Jurnal Teknologi

Full Paper

A NORM TRUST FRAMEWORK FOR NORMS ADOPTION IN NORMATIVE MULTI-AGENT SYSTEMS

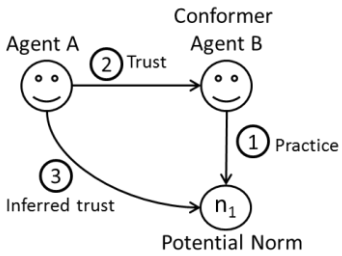
Nurzeatul Hamimah Abdul Hamid[a*], Mohd Sharifuddin Ahmad[a], Azhana Ahmad[a], Aida Mustapha[b]

[a]Center for Agent Technology, College of Computer Science and Information Technology, Universiti Tenaga Nasional, Putrajaya, Malaysia
[b]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn, Batu Pahat, Malaysia

Article history
Received
15 July 2015
Received in revised form
1 October 2015
Accepted
25 October 2015

*Corresponding author
nurzeatul@tmsk.uitm.edu.my

Graphical abstract



Abstract

Researchers in normative multi-agent systems have emphasized the importance of equipping agents with the ability to detect and learn the norms of a new environment. They propose active learning approaches and prove that agents are capable of detecting norms using these approaches. However, most of their works entail agents that detect one norm in an event. We argue that these approaches do not help agents to decide in cases of norms coexistence is detected in an event. To solve this problem, we introduce the concept of norms trust to help agents decide which detected norms are credible in a new environment. In this paper, we propose a conceptual norms trust framework by inferring norms trust through two-tier assessment; credible agent evaluation and norms trust assessment. Norms trust assessment is based on filter factors of norm adoption ratio, norm adoption risk, and norms salience. The framework assesses norms trust value for each detected norm. This value is then used by the agent to decide either to only emulate or fully internalize the detected norms.

Keywords: Normative multi-agent systems; trust; norm life cycle; norms' trust, norm adoption

Abstrak

Kajian di dalam sistem berbilang agen normatif menekankan kepentingan bagi melengkapi agen-agen dengan keupayaan untuk mengesan dan mempelajari norma-norma dalam persekitaran yang baharu. Kajian tersebut adalah berasaskan kaedah pembelajaran aktif, bila mana pembelajaran dicapai berasaskan perbuatan, pemerhatian dan komunikasi. Walau bagaimanapun, kebanyakan kajian tersebut hanya menumpukan kepada agen yang mengesan satu norma di dalam sesuatu peristiwa. Kami mempertikaikan bahawa kaedah tersebut tidak dapat membantu agen-agen membuat keputusan sekiranya terdapat lebih daripada satu set norma yang dikesan di dalam sesuatu peristiwa. Bagi menyelesaikan masalah ini, kami memperkenalkan konsep kebolehpercayaan norma bagi membantu agen-agen memilih norma yang boleh dipercayai setelah dikesan di dalam sesuatu persekitaran yang baharu. Dalam penyelidikan kami, kami mencadangkan satu kerangka konsep bagi kebolehpercayaan norma melalui penilaian dua peringkat; penilaian ejen yang boleh dipercayai dan penilaian norma berdasarkan faktor turas nisbah penerimaan norma, risiko penerimaan norma dan kepentingan norma di dalam masyarakat. Nilai ini kemudiannya akan digunakan oleh agen tersebut bagi membuat keputusan sama ada akan menggayakan sahaja atau menerima sepenuhnya norma baharu yang telah dikesan.

Kata kunci: Sistem agen berbilang normatif, kebolehpercayaan, kitar hayat norma, kebolehpercayaan norma, adopsi norma.

1.0 INTRODUCTION

In multi-agent systems' environments, agents interact to solve tasks to achieve their goals. The agents' goal-directed behaviors occasionally create conflicts or inconsistencies [1]. For example, agents are deployed to engage with other agents or humans to provide online web services such as information retrieval. As they arrive from various hosts, there is a high probability of conflicts and inconsistencies that could occur. Researchers use a few strategies such as norms and partial global planning to overcome these inconsistencies. In the early phase of norms research, researchers designed norms as soft-constraints that are off-line programmed into agents. In such systems, agents are considered as benevolent with less autonomy and assumed to work on common goals [1], [2]. However, these assumptions become a limitation in dynamic open systems and do not reflect the social reality [3]. In open systems' environments, agents are owned by different organizations with individual access to different resources. They are also assumed to be self-interested which bear different beliefs, goals, and preferences. Moreover, the open systems' heterogeneous characteristics and agents' autonomy transform multi-agent societies to become highly dynamic. The restriction of the number of members in an agent society depends on the level of openness of the system (dynamic, static or off-line) [4].

The shift of research in normative multi-agent systems (NMAS) to open systems settings poses new challenges. The first challenge is to confirm and satisfy normative goals in multiple societies. As agents in open systems belong to different societies, they have different architectures, beliefs and goals [5]. Even though the best strategy is to maximize their individual utility, the needs to satisfy society's normative goal are of utmost priority.

The second challenge is heterogeneity and conflicts. In open systems, agents can freely join and leave a society. As they have multiple societies' memberships, it increases the possibilities of potential conflicts such as contradicting norms and goals [5]. The third challenge is the norm dynamics. In open systems, norms are dynamic entities. Norms can emerge as new interpretations of existing norms [6]. Agents then encounter variations of norms for events or tasks. Thus, they have to reconcile between prioritizing personal goals and conflicting norms.

