, and quasi-moral norms, which are followed as long as others are observed complying with them. Other connected concepts are legal norms (where special agents enforce the norms) and conventions that are independent of external agent action. In his text, Elster discusses in detail some examples of norms such as: norms about etiquette, norms as codes of honour, and norms about the use of money. For the purpose of this chapter, we will confine the social mechanisms and concepts to social norm related ones.

In 1956 Morris [31] proposed one definition of norms based on 17 characteristics which he grouped in 4 categories. The categories and characteristics can be understood as the first set of conceptual challenges when doing NorMAS-related simulations, as the simulation designer – for each of them – needs to make a decision whether and how to model the respective normative aspect. Summarizing Morris' characteristics, we can derive four building blocks that can be combined in a normative simulation. Different (implementations of) normative systems may have different distributions of these building blocks over the constituents of the system:

1. Ways to distribute/communicate norms
2. Ways of deliberating on norm following (or violating)

3. Ways of detecting violations of norms and norm following
4. Ways of implanting positive and negative sanctions as a consequence of norm compliance or violation (enforcing the norms)

This rather extensive list of characteristics (and indirectly conceptual challenges) was extended by Gibbs, who in 1965 published a norm typology based on Morris' earlier work [24]. One important additional building block Gibbs highlighted was:

5. Ways to detect norms

This building block could be located anywhere on the scale between collective evaluations of actions and situations to the imposed by external sources. This is closely linked to the question about the origin of norms: bottom-up (emergence from behaviour regularities) or top-down (i.e. mostly as a control mechanism aiming for a certain social order). In the former case, further questions arise such as how agents recognize norms and learn about them and how they internalize them after recognition. And even if individual agents recognize something as a norm, how do group norms or generally accepted norms emerge?

In the Dagstuhl NorMAS Seminar in 2007 a final building block was identified:

6. Ways to modify norms

This building block highlights the question of whether norms can be modified in the simulation (or whether they are static), which norms can be modified and furthermore who can modify which norms under which circumstances.

1.4 Simulation and norms

For the NorMAS (Normative Multi-Agent Systems) community, agent-based simulations offer a platform to evaluate the behaviour of different models of norms and normative processes in a dynamic environment. Vice versa, the NorMAS community can supply (social) agent-based simulation studies with formal models of social concepts and mechanisms, especially those related to normative concepts, such as norms proper, roles, values, morals and conventions, and their transmission within a society. Agent-based simulation has had great success in modelling normative behaviour, due to its ability to address the fundamental role of norms by reconstructing the micro-macro link: generating macro-level phenomena from micro-level specifications and vice versa, modelling micro-level behaviour and choices as bound by macro-level phenomena. To date, this has largely been achieved through models based on individualist agents that behave according to their own internal goals, with social behaviour resulting as an emergent phenomenon. For example, the BDI architecture on which many models are based is a strongly individualist architecture. An agent is defined by its individual beliefs, desires and intentions and any social behaviour results by emergence [21] or deterrence [7].

While explicit social knowledge can be added to the BDI architecture, e.g. by explicitly defining a set of obligations an agent has to follow, as in the BOID architecture [13], more advanced models of normative behaviour such as the EMIL-A architecture [4] have recently been proposed to transcend the individualistic nature of an agent to some extent by incorporating both perception of norms and reasoning with norms into the agent. Now the agents are able to avail themselves of a normative interface with the world rather than just a factual one as is the case for a BOID agent. Still, desires and intentions of the agent are defined individualistically, with normative knowledge evaluated according to these desires and intentions. But what if the agent was not quite as individualistic? What if agents have an active interest in social behaviour, in sharing goals, in cooperating? What if agents can explicitly reason about their group structure, their friendship relations, and use them both as a source of norms and a context that drives norm interpretation? And how do we integrate emotions into these frameworks or open them up to create glass-box cognitive models to replace the black box of BDI? And what about emotions? We advocate work on these issues to improve agent simulation models so that:

a) models will no longer analyse whether social behaviour is possible but rather what kind of social behaviour might emerge;
b) models will no longer be based on the long-standing paradigm of 'atomism', i.e. individual agents are no longer seen as social atoms connected by chance but holistically, as inseparable parts of a social entity;
c) models will no longer be purely behavioural, allowing agents to understand their own intentions and goals and those of other agents; and
d) models of human agency will address social, psychological and emotional aspects simultaneously.

2 Overview of Norm Simulation Research

This section provides a brief background on different phases of norm life-cycle and the different mechanisms that have been employed to study those using simulation-based research. Broadly, from the viewpoint of the society, the three important stages of norms are the formation stage, propagation stage and the emergence stage. Researchers employing simulation-based studies of norms have investigated various mechanisms associated with norms with each of these stages. Mechanisms employed in the *norm formation* stage aim to address how agents can create norms in a society and how individual agents can identify the norms that have been created. Mechanisms used in the *norm propagation* stage aim to explain how norms might be spread and enforced in the society. The *emergence* stage is characterized by determining the extent of the spread of a norm in the society.

Based on these three important stages of norms, Savarimuthu et al. [36] have identified five phases (i.e. expanded stages) of the norm life-cycle which are norm creation, identification, spreading, enforcement and emergence. Figure 1 shows these five phases of the norm life-cycle on the left and the mechanisms investigated by researchers for each of the phases on the right. Note that not all the mechanisms shown in the figure are discussed in this chapter. For details refer to [36]; compare also with the model for norm dynamics presented in Chapter 5 of this volume.

2.1 Norm creation

The first phase of the life-cycle model is that of norm creation. The norms in multi-agent systems are created by one of the three approaches. The three approaches are (a) a designer specifies norms (off-line design) [15], (b) a norm leader specifies norms [12, 45], and (c) a norm entrepreneur considers that a norm is good for the society [26].

In the off-line design approach, norms are designed off-line, and hard-wired into agents. This approach has been used by researchers to investigate norms that might be beneficial to the society as a whole using social simulations. An example of a hard-wired norm includes investigations on how different traffic rules emerge by fixing tendencies of agents to drive on the left or the right [43, 41]. Researchers have also investigated a pre-specified *finder-keeper* norm where they compare a society that does not have a norm [16] with a society

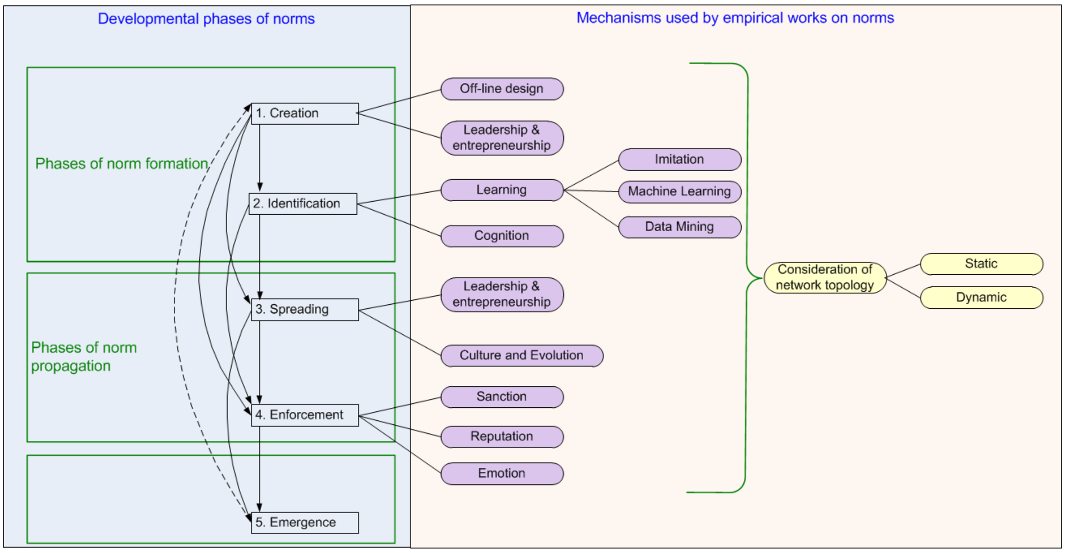


Figure 1 Developmental phases of norms.

that has a norm. In the leadership approach, some powerful agents in the society (the norm leaders) create a norm. The leadership approach can be based on authoritarian or democratic leadership. The leader can provide these norms to the follower agents [12, 44]. Leadership mechanisms are based on the notion that there are certain leaders in the society, who provide advice to the agents in the society. The follower agents seek the leaders advice about the norm of the society.

Boman [12] has used a centralised leadership approach, where agents consult with a normative advisor before they make a choice on actions to perform. Verhagen [45] has extended this notion of normative advice to obtaining normative comments from a normative advisor (e.g. the leader of the society) on an agent's previous choices. The choice of whether to follow a norm and the impact of the normative comment on an agent are determined by the autonomy of the agent. Once an agent decides to carry out a particular action, it announces this decision to all the agents in the society, including the leader of the society, and then carries out that action. The agents in the society can choose to send their feedback to this agent. When considering the received feedback, the agent can choose to give a higher weight to the feedback it received from the leader agent.

In the entrepreneurship approach to the creation of norms, there might be some norm entrepreneurs who are not necessarily the norm leaders but create a proposed norm. For example, Henry Dunant, the founder of Red Cross was the entrepreneur of the norm to treat wounded soldiers in a war as neutrals. When an entrepreneur agent creates a new norm it can influence other agents to adopt the norm [23, 26]. Hoffmann [26] has experimented with the notion of norm entrepreneurs who think of a norm that might be beneficial to the society. An entrepreneur can recommend a norm to a certain percentage of the population (e.g. 50%) which leads to varying degrees of establishment of a norm.

2.2 Norm identification

If a norm has been created in the society using one of the explicit norm creation approaches (e.g. leadership and entrepreneurship) then the norm may spread in the society. However, if the norms have not been explicitly created (i.e. norms are derived based on the interactions between agents), then an agent will need a mechanism to identify norms from its environment based on the interactions with other agents. In game-theory based empirical works [42, 41], agents have a limited number of actions that are available, and they choose the action that maximizes their utility based on some learning mechanism about the behaviour of other successful agents. These mechanisms could be imitation, machine learning (e.g. genetic algorithms) or data mining. The second approach to norm identification considers the cognitive capabilities of an agent to infer what the norms of the society are [5, 3]. In the cognitive approach, one or more cognitive agents in a society may come up with norms. At this stage, the norm exists only in the mind of one agent. It can become a social norm if that is accepted by other agents based on the deliberative processes that they employ. In this approach the other agents have the cognitive ability to recognize what the norms of a society are based on the observations of interactions. Agents have normative expectations, beliefs and goals. It should be noted that the norms inferred by each agent might be different (as they are based on the observations that the agent has made). Thus, an agent in this model creates its own notion of what the norms are based on inference.

2.3 Norm spreading

Norm spreading relates to the distribution of a norm among a group or in the system. Once an agent forms a belief on what the norm in the society or in its group is (i.e. either based on norm creation or identification), several mechanisms help in spreading the norms such as leadership, entrepreneurship, cultural, and evolutionary mechanisms.

Leadership mechanisms are based on the notion that there are certain leaders in the society. These leaders provide advice to the agents in the society. The follower agents seek the leaders' advice about the norm of the society. Thus, the norm spreads in the society. Researchers have experimented with centralized and decentralized models of leadership [45, 38] for norm spreading. Hoffmann [26] has experimented with the notion of norm entrepreneurs who think of a norm that might be beneficial to the society. An entrepreneur can recommend a norm to certain percentage of the population which leads to varying degrees of norm spreading in the society.

Boyd and Richerson [35] proposed that norms can be propagated through cultural transmission. According to them, there are three ways by which a social norm can be propagated from one member of the society to another. They are

- Vertical transmission (from parents to offspring)
- Oblique transmission (from a leader of a society to the followers)
- Horizontal transmission (from peer to peer interactions)

Of these three kinds of norm transmission mechanisms, vertical and oblique transmissions can be thought of as leadership mechanisms in which a powerful superior convinces the followers to adopt a norm. Horizontal transmission is a peer-to-peer mechanism where agents learn from day-to-day interactions from other peers.

Norm spreading based on evolution involves producing offspring that inherit the behaviour of their parents. One well known work in this category is by Axelrod [7]. Other researchers have also experimented with evolutionary models for norm spreading [14, 46].

2.4 Norm enforcement

Norm enforcement refers to the process by which norm violators are discouraged through some form of sanctioning or norm compliers are encouraged by some form of benefits (i.e. the utilization of carrots and sticks). A widely used sanctioning mechanism is the punishment of a norm violator (e.g. monetary punishment which reduces the agent's fitness or a punishment that invokes emotions such as guilt and embarrassment) through regimentation (where the agents are considered as white boxes or it is assumed that their actions can directly be controlled) or with the help of some kind of police agents. Reputation mechanisms or image information have also been used as sanctions (e.g. an agent is black-listed for not following a norm). One final method of enforcement can be administered through the norm violators themselves (by means of self-enforcement[1]). The process of enforcement helps to sustain norms in a society. Note that enforcement of norms can influence norm spreading. For example, when a powerful leader punishes an agent, others observing this may identify the norm. Hence, the norm can be spread. Norms can also be spread through positive reinforcements such as rewards. Some researchers have considered enforcement as a part of the spreading mechanism [8].

2.5 Norm emergence

The fifth phase is the norm emergence phase. We define norm emergence to be reaching some significant threshold in the extent of the spread of a norm; that is a norm is followed by a considerable proportion of an agent society and this fact is recognised by most agents. For example, a society could be said to have a norm of gift exchange at Christmas if more than $x\%$ of the population follows such a practice. The value of x varies from society to society and from one kind of norm to another. The value of x has varied from 35 to 100 [36] across different empirical studies of norms. Emergence can be detected either from a global view of the system or through a local view of an agent (e.g. an agent might only see agents that are one block away on all directions in a grid environment). Spreading of norms with or without enforcement can lead to emergence. Once a norm has emerged, the process can continue when an entrepreneur or a leader comes up with a new norm that replaces the old one. This is indicated by a dotted arrow in Figure 1. The adoption of a norm may decrease in a society due to several reasons. A norm that has emerged may lose its appeal when the purpose it serves does not hold or when there are not enough sanctions or rewards to sustain the norm or when other alternate effective norms emerge. Note that the model presented here is from a bird's-eye view (i.e. an external agent that observes the society). An external agent will able able to observe the norm establishment and de-establishment in the society based on the emergence criterion (i.e. the extent of spread of the norm).

2.6 Consideration of network topologies

An important attribute of the research on norms is the consideration of network topology. The underlying interaction topology of agents has an impact on all phases of norm development. For example the interactions between a leader and his followers have an implicit network topology (i.e. a star network) which governs how norms created by the leader may spread and may lead to norm emergence in an agent society. Hence the consideration of network topology is included as one of the nine main categories in Figure 1. The network structure of the society can either be static or dynamic (i.e. can evolve due to agents joining and leaving).

[1] Balke and Villatoro [9] have pointed out that enforcement mechanisms should not only consider who sanctions, but also observe the violations and who determines the sanctions.

For a detailed overview of different mechanisms employed in the simulation-based study of norms which have not been covered in this chapter such as learning approaches, cultural and evolutionary approaches, reputation-based approaches please refer to the work of Savarimuthu et al. [36].

3 Research challenges

3.1 Methodological Challenges

The methodological challenges for simulation of normative multi-agent systems, apart from the challenges inherent to simulation as a method in general, include the development of measures for the mechanisms described in section 1.4 on Simulation and Norms. Few measures exist in the social science studies of norms and normative processes that can be used here. The development of such measures for use in simulation studies will thus in itself be a contribution to social science in general. The process of *norm adoption* – defined as the processing of an obligation, prohibition, or permission from the part of the individual agent – is a central element in normative reasoning and directly related to the behavioural outputs in norm emergence and enforcement. Its study aims at the *micro-foundation of a normative agent architecture*, an endeavour that connects social simulation with the efforts made in other social sciences, and brings in results from a diverse array of disciplines – psychology, sociology, anthropology, economics, etc. – in order to increase the plausibility of the implemented models.

Validation of simulation results is another important methodological point. Game theory offers a consistent set of tools to capture and analyse the results of agent interactions, especially in the case of social dilemmas. The implementation of the relative game-theoretical concepts in computational models assumes the introduction of principles of economic rationality, however relaxed they might be. For a discussion of the treatment of norms from the standpoint of game-theory, see Chapter 2 in this same volume.

Also, the development of tools for the *visualization of effects* going beyond, e.g., the use of different colours for agents to represent different norms, is one solution will be beneficial for the *interpretation and communication of simulation results*. The same is true for the development of mathematical measures for e.g. norm salience such as developed in [45].

Finally, in research on norm identification, there is a *norm bootstrapping* problem: in order for agents to learn norms, there must already be norms present in the environment or agents must be provided with the means to invent new norms and act on them. In current simulation experiments the researchers already know the norms that their agents are intended to learn, as these were pre-engineered into the system. This can lead to ad hoc solutions that do not generalise to address other scenarios and problems. One possible solution is to designing mechanisms and protocols for humans to control some of the agents in a simulation. Human participants could then be instructed to exhibit specific norms, given tasks that should implicitly lead to normative behaviour, or left to bring their own real-world norms into the simulation. This would provide a much greater challenge for the design of norm learning mechanisms, and may lead to the development of more powerful approaches.

3.2 Topologies

In addition to the type of game that is being played to model agent interactions (e.g. a coordination or cooperation game), the network structure [33] is also a fundamental component of any agent-based simulation, since real-world interactions between individuals are governed by the underlying topology (i.e. we interact with people from our family network, work network and so on). Apart from identifying connections between different components (or agents) within the the system, the network structure also imposes constraints on agent actions, interactions and observations. Networks have been investigated from two different perspectives: as *static* networks or *dynamic* networks.

In static networks, agents have fixed connections to other agents in their environment. Depending on the simulation, these connections define the ability to interact and observe the interactions of others. Various types of static networks [2] have been studied, and their effects analysed in the literature of social norms and agent-based simulation. The main types of static networks commonly analysed are random, lattices, small worlds and scale-free networks. Indeed, there has been a considerable recent effort on analysing the impact of each of these network types on the effectiveness of social norms [18, 40, 30, 47].

In dynamic networks, agent connections are modifiable during the course of the simulation, possibly due to the dynamism of the system under investigation. Dynamic networks are representative of phenomena that can be observed in the real world (agents dying, relocating etc.). For example, Savarimuthu et al. [37] describe the emergence of norms in a scenario in which the network of interactions dynamically changes as the agents move in a two-dimensional abstract social space, with agent connections being formed through collisions in this space. Alternatively, Mungovan et al. [32] propose an interesting scenario in which agents have a fixed interaction network (small-world) and are provided with the possibility of having random interactions.

