

Norms in MAS: Definitions and Related Concepts

Tina Balke¹, Célia da Costa Pereira², Frank Dignum³,
Emiliano Lorini⁴, Antonino Rotolo⁵, Wamberto Vasconcelos⁶, and
Serena Villata⁷

- 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



© T. Balke, C. da Costa Pereira, F. Dignum, E. Lorini, A. Rotolo, W. Vasconcelos, and S. Villata; licensed under Creative Commons License CC-BY

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 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 impossible¹. 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 save a victim or stay out to save 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 triggered². 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 law⁴:

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 norms⁵. 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 transferred 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 actions 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 science⁷. 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 richer 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 Carneades⁸, 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 $b_1, \dots, b_n \Rightarrow l$ 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 $b_1, \dots, b_n \Rightarrow l$ (if b_1, \dots, 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, \dots, 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 $b_1, \dots, b_n \Rightarrow l$ 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 norms¹⁰. 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.

¹⁰These 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 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 Forschungszentrum 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. Johnson. *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.