Finally, the fourth challenge is the agents' cognitive limitation. Agents have limitations, including the amount of data, information, and knowledge they can handle [7]. Thus, it is impossible for agents to acquire all the society's information including the norms that local agents comply with due to the high cost of obtaining such a global view [7, 8]. On the other hand, open systems implement indirect sanction, which progressively affects the agent's reputation and emotion [9]. This means any non-compliance behavior may negatively influence other agents from complying with such norms. Agents need to be able to infer all credible norms that are perceived as beneficial for them to avoid such adverse effect of failure to comply with them as well as to achieve their goals [8]. Therefore, the agents need a decision making mechanism to infer the correct norms using the available information.

To overcome these challenges, researchers propose algorithms that enable agents to observe and detect norms in a new society [10, 11, 12, 13]. However, these algorithms assume one set of norms enacted by local agents [10, 11, 12]. Similarly, there is a limited study that investigates norm coexistence in an environment [13].

Consequently, the motivation of this research is to conceive a framework for agents to have an internal decision-making mechanism that enables them to decide which norms to comply with and to adopt when they detect multiple norms for an event or task in a new society. This mechanism facilitates and guarantees that the decision to comply with the norm preserves their personal goals. To be able to do this, agents need sufficient information about the new society. Nonetheless, agents face a significant degree of uncertainties in making decisions because of their cognitive limitations [14].

Hence the objectives of this paper are (i) to identify the main motive of norm adoption behavior in agents' and humans' societies, (ii) to propose a norms trust adoption framework.

We structure this paper into five main sections. Section 2 discusses the norms' concept in normative multi-agent systems, related normative processes, norm adoption motives and its strategies, and related works. We present an example scenario of a variation of prohibition norm of double parking and discuss the relation of trust and norms in Section 3. Section 4 proposes the norms trust framework and Section 5 concludes this paper.

2.0 RELATED WORKS

This section discusses the important concepts related to the study. We begin with the main definition of norms in human and multi-agent society, norm enforcement, norm adoption and norm internalization. We then present the norm adoption motives, norm adoption strategies, and related studies.

2.1 Norms in Normative Multi-Agent Systems

Social norms guides humans by specifying which behavior is permitted, prohibited and obligated in the absent of a central authority [3]. The term, 'norms' refer to an established expected pattern of behavior. According to the Merriam-Webster Online Dictionary [15], there are three definitions of norms:

1. *Standards of proper or acceptable behavior*

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 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.*

The second definition describes how norms are bound with a society. It implies the expected pattern of behavior from the society.

Similarly, in a normative multi-agent context, norms are informal rules or soft constraints that permit, prohibit and oblige agents' behaviors according to their roles and responsibilities [2, 16, 17]. In [12], Andrighetto, Villatoro, and Conte view norms as behaviors which spread in a society that formalize a set of normative goals and beliefs. Normative beliefs refer to collective beliefs of the kinds of behaviors that are permitted, obliged or prohibited in a given social context. Normative beliefs relate to the normative goals as when an agent adopts a norm, the normative beliefs will be part of its normative goals. Bicchieri [18] further describes this concept into empirical expectations and normative expectation. The empirical expectation is the expectation that others will do the same action in the same situation. Whereas normative expectation is the 'belief' that others expect them to do the same action in the same social context.

Another important aspect of norms is that they are shared among members of a society. Norm enforcement through punishment and reward ensures that members of the society comply with the regulated norms [2, 9]. Norms are shared through social interactions among agents as well as their environments through the socialization process [9]. The dynamic factors such as changing in beliefs and trust cause norms to grow or decay. In agent domain, Lopez [19] visualizes transitions between normative processes as in Figure 1.

A norm is activated when an agent adopts it. At this point, a compliance act generates reward and a non-compliance causes the agent to be penalized. A norm is considered as stable when a majority or the whole population of the society adopts it. Over time, this norm may be abolished and replaced by new norms, evolve into new norms or promoted by the society as law. Upon an agent's acceptance and adoption, an instance of the norm is created.

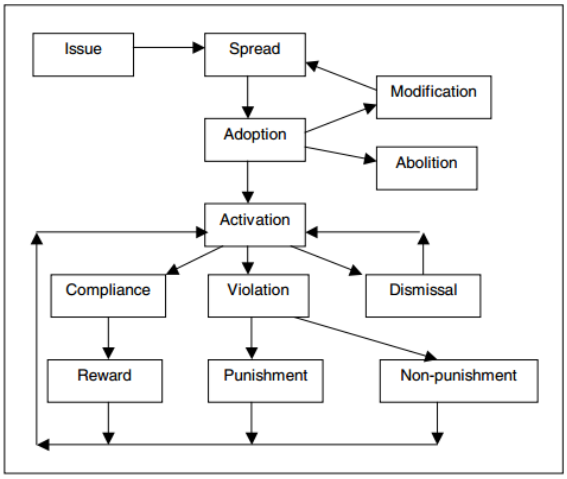


Figure 1 Norms dynamics adapted from Lopez [19]

2.2 Norm Enforcement

Norm enforcement refers to the process of influencing or forcing an agent to comply with a set of norms [2, 9]. A sanction or punishment is applied to an agent if it violates the specified norm, or a reward is otherwise granted. Norm enforcement is an important process that indicates the salience of norms in a particular society. The level of the enforcement mechanism in normative multi-agent systems relates to the different level of interactions among agents. An effective enforcement mechanism shows a strong signal to new agents that norms are applied in the society [2]. However, before norms can be enforced, it must be autonomously adopted by agents.

