



Christopher D. Hollander and Annie S. Wu (2011)

The Current State of Normative Agent-Based Systems

Journal of Artificial Societies and Social Simulation 14 (2) 6
<<http://jasss.soc.surrey.ac.uk/14/2/6.html>>

Received: 07-Sep-2010 Accepted: 20-Feb-2011 Published: 31-Mar-2011



Abstract

Recent years have seen an increase in the application of ideas from the social sciences to computational systems. Nowhere has this been more pronounced than in the domain of multiagent systems. Because multiagent systems are composed of multiple individual agents interacting with each other many parallels can be drawn to human and animal societies. One of the main challenges currently faced in multiagent systems research is that of social control. In particular, how can open multiagent systems be configured and organized given their constantly changing structure? One leading solution is to employ the use of social norms. In human societies, social norms are essential to regulation, coordination, and cooperation. The current trend of thinking is that these same principles can be applied to agent societies, of which multiagent systems are one type. In this article, we provide an introduction to and present a holistic viewpoint of the state of normative computing (computational solutions that employ ideas based on social norms.) To accomplish this, we (1) introduce social norms and their application to agent-based systems; (2) identify and describe a normative process abstracted from the existing research; and (3) discuss future directions for research in normative multiagent computing. The intent of this paper is to introduce new researchers to the ideas that underlie normative computing and survey the existing state of the art, as well as provide direction for future research.

Keywords: Norms, Normative Agents, Agents, Agent-Based System, Agent-Based Simulation, Agent-Based Modeling



Introduction

- 1.1 The past decades have seen an increase in the use and acceptance of agent-based systems across a broad number of fields that range from economics to biology (Galan and Izquierdo 2005). This growth is partly due to an interest in modeling social problems and simulating complex social conditions that do not easily lend themselves to traditional mathematical models. As a result, ideas from the various branches of the social sciences have been explored to solve hard problems in computer science. In this paper, we are interested in a problem that is shared between sociology and multi-agent systems known as social control (Mukherjee et al. 2007). Research on social control addresses the challenge of ensuring that a system operates efficiently while at the same time allowing the individual agents maintain their freedom (Verhagen 2000). On the frontier of this research, *social norms* are being investigated for their potential use in implementing social control in multi-agent systems (Therborn 2002). In this approach, which we call the *normative approach*, social norms act as behavioral constraints that regulate and structure social order within a multi-agent system and promote cooperation and coordination between heterogeneous agents in open systems (Boella and Torre 2007).
- 1.2 The normative approach can be realized through *normative multi-agent systems* that combine social norms and multi-agent systems. Normative multi-agent systems offer the ability to integrate social and individual factors to provide increased levels of fidelity with respect to modeling social phenomenon such as cooperation, coordination, group decision making, organization, and so on in human and artificial agent systems (Boella et al. 2007). Additionally, normative multi-agent systems offer a tool to examine sociology through the perspective of methodological individualism (Neumann 2008). Methodological individualism is an attempt to build the foundations of sociology using

individual actors and study the emergent phenomenon. To accomplish this, methodological individualism investigates the feedback mechanisms present in society as well as the system dynamics.

- 1.3 Normative multi-agent systems are composed of normative agents. Normative agents must satisfy the regular notions associated with artificial agents and possess the capability to: represent norms in a format that allows them to be reasoned over and modified during the lifetime of the agent (knowledge representation); recognize and infer the norms of other agents based on observations and interactions while not confusing the norms with individual rules and constraints (learning theory); transmit norms, in both an active and passive fashion, to other agents (communication and network theory); and sanction other agents who do not comply with known norms if those norm require it (morality and law). To date, implementations of normative agent architectures are largely based on the belief, desire, and intention (BDI) architectures of computer science and remain uninfluenced by the conclusions of psychology, pedagogy, and neurophysiology (Neumann 2010).
- 1.4 In this paper, we extend the current body of research on normative multi-agent systems by introducing a process driven norm life cycle model that is based on the most recent empirical results and theoretical and philosophical ideas. In addition to describing and discussing our model, this paper is intended to serve as survey of the current field of normative systems and as an introduction for new researchers in the field of normative systems and the computational study of norms. These survey and introduction goals are accomplished by highlighting the major areas of current research and discussing how they fit together into a holistic picture.
- 1.5 The structure of this paper can be briefly summarized as follows. First, we introduce the concept of a norm and identify the essential characteristics. We then illustrate the richness of the philosophical normative landscape by discussing the past approaches used to categorize norms and create norm typologies. Next we introduce the methods used to represent norms in computational systems before finally defining normative multi-agent systems. Following our presentation of normative multi-agent systems, we introduce our normative process model and structure its description around previous research that supports each component. After describing each process, we look ahead to the challenges that still exist and discuss where future research might be best focused.
- 1.6 We have tried to restrict our focus to work done in the field of normative systems and not the broader social sciences. Through this approach we hope to better underscore the necessity of interdisciplinary cooperation and illustrate, through leaving explicit gaps and using general explanations, where there is a need for further refinement in our theories. Some of these refinements may already exist in fields unknown and unexplored by researchers of normative systems. Others still may not yet be conceptualized.



What is a Norm?

- 2.1 The literature is populated with numerous of definitions and uses of the term norm (Horne 2001). The lack of a consistent definition makes it difficult to describe any sort of general normative process or discuss the life cycle of norms in general. However, by closely examining the literature on normative agents and normative multi-agent systems, we have identified a number of varied but conceptually consistent definitions that, taken together, form the general concept of a norm as it is used in much of the work surveyed for this paper. To facilitate an understanding of generalized norms, we first examine how the existing definitions of a norm are reflected in the literature, starting with its use in the sciences and then looking at its use in multi-agent systems. We then identify the essential characteristics of a norm and conclude with a short discussion on how two particular aspects of a norm can create scenarios that are not typically encountered in traditional multi-agent systems that use rational agents.

Normative Definitions

- 2.2 The Webster online dictionary (Merriam-Webster 2010) provides three definitions for the term norm:
 1. an authoritative standard
 2. a principle of right action binding upon the members of a group and serving to guide, control, or regulate proper and acceptable behavior
 3. average:
 - a. as a set standard of development or achievement usually derived from the average or median achievement of a large group
 - b. as a pattern or trait taken to be typical in the behavior of a social group
 - c. as a widespread or usual practice, procedure, or custom

These definitions are representative of the term as used in the various areas of normative research, including deontic logic, legal theory, sociology, social psychology and social philosophy, decision theory, and game theory (Verhagen 2000).

- 2.3** In deontic logic, a norm is viewed as an obligation or a permission that an individual has to a larger social system (Boella et al. 2007). Obligation can take a negated form, in which case it is referred to as prohibition. In legal theory, a norm is any behavioral rule dictated by a ruling body and enforced through the use of sanctions (Verhagen 2000). In sociology and the social sciences, norms are rules of behavior or behavioral constraints that are socially enforced and considered valid by the majority of the group (Bendor and Swistak 2001; Ehrlich and Levin 2005; Horne 2001; Therborn 2002; Young 2008)—though the quantitative meaning of majority is never strictly defined and varies in accordance to the context. Decision theory, game theory, and other theories based on rational actors treat norms in a similar way; as successful behaviors that have been adopted by a majority of the population (Bendor and Swistak 2001).
- 2.4** Regardless of the specific domain, there is a common theme in which norms are treated as behaviors which *ought* to be displayed by members of a group when in a given context (Boella et al. 2007; Horne 2001); this implies that the observed behavior does not always match the expected normative behavior. When the observed behavior of an agent does not match the expected behavior then the acting agent is said to be *deviant*. That is to say, an agent is *deviant* when it does not abide by the generally accepted rules for behavior (the norm) for the context in which it is acting. Deviant agents can be lone actors, or part of a larger deviant subculture. One can even think of deviance as an emergent property that results from the failure of two groups to come to terms on accepted behaviors. In this perspective, deviance is not necessarily bad, and can be just as important as norm obedience to a healthy society. For instance, deviance provides a source of alternative behavioral rules that can be called upon at a later time. The topic of deviance is an important concept in sociology and has far reaching implications to normative multi-agent systems (Meneguzzi and Luck 2009); however, a full treatment of deviance and its related theories lies outside the scope of this paper. Normative systems attempt to minimize deviance through the use of social enforcements such as *sanctions*. Sanctions are actions that levy punishment on deviant agents or reward conforming agents. In many cases, sanctions are also norms. Norms about norms have been termed meta-norms within the normative systems community (Axelrod 1986; Ehrlich and Levin 2005).
- 2.5** Specific to the literature on multi-agent systems, norms typically refer to: constraints on behavior (Shoham and Tennenholtz 1992), solutions to a macro-level problem (Zhang and Leezer 2009), obligations (Verhagen 2000); and regulatory or control devices for decentralized systems (Savarimuthu et al. 2008). One aspect of norms that is frequently left unaddressed in artificial systems is their dynamic nature and tendency to change over time (Neumann 2008). An examination of human society yields clear examples of this phenomenon. The norms of one generation are rarely identical to the next.

Key Normative Concepts

- 2.6** The various definitions and treatments of norms can be distilled down to a small set of straightforward concepts:
- A *norm* is any behavioral rule that is considered valid by the majority of a population—where the quantitative definition of majority is problem dependent. The precise size and composition of a population is also problem dependent. It is not unusual to divide a large population into multiple sub-populations such that the original norm is considered invalid by one or more of the new collectives. The sub-populations may in turn have their own norms that did not exist in the unified population.
 - Norms are acquired through a social learning process where an agent interacts with other agents as well as its environment. In sociology, this process is known as socialization; in anthropology it is called enculturation.
 - Norms are socially enforced through external sanctions or other measures until they become internalized by an agent. Once internalized, norms are enforced primarily through internal mechanisms.
 - Norms spread. The ability for norms to spread is a consequence of the system's underlying network topology in conjunction with active and passive transmission mechanisms.
- 2.7** With these concepts, we are able to discuss norms at the individual level. If an agent associates a specific set of behavioral rules with a particular context, we call that set of rules and their associated context a *potential norm*. When we consider a specific agent in isolation, potential norms are norms in the traditional macro-level sense. However, when we consider a specific agent in the context of a larger population, potential norms act as micro-level forces that produce macro-level social phenomenon. Potential norms spread through a population via social learning processes and enforcements until adopted by a sufficient proportion of the population, at which time we refer to them as *societal norms* (or simply *norms*) and once again treat them in the traditional macro-level sense of the term. By introducing the idea of potential norms, we are able to examine normative processes in dynamic populations from a recursive perspective, with potential norms serving as the base case.

Potential Sub-optimality and Normative Choice

- 2.8** *Potential sub-optimality* and *normative choice* are two distinct characteristics that make norms particularly interesting in the context of agent-based problem solving when the agents under consideration are rational agents.
- 2.9** Potential sub-optimality is associated with norms that solve macro-level problems. It refers to the idea that while a normative solution can be acceptable, it may not be optimal (or even Pareto optimal.) A norm can exhibit potential sub-optimality if the problem does not require an optimal solution (Henrich et al. 2008), if multiple solutions exist for the problem and all are acceptable (Bendor and Swistak 2001), or if the cost of obtaining an optimal solution for the problem is too high (Boella et al. 2007). One consequence of the potential sub-optimality property is that it may be possible to find a normative solution quicker than an optimal solution. This can be used to define an evolutionary process where an acceptable solution is quickly found and implemented and then improved upon as the system continues to operate, eventually converging to an optimal solution over time.
- 2.10** Normative choice refers to the ability of an agent to willfully violate a norm and assume the role of deviant. The agent is able to make a choice as to whether or not it will obey a norm. Unlike the situation in many multi-agent systems where the violation of rules is neither permitted by design nor expected, the violation of norms is a common occurrence that can guide a normative system towards a better solution (this idea is at the center of theories on social change) (Boella et al. 2007; Dignum 1999). In order to support normative choice, normative agents must be able to recognize and reason about norms. This enables the agents to make intelligent decisions about the acceptance or rejection of a norm and detect when other agents obey or violate norms so as to respond appropriately.
- 2.11** Any system attempting to make use of norms with a high degree of fidelity must account for both potential sub-optimality and normative choice.

