Multiagent and Grid Systems – An International Journal 7 (2011) 21–54
DOI 10.3233/MGS-2011-0167
IOS Press

Norm creation, spreading and emergence: A survey of simulation models of norms in multi-agent systems

Bastin Tony Roy Savarimuthu* and Stephen Cranefield
Department of Information Science, University of Otago, Dunedin, New Zealand

Abstract. Norms in human societies are expectations of behaviours of the individuals. In human societies, there are several types of norms such as moral norms, social norms and legal norms (laws). In multi-agent systems, software agents are modelled as possessing characteristics and behaviour borrowed from human societies. In order to design and develop robust artificial agent societies, it is important to understand different approaches proposed by researchers by which norms can spread and emerge within agent societies. This paper makes three contributions to the study of norms. Firstly, based on the simulation research on norms, we propose a life-cycle model for norms. Secondly, we discuss different mechanisms used by researchers to study norm creation, identification, spreading, enforcement and emergence. We also discuss the strengths and weaknesses of each of these mechanisms. Thirdly, in the context of identifying the desired characteristics of the simulation models of norms we discuss the research issues that need to be addressed.

Keywords: Norms, agents, simulations, multi-agent systems, survey

1. Introduction

In human societies, norms have played an important role in governing the behaviour of the individuals in a society [41]. Norms are the societal rules that govern the prescription and proscription of certain behaviour. Norms improve cooperation [45] and coordination among individuals [107].

In multi-agent systems, software agents are modelled as possessing characteristics and behaviour borrowed from human societies such as autonomy, pro-activeness and the ability to communicate with other agents [116]. The social construct of norms has been used by researchers in multi-agent systems to study how cooperation can be achieved [99]. Norms reduce the amount of computation required by the agents [46] as the agents do not have to search their entire state space of possible actions and their effects if they choose to follow norms, and the behaviour of other agents should be more predictable than when norms are absent.

Artificial multi-agent societies are societies in a networked environment where agents share a virtual space and perform certain actions in a particular context (e.g. auctions). These agent societies are modelled using some of the social constructs borrowed from human society. There have been two approaches for building normative behaviour in an agent. The first approach is the prescriptive approach where an institutional mechanism specifies how the agents should behave. The second approach is the

*Corresponding author. E-mail: tonyr@infoscience.otago.ac.nz.

bottom-up approach by employing mechanisms that can help norms to emerge and govern the behaviour of agents.

The advent of digital virtual environments such as Second Life [85] call for a distributed approach to norm spreading and emergence. A centralised policing mechanism for such digital societies would be expensive from the viewpoint of the computation required, due to the explosion of the combined states of the agents. It is computationally infeasible or at the least resource intensive to monitor and control millions of agents playing numerous roles through a centralised enforcer. A distributed approach to norms addresses these problems. Both centralised and distributed approaches have been studied by researchers and the recent works in this area focus on the issues associated with the distributed approach.

A "norm-capable" agent society is one that is able to generate, distribute, enforce and modify norms. Building robust agent societies that can create and evolve norms is important because the framework that helps in recognizing these norms will also be helpful for the agents to dynamically change these norms if situations warrant it. A good approach for testing models of norm-capable societies is simulation. So, a first step towards building such norm-capable societies is to understand the existing simulation research on norms. The motivation for considering the simulation-based research on norms stems from the fact that implemented systems primarily use *designer-specified norms* [100]. On the other hand, a wide variety of mechanisms are explored using simulation.

This paper is organised as follows. A background on norms in human societies and multi-agent societies are provided in Sections 2 and 3 respectively. Based on the simulation works on norms, we propose a life-cycle model for norms in Section 4. In Section 5 we categorise the research work on norms based on the mechanisms employed by each research. In Section 6 we discuss some of the research issues that need to be addressed in the simulation-based works on norms.

2. What are norms?

Norms are expectations of an agent about the behaviour of other agents in the society [14]. Human society follows norms, such as the exchange of gifts at Christmas. Norms have been so much a part of different cultures, it is not surprising that it is an active area of research in a variety of fields (see Fig. 1) including Sociology [33,44,61], Economics [2,46,78], Biology [10,21,31,75], Philosophy [14,58,113], Law [42,43] and Computer Science [99,114].

2.1. Norms in human societies

Due to the multi-disciplinary nature of norms, several definitions for norms exist. Habermas [57], a renowned philosopher, identified norm-regulated actions as one of the four action patterns in human behaviour. A norm to him means *fulfilling a generalised expectation of behaviour*, which is a widely accepted definition for social norms. A behavioural expectation is *generalized* if every member of a social group expects all others to behave in a certain way in a given situation. Ullmann-Margalit [107] describes a social norm as a prescribed guide for conduct or action which is generally complied with by the members of the society. She states that norms are the result of complex patterns of behaviour of a large number of people over a protracted period of time. Coleman [33] writes *"I will say that a norm concerning a specific action exists when the socially defined right to control the action is held not by the actor but by others"*. Elster notes the following about social norms [45]: *"For norms to be social, they must be shared by other people and partly sustained by their approval and disapproval. They are sustained by the feelings of embarrassment, anxiety, guilt and shame that a person suffers at*

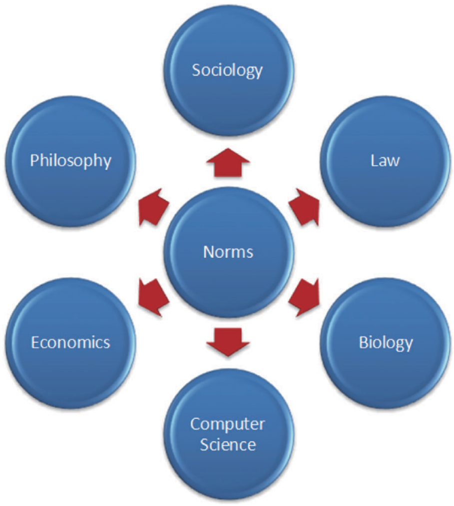

Fig. 1. Fields of study of norms.

the prospect of violating them. A person obeying a norm may also be propelled by positive emotions like anger and indignation... social norms have a grip on the mind that is due to the strong emotions they can trigger". What is common in these definitions is the expectation that an agent behaves in a certain way in a situation and the appropriate behaviour is dictated by the group. There is not any consensus on the level of social control on norm violation. Habermas's definition does not talk about norm violation. Coleman's definition mentions social control without any specifics. Elster's definition explicitly mentions the approval and disapproval of agents on other agent's behaviour.

Researchers have divided norms into different categories. Tuomela [106] has grouped norms into two categories: social norms and personal norms. Social norms define the behaviour of the group and are associated with sanctions. Personal norms are based on the personal beliefs of the individuals. Personal norms are the potential social norms. These norms could become social norms if they were to be observed by other agents and if sanctions were associated with not following the norm. Social norms are further classified into r-norms (rule norms) and s-norms (social norms). Personal norms are categorised into m-norms (moral norms) and p-norms (prudential norms). Rule norms are imposed by an authority based on an agreement between the members (e.g. one has to pay taxes). Social norms apply to large groups such as a whole society and they are based on mutual belief (e.g. one should not litter). Members of a society expect that a social norm be followed by other members of the society. Moral norms appeal to one's conscience (e.g. one should not steal or accept bribes). Prudential norms are based on rationality (e.g. one ought to maximize one's expected utility). When members of a society violate societal norms, they may be punished or even ostracised in some cases [39].

Many social scientists have studied why norms are followed. Some of the reasons for norm adherence include:

– fear of authority or power [10]
– rational appeal of the norms [1,13]
– emotions such as shame, guilt and embarrassment that arise because of non-adherence [45]
– willingness to follow the crowd [46].

In this paper, we focus on social norms because the agents in multi-agent systems have been modelled using ideas borrowed from social concepts such as speech act theory [96], collaboration and

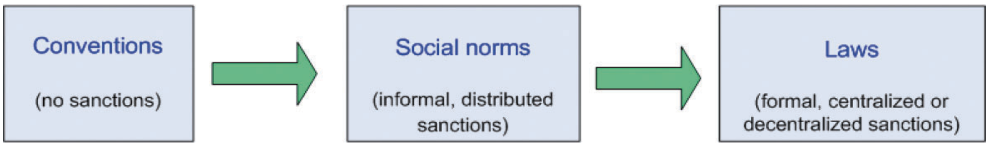

Fig. 2. Progression from conventions to laws.

cooperation [79].

Based on the definitions provided by various researchers, we note that the notion of a social norm is generally made up of the following three aspects:

- **Normative expectation of a behavioural regularity**: There is a general agreement within the society that a behaviour is expected on the part of an agent (or actor) by others in a society, in a given circumstance.
- **Norm enforcement mechanism**: When an agent does not follow the norm, it could be subjected to a sanction. The sanction could include monetary or physical punishment in the real world which can trigger emotions (embarrassment, guilt, etc.) or direct loss of utility. Other kind of sanctions could include agents not being willing to interact with an agent that violated the norm or the decrease of its reputation score. Agents that follow the norm might be rewarded.
- **Norm spreading mechanism**: Examples of norm spreading factors include the advice from powerful leaders and entrepreneurs, and the cultural and evolutionary influences. For an external observer, agents identifying and adopting norms through learning mechanisms such as imitation may also appear to spread norms in agent societies.

2.2. *Conventions vs. social norms vs. laws*

It should be noted that researchers are divided on what the differences between a social norm and a convention are. Gibbs [53, p. 592] notes that *the terms "convention" and "custom" are frequently employed in the discussions of norms, but there does not appear to be any consensus in definitions of them beyond the point that they may not be sanctioned*. We will assume that a convention is a common expectation amongst (most) others that an agent should adopt a particular action or behaviour (e.g. the convention in ancient Rome was to drive on the left). As conventions gain force, the violations of conventions may be sanctioned at which point a social norm comes into existence. For example, if driving on the right is sanctioned, the left-hand driving becomes a norm. A norm may become a law when it is imposed by an institution (e.g. laws that govern driving behaviour). Our view on the relationship between conventions, social norms and laws is given in Fig. 2. Note that when conventions are established, they are not associated with sanctions (e.g. style of dress, dinner table etiquette). However, the conventions may become social norms once they are enforced by other agents due to the *expectation* of a particular behaviour. The enforcement happens at the peer-to-peer level (decentralized enforcement). When a norm has emerged in a society it may be institutionalized (i.e. it becomes a law [48]), which will then be formally enforced by a central authority (e.g. the government) that makes use of distributed legal entities such as a police department and the justice department.

2.3. *Norm life-cycle*

In the body of research literature on social norms, there is no unified view on how norms are created and spread in a society and various models of norms have been proposed [14,33,45,48,63,81]. According to Coleman [33] *"Norms are macro level constructs based on purposive actions at the micro level but coming into an existence through a micro-to-macro transition. Once in existence they lead, under certain conditions to actions of individuals which affect the utilities and thus the actions of the individuals to whom the sanctions have been or might be applied"*.

In the context of their study on international relations, Finnemore and Sikkink [48] have proposed a three-stage model of the norm life-cycle: the norm emergence stage, the norm cascade stage and the norm internalization stage. The first stage is the norm emergence stage which is characterised by the persuasion of other agents to follow the norm by some norm entrepreneurs or norm innovators. Norm entrepreneurs are the innovators who think about new norms in a society (e.g. Henry Dunant the founder of Red Cross was the entrepreneur of the norm to treat wounded soldiers in a war as neutrals). Norm entrepreneurs attempt to convince a critical mass of norm leaders to embrace new norms. The motives of the entrepreneurs to come up with these norms include altruism and empathy. The second stage is the norm cascade stage characterised by the dynamics of imitation as the *norm leaders* attempt to socialise with other agents whom they might have influence over, so they might become followers. Followers may take up the norm because, following the norms might enhance their reputation and also their own esteem. They may also follow the norm because of the peer pressure from other followers. The third stage is the norm internalization stage where the norms are widely accepted by the agents in the society to the extent that they might be taken for granted (i.e. the norm following becomes an automatic task for the followers). An issue with the internalized norms is that these norms can then become hard to discern from some behavioural regularity as there is no discussion in the society about whether a norm should be followed.

Note that the researchers have broadly identified three phases of the norm life-cycle. In the context of describing the simulation-based works on norms we will revisit and extend the phases of the norm life-cycle model and discuss various mechanisms employed by the researchers in Section 4.

3. Normative multi-agent systems (NorMAS)

Research on norms in multi-agent systems is about two decades old [19,26,37,40,99,100]. Norms have been of interest to multi-agent system (MAS) researchers as they help in maintaining social order [36] and facilitating cooperation [11] and coordination [99,114]. Since norms enable smoother functioning of the societies by facilitating social order, MAS researchers have used this concept to build multi-agent systems. They have also investigated how norms may evolve in response to environmental changes.

3.1. Definitions

The definition of normative multi-agent systems (NorMAS) as described by the researchers involved in the NorMAS 2007 workshop is as follows [17]. *"A normative multiagent system is a multiagent system organised 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"*.

In NorMAS, many research works treat norms as constraints (hard or soft) on actions that an agent can perform [15]. Some researchers view norms as hard constraints where agents are not permitted to violate norms [18]. A more common view is that norms are soft constraints where an agent has the ability to violate norms.