2.3 Norm Adoption And Internalization

Norm adoption and norm internalization have various definitions in the literature. Norm adoption is the act of conforming to a norm [20]. Based on Lopez [19], norm adoption is the process in which agents are aware of their responsibilities towards another agent by internalizing relevant norms that specify these responsibilities [19]. This process is a part of the normative reasoning process that consists of adoption, deliberation, and compliance processes.

On the other hand, Andrighetto et al. [12] define norm internalization as a process of taking a norm into its mental representation. Norm internalization requires a deeper understanding of an agent. When the agent internalizes a norm, it fits the norm into its reasoning structure causing what is referred by Epstein [21]as a "mindless conformity". At this stage, reward and penalty no longer become the main motivation for the agent to comply with that norm. The agent no longer needs any external mechanisms of enforcement [12], [13].

For the purpose of this work, we define norm adoption as a normative process which an agent conforms to norms that exist in a society or environment. In this process, the agent chooses to

accept responsibilities specified by the norm. Based on [19], the conditions for adoption are as follows:

i. the agent accepts that it is the target (or addressee) of the norm;
ii. it has not adopted the norm before;
iii. the norm is issued by an agent with greater authority.

Concerning norm adoption condition, we further explore the motives that drive norm adoption. One of the first analyses of norm adoption motives in normative multi-agent systems outlines agents choose to follow norms for several reasons [16, 20]:

i. Instrumental reason - The agent believes that by adopting the normative goal as a result of adopting the norm; it will receive something in exchange (i.e. reward or avoid sanction or punishment);
ii. Cooperative reason - In the case of the agent believes by adopting the normative goal, it is capable of achieving a common goal. In other words, the norms ultimately help it to achieve its goal;
iii. Terminal reason - To achieve terminal goals, such as altruism or friendship. No material value in return, but for self-satisfaction or benefits.

These observations are similar to findings by Therborn [22]. He suggests that humans follow norms for the instrumental reason, such as for the sake of reward or fear of sanction (the cost of violations) and personal satisfaction (i.e. altruism). Based on our literature survey, the list of norm adoption motivations in various domains is presented in Table 1.

Table 1 The motivations for norm adoption

Motivations	Example of works	Domain
Detected norms are similar to internalized routines	Therborn [22]	Sociology (Human)
	Opp[23]	
New norms are consistent with internalized norms	Bicchieri [18]	Philosophy (Human)
	Agotnes and Wooldridge [24]	MAS (Agent)
	Andrighetto et al. [12]	MAS (Agent)
Instrumental reasons (punishment and reward)	Andrighetto et al. [12], [16]	MAS (Agent)
	Castelfranchi[20]	MAS (Agent)
	Criado et al. [25, 26]	MAS (Agent)
	Therborn [22]	Sociology (Human)
	Opp[23]	
Self-enhancing effects	Andrighetto et al. [12], [16]	MAS (Agent)
Cost of adoption	Agotnes and Wooldridge [24]	MAS (Agent)
Rational knowledge (Cooperation)	Ahmad et al. [27]	MAS (Agent)
	Therborn [22]	Sociology (Human)
Cost saving	Epstein [21]	Economics, Agent Based Model
Norms' source and value	Therborn [22]	Sociology (Human)
Terminal reason	Andrighetto et al. [12], [16]	MAS (Agent)
	Castelfranchi[20]	MAS (Agent)
	Bicchieri [18]	Philosophy (Human)
	Therborn [22]	Sociology (Human)
Norm's salience	Andrighetto et al. [12], [16]	MAS (Agent)
	Bicchieri [18]	Philosophy (Human)
	Savarimuthu et al. [11]	MAS (Agent)

Although in the context of agents compliance behaviors, researchers group three main reasons (instrumental, cooperative and terminal) for norm adoption, we explore more specific motives such as *new norms are consistent with internalized norms* [24]. Based on this motive, agents are inclined to adopt such norms due to the least modification to its internal structure. We then map the motives that relate to the trust concept as a foundation for our framework. In the next section, we present some related studies to norm adoption strategies used in multi-agent societies. These strategies utilize the norm adoption motivations to facilitate agent decision-making.

2.4 Norm Adoption Strategies

Lopez [19] proposes three strategies for norm adoption. The first strategy is based on the assumption that an agent does not consider whether the adopted norms have impacts on its goals. It consists of four methods:

i. *automatic* strategy – the agent accepts all norms;
ii. *rebellious* strategy – the agent rejects all norms;
iii. *fearful* strategy – only accept norms that involve sanctions although it may affect its goal;
iv. *Greedy* strategy – adopts all norms with rewards.

Another more complex approach of motivated strategies considers the possible impact the norms would have on the agent's goal. Based on this strategy, the agent considers the norm condition, reward and sanction to maximize its utility. The strategy includes (i) egoist – only accept beneficial norms, (ii) pressure – consider norms' sanction rather than condition and (iii) opportunist – consider norms' reward than a negative impact.

Meanwhile, the social strategy consists of (i) cooperative – adopts all norms that benefit the whole society and (ii) benevolent – adopt all norms that benefit the agents it wants to please.

Criado et al. [25] propose a mixed and pondered strategy. The strategies are the combination of simple and complex strategies based on [19]. In the mixed strategy, an agent only adopts a norm if the benefit of norm compliance is higher than the cost of violation. On the hand, the pondered strategy considers the observed probability of the norm's effects on top of norm's benefit and cost of violation.