Norms in the Social Science and Engineering

- 2.12** Throughout the years researchers have conceived of numerous typologies, categorizations, and specialized definitions of norms that can often make it difficult to determine which one to use when approaching a problem. Because the application of norms to solve real world problems depends on the context, broad categorizations can be immediately useful for finding related literature and understanding how others have approached similar problems. Towards this end, and to serve as a continuation of our discussion on the nature of norms, we briefly discuss norms from the perspective of the social sciences and engineering.

Norms in the Social Sciences

- 2.13** In the primary social sciences (psychology, sociology, and economics) and philosophy, the research and interest on norms has shifted throughout the years between the social function of norms (inspired by sociologists like Durkheim, Parsons, and Merton), the social impact of norms (inspired by economics), and the mechanisms leading to the emergence and creation of norms (inspired in part by complexity science).
- 2.14** In the context of social function, norms are often concerned with the *oughtness* and *expectation* of agent behavior; where oughtness refers to the notion that there are behaviors an agent should (or should not) perform regardless of the possible consequences and expectation refers to the behaviors other agents anticipate when observing an agent. Expectation is created when an agent displays behavioral regularity each time it encounters a specific context (Hechter and Opp 2001a). In the literature relating to the social function of norms, we find norms that address individual action in isolation, norms that govern interaction between members of a group, and norms that dictate responses to behaviors observed in others (Gibbs 1965; Horne 2001). Norms are also seen to provide social function in terms of obligations; be they legal, moral, or conditional. Aside from referring to behaviors that agents ought to do, or are expected to do, norms have functionality in regards to telling us what something is (definition) and telling us what is "normal" in a population (distribution) (Boella and Torre 2006; Duangsuwan and Liu 2009; Therborn 2002).
- 2.15** In the context of social impact, norms are considered terms of cost provided to or imposed on the parties involved in a social interaction. Under this context, cost is associated with the amount of resources that are gained or lost during an interaction. These resources can be internal, such as emotion levels, energy, etc, or external in the case of food, money, etc. The social impact perspective identifies norms that benefit the agent and society and incur a cost to both, norms that benefit individuals and cost society, norms that cost individuals and benefit society, and norms that benefit

both individuals and society without cost (Horne 2001; Savarimuthu et al. 2008).

- 2.16** Finally, and more recently, there has been a growing interest in norms from the context of norm emergence and creation (norm origin). This type of research is concerned with the "how" of norms more than the "why." In particular, the literature identifies two general methods by which norms can come into existence (Boella and Torre 2006; Boella et al. 2008b; Tuomela 1995; Verhagen 2000). The first explanation of norm origin is that they are explicitly created (as potential norms) and enforced by an authority structure to regulate and order interactions with a society. The second explanation is that norms emerge from regularities in behavior when agents interact with one another and mutual belief that the behavior is correct with regards to the goals of the agents. We will call norms that originate from an authority structure *Type I* norms, and norms that emerge from the interaction of agents *Type II* norms. Both types of norms have implications in the control, organization, and structure of agent societies.

Norms in Engineering (and Computational Science)

- 2.17** An engineering perspective typically places importance on the application of norms to engineering processes, usually as a form of regulation and control within a system. In this way, norms are viewed as another tool (in the same vein as contracts, protocols, etc) to accomplish some specific task.
- 2.18** In normative multi-agent systems, the focus is typically on rules of action that can be used to constrain an agent's behavior and thus reduce the size of its search space. These constraints can either be rigid, in which case the norm must be obeyed and is viewed as a global constraint; or flexible, in which case the obedience of the norm is dependent on the agent making the decision (Wu 2008).

Representation of Norms in Agent-Based Systems

- 2.19** In order for norms to be used in agent-based systems, norms must be specified in a way that enables them to be processed by artificial agents. Research in this area is still in its infancy due to a limited focus on the practical application of normative agents, but progress is being made nonetheless. We briefly describe four major representation schemes used in recent research: modal logic, condition/action pairs in rule-based systems, binary strings, and game theory.
- 2.20** Modal logic is an extension of classical formal logic that reasons about the necessary and the possible. Deontic logic is a derivative of modal logic that reasons about obligations, permissions, and prohibitions (Meyer and Wieringa 1993; Wright 1951). Deontic logic is tightly coupled with the normative systems from legal theory, and as a result it has become a popular representation scheme within the normative agent community. In the existing literature, deontic logic, other variations of modal logic, and first-order logic, have been used to develop normative agent architectures (Boella and Torre 2006; Castelfranchi et al. 1999), extend existing agent models (Alberti et al. 2005; Meneguzzi and Luck 2009; Sadri et al. 2005), and specify illegal behavior and its consequences (Meyer and Wieringa 1993).
- 2.21** Rule-based systems are collections of condition/action pairs together with an inference engine and a working memory. In normative systems, it is common to find that the condition/action pairs encode normative behaviors and their associated contexts (Boella et al. 2007). This representation format is commonly used by systems that take advantage of offline design, where the norms are coded directly into the agent's decision making system (Castelfranchi et al. 1998; Conte and Castelfranchi 1995; Hales 2002; Saam and Harrier 1999; Savarimuthu and Cranefield 2009; Schelling 1978; Shoham and Tennenholtz 1992; Staller and Petta 2001; Younger 2004).
- 2.22** Binary strings are sequences of ones and zeros, where each digit represents the presence (in the case of a one) or absence (in the case of a zero) of a norm. This scheme enables norms to be dealt with on an abstract level, and it is often used in research that examines the transmission and emergence of norms in a population (Caldas and Coelho 1999; Epstein 2001; Flentge et al. 2001a; Galan and Izquierdo 2005; Nakamarua and Levin 2004).
- 2.23** In normative multi-agent systems based on game theory, each agent is capable of making a simple choice that yields a corresponding payoff. At each round of the game, the agents attempt to maximize their payoff by choosing an action based on what they anticipate their opponent to choose. Norms are represented by the strategies that an agent uses to make these decisions (Mukherjee et al. 2007; Savarimuthu et al. 2008). A norm *emerges* when the number of agents in the population playing by the same strategy exceeds some tolerance value. Like the condition/action pairs of rule-based systems, this scheme is commonly used in systems where the norms are designed and encoded offline.

Normative Multi-agent Systems

- 2.24** On an abstract level, a normative system can be defined as any system where norms and normative concepts are required to accurately describe and specify the system's behavior (Meyer and Wieringa 1993). Of the many types of normative systems that exist, we are concerned with one particular type referred to as a *normative multi-agent system*.
- 2.25** A normative multi-agent system combines concepts of norms with an explicit representation scheme for normative information in order to provide a solution to problems relating to openness in multi-agent systems, where an *open multi-agent system* (Artikis and Pitt 2009) is a multi-agent system in which agents may not share the same architecture or the same goals, interactions between agents cannot be predicted in advance, and agents are able to join and leave the system freely. To accomplish this, a normative multi-agent system is built from normative agents that must be able to create, modify, detect, transmit, and reason about norms, as well as enforce existing norms through sanctions or some other mechanism (Boella et al. 2007; Boella et al. 2008a; Boella et al. 2008b).
- 2.26** Research specific to normative multi-agent systems, both in relation to open multi-agent systems and as entities in their own right, has grown in recent years. Partly responsible for, and emerging from this growth, are the COIN workshops (<http://www.pcs.usp.br/~coin/>) that investigate coordination, organization, institutions, and norms in agent systems; the Normative Multi-agent Systems workshops at Dagstuhl (<http://www.dagstuhl.de/09121>) which cover a wide variety of topics on normative multi-agent systems; the Emil Project (<http://emil.istc.cnr.it/>) that seeks to understand norm innovation; and the COST Action OC0801 working group on Norms (<http://www.agreement-technologies.eu/wg2>).



Normative Processes and the Norm Life Cycle

- 3.1** The published research on normative multi-agent systems contains a remarkable amount of structure, similarity, and connectivity. This underlying organization enables us to create a process oriented model of norm life cycle. In our model, we identify the following *normative processes*: *creation, transmission, recognition, enforcement, acceptance, modification, internalization, emergence, forgetting, and evolution*. From these, we identify *enforcement, internalization, and emergence* as three super-processes and *evolution* as an end-to-end process. The interaction and relationship between these processes is illustrated in figure 1, where each box represents a normative process and the thin arrows represent the flow of information between processes. The larger arrows can be read as "consists of the following subprocesses."
- 3.2** Recent research involving norm taxonomies and typologies reinforces our decision to use a process oriented approach and supports our claims that the existence of a norm life cycle is alluded to in the literature on norms and normative multi-agent systems (Finnemore and Sikkink 1998; Savarimuthu and Cranefield 2009; Verhagen 2007). As a comparison to our model, Savarimuthu and Cranefield (2009) describe a norm life cycle in which norm creation, spreading, enforcement, and emergence act as the primary stages. Each of these stages is supported by a set of relevant research publications. Internalization is mentioned, but not emphasized. The evolutionary nature of norms as described in our model is not mentioned, nor is importance of forgetting. There are, however, many commonalities in our findings which we consider to be important since we developed our model prior to knowledge of this particular paper. For example, the main ideas about how norms spread and emerge are similar, as are our conclusions and interpretations of the shared references. It is partly on this basis of similarity that we claim the fundamentals of our process model are justified.