3.2. Two branches of research in NorMAS

Researchers in normative multi-agent systems have been influenced from two different perspectives: philosophy of law (prescriptive approach) and conventionalistic approach (emergence approach) [35]. Based on these two perspectives, research in normative multi-agent systems can be categorized into two branches. The first branch focuses on normative system architectures, norm representations, norm adherence and the associated punitive or incentive measures. The second branch is concerned with the emergence of norms.

Several architectures have been proposed for the study of norms [16,72]. Researchers have created these different architectures to study and test their intuitions about norms [77]. Lopez et al. [72] have designed an architecture for normative BDI agents. Boella et al. [16] have proposed a distributed architecture for normative agents. Some researchers have used deontic logic to define and represent norms [16, 52,65,115]. Other work has investigated mechanisms for norm compliance and enforcement [4,10, 71]. A recent development is the research on emotion-based mechanisms for norm enforcement [49,95, 104]. For a detailed comparison of selected normative architectures, refer to Neumann's article [77]. Neumann has categorized selected normative architectures based on a) the theoretical backgrounds of the architectures, b) the viewpoints of the architectures (single agent vs. agent society), c) the reason for following a norm (deontic vs. consequentialistic view),[1] d) the consideration of static vs. dynamic societies, and e) the ability to deal with conflicts.

Even though the first branch studies how norms are formalized and represented, it does not address the question of where the norm comes from (i.e. how a norm emerges in a society). Some researchers have proposed mechanisms by which norms can emerge in an agent society [98,110]. Thus the second branch of research deals with the simulation-based approaches to norms. This branch of work differs from the first branch in terms of the different mechanisms explored by the researchers (e.g. leadership, reputation, machine learning, imitation) and the experiments that are conducted using these mechanisms. The emergence of norms is explored only by the research of this branch. We note that much of the work on norms in this branch does not make any distinction between conventions and norms [88,98,99,111] – both conventions and norms are included under the umbrella of norms. Most work on the emergence of norms (mainly conventions) are from a game-theory perspective [10,98,99]. Neumann has presented a case study of four research works on simulated models of norms from the perspective of foundations of social theory [76]. Four papers were investigated in detail resulting in the identification of three methodological core problems which are norm transmission, norm transformation and the function of the norm. The first two problems correspond to the causal aspect of the norm (i.e. what causes the norm to spread). The last problem deals with the purpose of the norm. The author concludes that no model has been able to fully explain both the causal and functional reasons behind norm emergence, however, the current trend is towards trying to bridge this gap.

Conte et al. [35] have worked on an integrated view of norms. Their work tries to bridge the gap between the prescriptive view of norms (first branch) and the emergence of conventions (second branch), using the cognitive abilities of an agent. They have proposed a logic-based framework to integrate these two perspectives. However, concrete implementations of this integrated approach are yet to be seen.

[1] Deontic view of norms advocates that norms are in itself a reason for action. On the other hand in the consequentialist view, actions are judged by their consequences (e.g. based on the utility of the actions).

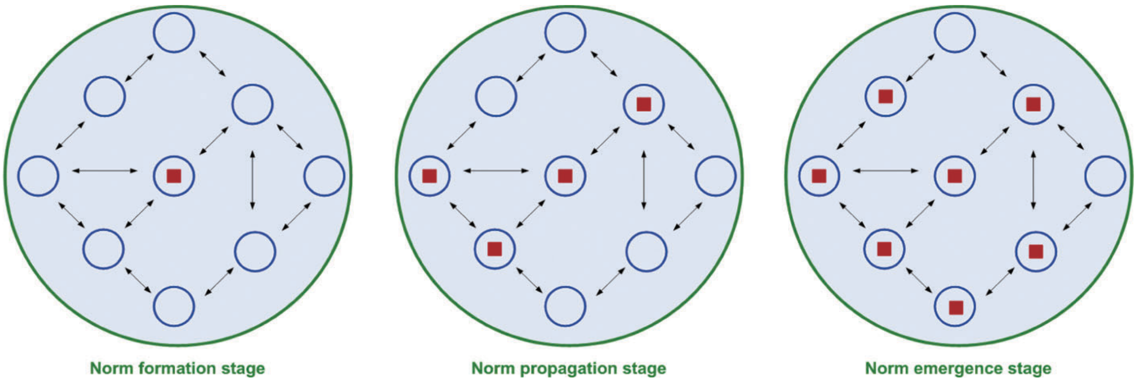

Fig. 3. Three main stages of the norm life-cycle.

4. Developmental phases of norms based on simulation studies

Broadly, from the view point of the society, the three important stages of norms are the formation stage, propagation stage and the emergence stage. Researchers employing simulation-based approaches to norms have investigated various mechanisms associated with norms with each of these stages. Mechanisms employed in the norm formation stage aim to address how agents can *create* norms in a society and how individual agents can *identify* the norms that have been created. Mechanisms used in the norm propagation stage aim to explain how norms might be *spread* and *enforced* in the society. The emergence stage is characterized by determining the extent of the spread of a norm in the society.

Figure 3 shows an overview of these three stages from the view point of a society (or a bird's eye view). The larger green circles represent agent societies. The smaller blue circles represent agents and the red squares inside the blue circles represent norms. The bi-directional arrows represent interactions between agents. The first green circle depicts an agent society in the norm formation stage. This society has nine agents. An agent in this society has created a norm (the blue circle that has a solid red square inside). The second green circle shows the agent society in its norm propagation stage. The norm from the agent in the middle of the society has propagated to other agents it is connected to (i.e. the agents it interacts with). The last green circle shows the agent society in the norm emergence stage assuming that the threshold for norm emergence from the viewpoint of the external observer of the society is 75%.

Based on these three important stages of norms, we identify five phases (i.e. expanded stages) of the norm life-cycle[2] which are norm creation, identification, spreading, enforcement and emergence as shown in Fig. 4. Even though there has not been any agreement on these phases by researchers, we use these five phases, as they broadly capture the processes associated with the norm life-cycle. Figure 4 shows the five phases of the norm life-cycle on the left and the mechanisms investigated by researchers for each of the phases on the right. We have categorized the mechanisms used in the simulation-based works on norms into nine main categories (marked with a * in Fig. 4).

This section provides an overview of the five developmental phases of norms, and the next section provides a detailed discussion of the mechanisms studied by researchers in each of these phases.

[2]We use the term norm life-cycle to capture the important aspects of a norm from its creation to its establishment and de-establishment in the society.

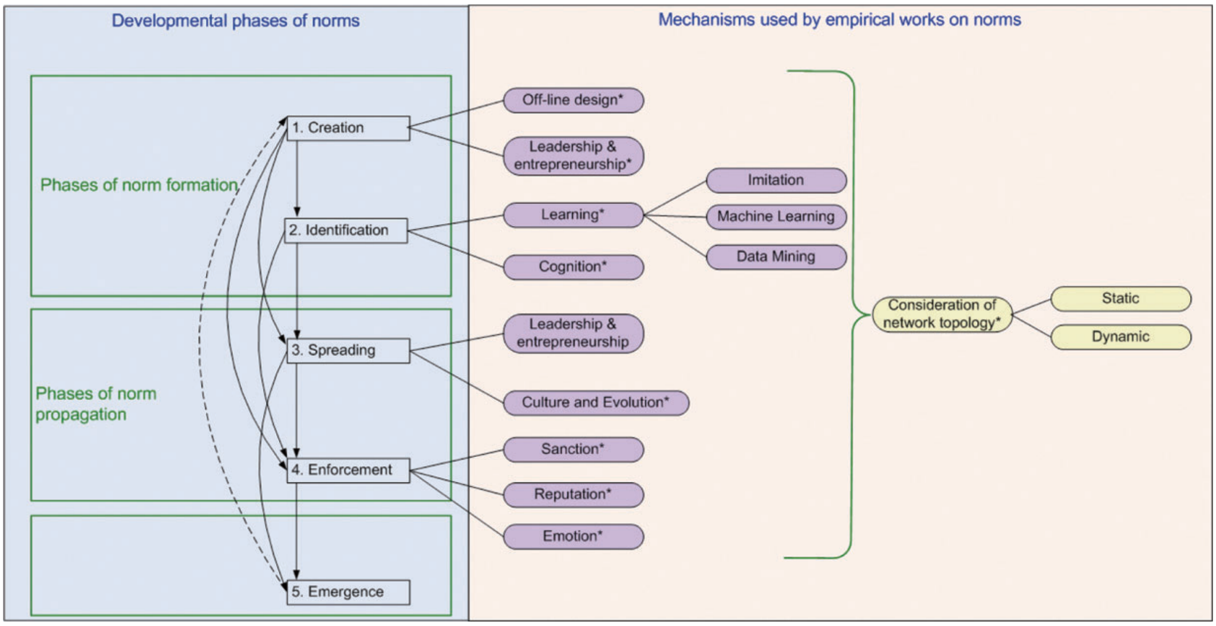

Fig. 4. Phases of norm life-cycle and categories of simulation models.

4.1. Norm creation

The first phase of the life-cycle model is that of norm creation. The norms in multi-agent systems are created by one of the three approaches. The three approaches are a) a designer specifies norms (off-line design) [34], b) a norm-leader specifies norms [19,110], c) a norm-entrepreneur considers that a norm is good for the society [62]. In the off-line design approach, norms are designed off-line, and hard-wired into agents. This approach has been used by researchers to investigate norms that might be beneficial to the society as a whole using social simulations. In leadership approach, some powerful agents in the society (the norm-leaders) create a norm. The leadership approach can be based on authoritarian or democratic leadership. The leader can provide these norms to the follower agents [19,91,109]. In the entrepreneurship approach to the creation of norms, there might be some norm entrepreneurs who are not necessarily the norm leaders but create a proposed norm.[3] When an agent creates a new norm it can influence other agents to adopt the norm [48,62].

4.2. Norm identification

If a norm has been created in the society using one of the explicit norm creation approaches discussed Section 4.1, then the norm may spread in the society. However, if the norms have not been explicitly created (i.e. norms are derived based on the interactions between agents), then an agent will need a mechanism to identify norms from its environment based on the interactions with other agents. In game-theory based empirical works [98,99], agents have a limited number of actions that are available,

[3]At this stage, the norm exists only in the mind of one agent (i.e. it is a personal norm). It hasn't become a social norm that is accepted by other agents.

and they choose the action that maximizes their utility as the norm, based on some learning mechanism such as imitation, machine learning or data-mining.

The second approach to norm identification considers the cognitive capabilities of an agent to infer what the norms of the society are [6,7]. In the cognitive approach, one or more cognitive agents in a society may come up with norms based on the deliberative processes that they employ [6,7,93,94]. In this approach the agents have the cognitive ability to recognize what the norms of a society are based on the observations of interactions. Agents have normative expectations, beliefs and goals. It should be noted that the norms inferred by each agent might be different (as they are based on the observations that the agent has made). Thus, an agent in this model creates *its* own notion of what the norms are based on inference.

4.3. Norm spreading

Norm spreading relates to the distribution of a norm among a group. Once an agent knows what the norm in the society is (i.e. either based on norm creation or identification), several mechanisms help in spreading the norms such as leadership, entrepreneurship, cultural, and evolutionary mechanisms (explained in detail in Section 5). It should be noted that for an external observer, agents identifying norms through learning mechanisms such as imitation appear to spread norms in agent societies.

4.4. Norm enforcement

Norm enforcement refers to the process by which norm violators are discouraged through some form of sanctioning. A widely used sanctioning mechanism is the punishment of a norm violator (e.g. monetary punishment which reduces the agent's fitness or a punishment that invokes emotions such as guilt and embarrassment). Reputation mechanisms have also been used as sanctions, such as where an agent is black-listed for not following a norm. The process of enforcement helps to sustain norms in a society.

Note that enforcement of norms can influence norm spreading. For example, when a powerful leader punishes an agent, others observing this may identify the norm. Hence, the norm can be spread. Norms can also be spread through positive reinforcements such as rewards. Some researchers have considered enforcement as a part of the spreading mechanism [10] (see Section 5.6 for a detailed discussion).

4.5. Norm emergence

The fifth phase is the norm emergence phase. We define norm emergence to be reaching some significant threshold in the extent of the spread of a norm; that is a norm is followed by a considerable proportion of an agent society and this fact is recognised by most agents. For example, a society could be said to have a norm of gift exchange at Christmas if more than x% of the population follows such a practice. The value of x varies from society to society and from one kind of norm to another. The value of x has varied from 35 to 100 across different simulation studies of norms (see Table 3).

Emergence can be detected either from a global view of the system or through a local view[4] of an agent (e.g. an agent might only see agents that are one block away on all directions in a grid environment[5]). Spreading of norms with or without enforcement can lead to emergence. Once a norm has emerged, the

[4] Note that the norms observed in the local view could be different from the norms that can be observed from a global view.
[5] Agents in a particular environment can be connected to one another using one of many topologies such as regular, small-world, scale-free and fully-connected topologies [74].

process can continue when an entrepreneur or a leader comes up with a new norm that replaces the old one. This is indicated by a dotted arrow in Fig. 4.

The adoption of a norm may decrease in a society due to several reasons. A norm that has emerged may lose its appeal when the purpose it serves does not hold or when there are not enough sanctions or rewards to sustain the norm or when other alternate effective norms emerge. Note that the model presented here is from a bird's-eye view (i.e. an external agent that observes the society). An external agent will able able to observe the norm establishment and de-establishment in the society based on the emergence criterion (i.e. the extent of spread of the norm).