- **Analysing the influence of different locations in the network**. Much work has focused on identifying the effects of general network characteristics on social norms, yet there has been very little attention on the influence of different locations within these networks. There may be agents in certain powerful locations in the network that can either positively or negatively influence norm emergence. For example, in scale-free networks, *hubs* (nodes with a vast amount of connections) are known to play an important role by either supporting or obstructing social norms. A more detailed analysis of the specific effects of different locations is thus crucial in overcoming any obstacles in the spreading of a social norm in a complex system.
- **Investigating the influence of dynamic networks**. Most current social norm models consider static networks, and seem to consider only interactions between two agents resulting from their unchanging static connection. However, since dynamic networks incorporate many characteristics of real world complex systems, it is important to determine the impact of these types of networks on the evolution of social norms. Dynamic networks thus provide ample research opportunities for work involving social norms and complex systems. In addition, and as done initially for the construction of in-silico social networks ([33, 10, 18, 19]), we need models and algorithms that simulate the dynamic evolution of social networks over time.
- **The existence of different groups and multiple topologies**. In most experimental models, individuals have been considered to belong to only one group, but in reality agents can be influenced from many angles as an individual might belong to several groups (e.g. work group, neighbours and hobby group). Each of these groups can have a different topology and the existence of such many topologies might influence agents' actions. Additionally, the strength of the links from an agent to other agents might be different which may influence norm emergence. The strength of the links can be modeled by assigning weights.

4 Outline of a NorMAS platform

Tools and platforms dedicated to agent-based simulation are now widely diffused in the simulation community. Besides making simulations easier, they favour standardization, sharing of code, and replication of experiments. However, existing tools focus mostly on issues of synchronization, reporting, communication and topology. With the exception of Jason[2], which implements a BDI architecture, these tools are agnostic on the matter of mental structures as the norms that we are describing. We believe that the field could benefit by the introduction of a specialized platform dealing with the whole lifecyle of norms. In this section, leaning on some scenarios presented next, we point out some of the demands and requirements that a NorMAS platform should satisfy.

4.1 Scenarios

Scenario 1: The culture of graffiti artists has a rich social structure that has emerged over time [48]. Based on the goal of "getting up" (gaining reputation), practitioners ("writers") develop their skills by creating graffiti works in various named styles (such as *tags*, *throw-ups* and *pieces*) of increasing complexity and in locations with varying levels of risk of detection and/or injury. Writers can work together in teams (named "crews"), which gives them the chance to participate in more complex works. The creation of works signed with a group's tag also reduces the risk of any individual being held responsible if arrested. There are norms governing when a work can cover another (based on a complexity hierarchy of named styles) and how writers can gain and claim social status, to potentially rise from the status of a *pawn* (or *toy*), through the level of a *knight*, to become a *king* or *queen*. Sanctions such as "slashing" (painting a line through or tagging over another's graffiti) or physical violence may be applied if these norms are breached. There is also a notion of honour among thieves and associated social recognition of those who are trusted not to give any information about other writers to authorities. This culture can be viewed as a microcosm of human society, exhibiting some significant social structures and processes. A simulation implementing an abstraction of graffiti culture would be a useful testbed for evaluating theories of the creation, acquisition and recognition of norms by agents and the spreading and emergence of norms within a society.

Scenario 2: Divya Chandran, a new resident in a virtual environment (e.g. Second Life) wants to explore the rich interactions offered by the new medium. She wants to go to a virtual park and relax by the fountain and listen to chirping birds. She flies to the virtual park and upon looking at the layout starts wondering if there are norms that regulate where to sit in the park (e.g. benchs, wall-like structures, and the grass). She notices some water fountains and some soft-drink fountains from the sponsor of the park. She would like to get a drink, but does not know if there is a norm governing the usage of the fountain. She wonders if she should get a cup from the jazzy sponsor's booth by paying virtual money or if she needs to acquire the skill of making a cup object. Can she take her drink to all the areas of the park or is she restricted to a particular zone (e.g. food permitted zone)? And finally, what should she be doing with the cup – store it in her inventory for further use, destroy it or just leave it around? How can she find what is the norm associated with littering in the park? Can she leave the cup anywhere for some mechanism to collect it or should she find a rubbish bin and drop it there? Also, she is curious to know the social interaction rules that

[2] `http://jason.sourceforge.net/wp/`

apply in this setting. Is she supposed to send a greeting 'signal', to engage in conversation with strangers? If so, what should this signal be (e.g. uttering a 'hi', waving a hand, or an obscure ritual specific to the virtual environment)? Should she also discretely walk away from people engaged in conversation, and if that is the case what distance is considered polite? Finally, once she has learned the norms of the park, will the norms of this particular park be applicable to all the parks in the virtual environment? When she visits the park at a later date will the norms of the park still be the same or will there be new norms?

Scenario 3: As defined by Schollmeier [39], a peer to peer system is a distributed system that consists of members, each of which share some resources (hardware, software or information) with other members. There are many research efforts that address the issue of free riding behaviours in P2P file sharing systems [29, 25, 28], but here we provide an explanation of the problem in relation to the *Gnutella* system. *Gnutella* is a P2P file sharing network, in which each peer plays the role of both a client and a server. As a client, the peer requests files from other peers, while as a server the peer provides files to other peers. When a peer needs access to a file in the network, it creates a query regarding the desired file and passes this query to its neighbours. If the neighbouring peer has the file, it replies to the request. If not, the peer passes the request to its neighbours and returns the response of those neighbours back to the requesting peer. When a peer downloads a file, the informal norm of the system is that it should make the file available to others. A peer does not pay anything to access files and there is no limit on the amount of files that a peer can access nor the proportion of the files it shares. Therefore, it might be rational for peers to not waste their bandwidth responding to other peers' requests as they can access different files on the network without sharing any of their own files, which is known as the problem of *free riding.* As shown by Adar and Huberman [1], 70% of Gnutella peers share do not share any file; they receive files from the network without sharing. A simulation platform can provide vital help in investigating this phenomenon and in analysing various mechanisms that can lead to the preemption of free riding. Researchers can investigate both top-down policy based mechanisms for handling free riding and also can investigate the more interesting bottom-up approaches where norms against free riding may emerge dynamically. This scenario can be used not only to study the emergence of norms but also how the emerged norms can be spread, enforced and eventually be updated (depending upon changing circumstances).

4.2 Agent Internals

To be capable of dealing with norms in a non-trivial way, agents must be endowed with extensive capabilities to deal with cognitive representation of artifacts related to the norm life-cycle. The processes that are executed in the mind of the agent, and that a NorMAS library should provide, are exemplified in the rest of this section, describing norm formation and norm propagation.

Norm formation. Explicit mechanisms of norm formation should be available to the designer. These should cover both norm creation and norm identification. Norm creation is the creative process of an agent generating a new candidate norm. This may happen at the beginning of the simulation, when all agents might randomly draw norms from a more or less large set, or during the simulation, e.g. if "norm entrepreneur" agents invent new norms that they wish to propose to the society, or if agents have their own personal norms that may be motivated by personal emotions and/or goals. For norm creation, following the mechanisms proposed in section 2.1, we could imagine the following:

(a) functions to select, possibly at random, one of the norms provided off-line by the system designer, a selection that could be changed according to the performance of the agent under that norm; for example, in a simulation of peer to peer sharing, choosing between pre-constituted norms defining the appropriate time length one should share a downloaded file.

(b) Functions to create a norm to be induced through authority by special agent figures; for the graffiti example, consider how figures of prominence (artists, *kings* or *queens*) could dictate norms about slashing, enacting thus a collective behaviour in the group they lead.

(c) Functions describing how a norm entrepreneur could envisage the possibility of creating a new norm – possibly by foreseeing some of the collective effects that could benefit himself, the group he belongs to, or both.

This involves inducing norms that may hold in the society, based on the agent's experience and observations. For complex scenarios where many different actions may be observed, this will require a module for the detection of frequent behaviour (regularity detection).

In the virtual world example, some of the regular visitors in the park could try to instantiate an informal norm of precedence in greeting (as in, you wait for the newcomer to greet first, or the opposite, or one based on avatar creation date), trying to assess which norm would contribute better to create a pleasant mood.

Here, we also need to distinguish between personal norms (recognized only by isolated agents) and actual ones. At the collective level, the generated norms are just candidate norms until the population contains a sufficient number of agents with similar personal norms to trigger norm recognition by other agents.

The proposed regularity detection module could be used also as the basis of norm identification (see section 2.2), with different implementations for learning-based identification and cognition-based one. Agents could also be endowed with mental constructs of *candidate norms* populated by the regularity detection module.

At the cognitive level, regularities can be caused by the preferences or goals of the agents involved without the need for a norm to be present; but a norm can be hypothesized if we can interpret an action by an agent as a signal that a violation of accepted behaviour has occurred. Examples of such signals include negative emotional reactions (indicating a potential need for emotion detection) and the explicit invocation of sanctions.

Thus, agents should be endowed with an internal process for keeping track of the relevance of a norm to the agent and the community over time, thus helping the agent to determine when to observe the norm and when the norm should be forgotten [6]. As sanctions cannot, in general, be distinguished from punishments motivated by individual goals, a threshold of independent evidence should be achieved before an agent induces a norm from observation.

However, both candidate and confirmed norms should be available to an agent's introspection, as it may wish to plan actions that test whether a candidate norm is prevalent in the community. With reference to the examples provided, a new agent in the graffiti scenario would need a routine to identify, between the regularities he can detect in the "slashing" acts, which rules of priority are in action. In the virtual world example, the visitor could observe the behaviour of other players, come up with candidates to a norm, and then discuss them explicitly, or even try to enact some slight violation to get confirmed by sanction.

Norm propagation. Norm propagation includes spreading and enforcement (see section 2.3). Norm propagation requires communication between agents and an explicit, transmissible representation for norms. A library for simulation of this process would include, with regard to the agent internals, both agent-to-agent communication (possibly connected by one of the topologies discussed above) and some broadcast mechanism. Propagation should also keep into account possible differences in value structure, for example, when entering in contact with different groups or cultures. The justification of a norm could be based on values that are different between the new agent and the community.

For norm enforcement (see section 2.4), agents would also need routines for performing sanctioning behaviour in relation to norms that they believe to be in effect.

In addition to the specifically normative components above, there are other ingredients that may be necessary or useful for agents internals. These are:

- a connection to the object/enviroment/context level (for agent sensing and acting, for measures of regularity, etc.)
- representation of cognitive artifacts related to norms (i.e. reputation, trust, emotions,and values), in their role of indirect sanctions, salience indications, or conflict resolution between norms. About trust and reputation, identifying oneself with a group can lead the agent to generate goals instrumental to the preservation of this identity, to the strengthening of his affiliation (and of the related trust) through the acquisition of a more central position in the group, and other goals related to reputation building and maintenance. Values confer the agents a moral sense (proto-morality – they are only one component among others along the way a moral architecture). In a simplified process they are able to flag a certain state of the world (W) as being good or bad.
- a norm-aware practical reasoning architecture: the use of a practical reasoning architecture requires extensions to the classic BDI architecture (as with the BOID architecture, [13]); but this is not yet common in the practice of multi-agent simulations which often focus on rather simple scenarios in which agents have only a few actions to choose from, and the choice between them is governed by one or more numeric variables (e.g. representing probabilities of actions). However, to gain a deeper understanding of how norms are created, evolve and emerge in a society, and when and how agents choose to follow norms, it seems necessary to have an explicit connection between practical reasoning and normative behaviour. A NorMAS simulation toolkit could support research in this direction by providing an optional explicit BDI-style execution cycle (along the lines of AgentSpeak [34]), to which additional normative reasoning steps could be added.
- Group identity: affiliations and belonging, besides being a primitive social drives, is also input for the agent cognitive processing. Sharing (a set of) values, adopting and following the same norms, and consequently enforcing them are together process on which group identity is based.

4.3 Interaction Mechanisms and Topologies

In a social simulation, there can be various ways that agents can directly interact (as opposed to influencing each other indirectly by acting in the environment). An MAS simulation toolkit should provide support for a range of interaction mechanisms, including direct agent-to-agent messaging, agent-to-group messaging, and the exchange of utility such as making payments. It may also be useful to provide an interface for registering gossip to be spread automatically to all encountered agents.

Figure 1 shows various mechanisms researchers have been proposed for norm establishment in a society. A good normative simulation system should allow one or more of these mechanisms to be plugged in and experimented with. For example, researchers should be able to experiment with a scenario by setting up an entrepreneurial mechanism for norm creation, an imitation model for norm spreading, a reputation mechanism for norm enforcement, and then observe how norms are established in the society. After achieving this in one particular experiment, the researchers should be able to replace the reputation mechanism for enforcement with a punishment mechanism in the next experiment and study the effects (such as time to converge and efficiency). Such a pluggable simulation architecture will be beneficial to the research community, both for sociologists to understand human societies and computer scientists to deal with artificial agent societies.

The set of agents that a given agent can observe and interact with is dependent on the underlying physical environment and/or social network topology. However, an agent may need to reason about higher level social structures such as connections formed in different social and organisational contexts. These can be viewed as overlay networks on top of the underlying topology, and the platform should offer support for agents to query and update information about these personalised and contextualised views of society.

Support should also be provided for an agent to store its own private data about other agents and to query and configure (e.g. by setting the history length) its record of past interactions with other agents. In a P2P system, an agent might store some information about a free rider agent in order to avoid interacting with this agent again.

As discussed in Section 3.2, topologies impact norms spreading and emergence. Therefore, a NorMAS simulation toolkit should allow the integration of tools that automatically generate network topology. The toolkit should also provide support of various types of topologies that can be used to simulate popular social or computation systems. Such topologies include random, regular, small world and scale-free topologies. Moreover, generating dynamically changing network topology is an important functionality that a simulation toolkit should support in order to facilitate new agents joining, reordering connections or even leaving the network. In the example of P2P, agents might decide to move away from a free rider agent by rewiring the connection with this agent to another agent.

In addition, the definition of multiple topologies that can represent different semantics should be supported. For example, it should allow the definition of a physical topologies, social topologies and observation topologies. A physical topology can refer to agents' physical connection, while social topologies can indicate the existence of multiple social interest topologies (films, music and sport), each represent the relationship and interest of its members. An observation topology can define how agents observe each others behaviour and it can be derived from the constraints imposed by the interaction mechanism. Over the physical network of P2P system, agents might establish different interest network related to their shared interest type of contents.

4.4 The Object Level

The object level is where the basic interaction happens; the norms regulate the object level, normally forbidding or prescribing a course of action. For optimizing agents endowed with an utility function, actions at the object level act on utility directly. For scenarios with fitness parameters (energy, strength, and the such) actions at the object level act on these parameters directly. Finally, for agents with a BDI-like structure, goals will describe an object-level structure. In this case, we have a *pure* object level. Thus, (candidate) norms will be created as restrictions of the agents' autonomy in the operations performed on the object level.

Regarding the examples presented above, the graffiti example would have an object level composed of basic actions as creation of graffiti in the different styles, joining or leaving a crowd, and covering/slashing actions, with reward/penalty. To simulate the object level in the Second Life example, explicit rewards and penalties can be used; but if more details on the internals of the agent are desired, one can imagine agents having as a goal the absence of litter in the park, or, if we start directly at a social level, the goal of being welcomed in a friendly way. In a NorMAS library, these object-level actions would be provided by a set of APIs for fitness modifying behaviors.

The object level will be the target of reasoning at the meta level; in this case, we are talking about a NorMAS meta level. Thus, the constructs at the object level will be examined and recognized by a normative reasoner.

5 Conclusions

This chapter has presented the potential contribution that agent-based simulation techniques and methods can bring to the field of norms, and especially, to norms in multi-agent systems. Based on the descriptions of simulation, the definition of norms with a social science background and the discussion of aspects of norms that are special for artificial agent systems, we have described the relationship between simulation and norms before giving an overview of the norm simulation research centered around the stages of the norm lifecycle presented in figure 1. Following this we presented the research challenges.

First we adressed the methodological challenges. These include the development of measures of spreading of norms, simulation validation, visualisation, and the norm bootstrapping problem. Following this we discussed topological issues. Apart from dynamic versus static networks we also describe the network characteristics such as location influences and the consequences of multiple topologies and multiple groups which an agent can be a member of, causing interacting topologies. While striving for a simulation platform, we finally presented three different scenarios illustrating what we think to be the essential components of such an architecture, that will be the subject of our future work.

References

1 E. Adar and B. A. Huberman. Free riding on gnutella. *First Monday*, 5(10), 2000.

2 R. Albert and A.-L. Barabasi. Statistical mechanics of complex networks. *Review of Modern Physics*, 74(1):47–97, 2002.

3 Giulia Andrighetto, Marco Campenn, Federico Cecconi, and Rosaria Conte. The complex loop of norm emergence: A simulation model. In Shu-Heng Chen, Claudio Cioffi-Revilla, Nigel Gilbert, Hajime Kita, Takao Terano, Keiki Takadama, Claudio Cioffi-Revilla, and Guillaume Deffuant, editors, *Simulating Interacting Agents and Social Phenomena*, volume 7 of *Agent-Based Social Systems*, pages 19–35. Springer, 2010.

4 Giulia Andrighetto, Marco Campennì, Rosaria Conte, and Marco Paolucci. On the immergence of norms: a normative agent architecture. In *Proceedings of AAAI Symposium, Social and Organizational Aspects of Intelligence Washington DC*, 2007.

5 Giulia Andrighetto, Rosaria Conte, Paolo Turrini, and Mario Paolucci. Emergence in the loop: Simulating the two way dynamics of norm innovation. In Guido Boella, Leon van der Torre, and Harko Verhagen, editors, *Normative Multi-agent Systems*, number 07122 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, 2007.

6 Giulia Andrighetto, Daniel Villatoro, and Rosaria Conte. Norm internalization in artificial societies. *AI Communications. (In press)*, 2010.

7 Robert Axelrod. An evolutionary approach to norms. *The American Political Science Review*, 4(80):1095–1111, 1986.

8 Robert Axelrod. An evolutionary approach to norms. *The American Political Science Review*, 80(4):1095–1111, 1986.

9 Tina Balke and Daniel Villatoro. Operationalization of the sanctioning process in hedonic artificial societies. In *Workshop on Coordination, Organization, Institutions and Norms in Multiagent Systems*, 2011.

10 AL Barabasi and E. Bonabeau. Scale-free networks. *Scientific American*, 288(5):60–9, 2003.

11 Cristina Bicchieri. *The Grammar of Society: The Nature and Dynamics of Social Norms.* Cambridge University Press, New York, 2006.

12 Magnus Boman. Norms in artificial decision making. *Artificial Intelligence and Law*, 7(1):17–35, 1999.

13 Jan Broersen, Mehdi Dastani, Joris Hulstijn, Zisheng Huang, and Leendert van der Torre. The BOID architecture. Conflicts between beliefs, obligations, intentions and desires. In *In Proceedings of the fifth international conference on Autonomous agents, Montreal, Quebec, Canada*, pages 9 – 16, 2001.

14 Fabio A. C. C. Chalub, Francisco C. Santos, and Jorge M. Pacheco. The evolution of norms. *Journal of Theoretical Biology*, 241(2):233 – 240, 2006.

15 Rosaria Conte and Cristiano Castelfranchi. Understanding the effects of norms in social groups through simulation. In Nigel Gilbert and Rosaria Conte, editors, *Artificial societies: the computer simulation of social life*, pages 252–267. UCL Press, London, 1995.

16 Rosaria Conte and Cristiano Castelfranchi. Understanding the functions of norms in social groups through simulation. In N. Gilbert and R. Eds Conte, editors, *Artificial Societies: The Computer Simulation of Social Life.*, pages 74–118. UCL Press, 1995.

17 Rosaria Conte and Mario Paolucci. On Agent Based Modelling and Computational Social Science. *Social Science Research Network Working Paper Series*, jul 2011.

18 Jordi Delgado. Emergence of social conventions in complex networks. *Artificial Intelligence*, 141(1-2):171–185, October 2002.