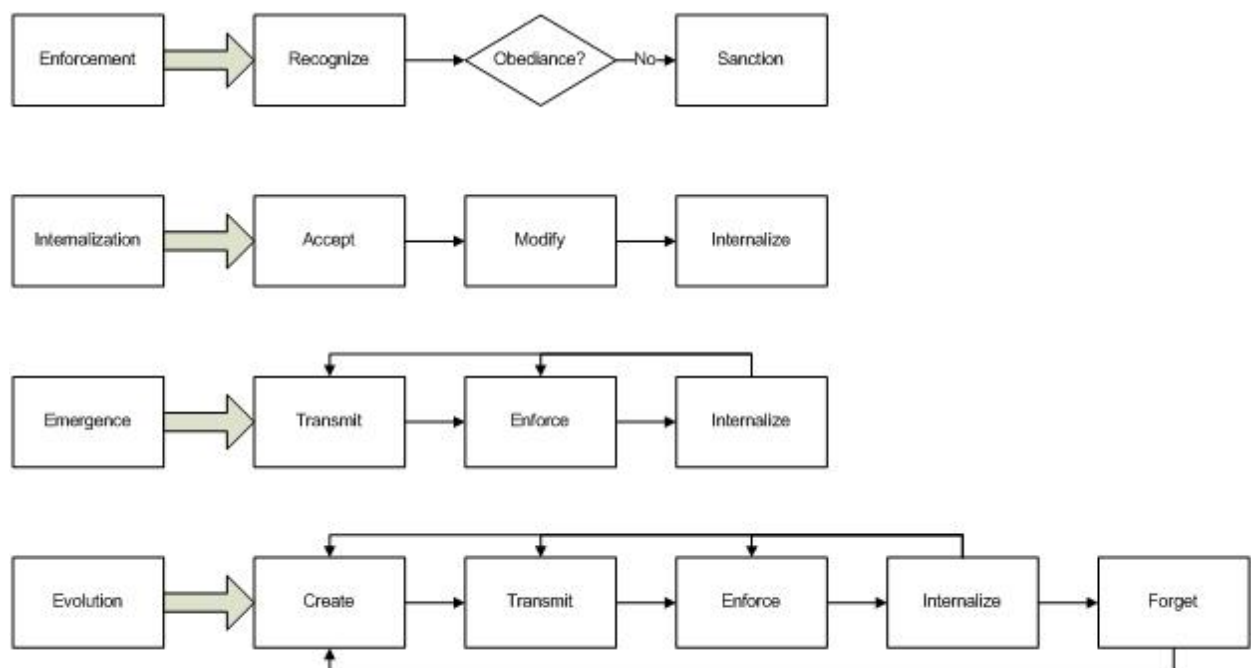


Figure 1. Normative Processes

- 3.3** In the proposed norm life cycle, ideas that will become norms are created as part of an *evolutionary* process. These new *potential norms* are then spread through active or passive *transmission*, depending on the system organization and allocation of control. As neighboring agents are exposed to the new norms, social *enforcement* ensures that those norms are acquired and internalized. *Internalization* refers to the shift in preference from the agent's original set of norms to the newly acquired norms. It also signals a shift in enforcement from external pressures to internal desires. This chain of transmission, enforcement, and internalization is known as normative *emergence*. The emergence sub-process continues until the potential norms are acquired, internalized, and rebroadcast by a sufficiently large subpopulation; at which point the potential norm becomes an actual norm (which we simply call a *norm*.) Eventually, conditions change and it becomes unreasonable to obey a particular norm. Consensus with regard to the norm disappears and the norm becomes invalid. When existing norms are no longer suited to the current conditions, they are candidates to be forgotten and new norms are created through an evolutionary process that begins the cycle anew.
- 3.4** Given our normative process model and its associated life cycle, we next describe in detail the major processes of enforcement, internalization, emergence and evolution as illustrated in Figure 1. First and foremost, however, we discuss transmission since it is an integral part of the normative process as a whole.

Transmission

- 3.5** Transmission (not directly pictured in figure 1) is responsible for the spread of norms from one agent to another. Over time, this process results in norms being diffused throughout an entire population. In the literature, we identify three core components that make this possible: agent relationship, transmission technique, and connectivity structure. Transmission (also called spreading) is a much larger topic than we are able to fully discuss here. It has applications in virtually every scientific field, from statistical mechanics to economics to biology to sociology, and is tightly connected to the field of networks. Audiences interested in knowing more about transmission in the general sense are encouraged to read papers on percolation theory and diffusion (Stauffer and Aharony 1994), as well as networks in general (Newman 2010).

Agent Relationship

- 3.6** There are three relationships between sender and receiver that dictate how norms spread: vertical, horizontal, and oblique (Boyd and Richerson 1985; Ehrlich and Levin 2005). Vertical transmission occurs in the presence of reproduction, when the relationship between sender and receiver is one of parent and offspring. Social learning processes ensure that the offspring acquire some or all of their parent's norms, resulting in the directed transmission of norms from one generation to another (Younger 2004). Horizontal transmission takes place when norms are transmitted between peers of the same generation (Boyd and Richerson 2002; Henrich et al. 2008). This process results in norms that spread laterally through a population, enabling agents to acquire norms from their unrelated

neighbors. Horizontal transmission has the potential to increase the diversity of an agent's behavior during the agent's own lifetime. The final transmission pattern we mention, oblique transmission, occurs when norms are transmitted from an authority figure to a set of subordinates. This process can result in norms spreading both vertically and horizontally. Oblique transmission is the approach used by centralized multi-agent systems and normative systems that concerned with Type I norms (Hoffmann 2005).

- 3.7 In normative simulations and systems that attempt to model real world phenomena, transmission patterns are not restricted to a single relationship. In models of human society, horizontal and vertical transmission occurs at the same time, with entities learning from their peers as well as their parents (Flentge et al. 2001a). The degree to which one relationship influences norm adoption over another is still under investigation within the social sciences.

Transmission Techniques

- 3.8 We identify two transmission techniques from the literature: *active transmission* and *passive transmission*. Active transmission is similar to the idea of push techniques in event-driven programming, and passive transmission is similar to the idea of pull techniques in event-driven programming.
- 3.9 Active transmission occurs when one agent purposefully broadcasts a set of norms to neighboring agents. Active transmission is typically accompanied by social enforcements in the form of sanctions that are intended to persuade neighbors to adopt the behavior. At the time of this research, normative multi-agent systems that use active transmission do not appear to be as common as those that use passive transmission, though one example of active transmission can be found in the use of norm entrepreneurs (Hoffmann 2005). It is our opinion that challenges associated with norm representation are partly responsible for the lack of research in active transmission techniques.
- 3.10 Passive transmission, on the other hand, occurs when an agent observes one of its neighbors performing some behavior. In passive transmission, the observer must infer the norms governing the observed behavior. Because the justification of a behavior is not explicitly given, it is possible for the acquired norm to be different from the original norm due to bias, error, or some other factor (Henrich et al. 2008). As with the process of active transmission, social enforcements are often used to coerce one agent to acquire the behavior of another. However, in the case of passive transmission these enforcements are often internal to the observer. The simplest approach to passive transmission can be seen in normative multi-agent systems where agents copy the norms of their more successful neighbors (Flentge et al. 2001b).

Connectivity Structure

- 3.11 The final component of transmission that we discuss is the social connectivity structure between agents (we simply call this the agent topology). This structure is often referred to as a social network in sociology (Hechter and Opp 2001b) or more generally, a network (Newman 2010). Research in this area of normative multi-agent systems has been primarily concerned with the effects of agent topology on the emergence of norms. This has resulted in experiments that examine the performance of systems where the agents are placed in 2-D lattices (Mukherjee et al. 2007; Younger 2004) and complex networks (Nakamarua and Levin 2004), specifically scale-free and small world networks (Zhang and Leezer 2009).
- 3.12 Simulations of normative multi-agent systems designed to investigate the effects of social network structure on norm emergence suggest that the agent topology determines the number of agents that cooperate with one another (as defined by how many agents play (Cooperate, Cooperate) in a game such as the Prisoner's Dilemma or Stag Hunt) as well as the rate at which cooperation is achieved (Zhang and Leezer 2009). These simulations also suggest that while the initial transmission patterns vary depending on network, the asymptotic behavior of the networks is equivalent; except in the case of a 1-D lattice structure (Nakamarua and Levin 2004). There is still some question as to the applicability of these results to dynamic networks such as those found in open multi-agent systems.
- 3.13 Research on agent topology in normative multi-agent systems has also identified two important properties that relate to the edge structure of the network; information sparseness and the interaction sparseness (Bendor and Swistak 2001). When agents are connected in such a way that information is sparse (agents only know what some of the other agents are doing, they are not omniscient) it becomes harder to impose sanctions since deviant acts are not always observed by those with the power to sanction. If, however, there is sparse interaction (there are some agents that cannot interact with each other) then the potential for a norm to emerge is largely unaffected so long as the network is dense in terms of information (every agent knows what every other agent has done). This result is due to the ability to maintain social enforcement. Currently, there have been no experiments to test

the effects of a network that possesses both sparse information and sparse interaction.

Enforcement

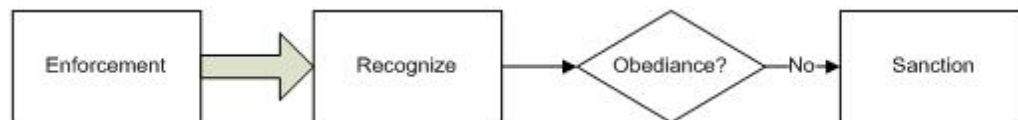


Figure 2. The Enforcement Processes

- 3.14** In the enforcement process (figure 2), one set of agents attempts to coerce another set of agents to adopt and/or obey a set of norms. To accomplish this, agent behavior must first be recognized and classified as normative or deviant. If the observed behavior is deviant, then the associated agent is negatively sanctioned (and potentially positively sanctioned if the behavior is normative.) Enforcement can be externally directed, internally directed, or motivational (as in the case of coordination motives) (Young 2008).

Norm Recognition

- 3.15** Norm recognition is one of the core challenges faced when building a normative multi-agent system (Boella et al. 2008b; Conte and Dignum 2001). Norm recognition refers to the ability of an agent to observe or interact with a group of agents and infer the correct norms of the agents in that group; humans are often able to accomplish this through conversation (Henderson 2005). Norm recognition is also concerned with the ability to detect deviant agents within a group. The detection of deviant agents is critical to proper enforcement (Therborn 2002). Many of the normative systems used to answer basic research questions use a very simple strategy to overcome the challenges of norm recognition (Castelfranchi et al. 1998; Hales 2002). In these systems, there is often only a single norm being used to investigate the questions of interest. The agents are divided into two groups, the normative group and the deviant group. The norm is known by all agents in the normative group. An agent is identified as part of the deviant group if it does not prescribe to the known norm. Information about the group orientation of agents is identified through interaction between agents and then communicated to other agents. Typically group orientation is only of concern to agents in the normative group. This approach works for detecting deviant agents in simple systems where all of the norms are known and violation has a direct and immediate impact, but it may not scale to more complex systems where multiple norms exist throughout multiple groups. Research into alternative techniques for norm recognition is required and starting to be pursued as part of the EMIL project (<http://emil.istc.cnr.it/>) (Andrighetto et al. 2007; Campenni et al. 2009) and the COST Action OC0801 working group on Norms (<http://www.agreement-technologies.eu/wg2>).

Sanctions

- 3.16** The coercive acts used in social enforcement are known throughout the literature as *sanctions*. A sanction can be a punishment for disobeying a norm or a reward for obeying it; though the common use is in the sense of punishment and the imposition of penalties for deviance. In general, sanctions are issued by either the deviant agent guilty of the violation, a neighboring agent that observes the violation, or an authority structure made aware of the violation. The act of sanctioning is often associated with a cost that must be paid by the sanctioning agent. This cost can be complex and take the form of relationship damage such as a loss of trust or friendship (Horne 2001), or it can be straight forward and result in the loss of some utility value in the case of rational agents. In computational systems, if the cost to sanction is too high, agents will not impose the sanctions and the desired behaviors will never be enforced and never stabilized in the system. Similarly, if the effect of the sanction is too small, agents will not have enough incentive to behave in the desired manner, and thus the desired behavior will never stabilize in the system (Axelrod 1986; Savarimuthu et al. 2008). If behavior fails to stabilize, then a norm will not emerge.
- 3.17** All of the research reviewed for this paper examines at most only two groups interacting with each other at the same time. While this approach is important to establish the fundamental ideas for modeling sanctions, it results in a limited view of social interaction in which one set of agents is normative and the other set is deviant. To fully appreciate the potential that normative multi-agent systems offer to solving practical problems and accurately modeling social interaction, there must be an investigation into the dynamics that arise when three or more groups, each with conflicting norms, interact with one another.