4.6. Consideration of network topologies

An important attribute of the research works on norms is the consideration of network topology. The underlying interaction topology of agents has an impact on all phases of norm development. For example the interactions between a leader and his followers have an implicit network topology (i.e. fully-connected network) which governs how norms created by the leader may spread and may lead to norm emergence in an agent society. Hence the consideration of network topology is included as one of the nine main categories.[6] The network structure of the society can either be static or dynamic (i.e. can evolve due to agents joining and leaving).

4.7. Discussion

The life-cycle that we have presented is similar to Finnemore and Sikkink's model [48] described in Section 2.3. As the reader may observe, their model is a subset of the life-cycle model that we have proposed. Finnemore and Sikkink's model caters only for the entrepreneurial approach for norm creation and the imitation approach for norm spreading. However, in our life-cycle model more mechanisms are brought under each of the phases. For example mechanisms based on emotions, culture and evolution are included in our model.

Another distinction between the models is the view point of the life-cycle. Finnemore and Sikkink's model includes the social mechanisms employed by *human agents* (e.g. entrepreneurship, imitation, reputation). Our model is based on a socio-computational viewpoint which includes modeling social mechanisms from a computational viewpoint and also studying pure computational techniques in the simulation study of norms which are enacted by *software agents*. For example offline design approaches and machine learning mechanisms are only applicable to our model. These mechanisms can be used to study phenomena which may otherwise be difficult without the help of computational modeling and abstractions.

While addressing how norms are created and spread, the proposed life-cycle model can also accommodate the process of norm change. A norm entrepreneur can come up with a modified norm which can be spread by one of the spreading mechanisms, which may lead to the replacement of the older norm with a new one, or cognitive agents might notice a change in norms due to a change in the society's membership.

---

[6]Unlike other categories such as off-line design and learning, the consideration of network topology is not strictly a *mechanism* that is relevant to one or two phases of norm development but a consideration that may have an impact on any phase of norm development (i.e. network topology is an orthogonal consideration). However, as we believe *network topology* is an important aspect in the study of norms we have included it as one of the categories in our categorization (see Fig. 4).

Table 1
Mechanisms used in different phases of the norm life-cycle (Yes: considered in the model, –: not considered/specified)

Empirical works	Creation	Identification	Spreading	Enforcement	Emergence
Axelrod, 1986 [10]	–	–	Evolution	Sanction	Yes
Shoham and Tennenholtz, 1992 [99]	–	Machine learning	–	–	Yes
Kittock, 1993 [68]	–	Machine learning	network topology	–	Yes
Conte and Castelfranchi, 1995 [34]	Off-line	–	–	–	–
Walker and Wooldridge, 1995 [114]	–	Machine learning	–	–	Yes
Shoham and Tennenholtz, 1995 [100]	Off-line	–	–	–	–
Castelfranchi et al., 1998 [27]	Off-line	–	–	Reputation	–
Verhagen, 2001 [110]	Leadership	–	Leadership	–	–
Epstein, 2001 [46]	–	Imitation	–	–	Yes
Flentge et al., 2001 [50]	–	–	Cultural transmission	Sanction	Yes
Hales, 2002 [60]	Off-line	–	–	Reputation	–
Hoffmann, 2003 [62]	Entrepreneurship	Machine learning	Entrepreneurship	–	Yes
Lopez et al., 2002, Lopez 2003 [70,71]	Off-line	–	–	Sanction and reward	–
Nakamaru and Levin, 2004 [75]	–	Machine learning, imitation	network topology	–	Yes
Chalub et al., 2006 [31]	–	Machine learning	Evolution, network topology	–	Yes
Fix et al., 2006 [49]	–	–	–	Emotion	–
Pujol, 2006 [82]	–	Machine learning	network topology	–	Yes
Sen and Airiau, 2007 [98]	–	Machine learning	–	–	Yes
Andrighetto et al., 2010, Campenní et al., 2008 [6,24]	–	Cognition, imitation	–	–	Yes
Savarimuthu et al., 2009 [90]	–	Machine learning	Leadership, network topology	–	Yes
Savarimuthu et al., 2010 [93,94]	Off-line	Cognition, Data mining	–	Sanction	No

Table 1 provides an overview of the contributions of selected simulation-based research works to these different phases of the norm life-cycle. It should be noted that not all phases of the norm life-cycle have been taken into account by most works, and some works make use of more than one mechanism in a single phase.

5. Categorization of simulation works on norms

In this section we categorize empirical works on norms[7] into nine main categories as shown in Fig. 5 (marked with a *).[8] For each of these categories, we provide a brief description and discuss a few key papers. It should be noted that some papers have made use of mechanisms that fall under more than one category (e.g. Axelrod's work [10]), and also a category may contribute to different phases of the norm life-cycle (i.e. leadership and entrepreneurship mechanisms can be used to facilitate norm creation and spreading).

5.1. Off-line design approaches

Off-line design models are characterised by the agents of the society possessing explicit knowledge of the norms. The intention of this approach is to seed agents with norms and compare how the society performs when the whole society possesses certain norms as opposed to agents behaving strategically (without possessing the notion of norms). One of the well-known works on norms specified by the designer is by Shoham and Tennenholtz [100], who experimented with norms associated with traffic. Several other researchers [34,60,114] have experimented with an off-line design approach borrowing the basic experimental set-up proposed by Conte and Castelfranchi [34]. Conte and Castelfranchi [34] have shown using their simulation experiments what the function of a norm is in the context of agents finding food in a grid environment characterised by simple rules for movement and food collection. An agent's strength increases when it consumes food and its strength decreases when it moves from one cell to another. This work compared the utilitarian strategy with the normative strategy and showed that norms reduce the aggression level of the agent (when a finders-keepers norm is followed) and also increase the average strength of an agent.

The work of Conte and Castelfranchi [34] assumes that an agent society is either made up of strategic agents or normative agents. Lopez et al. [71] have extended the notion of off-line design by experimenting with an agent society by varying the types of agent personalities.

Discussion – Off-line design models are best suited for studying and comparing different normative schemes in a closed agent society. However, agents inhabiting open and distributed environments may not have the privilege of consulting a designer. Walker and Wooldridge [114] note the following about the off-line design of norms. "*[This] approach will often be simpler to implement and might present the designer with a greater degree of control over system functionality. However, there are a number of disadvantages with this approach. First, it is not always the case that all the characteristics of a system are known at design time; this is most obviously true of open systems. . . . . Secondly, in complex*

[7] Even though we distinguish norms from conventions (see Section 2.2), for the purpose of categorization, we have incorporated both conventions and norms under the umbrella of *norms*.

[8] These are the same nine categories shown in Fig. 4.

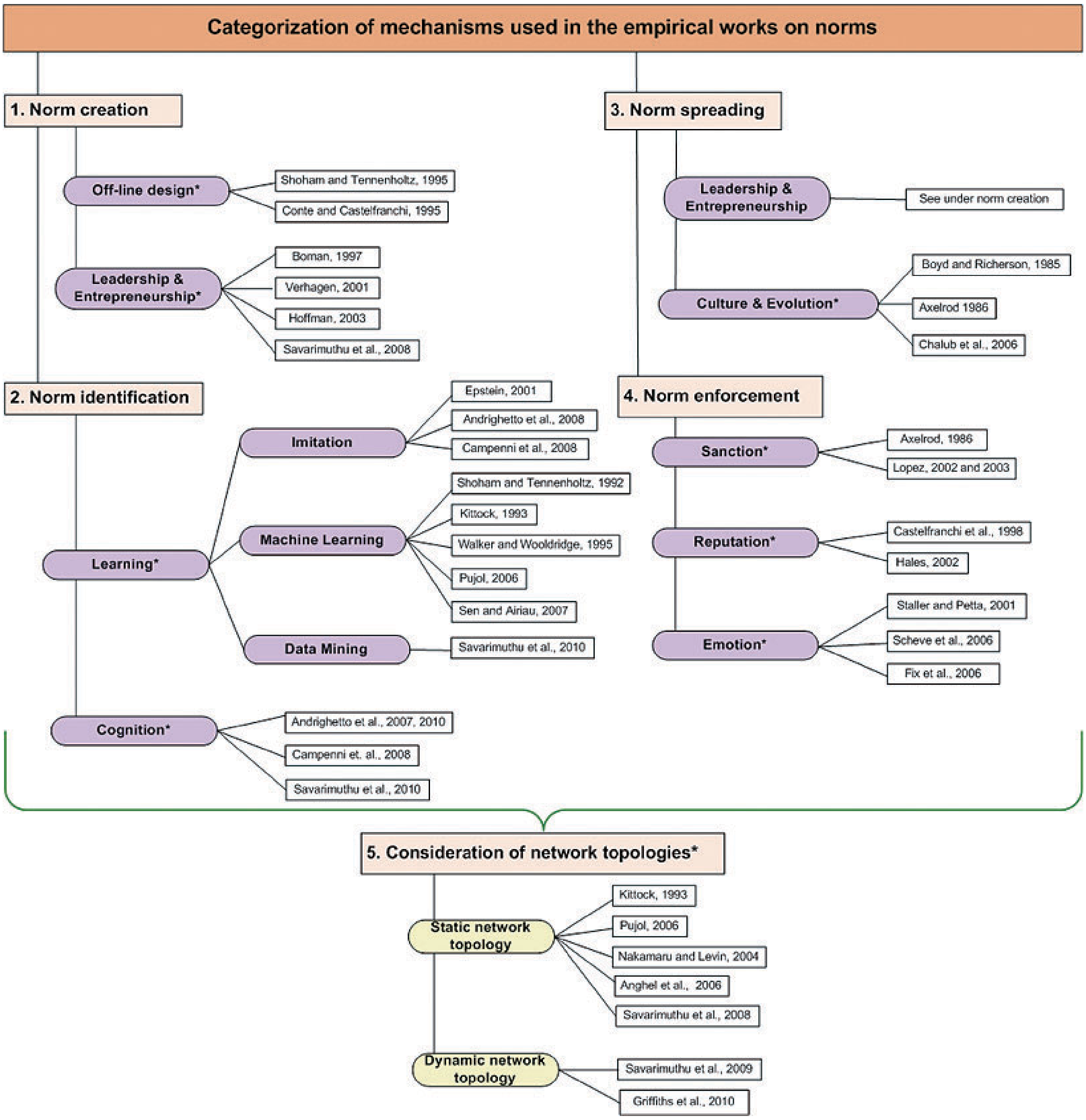

Fig. 5. Categorization of empirical works on norms.

systems the goals of agents... might be constantly changing. To keep reprogramming agents in such circumstances would be costly and inefficient. Finally, the more complex a system becomes, the less likely it is that system designers will be able to design effective social laws". Some researchers have used this approach to compare the performance of a normative system with a non-normative one [34].

Another limitation of the off-line design mechanism is that it assumes all the agents in the society will follow a norm (e.g. the finders-keepers norm) which might not be realistic. Although this is possible for a well entrenched norm in a society, open societies might have different competing norms that are present at a given point of time.

5.2. Leadership and entrepreneurship mechanisms

Social power plays an important role in societies in establishing order and enabling smoother functioning. Several researchers of normative multi-agent systems have focused on the notion of power [25, 29,66] such as institutional power. López in her thesis on social power and norms notes that the social powers of an agent are expressed through its abilities to change the beliefs, the motivations, and the goals of other agents in such a way that its goals can be satisfied [70]. Sources of power could motivate, encourage, or persuade their followers to take up a particular norm (the leadership approach) or coerce them to adopt a particular norm based on sanctions (the punishment approach). Researchers have used leadership approaches for norm creation and spreading and have also experimented with sanction approach for norm enforcement (see Section 5.6).

Leadership mechanisms are based on the notion that there are certain leaders in the society, who provide advice to the agents in the society. The follower agents seek the leaders' advice about the norm of the society.

Boman [19] has used a centralised approach, where agents consult with a normative advisor before they make a choice on actions to perform. Verhagen [110] has extended this notion of normative advice to obtaining normative comments from a normative advisor (e.g. the leader of the society) on an agent's previous choices. The choice of whether to follow a norm and the impact of the normative comment on an agent are determined by the autonomy of the agent. Once an agent decides to carry out a particular action, it announces this decision to all the agents in the society, including the leader of the society, and then carries out that action. The agents in the society can choose to send their feedback to this agent. When considering the received feedback, the agent can choose to give a higher weight to the feedback it received from the leader agent. Verhagen has experimented with the internalization of norms in an agent society. Internalization refers to the extent to which an agent's personal model of a norm matches the group model of a norm.

Savarimuthu et al. [91] have adopted a distributed approach for norm emergence. In their mechanism, there could be several normative advisors (called role models) from whom other agents can request advice. In this model, an agent can be a leader for some agents while that agent itself can be a follower of some other agent. The model is based on a utilitarian notion that the agent that performs the best in a given neighbourhood is chosen to be the norm leader. The role model can recommend its norm to its followers who ask for advice. Depending upon a leader's neighbourhood, this approach may allow different norms to appear in each of the neighbourhoods.