19 Jordi Delgado, Josep M. Pujol, and Ramón Sangüesa. Emergence of coordination in scale-free networks. *Web Intelli. and Agent Sys.*, 1(2):131–138, 2003.

20 Jon Elster. *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences.* Cambridge University Press, 1 edition, apr 2007.

21 J.M. Epstein. Learning to be thoughtless: Social norms and individual computation. *Computational Economics*, 18:9 – 24, 2001.

22 Joshua Epstein, Robert Axtell, and Peyton Young. The emergence of economic classes in an Agent-Based bargaining model. In Steven Durlauf and Peyton Young, editors, *Social Dynamics*. Brookings Press / MIT Press, 2001.

23 Martha Finnemore and Kathryn Sikkink. International Norm Dynamics and Political Change. *International Organization*, 52(4):887–917, 1998.

24 Jack P. Gibbs. Norms: The problem of definition and classification. *American Journal of Sociology*, 70(5):586–594, 1965.

25 R. Guerraoui, K. Huguenin, A. Kermarrec, and M. Monod. On Tracking Freeriders in Gossip Protocols. In *P2P'09: Proceedings of the 9th International Conference on Peer-to-Peer Computing*, 2009.

26 Mathew J. Hoffmann. Entrepreneurs and Norm Dynamics: An Agent-Based Model of the Norm Life Cycle. Technical report, Department of Political Science and International Relations, University of Delaware, USA, 2003.

27 Matthew Interis. On norms: A typology with discussion. *American Journal of Economics and Sociology*, 70(2):424–438, 2011.

28 Sebastian Kaune, Konstantin Pussep, Gareth Tyson, Andreas Mauthe, and Ralf Steinmetz. Cooperation in p2p systems through sociological incentive patterns. In *Proceedings of the 3rd International Workshop on Self-Organizing Systems*, IWSOS '08, pages 10–22, Berlin, Heidelberg, 2008. Springer-Verlag.

29 R. Krishnan, D. M. Smith, Z. Tang, and R. Telang. The impact of free-riding on peer-to-peer networks. In *HICSS '04: Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, page 70199.3. IEEE Computer Society, 2004.

30 S. Mahmoud, J. Keppens, M. Luck, and N. Griffiths. Norm establishment via metanorms in network topologies. In *Proceedings of the 2011 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, volume 3, pages 25 –28, aug. 2011.

31 Richard T. Morris. A typology of norms. *American Sociological Review*, 21(5):610–613, 1956.

32 Declan Mungovan, Enda Howley, and Jim Duggan. The influence of random interactions and decision heuristics on norm evolution in social networks. *Computational &amp; Mathematical Organization Theory*, pages 1–27, 2011. 10.1007/s10588-011-9085-7.

33 M. E. J. Newman. The structure and function of complex networks. *SIAM REVIEW*, 45:167–256, 2003.

34 Anand S. Rao. AgentSpeak(L): BDI agents speak out in a logical computable language. In *Agents Breaking Away: Proceedings of the 7th European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, volume 1038 of *Lecture Notes in Computer Science*, pages 42–55. Springer, 1996.

35 Peter J. Richerson Robert Boyd. *Culture and the evolutionary process.* University of Chicago Press, Chicago, 1985.

36 Bastin Tony Roy Savarimuthu and Stephen Cranefield. Norm creation, spreading and emergence: A survey of simulation models of norms in multi-agent systems. *Multiagent and Grid Systems*, 7(1):21–54, 2011.

37 Bastin Tony Roy Savarimuthu, Stephen Cranefield, Martin Purvis, and Maryam Purvis. Norm emergence in agent societies formed by dynamically changing networks. In *Proceedings of the 2007 IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, IAT '07, pages 464–470, Washington, DC, USA, 2007. IEEE Computer Society.

38 Bastin Tony Roy Savarimuthu, Stephen Cranefield, Maryam Purvis, and Martin Purvis. Role model based mechanism for norm emergence in artificial agent societies. In *Coordination, Organizations, Institutions, and Norms in Agent Systems III*, volume 4870 of *Lecture Notes in Computer Science*, pages 203–217. Springer, Berlin/Heidelberg, 2008.

39 R. Schollmeier. A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications. In *P2P'01: Proceedings of the 1st International Conference on Peer-to-Peer Computing*, pages 101–102. IEEE Computer Society, 2001.

40 Onkur Sen and Sandip Sen. Effects of social network topology and options on norm emergence. In *Proceedings of the Fifth International Conference on Coordination, Organizations, Institutions, and Norms in Agent Systems*, pages 211–222, 2010.

41 Sandip Sen and Stephane Airiau. Emergence of norms through social learning. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1507–1512. AAAI Press, 2007.

42 Yoav Shoham and Moshe Tennenholtz. Emergent conventions in multi-agent systems: Initial experimental results and observations. In *Proceedings of the Third International Conference on the Principles of Knowledge Representation and Reasoning (KR)*, pages 225–231, San Mateo, CA, USA, 1992. Morgan Kaufmann.

43 Yoav Shoham and Moshe Tennenholtz. On social laws for artificial agent societies: off-line design. *Artificial Intelligence*, 73(1-2):231 – 252, 1995. Computational Research on Interaction and Agency, Part 2.

44 Harko Verhagen. *Norm Autonomous Agents*. PhD thesis, Department of System and Computer Sciences, The Royal Institute of Technology and Stockholm University, Sweden, 2000.

45 Harko Verhagen. Simulation of the Learning of Norms. *Social Science Computer Review*, 19(3):296–306, 2001.

46 Daniel Villatoro and Jordi Sabater-Mir. Categorizing social norms in a simulated resource gathering society. In *Coordination, Organizations, Institutions and Norms in Agent Systems IV: COIN 2008 International Workshops, Revised Selected Papers*, pages 235–249. Springer-Verlag, Berlin, Heidelberg, 2009.

47 Daniel Villatoro, Sandip Sen, and Jordi Sabater. Topology and memory effect on convention emergence. In *Proceedings of the International Conference of Intelligent Agent Technology (IAT)*. IEEE Press, 2009.

48 Wikipedia. Glossary of graffiti — Wikipedia, The Free Encyclopedia. `http://en.wikipedia.org/w/index.php?title=Glossary_of_graffiti&oldid=491070208`, 2012. Accessed 8 May 2012.

49 Michael Wooldridge. *An Introduction to MultiAgent Systems*. Wiley, 2nd edition, jul 2009.

The Uses of Norms

Munindar P. Singh[1], Matthew Arrott[2], Tina Balke[3], Amit Chopra[4], Rob Christiaanse[5], Stephen Cranefield[6], Frank Dignum[7], Davide Eynard[8], Emilia Farcas[2], Nicoletta Fornara[8], Fabien Gandon[9], Guido Governatori[10], Hoa Khanh Dam[11], Joris Hulstijn[12], Ingolf Krueger[2], Ho-Pun Lam[10], Michael Meisinger[2], Pablo Noriega[13], Bastin Tony Roy Savarimuthu[6], Kartik Tadanki[14], Harko Verhagen[15], and Serena Villata[16]

1 North Carolina State University, USA
2 University of California at San Diego, USA
3 University of Surrey, UK
4 University of Trento, Italy
5 Vrije Universiteit, The Netherlands
6 University of Otago, New Zealand
7 Utrecht University, The Netherlands
8 Università della Svizzera italiana, Switzerland
9 INRIA Sophia Antipolis, France
10 NICTA, Australia
11 University of Wollongong, Australia
12 Delft University of Technology, The Netherlands
13 IIIA-CSIC, Spain
14 Deutsche Bank, USA
15 Stockholm University, Sweden
16 University of Luxembourg, Luxembourg

Abstract

This chapter presents a variety of applications of norms. These applications include governance in sociotechnical systems, data licensing and data collection, understanding software development teams, requirements engineering, assurance, natural resource allocation, wireless grids, autonomous vehicles, serious games, and virtual worlds.

1998 ACM Subject Classification I.2.11 Distributed Artificial Intelligence

Keywords and phrases Norms, MAS, Governance, Requirements engineering

Digital Object Identifier 10.4230/DFU.Vol4.12111.191

1 Introduction

This chapter presents a compendium of several uses of norms. Each writeup follows a more or less fixed pattern where it first brings out of the application scenario and its importance; second the suitability of a normative model for the problem at hand; third some technical challenges for norms brought to the forefront by that scenario; and fourth a description of its status and some speculation about its prospects. The common notion of norms in these works is that norms represent a standard of correct behavior and correspond loosely to the family of concepts that includes commitments, obligations, and prohibitions.

Normative Multi-Agent Systems. *Dagstuhl Follow-Ups, Vol. 4.* ISBN 978-3-939897-51-4.
Editors: Giulia Andrighetto, Guido Governatori, Pablo Noriega, and Leendert W.N. van der Torre; pp. 191–229
DAGSTUHL FOLLOW-UPS Dagstuhl Publishing
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany

Note that the use of norms reported here are research efforts, in early stages of development. They are inspired by real-life applications and mostly go to demonstrate the potential value of norms in the field. We hope that a collection of these uses will provide some inspiration to researchers in norms and potentially a basis for usage scenarios that might be used in further study or evaluation.

The uses of norms presented next are organized as follows. The contributions by Singh et al., Villata and Gandon, and Fornara and Eynard all deal with norms as they relate to policies in distributed systems. The contributions by Savarimuthu and Dam, Christiaanse and Hulstijn, and Chopra relate norms to software engineering showing how to mine norms, how to map norms to an architecture, and how to base requirements on norms. The contributions by Noriega, Balke, and Governatori and Lam apply norms to modeling scenarios placing agents in real-life settings such as sharing water, wireless connectivity, and physical space (by unmanned vehicles). The contributions by Dignum and Cranefield and Verhagen discuss norms in virtual environments, such as for gaming and virtual worlds.

2 Singh et al.: Governance in Sociotechnical Systems[1,2]

We address the challenge of administering sociotechnical systems, which inherently involve a combination of software systems, people, and organizations. Such systems have a variety of stakeholders, each in essence autonomous. Traditional architectural approaches assume that stakeholder concerns are fixed in advance and addressed out-of-band with respect to the system. In contrast, the sociotechnical systems of interest have long lifetimes with changing stakeholders and needs. We propose addressing stakeholders' needs during the operation of the system, thus supporting flexibility despite change. Our approach is based on normative relationships or norms among stakeholders; the norms are streamlined through a formal notion of organizations. We demonstrate our approach on a large sociotechnical system we are building as part of the Ocean Observatories Initiative.

We define governance as the administration of collaborations among autonomous and heterogeneous peers by themselves. Because each participant is independently implemented and operated, governance must be captured in terms of high-level normative relationships that characterize the expectations that each participant may place on the others. Norms are standards of correctness and may be aggregated into contracts.

Further, our interest lies in sociotechnical systems, which arise in a variety of domains such as scientific investigation, health care and public safety, defense and national security, global business and finance. We define a sociotechnical system as a system-of-systems (SoS). Its value and complexity arise from the combination of capabilities provided by their (heterogeneous) constituent systems.

2.1 Application Scenario

An excellent example of a sociotechnical system is the one being built as part of the NSF-funded Ocean Observatories Initiative (OOI), a thirty-year $400 million project [54]. OOI provides novel capabilities for data acquisition, distribution, modeling, planning and control

[1] The authors of this section are Munindar P. Singh, Emilia Farcas, Kartik Tadanki, Ingolf Krueger, Matthew Arrott, and Michael Meisinger.

[2] Acknowledgments: This work was partially supported by the OOI Cyberinfrastructure program, which is funded by NSF contract OCE-0418967 with the Consortium for Ocean Leadership via the Joint Oceanographic Institutions.

of oceanographic experiments, with the main goal of supporting long-term oceanographic and climate research. The OOI stakeholders include ocean scientists, resource providers, technicians, operators, policy makers, application developers, and the general public.

The OOI presents system requirements that involve supporting thousands of stakeholders, tens of thousands of physical resources such as autonomous underwater vehicles (AUVs), and potentially millions of virtual resources such as datasets. The resources are independently owned and operated. Moreover, OOI facilitates virtual collaborations created on demand to share access to ocean observatory resources, including instruments, networks, computing, storage, databases, and workflows.

The stakeholders have complex needs and objectives in this setting. These include how they can benefit from individual and shared resources, monitor the health of such resources, control their functioning, and administer their usage. Additional considerations include entering into scientific collaborations, managing resource conflicts, achieving and enforcing accountability of colleagues and staff. Importantly, the specifics can differ for each stakeholder individual or organization, and are influenced by whom the stakeholder interacts with. Such concerns are not readily enumerated during design, especially when dealing with long-lived sociotechnical systems. Not treating them would waste opportunities for improving the social and scientific value of oceanographic research. Indeed, this is the current situation and its weakness has motivated the creation of the OOI.

Consider the following important OOI use cases for governance, which highlight the autonomy of the participants and the business relationships among them.

Collaboration. The stakeholders of OOI include research scientists or investigators as well as educators from middle and high schools. Consider a situation where a teacher in a school near Chesapeake Bay would like to present some information about the students' local environment. This data could be as simple as acidity levels in the Bay. Clearly, the teacher would need to access data that a researcher with the appropriate sensors would have gathered. The researcher may have entirely different interests from the teacher; for example, she may be interested in multiyear trends. To this end, the researcher would participate in a resource-sharing community where she would have shared the data streams being generated by her sensors. The teacher would also authenticate with OOI, discover the appropriate community, and enroll in it. Therein the teacher would discover the desirable data stream and extract the information he needs.

Affiliation. The stakeholders of OOI include not only investigators but also research institutions and laboratories. Two institutions may decide to share their resources on a reciprocal basis, and thus enter into a suitable contract, viewed as an aggregation of normative relationships. A researcher at one of those institutions would be able to discover with which institutions her institution is affiliated. She would then be able to access an affiliate institution and further discover a research laboratory based at the second institution. Lastly, she would be able to take advantage of resources belonging to the laboratory. Either institution may decide to end the affiliation but even its exit could be subject to the existing norms, e.g., that ongoing experiments not be aborted.

Existing IT or SOA approaches treat governance primarily as a slow, ponderous, out-of-band activity, whereby stakeholders negotiate their concerns only during the design of a system, not during its operation. Such approaches are ill-suited for specific concerns arising during collaboration. In contrast, automation is essential to improve the quality (such as the precision, timeliness, productivity, and comprehensibility) and scale of governance. For this reason, we approach governance as a central endeavor carried out in-band in a sociotechnical system.

We propose a novel approach that gives first-class status to stakeholders as principals of the resulting system and to their concerns expressed via norms and policies. A policy is what determines a principal's interactions, which may or may not comply with an applicable norm. A norm itself could be operationalized via rules applied by any principal to check if the norm is being respected. Our approach is compatible with traditional approaches, and thus helps leverage existing tools and experience where appropriate.

2.2 Suitability of a Normative Model

Our conceptual model is centered on the concept of principal. Principals include users, resources, and organizations (termed Orgs in our model). Each principal possesses a unique identity within OOI. Governance is achieved through interactions among principals: realized through their local policies and constrained by their normative relationships with each other. Each principal may adopt roles in one or more Orgs. In essence, each role corresponds to a set of norms between a principal who adopts it and the Org (also a principal). The norms constrain the subsequent interactions between two principals present in the same Org. In general, a normative relationship may arise as the result of a successful negotiation or may be implicitly imposed due to the parties adopting complementary roles in the same Org. Each norm references an Org that serves as its context [64].

Each principal is represented in the computational system by an agent. The principal, e.g., a human, exists outside the computational system; the agent is all that exists computationally: there is no principal. Each principal's agent supports bookkeeping regarding the norms in which the principal participates. The agent helps determine (1) if its principal is complying with the applicable norms and (2) if others with whom it deals are complying as well. The agent continually tracks the state of each norm by updating the state for each observable action, such as sending or receiving a message (including making an observation of the environment).

Orgs serve multiple purposes in our architecture, specifically providing a backdrop for norms, a locus for identity, and a venue to share resources. Each Org defines the rules for adopting each of its roles. Joining an Org means adopting at least one role in that Org. Adopting a role means accepting the rules of the Org for that role. Thus, we understand enrollment in an Org as involving the creation of one or more norms and treat the subsequent interactions of the participants as arising within the scope of the given Org. An example of enrollment is someone joining eBay; an example of additional norms is when two eBay members carry out a transaction. The members are subject to eBay's rules such as accepting the price announced by eBay at the end of an auction.

The above interactions, including enrollment, inherently involve the creation and manipulation of normative relationships and can potentially be operationalized in multiple ways. For example, for enrollment, (1) the prospective enrollee may request membership; (2) the prospective enroller may invite the enrollee; (3) a third party may introduce the enrollee and enroller; or (4) a third party may require the enrollee and enroller to carry out the enrollment. Such flexibility facilitates separating stakeholder concerns from each other and from the implementation, thereby improving how stakeholders comprehend the architecture and enhancing the confidence they can place in it.

Each principal applies its own policies to determine what actions to take. Thus, a principal can decide whether to adopt a role in an Org and, conversely, the Org can decide whether to admit a principal to a role. Each principal's decisions are subject to constraints such as the requirements imposed by the roles that it has adopted.

The model from Figure 1 relates an Org specification with a set of roles. Each norm

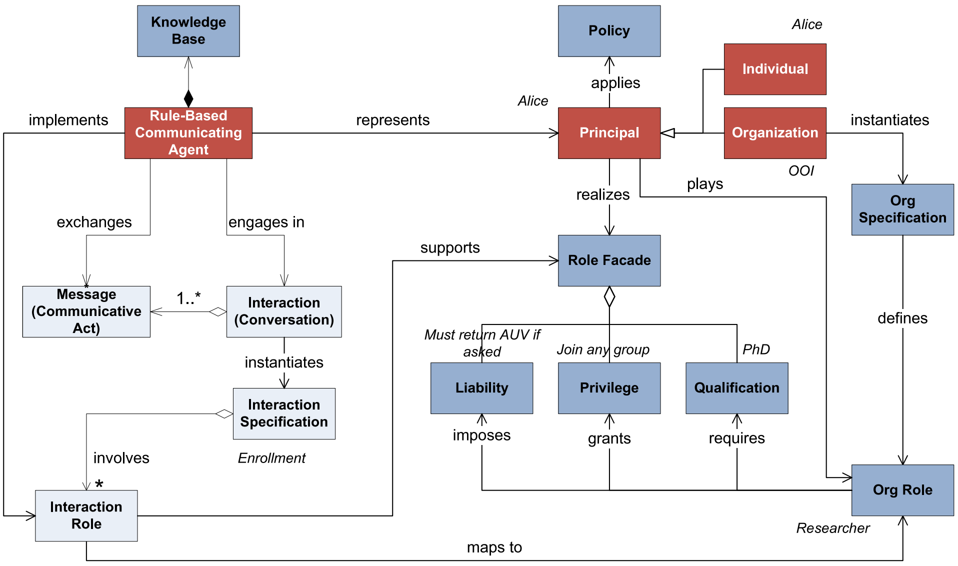


Figure 1 Overview of our governance model.

involves two or more Org roles. In effect, each Org role partitions its view of the relevant parts of the set of norms that characterize the Org. We model the role-relevant parts of each Org specification as consisting of three components, assembled into a *role façade* [65], which helps us provide a normative basis for roles:

Qualifications, which specify eligibility requirements for a principal to take on a role. For example, a professor must have a university identity to join a PhD committee.