We also review some norm internalization models that explore similar motives in Table 1. Andrighetto et al. [12] and Savarimuthu et al. [28] exploit norms' salience as one of the factors for an agent to internalize a norm. In addition to norms salience, Andrighetto et al. [12] also focus on the norm consistency (with internalized norms), self-enhancing effects, urgency and cost saving. Savarimuthu et al. [28] combine norms salience with norms detection algorithm to provide a recommendation to the agent.

So far, we have discussed norms in normative multi-agent systems and the way the shift of environment from closed to open systems setting affects norms adoption process in NMAS. Because an agent can join multiple groups or societies simultaneously, researchers not only have to develop mechanisms to enable to detect norms, but also to evaluate the right norm to adopt. On another note, research in other disciplines found that heterogeneity in human societies indicates a possibility of various norms (set of actions) that are applied to a task or an event. Subsequently, a different set of actions generates new goals for a normative agent to fulfill respectively. Therefore, it raises a need to have a mechanism that can help agents to decide which norms to adopt. In the next sub-section, we present an example of a norm coexistence in a domain. In the later stage of this research, we shall simulate a scenario to observe agents' norm adoption behaviors when a norm coexistence occurs.

2.5 Norms And Trust

We illustrate the following scenario to describe an example of a coexistence of three norms ~~for~~ in a double parking situation. Double parking refers to the situation of parking a vehicle at the side or behind another parked vehicle. In many cities, double parking is considered illegal because it prevents the parked vehicles to depart or causes traffic obstructions. Double parking is one of the examples of prohibition norms. Depending on the type of vehicles, in some cities commercial vehicles are permitted to double park for the purpose of loading or unloading of merchandise or passengers.

Recently, due to the heavy traffic in certain cities (e.g., in Malaysia, Subang Jaya, Puchong Jaya), even though it remains illegal, this prohibition norm has a new interpretation. It is observed that double parking is allowed with the following conditions:

i. The double parkers leave their contact numbers on the vehicles' windscreen. In the case when the blocked vehicle wants to depart; the driver can be contacted immediately.
ii. The parking lots are located in a public area. This norm is not observed in paid or private parking area such as a shopping mall.
iii. It some areas, the drivers do not display–their contact numbers on the windscreen but leave their cars without pulling their handbrakes. To ensure safety, they place a stopper at each vehicle's tyre (sometimes bricks are used).

For some individuals, the new re-interpretation of double parking becomes very practical in the case when the driver do not need to drive around to search for a parking lot. However, for others, due to security reasons, they do not trust in displaying their contact numbers or to leave their vehicles without pulling the handbrakes.

Based on this scenario, we find it very interesting to observe the reasons this norm is only applied in certain areas. Why does an individual decides to double park, leave their contact numbers and some do not? Why do individuals trust in double parking and leaving their vehicles without pulling the handbrakes?

2.6 Trust Concepts

Trust is an important element in human interactions. We often rely on trust and reputation to decide whom to interact and cooperate with. The notion of trust is discussed across many fields (i.e. sociology, business, management, computer science) and is defined with a variety of meanings. Trust is also considered as an important mechanism for handling uncertainty in agent-based systems [29]. Trust is mainly used in MAS as a mechanism for an agent to evaluate potential partner and to reason over its trusting behavior. It facilitates the agent to decide how, when and whom to interact with in an environment [14, 30].

According to Castelfrachi and Falcone [31], "*trust is a level of risk accepted by an individual to allow such a transaction to take place*." They highlighted that the trustee is not limited only to cognitive systems or autonomous agents, but also to entities we rely upon such as rules and procedures.

In general, trust models can be classified into two main paradigms. The first paradigm is the numerical models which are based on games theoretical approach. These models do not highlight any explicit representation of cognitive attitudes towards trusting behavior [32, 33, 34]. Josang et al. [33] define reliability trust as "*the subjective probability of a trustor, A, expects the trustee, B, to perform a given action on which its welfare depends*". Meanwhile, a decision trust is "*the extent to which one party is willing to depend on something or somebody in a given situation with a feeling of relative security, even though negative consequences are possible*" [33]. In both definitions, the element of dependence and risk is implicit due to the uncertain positive or negative outcomes.

The second paradigm is the cognitive trust models. Cognitive trust models are based on the socio-cognitive concept that is much more complex than the numerical paradigm. Based on a socio-cognitive trust model [31], trust is describe as a layered notion which consists of three layers; Trust Attitude (TA); Decision to Trust (DtT) and Act of Trusting (AoT).

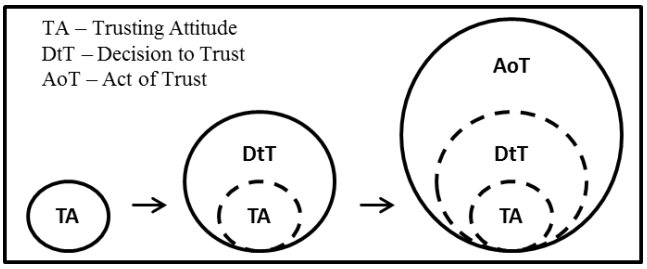


Figure 2 Layers of trust adapted from [31].

As shown in Figure 2, consider a trustor, X and trustee, Y. It starts from a simple layer; Trusting Attitude of X's cognitive mental attitude and disposition of X's belief and expectation towards Y. It moves to the Decision to Trust layer, in which X makes decision and intention based on its disposition towards Y. It then reaches the act of relying on B's expected behavior that initiates the Act of Trust layer.