Hoffmann [62] has experimented with the notion of norm entrepreneurs who think of a norm that might be beneficial to the society. An entrepreneur can recommend a norm to a certain percentage of the population (e.g. 50%) which leads to varying degrees of establishment of a norm. This model assumes that the agents in the society are willing to converge towards a norm. If their current norm deviates from the group norm (which is published by a centralised mechanism), an agent decrements the usefulness of its norm and may even choose another norm from the list of norms (or rules) available from a pool. Hoffmann's experiments explore the entrepreneurial norm dynamics and provide some initial evidence for Finnemore and Sikkink's norm life-cycle model [48]. Some shortcomings of this model as acknowledged by the authors include the assumption of a single norm entrepreneur, the lack of communication between agents about norms, and the use of a centralised monitor to compute consensus.

Discussion – The leadership models assume that a powerful authority is present in the society and all agents in the society acknowledge the power of such agents. Both centralised and distributed notions of norm spreading using *power* have been employed. The centralised approach is suitable for closed societies. However, this might not work well for open, flexible and dynamic societies. Distributed approaches for norm spreading and emergence are promising because the computational costs required to spread, monitor and control a norm are distributed across the individual agents.

Another criticism of the centralised leadership mechanism in empirical models is that it assumes that an *all knowledgeable* authority is present in the society. Though this might be how some aspects of human societies are modelled (e.g. an institution), it would be challenging to model an agent (human or artificial) that might think of possible norms and recommend the one that might be the best so that others could use them.

5.3. Learning mechanisms

Three types of learning mechanisms have been employed by researchers: imitation, machine learning and data mining.

5.3.1. Imitation mechanisms

The philosophy behind an imitation mechanism is *When in Rome, do as the Romans do* [46]. These models are characterised by agents mimicking the behaviour of what the majority of the agents do in a given agent society (following the crowd). Epstein's main argument [46] for an imitation mechanism is that individual thought (i.e. the amount of computing needed by an agent to infer what the norm is) is inversely related to the strength of a social norm. This implies that when a norm becomes entrenched the agent can follow it without much thought. Epstein has demonstrated this in the context of a driving scenario in which agents can observe each other's driving preference (left or right) based on a certain observation radius r. If the agent sees more agents driving on the right within the observation radius, it changes to the right. When a norm is established, the observation radius becomes one (i.e. the agent looks at one agent on its right and left to update its view about the norm). Other researchers have also experimented with imitation models [6,70,82].

Discussion – Imitation might be a good mechanism when agents want to avoid the cost of reasoning about what the norm of the society is. An agent using the imitation model is not involved in the creation of the norm. It is just a part of the norm spreading effort based on lazy identification (i.e. copying others without much thought). Other researchers have noted that an imitation approach cannot bring about the co-existence of multiple norms in a society [24,75]. This issue has to be scrutinised further because Epstein has shown that imitation can result in the creation of certain pockets of local norms even though a global consensus has not been arrived at. Another issue for debate is whether imitation-based behaviour (solely) really leads to norms as there is no notion of generalized expectation.

5.3.2. Works based on machine learning

Several researchers have experimented with agents finding a norm based on learning on the part of an agent [98,99,114].

Shoham and Tennenholtz [99] were the first in multi-agent systems research to experiment with norm emergence. They viewed a norm as a social law which constrains actions or behaviours of the agents in the system. They used a mechanism called co-learning which is a simple reinforcement learning mechanism based on a Highest Cumulative Reward (HCR) rule for updating an agent's strategy when playing a simple coordination game and a cooperation game (prisoner's dilemma). According to this rule, an agent chooses the strategy that has yielded the highest reward in the past *m* iterations. The history of the strategies chosen and the rewards for each strategy is stored in a memory of a certain size (which can be varied). They experimented with the rate at which the strategy is updated (after one iteration, two iterations, three iterations, etc.). When the frequency of update decreases, convention emergence decreases. They experimented with flushing the memory of the agent after a certain number of iterations and retaining only the strategy from the latest iteration. They found that when the interval between memory flushes decreases, the efficiency of the convention emergence decreases. One limitation of this model is that the agents do not have knowledge of the function of the norm (they are merely following an algorithm). Also, no notion of sanction or reward for violating or following the norms is included.

The experimental model of Walker and Wooldridge [114] is based on the work done by Conte et al. [34] where agents move about a grid in search of food. They experimented with 16 mechanisms for norm emergence. Their model used two parameters: the majority size and the strategic update function. Each of these parameters can be varied across four values. 16 experiments were based on the size of the majority (simple, double, quadruple, dynamic) and the nature of the update function (using majority rule, memory restart, communication type and communication on success). The model's novelty is that it considered the role of memory and communication mechanisms. Also, the agents learn from their own interaction experience with other agents. Visibility of an agent is restricted to a particular region (or neighbourhood) which is governed by the "extroversion radius" (similar to the one suggested by Shoham and Tennenholtz [101]). They concluded that further experiments should be done to enhance the understanding of this complex topic. They noted that the role played by social structure (i.e. the network topology) and communication should be considered in future work.

Sen and Airiau [98] proposed a mechanism for the emergence of norms through social learning. They experimented with three reinforcement learning algorithms and the agents learned norms based on private local interactions. They observed that when the population size is larger, the norm convergence is slower, and when the set of possible action states is larger, the convergence is slower. They also studied the influence of the dynamic addition of agents with a particular action state to a pool of existing agents, as well as norm emergence in isolated sub-populations.

Discussion – Machine learning mechanisms employ a particular algorithm to identify a strategy that maximizes an agent's utility and the chosen strategy is declared as the norm. Since all agents in the society make use of the same algorithm, the society stabilises to a uniform norm. Agents using this approach cannot distinguish between a strategy and a norm. These agents accept the strategy that maximizes its utility as its norm. However, the agents do not have a notion of normative expectation associated with a norm (i.e. when agents expect certain behaviour on the part of other agents). This issue can be addressed by making an agent deliberate about a learnt norm (i.e. how the newly learnt norm might affect its desires, goals and plans).

Another weakness of these works is that agents lack an explicit representation of norms. Hence, they are not able to communicate norms with others. This weakness has been noted by several researchers [28, 89]. Although the issue of norm communication between agents has been considered by an early work using an appropriate communication protocol [114], explicit norm representation has not been considered by most empirical works on norms.

5.3.3. Data mining mechanism

Agents can also use a data mining approach to identify norms in agent societies. Agents in open agent societies can identify norms based on what they infer based on their observations of the society. The repository of an agent's observations can then be mined for patterns of behaviour. There has been a proposal of an agent architecture for normative systems to employ data mining for citizens of a country to find information and norms from official documents [105]. However, the work does not describe what types of norms are discovered and also the mechanisms used in the identification of norms.

Savarimuthu et al. [93,94] have proposed an architecture for norm identification which employs association rule mining, a data mining approach. The architecture makes use signals (sanctions and rewards) as the starting points for norm identification. Mechanisms for identifying two types of norms, prohibition norms and obligations norms have been studied. The details on how an agent identifies a prohibition norm are explained in the context of a public park scenario, where the norm against littering is identified by the agent. The obligation norm inference is explained in the context of a tipping norm in a restaurant scenario. They have demonstrated that an agent using the proposed architecture can dynamically add, remove and modify norms based on mining the interactions that take place between agents. They have show that agents can identify co-existing and conditional norms. The agents can also identify the normative pre-conditions (conditions that have to be true for the norm to hold).

Discussion – Data mining is a promising approach for the identification of some types of norms that can be inferred based on observing the interactions between agents in the society. However, if actions that explicitly signal a sanction or reward are absent or other mechanisms such as reputation are used instead of explicit signals (i.e. reduction in the reputation score of a rogue agent instead of sanctioning), then it is difficult to identify norms.

5.4. Cognitive approach

Researchers involved in the EMIL project [7] are working on a cognitive architecture for norm emergence to explore how the mental capacities of agents play a role in the emergence of norms. We note that this is the only architecture that has been proposed to study the emergence of norms.

The EMIL project aims to deliver a simulation-based theory of norm innovation, where norm innovation is defined as the two-way dynamics of an inter-agent process and an intra-agent process. The inter-agent process results in the emergence of norms where the micro interactions produce macro behaviour (norms). The intra-agent process refers to what goes on inside an agent's mind so that they can recognise what the norms of the society are. This approach is different from learning models, as the agents in the cognitive approach are autonomous and have the capability to examine interactions between agents and are able to recognise what the norms could be. The agents in this model need not necessarily be utility maximizing like the ones in learning models. The agents in this model will have the ability to filter external requests that affect normative decisions and will also be able to communicate norms to other agents. Agents just employing learning algorithms lack these capabilities.

Andrighetto et al. [6] have demonstrated how the norm-recognition module of the EMIL platform answers the question "how does a agent come to know of what a norm is?". In particular they have experimented with an imitation approach versus the norm-recognition approach that they have come up with. The norm recognition module consists of two constructs: the normative board and a module for storing different types of messages (which the authors call "modals") that can be used to infer norms. The messages that are exchanged between agents can be of five different types (e.g. the deontics modal refers to partitioning situations as either acceptable or unacceptable). The normative board consists of normative beliefs and normative goals, which are modified based on the messages received. They have shown that norm recognisers perform better than social conformers (imitating agents) because the recognisers were able to identify a pool of potential norms while the imitators generated only one type of norm.

The limitation of this approach is that agents just observe actions performed by other agents. In practice they should be able to learn from their own experience as well. Perhaps, their own experience can be given a higher weight. At present, agents in this model do not have the capability of violating norms and hence there are no costs associated with sanctions. The authors note this as a potential extension.

In the work done by Campenni et al. [24], the experimental set-up used by Andrighetto et al. [6] was extended.[9] In the original set-up, agents moved across four different contexts. In each context an agent chose one of the three available actions. In each context, two of the three actions were unique to that context, while one action was common to all four contexts. So, there were nine actions in total. In an extension of this work [24], barriers were used to prevent agents moving from one context to another. The agents were restricted to one of the contexts. The authors then compared how norms emerged with and without the barriers. It was noted that without the barriers, only one type of norm emerged (all the agents converged to the action that was common to all contexts), but when the barrier was used, three possible normative beliefs were generated (one for the common action, and two for two other actions). This result is interesting because, one would think, five different types of normative beliefs would be generated (one for the common context and one for each of the four contexts). However, it is not clear from the paper why only a total of three normative beliefs were created.

The work done by Savarimuthu et al. [92–94] parallels the work that is being carried out by the researchers involved in the EMIL project. They proposed an architecture where agents can identify what the norms of the society are in a bottom-up fashion. While many well-known works identify one norm that exists in the society [10,46,62] this architecture is able to identify several norms that might exist in the society. This work has also addressed how an agent might be able to identify whether a norm is changing in a society and how it might react to this situation. In this model, the agents are able to identify the norm change and dynamically add, remove and modify norms. This architecture can be used to study norm emergence.

Discussion – Cognitive mechanisms are promising because agents with this type of mechanism have the notion of normative expectation. This mechanism focuses on what goes on inside the mind of an agent to infer norms (i.e. the computational machinery used by the agent to infer norms). Agents infer norms when they join new societies and deliberate about norms. Agents can also suggest a new norm based on their past experience and may bring about norm change.

5.5. Cultural and evolutionary mechanisms

Researchers have proposed cultural and evolutionary models [20,31] for norm spreading and emergence. Boyd and Richerson [20] proposed that norms can be propagated through cultural transmission. According to them, there are three ways by which a social norm can be propagated from one member of the society to another. They are

- Vertical transmission (from parents to offspring)
- Oblique transmission (from a leader of a society to the followers)
- Horizontal transmission (from peer to peer interactions).

[9] The details of the experimental set-up originally appeared in the paper [5] for which a copy cannot be found on the Internet.

Of these three kinds of norm transmission mechanisms, vertical and oblique transmissions can be thought of as leadership mechanisms in which a powerful superior convinces the followers to adopt a norm. Horizontal transmission is a peer-to-peer mechanism where agents learn from day-to-day interactions from other peers. Few researchers have used this idea to experiment with norm spreading (e.g. the idea of normative advice used by Verhagen [109] and Savarimuthu et al. [91]).

Norm spreading based on evolution involves producing offspring that inherit the behaviour of their parents. One well known work in this category is by Axelrod [10]. Other researchers have also experimented with evolutionary models for norm spreading [31,111].

Chalub et al. [31] studied how norms might spread in different societies (e.g. an archipelago of islands). In their experiments, agents in an island were fully connected to each other. Each agent played the donor-receiver game once with all other agents in the island. Then an agent reproduced by choosing a connected agent at random and comparing their payoffs. If its payoff was higher than the other agent's, then the other agent inherited the strategy of the winning player. Each island had a Gross Domestic Product (GDP) (i.e. the score of the island was represented as the GDP) which was a normalised average payoff of the entire island population at end of playing the game. Islands which were fully connected competed against each other. There were times of war and peace. During peace, the norms of the islands did not change. When the islands were at war, they played the Hawk and Dove game [103]. The losers changed their norm based on a probabilistic norm-update rule. This model resulted in the emergence of one common social norm which induces reputation-based cooperative behaviour in these islands.

Discussion – Even though cultural models provide an answer to how norms are spread (e.g. normative advice from parents to children), they do not explain how a norm is internalized in the first place. There is an implicit assumption that the norms are internalized by the parents and are passed on to their progeny. A limitation of evolutionary models is that they do not distinguish a strategy from a norm. In other words, these models lack the conceptualization of normative expectation. Some researchers have studied the effect of gene-cultural co-evolution [54]. Kendal et al. [67] have studied the cultural co-evolution of norm adoption and enforcement when punishers are rewarded or non-punishers are punished. Still, the issues that are not addressed remain the same. Chalub et al. [31] note that they believe it is important to investigate how the cognitive capacity of an agent impacts the evolution of social norms as the cognitive capacity of humans allows them to set and enforce norms. Work along these lines is in progress (see works of Andrighetto et al. [6,7] and Savarimuthu et al. [92–94]).