Privileges, which specify what authorizations and powers a principal gains in adopting the role. A professor as committee member is authorized to review the student's lab notebook and empowered to determine if the student passes.

Liabilities, which specify what a principal becomes subject to in adopting the role. A committee member must attend a PhD defense.

Each principal applies its policies, to determine whether to enroll, potentially to take advantage of its privileges, and ideally to satisfy its liabilities. In general, one cannot guarantee compliance in a sociotechnical system, but we address compliance in two main ways:

- Conservatively, ensure that actions taken by a principal are compliant. This can work where the principal is not autonomous and heterogeneous. We can subject the principal to a guard that allows only the policy-compliant (attempted) actions to proceed.
- Optimistically, recognize that a principal may proceed as it would, but detect and handle noncompliant behavior. We can accomplish detection either by introducing a monitor in the architecture or through the principals monitoring each other. We can respond to detected violations by escalating them to the nearest scoping Org.

2.2.1 Architectural Case Study

OOI enables its primary stakeholders (scientists) the opportunity to seamlessly collaborate with other scientists across institutions, projects, and disciplines and to access and compose resources as needed. To address complexity, mitigate risks, and accommodate requirement

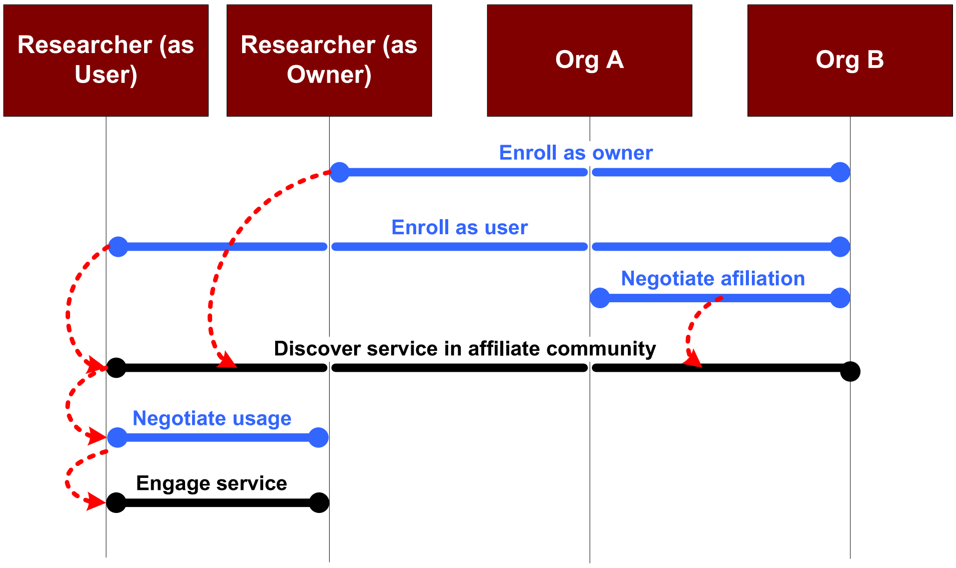


Figure 2 Governance of resource sharing across affiliated Orgs.

changes, OOI uses a spiral development process, a variant of the Incremental Commitment Model (ICM) [7]. ICM includes iterative development cycles focusing on incremental refinement of system definition and stakeholder commitment and satisfaction. We have adopted selected architectural views from the Department of Defense Architecture Framework (DoDAF) [23] to document the OOI architecture.

OOI resources are distributed both physically and virtually among different organizations, each with its own policies for resource access and data delivery or consumption. We model OOI itself as an Org that is the highest scope for all OOI users and their interactions. The OOI Org serves as the root Org for the identities for all OOI principals and helps monitor and enforce the applicable norms among them.

Figure 2 illustrates the use case where two research organizations (each an Org) form an affiliation relationship with each other. Both Org A and Org B are what we term resource-sharing Orgs, and define two main roles: owner (of a resource) and user (of a resource). Each principal who adopts owner can contribute its resources to the Org, so those resources can be discovered by any principal who adopts the role user. In addition, to form affiliations, each Org supports additional roles capturing the affiliation relationship. These roles are AFFILIATEORG to capture the clauses for the affiliated community, and AFFILIATEMEMBER to capture the clauses for the members of the affiliated community, which could have weaker rights than the Org's own members.

The affiliation relationship between Orgs propagates to their respective members. As a result, a member of Org A can discover services offered by members of Org B. Once it has discovered such services, it may negotiate with and engage them as appropriate.

Our notation is similar to message sequence charts in terms of having a swim lane for each principal. However, instead of messages, we use horizontal lines to show joint (governance) actions that create or modify relationships among the (two or more) parties whose lifelines they connect. Any temporal order requirements are captured via the dashed arrows that connect some pairs of the horizontal lines. In general, the parties would realize a governance interaction such as enrollment by exchanging multiple messages, e.g., propose, counterpropose, accept, or reject.

2.2.2 Benefits of Norms and Allied Constructs

We attribute the benefits of our architectural treatment of governance to the following main principles that it respects.

- Centrality of organizations in modeling communities; modeling the OOI itself as an runtime entity or agent; specifying rules of encounter; monitoring norms; sanctioning violators.
- Autonomy of stakeholders; representing stakeholders as agents that apply autonomous policies and are subject only to the applicable organizational rules of encounter.
- Emphasizing normative relationships and modeling them explicitly to make them easy to inspect, share, and manipulate; accommodating openness of the system by recognizing that autonomous parties may violate rules of encounter and, thus, may need enforcement ex post facto, such as via sanctions.

In the OOI, policies specified in the norms within an Org govern the circumstances under which resources can be discovered, accessed, and utilized. In the example, we considered two classes of stakeholder roles: user and owner of a resource. The user is concerned with accessing a resource, without facing any hidden obligations. The owner is concerned with providing resources (with spare capacity) to expand the impact of the resources on others and to treat the resources as a basis for negotiating value in exchange.

Our governance approach addresses stakeholder (user and owner) concerns as follows:

- The resource sharing community provides access to needed resources and clarifies what restrictions are imposed on the user as a result; guarantees that the user will not be subject to the whims of the resource owner once the user begins an allowed interaction with a resource.
- The affiliation community expands resource sharing to external organizations and provides access to remote resources on a reciprocal basis.
  - The user and owner can negotiate detailed terms as norms that go beyond the basic norms imposed by being members of a community.
  - The user and owner can accommodate changing needs, renegotiate the set of norms, or may decline to continue to participate.
  - In deployment, policies are separated from the business functionality, allowing them to be changed easily over time according to stakeholder needs.

2.3 Challenges for Norms

The above exercise makes apparent some important challenge for norms.

An engineering challenge is to harmonize norms with approaches to software architecture and methodology, so that norms can be naturally incorporated into practice.

- The OOI effort builds upon methodologies such as Model-Driven Architecture (MDA) and goal-oriented requirements engineering, and goes beyond them by providing a systematic treatment of governance from the modeling level to implementation. We understand sociotechnical systems to be ultra-large-scale systems (ULSSIS) because they inherently involve multiple stakeholders use the system, contribute resources, form virtual communities, and determine the rules that govern their interactions [27]. Our approach applies naturally to ULSSIS because it dynamically captures stakeholder concerns by (1) defining patterns of interaction based on Orgs; (2) enabling stakeholders to select roles in desirable Orgs; and (3) supporting the specification and application of policies potentially customized to each stakeholder. How can ULSSIS incorporate norms in general?

- Addressing the inherent complexities of sociotechnical systems involves going beyond traditional Service-Oriented Architecture (SOA), specifically in accommodating multiple ownership domains [26]. Following Singh et al. [66], we view services as real-life services, not computational objects. We identify principals as the participants in service engagements described in terms of the normative relationships, and define patterns on the creation, propagation, and manipulation of such norms. Our approach coheres with recent advances in goal-oriented methodologies, specifically Tropos [13]. Tropos emphasizes the goals of the actors whereas we emphasize their norms and would capture their goals both in what norms they enter and how they choose to act based on those norms. How may we expand the above-mentioned normative patterns and place them within a comprehensive methodology for developing normative systems?

Theoretical challenges are highlighted by this effort.

- To use norms for specifying sociotechnical systems, we would need algorithms that help validate normative specifications, so as to identify conflicts early in development.
- We need techniques to determine whether a principal complies with applicable norms, especially using information that other principals can access.

2.4 Status and Prospects

The OOI development effort is underway. At a conceptual level, normative thinking has guided the software architecture right from the beginning of the OOI. The early part of its development effort has dealt with providing the software infrastructure to realize scientific collaboration. Normative concepts are now being introduced into the development.

We apply the Rich Services architecture [3], a type of SOA that provides decoupling between concerns and allows for hierarchical service composition. Rich Services is a logical architecture that can be mapped to possible deployments such as Enterprise Service Buses or multi-agent systems. For the affiliation use case in OOI, each Org and the User itself are modeled as a Rich Service within the root OOI Rich Service. Infrastructure Services include identity and policy management, logging of all conversations and actions, as well as repositories for the community specification and the norms already established with other parties. Each Rich Service has its local policies and a local representation of the norms in which it participates.

Rich Services provide a clear separation between the business logic and its external constraints, supporting their composition at the infrastructure level through specialized interceptors. When requirements change during the lifetime of the system, they often affect policies and not core services; therefore, the decoupling between them allows to update Infrastructure Services without modifying the services that are composed.

A specific implementation of governance may be realized via a rule-based communicating agent, which maintains the applicable rules and information about the state of the world and any ongoing interactions in a knowledge base. Each agent represents one principal—thus the approach is decentralized. An agent represents a principal in an Org as a locus of autonomy and identity. We have prototyped such an agent using an agent platform (specifically, Java Agent Development Framework (JADE)) and a rule engine (specifically, Java Expert System Shell (JESS)). An agent platform provides a container for the execution of agents, communication infrastructure to enable agent communication, and directory services. A rule engine maintains and applies the facts and rules for an agent and, thus, enables reasoning and reaction.

Rules lead to a simple implementation where an agent loads the rules corresponding to each role that it adopts. The rules are generated from the norm specifications for which we developed a domain Specification Language; its constructs are based on properties and predicates.

3 Villata and Gandon: Data Licensing in the Web of Data

3.1 Application Scenario

A common assumption in the Web is that the publicly available data, e.g., photos, blog posts, videos, can be reused without restrictions. However, this is not always true, even when the licensing terms are not specified. Consuming Linked Open Data includes the fact that the data consumer has to know the terms under which the data is released. The licensing terms in the Web of Data are specified by means of machine-readable expressions, such as additional triples added to the RDF documents stating the license under which the data is available. We highlight the future trends in data licensing and the possible connections with normative multi-agent systems.

The JISC Linked Data Horizon Scan[3] states about the link between Linked Data and Open Data: “Linked Data may be Open, and Open Data may be Linked, but it is equally possible for Linked Data to carry licensing or other restrictions that prevent it being considered Open, or for Open Data to be made available in ways that do not respect all of Berners-Lee’s rules for Linking.”[4]. Licensing of data needs to be explicit to avoid any ambiguity in terms of use and reuse for the data consumers. The absence of clarity for data consumers about the data terms of reuse does not encourage the reuse of that data. There are many differences worldwide related to the copyright of data, and not all data is copyrightable. Some of the most popular licenses on the Web include Creative Commons[5], GNU Free Documentation License[6], Open Data Commons[7], Science Commons Database Protocol[8], and Freedom to Research: Keeping Scientific Data Open, Accessible, and Interoperable[9].

The Linked Data cloud[10] presents various examples of use of different licensing terms. Table 1 shows the absence of a common approach for data licensing in the Web of Data. This is one of the main problems in the context of licensing for the Web of Data [6].

Heath and Bizer [40] underline that “the absence of clarity for data consumers about the terms under which they can reuse a particular dataset, and the absence of common guidelines for data licensing, are likely to hinder use and reuse of data”. Therefore, all Linked Data on the Web should include explicit license, or waiver statements, as discussed also by Miller et al. [51]. In this paper, we briefly introduce the licenses schemas proposed in the Web of Data, then we describe the open challenges in data licensing for the Web of Data. We conclude by suggesting some further challenge bridging the gap between the Semantic Web and Normative Multi-Agent Systems (NorMAS).

[3] `http://linkeddata.jiscpress.org/`
[4] `http://wiki.cetis.ac.uk/images/1/1a/The_Semantic_Web.pdf`
[5] `http://creativecommons.org/`
[6] `http://www.gnu.org/copyleft/fdl.html`
[7] `http://www.opendatacommons.org/licenses/`
[8] `http://sciencecommons.org/resources/faq/database-protocol`
[9] `http://sciencecommons.org/wp-content/uploads/freedom-to-research.pdf`
[10] `http://richard.cyganiak.de/2007/10/lod/`

Table 1 Examples from the Linked Data cloud and their licenses.

	CC	ODC	Country	GNU	Commercial	No licenses
MusicBrainz				X		
Guardian Data Store						X
OpenStreetmap	X					
BBC Backstage					X	
DBpedia	X			X		
legislation.gov.uk			X			

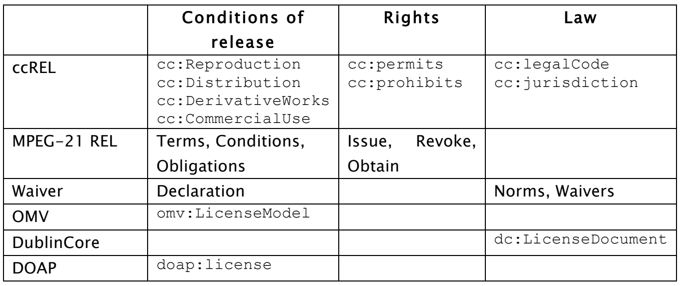
3.2 Suitability of a Normative Model

The applications that consume data from the Web must be able to access explicit specifications of the terms under which data can be reused and republished. The availability of appropriate frameworks for publishing such specifications is an essential requirement in encouraging data owners to participate in the Web of Data, and in providing assurances to data consumers that they are not infringing the rights of others by using data in a certain way. Initiatives such as the Creative Commons have provided a framework for open licensing of creative works, underpinned by the notion of copyright. However, as discussed by Miller et al. [51], copyright law is not applicable to all data because not all data are creative works, which from a legal perspective is also treated differently across jurisdictions. Therefore frameworks such as the Open Data Commons can be adopted to state the revise conditions.

The most diffused machine-readable licensing languages are Creative Commons, Open Data Commons, and MPEG-21 REL. The Creative Commons Rights Expression Language (ccREL) [1] is the standard recommended by Creative Commons (CC) for machine-readable expression of copyright licensing terms and related information. Miller et al. [51, 69] propose the Open Data Commons waivers and licenses[11] that try to eliminate or fully license any rights that cover databases and data. The Waiver vocabulary[12] defines properties to use when describing waivers of rights over data and content. A waiver is the voluntary relinquishment or surrender of some known right or privilege. As discussed by Heath and Bizer [40], "licenses and waivers represent two sides of the same coin: licenses grant others rights to reuse something and generally attach conditions to this reuse, while waivers enable the owner to explicitly waive their rights to a dataset". In MPEG-21, a Rights Expression Language (REL)[13] is a machine-readable language that can declare rights and permissions using the terms as defined in the Rights Data Dictionary[14]. Two further vocabularies which can be used also to define the licensing terms of the data on the Web are the Description of a Project vocabulary (DOAP)[15], and the Ontology Metadata vocabulary (OMV)[16] [55]. The former is an RDF/XML vocabulary to describe software projects, in particular open-source. It defines a property `doap:license` for defining the licensing terms of the project. The latter, instead, describes a particular representation of an ontology, and it captures the key aspects of the

[11] http://opendatacommons.org/licenses/
[12] http://vocab.org/waiver/terms/.html
[13] http://mpeg.chiariglione.org/standards/mpeg-21/mpeg-21.htm
[14] http://iso21000-6.net/
[15] http://usefulinc.com/ns/doap
[16] http://omv2.sourceforge.net/index.html

ontology metadata information, e.g., provenance, availability, statistics. OMV defines the property `omv:hasLicense` which provides the underlying license model. Moreover, OMV introduces a class `omv:LicenseModel` which describes the usage conditions of an ontology. Finally, we mention also the Dublin Core license document class `dc:LicenseDocument`[17] which provides the legal document giving official permission to do something with the resource and the license property `dc:license`[18]. A mapping among the concepts used in these schemas is provided in Figure 3.



	Conditions of release	Rights	Law
ccREL	cc:Reproduction cc:Distribution cc:DerivativeWorks cc:CommercialUse	cc:permits cc:prohibits	cc:legalCode cc:jurisdiction
MPEG-21 REL	Terms, Conditions, Obligations	Issue, Revoke, Obtain	
Waiver	Declaration		Norms, Waivers
OMV	omv:LicenseModel		
DublinCore			dc:LicenseDocument
DOAP	doap:license		

Figure 3 Mapping among the different licensing languages.

Licenses such as Creative Commons and Open Data Commons have common features, but also differ from each other. The requirement to mention the author (BY) seems to be one of the best shared features, since it is absent only in the license PDDL 1.0[19]—which in essence, exempts any obligation. Most legal frameworks allow commercial use: that is, they make it possible for re-users to sell public data without transforming or enriching them. Other features are adopted by some legal framework, and not by others.

Figure 4 visualizes the results of a search on Watson[20] of the licensing terms we have presented. The results show that the Creative Commons Attribution license is the most used one among the Creative Commons licenses. The other well diffused way to express licenses adopts the Dublin Core `license` property. These results make even clearer the lack of a uniform approach to data licensing.

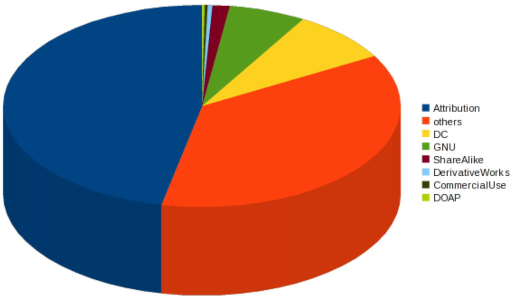


Figure 4 The use of licenses in the Web (Watson search).

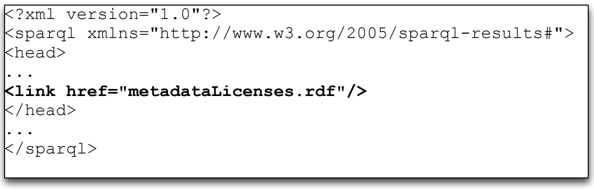
[17] http://dublincore.org/documents/dcmi-terms/#classes-LicenseDocument
[18] http://dublincore.org/documents/dcmi-terms/#terms-license
[19] http://www.opendatacommons.org/licenses/pddl/
[20] http://watson.kmi.open.ac.uk/WatsonWUI/

3.3 Challenges for Norms

The challenge about treating licenses in the Web of Data can be decomposed in a number of sub-tasks as follows:

1. Selection of n license schemas;
2. Alignment of these n license schemas;
3. Returning the requested data together with the license under which it is released.