Types of Enforcement

- 3.18** Enforcement can be externally directed, internally directed, or motivational (as in the case of coordination motives) (Young 2008). Initially, enforcement is used during the transmission process to create an incentive for agents to adopt a new set of norms. After transmission, enforcement is used to ensure that agents continue to obey the acquired norms (Therborn 2002) until they are eventually internalized and external enforcement is no longer required. Once internalized, norms are enforced internally through intrinsic motivations (Epstein 2001).
- 3.19 Externally Directed Enforcement** External enforcement occurs when one agent observes another agent violating a norm (Flentge et al. 2001a; Galan and Izquierdo 2005; Savarimuthu et al. 2008) or during norm transmission when one agent refuses to adopt the norms of another. In response, the observing agent or an associated authority structure sanctions the offender. In artificial systems, sanctions are typically realized through a reduction in some type of problem dependant resource. The reduction is intended to be detrimental to the agent's ability to achieve its goal. In systems that use vertical transmission (Flentge et al. 2001a), externally imposed sanctions are sometimes used to reduce the chance that a given agent will mate; and thus his deviant behavior will not be passed on to the next generation, thereby reducing the overall level of deviance in the population (Caldas and Coelho 1999).
- 3.20** An alternative to direct resource reduction as a form of external enforcement is through the use of reputation (Axelrod 1986; Castelfranchi et al. 1998; Hales 2002; Younger 2004). In reputation-based enforcement systems, each agent maintains a list of other agents in the system—either agents that it has personally met, or agents that it has learned of from communication with other agents. Each agent in the list has a reputation value associated with it. If the reputation value is less than some tolerance value, the agent maintaining the list will either sanction the associated agent or refuse to interact with it. Research into the effects of reputation suggests that it can improve the stability of a norm, especially when agents are able to communicate with others and share their reputation lists (Hales 2002; Younger 2004). The use of stereotyping has also been investigated, with results showing that while it is effective in establishing reputation, it only works when the group being stereotyped is homogeneous (Hales 2002). In the past, reputation has been measured as a binary value, but it has been suggested that using an interval measurement can result in increased performance and the capability for forgiveness among agents in the system (Younger 2004). Forgiveness, however, has not yet been experimentally examined. We purpose that the true power of forgiveness emerges when agents are able to acquire new norms on-line through oblique or horizontal transmission. To our knowledge, this process has not yet been implemented in a real system.
- 3.21 Internally Directed Enforcement** In contrast to external enforcement, whereby a deviant agent is sanctioned by a third party, internal enforcement occurs when an agent sanctions itself for disobeying a norm. This type of enforcement is typically the case once an agent has internalized a norm. Emotions are thought to be one critically important factor in motivating internal enforcement (Scheve et al. 2005; Staller and Petta 2001), but in practice the sanction for deviance is typically encoded as part of the cost for executing an action in the presence of a norm. This behavior is particularly widespread in normative systems where the agents are considered rational. In these systems, sanctions are implicitly specified during the construction of the payoff functions (Axelrod 1986; Mukherjee et al. 2007; Savarimuthu et al. 2008). Research on more realistic mechanisms for internal enforcement such as the use of emotion, morality, and personal goals are still in their infancy when compared to the state of the art in external enforcement.
- 3.22 Motivational Enforcement** Motivational enforcement (in particular, coordination motives) is a type of social enforcement resulting from "common sense" in a system of rational agents. If a particular norm is expected by all agents in the system, then there is no benefit for an agent to violate it; thus the only rational choice for the agent is to obey the norm. There is no need for an explicit sanction since any deviant choice would lead to suboptimal behavior and all the agents are rational (thus the agents would not choose the deviant action when the other yields a higher reward.) This method of enforcement is often seen in problems centered on cooperation and coordination where only one choice is the "correct" choice (Young 2008). As a concrete example, consider an autonomous car that needs to drive on a highway—it can drive on the left side of the road or the right, but there is only one correct choice—the social norm of the system. If the control system attempts to deviate from the established norm it will suffer severe physical damage and fail at its task.

Internalization



Figure 3. The Internalization Process

3.23 Internalization (which can be viewed as a type of immergence (Conte et al. 2007; Andrighetto and Conte 2009)) is the process (figure 3) in which agents acquire and integrate new information into their cognitive structure (Verhagen 2001). Initially, an agent becomes aware of a norm through the transmission process. Social enforcements result in pressure on the agent to adopt and obey the new norm over any existing desires. To handle these pressures, an agent invokes a conflict resolution process (acceptance) that results in the acceptance or rejection of the transmitted norm. If the agent decides to accept the norm, there is a chance that the information is modified during the transcription process (modification), when the agent is integrating the norm into its own knowledge base—this may be due to bias, error, or some other unforeseen scenario. Once the norm has been acquired, social enforcements continue to exert pressure against internal desires and motivations and ensure that the agent continues to obey the norm. Over time, the norm is integrated into the desires of the agent and priority shifts from the original norms possessed by the agent to the newly acquired norm (internalization). Defined in a more concise form, internalization can be seen as the measure of commitment an agent feels to the execution of new norms (Campbell 1964). In humans, theories such as Self-Determinism Theory (Deci and Ryan 2000; Ryan and Deci 2000) attempt to explain how internalization may occur and identify key components, such as motivation and need, which appear to be critical in the internalization process. It is essential to understand these theories if the goal of a normative system is to accurately model human behavior. However, the details of Self-Determinism Theory and other psychological and sociological theories of motivation (see Deci and Ryan (2000) for a summary) are outside the scope of this current paper.

Norm Acceptance and Modification

- 3.24** Norm acceptance and modification are precursors to the actual act of internalization and integration of information into the agent's deep cognitive structures. Before a norm can be internalized, it must first be accepted by an agent. Norm acceptance is a conflict resolution process in which external social enforcements compete against the internal desires and motivations of the agent. If the new norm is in conflict with existing norms and may lead to inconsistent behaviors, or if the cost of accepting the new norm is too high, it will be rejected (Meneguzzi and Luck 2009). However, it is feasible that there are circumstances where it is desirable for an agent to accept conflicting norms, especially if an agent is able to forget existing norms.
- 3.25** Once an agent has made the decision to accept a norm, it must then undergo a process of transcription in which the norm is added to the agent's knowledge base. During this phase, it is possible that the norm might undergo modifications due to bias, inferential errors, or some other unforeseen scenario (e.g. perhaps there is a chance for mutation). These modifications can be seen as a side effect of possessing incomplete information due to autonomous operation; the results are akin to human misunderstanding.
- 3.26** Once a norm has been accepted and integrated into an agent, it is then reinforced through techniques such as social monitoring (Conte and Dignum 2001) to ensure it is in compliance with the rest of the system. Failure to properly obey a norm may result in sanctions that force the agent to re-evaluate its behavior and potentially attempt to reacquire one or more norms. This reinforcement process continues even after a norm has been internalized.

Methods of Internalization

- 3.27** The actual mechanisms used in human norm internalization are still unknown. This leaves designers of normative systems free to explore their own solutions, though in general internalization methods for normative agents do not appear to be a popular topic of research and internalization itself is seldom considered in the normative systems literature (Meneguzzi and Luck 2009), although multiple authors have cited the need for investigation (Axelrod 1986; Neumann 2008; Saam and Harrier 1999) and the topic is starting to achieve more recognition (Andrighetto and Conte 2009). In the existing literature, norms are typically internalized as soon as they are acquired, without any sort of modification. This is most obvious when the norms are designed offline, but also applies to systems that allow norms to spread between agents. Precious few alternatives to this approach have been tried. In one system developed by Verhagen (2001), agents maintain a decision model for themselves and the group to which they belong. Internalization is then measured by the degree of similarity the two models have to one another, and a norm is said to be internalized when the self-model matches the group model. Another approach is to measure internalization as a function of the neighborhood size an agent observes when considering social retaliation for violating a norm. In this method, a norm is said to be internalized when the agent is no longer concerned with external pressures and only internal desires are used to decide if the agent will obey or violate a norm (Epstein 2001). The Belief-Obligations-Intentions-Desires (BOID) agent architecture (Broersen et al. 2002; Neumann 2010) offers hope for future investigations into norm internalization. It specifically identifies internalization and enables it by separating group norms from traditional BDI agent components and specifying a conflict resolution mechanism that can then be used to select

behaviors based on autonomy values that represent the priority of an agent's own desires versus the desires of the group.

Emergence

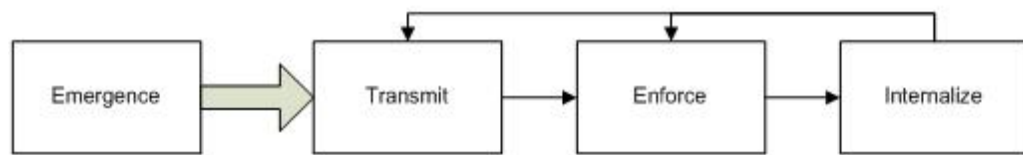


Figure 4. The Emergence Process

- 3.28** When speaking of norms and agent-based systems, the term *emergence* is used to describe two phenomena. The first use of emergence is to describe norm creation on a micro scale; the second use refers to norm establishment on a macro scale. We use emergence in the latter sense and say that a norm has *emerged* when it has been acquired by a sufficient number of agents in a population. Thus *emergence* is the process under which potential norms transform into societal norms. This process is also referred to as *evolutionary* norm emergence (Flentge et al. 2001a). In terms of the process model presented in figure 1, normative emergence (figure 4) is the result of interactions between transmission, enforcement, and internalization.
- 3.29** Research on norm emergence within the normative multi-agent system literature has centered around three main areas. The first area of research deals with investigating into the use of game theory to explain the dynamics of norm emergence. The second area examines the relationship between sanctions and norm emergence. The third area attempts to understand the impact of transmission on norm emergence.

Game Theoretical Analysis

- 3.30** The popularity of research on norm emergence is due in part to research on the emergence of cooperation and coordination in multi-agent systems (Axelrod and Hamilton 1981). Cooperation and coordination are often approached from a game theory perspective. A population of agents plays a game where each action has an associated payoff. The selection of actions is guided by the agent's strategy, which can be viewed as a potential norm. Typically (in the case of rational agents), the agent attempts to maximize its payoff in every round. In evolutionary game theory (Bendor and Swistak 2001), the agents are able to adapt their strategies through progressive rounds in an attempt to achieve higher payoffs during iterated games. Cooperation or coordination is said to emerge in a system once a sufficient number of agents all play the same strategy (now a societal norm) or interact in such a way that the system reaches a steady state. Game theory is used to study norms other than cooperation or coordination in the same manner. Norms are said to emerge once a sufficient proportion of the population settles on the same set of behaviors. One way to measure the degree of emergence is to monitor the average payoff of the population and observe when it exceeds a specific tolerance value. Norms are stable once they have emerged and the number of agents obeying the norms does not decrease below the tolerance value (Axelrod 1986; Mukherjee et al. 2007; Zhang and Leezer 2009).