5.6. Sanction mechanisms

Even though the models discussed in Section 5.2 are based on the notion of leadership, they do not include the ability to sanction agents that do not follow the norm specified by a norm leader. Several works on norms have used the notion of social power to inflict sanctions on agents that do not follow a norm [10,50,71].

In his well known work, Axelrod [10] has shown how a meta-norm of *punishing an agent who did not enforce a norm* can stabilise the norm of cooperation in an agent society. When an agent A sees another agent B violating a norm, then the agent A observing this violation can either punish or ignore this violation. If another agent C sees A not punishing B, it can choose to punish B with a certain probability. When this mechanism was introduced, the norm stabilised in every run of the experiment. The norm of cooperation was not always sustained without the use of a meta-norm mechanism. This shows that a meta-norm mechanism is useful for sustaining a norm. Axelrod also notes that when the cost of punishments are low for the punishers a norm can be sustained. This has been verified by other researchers (e.g. [50]).

López et al. [70,71] have used the motivation of an agent and autonomy to be the main driving forces behind norm compliance and have considered punishments and rewards in their model. In their scheme agents have goals and norms. Their framework models agents with different personalities (social, pressured, opportunistic, greedy, fearful, rebellious, simple imitation, reasoned imitation and reciprocity). The model assumes that sanctions or rewards can only be used as an enforcing mechanism if they affect an agent's goals. The authors have shown the effect of norm compliance by varying different types of agents through simulations based on two parameters, the social satisfaction and individual satisfaction which respectively indicate the significance of norm compliance from the points of view of the society as well as an individual agent. One limitation of this work is that the cost of sanctions is not explicitly considered on the part of the agent imposing this punishment. They assume that all punishments and rewards are applied by someone else in a society.

Flentge et al. [50] have shown how an agent comes to acquire a possession norm. In this system an agent can possess a piece of land. The authors have shown that the probability for the survival of the population is much higher when possession claims of others are respected. In the short term, an agent can benefit from violating the possession norm. They have noted that sanctions help in the establishment of the possession norm when the sanctioning costs are low or when there is no cost for sanctioning.

Discussion – The role of sanctions is to help sustain norms. In centralised institutional mechanisms there are special *norm policemen* who ensure that norm violations do not take place (e.g. governor agents in the AMELI framework [9]), and the institution bears the cost of enforcement. However, in a distributed punishment environment as employed in open agent societies, except for altruistic agents, the cost of sanctions may be high for the punishing agents. Altruistic agents have their own utility function that is different from the selfish agents (e.g. an altruistic agent's utility increases when it performs a good deed). Even though the work in this area has begun [47,80], a proper account of the cost of punishment should be provided by future works that employ a sanction mechanism.

Another issue is that sanctions have an impact on an agent's autonomy. Agents may have to give up some of their autonomy when they are forced to take up a norm in order to avoid the reduction of their utilities through sanctions. Researchers have modelled different personalities of agents to address this issue (e.g. social agents, pressured agents) [38,71].

5.7. Reputation mechanisms

Reputation refers to the positive or negative opinion about a person or agent based on their interactions with others in the society. Researchers have addressed how reputation models are beneficial in sustaining norms in an agent society [27,60]. They have experimented with the effect of a notion of "normative reputation" on the compliance costs of the norm. They have shown that providing a mechanism to record the normative reputation of agents of the society helps in redistributing the costs of norm compliance to both the agents that follow the norms as well as those who do not follow the norms.

The context of interaction of agents in the work of Castelfranchi et al. [27] was a food-consumption problem where the food source randomly appears in a grid and the agent closest to it can either see or smell the food. Strategic agents were compared to normative agents. Strategic agents would attack another agent consuming food if that agent was weaker. Normative agents on the other hand were those that followed norms and in this case they followed the finders-keepers norm. Castelfranchi et al. [27] showed that if the population had pure strategies (either completely strategic or normative), the normative agents always had a better utility at the end of the game as aggression costs both the aggressor and the victim. For mixed strategies, the strategic agents always fared better than the normative agents because the strategic agents took advantage of the normative compliance of the agents. The researchers then introduced the notion of normative reputation where each agent learns whether another agent is normative or strategic. A normative agent shares this information with another normative agent. This helped the normative agents to identify the strategic agents. When the normative agents encountered a strategic agent, they did not follow the norm. In this way, they distributed the cost of norm compliance to the agents that did not observe the norm. Thus, the agents that used the normative reputation fared better in mixed strategy scenarios.

Hales [60] has shown that group reputation supports the creation of beneficial norms. He extended the model described by Castelfranchi et al. [27] to address the food-consumption game by stereotyping agents into homogeneous groups based on reputation. Agents belonged to one of two groups: the normative group or the strategic group (i.e. cheaters). The concept of group is similar to the tagging mechanism where each group can be thought of as having a common tag (i.e. each agent belongs to a certain tag group [59]). When an agent in the normative group interacts with an agent in the cheating group, the agent will sense that the opponent is a cheater. It then associates the cheating action with the group and communicates this information to all the other agents in the group. When other norm-following agents interact with agents in the cheating group, they will use the cheating strategy instead of the normative strategy. By this process, the agents are able to distribute the cost of norm compliance to the cheaters. The weakness of this approach is that even if one of the agents in the norm-following group becomes a *black sheep*, then the whole society is labelled as a cheating society because this model allows one agent's misbehaviour to tarnish the reputation of the whole society. Also, it does not allow a group to revert back if it was stereotyped as a cheating group.

Discussion – Agents using reputation models store the reputation of the agents that they interact with. This can be either an individual reputation or a group reputation. Other researchers in multi-agent systems field have used multi-level reputation through referrals [117]. Both the models discussed above do not take into account the personal cost incurred by an agent for maintaining the history of interactions with other agents that is used for computing reputation. However, reputation is a simple decentralised mechanism and agents can deliberate on the cost of computing the reputation by varying the amount of history they would like to maintain.

5.8. Emotion-based mechanisms

Some researchers have modelled agents with emotions [49,95,104]. Based on the work done by Scheve et al. [95], Fix et al. [49] discuss the micro-macro linkage between emotions at the micro-level and norm enforcement at the macro-level. The authors argue that emotions have a norm-regulatory function in agent societies. An emotional agent that observes deviation of another agent from a norm might experience emotions such as contempt or disgust which can be the motivation behind sanctions. Agents that are sanctioned might experience emotions such as shame, guilt or embarrassment which might lead to norm internalization. The authors have used a Petri net model [64] to capture the micro-macro linkage. It should be noted that the proposed model has not been implemented in the context of a simulation experiment. Staller et al. [104] extended the experimental set-up of Conte and Castelfranchi [34] by including emotion-based strategies.

Discussion – Not many works in MAS have considered emotional agents since not all domains require agents to be emotional. Emotional agents are interface agents that interact with human users (e.g. a virtual patient avatar interacting with a doctor trainee). In emotion-based models for norms, it is assumed that the punisher and the norm-violator know what the norms of the society are. Only if they had known what the norms are, could any kind of emotion be generated. Initially, a norm could be a personal norm in the mind of one agent (e.g. a norm entrepreneur or a normative advisor). To start with, an agent might not know what the norms of a society are until it is punished a certain number of times for the deviant behaviour. It should be noted that emotion detection itself can be a complex problem, but this can be simplified if it can be assumed that these emotions are visible through external actions (e.g. an agent *yells* at a litterer in a park). Emotions can serve as a starting point for norm inference (see the work of Savarimuthu et al. [92]).

5.9. Research using network topologies

Social networks are important for norm spreading and emergence because in the real world, people are not related to each other by chance. They are related to each other through the social groups that they are in, such as work groups, church groups, ethnic groups and hobby groups. Information tends to percolate among the members of the group through interactions. Also, people seek advice from a close group of friends and hence information gets transmitted between the members of the social network.

In most empirical works, the treatment of norms has been mostly in the context of an agent society where the agents interact with all the other agents in the society in a random fashion [19,109]. Few researchers have considered the actual topologies of the social network for norm emergence [82]. We believe such an approach is important for the study of norm spreading and emergence, as networks provide the topology and the infrastructure on which the norms can be exchanged. Researchers have studied different kinds of network topologies and their applications in the real world (an overview of different topologies is given by Mitchell [74]). These application areas include opinion dynamics [51] and the spread of diseases [32]. Researchers in normative multi-agent systems have started to look at the role of network topologies [68,82,97,111]. Network topologies have also been explored by other multi-agent system researchers in other contexts, such as reputation management [83,117].

Research that has considered network topologies can be categorised into static and dynamic network topology approaches. In the static approach, the network topology is fixed. In the dynamic topology approach, the underlying network can change when the simulation experiments are conducted.

5.9.1. Research using a static network topology

Kittock was the first to experiment on the role of network topology in convention emergence [68]. Agents interacted with each other based on their position in a circular lattice. The agents learned about the best strategy to choose based on the HCR algorithm proposed by Shoham and Tennenholtz [101]. Kittock noted that the choice of the global structure has a profound effect on the evolution of the system. Depending upon the topology of the network, the emergence of a convention varies. In particular, he conjectured that the diameter of a network is directly related to the rate of convergence, which was observed by other researchers later [3,12].

Nakamaru and Levin [75] studied how two related norms evolved in networked environments. The two related norms were based on a) the background of an agent (e.g. the religion that the agent follows) and b) the opinions that the agent holds (e.g. opinions about fast food, political affinity). The background is a norm that the whole population shares but different opinions can be held by agents in that population. They note that when people of the same background meet, they might change some of their opinions and when all opinions of two agents are the same, they can change their background if they are different. They conducted experiments using four different types of network topologies and observed how the two related norms evolved on those networks. They showed that the spread of social norms is influenced not only by the structure of the network topology but also by the learning mechanism that is used. They experimented using mechanisms of social learning (imitation) and individual learning (learning based on interactions with other agents). They showed that imitation does not lead to co-existence of social norms (which has also been reported by Campenní et al. [24]), while individual learning does. They also showed that individuals learn faster using imitation than when using individual learning. They observed that norms tend to propagate the fastest on a scale-free network.

Anghel et al. [8] investigated the effects of inter-agent communication across a network in the context of playing the minority game [30]. They showed that a scale-free leadership structure emerges on top of a random network. Pujol [82] dealt with the emergence of conventions on top of social structures. He used the HCR mechanism proposed by Shoham and Tennenholtz [101] to test norm emergence in connected, random, small world and scale-free networks. He also demonstrated that the structure of the network is crucial for norm emergence. Recently several researchers [87,97,112] have investigated the role of static network topologies on norm emergence in different domains.

5.9.2. Dynamic topology works

Very few researchers have investigated the role of dynamic network topologies on norm spreading and emergence. Griffiths and Luck [56] have experimented on how norms might emerge on a network topology where an agent rewires its links by replacing its worst neighbours with the best neighbours of its immediate neighbours. Even though the work considers the links of a node to be rewired, the work does not model an open agent society where new agents can join and leave the society. Another limitation of this work is that the concept of tag is viewed as a norm (i.e. agents belonging to a tag group have the norm). This approach does not allow multiple norms to co-exist in a group.

Savarimuthu et al. [90] used a model of Gonzalez et al. [55] to create dynamic network topologies. Gonzalez et al. developed a model for constructing dynamically changing networks using the concept of agents (modelled as particles) colliding in an abstract social space to construct evolving networks. Savarimuthu et al. [90] created dynamic network topologies using this model to test their agent-based role model leadership mechanism. They showed how different types of norms emerge when societies with different norms for the same context (playing the Ultimatum game [102]) are brought together. In particular, they showed that under certain conditions norms can co-exist in an agent society.

5.9.3. Discussion

One limitation of static topologies is that they assume that the underlying network topology does not change. In the real world and in artificial agent societies, the social network topologies are dynamic. People join and leave social groups, so any mechanism for norm emergence should be applicable to dynamically changing network topologies. Thus, models for norm spreading and emergence should be tested on top of dynamically changing network topologies. Another limitation is that the weights of the links are not considered. The strength of the links from an agent to other agents might be different which may influence norm spreading and emergence. Autonomy of the agents is also not considered in most of these works. Some agents can just refuse to participate in norm spreading. If they are the important nodes such as a hub, this will have implications on the spread of the norm. In most of the models, individuals belong to only one group but in reality the agents can be influenced from many angles as an individual might belong to several groups (e.g. work group, neighbours and hobby group). Each of these groups could have a different topology.