Each of the points above presents a challenge in the research area of the Semantic Web. First, the selection of the n license schemas is a complex task since the vocabularies are not all indexed and the licensing terms of some data may be expressed by various vocabularies, and not only by ad-hoc vocabularies. Second, these vocabularies may define concepts and properties about licenses which have been already defined by other vocabularies. At this point, an alignment step is necessary to establish which are the "equivalent" licensing terms defined in the various vocabularies. Third, the aim of a license model for the Web of Data is to provide, after a user query, the data resulting from the query together with the licensing terms under which the data is available.



```
<?xml version="1.0"?>
<sparql xmlns="http://www.w3.org/2005/sparql-results#">
<head>
...
<link href="metadataLicenses.rdf"/>
</head>
...
</sparql>
```

Figure 5 A sample of the SPARQL query results XML format providing information about the licenses on the data.

A possible solution concerning the third point would be to adopt the standard SPARQL query results XML format[21], and to introduce thanks to the `<link>` element the information about the license under which the data returned by the SPARQL query is released. The problem which arises at this point is that we cannot express more than one license on the data. Thus we should choose the more restrictive license among the set of licenses constraining the data returned by the query. If this solution is adopted, this leads to a lack of satisfaction of the less restrictive licensing terms which allow, for instance, a free distribution and reuse of the data. The possible alternative consists in changing the SPARQL query results XML format in order to associate to different sets of data different licenses. This would allow a better representation of the licensing terms, but this includes also the definition of a new standard for the SPARQL query results XML format.

The Linked Data initiative aims at improving data publication on the Web, thereby creating a Web of Data: an interconnected, distributed, global data space. The Web of Data enables people to share structured data on the Web as easily as they can currently share documents on the Web of Documents (WWW). The basic assumption behind the Web of Data is that the value and usefulness of data increases with the amount of interlinking with other data. The emerging Web of Data includes datasets as extensive and diverse as DBpedia, and DBLP. The availability of this global data space creates new opportunities for the exploitation of techniques in relation with knowledge representation and intelligent agents. In particular,

[21] http://www.w3.org/TR/rdf-sparql-XMLres/

a new challenge in this view consists in having intelligent agents exploiting the Web of Data. This is a challenge which involves the NorMAS community as well concerning for instance the licenses issue. In particular, normative multi-agent systems may be adopted to support the alignment phase among the different licenses schemas. An open issue in ontology matching is how to have a consistent alignment. For instance, Santos and Euzenat propose a model based on argumentation theory [24]. An idea would be to use techniques developed in the field of normative multi-agent systems to check the consistency of an alignment of licensing schemas, following the approach proposed by Fornara and Colombetti for obligations [33]. Moreover, the reasoning techniques developed in the NorMAS community can be used to reason over the Web of Data on order to infer, starting from the links among the different schemas and datasets, further normative constraints among the datasets and further links among the (licenses) schemas.

4 Fornara and Eynard: Web-based Data Collection using Norms and Semantic Web Technologies[22]

4.1 Application Scenario

Web-based data collection is becoming more and more important for many social science fields like economy, sociology, social media, market research, and psychology. This fact is clearly highlighted for example by the WEBDATANET COST Action[23]. Web-based data collection is not restricted to Web surveys, but it also includes *non-reactive data*, collected by means of log files analysis, data mining, text mining, and data crawling from heterogeneous Web sources (i.e., blogs, social networks, consumer reviews, folksonomies, and search results). These approaches require new techniques, algorithms, and tools whose application to the problem of Web-based data collection represents a crucial multidisciplinary problem, involving both social scientists and computer engineers.

The design of the tools for collecting non-reactive Web data is strongly infuenced by the perspectives, the constraints, and the desires of the different actors involved in the creation, publication, collection, storage and manipulation of those data. In particular it is crucial to take into account: (i) the point of view of those who will analyse these data, whose requirements are data validity, reliability, quality and, as discussed in [43], the need to be able to guarantee integrity, transparency, and to respond to privacy and confidentiality wishes of the users; (ii) the perspective of data providers (i.e., data publishers and companies or single users that produce those data), who consider essential the possibility to constrain the access to their data, together with the possibility of being aware of how they will be stored, used, and combined. Guidelines for professional ethics and practices can be found, for example, in [30]. Some of them express norms, that is, obligations ("we shall") and prohibitions ("we shall not") on how data can be obtained, stored, used, disclosed, and so on. Examples policies stating how the resources available on a social software like Facebook can be used for automatic data collection can be found at `http://www.facebook.com/apps/site_scraping_tos_terms.php`. Other examples of constraints on how data –and as a particular case Linked Data– can be accessed and reused by consumers are represented by popular licenses like Creative Commons and GNU Free Documentation License, plus many others referred and discussed in detail in the previous section.

[22] Acknowledgments: Dr. Fornara is supported by the Hasler Foundation project nr. 11115-KG and by the SER project nr. C08.0114 within the COST Action IC0801 Agreement Technologies.

[23] `http://webdatanet.cbs.dk/`

4.2 Suitability of a Semantic Normative Model

Very often these guidelines, norms, policies, and constraints are expressed in natural language (e.g. English). Therefore, in order to comply with them researchers need to read, understand, and finally apply them. The problem of applying these norms is complicated by the fact that top-down policies (provided by data publishers and users) also need to be integrated by additional (bottom-up) constraints, provided by data collectors and declaring what, within a set of Web sources, can or cannot be accessed in the context of a given research. Moreover, the licenses that regulate access to datasets are often different from one another. Finally, a relevant aspect of the problem is that this application scenario involves different actors (social researchers, data collectors, data publishers, and data producers) having different interests. Therefore there is always the possibility that some of those norms are not fulfilled, thus it is important to implement mechanisms for their monitoring and enforcement.

When big amounts of data are treated for automatic extraction by means of specialized software (i.e., both site-specific and generic crawlers, or survey software), being compliant with those norms becomes very difficult. This problem even worsens when data to be used in one survey are collected from many sources by different people and organizations, with different techniques, and in different instants of time. *Formal models* of guideline, policies, and norms may be used to guide specialized software with a specific focus on (i) guaranteeing that activities performed during data collection are compliant with given norms, (ii) guaranteeing that the way those data are stored and re-used is compliant with given policies and guidelines, and (iii) providing a way to keep track on how data is accessed and used and issue warnings when a potential misbehavior, with respect to a set of norms, occurs.

Studies on normative multi-agent systems (NorMAS) and on automatic extraction and representation of knowledge from semi-structured and unstructured data may be crucial to tackle those problems. Studies on the formalization of obligations, permissions, and prohibitions in particular, and of agreements[24] and contracts in general [21] may be used to express norms or licenses (e.g. in the Web of Data as discussed in Section 3) for accessing, using, and storing data. Moreover if those norms are expressed and manipulated using Semantic Web Technologies, [41] like OWL as proposed in [32, 31] and in [63], it will be possible to easily merge sets of norms coming from different sources (by merging their OWL ontologies) and use techniques for OWL ontologies alignment (as discussed in the previous section) for solving possible differences and for checking the consistency of the resulting set. Similarly techniques for the monitoring of those semantic norms may be used to develop software able to issue warning when violations occurs. Studies on mechanisms for developing software that are compliant with a given set of norms [35] or on mechanisms for developing agents able to plan their actions on the basis of certain norms [19] can be used to develop software agents able to reason on semantic norms.

Those formal representation of norms can be viewed as formal data attached to semi-structured and unstructured data that is published on the Web on a daily basis. The problem of structuring knowledge can be addressed in two different ways. On the one hand, knowledge representation standards and techniques could be adopted to provide knowledge in a form that is directly consumable by machines. These techniques mainly relate to the use of Semantic Web Technologies [41] like RDF and OWL for knowledge representation. On the other hand, given the amount of data already provided on the Web as semi-structured or unstructured text, the study of tools and techniques for automatic knowledge extraction from these sources is becoming more and more important.

[24] http://www.agreement-technologies.eu/

4.3 Challenges for Norms

The main challenges for norms formalization and monitoring related to the proposed application scenario are:

- The development of models and techniques to express Web-based data collection guidelines, rules and policies at different level of abstraction, that is, representing high-level general guidelines and transforming them in low-level concrete descriptions of allowed and disallowed actions;
- Given that those rules need to be automatically processed by software tools, studying how to formalize them using decidable logics, like for example the Description Logics (DLs) that are at the basis of the Ontology Web Language OWL;
- The design of systems able to plan their actions on the basis of a set of semantic norms;
- The design of systems able to keep track of how data are accessed and used in an intrinsically partial observable world like the Internet is and raise warnings when inconsistencies among the expected behavior and the actual behavior arise.

4.4 Status and Prospects

The idea of applying semantic norms to the formalization of guidelines for Web data collection and of norms and policies for regulating how, where, and from who those data may be collected is underway. We plan first of all to study models and techniques to express Web-based data collection guidelines, norms and policies at different levels of abstraction, from very high-level general guidelines to low-level concrete descriptions of allowed and disallowed actions. To this purpose we plan to extend our model of obligations [31] and norms [32], expressed using Semantic Web technologies. Based on such models and techniques we plan to design and implement a demonstrative system able to monitor processes of Web data collection, in order to guarantee that data collection guidelines, norms and policies are actually satisfied.

5 Savarimuthu and Dam: Norms in Open Source Software Repositories

5.1 Application Scenario

Extracting valuable information from Open Source Software (OSS) repositories is gaining popularity since huge volume of data is available for free. Open source projects such as Linux and Andorid OS are used by millions and developed by hundreds of developers over extended period of time have produced rich, extensive, and easy-to-access data from which valuable information can be *mined*. We believe open source repositories present an interesting context for the extraction and the study of norms. The reasons are manifold. Firstly, there are a substantial number of OSS projects in various sizes (ranging from a few developers to hundreds of contributors), different coding cultures (e.g., Java vs. C), or different application domains (browsers vs. operating systems). These projects would allow us to understand *how norms emerge and how they are enforced in different settings*. Secondly, OSS projects involve communication and coordination of contributors from different backgrounds, cultures and geographical regions, which makes OSS an exciting domain for exploring *how norms affect the success or failure of a particular software project or community*. Thirdly, such rich, extensive and readily available data from OSS projects allow us to *extract norms from different sources*. For instance, we can directly observe developer discussions, identify their contents (e.g., patches, bugs, reviews) on mailing lists or forums. We can build social networks, and cross-check associated discussion and programming activity. In addition, we can leverage existing mining software repositories (MSR) technologies [39][25] such as data preprocessing, cross-linking data from different sources for mining norms.

5.2 Suitability of a Normative Model

We believe the techniques and tools developed by NorMAS researchers to identify and extract norms can be leveraged and extended. Researchers in NorMAS [60, 61] have developed mechanisms for extracting norms from agent interaction data. Open source software repositories available in various forms such as historical repositories (e.g., SVN or CVS repositories, archived communication records, bug repositories), code repositories (e.g., Sourceforge.net or Google code), and run-time repositories (e.g., deployment and/or execution logs) remain largely unexplored in the context of norms. Since these repositories are populated by humans, these repositories contain explicit or implicit information on *norms* relevant to the communities involved in the process of software development. These repositories need to be *mined* using the techniques developed in the NorMAS community to uncover useful and important patterns and information about the normative processes associated with the development of software systems. Such information might offer insights and predictions about the future of software systems.

5.3 Challenges for Norms

This subsection discusses the research opportunities for NorMAS researchers in applying the concepts and mechanisms developed to extract different types of norms from software repositories and the associated challenges.

5.3.1 Challenge 1: Norm Types and Classification

The first challenge is to answer the question of *what types of norms exist in open source software development communities.* Several research work in NorMAS have treated both conventions and norms under the same umbrella of norms despite the differences between the two. We briefly discuss the distinction between the two using the examples from Open Source Software Development (OSSD).

Conventions of a community are the behavioural regularities that can be observed. Coding standards of a project community is an example of a convention. The specifications of these conventions may be explicitly available from the project websites[26] or can be inferred implicitly (e.g., a wide spread convention that may not be explicitly specified in project websites).

Norms are conventions that are enforced. A community is said to have a particular norm, if a behaviour is expected of the individual members of the community and there are approvals and disapprovals for norm abidance and violation respectively. There have been several categorizations of norms proposed by researchers (c.f. [59]). We believe that *deontic norms* - the norms describing prohibitions, obligations and permissions studied by the NorMAS community [73] is an appropriate categorization for investigating different types

[25] A extensive review of the work in MSR can be found from the “Bibliography on Mining Software Engineering Data” available at `http://ase.csc.ncsu.edu/dmse`.

[26] Refer to `http://source.android.com/source/code-style.html` for the coding guidelines for Android development.

of norms that may be present in OSSD communities. We believe most norms in software repositories will either be prohibitions or obligations.

Prohibition norms: These norms prohibit members of a project group from performing certain actions. However, when those prohibited actions are performed, the members may be subjected to sanctions. For example, the members of an open source project may be prohibited to check-in code that does not compile, and they may be prohibited to check-in a revised file without providing a comment describing the change that has been made.

Obligation norms: Obligations describe activities that are expected to be performed by the members of a project community. When the members of a community fail to perform those, they may be subjected to sanctions. For example, the members may be expected to follow the coding convention that has been agreed upon. Failure to adhere to this convention may result in the code not being accepted by the repository (e.g., based on automatic checking) or a ticket may be issued by a quality assurance personnel. Another obligation may be that the members should complete a task within a time frame. Failure to do so may result in a warning message (issued either automatically or manually) in the first instance.

We note that recognizing sanctions (a starting point to infer norms) is a key challenge since it involves natural language processing. Verbose text may be used in the construction of sanction messages. For example, the messages may involve terms that are well beyond the deontic terms such as 'should not', 'must not', 'ought not' in the case of prohibitions. One way to address this problem is to use existing tools such as WordNet [28] to extract synonyms of terms used in the text to infer deontic terms and also use information retrieval tools that offer data manipulation functions such as cleaning and disambiguating the verbose text in order to extract sanctions. Suitability of tools such as OpenCalais (`http://www.opencalais.com`) and AlchemyAPI (`http://www.alchemyapi.com`) for this purpose can be investigated. We believe recognizing sanctions is indeed a huge challenge. At the same time, it presents opportunities such as the construction of normative ontologies that can be used across projects for recognizing sanctions.

5.3.2 Challenge 2: Norm Identification

In NorMAS, researchers have proposed a life-cycle for norms [59] which broadly consists of five phases namely norm creation, identification, spreading, enforcement and emergence. Chapters 5 and 6 of this volume also presents a similar norm life-cycle model. We believe various phases of norm development can be studied based on the data available from software repositories. Specific research challenges in the context of mining software repositories for norms are given below.

- What are the modes of norm creation in a project community?
- How are prespecified norms enforced? What kinds of sanctions exist for norm violations? What is the uptake of a norm in a community (i.e., level of conformance)?
- How can emergent norms be detected? How are these norms spread? What contributes to the acceptance or rejection of these norms in a community?

5.4 Status and Prospects

This section offers some initial thoughts on addressing the research questions described in Section 5.3.2. It also reports the progress made by relevant research works.

5.4.1 Modes of Norm Creation

There could be two modes for norm creation in a software development community. They are 1) explicitly specified norms which every project member is expected to know (prespecified norms) and 2) norms that arise due to interactions between agents (emergent norms).

5.4.2 Enforcement of Prespecified Norms

In a project, both conventions and norms may exist. Conventions agreed upon by project members can be easily monitored. Examples include coding conventions and the convention of not uploading files that do not compile to a version control system. It should be noted that coding conventions can be checked for compliance by evaluating the code using an automated software program such as CheckStyle[27].

Norms on the other hand are enforced. Enforcement involves the delivery of appropriate sanctions. In the domain of software repositories these sanctions are present in artifacts. For example, a bug report on a module that does not deliver the functional requirements can be viewed as a sanction. Additionally, tickets issued for not resolving a bug completely can also be considered as a sanction. Therefore, sanctions that follow violations act as triggers to infer norms. Frequency of norm violations over time may provide evidence for the uptake of a norm in a society. We note that identifying and categorizing different types of sanctions from different types of artifacts is a challenge since the extraction of sanctions involves natural language processing.

5.4.3 Identifying Emergent Norms

Norms that are not prespecified but that emerge at run-time will be challenging to identify. We believe that emergent norms can be identified by identifying violations first and then inferring what the norms might be. The machinery proposed for norm identification by Savarimuthu et al. [60, 61] can be used as a starting point to infer prohibition and obligation norms. In their work, prohibition norms are identified by extracting sequence of action (or actions) that could have caused the sanction by using a data mining approach [60]. Sanctions form the starting point for norm identification. In the case of obligation norms, missing event sequence (or sequences) that was responsible for the occurrence of a sanction, is identified [61]. While these work on norm identification can be used as a starting point for the extraction of emergent norms in simple cases, the domain of MSR poses more challenges. For example, correlating or linking different types of documents containing relevant information is required before a sequence of actions can be constructed. For example, an email message may contain the sanction message exchanged between developers A and B. Let us assume that A sanctions B for not adding a valid comment to the latest version of the uploaded file. The problem in extracting the norm in this case is that, first, the verbose message sent from A to B should be understood as a normative statement which involves natural language processing. Second, a cross-check should be conducted to evaluate whether the normative statement is indeed true (i.e., checking whether the comment entered by B is invalid by investigating the log)[28]. Third, the support for endorsements or oppositions to such normative positions need to be evaluated in order to extract this as a norm.

[27] `http://checkstyle.sourceforge.net/`

[28] In this example only two artifacts, the email message and the log are involved. But in practice, several different types of documents may need to be traversed to find the relevant information. Techniques developed in the field of MSR (e.g., [5], [53]) can be employed for cross-linking documents.

Norms that are identified through this process can then be made available to the project community (e.g., on the project websites) once it has been verified by the project administration team.

5.4.4 Need for a Norm Extraction Framework

A first step towards addressing these issues is to create a framework that can extract both conventions and norms from software repositories. The framework should be equipped with appropriate libraries for a) information retrieval techniques (including natural language processing) in order to identify sanctions b) mining software repositories (e.g., cross-linking different sources) and c) norm extraction (e.g., inferring norms from sequences of events). Additionally, it should be able to track and trace the life-cycle of a norm. For example, it should provide appropriate features to capture the waxing and waning of a norm across different periods of time.

5.4.5 Cross-Disciplinary Research Questions

The following are interesting research questions that can be considered in the future.

- How different are norms in large projects (e.g., measured based on total number of members or size of the project in kilo-lines of code) than the smaller projects? Are norm violation and enforcement patterns different in these projects?
- What are the relationships between roles of individuals in software development and norms (e.g., contributor vs. reviewer vs. module administrator)?
- Are there cultural differences within members of a project with regards to norms (inter- and intra-project comparisons) since individuals from different cultures may have different norms?
- Is there a difference between norm adoption and compliance between open-source and closed-source projects?

The above mentioned questions may interest both humanities researchers and computer scientists. Synergy between the two is required for addressing these questions. As computer scientists we can employ our expertise in several areas (i.e., normative multi-agent systems, information retrieval and MSR) to help answering these questions.