Effects of Enforcement on Emergence

- 3.31** Recent research centered on norm emergence builds on ideas from evolutionary game theory and incorporates cognitive models, learning mechanisms and specific connectivity structures. In particular, the use of cognitive models and agent architectures such as the BDI model allows researchers to examine the role of enforcement and social learning in norm emergence.
- 3.32** The results from research on the relationship between sanctions and emergence (Axelrod 1986; Flentge et al. 2001a; Galan and Izquierdo 2005) suggest that the cost of a sanction, both to the enforcer and deviant, is key to its usefulness as a control element; but the dynamic nature of social interaction makes it hard to generalize results obtained from static environments. Sanctions in and of themselves do not appear to be enough to cause the emergence of norms. The degree of punishment or reward due to a sanction must also be taken into account, and it must change with the degree of emergence to ensure that deviant agents want to acquire the norms and normative agents want to continue obeying them. If the cost to sanction is too high, or the effect of the sanction is not severe enough, the norms will not emerge (Caldas and Coelho 1999; Savarimuthu et al. 2008).

Effects of Transmission on Norm Emergence

- 3.33** More recently, the effects of agent connectivity and the underlying social networks of normative multi-agent systems has been a growing topic of study. Research on the relationship between agent topology and norm emergence has shown that in asymptotic time, the underlying network topology contributes minimal difference to the emergence of norms if it is static. Additionally, it is shown that when agents acquire norms through a social learning processes based on the composition of their neighbors, one norm tends to become dominate the group; however when each agent learns individually and selects a norm to follow in an independent fashion, multiple norms are able to co-exist within the same group (Boyd and Richerson 1985; Boyd and Richerson 2005; Nakamarua and Levin 2004). The precise relationship between the number of social and individual learning agents required to produce multiple norms in a heterogeneous population is still unknown. In the case of dynamic networks, research has shown that by using specific learning heuristics, norms can emerge in a system of selfish agents that attempt to maximize their own utility (Zhang and Leezer 2009). Further research is required to discover additional factors responsible for the emergence and stability of norms in both human and artificial populations.

Evolution

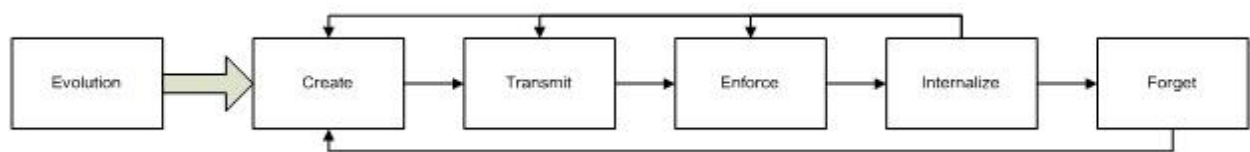


Figure 5. The Evolution Process

- 3.34** Normative evolution (figure 5) is an end-to-end process that fully encompasses the norm life cycle. It is an emergent phenomenon arising from the interaction of norm creation, norm emergence, norm adaptation, and norm removal. However, in the literature on normative multi-agent systems, evolution is often treated as a synonym of emergence with the majority of work focused on how a norm spreads throughout a population. To oppose this perspective, we focus on the processes of norm creation, norm adaptation, and norm removal so as to illustrate that while the evolutionary process contains elements of norm emergence, it is a much broader concept.

Norm Creation

- 3.35** Norm creation (or innovation) is the process by which new norms are introduced into a normative system. Current research in this area is inspired by methods of innovation within natural systems, but when applied to artificial systems these approaches can be generalized. Our current understanding of norm creation is restrained by our lack of knowledge in the areas of innovation, imagination, and creativity.
- 3.36** There are three methods that lead to the creation of norms in the natural world (Boella et al. 2007; Finnemore and Sikkink 1998; Lopez et al. 2005; Savarimuthu and Cranefield 2009): spontaneous emergence from social interaction, decree by an agent in power, and negotiation by agents within a group. However, research on norm creation is typically conducted at the macro-level and aimed at investigating the relationship between social and individual learning strategies, typically using imitation and innovation, and their effect on the dynamics of emergence (Boyd and Richerson 2002; Boyd and Richerson 2005; Nakamarua and Levin 2004). Results from this type of experiment suggests that both learning components are needed in normative systems, but an optimal combination has not been discovered and is more than likely context dependent. More recently, the idea of social enhancement has been proposed as a mechanism for norm creation (Franz and Matthews 2010). In this approach, agents do not imitate their neighbors; instead they observe the objects involved in nearby interactions and play with them. Reinforcement learning is then used to guide the agents towards the correct use of the objects. Social enhancement is weakly connected to the idea of combining imitation with reinforcement learning to allow further exploration of a learned concept.
- 3.37** From the literature specific to normative multi-agent systems, we identify two general methods of norm creation in artificial agents: offline design and autonomous innovation, where offline design is by far the most common approach to norm creation (Savarimuthu and Cranefield 2009). In offline design, system designers specify what norms a system will follow and encode them directly into the agents. If the system requires new norms, they must be inserted by the designers. This approach works well for simple systems, but for any reasonably complex system offline design can fail to

capture the intricacies and minutiae required for realistic performance. In contrast to offline design, autonomous innovation requires the agents of a system to create new norms without external interference. In order for this to occur, the researchers must address the challenges of *ideation* ("how an idea for a behavior that becomes a norm gets invented in the first place") and *filtering* ("which ideas are accepted and which are rejected") (Ehrlich and Levin 2005).

- 3.38** Current efforts to investigate autonomous innovation and address the challenges of ideation and filtering in artificial systems have been focused on machine learning and game theory (Bendor and Swistak 2001; Mukherjee et al. 2007; Savarimuthu et al. 2008; Savarimuthu and Cranefield 2009). In game theoretical approaches, ideation is often reduced to offline design and filtering to the selection of the most successful behavior. Machine learning approaches handle ideation through search, but take the same selection-based approach as game theory. The EMIL project (Andrighetto et al. 2007; Campenni et al. 2009) proposes to create alternative innovation methods based on cognitive architectures and further research into norm creation by creating a simulator that will allow exploration and experimentation on norm innovation theories.

Norm Adaptation

- 3.39** Norm adaptation refers to the process by which the norms of a system change over time. In theory, this is accomplished through the use of social learning processes such as imitation and socialization. Once a new norm is created, it has the potential to emerge. If it successfully emerges, it either competes against or replaces the existing norms that are internalized within the same context. Norm adaptation can be illustrated with the Iterated Prisoner's Dilemma by dividing agents into one or more groups and assigning an IPD strategy such as Grim Trigger, Tit for Tat, or Tit for Two Tats to each agent (Bendor and Swistak 2001). The agents play against one another until one strategy appears to be stable. Then, a different strategy can be introduced into the stable system and play resumed. By introducing a new strategy into a stable system, it is possible to test the ability for the new strategy to penetrate the society and establish itself as dominant. If the new strategy is able to stabilize, then it can be claimed that the norms of the system have adapted to take advantage of new information. This is adaptation at the macro level.
- 3.40** Alternatively, a norm can be modified when it is acquired by an agent. During the transmission phase, bias from the receiving agent can result in subtle changes. This is mainly a concern in systems where the norm cannot be copied over directly—in artificial systems this would require an ingenious method of implementation that we have not found in the existing literature. A question of practicality and usefulness also arises when this notion is applied to artificial systems. Modifications that occur during transmission and internalization are adaptations at the micro-level.

Norm Removal

- 3.41** Norms that are replaced as a result of norm modification may or may not be forgotten. The ability to forget a norm (removal) is conceptually important when a system becomes complex or is limited in resources. We have not found any specific research on the performance effects of norm removal, but the process is often implicit in many systems that implement norm modification (Bendor and Swistak 2001; Mukherjee et al. 2007). More recently the larger question of norm adaption, of which removal is essential, is beginning to receive more attention by researchers (Lopez-Sanchez et al. 2009). As more academics become interested in this topic, we have no doubt that all the components of the norm evolution process will be examined.



Applications of Normative Multi-agent Systems

- 4.1** The research on and related to normative multi-agent systems contains a growing amount of work on the application of social norms to computational systems. This body of applied work currently centers around two main areas: the role of norms in society and social simulation; and the use of normative concepts to improve cooperation, coordination and control in engineered multi-agent systems.

The Role of Norms in Society

- 4.2** The use of multi-agent simulation to investigate social phenomenon is not new, and a survey of the complete domain is outside the scope of this paper. Instead, we focus on four common themes that appear in the work specific to normative multi-agent systems: social control, benevolence, reciprocity, and institutions.
- 4.3** Social control through the use of normative agents and reputation is one of the more researched

domains in normative multi-agent systems. In one of the earliest experiments (Conte and Castelfranchi 1995), a "finders-keepers" norm is used to try and reduce aggression (the number of attacks in a simulated run) and increase social equality (how evenly resources are distributed) within a population of foraging agents. To experiment with the finders-keepers norm, a simulated world was constructed on a 2-D lattice with resource nodes placed randomly on a subset of the available cells. At each time step, agents move through the environment in search of resources. At the start of the simulation every agent is assigned ownership over one or more nodes, and every agent knows who owns every node. When an agent encounters a resource, its behavior depends on the strategy it is following. The simulation allows agents to be blind (they harvest or attack any agent that is on the resource node), strategic (they harvest or attack any agent on a resource node as long as that agent is weaker), or normative (they will only harvest nodes they own). The results from this simulation show that when all agents are normative the aggression of the population dramatically decreases and the overall equality of agents increases when compared to the other strategies. The original research has since been extended (Castelfranchi et al. 1998) to investigate the effects of a binary reputation value on aggression control and inequality in artificial societies. Other extensions investigate the use of stereotypes with reputation to control aggression and social inequality (Hales 2002; Younger 2004). In the more recent simulations, reputation is measured on an interval instead of being binary. It is shown by these experiments that reputation is a powerful tool for social enforcement, with interval reputation allowing smoother control and the ability to easily forgive agents over time. Stereotypes can also play a powerful role in classifying deviant agents, but care has to be taken not too over-fit the population and the impact is most beneficial when all groups are homogeneous. A combination of stereotypes and interval reputation may provide an even more powerful mechanism of control, allowing one agent in a group to serve as a warning to others.