6. Discussion of desired characteristics of simulation models and future work

In the NorMAS 2007 workshop [17] it was discussed that social norms will play a major role in regulating behaviour in virtual environments such as Second Life [85]. It was reiterated that the current trend of research in normative agent systems is that of a interactionist view (a bottom-up approach) as opposed to the earlier notion of the legalistic view (a top-down approach). Recently, Boella et al. [15] have proposed ten guidelines for the development of normative multi-agent systems. In this section of the paper, we focus on the desired characteristics of simulation models of norms. As a motivation for the discussion on desired agent characteristics to operate with norms, let us consider a hypothetical scenario relating to a social norm in a virtual environment:

Divya Chandran, a new Second Life resident, wants to explore the rich interactions offered by the new medium. She wants to go to a virtual park and relax by the fountain hearing chirping birds. She flies to the virtual park and sees people enjoying the sun. She notices some water fountains and some soft-drink fountains from the sponsor of the park. She would like to get a drink, but does not know if there is a norm associated with using the fountain. She wonders if she should get a cup from the jazzy sponsor's booth by paying money or if she needs to acquire the skill of making the cup object. Once she fills the cup with the drink, can she enjoy her drink in all areas of the park or is she restricted to a particular zone. And finally, what should she be doing with the empty cup? What is the norm associated with littering in the park? Can she drop it anywhere for an autonomous robot to pick it once the cup is dropped or should she find a rubbish bin and drop it? Will the norm of this park be applicable to all the parks in Second Life? When she visits the park at a later date will the norm of the park still be the same?

The important questions that the scenario described above poses is how will Divya know what the norms of using the park are and what would happen if she does not follow the norm of the society. In essence, what should be the capabilities that should be built into an agent to recognise and reason about the implications of the norm.

In this section we describe the characteristics that are desirable for the inclusion of norms in a simulated multi-agent system world. Tables 2 and 3 show a comparison of the characteristics of norm simulation studies.

6.1. Richer representations of norms

Although norm representation has been well studied by deontic logic researchers[10] in the multi-agent system research community [65,115], the norm representations used in the simulation models are all based on simple data types (see column 8 of Tables 2 and 3). This is mainly because keeping a representation simple reduces the complexity of the model and reduces the computation required. Additionally it allows exploration of certain properties which may not be easily understood with more complexity.

In Axelrod's work [10], the norm against cheating was inferred based on the boldness and vengefulness parameters possessed by the agent. These parameters were integer values. The norm representation was not explicitly stated by Axelrod, but it can be inferred by the reader. In Tables 2 and 3, whenever the reader has to infer what the norm representation might have been, we have used italicised text (see column 8).

[10] Norms in deontic logic are expressed as permissions, prohibitions and obligations. For example, $P_A(litter|home)$ implies that agent A is permitted to litter given that she is at her home. $F_A(litter|park)$ implies that A is prohibited from littering the park. $O_A(pay|win)$ implies that A is obliged to pay after winning an item in an auction.

Table 2
Comparison of characteristics of simulation works on norms (Yes – considered, No – not considered, NA – Not Applicable, NS – Not Specified)

Model	Context of Interaction	Autonomy	Fixed or Dynamic society (size)	Norm communication	Norm co-existence	Network topology	Norm representation	Emergence criterion
Axelrod, 1986 [10]	Norms game	Yes	Fixed (20)	No	No	Random	*Integers*	100%
Shoham and Tennenholtz, 1992 [99]	Coordination game	No	Fixed (100)	No	No	Random	*Integer*	85–95%
Kittock, 1993 [68]	Coordination game, Prisoner's dilemma game	No	Fixed (10–10000)	No	No	Circular lattice	*Integer*	90%
Conte and Castelfranchi, 1995 [34]	Food searching game	No	Fixed (50)	No	No	Random	*Boolean*	NA
Walker and Wooldridge, 1995 [114]	Food searching game	No	Fixed (100)	Yes	No	Random	*Boolean*	73–99%
Castel-franchi et al., 1998 [27]	Food searching game	No	Fixed (50)	Yes	No	Random	*Boolean*	NA
Verhagen, 2001 [110]	Resource consumption game	Yes	NS	Yes	No	Fully connected	Decision tree	NS
Epstein, 2001 [46]	Coordination game	No	Fixed (191)	No	Yes	Ring	*Binary*	Varies
Hales, 2002 [60]	Food searching game	No	Fixed (50)	Yes	No	Random	*Boolean*	NA
Hoffmann, 2003 [62]	Pick a number game	No	Fixed (10–50)	No	No	Fully connected	Integer	70%
López et al., 2002, 2003 [71,70]	NS	Yes	NS	No	No	NA	NS	NS

In some research, a norm is a binary variable, e.g. to indicate whether an agent drives on the left or right [46,98]. Castelfranchi et al. [27] used a Boolean variable to represent a norm. Chalub et al. [31] used a string representation for norms. In complex, dynamic and open agent societies, these representations may not be sufficient even at the simulation level. In the context of park littering, agents observe what kinds of norms might evolve in the usage of a park. It could happen that littering is prohibited in general. One kind of norm could be that whenever an agent litters within 20 metres from the location of a rubbish bin, it should be punished. It could also be that whenever there is no one within ten metres of the agent's vicinity and no rubbish bin could be found within 20 metres of the agent then littering would be permitted. In these cases, the representation of the norm needs to be flexible and dynamic and for this reason a richer model of norm representation would be required. This dynamic model of norm representation should allow for norms to change at run-time. It should also allow more subtle contextual norms to emerge.

6.2. *Adjustable agent autonomy*

Even though autonomy is one of the core characteristics of an agent, not all the normative agent models

Table 3
Comparison of characteristics of simulation works on norms (Yes – considered, No – not considered, NA – Not Applicable, NS – Not Specified)

Model	Context of Interaction	Auto-nomy	Fixed or Dynamic society (size)	Norm commun-ication	Norm co-exist-ence	Network topology	Norm repre-senta-tion	Emer-gence criterion
Nakamaru and Levin, 2004 [75]	Background and opinion exchange	Yes	900 to 10000	No	No	Complete mixing, lattice, random, scale-free	Binary string	Varies
Chalub et al., 2006 [31]	Donor-Receiver game, Hawk-Dove game	Yes	Fixed (128)	No	Yes	Fully connected	8-bit string	95–98%
Pujol, 2006 [82]	Coordination game	No	Up to 100000	No	No	Random, regular, small-world, scale-free	Binary	90%
Sen and Airiau, 2007 [98]	Coordination game	No	Fixed (100), Dynamic	No	Yes	Random	*Binary*	Mostly 100%
Campenní et al., 2008 [24]	4 contexts	Yes	Fixed (100)	Yes	Yes	Random	*Boolean*	NS
Savarimuthu et al., 2009 [90]	Ultimatum game	Yes	Fixed (200)	Yes	No	Fully connected, random, scale-free	Integer	100%
Andrighetto et al., 2010 [6]	4 contexts	Yes	Fixed (100)	Yes	No	Random	*Boolean*	NS
Savarimuthu et al., 2010 [93,94]	Park littering, Tipping in restaurants	No	Fixed (100)	Yes	Yes	Random	*String*	NA

have considered the autonomy of the agents (see column 3 of Tables 2 and 3). An agent participating in a normative system should have the ability to say *no* and violate a norm when it does not want to obey the norms (because the norm may prevent the satisfaction of an urgent goal or may lower the agent's utility drastically). The notion of adjustable autonomy has been discussed by Verhagen [109]. There is always the dilemma on the part of an agent whether to give up some of its autonomy in order to be a part of a group or to achieve a common goal. Still, the agents that are a part of a group that obey certain norms should be able to identify whether a norm applies in certain situations. For example, when the traffic light turns red at 3am and the agent does not see any other agent in an intersection, it should be able to decide whether to respect the norm. Similarly, in the park-littering scenario the agent should decide whether to litter when it does not see anybody within its vicinity or when it does not want to waste its energy to find a rubbish bin. In essence, the notion of dynamic autonomy should be built into the simulation models.

6.3. *Dynamic change of norms*

One issue that has not been addressed by the simulation work on norms is the dynamic change of norms. None of the works that are based on simulations address how a norm is made obsolete or how a new norm replaces an old one.[11] Similar to human societies, virtual agent societies should allow norms to change or evolve (either based on change in the environment or due to finding a superior norm). Normative simulation systems that cater for the dynamic change of norms will be useful for understanding the conditions under which new norms take over the older ones both in human and artificial agent societies.

6.4. *Explicit communication of norms*

Research employing game-theoretical models assumes that an agent explicitly knows the setting and options available to other agents (i.e. the pay-off matrix). In these works, there is no explicit norm communication. Hence, there is no need for explicit representation of norms. Only few works have considered the notion of norm communication [60,114] (see column 5 of Tables 2 and 3). Norm communication will be beneficial to newly joining agents as well as other agents in the society. Newly joining agents can ask other agents about the norms in the society and the existing agents can use communication channels to enquire if a norm is currently in force. Communication of norms will require explicit representation of norms in the mind of an agent.

6.5. *Facilitating an open environment*

Even though most researchers on norm simulation point to open systems as their target environment, their simulations only used a fixed number of agents (see column 4 of Tables 2 and 3). Under an open world assumption, agents should be able to join and leave dynamically. In one of the studies, Sen and Airiau [98] have experimented with adding agents dynamically to a society. The work done by Savarimuthu et al. [90] considered agents living for certain amount of time then being replaced with new agents. We believe there is a lot of scope for experimentation in this direction.

Another characteristic of an open agent society is that the new agents that join the society can be of different personality types. Hence the composition of agents in a society will play an important role in norm dynamics. Simulation models need to cater for this kind of dynamism in the society. The work by López et al. [70] is one such approach in this direction (varying the personality of the agents). We believe experimenting with dynamic composition of agents in open agent societies is a good avenue for further research.

6.6. *Consideration of network topology*

Agents form links with other agents that they interact with. In the context of norms these links are important because normative influence and spreading occurs through these links. While the emphasis is on static network topologies (such as the work by Kittock [68] and Pujol [82]) not many researchers have considered the role of dynamic network topologies and the effect of norm spreading and emergence (see column 7 of Tables 2 and 3). Though the work in these lines has begun [56,90], we believe this is fertile ground for further research since agents in open societies constantly acquire and lose links forming dynamic networks which will influence how norms are spread in an agent society.

[11] This has been studied by deontic logic researchers in the context of an open institution where norms are added and removed [108].

6.7. *Experimenting with co-exiting and conflicting norms in agent societies*

Another area that lacks a lot of research is the consideration of societies where different norms co-exist in a given context (see column 6 of Tables 2 and 3). Most work has concentrated on how one particular norm emerges in a society. Burke et al. [23] experimented with how multiple norms can co-exist. Savarimuthu et al. [90] explored how different societies with different norms for the same context might arrive at a common norm due to environmental triggers. Chalub et al. [31] experimented on how a norm might emerge when agents living in different societies (islands) are at war. We believe more research in this area can be undertaken. Especially considering the richness of virtual agent societies with different norms for the same context, further extensions can be made, such as investigating when a norm will take over another when two societies interact and when these norms can co-exist.

Using a formal approach, some researchers have worked on resolving normative conflicts in open agent societies [69,108]. Simulation models can be used to simulate, identify and resolve normative conflicts that have richer and complex constructs.

6.8. *Implicit norm inference*

An agent joining a new society may infer norms based on observing other agents and through its personal experiences in the society. The learning mechanisms used so far are based on machine learning algorithms such as Highest Cumulative Reward (HCR) [68,99,114] or reinforcement learning algorithms [98]. An autonomous social agent needs to have more capabilities than just following a given algorithm. One may argue that these software agents can handle the level of computation needed by these algorithms to infer the norms. Still, these algorithms alone do not help in developing a full model of norm emergence. An agent needs to possess cognitive abilities to reason about its goals and desires before and after finding out what the norms of the society are. A reasonable social learning mechanism for an agent will be one that infers norms based on observation of interactions and personal experience, that are recorded in the belief base of the agent and which chooses whether to follow the norm based on its intentions and goals. Researchers have developed architectures based on the Belief-Desire-Intention (BDI) model [84] such as the BOID architecture [22], which takes into account the obligation of an agent. There have been other architectures developed for normative systems [16,28,72,86]. What is missing so far is a full account of what goes on in an agent's mind (the internal architecture of the agent[12]) and how it affects norm spreading in the entire society. Although such work has begun [7,73,93,94], more work in this direction will be beneficial to create agents that are 'norm capable'.

6.9. *Pluggable component framework for the norm life-cycle*

Various mechanisms have been proposed for norm establishment in a society. Each of these mechanisms fits into one or more phases of the life-cycle model. A good normative architecture should allow one or more of these mechanisms to be plugged in and experimented with. For example, researchers should be able to experiment with a scenario by setting up an entrepreneurial mechanism for norm creation, an imitation model for norm spreading, a reputation mechanism for norm enforcement, and

[12] The internal agent architecture should facilitate the identification of different types of norms such as permission norms, prohibition norms and obligation norms. The architecture should be able to identify norm conflicts and provide mechanisms to resolve these based on the agent's utility. Utility for each agent can be different. For example, an agent can be selfish while another can be benevolent.

then observe how norms are established in the society. After achieving this in one particular experiment, the researchers should be able to replace the reputation mechanism for enforcement with a punishment mechanism in the next experiment and study the effects (such as time to converge and efficiency). Such a pluggable architecture will be beneficial to the research community, both for sociologists to understand human societies and computer scientists to deal with artificial agent societies.

6.10. Other relevant characteristics

It should be noted that researchers have different criteria for norm emergence (see column 9 of Tables 2 and 3). Finnemore and Sikkink [48] noted that when a critical mass of agents has accepted a new idea (a norm) as appropriate, then a norm is said to have emerged. The emergence criterion needs to be parameterised in a norm simulation architecture. Another feature that could be incorporated in a simulation architecture is that of considering a noisy environment. Only Hoffmann [62] considered the possibility of a noisy environment. This could be easily parameterised in the simulation architecture. For example, in the park-littering scenario, an agent might perceive another agent littering correctly 80% of the time.