6 Christiaanse and Hulstijn: Automation of Control Measures

6.1 Application Scenario

Management of corporations will delegate tasks. Delegating tasks raises specific control problems, as studied in agency theory[29]. In particular, delegation raises the problem of *private information* [47]: the agent to which the task is delegated has private access to information about execution, which the principal, who delegated the task, does not have. Private information problems can be of several types, namely *moral hazard* (hidden action), the agent may perform differently from what was expected, and *adverse selection* (hidden knowledge), the principal may have chosen the agent on the wrong grounds. Control problems are usually

[29] Agency theory is not the same as multi-agent systems theory. It is used in economics and sociology to study the delegation of tasks and ways of dealing with the resulting control problems [25]. For example, it explains the nature of remuneration and the set up of contracts.

mitigated by the principal, who may demand additional *control measures* to be implemented, such as supervision, formal procedures and guidelines, budget constraints, verification measures, software application controls, input controls, etc. Control measures must fulfill a purpose: a *control objective.* Generally, control objectives require a combination of organizational, procedural and automated control measures. A control objective corresponds to the notion of norms used in Normative Multi-Agent Systems. Just like norms, control objectives prescribe particular behavior, and clearly define deviations and violations.

However, adding control measures may have large *costs.* Consider the costs of implementing controls, the costs of resources that cannot be spent otherwise and the costs of reduced efficiency, usability or flexibility in execution of a task. In an attempt to reduce the costs of control, automated controls are becoming increasingly important. Control measures are for example built into ERP systems and workflow management systems to prevent undesirable behavior. This preventative approach may be called *compliance by design* [34]. Security logs and existing systems for monitoring process quality are extended and also used to verify effective implementation of control measures. Such a continuous approach to monitoring and detection may be called *continuous control monitoring* [72], or *continuous auditing* [46].

In general, providing assurance that an organization is compliant, involves several tasks: determining the control objectives (norms), determining specific control measures (actions) to be taken by the organization, determining control indicators (evidence), evidence collection, monitoring, warning in case of deviations, adjusting behavior when necessary, and applying sanctions when necessary. When parts of control systems are automated, the various tasks involved in providing assurance are re-distributed [14]. Tasks like data collection, monitoring, and triggering warnings can be automated. But even in a fully automated control system, an auditor must assess the appropriateness of the design and verify operating effectiveness of the system as specified.

Question: what are the effects of control automation on assurance provision?

Traditionally, auditors are responsible for providing reasonable assurance that (financial) information is free from material misstatements [45]. In doing so, auditors often use the *audit risk model.* This model helps to determine the amount and kind of substantive testing the auditor must perform for a given audit assignment, given the nature of the business, the strength of internal controls and the quality of evidence. Substantive testing is done manually, and is therefore relatively expensive.

$$Audit\ Risk = Inherent\ Risk \times Control\ Risk \times Detection\ Risk \tag{1}$$

Audit risk is the risk for an auditor that material misstatements remain undetected. Usually, an acceptable audit risk is set beforehand. *Inherent risk* is the a priori likelihood for a misstatement. This is based on the nature of the business. *Control risk* is the likelihood that (internal) controls will not prevent or detect a misstatement. This depends on the strength of preventative controls. *Detection risk* refers to the risk that an auditor will not detect a material misstatement. This depends on the amount of substantive testing and the persuasiveness of audit evidence. In general, there are six ways of obtaining audit evidence: Inspection, Confirmation, Observation, Re-performance, Analytical evidence and Client inquiry [45]. Of these types Inspection, Confirmation and Re-performance are considered most reliable, but they are also the most labour intensive and therefore the most expensive. So there is a trade-off between the quality of evidence and the costs of control.

We argue that automating controls may have two kinds of effects on the costs of control. Obviously, it will increase control effectiveness (prevention). In terms of the audit risk formula, this means a smaller control risk. Given a fixed audit risk and inherent risk, this means that the detection risk is allowed to be higher, and less substantive testing is required, reducing the costs of control. Second, controversially, we also believe control automation will enhance the quality of evidence (detection). This means a smaller detection risk. Therefore, less additional substantive testing is needed, which will reduce the costs of control.

6.2 Example Case Study

We investigated this claim by analysis of a case study [17]. The case concerns the procurement process for care-related public transport services. The case is representative for a large class of complex purchasing processes, which can benefit from control automation.

The case involves the following parties. SRE[30] is the name of a public agency acting on behalf of several municipalities in the south of the Netherlands, which must purchase care-related transport services for elderly and disabled citizens. As you can imagine, care-related transport services are highly regulated. There are many legal requirements concerning the vehicle, the driver and how to deal with patients. The transport service provider (TSP) provides care-related taxi bus services on a demand basis. It receives a monthly fee from SRE for all patients being transported, as well as individual contributions from non-patient passengers. TSP keeps a record of all trips being requested, executed and cancelled, including data about individual patients and other travelers.

What is the norm? SRE must ensure the accuracy and legitimacy of the monthly invoice from TSP. The norm that needs to be verified, is whether the invoice is calculated correctly and all trips are executed according to legislation. Therefore the contract stipulates that TSP must provide a data file about the executed trips. The contract also contains a data-protocol about the expected format of the data file. In the context of the contract, the data file counts as evidence of accuracy and legitimacy of the invoice. How can SRE verify adherence to this norm? In other words: how can SRE establish reliability of the data file?

First, with the help of an auditor, SRE has set up a system of automated controls, to verify syntactic and semantic well-formedness of the data according to the data protocol. Verification is executed automatically by a PHP script. For example, for all trips a patient number must be recorded, and the patient must be known to be eligible for transport. In addition, the script can also test for coherence of the data file according to *reconciliation relationships* [67]. For example, the set of executed trips should equal the set of requested trips, minus the cancelled and no-show trips. Or, the total length of all executed trips should equal the sum total of kilometers registered by the taxi company. These reconciliations make sure that the data represent well-formed transactions.

Second, the contract stipulates that once every year, SRE must provide an audit opinion from an external auditor about the reliability of the processes and computer systems which generate both the invoice and the data file. In particular, reliability depends on segregation of duties and general IT-related control measures: change management, access control, logging and monitoring, and baseline security.

6.3 Challenges for Norms

The case study concerns *control automation*: automatic verification of evidence against a norm. Although there are many techniques for automated verifications, it is unclear how

[30] Samenwerkingsverband Regio Eindhoven (Cooperation Eindhoven Region)

these techniques will affect the role and responsibility of auditors in providing assurance. In fact, the case is representative of a large number of cases, where audit evidence will be provided by an automated system, partly under the supervision of the company being audited. This requires trust, that can be founded on control measures. Auditors are still important, but only at the set-up and periodic assessment of the automated collection of evidence. In a sense, auditors will now perform a meta-audit of the automated controls, rather than a direct audit of behavior. Clearly, the tasks in assurance provision can be redistributed, partly to the automated system and partly to the company being audited. How does this redistribution of tasks affect assurance provision?

The challenge is to come up with a suitable *assurance architecture*: a set of modules in a specific configuration, which together provide assurance that some process or system is 'in control', i.e., will meet specified control objectives (norms). Some modules may be implemented by procedures carried out by humans and others by automated systems.

6.4 Use of NMAS: Status and Prospects

Currently, automated verification of controls is often addressed within the field of business process management (BPM). Here, people tend to focus on *conformance testing* [58]: can we prove that process designs meet specific constraints? However, the basic audit questions remain relevant: who translates a general regulatory objective into appropriate process constraints? (testing of design) Who decides that a specific system does in fact implement the processes as specified? (operating effectiveness).

We believe that essentially, these questions are about the automated collection of evidence. We believe the notion of *constitutive norms* [62], can be fruitfully used to further investigate the conditions under which automatically generated evidence becomes legally acceptable. In the case study, we have seen that a monthly datafile may under some circumstances (verified to be well-formed; yearly external audit opinion) count as sufficient evidence.

The notion of *organizational roles*, which is explored at length in Normative Multi-Agent Systems [8], will also play an important role. After all, who is authorized to state that certain automatically generated data counts as evidence of compliance to norms?

Finally, the notion of a *contract* will be crucial. Much compliance issues develop between companies, although set in a legal context of enforced contract law. There are various proposals for the representation of contractual clauses, and subsequent translation into business process constraints, e.g. [38]. The process of contract negotiation between parties about the required strength of additional controls can be fruitfully studied using qualitative game theory, as developed with normative Multi-Agent Systems [8].

7 Chopra: Norms in Requirements Engineering

Requirements Engineering (RE) either treats requirements as properties of the environment [76] or as stakeholder goals [52]. In this note, I present a novel conceptual take on requirements.

I take a broader communication-centric view of RE than is customary. In this view, RE is itself a *social application* [15, 16] in which software engineers and stakeholders represent autonomous participants. Taking this perspective sheds new light on the nature of requirements.

7.1 Suitability of a Normative Model

A communication-oriented view leads naturally to the idea that a requirement is a normative relation between the communicating parties. I use the notation $\mathsf{R}(x, y, p, q)$ to mean that x *requires* of y that if p then q (p and q are propositions whose satisfaction or violation can be determined by observing the environment). For example, BestWines requires that CellarSys raise an alarm if the temperature in the cellar rises above 12°C.

$\mathsf{R}(\mathit{BestWines}, \mathit{CellarSys}, \mathit{temp} > 12^\circ, \mathit{alarmRaised})$

Requirements, like commitments [64], are established and manipulated by communications. Table 2 shows a partial list.

Table 2 Communicating Requirements.

Communication	From	To	Desired Effect
CreateReq(x,y,p,q)	x	y	$\mathsf{R}(x, y, p, q)$
ReleaseReq(x,y,p,q)	x	y	$\neg\mathsf{R}(x, y, p, q)$
CancelReq(x,y,p,q)	y	x	$\neg\mathsf{R}(x, y, p, q)$

The above normative view of requirements brings forth the nature of a requirement. It makes the parties to a requirement explicit. In influential RE literature dealing with the conceptual treatment of requirements, the parties are either implicit, or worse, missing altogether. There are at least two pragmatic advantages in making the parties explicit: (1) large projects may involve multiple stakeholders and engineers, and (2) contracts among parties will be established based on the requirements. Further, the above view of requirements does not make any assumptions about stakeholder goals; their existence is grounded instead in communication.

Requirements may be satisfied or violated. Further, once a system that presumably meets a requirement has been deployed, if the requirement is violated in perfect operating conditions for the system (see discussion on *domain assumptions* below), then the stakeholder will hold the engineer responsible for the violation. And lastly, a stakeholder cannot repudiate his or her requirement arbitrarily: the stakeholder is bound to the statement of the requirement (nonrepudiation does not mean that a stakeholder cannot change requirements; it just means that he or she cannot deny their communication).

7.2 Challenges for Norms

Besides requirements, RE also deals with domain assumptions and specifications [76]. A *domain assumption* describes what one can safely assume to hold in a stakeholder's operational environment. Specifications describe a machine's interface with the environment such that when an implementation of the machine is introduced in the environment, the requirements are satisfied. In other words, specifications are the bridge between RE and the rest of software engineering (SE). A normative description of RE would have to account for not just requirements, but also domain assumptions and specifications. Further, these concepts would have to be formalized so that one could perform reasoning over them.

7.3 Status and Prospects

The above normative view of requirements engineering breaks from the prevalent tradition in RE where requirements are either described in low-level terms or as the goals of stakeholders.

A communication-oriented view of RE can potentially have a big impact on the practice of both RE and SE.

- It could provide a basis for formulating business contracts between engineers and stakeholders.
- It provides a conceptual framework within which to place *requirements evolution*. Requirements would be created because of communications from the stakeholder to the engineer and would evolve only when the stakeholder communicated modified requirements. Further, any evolution would need to be accommodated by a change in the contractual relationships between the parties. The operationalization of the communication primitives would result in *requirements management systems*.
- It provides a basis for requirements evolution to be understood in the broader context of *requirements negotiation*. The stakeholder may change his requirements; however, that does not mean the engineer is committed or will commit to meeting them. The engineer would normally also take into account the cost and time required (among other things) to meet the requirements and possibly make counter-proposals.

8 Noriega: mWater as a Normative MAS[31]

8.1 Application Scenario

Water use—because of scarcity and stake-holders' conflicting goals–is a conflict-prone domain. Not surprisingly, it is a highly regulated one. One way that water policy-makers have to foster better water use and avoid conflicts is to regulate demand, and one such way is establishing a "water bank" to trade water rights.

mWater is a normative MAS that models the use of water rights in a closed basin. It focuses, on one side, on the process of trading those rights (regulating conditions that make the rights tradable, thus affecting demand and use behavior) and, on the other, on the process of using those rights (thus affecting conflict and conflict resolution).

The system has three objectives:

1. As a testbed of agent technologies developed within the Spanish Agreement Technologies project.
2. To simulate effects of different normative corpora on user behavior, for water management policy design.
3. To build a realistic prototype of an on-line water bank for a closed basin.

The design is rooted on traditional practices and actual regulations, although it is slightly idealized to be a malleable platform (for testbed and simulation). The model departs from current legislation by allowing trading and usage with added flexibility (contract, use and misuse of rights, grievances and corrective actions) and under different price-fixing and conflict resolution mechanisms. The model makes obvious simplifications on the anchoring and constitutive conventions.

In abstract terms, *mWater* is defined as an agile market (mWater) of water rights and an agreement space for the management of rights. It is modeled and implemented as a regulated open MAS using the electronic institution (EI) meta-model and the EIDE platform [22].

[31] Acknowledgments: The mWater prototype was partially funded by the Consolider programme of the Spanish Ministry of Science and Innovation through project AT (CSD2007-0022, INGENIO 2010) and MICINN project TIN2008-06701-C03-03. This research has also been partially funded by the Generalitat de Catalunya under the grant 2009-SGR-1434 and Valencian Prometeo project 2008/051.

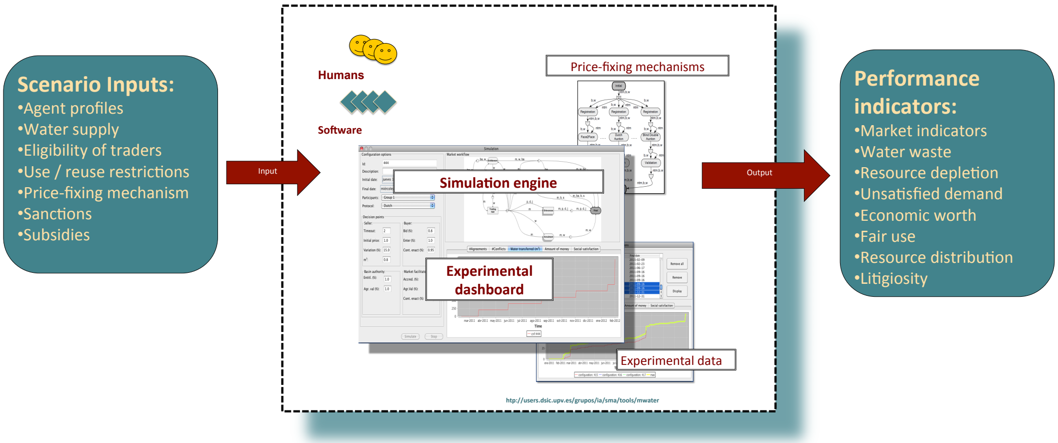


Figure 6 mWater used as a water-policy simulator.

8.2 Suitability of a Normative Model

From a normative point of view, *mWater* has the following features:

- *Norm sources:* Laws and regulations issued by governments; policies and local regulations issued by basin managers; traditional social norms.
- *Norm Expression:* Some are regimented in the EI framework or embedded in the decision-making processes of institutional agents. Other expressed are expressed in declarative form so that individuals may reason about them. The idea is that agents and designers may reason about these norms on and off-line; at design and at run time; and from the institutional (or legislative) perspective and the agent's individual perspective.
- *Issues dealt with in modeling:* Choice of expressive formalisms, institutional and social governance; norms dynamics (legislative and individual's perspectives), agent's decision-making strategies for compliance and, finally, criteria to evaluate effectiveness of norms.

The following, exemplify the types of norms that govern the mWater domain:

- Ponds in private land plots are considered part of it as long as they are used exclusively within that plot.
- Mineral and thermal waters have their own regulation.
- Any person may treat sea water provided proper administrative permission is granted id it has been proved that standards of residual disposal and quality for the intended use are met.
- When the Ministry of the Environment declares a state of emergency drought, all entitlements to water rights are suspended while the state is active.
- Rights are tagged for a type of use and may be exploited for that or for uses of higher priority. Water use priorities in descending order are: human, agriculture, energy, and recreational.
- The basin authority may at its discretion allow trading of suspended rights during states of emergency, given priority to holders of lower priority rights for higher priority use.
- Unused rights may be challenged and expropriated.

8.3 Challenges for Norms

As a normative MAS, *mWater* provides three main contexts of interest where challenges may be properly differentiated and to a large extent isolated and addressed independently:

- *Negotiation*: Protocols (conventions, eligibility of participants, tradable rights; judgment aggregation (decide when drought may be officially declared), negotiation heuristics (from an individual's decision to comply or not with given norms, suitability), agent with norm-aware architectures; argumentation about suitability, non-compliance, and others.
- *Contracting*: actual clauses of contracts, conflict identification and impact assessment. Transient character of norms, commitment conflicts, new contracts co-exist with other active contracts.
- *Agreement management:* Agreements and contracts may be contested by agreeing parties, observing third parties, authorities. System and agents capabilities to address conflict, online dispute resolution (ODR) and other forms of conflict management. Post-trading activities to study conflicts (detonators, structure, types) and conflict resolution (intervention strategies, rhetorical moves, settlement strategies).

Generally speaking, then, the application provides a convenient example to study the interplay between formal institutional components (laws, ontologies, sanctioned practices) organizations that enforce or should abide by them, individuals that form those organizations or participate in the regulated activities. An interesting, and perhaps not so frequently seen in normative MAS, mWater involves collective actors with their own group rules for allocating rights and solving conflicts. Collective actors that are involved in negotiation as collective entities in collective decision-making, judgment aggregation and other ways of social choice. Moreover, the problem domain needs to reflect organizational and institutional dynamics, is a good example to study an individual's immersion of norms and collective emergence of norms, and a natural context where there are empirical grounds for playing with notions like trust and reputation, moral authority, power and force.

8.4 Status and Prospects

A crude prototype of the complete system (including both trading and conflict management) was been implemented using the EIDE platform. Then a second version consisting of a more thoroughly developed trading part was implemented, and then, on top of it, a running simulator was built. Recently, a new version of this simulator with suitable agent populations is being developed using the GORMAS meta-model and tools.

On the other hand, mWater provided grounds for two other systems currently under development: a system for on-line trading of waste products and an "agreement space" for open innovation (in the "green economy" domain). Both have commercial interests behind and, in both cases, the need to approach the problem as a normative MAS is not only adequate but unavoidable if an effective sociotechnical system is to exist.

9 Balke: Wireless Mobile Grids

9.1 Application Scenario

The current deployment of the third generation (3G) of mobile network systems is in progress, but a quite different next generation network (called Fourth Generation or 4G) is under development. This latter is intended to bring about a paradigm shift in the cooperation architecture of wireless communication [44]. Whereas for 3G the industry focused on technology for enabling voice and basic data communications (technology-centric-view), the emphasis in 4G is more user-centric [75]. One issue that, according to several studies [70], is of very high importance to users is the battery capacity of mobile phones.

Batteries have fixed capacity that limits the operational time for a device in one charge cycle. The increasing sophistication of mobile phones and their evolution into smart phones offering Internet access, imaging, audio and access to new services, has a significant impact on power consumption, leading to shorter stand-by times.