Consequently, we believe that trust serves as an important element for agents to decide on which norms to comply with. However, there are limited research works that deliberate on trust in norms selection process. In this work, we focus on an agent decision-making process. It facilitates norms' adoption selection when more than one credible norm is detected.

2.7 Discussion

The research and development in normative multi-agent systems have progressed significantly in the last decade. The normative multi-agent environment is gearing towards open systems and implementing social norms close to human settings. In such systems, the decision is made based on norm-aware context, focusing mainly on moral ethics with minimal enforcement mechanism (i.e. sanctions). Consequently, the mechanisms and services in NMAS should be designed to support the new environment settings. Several researchers have proposed algorithms [8, 11, 12] to detect norms. However, these frameworks have two limitations:

i. Firstly, while these algorithms can detect new norms observed by agents in a new environment, the detection process assumes only one set of norms enacted by local agents. In human societies, however, more than one norm is attributable to an event [13, 35].

ii. Secondly, the limitation deals with the verification of the detected norms. Mahmoud et al. [8] perform such verification by only asking the validity of the detected norms to the nearest local agent. By doing so, they assumed that all agents in the environment are trustworthy. In contrast, the heterogeneous criteria of OMAS assume that all agents are not trustable.

To solve these limitations, we propose a new framework that incorporates a norms trust concept. The advantage of adopting the norms' trust is minimizing uncertainty (i.e. better prediction) associated with agents' decision-making. Moreover, a stable state of the norm (i.e. it is practiced by the majority and becoming more sustainable) in NMAS can be achieved.

3.0 A PROPOSED NORMS TRUST FRAMEWORK FOR NORM ADOPTION

We propose a framework that supports agents' decision-making the process by providing the agents with norms' trust information that enables the agents to adopt, emulate or ignore a detected norm. We define a norm's trust as the degree an agent can expect positive outcomes of fulfilling a normative goal with minimal impacts on its goal. The impacts include adoption risks and conflicts with its goal.

The proposed framework comprises two main modules. Figure 3 illustrates the architecture of the proposed framework. Based on an agent's observations in the environment, the norm detection process detects and classifies the set of actions perform by a group of local agents. At this stage, the detected norm is referred to as potential norms. Ultimately, the agent can reason and decide to comply with or even adopt a potential norm.

Before the potential norms' trust values are assessed, the agent evaluates the trust value of a local agent that complies with the potential norms (we refer this agent as a Conformer Agent). From this process, only potential norms that are performed by the conformer agent are assessed in a Norms' Trust Assessment.

In the Norms' Trust Assessment process, we use the norm's salience, adoption risk, and norm adoption ratio to assess a norm's trust value. In the Norm Adoption Willingness Module, the norm's trust value determines the adoption decision and the willingness level.

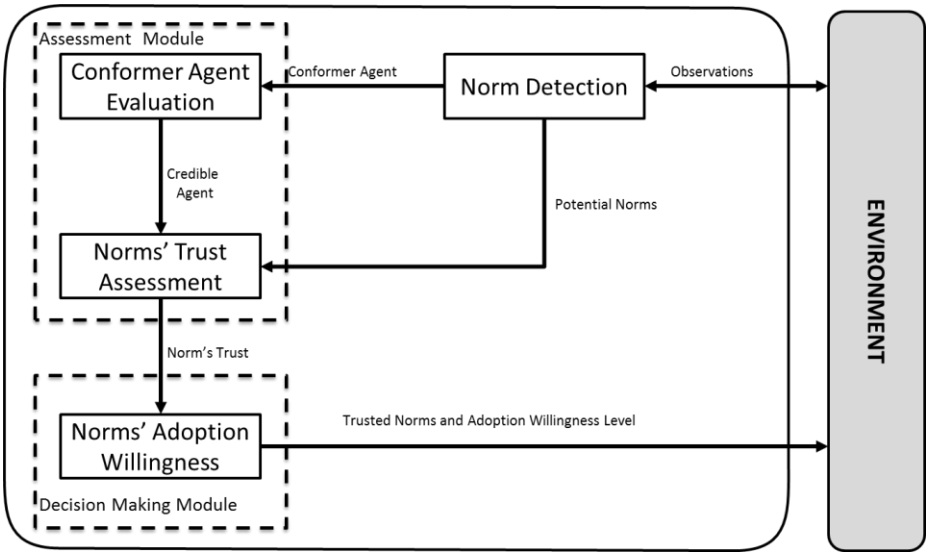


Figure 3 The Proposed Norm's Trust Framework

3.1 The Assessment Module

The assessment module consists of the Conformer Agent Evaluation and the Norms' Trust Assessment process. The Conformer Agent Evaluation is based on the transitive trust concept.

3.1.1 Conformer Agent Evaluation

We introduce the norms trust concept as an inferred trust of an agent which trusts another agent that conforms to a specific normative behavior (in Figure 4 we show this as norm n_1) in a particular event with a specific purpose (e.g. queuing behavior to board a bus). The condition for norms' trust is that an agent A trusts a conformer agent B, which performs a set of behaviors of the norm, n_1.

This trust value serves as a cue that the agent B is a credible agent and agent A can evaluate the potential norm, n_1, trust level. The credibility of the conformer agent is based on its reputation and authoritative level in the society.