6.11. Bringing it all together: A norm simulation architecture

What would bring the most benefit for the study of norm spreading and emergence is the design and development of an architecture that would include attributes (e.g. through parameterization) that have been tabulated in Tables 2 and 3, and which also has a pluggable component for mechanisms for norm establishment as described in Section 6.9.

Some important work in this direction is being currently carried out as a part of the EMIL project [6,7, 24]. As it stands, the researchers in the EMIL project have proposed an architecture for norm emergence. However, their design does not cater for the plug and play of different mechanisms for norm establishment.

7. Conclusions

This survey paper has made three contributions to the normative multi-agent system field in the context of norm simulation. Firstly, five phases of the norm life-cycle based on a socio-computational viewpoint were proposed. They were norm creation, identification, spreading, enforcement and emergence. Secondly, various norm-based simulation works were categorized based on the mechanisms employed by each of the works. Nine main categories were identified and some key papers in each of these categories were discussed. The strengths and weaknesses of each category were also discussed. Thirdly, based on the empirical models of norms, important characteristics of simulation models for norms were identified and discussed. These identified characteristics were then used to compare different research works. In this context, pointers to future research directions have also been presented.

Acknowledgments

We would like to thank the three anonymous reviewers for their valuable suggestions. We would also like to thank Peter Whigham, Christian Müller-Schloer and Michael Winikoff for their comments.

References

[1] George A. Akerlof, The economics of caste and of the rat race and other woeful tales, *The Quarterly Journal of Economics* **90**(4) (November 1976), 599–617.

[2] George A. Akerlof, A theory of social custom, of which unemployment may be one consequence, *The Quarterly Journal of Economics* **94**(4) (June 1980), 749–775.

[3] Reka Albert and Albert-Laszlo Barabasi, Statistical mechanics of complex networks, *Reviews of Modern Physics* **74** (2002), 47–97.

[4] Huib Aldewereld, Frank Dignum, Andrés García-Camino, Pablo Noriega, Juan A. Rodríguez-Aguilar and Carles Sierra, Operationalisation of norms for usage in electronic institutions. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems* (*AAMAS*), pages 223–225, New York,NY, USA, 2006. ACM Press.

[5] Giulia Andrighetto, Marco Campenní, Federico Cecconi and Rosaria Conte, How agents find out norms: A simulation based model of norm innovation, In *Third International Workshop on Normative Multi-agent Systems* (*NorMAS*), pages 16–30, 2008.

[6] Giulia Andrighetto, Marco Campenni, Federico Cecconi and Rosaria Conte, The complex loop of norm emergence: A simulation model. In Shu-Heng Chen, Claudio Cioffi-Revilla, Nigel Gilbert, Hajime Kita, Takao Terano, Keiki Takadama, Claudio Cioffi-Revilla, and Guillaume Deffuant, editors, *Simulating Interacting Agents and Social Phenomena*, volume 7 of *Agent-Based Social Systems*, pages 19–35. Springer, 2010.

[7] Giulia Andrighetto, Rosaria Conte, Paolo Turrini and Mario Paolucci, Emergence in the loop: Simulating the two way dynamics of norm innovation. In Guido Boella, Leon van der Torre, and Harko Verhagen, editors, *Normative Multi-agent Systems*, number 07122 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, 2007.

[8] M. Anghel, Zoltán Toroczkai, Kevin E. Bassler and G. Korniss, Competition-driven network dynamics: Emergence of a scale-free leadership structure and collective efficiency, *Physical Review Letters* **92**(5) (Feb 2004), 058701.

[9] Josep L. Arcos, Marc Esteva, Pablo Noriega, Juan A. Rodríguez-Aguilar and Carles Sierra, Environment engineering for multiagent systems, *Engineering Applications of Artificial Intelligence* **18**(2) (2005), 191–204.

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

[11] Robert Axelrod, *The complexity of cooperation: Agent-based models of competition and collaboration*, Princeton University Press, 1997.

[12] Albert-Laszlo Barabasi and Reka Albert, Emergence of scaling in random networks, *Science* **286**(5439) (October 1999), 509–512.

[13] Gary S. Becker, *The Economic Approach to Human Behavior*, University of Chicago Press, September 1978.

[14] Cristina Bicchieri, *The Grammar of Society: The Nature and Dynamics of Social Norms*, Cambridge University Press, New York, USA, 2006.

[15] Guido Boella, Gabriella Pigozzi and Leendert van der Torre, Normative systems in computer science – ten guidelines for normative multiagent systems, in: *Normative Multi-Agent Systems*, Guido Boella, Pablo Noriega, Gabriella Pigozzi and Harko Verhagen, eds, Number 09121 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2009. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

[16] Guido Boella and Leendert van der Torre, An architecture of a normative system: counts-as conditionals, obligations and permissions, In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems* (*AAMAS*), pages 229–231, New York, NY, USA, 2006. ACM Press.

[17] 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) (2008), 1–10.

[18] Magnus Boman, Norms as constraints on real-time autonomous agent action, *Multi-agent rationality*, pages 36–44, 1997.

[19] Magnus Boman, Norms in artificial decision making, *Artificial Intelligence and Law* **7**(1) (1999), 17–35.

[20] Robert Boyd and Peter J. Richerson, *Culture and the evolutionary process*, University of Chicago Press, Chicago, 1985.

[21] Robert Boyd and Peter J. Richerson, Group beneficial norms can spread rapidly in a structured population, *Journal of Theoretical Biology* **215**(3) (2002), 287–296.

[22] Jan Broersen, Mehdi Dastani, Joris Hulstijn, Zisheng Huang and Leendert van der Torre, The boid architecture: conflicts between beliefs, obligations, intentions and desires, In *AGENTS '01: Proceedings of the fifth international conference on Autonomous agents*, pages 9–16, New York, NY, USA, 2001. ACM.

[23] Mary A. Burke, Gary M. Fournier and Kislaya Prasad, The emergence of local norms in networks: Research articles, *Complexity* **11**(5) (2006), 65–83.

[24] Marco Campenní, Giulia Andrighetto, Federico Cecconi, and Rosaria Conte. Normal = Normative? The role of intelligent agents in norm innovation, *Mind & Society* **8**(2) (2009), 153–172.

[25] Cristiano Castelfranchi, The Micro-Macro Constitution of Power, *ProtoSociology ?An International Journal of Interdisciplinary Research* **18–19** (2002), 208–268.
[26] Cristiano Castelfranchi and Rosaria Conte, *Cognitive and social action*, UCL Press, London, 1995.
[27] Cristiano Castelfranchi, Rosaria Conte and Mario Paolucci, Normative reputation and the costs of compliance, *Journal of Artificial Societies and Social Simulation* **1**(3) (1998).
[28] Cristiano Castelfranchi, Frank Dignum, Catholijn M. Jonker and Jan Treur, Deliberative normative agents: Principles and architecture. In *Sixth International Workshop on Intelligent Agents VI, Agent Theories, Architectures, and Languages* (*ATAL'99*), pages 364–378, London, UK, 2000. Springer-Verlag.
[29] Cristiano Castlefranchi, Maria Miceli and Amedeo Cesta, Dependence relations among autonomous agents (abstract), *ACM SIGOIS Bulletin* **13**(3) (1992), 14.
[30] Damien Challet and Yi-Cheng Zhang, On the minority game: Analytical and numerical studies, *Physica A: Statistical and Theoretical Physics* **256**(3–4) (1998), 514–532.
[31] F.A.C.C. Chalub, F.C. Santos and J.M. Pacheco, The evolution of norms, *Journal of Theoretical Biology* **241**(2) (2006), 233–240.
[32] Reuven Cohen, Shlomo Havlin and Daniel ben-Avraham, Efficient immunization strategies for computer networks and populations, *Physical Review Letters* **91** (2003), 247901.
[33] James Coleman, *Foundations of Social Theory*, Belknap Press, 1990.
[34] Rosaria Conte and Cristiano Castelfranchi, Understanding the effects of norms in social groups through simulation, In Nigel Gilbert and Rosaria Conte, editors, *Artificial societies: the computer simulation of social life*, pages 252–267. UCL Press, London, 1995.
[35] Rosaria Conte and Cristiano Castelfranchi, From conventions to prescriptions. Towards an integrated view of norms, *Artificial Intelligence and Law* **7**(4) (1999), 323–340.
[36] Rosaria Conte and Chrysanthos Dellarocas, *Social order in multiagent systems*, Springer, Netherlands, 2001.
[37] Rosaria Conte, Rino Falcone and Giovanni Sartor, Agents and norms: How to fill the gap? *Artificial Intelligence and Law* **7**(1) (1999), 1–15.
[38] Natalia Criado, Estefania Argente and Vicent Botti, Rational strategies for autonomous norm adoption. In *Proceeding of the Ninth international workshop on Coordination, Organization, Institutions and Norms in agent systems* (*COIN@AAMAS 2010*), pages 9–16, 2010.
[39] Adrian Perreau de Pinninck, Carles Sierra and W. Marco Schorlemmer, Distributed norm enforcement: Ostracism in open multi-agent systems, In Pompeu Casanovas, Giovanni Sartor, Nuria Casellas, and Rossella Rubino, editors, *Computable Models of the Law: Languages, Dialogues, Games, Ontologies*, volume 4884 of *Lecture Notes in Computer Science*, pages 275–290. Springer, 2008.
[40] Frank Dignum, Autonomous agents with norms, *Artificial Intelligence and Law* **7**(1) (1999), 69–79.
[41] Emile Durkheim and George Simpson, *Emile Durkheim on The division of labor in society / being a translation of his De la division du travail social, with an estimate of his work by George Simpson*. Macmillan, 1933.
[42] Robert C. Ellickson, *Order without law: How neighbors settle disputes*, Harvard Univ Press, 1991.
[43] Robert C. Ellickson, Law and economics discovers social norms, *The Journal of Legal Studies* **27**(2) (1998), 537–552.
[44] Jon Elster, *The cement of society: A study of social order*, Cambridge University Press, December 1989.
[45] Jon Elster, Social norms and economic theory, *The Journal of Economic Perspectives* **3**(4) (1989), 99–117.
[46] Joshua M. Epstein, Learning to be thoughtless: Social norms and individual computation, *Computational Economics* **18**(1) (2001), 9–24.
[47] Ernst Fehr and Urs Fischbacher, Third-party punishment and social norms, *Evolution and human behavior* **25**(2) (2004), 63–87.
[48] Martha Finnemore and Kathryn Sikkink, International Norm Dynamics and Political Change, *International Organization* **52**(4) (1998), 887–917.
[49] Julia Fix, Christian von Scheve and Daniel Moldt, Emotion-based norm enforcement and maintenance in multi-agent systems: foundations and petri net modeling, In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems* (*AAMAS*), pages 105–107, New York, NY, USA, 2006. ACM Press.
[50] Felix Flentge, Daniel Polani and Thomas Uthmann, Modelling the emergence of possession norms using memes, *Journal of Artificial Societies and Social Simulation* **4**(4) (2001).
[51] Santo Fortunato, Vito Latora, Alessandro Pluchino and Andrea Rapisarda, Vector Opinion Dynamics in a Bounded Confidence Consensus Model, *International Journal of Modern Physics C* **16**(10) (2005), 1535–1551.
[52] Andrés García-Camino, Juan A. Rodríguez-Aguilar, Carles Sierra and Wamberto Weber Vasconcelos, Norm-oriented programming of electronic institutions, In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems* (*AAMAS*), pages 670–672, New York, NY, USA, 2006. ACM Press.
[53] Jack P. Gibbs, Norms: The problem of definition and classification, *American Journal of Sociology* **60** (1965), 586–594.
[54] Herbert Gintis, The hitchhiker's guide to altruism: Gene-culture coevolution, and the internalization of norms. Working Paper 01-10-058, Santa Fe Institute, October 2001.