- 4.4 The effects of private property, heritage and power have also been investigated for their effects on aggression and social inequality (Saam and Harrier 1999). Results from these investigations suggest that under certain social conditions, the finders-keepers norm can lead to an increase in inequality; but the more important result is the observation that power appears to play an integral role in normative transmission and enforcement, and requires more in depth study. The finders-keepers environment has also been used to study the effect of emotion on normative compliance (Staller and Petta 2001). By adding an emotional component to normative agents, they are able to deviate from the finders-keepers norm when the agent's desire to feed outweighs its desire to comply. The results of experiments on emotion in normative agents reinforce the idea that norms bring order and overall well-being to a population. The true potential of emotions lies in their ability to act as internal sanctioning mechanisms.
- 4.5 Research into the social function of norms has also lead to investigations of benevolence norms. There is not a fully agreed upon definition of benevolence, but one attempt at a definition of compromise (Mohamed and Huhns 2001) states that an agent is benevolent if it voluntarily helps other agents, acts towards the benefit of society, does not expect an immediate reward for its actions, and does not let benevolent actions stop it from attaining its own personal goals. Benevolent agents (agents implementing the benevolence norm) are rational because they act without being asked and will not act in a benevolent way if it prevents their personal goals from being accomplished (Mohamed and Huhns 2001). To measure the functionality of benevolence, the "Mattress in the Road" simulation was developed. In this simulation, agents drive along a road and occasionally drop an object that slows down traffic. The agents are divided into two groups, those following a benevolence norm and those who do not. Benevolent agents take the time to remove the object, paying a cost to do so, while non-benevolent agents pay a smaller cost to avoid it. When an agent enters the road, it is assigned a random time limit in which it must get to the other side, if an agent uses more than its allotted time it increases the chance that other agents will be late. The overall fitness of the society is rated in terms of the average time taken to drive down the road. Through repeated runs of the simulation, it is shown that the benefits of benevolent behavior depend on the context. Under the definition of *rational benevolent agents* (agents subscribing to the compromised definition of benevolence), benevolent agents are at worst equal to non-benevolent agents and often result in better performance for the entire society.
- 4.6 Reciprocity (the idea that an agent responds to an action with a similar action, often represented by the concept of sharing) is another area in which normative multi-agent systems have been used to try and understand the impact of a particular social norm. Research in this domain suggests that reciprocity can result in agents being obligated to one another, and thus increasing the level of cooperation when compared to agents that do not reciprocate (Younger 2004). The idea of reciprocity is in linked to trust, another area that has begun to see more interest by the normative multi-agent systems community within recent years. However, a discussion of trust and its role in normative systems is beyond the scope of this current paper.
- 4.7 The final area of research involving norms at a social level that we mention in this paper is electronic institutions. An electronic institution is an "agent environment that can regulate and direct the interactions between agents" (Grossi et al. 2005) by providing a set of rules that dictate what agents are permitted and forbidden to do under various circumstances (Bou et al. 2009). The core notions underlying an electronic institution include roles, dialogic frameworks, scenes, performance

structures, and normative rules (Artikis and Pitt 2009; Bou et al. 2009; Esteva et al. 2001). In an electronic institution, institutional ontologies are often used to allow norms to be defined at general level and then redefined at a specific level depending on context they are used in. These norms, and their associated sanctions, can then be used to provide social control (Blake 2009). Institutions are also used in BDI-inspired architectures to control agent behavior in open marketplace systems (Hahn et al. 2005), systems that allow the formation of coalitions through social contracts (Sauro 2005), and as gate keepers to restrict interactions in artificial societies (Davidsson and Johansson 2005). At the forefront of research on electronic institutions are adaptive electronic institutions (Bou et al. 2007; Bou et al. 2009) in which the institution itself is able to change its own regulations in response to agent behavior so that the institutional goals can be accomplished in an open multi-agent system. Electronic institutions are a growing research area with wide ranging implications for normative and open multi-agent systems; however a full discussion of this topic is outside the scope of this paper. A related area of research is multi-agent organizations, where one goal is to identify how to structure the relationships between agents. Many multi-agent organizations also incorporate normative ideas, traditionally as constraints on the behavior of agents within the system. A discussion on multi-agent organizations is outside the scope of this paper, but interested readers are encouraged to examine the Moise Organization Model (<http://moise.sourceforge.net/>) by J. F. Hubner, J. S. Sichman, and O. Boissier as a starting place. For a more general introduction to Multi-Agent Organizations and their uses, a good starting point can be found in the work of Victor Lesser, in particular Horling and Lesser (2005).

Application of Normative Concepts

- 4.8** In addition to simulating social phenomenon, research on normative multi-agent systems and norm-inspired computing is also applied to the challenges faced when designing and implementing open multi-agent systems, where an open multi-agent system is a multi-agent system in which agents may not share the same architecture or the same goals, interactions between agents cannot be predicted in advance, and agents are able to join and leave the system freely (Artikis and Pitt 2009). Related to the implementation of open multi-agent systems, there is a large body of research devoted to designing normative agent architectures and frameworks (Broersen et al. 2002; Castelfranchi et al. 1999; Dignum 1999; Lopez et al. 2005; Meneguzzi and Luck 2009; Neumann 2010), but a survey of normative agent architectures is outside the scope of this current paper.
- 4.9** With regards to the challenge of designing an open multi-agent system, the primary role of normative concepts has been to improve coordination, cooperation, and control among the agents (Carvalho et al. 2005; Duangsuwan and Liu 2009; Kamara et al. 2005; Villata 2010; Wu 2008). A general method to accomplish this involves foreign agents being able to recognize and acquire the local norms of the system through direct or indirect interaction. Once the norms are acquired, the foreign agents are able to interact with the local agents and assume task-related roles to provide additional resources towards solving the current problem. Once the problem is solved, agents can remain in the local system and take on a new task, or leave the system. Practical applications that incorporate this method to one degree or another include norm-based Contract Net Protocol (Wu 2008), a norm-based system for personal environmental control in modern buildings (Duangsuwan and Liu 2009), an agent architecture that can enable ad-hoc networks to self-regulate and self-manage (Kamara et al. 2005), a graph-based tool for metamodeling agent-based system requirements called NorMAS-RE (Villata 2010), and a law enforcement architecture aimed at reducing risk and increasing the dependability of software infrastructures (Carvalho et al. 2005).
- 4.10** Outside of a few specific social contexts and open multi-agent systems, the application domain for normative multi-agent systems is still relatively unknown and not completely understood. As future research begins to answer fundamental questions about the nature of norms and their role in society, this will change. Over time, we predict that the application domain of normative multi-agent systems will grow as their related problem space is defined.



Future Research Directions

- 5.1** Significant advancements have been made in regards to normative multi-agent systems over the past twenty years, many of which directly contribute to our ability to create a process driven norm life cycle. Though as evidenced by the gaps in our model, there are still many domains that require further research before we can fully understand how to best model and make use of norms in artificial systems.
- 5.2** Five levels of norm development have been proposed to classify for normative multi-agent systems (Boella et al. 2008a):
1. Norms are designed off-line and implicitly part of the agent's behavior
 2. Norms are explicitly represented in the agents
 3. Agents can add and remove norms according to some predefined rules
 4. Agents are able to create and manage their own norms

5. Agents are able to use norms to create a "moral reality"

- 5.3** Most existing normative systems belong in the first two levels. In order to enable advancements in future systems, the following areas have been suggested (Boella et al. 2008a) for further research: norm representation, recognition, reasoning, deviance, institutions, and autonomy and internalization. Based on our own research, we identify several additional areas that deserve further research: norm creation, modification, and removal; norm transmission and social learning; the application of social networks analysis to normative emergence; norm enforcement; and the fundamental nature of norms.

Suggested Areas for Future Research

- 5.4** The first area requiring additional research is norm representation. Without an adequate representation scheme, it is impossible to reason about or manipulate norms. Currently many systems are implicit in their representation, but we have seen a trend towards attempting to develop external formats and specifications; many of which are based on deontic logic. In some ways, this is similar to the early days of logic-based AI. Representing norms as condition/action pairs of logical statements seems fairly natural, but is it the most appropriate? Are there alternatives that can be more easily manipulated by automated processes similar to evolutionary mechanics such that norm creation and modification are less challenging?
- 5.5** Research on norm creation, adaptation, and removal is currently under-represented in the published literature, but recent research (Lopez-Sanchez et al. 2009) suggests that these topics are beginning to receive more attention. In order to understand creation at the micro-level, we must first understand the process of *ideation* (how the idea for a behavior that becomes a norm initially appears) (Ehrlich and Levin 2005). There is also the question of transmission's role in the norm creation process; can the subjective (biased) interpretation of the observed behavior partly explain the emergence of new norms (Henrich et al. 2008)? Furthermore, what is the role of the environment in norm creation (Neumann 2010; Zhang and Leezer 2009)? Finally, do any existing search strategies, such as Novelty Search (Stanley 2010) hold promise for discovering new behaviors? We have found a stark lack of publications on the modification of norms during the lifetime of an agent and the effect of personal bias on the learning of new norms. We have also not seen any research concerned with the ability to forget a norm after it has been learned. Both of these areas are critical to a theory of normative evolution and may have far reaching applications in industrial systems where resources are scarce or expensive.
- 5.6** Research into norm transmission through social learning mechanisms such as imitation has increased over the past decade, but only a handful of papers (Nishizaki et al. 2008) investigate methods other than imitation. Can the same learning mechanisms used by humans be applied to artificial systems? What are those mechanisms? What lessons can we take from decades of research on machine learning? Two interesting phenomenon that deserve further attention are *cognitive attractors* and *selective attention* (Henrich et al. 2008). Cognitive attractors are clusters of behavior form that serve as a template for what is "correct" to nearby agents seeking to conform. Selective attention is the idea that agents are particular about who they learn from. In human systems, it is not the case that an agent will be open to transmission from every neighbor in its social network. The role of power in the transmission process has been identified as a critical factor in human social systems, but it has not received as much attention as some (Ehrlich and Levin 2005; Saam and Harrier 1999) deem necessary.
- 5.7** The use of social networks to analyze the underlying connection topologies of multi-agent systems has also started to come into favor more recently. Experiments comparing emergence and spread over lattice vs. power-law vs. random networks has shown that while there may not be extreme differences in asymptotic time when the network is static, there is an effect in the short term and when the networks are dynamic (Nakamarua and Levin 2004; Zhang and Leezer 2009). Do ideal network structures exist for certain problem classes? How do hybrid networks perform (where a hybrid network is defined as a network made up of multiple sub-networks, each with a specific topology)? What are the other differences between static and dynamic networks, and do the tools exist to analyze norm evolution and emergence in dynamic networks?
- 5.8** One of the most critical areas that identified for research is the internalization process. At the moment norms are generally internalized as soon as they are accepted. The BOIDS architecture (Neumann 2010) isolates norms from personal desires, but does not provide a mechanism for their transfer. There has been some work (Verhagen 2000) using the notion of an autonomy threshold for the agent, where the autonomy value is used to influence the probability that an agent obeys a norm over a personal desire, but this seems overly simplistic. Additionally, the choice to obey a norm is typically deliberative and not random. Motivation is also missing from many of the current systems. A theory of normative reasoning needs to be developed to address this challenge. However, in order for internalization to happen at all the agent must first be able to recognize and acquire a norm. With regards to recognition, how does an agent know that a norm is guiding observed behavior? How

does it infer what that norm may be? Perhaps more importantly, how does an agent know when there *should* be a norm? Going back to the notion of normative reasoning, what is the role of *filtering* (Ehrlich and Levin 2005) in determining which norms are accepted or rejected? Finally, how do agents solve conflicts that may arise when a norm is inconsistent with personal values? What if it is unable to bear the cost of rejection?

- 5.9 The effect of social enforcement through the use of sanctions is fairly well established in the literature. Experiments have shown that sanctions lead to norm stability, but only if the cost and force of the sanction is within a "goldilocks zone" that appears to completely depend on the problem context. Are there underlying models that can be used to explain and predict the effect of sanctions in general? It has been mentioned that one of the defining attributes of a norm is that it can be disobeyed. How do sanctions impact this ability while still ensuring norm stability? Furthermore, when should an agent disobey a norm and suffer the cost of a sanction? How much control should norms have over an agent's behavior? Finally, are there alternatives to sanctions and social enforcements, such as persuasion or altruism or benevolence, that can provide norm emergence and stability (Ehrlich and Levin 2005)?
- 5.10 In addition to the challenges related directly to the normative processes we present in this paper, there are still a number of basic science questions about norms must still be answered. For instance, how do norms interact with other cultural and psychological factors? How do norms actually result in coordination, cooperation, and organization? There are also challenges faced at a more practical level, such as how do you conduct verification and validation on normative systems and ensure that the results are correct (Galan and Izquierdo 2005; Saam and Harrier 1999)? In particular, what degree of sensitivity analysis and exploration of a normative system's parameter space is appropriate? How important is it to let the system operate for extended periods of time to observe the asymptotic effects, given the dynamic nature of social interaction and speed at which decision trajectories change? Multi-agent systems often handle communication through specialized protocols such as FIPA or KQML. Could this approach be extended to represent norms for inter- and intra-agent communication?
- 5.11 We end our suggestions for future research with a final question that connects normative systems to their complex cousins: Is the emergence of norms an expected phenomenon of social systems? Is it the natural result of social interaction under all possible conditions?