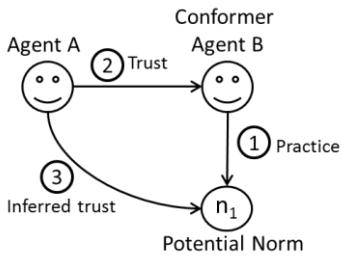


Figure 4 Inferring trust of a conformer agent of potential norms

When an agent joins another society, one of the resources for determining the trust value of a particular norm is by observing authorized bodies. Authorized bodies are trusted because they represent their societies and have the power to reward or sanction societies' members. Consequently, a norm that has been verified by an authorized body has a higher trust value. The verification is justified by a reward to an agent that enacts the norm indicating that the norm is trusted by the authorized body. Moreover, an individual is more likely to adopt norms if he/she identifies the source is credible (i.e. organization's authority, parents) [22, 36].

Reputation is also a type of social evaluation. Conte et al. [6] describe reputation as a meta-belief that acknowledges the existence of a positive evaluation of a person that diffuses within a society. In short, it is a belief about others' evaluation of a person [33]. Reputation is generated from interactions between people. Each member of a society sets reputation values for others based on the experience gained from interactions. In normative multi-agent systems, reputation also builds upon norm compliance [37]. Although punishment is used to enforce norms, the reputation-based mechanism also works in a certain society. In such society, individuals comply with the norms because they are motivated to achieve good reputation [37]. A norm practiced by a reputable agent gives a high possibility of being a credible norm in the society. Moreover, the sources of norms play an important role in influencing other individuals to follow a norm [18, 22]. Thus, we believe that the reputation of a local agent that practices a norm in a new environment influences the norm's trust value. The norm is verified, if it is observed to be enacted by a reputable agent.

3.1.2 Norm's Trust Assessment

In the Norms' Trust Assessment process, norms that are only performed by credible agents are considered. A norms' trust value is derived from the norm's salience, adoption risk, and adoption ratio. Depending on the respective factor, an Evidence Database matches and verifies a norm's trust value.

Adoption ratio is the first filter used to assess a norm's trust. Adoption ratio refers to the ratio of the number of agents that comply with a norm to the number of all agents in a society. In this case, if a majority of the population performs a particular norm, it has a high adoption ratio. The norm is verified, if it is practiced by a significant number of the population that exceeds an adoption threshold. This information can be obtained via two ways. First is via direct interaction (DI), which is gathering information by the agent and directly engaging with another agent in the society. Second is via direct observation (DO), by observing other agents' interactions and behaviors in the society.

The second filter we consider is norm salience. Norm salience indicates the degree of importance of a norm within a social group. It influences the level of norms' adoption willingness. The more salient a norm is, the higher the probability that an agent will adopt the norm [28]. Hence, a majority of the society's population will adopt the norm, N. In such a case, by violating the norm N, third party sanction (sanction enforced by another agent who has adopted the norm) is applicable. We postulate that the evidence of norm salience (next section) influences a norm's trust value.

The third filter is the adoption risk of potential norms. We adapt Castelfranchi's and Falcone's [31] risk concept. In the context of agent A trusts norm, N, an adoption risk is:

i. The probability of not achieving a goal, g_x, if A adopts norm N. Norm N conflicts with goal g_x; From A belief's base or existing norm, N_x.
ii. The probability of norm N is a weak norm or decaying in a society, S. Thus, the effort of adopting N is a wasted commitment.

In short, adoption risk filters assess the possibility of not achieving a goal if a potential norm is adopted.

A norm's trust level is comprised of five levels: Norm Full Trust (NT_F), Norm High Trust (NT_H), Norm Medium Trust (NT_M), Norm Low Trust (NT_L) and Norm Distrust (NT_D) with values ranging ~~of~~ between 0 to 1. A threshold value, ρ, is applied to Norm's Trust. Table 2 shows the norms' trust levels classification and descriptions.

Table 2 The Norms' Trust level and descriptions.

Norm's Trust Levels	Description
Level 1 – Full Trust (NT_F)	A norm is fully trusted when it is practiced by a credible agent and all the three parameters (norm adoption ratio, norm adoption risk and norm salience) each hold a value that jointly produces a high value of norms trust. There is no conflict between the values of the parameters and the agent positively verifies the norm with all factors
Level 2 – High trust (NT_H)	A norm is highly trusted when it is practiced by a credible agent and many of the parameters (norm adoption ratio, norm adoption risk, and norm salience) each holds a value that jointly produces a high value of norm's trust. In this case, the agent positively verifies the norm with many factors exceed the threshold and accepted level with minor conflicts/risks.
Level 3 – Medium trust (NT_M)	A norm is moderately trusted when the agent only observes a majority of credible agents practices the norm and it is not able to positively verify the norm or when there are conflicts between the results of the parameters. However, this decision is based on the Adoption ratio, parameter. If the norm is observed practiced by a majority of agents or exceeds the Adoption ratio threshold, the norm's trust value is medium even though other parameters' values are negative.
Level 4 – Low Trust (NT_L)	A norm is poorly trusted when the agent only observes a minority of credible agents practices the norm, does not exceed the Adoption ratio threshold and it negatively verifies the norm parameters (norm adoption ratio, norm adoption risk and norm salience).
Level 5 – Distrust (NT_D)	A norm is distrusted when no credible agents practice the norm and all the three parameters negatively produce a very low value.