[55] Marta C. Gonzaléz, Pedro G. Lind and Hans J. Herrmann, Networks based on collisions among mobile agents, *Physica D* **224** (2006), 137–148. e-print: physics/0606023.
[56] Nathan Griffiths and Michael Luck, Norm emergence in tag-based cooperation, In *Proceeding of the AAMAS 2010 workshop on Coordination, Organization, Institutions and Norms in agent systems* (*COIN*), pages 80–87, 2010.
[57] Jürgen Habermas, *The Theory of Communicative Action, Volume 1: Reason and the Rationalization of Society*. Beacon Press, 1985.
[58] Jürgen Habermas, *Between facts and norms*, MIT Press Cambridge, MA, USA, 1996. translated by William Rehg.
[59] David Hales, Cooperation without memory or space: tags, groups and the prisoner's dilemma, In Scott Moss and Paul Davidsson, editors, *Multi-agent based simulation: second international workshop* (*MABS 2000*)*, revised and additional papers*, volume 1979 of *Lecture Notes in Computer Science*, pages 157–166. Springer, 2001.
[60] David Hales, Group reputation supports beneficent norms, *Journal of Artificial Societies and Social Simulation* **5** (2002).
[61] Michael Hechter and Karl Dieter Opp, editors, *Social Norms*, Russel Sage Foundation, New York, 2001.
[62] Mathew J. Hoffmann, Entrepreneurs and Norm Dynamics: An Agent-Based Model of the Norm Life Cycle. Technical report, Department of Political Science and International Relations, University of Delaware, USA, 2003.
[63] Christine Horne, Sociological perspectives on the emergence of norms, in: *Social Norms*, M. Hechter and KD Opp, editors, Russel Sage Foundation, New York, 2001, pp. 3–34.
[64] Kurt Jensen, editor, *Application and Theory of Petri Nets 1992*, volume 616 of *Lecture Notes in Computer Science*. Springer, 1992.
[65] Andrew J.I. Jones and Marek Sergot, On the characterisation of law and computer systems: The normative systems perspective, In *Deontic Logic in Computer Science: Normative System Specification*, pages 275–307. John Wiley and Sons, 1993.
[66] Andrew J.I. Jones and Marek J. Sergot, A formal characterisation of institutionalised power, *Logic Journal of the IGPL* **4**(3) (1996), 427–443.
[67] Jeremy Kendal, Marcus W. Feldman and Kenichi Aoki, Cultural coevolution of norm adoption and enforcement when punishers are rewarded or non-punishers are punished, *Theoretical Population Biology* **70**(1) (2006), 10–25.
[68] James E. Kittock, Emergent conventions and the structure of multi-agent systems, in: *1993 Lectures in Complex Systems, Santa Fe Institute Studies in the Sciences of Complexity*, Lynn Nadel and Daniel L. Stein, editors, Addison-Wesley, 1995.
[69] Martin J. Kollingbaum, Wamberto Weber Vasconcelos, Andrés García-Camino and Timothy J. Norman, Managing conflict resolution in norm-regulated environments. In *Engineering Societies in the Agents World VIII*, volume 4995 of *Lecture Notes in Computer Science*, pages 55–71, 2008.
[70] Fabiola López y López, *Social Powers and Norms: Impact on Agent Behaviour*, PhD thesis, Department of Electronics and Computer Science, University of Southampton, United Kingdom, 2003.
[71] Fabiola López y López, Michael Luck and Mark d'Inverno, Constraining autonomy through norms, In *Proceedings of The First International Joint Conference on Autonomous Agents and Multiagent Systems* (*AAMAS*), volume 2, pages 674–681, New York, NY, USA, 2002. ACM.
[72] Fabiola López y López and Amauri Arenas Márquez, An architecture for autonomous normative agents. In *Proceedings of the Fifth Mexican International Conference in Computer Science* (*ENC*), pages 96–103, Los Alamitos, CA, USA, 2004. IEEE Computer Society.
[73] Felipe Rech Meneguzzi and Michael Luck, Norm-based behaviour modification in BDI agents. In *Proceedings of the Eighth International Conference on Autonomous Agents and Multiagent Systems* (*AAMAS*), pages 177–184. IFAAMAS, 2009.
[74] Melanie Mitchell, Complex systems: Network thinking, *Artificial Intelligence* **170**(18) (2006), 1194–1212.
[75] Mayuko Nakamaru and Simon A. Levin, Spread of two linked social norms on complex interaction networks, *Journal of Theoretical Biology* **230**(1) (2004), 57–64.
[76] Martin Neumann, Homo socionicus: a case study of simulation models of norms, *Journal of Artificial Societies and Social Simulation* **11**(4) (2008).
[77] Martin Neumann, A classification of normative architectures, In Keiki Takadama, Claudio Cioffi-Revilla, and Guillaume Deffuant, editors, *Simulating Interacting Agents and Social Phenomena: The Second World Congress*, volume 7 of *Agent-Based Social Systems*, pages 3–18. Springer, Berlin/Heidelberg, 2010.
[78] Douglass C. North, *Institutions, institutional change, and economic performance*, Cambridge Univ Press, 1990.
[79] Hyacinth S. Nwana, Software agents: an overview, *The Knowledge Engineering Review* **11**(03) (1996), 205–244.
[80] Hisashi Ohtsuki, Iwasa Iwasa and Martin A. Nowak, Indirect reciprocity provides only a narrow margin of efficiency for costly punishment, *Nature* **457**(7225) (2009), 79–82.
[81] Karl-Dieter Opp, How do norms emerge? An outline of a theory, *Mind & Society* **2**(1) (2001), 101–128.
[82] Josep M. Pujol, *Structure in Artificial Societies*, PhD thesis, Departament de Llenguatges i Sistemes Informàtics, Universitat Politènica de Catalunya, 2006.

[83] Josep M. Pujol, Ramon Sangüesa and Jordi Delgado, Extracting reputation in multi agent systems by means of social network topology, In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems* (*AAMAS*), pages 467–474, New York, NY, USA, 2002. ACM Press.

[84] Anand S. Rao and Michael P. Georgeff, Modeling rational agents within a BDI-architecture, in: *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, James Allen, Richard Fikes and Erik Sandewall, eds, Morgan Kaufmann publishers Inc.: San Mateo, CA, USA, 1991, pp. 473–484.

[85] Michael Rymaszewski, Wagner James Au, Mark Wallace, Catherine Winters, Cory Ondrejka, Benjamin Batstone-Cunningham and Philip Rosedale, *Second Life: The Official Guide*, SYBEX Inc., Alameda, CA, USA, 2006.

[86] Fariba Sadri, Kostas Stathis and Francesca Toni, Normative kgp agents, *Computational & Mathematical Organization Theory* **12**(2–3) (2006), 101–126.

[87] Norman Salazar, Juan A. Rodríguez-Aguilar and Josep L. Arcos, Infection-based norm emergence in multi-agent complex networks. In *Proceeding of the 2008 conference on ECAI 2008*, pages 891–892, Amsterdam, The Netherlands, The Netherlands, 2008. IOS Press.

[88] Norman Salazar, Juan A. Rodríguez-Aguilar and Josep L. Arcos, Handling uncertainty in the emergence of social conventions. In *Third IEEE International Conference on Self-Adaptive and Self-Organizing Systems* (*SASO'09*), pages 282–283, 2009.

[89] Giovanni Sartor, Why agents comply with norms and why they should, in: *Social Order in Multiagent Systems*, Rosaria Conte and Chrysanthos Dellarocas, eds, Kluwer Academic Publishers, Boston, 2001, pp. 19–44.

[90] Bastin Tony Roy Savarimuthu, Stephen Cranefield, Martin K Purvis and Maryam A Purvis, Norm emergence in agent societies formed by dynamically changing networks, *Web Intelligence and Agent Systems* **7**(3) (2009), 223–232.

[91] Bastin Tony Roy Savarimuthu, Stephen Cranefield, Maryam Purvis and Martin Purvis, Role model based mechanism for norm emergence in artificial agent societies, In *Coordination, Organizations, Institutions, and Norms in Agent Systems III*, volume 4870 of *Lecture Notes in Computer Science*, pages 203–217. Springer, Berlin/Heidelberg, 2008.

[92] Bastin Tony Roy Savarimuthu, Stephen Cranefield, Maryam Purvis and Martin K. Purvis, Internal agent architecture for norm identification, In *Coordination, Organizations, Institutions and Norms in Agent Systems V, COIN 2009 International Workshops*, pages 241–256, 2010.

[93] Bastin Tony Roy Savarimuthu, Stephen Cranefield, Maryam A. Purvis and Martin K. Purvis, Norm identification in multi-agent societies, Discussion Paper 2010/03, Department of Information Science, University of Otago, 2010.

[94] Bastin Tony Roy Savarimuthu, Stephen Cranefield, Maryam A. Purvis and Martin K. Purvis, Obligation norm identification in agent societies, *Journal of Artificial Societies and Social Simulation* **13**(4) (2010), 3.

[95] Christian Scheve, Daniel Moldt, Julia Fix and Rolf Luede, My agents love to conform: Norms and emotion in the micro-macro link, *Computational and Mathematical Organization Theory* **12**(2–3) (2006), 81–100.

[96] John R. Searle, *Speech Acts*, Cambridge University Press, Cambridge, UK., 1969.

[97] Onkur Sen and Sandip Sen, Effects of social network topology and options on norm emergence, in: *Coordination, Organizations, Institutions and Norms in Agent Systems V*, Julian Padget, Alexander Artikis, Wamberto Vasconcelos, Kostas Stathis, Viviane da Silva, Eric Matson and Axel Polleres, eds, volume 6069 of *Lecture Notes in Computer Science*, pages 211–222. Springer Berlin / Heidelberg, 2010.

[98] Sandip Sen and Stephane Airiau, Emergence of norms through social learning, In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence* (*IJCAI*), pages 1507–1512. AAAI Press, 2007.

[99] Yoav Shoham and Moshe Tennenholtz, Emergent conventions in multi-agent systems: Initial experimental results and observations, In *Proceedings of the Third International Conference on the Principles of Knowledge Representation and Reasoning* (*KR*), pages 225–231, San Mateo, CA, USA, 1992. Morgan Kaufmann.

[100] Yoav Shoham and Moshe Tennenholtz, On social laws for artificial agent societies: Off-line design, *Artificial Intelligence* **73**(1–2) (1995), 231–252.

[101] Yoav Shoham and Moshe Tennenholtz, On the emergence of social conventions: Modeling, analysis, and simulations, *Artificial Intelligence* **94**(1-2) (1997), 139–166.

[102] Tilman Slembeck, Reputations and fairness in bargaining – experimental evidence from a repeated ultimatum game with fixed opponents, Experimental working paper, Economics Working Paper Archive (EconWPA), http://ideas.repec.org/p/wpa/wuwpex/9905002.html, 1999.

[103] Maynard J. Smith and G.R. Price, The logic of animal conflict, *Nature* **246**(5427) (November 1973), 15–18.

[104] Alexander Staller and Paolo Petta, Introducing emotions into the computational study of social norms: A first evaluation, *Journal of Artificial Societies and Social Simulation* **4**(1) (2001).

[105] Stanislaw A.B. Stane and Mariusz Zytniewsk, Normative multi-agent enriched data mining to support e-citizens. In Longbing Cao, editor, *Data Mining and Multi-agent Integration*, pages 291–304. Springer, 2009.

[106] Raimo Tuomela, *The Importance of Us: A Philosophical Study of Basic Social Notions*, Stanford Series in Philosophy, Stanford University Press, 1995.

[107] Edna Ullmann-Margalit, *The Emergence of Norms*, Clarendon Press, 1977.

[108] Wamberto Weber Vasconcelos, Martin J. Kollingbaum and Timothy J. Norman, Normative conflict resolution in multi-agent systems, *Autonomous Agents and Multi-Agent Systems* **19**(2) (2009), 124–152.
[109] Harko Verhagen, *Norm Autonomous Agents*, PhD thesis, Department of Computer Science, Stockholm University, 2000.
[110] Harko Verhagen, Simulation of the Learning of Norms, *Social Science Computer Review* **19**(3) (2001), 296–306.
[111] Daniel Villatoro and Jordi Sabater-Mir, Categorizing social norms in a simulated resource gathering society, In *Coordination, Organizations, Institutions and Norms in Agent Systems IV: COIN 2008 International Workshops, Revised Selected Papers*, pages 235–249. Springer-Verlag, Berlin/Heidelberg, 2009.
[112] Daniel Villatoro, Sandip Sen and Jordi Sabater-Mir, Topology and memory effect on convention emergence, In *WI-IAT '09: Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, pages 233–240, Washington, DC, USA, 2009. IEEE Computer Society.
[113] Georg Henrik von Wright, *Norm and action: a logical enquiry*, Routledge & Kegan Paul PLC, 1963.
[114] Adam Walker and Michael Wooldridge, Understanding the emergence of conventions in multi-agent systems. In Victor Lesser, editor, *Proceedings of the First International Conference on Multi-Agent Systems* (*ICMAS*), pages 384–389, Menlo Park, California, USA, 1995. AAAI Press.
[115] Roel J. Wieringa and John-Jules Ch. Meyer, Applications of deontic logic in computer science: a concise overview. In *Deontic logic in computer science: Normative system specification*, pages 17–40. John Wiley & Sons, Inc., New York, USA, 1994.
[116] Michael Wooldridge and Nicholas R. Jennings, Intelligent agents: Theory and practice, *Knowledge Engineering Review* **10** (1995), 115–152.
[117] Bin Yu and Munindar P. Singh, Searching social networks, In *Proceedings of the second international joint conference on Autonomous Agents and MultiAgent Systems*, (*AAMAS'03*), pages 65–72, New York, USA, 2003. ACM Press.

Author's Bios

Bastin Tony Roy Savarimuthu received his Master of Engineering (ME) degree in Software Systems from Birla Institute of Technology and Science, Pilani, India. He is currently a Lecturer in Information Science at the University of Otago, Dunedin, New Zealand. His primary area of research is normative multi-agent systems. His PhD work focuses on how norms emerge in artificial agent societies and how agents can identify norms in open agent societies.

Stephen Cranefield studied mathematics and computer science at the University of Otago and then gained a PhD in artificial intelligence from the University of Edinburgh. He is currently an associate professor in information science at the University of Otago where he researches in the area of multi-agent systems, with a focus on the use of social concepts such as norms. He is on the editorial board of the journals Multiagent and Grid Systems and Knowledge Engineering Review and has been on the programme committee for many international conferences and workshops, and on the senior programme committee for the International Conference on Automonous Agents and Multiagent Systems and the Australasian Conference on Artificial Intelligence.