Summary and Conclusions

- 6.1 Normative multi-agent systems are a relatively new approach to addressing the challenges encountered in open multi-agent systems. Norms form the basis of these systems, but pinning down their exact nature and responsibilities is difficult due to the diversity of existing definitions with respect to the term *norm*. Norms have been defined as behavioral constraints, obligations, expectations, averages, solutions to macro-level problems, and regulatory and control devices for decentralized systems. We choose to define them in a manner that offers compromise between the differences, and suggest that norms are behavioral rules agreed upon by a sufficient proportion of a population; subject to willful violation and social enforcement. Norms can be acquired by local and foreign agents through social and individual learning processes and are able spread beyond their initial population and displace existing norms. Norms can further be defined by categorizing them based on morality, origin, economics, and function. Moral norms concerned with the behaviors an agent ought to perform, regardless of any consequences. When categorized by origin, norms are divided into those produced by a group in power, and those that emerge naturally through social interaction. Economic categorizations look at norms in terms of their cost to members of a group. A functional approach describes norms based on their effects within a society. Norms form the basis for normative agents and normative multi-agent systems.
- 6.2 A normative multi-agent system is itself a multi-agent system built from normative agents, where normative agents combine the concepts of norms with traditional agent concepts. Normative agents must be able to represent norms, and current approaches accomplish this through the use of logic, rule-based systems, binary strings, and implicit definition in game theory strategies. They must also support a handful of processes that constitute a norm life cycle and encapsulate the critical operations required to create, modify, detect, transmit, reason about, and enforce norms.
- 6.3 In the norm life cycle described in this paper, ideas that will become norms are created as part of an evolutionary process. These new potential norms are then spread through active or passive transmission techniques, depending on the system organization and allocation of control. As neighboring agents are exposed to the new norms, social enforcement ensures that those norms are acquired and internalized. Internalization refers to the shift in preference from the agent's original set of norms to the newly acquired norms. It also signals a shift in enforcement from external pressures to internal desires. This chain of transmission, enforcement, and internalization is known as normative emergence. The emergence sub-process continues until the potential norms are acquired, internalized, and rebroadcast by a sufficiently large subpopulation; at which point the potential norm

becomes an actual norm (which we simply call a norm.) Eventually, conditions change and it becomes unreasonable to obey a particular norm. Consensus with regard to the norm disappears and the norm becomes invalid. When existing norms are no longer suited to the current conditions, they become candidates to be forgotten and new norms are created through an evolutionary process that begins the cycle anew.

- 6.4** This norm life cycle model is constructed from the normative processes of creation, transmission, enforcement, internalization, emergence, and evolution. Transmission describes how agent relationship, transmission technique, and the network topology of agent connections impact the spread of potential norms through a population. Enforcement handles norm recognition and sanctions, including external and internal enforcement measures that coerce agents to adopt and obey new norms, even when it may be against their individual interests. Internalization is responsible for the principles behind norm acquisition and the effect that bias can have when adopting new norms. The most common approach to internalization in existing normative multi-agent systems is the idiom of *acquisition is internalization*. Emergence is the process in which transmission, enforcement, and internalization converge to transform potential norms into societal norms, spreading them through the population to produce macro-level cooperation behaviors. The final process, evolution, is an end-to-end process that unites emergence with norm creation, norm modification, and norm removal. Normative evolution describes the entire norm life cycle, from birth to death.
- 6.5** Normative multi-agent systems are still searching for their true application domain, but much of the existing work centers around using them to simulate social interaction and theories from the social sciences, as well as solve problems related to open multi-agent systems. To date, normative multi-agent systems have been used in research on coordination, cooperation, social control, benevolence, reciprocity, trust, electronic institutions, agent architecture design, software engineering, adaptive environmental control, and autonomous configuration of ad-hoc networks.
- 6.6** The past twenty years have seen the research on normative agents and normative multi-agent systems come a long way, but there are still significant gaps in our knowledge of fundamental concepts. There is a dire need to further research in the domains of norm representation; norm creation, modification, and removal; norm transmission and social learning; the application of social networks analysis to normative emergence; norm internalization; norm enforcement; and even the very definition of what it means for something to be a norm and the consequential behaviors and expectations that follow. Only by challenging ourselves to step into the unknown and cross from the world of the computational to the social can we hope to find the answers we require for significant progress. It is our hope that this paper serves as a guide to new researchers and presents the topic from an integrated perspective that illustrates the complexity and multidisciplinary nature of the subject matter. After all, social norms govern almost every form of interaction (Young 2008), and the future of agent-based computing is rooted in interaction.



Acknowledgments

This work was supported in part by ONR grant #N000140911043 and General Dynamics grant #100005MC.



References

- ALBERTI, M et al. (2005) Mapping Deontic Operators to Abductive Expectations. In *Proceedings of the Symposium on Normative Multi-Agent Systems*. Hatfield, UK: AISB.
- ANDRIGHETTO, G et al. (2007) On the Immersion of Norms: A Normative Agent Architecture. In *Proceedings of the AAI Symposium, Social and Organizational Aspects of Intelligence*. Washington DC.
- ANDRIGHETTO, G and Conte, R (2009) On Norm Internalization. <http://www.agreement-technologies.eu/wp-content/uploads/2009/10/internalization.pdf>.
- ARTIKIS, A and Pitt, J (2009) Specifying Open Agent Systems: A Survey. In A. Artikis, G. Picard, and L. Vercouter, eds. *Engineering Societies in the Agents World IX*. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, pp. 29-45.
- AXELROD, R., 1986. An Evolutionary Approach to Norms. *American Political Science Review*, 80(4), pp.1085-1111. [doi:10.2307/1960858]
- AXELROD, R and Hamilton, W (1981) The Evolution of Cooperation. *Science*, 211 (4489), pp.1390-

1396. [doi:10.1126/science.7466396]

BENDOR, J and Swistak, P (2001) The Evolution of Norms. *The American Journal of Sociology*, 106 (6), pp.1493-1545. [doi:10.1086/321298]

BLAKE, T (2009) A Taxonomy for Ensuring Institutional Compliance in Utility Computing. In *Dagstuhl Seminar Proceedings 09121 - Normative Multi-Agent Systems*. Dagstuhl, Germany.

BOELLA, G. and Torre, L.V.D., 2007. A Game-Theoretic Approach to Normative Multi-Agent Systems. In *Dagstuhl Seminar Proceedings 07122 - Normative Multi-Agent Systems*. Dagstuhl: LZI.

BOELLA, G and Torre, L V D (2006) An architecture of a normative system: counts-as conditionals, obligations and permissions. In *AAMAS '06: Proceedings of the fifth international joint conference on Autonomous agents and multi-agent systems*. Hakodate: ACM, pp. 229 - 231.

BOELLA, G, Torre, L V D and Verhagen, H (2007) Introduction to Normative Multi-agent Systems. In *Dagstuhl Seminar Proceedings 07122*. Dagstuhl: Internationales Begegnungs und Forschungszentrum für Informatik (IBFI).

BOELLA, G, Torre, L V D and Verhagen, H (2008a) Introduction to the special issue on normative multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 17 (1), pp.1-10. [doi:10.1007/s10458-008-9047-8]

BOELLA, G., Torre, L.V.D. and Verhagen, H. (2008b) Ten Challenges for Normative Multi-agent Systems. In R. Bordini et al., eds. *Dagstuhl Seminar Proceedings 08361*. Dagstuhl: Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

BOU, E, López-Sánchez, M and Rodríguez-Aguilar, J (2007) Adaptation of Autonomic Electronic Institutions Through Norms and Institutional Agents. In G. O'Hare et al., eds. *Engineering Societies in the Agents World VII*. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, pp. 300-319.

BOU, E. et al., 2009. Adapting Autonomic Electronic Institutions to Heterogeneous Agent Societies. In G. Vouros et al., eds. *Organized Adaption in Multi-Agent Systems*. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, pp. 18-35.

BOYD, R. and Richerson, P., 1985. *Culture and the Evolutionary process*, Chicago: University of Chicago Press.

BOYD, R and Richerson, P (2002) Group Beneficial Norms Can Spread Rapidly in a Structured Population. *Journal of Theoretical Biology*, 215, pp.287-296. [doi:10.1006/jtbi.2001.2515]

BOYD, R and Richerson, P (2005) *The Origin and Evolution of Cultures*, New York: Oxford University Press.

BROERSEN, J et al. (2002) Goal Generation in the BOLD Architecture. *Cognitive Science Quarterly*, 2 (3-4), pp.428-447.

CALDAS, J C and Coelho, H (1999) The Origin of Institutions: Socio-Economic Processes, Choice, Norms, and Conventions. *Journal of Artificial Societies and Social Simulation*, 2 (2) 1. <http://jasss.soc.surrey.ac.uk/2/2/1.html>.

CAMPBELL, E Q (1964) The Internalization of Moral Norms. *Sociometry*, 27 (4), pp.391-412. [doi:10.2307/2785655]

CAMPENNI, M, Cecconi, G A F and Conte, R (2009) Normal = Normative? The Role of Intelligent Agents in Norm Innovation. In *Dagstuhl Seminar Proceedings 09121 - Normative Multi-Agent Systems*. Dagstuhl, Germany: LZI.

CARVALHO, G et al. (2005) Increasing Software Infrastructure Dependability through a Law Enforcement Approach. In *Proceedings of the Symposium on Normative Multi-Agent Systems*. Hatfield, UK: AISB.

CASTELFRANCHI, C, Conte, R and Paolucci, M (1998) Normative Reputation and the Cost of Compliance. *Journal of Artificial Societies and Social Simulation*, 1 (3) 3. <http://jasss.soc.surrey.ac.uk/1/3/3.html>.

CASTELFRANCHI, C et al. (1999) Deliberative Normative Agents: Principles and Architecture. In *6th International Workshop on Intelligent Agents VI, Agent Theories, Architectures, and Languages (ATAL)*. London: Springer-Verlag, pp. 364 - 378.

CONTE, R et al. (2007) Emergent and Immergent Effects in Complex Social Systems. In

Proceedings of AAAI Symposium, Social and Organizational Aspects of Intelligence. Washington, DC.

CONTE, R and Castelfranchi, C (1995) Understanding the functions of norms in social groups through simulation. In N. Gilbert and R. Conte, eds. *Artificial Societies. The Computer Simulation of Social Life*. London: UCL Press, pp. 252-267.

CONTE, R and Dignum, F (2001) From Social Monitoring to Normative Influence. *Journal of Artificial Societies and Social Simulation*, 4 (2) 7. <http://jasss.soc.surrey.ac.uk/4/2/7.html>.

DAVIDSSON, P and Johansson, S (2005) On the Potential of Norm-Governed Behavior in Different Categories of Artificial Societies. In *Proceedings of the Symposium on Normative Multi-Agent Systems*. Hatfield, UK: AISB.

DECI, E and Ryan, R (2000) The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory*, 11 (4), p.227.