3.2 The Decision Making Module

The norms' trust (NT) value and norm salience level are mapped to determine norm adoption decision and willingness level. The norm decisions are as follows:

i. Adopt - an agent internalizes a norm and adds it to its cognitive structure. The decision to adopt a norm entails all or most factors confirming the norm's trustworthiness.
ii. Emulate - The decision to emulate a norm occurs when there is a conflict between the factors, e.g. a reputable agent confirms that a queuing norm is not trustworthy, but the agent observes that at some particular place the norm is practiced by the majority. In this case, the

agent would only emulate the norm, but not adopt or ignore it. However, once it observes that the norm is performed at the same place/ context in the next occurrence, the norm's trust value increases.

iii. Ignore - The decision to ignore occurs when there is a common negative agreement among the factors in confirming the norm's trust value. In this case, the agent ignores the detected norm.

We show the decision's willingness level definition that depends on the NT values in Table 3. Figure 5 shows the decision's willingness level based on the NT values.

Table 3 Definitions of agents' decision and willingness level

Decision	Willingness	
	Would	***Will***
Adopt	An agent does not fully trust the norm. The agent would adopt a norm if its norm's trust value is between the NT threshold, ρ and 1. It has the choice to emulate it only without adopting. However, the agent, in such case, is motivated by the high value of trust to adopt it. NT_H: $\rho < NT < 1$	An agent has no choice but to adopt the norm because such norm could belong to conventional norms that are commonly practiced in a society. The agent will adopt a norm if its norm's trust value is equal to the highest possible value, 1. NT_F: NT = 1
Emulate	An agent would emulate a norm when the NT value is medium or low. The medium value is the approximate threshold value, ρ and the low value is between the threshold value, ρ and 0. NT_M: NT ≅ρ NT_L: 0 < NT < ρ	An agent will emulate a norm when the NT value is equal to 1; or between the threshold value, ρ and the highest possible value, 1. NT_H: ρ < NT < 1
Ignore	An agent would ignore a norm when the NT value is between the threshold value, ρ and 0: NT_L: $0 < NT < \rho$	An agent will ignore a norm if its norm's trust value is equal to the lowest possible value, 0. NT_D: NT = 0

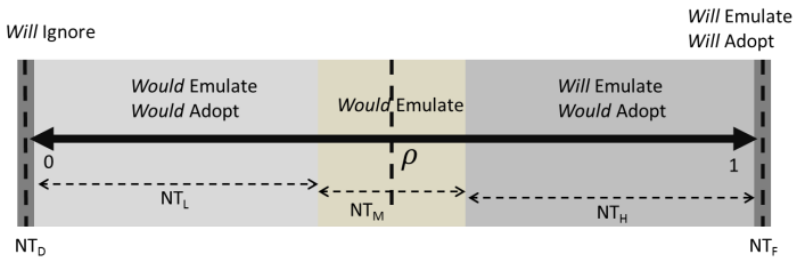


Figure 5 Decision Willingness Level

4.0 CONCLUSION

We explore the challenges pose by open community's settings in relation to norm detection and adoption mechanisms. In this paper, we present our research that establishes the concept of norms' trust to overcome the problems of norm coexistence. In line with researchers in NMAS, we argue that agents' capabilities to adopt correctly detected norms are vital to ensure that their goals are achieved. We exploit the transitive trust concept to infer credible agents as a means to initiate a norm's trust framework. By having a mechanism to compute a norm's trust value, it facilitates a normative agent decision-making process in the norm adoption stage. Once a norm's trust value is computed, it is used to facilitate a normative agent's decision-making process either to adopt, emulate or ignore the detected norms.

In our future work, we shall formalize the norms' trust concept and develop algorithms for each module. Subsequently, we shall validate the framework using agent-based simulation.

References

[1] Wooldridge, M. 2009. *An Introduction Multi-Agent Sytems.* John Wiley & Sons Ltd.

[2] Criado, N., E. Argente and V. Botti. 2011. Open Issues For Normative Multi-Agent Systems. *AI Communications*. 24(3): 233-264.

[3] Boella, G., L. Van Der Torre and H. Verhagen. 2006. Introduction To Normative Multiagent Systems. *Computational & Mathematical Organization Theory*. 12(2-3): 71-79.

[4] Shehory, O. and A. Sturm. 2014. *Multi-agent Systems: A Software Architecture Viewpoint*. *Agent-Oriented Software Engineering*. Springer. Berlin Heidelberg.