Fitzek and Katz [29] proposed a mechanism to address these issues with the concept of a "wireless mobile grid" (WMG), in which users share resources in a peer-to-peer fashion via the short-link connection devices built-in in the current mobile phone generation (e.g., WLAN or Bluetooth). The advantage of these it that they use significantly less power. However, the WMG idea of Fitzek and Katz requires collaboration between users that may be difficult to realize. The ensuing social dilemma is that network users can exhibit strategic behaviour, that places their own benefit above that of the collective. The main problem in WMG is that collaboration comes at a cost, in the form of battery consumption for contributing to the WMG. In consequence, rational users will prefer to access the resources without any commitment. However, if a substantial proportion of users follow this selfish strategy, the network itself would be at stake; depriving all users from the benefits, namely the potential battery saving arising from cooperation [75].[32]

9.2 Suitability of a Normative Model & Challenges for Norms

Following this brief explanation of WMGs and the possible contribution problems in them, they appear to be an interesting case study of a normative models in which one could analyse how different norms and enforcement mechanisms (reputation, police agents,. . .) affect the behaviour of the telephone users (agents) in the system and how norms could possibly alter their behaviour.

From a normative point of view WMGs have several interesting properties. These properties, which make them well suitable for normative models and pose interesting challenges for normative MAS—the major one being the study of encouraging collaboration (e.g., by means of enforcement)—will now be explained in brief.

Complex Open Distributed Setting. WMGs are systems in which any mobile phone user with an WMG-enabled phone can join or leave at any point of time. The users are not static, but they are moving, which changes the possible collaboration groups all the time and also reduces the number of possible repeated interactions. This poses interesting challenges for reputation-based enforcement mechanisms. In addition if reputation mechanisms were to be used, the problem arises that sending information comes at a battery cost again and thus poses the research challenge to reason about enforcement and "optimal" levels of enforcement when enforcement is not for free but comes at a cost.

[32] WMGs are similar to conventional P2P networks, but differ in physical aspects and emphasis. In particular, phones have specific resource restrictions (e.g., SIM card space) that limit the processing and storage capabilities of the WMG nodes. In addition, the WMG concept includes sharing processing power (omitted in the scenario discussed later for the sake of simplicity). Current scenarios from the mobile phone industry include big sport events, news and financial data in banking districts, IPTV, cooperative online gaming as well as maps and location information at airports, etc. Routing is a major issue in P2P systems, but of less significance in WMG, in part due to the relatively short-lived and transient nature of alliances. Lastly, we note that transmission failure is more common in wireless than in wired networks, making it harder to determine intent when non-cooperation is observed.

As real humans are envisioned to engage in WMGs, they need to be adequately represented in a normative MAS by means of agents. This poses huge challenges w.r.t. modelling their decision making behaviour, which is far from being rational at all times.

Beyond Micro-Macro. From a normative perspective, in a WMG both on the micro as well as on the macro level norms can be present and the norms on both levels possibly affect each other. Thus, in WMGs one expects norms to be defined at a macro-level specifying correct behaviour and for example which sanctions might be enacted if the users do defect (i.e., do not contribute to the WMG but use its resources). This is likely to influence the decision making of the users in the WMG. On the micro level, in addition norms as a result of the user interaction can emerge. These norms do also can have an effect on the users' decision making and actions and will not necessarily be in concordant with the norms defined on the macro-level. They might even effect the the macro-level norms. This results in a complex micro-macro link where both levels bi-directionally interactively influence each other. To model this continuous two-way interaction of the micro and the macro perspective is another challenge for normative MAS.

Multiple Stakeholde.r The final interesting challenge that WMGs pose for normative MAS is the number of different stakeholders involved in WMGs; all of which are dependant on one another, but which have different objectives. In the WMG one for example typically finds at least the mobile phone users, mobile phone manufacturers, infrastructure providers as well as telecommunication providers. Balancing the interest of these different WMG stakeholders adds another layer to the question of finding good mechanisms that encourage WMG collaboration—a challenge that so far has not been approached in normative MAS.

9.3 Status and Prospects

From a technological point of view, much has been done in respect to WMGs. Thus, currently prototypes of WMG-capable mobile phones exist and first tests for their deployment are being conducted. This allows to obtain real WMG mobile phone data which can be used in normative MAS models. Nevertheless, from a normative perspective many questions—such as how to encourage contribution and discourage defection in these networks—still need to be addressed, before WMG could be turned into a commercial product. This real world commercial focus as well as the numerous WMG-related interesting research challenges for normative MAS outlined before are the reason why research into the normative aspects of WMGs is both important as well as at the same time challenging, but worthwhile.

10 Governatori and Lam: UAV[33]

Governatori and Rotolo [36, 37] proposed a computationally oriented rule based approach to model (normative) agents. The framework, called BIO (Belief, Intention, Obligation) is an extension of defeasible logic with modal operators to model: the representation of the environment in which an agent is situated (beliefs), the norms governing the agent (obligations), and the goals of the agent (intentions). The BIO approach permits parametrised definition of different agent types, where an agent type corresponds relationships and preferences over the various modalities describing the mental attitudes of the agent and external modalities. The framework is intended to provide executable specifications for

[33] Acknowledgments: NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council.

an agent. This means the rules in which the agent and the environment are described provide the rules can be executed directly by a defeasible logic engine without the need to program the agent in an external language. To facilitate this, SPINdle [48], a modern and efficient Java based implementation of defeasible logic has been developed. The result is that the combination of of the BIO framework and SPINdle offers a flexible and agile tool for programming (normative) agent based applications.

10.1 Application Scenario

Typical complex system have to manipulate and react to different types of data (e.g., numerical and boolean), and in many occasions we have to integrate different types of reasoning process. In this sense, the focus of our research are of two-fold: (1) how a non-monotonic rules-based system can be integrated with numerical computations engines, and (2) how the behavior of an agent can be affected by the external contexts. To illustrate the combination of techniques, we have the following problem scenario:

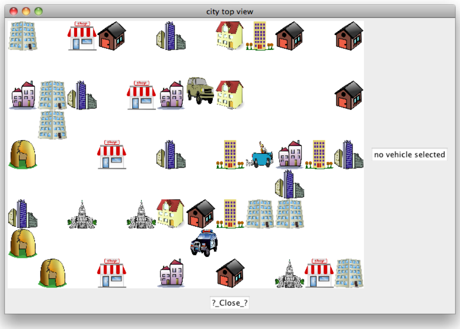


Figure 7 City map model.

> Given a city map with specific targets and obstacles (Figure 7), a number of UAVs has to navigate through the city from a starting location to a destination without colliding with each other. There is a GPS enabled application that informs the UAVs about the current traffic situations and the locations of other UAVs. To navigate successfully, the UAVs have to follow some guidelines about how and when they should alter their route.

The above scenario revealed how a UAV should interact with its environment. It presumes an execution cycle consisting of a phase where the UAV collects information through its sensors, decides an action and then applies this action [74].

10.2 Suitability of a Normative Model

In order to travel from one location to another, a UAV has to gather different types of information from the GPS monitor within a proximity range and detects if any traffic problems might appear. The Knowledge Base (KB) of a UAV is a well-documented limited set of behavioral interactions that describes the behavior of a UAV under different situations, in particular it contains (prescriptive) rules for determining who has right of way over who. It is complemented with formal logics (and in particular Defeasible Logic (DL)) to represent significant issues regarding the domain of operations.

10.3 Challenges for Norms

In case of a possible collision, a UAV will utilize the information in its KB and incorporate into it the set of context-related information (such as traffic situation, information about the vehicles nearby, etc) and derive a safe direction of travel or eventually to temporary stop its motion in *real-time fashion.*

Consider the scenario as shown in Figure 8 where vehicles V_3, V_4 and V_5 are moving towards the same location (the red circle) and collisions may occur if none of the vehicles alter their route. This perception-action cycle (Figure 9) can be conceived not only as an issue of control, but also lays out the interface on how the UAVs should interact with the environment [4]. Here, the *sensors* (in our case the GPS monitor) collect information about the environment (as described above) and which is then combined them with the knowledge base for common-sense decision making. The *behavior controller* then reasons on these information and triggers the associated *actuator(s)* (whether to change its current travel direction, speed or even to stop its current motion) based on the conclusions derived.

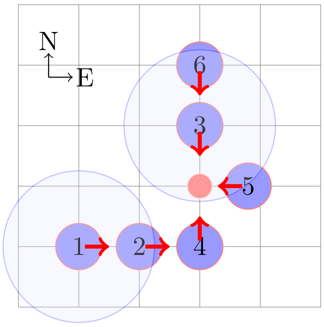


Figure 8 City with UAVs.

10.4 Status and Prospects

Besides the combination of numerical and logical techniques, in particular, an extension, which combines the ability of handling violations and degree of preferences, of BIO approach to the development of agents in DL has been established. In addition, a novel algorithm computing extensions of Modal Defeasible Logic has been devised. Readers interested please refer to [49] for details.

Future work includes the study of UAV negotiation, the use of Temporal Defeasible Logic and integrating the rule-based system with reaction-based mechanism.

The main aim of this application was to demonstrate the suitability of the use of the BIO approach to the development of agent based applications. The outcome is promising. It was possible to describe the behaviour of the UAV agents, and the norms defining the right of way in a theory of approximately fifty rules.

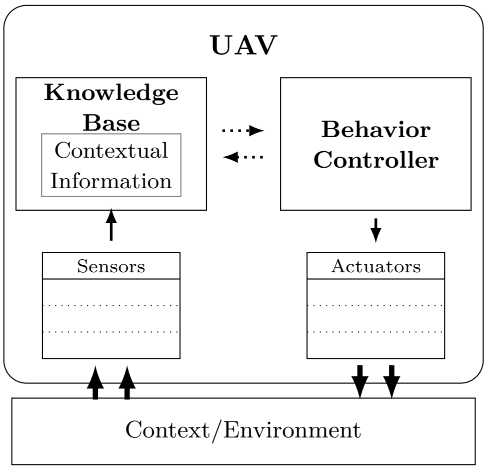


Figure 9 Behavior control under an perception-action cycle.

11 Dignum: Using Norms for Serious games

11.1 Application Scenario

When training skills like arithmetic or spelling the trainee has to learn to follow the rules that will deliver the right output for a given input. For example, $5^3 = 5 \times 5 \times 5 = 5 \times 25 = 5 \times 20 + 5 \times 5 = 100 + 25 = 125$.

In these types of serious games the rules can be implemented as constraints or given as formulas that can be used to calculate the answer. However, when training involves the behavior of other people we cannot suffice to use simple rules or constraints anymore. For example, when learning to drive a car in a simulator we also have to simulate realistic behavior of other traffic participants. Although all traffic participants are regulated by the same traffic laws this does not mean that they always obey those rules!

In order to train properly we should not just create traffic participants that do not follow the rules, but characters that violate the rules in realistic ways. For example, a kid might cross the road without looking when chasing its ball. So, this would happen if there is space next to the road where the kids can play with a ball (e.g., a park or garden), but not at a through road without playroom or houses. It would be completely weird to have a kid chasing a ball in a highway.

In order to model scenarios where characters violate behavioral rules in believable ways we need to be able to model the rules as norms which can be violated. Having the norms represented explicitly facilitates the representation of rules for violation as well. These rules indicates the type of circumstances in which violation is likely or at least possible. They might also indicate the consequences of the violation.

Using norms to represent behavioral rules in training scenarios forces the designers of the game to think carefully which type of violations (or unexpected events) the trainee should be able to handle. When some violations are not important for the purpose of the training they can be represented by fixed constraints or rules and thus will never be violated.

For many training scenarios it is exactly the coping with violation of rules that is important. In a first phase of the training one can make scenarios where every character keeps to the rules and thus the standard behavior is trained. Think about a driver that learns to cope with handling the car (gas pedal, brake, clutch, steering wheel, etc.). In the next phase the trainee is allowed to apply the behavioral rules, but, of course, can make mistakes when applying them. In this case the game should react properly in order to show the (possible) consequences of breaking a rule. For example, when a car turns right the driver should look over his right shoulder to check if a bike is approaching (typical in at least the Dutch situation). When the driver forgets to do this the game might react by having bikes appear and hit the car or fall when trying to avoid the car.

Finally, we want to model scenarios where characters violate the rules and the trainee has to react properly to the violation. For example, bikes going straight and ignoring a red light while the car has to turn right. In this case the driver might assume he does not have to look for the bikes because his light is green. However, in The Netherlands the car is still responsible to avoid the bikes even though they violate the traffic rules.

Many training scenarios follow the same pattern as sketched above. Whenever training skills in situations that involve other people it is important to be able to accurately react to violations of rules. In order to model these scenarios it is imperative to model rules as norms and have them explicitly available during the design.

Although many serious games are now used in class room situations where cognitive skills are trained there is a large interest to use serious gaming for training skills that involve social abilities and require adequate coping with people behaving unexpectedly. Training of these skills which range from the above driving lessons, to fire drills, team training and application training usually are now performed using hired actors or involving experienced professionals. These people are expensive to use for training and are only available at limited times. Therefore the future of serious games for this market is promising.

11.2 Suitability of a Normative Model

The suitability of the normative model is already discussed implicitly in the previous section. When the rules would have been implemented as constraints in the game, the trainee would have no opportunity to violate the rules. Although this might be good in the initial stage of the training, it would really impede the learning of the rules, which is an important part of training situations involving other persons.

One could argue that the system does not need to represent the norms explicit in order to react to violations of rules by the trainee. This is true. However, designing the system becomes easier if situations can be generated (or triggered) based on explicit violations of rules. This makes the system more modular and also facilitates later additions of extra scenarios. This might be necessary when certain violations occur often and can occur especially in different situations. Then one might want to design different scenarios covering all the prototypical violation situations. For example, not looking over your right shoulder when turning right in a car can happen at a traffic light, but also when changing lanes at the highway or returning in your lane after passing a car. It is easy if all these situations can be explicitly tied to the same norm. Additionally this also facilitates keeping track which rules a trainee has most trouble with and possibly giving him extra scenarios in which that rule plays a role.

Having the norms explicitly available becomes even more important when the characters in a scenario are supposed to violate the rules. Those violations should be realistic. It is very hard to implement realistic violations for many scenarios if the norms are not explicitly available and can be prioritized and balanced against other goals of the character. Thus having normative agents, that can explicitly reason about norms and make decisions based on norms and other elements of the situation becomes almost necessary in order to implement these complex scenarios.

11.3 Challenges for Norms

When using norms in serious games one of the main challenges will be the combination and prioritization of the norms. Usually there will be both general rules governing the behaviour and interactions between persons as well as specific rules that govern detailed situations. For example, in general the speed limit for cars within a build-up area is 50 km/hr, however when signs along the road indicate that the speed limit is 70 km/hr one can drive more than the 50 from the general rule. (In general, in traffic law, signs take priority over rules). However, this becomes different again when a truck drives in a highway, where the sign states that you can drive 100 km/hr. In spite of this sign a truck can still only drive 80 km/hr in the highway.

There has been quite some theory developed on deontic logic that can be used to combine and prioritize norms. However, little work has been done that makes this theory practically usable in software systems.

11.4 Status and Prospects

Although the use of norms in designing serious games seems very intuitive we have not been able to actually include it in a project yet. The main impediment is the fact that a different design methodology should be used to design the game. At the present time game developers are under too much pressure to make money in order to take any risks by using an unfamiliar methodology.

We have participated in the early stages of a serious game design project for training fire fighting on board ships. Introducing norms led to exactly the right questions about which scenarios would need to be made for the training to be effective. Thus it seemed very promising as part of the design methodology. Unfortunately, after the initial phase there was too little money to take any risk in developing the game and traditional methods were used (leading to a rather static and limited game play).

We hope to get funding from government in order to develop a real case study and get a prototype serious game that can be used as show case. Once we succeed in this the prospects are quite good. Industry indicates that games for training human skills will become more important. Normative behavior forms an integral part of this type of skills and it thus becomes imperative to include norms into the games.

12 Cranefield and Verhagen: Virtual worlds as an application area for normative multi-agent systems

12.1 Application Scenario

Consider an online meeting place in which geographically separated human users can interact with each other and with software agents in a human-friendly way, whilst also being secure in the knowledge that certain specified norms of interaction will be monitored and enforced. That is the type of scenario explored by research on extending 3D virtual worlds with e-institution and normative multi-agent systems middleware. For example, remote buyers can participate in auctions held in a real-life fish market by controlling avatars that move (when allowed by the auction house rules) through various virtual rooms representing different stages of the auction process (e.g., buyer registration, the auction itself, and settlement), using natural gestures such as raising their (virtual) hands to make bids [11]. A connection with the real-world fish market is provided by instrumenting objects in the virtual world objects with scripts that sense avatar actions and enable or disable virtual counterparts to institutional actions (e.g., doors will open to allow avatars to move between rooms if they have permission to move between the stages of the auction represented by those rooms) [12].

This scenario illustrates the potential of research on electronic institutions and normative multi-agent systems to enhance computer-mediated human interaction. While virtual worlds offer a rich medium for people to interact despite being physically distant, they currently offer little or no support for users to maintain an awareness of the social context in which they are interacting. Tools developed by researchers in the NorMAS research community hold promise to fill this gap.

12.2 Suitability of a Normative Model

While much of the use of virtual worlds such as Second Life [50] is for unstructured social interaction and exploration of novel constructed fantasy environments, virtual worlds are also used for interactions within societies or organisations with a predefined or emergent social structure. For example, virtual worlds are used as a venue for meetings, lectures, role-playing and training exercises, re-enactments of historical or fictional societies, and for buying and selling both virtual and real-world goods. All these activities involve social structure and, potentially, norms.

12.3 Challenges for Norms

The scenario above has been implemented using an existing model of *electronic institutions*. This was possible due to the use of a specially constructed virtual world environment generated from an electronic institution specification [12]. Key challenges in this approach include how to generate virtual world environments from institutions, and how to map institutional actions and the regimentation and enforcement of norms to suitable counterparts in the virtual world.

Another challenge is to integrate NorMAS technology with virtual world environments that have already been constructed and have existing social uses. Although it has been suggested that the popular virtual world Second Life does not facilitate the use of social norms as an efficient social order mechanism due to the ease of users changing their identities [68], social norms have been identified at the level of groups in Second Life [9]. NorMAS technology could therefore be applied to provide computational support for normative processes at the local group level in virtual worlds. This could be done in a way that requires users to adapt some of their real-world practices when interacting in virtual worlds, e.g. by provided instrumented virtual artifacts that users must use to ensure that their institutional actions are detected. An example would be an object that virtual meeting participants must pass between themselves to indicate who "has the floor". However, it would be less intrusive to develop techniques for detecting the existing domain-specific significant events in virtual worlds. As virtual worlds are real-time simulations with many possible avatar movements and actions available to users, detecting significant normative events requires the recognition of complex events from a rich (and not always consistent) sensory data stream. The richness of virtual worlds may also mean that more expressive languages for representing norms will be needed.

An even greater challenge is to develop techniques for *learning* norms that are present in existing virtual world societies. In this case, the high level significant events (such as the application of sanctions or other signals that indicate norm violation) cannot be predefined and must be learned from observation, with the norms then inferred by an inductive process. Norms that are never or rarely violated will be difficult to learn, unless agents can communicate with human avatars about possible norms and use natural language analysis of text chat to detect discussions about norms. Awareness of cultural difference may also be necessary, given the diverse range of users that may be present in virtual worlds. Research in this area could be undertaken not only to provide better tools for better social awareness and control in virtual world societies, but also to study norm-related processes amongst human users. In the chapter titled "(Social) Norm Dynamics" (page 135) the effects of cultural differences are discussed in more detail.