DIGNUM, F (1999) Autonomous Agents with Norms. *Artificial Intelligence and Law*, 7 (1), pp.69-79. [doi:10.1023/A:1008315530323]

DUANGSUWAN, J and Liu, K (2009) Normative Multi-Agent System for Intelligent Building Control. In *Pacific-Asia Conference on Knowledge Engineering and Software Engineering, 2009. KESE '09*. Shenzhen: IEEE, pp. 197-200.

EHRlich, P and Levin, S (2005) The Evolution of Norms. *PLoS Biology*, 3 (6). <http://www.plosbiology.org/article/info:doi%2F10.1371%2Fjournal.pbio.0030194>.

EPSTEIN, J (2001) Learning to Be Thoughtless: Social Norms and Individual Computation. *Computational Economics*, 18, pp.9-24. [doi:10.1023/A:1013810410243]

ESTEVA, M, Padget, J and Sierra, C (2001) Formalizing a Language for Institutions and Norms. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.24.1941> [Accessed December 13, 2010].

FINNEMORE, M and Sikkink, K (1998) International Norm Dynamics and Political Change. *International Organization*, 52 (4), pp.887-917. [doi:10.1162/002081898550789]

FLENTGE, F, Polani, D and Uthmann, T (2001a) Modelling the Emergence of Possession Norms using Memes. *Journal of Artificial Societies and Social Simulation*, 4 (4) 3. <http://jasss.soc.surrey.ac.uk/4/4/3.html>.

FRANZ, M and Matthews, L (2010) Social Enhancement can Create Adaptive, Arbitrary and Maladaptive Cultural Traditions. *Proceedings of the Royal Society B*. 277 (1698) pp. 3363-3372 [doi:10.1098/rspb.2010.0705]

GALAN, J M and Izquierdo, L R (2005) Appearances Can Be Deceiving: Lessons Learned Re-Implementing Axelrod's 'Evolutionary Approach to Norms'. *Journal of Artificial Societies and Social Simulation*, 8 (3) 2. <http://jasss.soc.surrey.ac.uk/8/3/2.html>.

GIBBS, J P (1965) Norms: The Problem of Definition and Classification. *The American Journal of Sociology*, 70, pp.586 - 594. [doi:10.1086/223933]

GROSSI, D, Aldewereld, H and Vazquez-Salceda, J (2005) Ontological Aspects of the Implementation of Norms in Agent-Based Electronic Institutions. In *Proceedings of the Symposium on Normative Multi-Agent Systems*. Hatfield, UK: AISB.

HAHN, C, Fley, B and Florian, M (2005) A Framework for the Design of Self-Regulation of Open Agent-based Electronic Marketplaces. In *Proceedings of the Symposium on Normative Multi-Agent Systems*. Hatfield, UK: AISB.

HALES, D (2002) Group Reputation Supports Beneficent Norms. *Journal of Artificial Societies and Social Simulation*, 5 (4) 4. <http://jasss.soc.surrey.ac.uk/5/4/4.html>.

HECHTER, M and Opp, K (2001a) Introduction. In M. Hechter and K. Opp, eds. *Social Norms*. New York: Russell Sage Foundation, pp. xii - xiii.

HECHTER, M and Opp, K (2001b) What Have we Learned about the Emergence of Social Norms? In M. Hechter and K. Opp, eds. *Social Norms*. New York: Russell Sage Foundation, pp. 394-415.

HENDERSON, D (2005) Norms, Invariance, and Explanatory Relevance. *Philosophy of the Social Sciences*, 35 (3), pp.324-338. [doi:10.1177/0048393105277989]

- HENRICH, J, Boyd, R and Richerson, P (2008) Five Misunderstandings about Cultural Evolution. *Human Nature*, 19, pp.119-137. [doi:10.1007/s12110-008-9037-1]
- HOFFMANN, M (2005) Self-Organized Criticality and Norm Avalanches. In *Proceedings of the Symposium on Normative Multi-Agent Systems*. Hatfield, UK: AISB.
- HORLING, B and Lesser, V (2005) A survey of multi-agent organizational paradigms. *The Knowledge Engineering Review*, 19 (04), p.281. [doi:10.1017/S0269888905000317]
- HORNE, C (2001) Sociological Perspectives on the Emergence of Norms. In M. Hechter and K. Opp, eds. *Social Norms*. New York, New York, USA: Russell Sage Foundation, pp. 3-34.
- KAMARA, L, Pitt, J and Sergot, M (2005) Towards Norm-Governed Self-Organizing Networks. In *Proceedings of the Symposium on Normative Multi-Agent Systems*. Hatfield, UK: AISB.
- LOPEZ, F, Luck, M and d'Inverno, M (2005) A Normative Framework for Agent-Based Systems. In *Proceedings of the Symposium on Normative Multi-Agent Systems*. Hatfield, UK: AISB.
- LOPEZ-SANCHEZ, M, Campos, J and Esteva, M (2009) Empirical Norm Adaptation. <http://www.agreement-technologies.eu/wp-content/uploads/2009/10/empirical-norm-adaptation.pdf>.
- MENEGUZZI, F and Luck, M (2009) Norm-based behaviour modification in BDI agents. In *Proceedings of The 8th International Conference on Autonomous Agents and Multi-agent Systems - Volume 1*. Budapest: International Foundation for Autonomous Agents and Multi-agent Systems, pp. 177-184.
- MERRIAM-WEBSTER (2010) norm. *Merriam-Webster Online Dictionary*. <http://www.merriam-webster.com/dictionary/norm>.
- MEYER, J J and Wieringa, R (1993) *Deontic Logic in Computer Science: Normative System Specification*, Chichester, UK: Wiley.
- MOHAMED, A M and Huhns, M N (2001) Multi-agent Benevolence as a Societal Norm. In R. Conte and C. Dellarocas, eds. *Social Order in Multi-agent Systems*. Boston: Kluwer Academic Publishers, pp. 65-84.
- MUKHERJEE, P, Sen, S and Airiau, S (2007) Emergence of Norms with Biased Interactions in Heterogeneous Agent Societies. In *Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Workshops*. Washington, DC: IEEE Computer Society, pp. 512-515.
- NAKAMARUA, M and Levin, S A (2004) Spread of Two Linked Social Norms on Complex Interaction Networks. *Journal of Theoretical Biology*, 230 (1), pp.57-64. [doi:10.1016/j.jtbi.2004.04.028]
- NEUMANN, M (2008) Homo Socionicus: a Case Study of Simulation Models of Norms. *Journal of Artificial Societies and Social Simulation*, 11 (4) 6. <http://jasss.soc.surrey.ac.uk/11/4/6.html>.
- NEUMANN, M (2010) Norm Internalisation in Human and Artificial Intelligence. *Journal of Artificial Societies and Social Simulation*, 13 (1) 12. <http://jasss.soc.surrey.ac.uk/13/1/12.html>.
- NEWMAN, M (2010) *Networks: An Introduction* 1st ed., Oxford University Press, USA.
- NISHIZAKI, I et al. (2008) Agent-based simulation analysis for social norms. In *IEEE International Conference on Systems, Man and Cybernetics, 2008. SMC 2008*. Singapore, pp. 2105-2110.
- RYAN, R and Deci, E (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist*, 55 (1), pp.68-78. [doi:10.1037/0003-066X.55.1.68]
- SAAM, N and Harrier, A (1999) Simulating Norms, Social Inequality, and Functional Change in Artificial Societies. *Journal of Artificial Societies and Social Simulation*, 2 (1) 2. <http://jasss.soc.surrey.ac.uk/2/1/2.html>.
- SADRI, F, Toni, F and Stathis, K (2005) Normative KGP Agents: A Preliminary Report. In *Proceedings of the Symposium on Normative Multi-Agent Systems*. Hatfield, UK: AISB.
- SAURO, L (2005) Formalizing Coalition Structures via Obligations. In *Proceedings of the Symposium on Normative Multi-Agent Systems*. Hatfield, UK: AISB.
- SAVARIMUTHU, R and Cranefield, S (2009) A Categorization of Simulation Works on Norms. In G. Boella et al., eds. *Dagstuhl Seminar Proceedings 09121 - Normative Multi-Agent Systems*. Dagstuhl:

LZI.

SAVARIMUTHU, R, Purvis, M and Purvis, M (2008) Social norm emergence in virtual agent societies (Short Paper). In *Proceedings of the 7th international joint conference on Autonomous agents and multi-agent systems - Volume 3*. Dunedin: International Foundation for Autonomous Agents and Multi-agent Systems, pp. 1521-1524.

SCHELLING, T (1978) *Micromotives and Macrobehavior*, New York: W. W. Norton and Company, Inc.

SCHEVE, C V et al. (2005) My Agents Love to Conform: Emotions, Norms, and Social Control in Natural and Artificial Societies. In *Proceedings of the Symposium on Normative Multi-Agent Systems*. Hatfield, UK: AISB.

SHOHAM, Y and Tennenholtz, M (1992) On the Synthesis of Useful Social Laws for Artificial Agent Societies (preliminary report). In *Proceedings Of The Tenth National Conference On Artificial Intelligence*, AAAI-92, San Jose, California.

STALLER, A and Petta, P (2001) Introducing Emotions into the Computational Study of Social Norms: A First Evaluation. *Journal of Artificial Societies and Social Simulation*, 4 (1) 2. <http://jasss.soc.surrey.ac.uk/4/1/2.html>.

STANLEY, K O (2010) Novelty Search Users Page. <http://eplex.cs.ucf.edu/noveltysearch/userspage/>.

STAUFFER, D and Aharony, A (1994) *Introduction To Percolation Theory* 2nd ed., CRC Press.

THERBORN, G (2002) Back to Norms! On the Scope and Dynamics of Norms and Normative Action. *Current Sociology*, 50, pp.863-880. [doi:10.1177/0011392102050006006]

TUOMELA, R (1995) *The Importance of Us: A Philosophical Study of Basic Social Norms*, Stanford: Stanford University Press.

VERHAGEN, H (2000) *Norm Autonomous Agents*, Stockholm: Department of System and Computer Sciences, The Royal Institute of Technology and Stockholm University.

VERHAGEN, H (2007) Normtypologies. In *Dagstuhl Seminar Proceedings 07122 - Normative Multi-agent Systems*. Dagstuhl, Germany: LZI.

VERHAGEN, H (2001) Simulation of the Learning of Norms. *Social Science Computer Review*, 19 (3), pp.296-306. [doi:10.1177/089443930101900305]

VILLATA, S (2010) A Normative Multi-agent Approach to Requirements Engineering. *Logic Journal of the IGPL - Special Issue: Normative Multi-agent Systems*, 18 (1), pp.245-274.

WRIGHT, G H V (1951) Deontic Logic. *Mind*, 60 (237), pp.1-15.

WU, J (2008) Contract Net Protocol for Coordination in Multi-Agent System. In *Second International Symposium on Intelligent Information Technology Application*. pp. 1052-1058.

YOUNG, H P (2008) Social Norms. In S. N. Durlauf and L. E. Blume, eds. *The New Palgrave Dictionary of Economics*. New York: Palgrave Macmillan.

YOUNGER, S (2004) Reciprocity, Normative Reputation, and the Development of Mutual Obligation in Gift-Giving Societies. *Journal of Artificial Societies and Social Simulation*, 7 (1) 5. <http://jasss.soc.surrey.ac.uk/7/1/5.html>.

ZHANG, Y and Leezer, J (2009) Emergence of Social Norms in Complex Networks. In *International Conference on Computational Science and Engineering, 2009. CSE '09*. Vancouver, pp. 549-555.