[5] Artikis, A., M. Sergot and J. Pitt. 2009. Specifying Norm-Governed Computational Societies. *ACM Trans. Comput. Logic*. 10(1): 1-42.
[6] Conte, R. and M. Paolucci. 2002. *Social Impact of Reputation. Reputation in Artificial Societies*. US. Springer.
[7] Di Marzo Serugendo, G., M.-P. Gleizes and A. Karageorgos. 2011. *Self-organising Systems. Self-organising Software*. Springer. Berlin Heidelberg.
[8] Mahmoud, M. A., A. Mustapha, M. S. Ahmad, A. Ahmad, M. Z. M. Yusoff and N. H. A. Hamid. 2013. *Potential Norms Detection in Social Agent Societies. Distributed Computing and Artificial Intelligence*. Springer International Publishing.
[9] Hollander, C. D. and A. S. Wu. 2011. The Current State of Normative Agent-Based Systems. *Journal of Artificial Societies and Social Simulation*. 14(2): 6.
[10] Mahmoud, M. A., M. S. Ahmad, A. Ahmad, A. Mustapha, M. Z. M. Yusoff and N. H. A. Hamid. 2013. Building Norms-Adaptable Agents from Potential Norms Detection Technique PNDT. *Int. J. Intell. Inf. Technol.* 9(3): 38-60.
[11] Savarimuthu, B. T. R., S. Cranefield, M. A. Purvis and M. K. Purvis. 2013. Identifying Prohibition Norms In Agent Societies. *Artificial Intelligence and Law*. 21(1): 1-46.
[12] Andrighetto, G., D. Villatoro and R. Conte. 2010. Norm Internalization In Artificial Societies. *AI Communications*. 23(4): 325-339.
[13] Villatoro, D., G. Andrighetto, R. Conte and J. Sabater-Mir. 2015. Self-Policing Through Norm Internalization: A Cognitive Solution to the Tragedy of the Digital Commons in Social Networks. *Journal of Artificial Societies and Social Simulation*. 18(2): 2.
[14] Huynh, T. D., N. R. Jennings and N. R. Shadbolt. 2006. An Integrated Trust And Reputation Model For Open Multi-Agent Systems. *Autonomous Agents and Multi-Agent Systems*. 13(2): 119-154.
[15] *Norm. Merriam-Webster.com.* 2015. http://www.merriam-webster.com/dictionary/norm. Accessed on July 6, 2015.
[16] Andrighetto, G., C. Castelfranchi, E. Mayor, J. Mcbreen, M. Lopez-Sanchez and S. Parsons. 2013. *(Social) Norm Dynamics. Normative Multi-Agent Systems*. Dagstuhl, Germany. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
[17] Villatoro, D. 2011. Self-Organization In Decentralized Agent Societies Through Social Norms. *10th International Conference on Autonomous Agents and Multiagent Systems. 3:* 1373-1374. Taipei, Taiwan.
[18] Bicchieri, C. 2006. *The Grammar of Society: The Nature and Dynamic of Social Norms*. New York. Cambridge University Press.
[19] Lopez, F. L. Y. 2003. *Social Power and Norms: Impact on Agent Behaviour*. 241. United Kingdom. University of Southampton.
[20] Castelfranchi, C. 1999. Prescribed Mental Attitudes In Goal-Adoption And Norm-Adoption. *Artificial Intelligence and Law*. 7(1): 37-50.
[21] Epstein, J. M. 2001. Learning to Be Thoughtless: Social Norms and Individual Computation. *Computational Economics*. 18(1): 9-24.
[22] Therborn, G. 2002. Back to Norms! on the Scope and Dynamics of Norms and Normative Action. *Current Sociology*. 50(6): 863-880.
[23] Opp, K.D. 2001. How Do Norms Emerge? An Outline Of A Theory. *Mind & Society*. 2(1): 101-128.
[24] Thomas Ägotnes, W. V. D. H., Michael Wooldridge. 2012. Conservative Social Laws. *242. ECAI 2012*.
[25] Criado, N., E. Argente and V. Botti. 2010. Rational Strategies For Autonomous Norm Adoption. *9th International Workshop on Coordination, Organization, Institutions and Norms in Agent Systems (COIN@ AAMAS 2010)*. Toronto, Canada. 9-16.
[26] Criado, N., E. Argente, P. Noriega and V. Botti. 2013. Human-Inspired Model For Norm Compliance Decision Making. *Information Sciences*. 245(0): 218-239.
[27] Ahmad, A., M. Z. M. Yusof, M. S. Ahmad, M. Ahmed and A. Mustapha. 2011. Resolving Conflicts Between Personal And Normative Goals In Normative Agent Systems. *7th International Conference on Information Technology in Asia 2011 (CITA 11)*. 12-13: 1-6. July 2011.
[28] Savarimuthu, B. T. R., J. Padget and M. A. Purvis. 2013. Social Norm Recommendation for Virtual Agent Societies. *PRIMA 2013: Principles and Practice of Multi-Agent Systems*. Springer. Berlin Heidelberg.
[29] Parsons, S., K. Atkinson, Z. Li, P. Mcburney, E. Sklar, M. Singh, K. Haigh, K. Levitt and J. Rowe. 2014. Argument Schemes For Reasoning About Trust. *Argument & Computation*. 5(2-3): 160-190.
[30] Ramchurn, S. D., D. Huynh and N. R. Jennings. 2004. Trust In Multi-Agent Systems. *The Knowledge Engineering Review*. 19(1): 1-25.
[31] Castelfranchi, C. and R. Falcone. 2010. *Trust Theory: A Socio-Cognitive and Computational Model*. John Wiley & Sons Ltd.
[32] Pinyol, I. and J. Sabater-Mir. 2013. Computational Trust And Reputation Models For Open Multi-Agent Systems: A Review. *Artificial Intelligence Review*. 40(1): 1-25.
[33] Jøsang, A., R. Ismail and C. Boyd. 2007. A Survey Of Trust And Reputation Systems For Online Service Provision. *Decision Support Systems*. 43(2): 618-644.
[34] Mohamed Firdhous, O. G., Suhaidi Hassan. 2014. Robust Multi-Dimensional Trust Computing Mechanism for Cloud Computing. *Jurnal Teknologi*. 69(2): 1-6
[35] Mcdonald, R. I., K. S. Fielding and W. R. Louis. 2013. Energizing and De-Motivating Effects of Norm-Conflict. *Personality and Social Psychology Bulletin*. 39(1): 57-72.
[36] Cialdini, R. B. and M. R. Trost. 1998. *Social Influence: Social Norms, Conformity And Compliance*. The Handbook Of Social Psychology, Vols. 1 And 2 (4th Ed.). New York, NY, US. McGraw-Hill.
[37] Teraji, S. 2013. A Theory Of Norm Compliance: Punishment And Reputation. *The Journal of Socio-Economics*. 44(0): 1-6.