However, the focus there is not on mixed culture groups but on potential differences between cultures as such.

12.4 Status and Prospects

Progress has been made on many of the research challenges above, but to the best of the authors' knowledge, no NorMAS technology has yet been deployed in real virtual world applications—only research prototypes have been developed to date.

Perhaps the most sophisticated constructed virtual institution is an interactive simulation allowing users to have an immersive experience of the culture of the ancient city of Uruk in 3000 B.C., complete with agent-controlled citizens playing different a variety of roles in the society [10]. Technology developed to facilitate the construction of such virtual institutions includes middleware for managing "intelligent objects" in virtual worlds [57] and a formal approach to generating environments in virtual worlds [71]. Another recent application of this technology is a prototype of a virtual institution for trading water rights [2].

Other work has investigated monitoring for the fulfillment and violation of conditional rules of expectation in existing environments in Second Life, based on the detection of changes in avatar animations [18] or predefined domain-specific complex events [56]. Agents can use a monitor service to track their personal expectations, which describe the expected future traces of the local region of the virtual world. These may correspond to norms, team tactics, or simply observed regularities in the environment that the agent assumes will hold, but wishes to monitor.

The concept of a *virtual nation* has been proposed "to guarantee the existence of a secure and safe virtual world" by incorporating into a virtual world real-world structures such as a constitution, government and monetary system [20]. Laws in a virtual nation are defined in natural language and then translated into technical implementations by a team consisting of (at least) a legal expert, a virtual world developer and a programmer. Advances in NorMAS technology will be necessary to make this vision a reality.

Virtual worlds also provide a generic 3D simulation platform in which games (particularly so-called "serious games", such as training scenarios [42]), can be implemented. The application of normative multi-agent systems to serious games is discussed in Section 11.

13 Summary

The preceding sections have illustrated the wide range of applications of norms in areas of significance not only to the computing discipline but also to society at large. These sections highlight various aspects of norms, e.g., in modeling complex systems, in engineering software systems, and in applying norms to capture and regulate interactions among humans. These could be potentially extended to apply to other real-life social entities such as organizations. Although the uses of norms reported here are research efforts, they are inspired by practical considerations. To fully develop an approach to the level where it could be deployed would require substantial effort, mostly in capturing elements of the domain over which norms can be employed.

14 Authors' Note

A remark about the writing. Like for the other chapters in this volume, a working group for this chapter was organized at Dagstuhl. The original contributors of that group (Savarimuthu, Singh, Villata) had drafted longish extended abstracts of their efforts, as did Christiaanse. Subsequently, other researchers were invited to contribute brief writeups on a use of norms. For this historical reason—and not anything to do with the relative importance of the application scenarios—the contributions by the above-named four authors (and their coauthors) are longer than the others.

15 Acknowledgments

We are indebted to the Dagstuhl organizers for the opportunity and venue to meet and discuss these topics. We would like to thank the anonymous reviewers for helpful comments.

References

1 Hal Abelson, Ben Adida, Mike Linksvayer, and Nathen Yergler. ccREL: The creative commons rights expression language. Technical report, Creative Commons, 2008.

2 P. Almajano, T. Trescak, M. Esteva, I. Rodriguez, and M. Lopez-Sanchez. v-mWater: a 3D virtual market for water rights (demonstration). In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, pages 1483–1484. IFAAMAS, 2012.

3 Matthew Arrott, Barry Demchak, Vina Ermagan, Claudiu Farcas, Emilia Farcas, Ingolf H. Kriiger, and Massimiliano Menarini. Rich services: The integration piece of the SOA

puzzle. In *Proceedings of the IEEE International Conference on Web Services (ICWS)*, pages 176–183, Salt Lake City, 2007. IEEE Computer Society.

4 David Billington, Vladimir Estivill-Castro, René Hexel, and Andrew Rock. Architecture for Hybrid Robotic Behavior. In Emilio Corchado, Xindong Wu, Erkki Oja, Álvaro Herrero, and Bruno Baruque, editors, *HAIS*, volume 5572 of *Lecture Notes in Computer Science*, pages 145–156. Springer, 2009.

5 Christian Bird, Alex Gourley, Prem Devanbu, Michael Gertz, and Anand Swaminathan. Mining email social networks. In *Proceedings of the 2006 international workshop on Mining software repositories*, MSR '06, pages 137–143, New York, NY, USA, 2006. ACM.

6 Christian Bizer, Tom Heath, and Tim Berners-lee. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, 5:1–22, 2009.

7 Barry Boehm and Jo Ann Lane. Using the incremental commitment model to achieve successful system development. TR 715, USC Center for Systems and Software Engineering, Los Angeles, July 2007. `http://csse.usc.edu/csse/TECHRPTS/2007/usc-csse-2007-715/usc-csse-2007-715.pdf`.

8 G. Boella and L. van der Torre. A game theoretic approach to contracts in multiagent systems. *IEEE Transactions on Systems, Man and CyBernetics - Part C*, 36(1):68–79, 2006.

9 Tom Boellstorff. *Coming of age in "Second Life": An anthropologist explores the virtually human.* Princeton, NJ: Princeton Univ. Press, 2008.

10 A. Bogdanovych, J. A. Rodríguez-Aguilar, S. Simoff, and A. Cohen. Authentic interactive reenactment of cultural heritage with 3D virtual worlds and artificial intelligence. *Applied Artificial Intelligence*, 24(6):617–647, 2010.

11 A. Bogdanovych, S. Simoff, and M. Esteva. Normative virtual environments: Integrating physical and virtual under the one umbrella. In *Proceedings of the Third International Conference on Software and Data Technologies (IC-Soft 2008)*, pages 233–236, 2008.

12 Anton Bogdanovych, Marc Esteva, Simeon J. Simoff, Carles Sierra, and Helmut Berger. A methodology for 3D electronic institutions. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 358–360. IFAAMAS, 2007.

13 Paolo Bresciani, Anna Perini, Paolo Giorgini, Fausto Giunchiglia, and John Mylopoulos. Tropos: An agent-oriented software development methodology. *Journal of Autonomous Agents and Multi-Agent Systems*, 8(3):203–236, May 2004.

14 Brigitte Burgemeestre, Joris Hulstijn, and Yao-Hua Tan. Value-based argumentation for justifying compliance. *Artificial Intelligence and Law*, 19(2-3):149–186, 2011.

15 Amit K. Chopra. Social computing: Principles, platforms, and applications. In *Proceedings of the 1st Workshop on Requirements Engineering for Social Computing*, pages 26–29. IEEE, 2011.

16 Amit K. Chopra and Paolo Giorgini. Requirements engineering for social applications. In *Proceedings of the 5th International i* Workshop*, volume 766 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2011. 138–143.

17 Rob Christiaanse and Joris Hulstijn. Control automation to reduce costs of control. In Marta Indulska, Michael zur Muehlen, Shazia Sadiq, and Yao Hua Tan, editors, *Proceedings of CAISE workshops – (GRCIS'2012)*, volume LNBIP XYZ, pages x – y. Springer Verlag, Berlin, 2012.

18 Stephen Cranefield and Guannan Li. Monitoring social expectations in Second Life. In J. Padget, A. Artikis, W. Vasconcelos, K. Stathis, V. Silva, E. Matson, and A. Polleres, editors, *Coordination, Organizations, Institutions and Norms in Agent Systems V*, volume 6069 of *Lecture Notes in Artificial Intelligence*, pages 133–146. Springer, 2010.

19 Natalia Criado, Estefania Argente, and Vicente J. Botti. A normative model for open agent organizations. In Hamid R. Arabnia, David de la Fuente, and José Angel Olivas, editors, *IC-AI*, pages 101–107. CSREA Press, 2009.

20 V. Čyras and F. Lachmayer. Technical rules and legal rules in online virtual worlds. *European Journal of Law and Technology*, 1(3), 2010.

21 K. da Silva Figueiredo, V. Torres da Silva, and C. de O. Braga. Modeling norms in multi-agent systems with NormML. In M. De Vos, N. Fornara, J. Pitt, and G.A. Vouros, editors, *COIN@AAMAS 2010, Toronto, Canada, May 2010, COIN@MALLOW 2010, Lyon, France, August 2010, Revised Selected Papers.*, volume 6541 of *LNCS*, pages 39–57. Springer, 2010.

22 Mark d'Inverno, Michael Luck, Pablo Noriega, Juan A. Rodriguez-Aguilar, and Carles Sierra. Communicating open systems. *Artificial Intelligence*, 186:38–94, 2012.

23 US Department of Defense Architecture Framework Working Group DoDAF-WG. Department of defense architecture framework (dodaf), version 1.5. volume i: Definitions and guidelines, April 2007. http://www.defenselink.mil/cio-nii/docs/DoDAF_Volume_I.pdf.

24 Cássia Trojahn dos Santos and Jérôme Euzenat. Consistency-driven argumentation for alignment agreement. In Pavel Shvaiko, Jérôme Euzenat, Fausto Giunchiglia, Heiner Stuckenschmidt, Ming Mao, and Isabel F. Cruz, editors, *OM*, volume 689 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2010.

25 K. M. Eisenhardt. Agency theory: An assessment and review. *Academy of Management Review*, 14(1):57–74, 1989.

26 Jeff A. Estefan, Ken Laskey, Francis G. McCabe, and Danny Thornton. Reference architecture for service oriented architecture. Version 1.0, OASIS, April 2008. http://docs.oasis-open.org/soa-rm.

27 Peter Feiler, Richard P. Gabriel, John Goodenough, Rick Linger, Tom Longstaff, Rick Kazman, Mark Klein, Linda Northrop, Douglas Schmidt, Kevin Sullivan, and Kurt Wallnau. Ultra-large-scale systems: The software challenge of the future. Technical report, CMU Software Engineering Institute, Pittsburgh, 2006. http://www.sei.cmu.edu/library/abstracts/books/0978695607.cfm.

28 Christiane Fellbaum. *WordNet: An Electronic Lexical Database.* Bradford Books, 1998.

29 Frank H. P. Fitzek and Marcos D. Katz. Cellular controlled peer to peer communications: Overview and potentials. In Frank H. P. Fitzek and Marcos D. Katz, editors, *Cognitive Wireless Networks*, pages 31–59. Springer, 2007.

30 American Association for Public Opinion Research AAPOR. The Code of Professional Ethics and Practices, (Revised May, 2010).

31 N. Fornara. Specifying and Monitoring Obligations in Open Multiagent Systems using Semantic Web Technology. In A. Elçi, M. Tadiou Kone, and M. A. Orgun, editors, *Semantic Agent Systems: Foundations and Applications*, volume 344 of *Studies in Computational Intelligence*, chapter 2, pages 25–46. Springer-Verlag, 2011.

32 N. Fornara and M. Colombetti. Representation and monitoring of commitments and norms using OWL. *AI Commun.*, 23(4):341–356, 2010.

33 Nicoletta Fornara and Marco Colombetti. Ontology and time evolution of obligations and prohibitions using semantic web technology. In Matteo Baldoni, Jamal Bentahar, M. Birna van Riemsdijk, and John Lloyd, editors, *DALT*, volume 5948 of *Lecture Notes in Computer Science*, pages 101–118. Springer, 2009.

34 G. Governatori and S. Sadiq. *The journey to business process compliance*, pages 426–445. IGI Global, 2009.

35 Guido Governatori, Francesco Olivieri, Simone Scannapieco, and Matteo Cristani. Designing for compliance: Norms and goals. In Frank Olken, Monica Palmirani, and Davide

Sottara, editors, *Rule - Based Modeling and Computing on the Semantic Web*, volume 7018 of *Lecture Notes in Computer Science*, pages 282–297. Springer Berlin Heidelberg, 2011.

36 Guido Governatori and Antonino Rotolo. Defeasible logic: Agency, intention and obligation. In Alessio Lomuscio and Donald Nute, editors, *Deontic Logic in Computer Science*, number 3065 in LNAI, pages 114–128, Berlin, 2004.

37 Guido Governatori and Antonino Rotolo. BIO logical agents: Norms, beliefs, intentions in defeasible logic. *Journal of Autonomous Agents and Multi Agent Systems*, 17(1):36–69, 2008.

38 Guido Governatori and Antonino Rotolo. Norm compliance in business process modeling. In *Proceedings of (RuleML'2010)*, volume LNCS 6403, pages 194 – 209. 2010.

39 A.E. Hassan. The road ahead for mining software repositories. In *The 24th IEEE International Conference on Software Maintenance, Frontiers of Software Maintenance.*, pages 48 –57, October 2008.

40 Tom Heath and Christian Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool, 2011.

41 P. Hitzler, M. Krötzsch, and S. Rudolph. *Foundations of Semantic Web Technologies*. Chapman & Hall/CRC, 2009.

42 M. Johansson, H. Verhagen, and M. Eladhari. Model of social believable NPCs for teacher training using Second Life. In *Proceedings of the 16th International Conference on Computer Games*, pages 270–274. IEEE, 2011.

43 L. Kaczmirek and N. Schulze. Standards in Online Surveys. Sources for Professional Codes of Conduct, Ethical Guidelines and Quality of Online Surveys. A Guide of the Web Survey Methodology Site. http://websm.org/, 2005.

44 Marcos D. Katz and Frank H. P. Fitzek. Cooperation in 4g networks - cooperation in a heterogenous wireless world. In Frank H. P. Fitzek and Marcos D. Katz, editors, *Cooperation in Wireless Networks: Principles and Applications*, pages 463–496. Springer, 2006.

45 W. Knechel, S. Salterio, and B. Ballou. *Auditing: Assurance and Risk*. Thomson Learning, Cincinatti, 3 edition, 2007.

46 John R. Kuhn and Steve G. Sutton. Continuous auditing in erp system environments: The current state and future directions. *Journal of Information Systems*, 24(1), 2010.

47 Jean-Jacques Laffont and David Martimort. *The Theory of Incentives: the Principal-Agent Model*. Princeton University Press, 2002.

48 Ho-Pun Lam and Guido Governatori. The making of SPINdle. In Guido Governatori, John Hall, and Adrian Paschke, editors, *Rule Representation, Interchange and Reasoning on the Web*, number 5858 in LNCS, pages 315–322, Berlin, 2009. Springer.

49 Ho-Pun Lam and Guido Governatori. Towards a model of UAVs Navigation in urban canyon through Defeasible Logic. *Journal of Logic and Computation*, 2011. to appear.

50 Linden Lab. Second Life home page. `http://secondlife.com/`, 2008.

51 Paul Miller, Rob Styles, and Tom Heath. Open data commons, a license for open data. In *Proceedings of LDOW*, 2008.

52 John Mylopoulos, Lawrence Chung, and Eric S. K. Yu. From object-oriented to goal-oriented requirements analysis. *Communications of the ACM*, 42(1):31–37, 1999.

53 Nachiappan Nagappan, Thomas Ball, and Andreas Zeller. Mining metrics to predict component failures. In *Proceedings of the 28th international conference on Software engineering*, ICSE '06, pages 452–461, New York, NY, USA, 2006. ACM.

54 OOI. Ocean Observatories Initiative, 2007. `http://www.oceanobservatories.org/`.

55 Raul Palma, Jens Hartmann, and Peter Haase. OMV Ontology Metadata Vocabulary for the Semantic Web. Technical report, OMV Consortium, 2008.

56 S. Ranathunga, S. Cranefield, and M. Purvis. Identifying events taking place in Second Life virtual environments. *Applied Artificial Intelligence*, 26(1–2):137–181, 2012.

57 Inmaculada Rodríguez, Anna Puig, and Marc Esteva. Cross-platform management of intelligent objects behaviors in serious virtual environments. *Journal of Visualization and Computer Animation*, 22(4):343–350, 2011.

58 A Rozinat and W.M.P. van der Aalst. Conformance checking of processes based on monitoring real behavior. *Information Systems*, 33(1):64–95, 2008.

59 Bastin Tony Roy Savarimuthu and Stephen Cranefield. Norm creation, spreading and emergence: A survey of simulation models of norms in multi-agent systems. *Multiagent and Grid Systems*, 7(1):21–54, 2011.

60 Bastin Tony Roy Savarimuthu, Stephen Cranefield, Maryam A. Purvis, and Martin K. Purvis. Norm identification in multi-agent societies. Discussion Paper 2010/03, Department of Information Science, University of Otago, 2010.

61 Bastin Tony Roy Savarimuthu, Stephen Cranefield, Maryam A. Purvis, and Martin K. Purvis. Obligation norm identification in agent societies. *Journal of Artificial Societies and Social Simulation*, 13(4):3, 2010.

62 J. R. Searle. *The Construction of Social Reality.* The Free Press, 1995.

63 Murat Sensoy, Timothy J. Norman, Wamberto W. Vasconcelos, and Katia Sycara. OWL-POLAR: A framework for semantic policy representation and reasoning. *Web Semantics: Science, Services and Agents on the World Wide Web*, 12-13:148–160, April 2012.

64 Munindar P. Singh. An ontology for commitments in multiagent systems: Toward a unification of normative concepts. *Artificial Intelligence and Law*, 7(1):97–113, March 1999.

65 Munindar P. Singh. Norms as a basis for governing sociotechnical systems. *ACM Transactions on Intelligent Systems and Technology (TIST)*, pages 1–30, 2013. To appear; available at `http://www.csc.ncsu.edu/faculty/mpsingh/papers`.

66 Munindar P. Singh, Amit K. Chopra, and Nirmit Desai. Commitment-based service-oriented architecture. *IEEE Computer*, 42(11):72–79, November 2009.

67 R. W. Starreveld, B. de Mare, and E. Joels. *Bestuurlijke Informatieverzorging (in Dutch)*, volume 1. Samsom, Alphen aan den Rijn, 1994.

68 Phillip Stoup. The Development and Failure of Social Norms in Second Life. *Duke Law Journal*, 58:331–344, 2008.

69 Talis. ODC Public Domain Dedication and License. Technical report, Open Knowledge Foundation, 2008.

70 TNS. Two-day batter life tops wish list for future all-in-one phone device. Technical report, Taylor Nelson Sofres, September 2004.

71 Tomas Trescak, Marc Esteva, and Inmaculada Rodríguez. A virtual world grammar for automatic generation of virtual worlds. *The Visual Computer*, 26(6-8):521–531, 2010.

72 M. A. Vasarhelyi, M. Alles, and A. Kogan. Principles of analytic monitoring for continuous assurance. *J. of Emerging Technologies in Accounting*, 1(1):1–21, 2004.

73 Roel J. Wieringa and John-Jules Ch. Meyer. Applications of deontic logic in computer science: a concise overview. In *Deontic logic in computer science: Normative system specification*, pages 17–40. John Wiley & Sons, Inc., New York, USA, 1994.

74 Michael Wooldridge. *An introduction to Multi-Agent Systems.* John Wiley & Sons, 2009.

75 Konrad Wrona and Petri Mähönen. Analytical model of cooperation in ad hoc networks. *Telecommunication Systems*, 27(2–4):347–369, October 2004.

76 Pamela Zave and Michael Jackson. Four dark corners of requirements engineering. *ACM Transactions on Software Engineering and Methodology*, 6(1):1–30, 1997.