Modelling Normative Awareness: First Considerations

Paul Rauwolf[1] and **Tina Balke**[2] and **Marina De Vos**[3]

Abstract.

As software agents are being employed in more complex situations, experimental findings in the social sciences are becoming increasingly relevant to the computational sciences. The social scientific concept of situation awareness is now being utilized to quantify the success of an agent's environmental perceptual comprehension and causative processing. In this paper, we suggest that awareness of one's situation is not sufficient to succeed in navigating the growing complexity of agent-based social interactions. In human societies norms (personal, legal, and social) have emerged as multi-faceted mechanisms for prescriptive pressures projected onto individual's beliefs and intentionality. Here we define the term normative awareness as the perceptual comprehension of norms and the prediction of the causative effect of actions on norms. In this paper, we suggest that such awareness of the creation and perpetuation of norms would prove advantageous to agent-based research and review to what extent the multi-agent system literature has implicitly utilized the concept of normative awareness. We recognize that a ubiquitous merger of the vernacular between the social and computational sciences is unnecessary, as such, we discuss when and how normative awareness should be extended to agent-based modelling and multi-agent systems.

1 A Case for Modelling Normative Awareness

Situation awareness, which includes perceptual processing, comprehension, and causative predictions [23], is a foundational skill in generating useful human action selection mechanisms. Recently, this concept has been projected onto the study of computational agents (see [42, 30, 31] for example), permitting quantified measurements of awareness, and thus opening a dialogue of the utility therein. However, the instantiation of multiple agents within the computational arena may lead to further complexities than those described in the situational awareness literature. Thus, we define a new term, *normative awareness* as projected onto the situation awareness definition, as the perceptual processing, comprehension, and causative predictions of norms.

The social interactions inherent in multi-agent systems generate normative complexities analogous to those described in the social sciences. Rather than perceiving a situation and processing the potential consequences of an action in isolation, agent's performance is increased through deftly navigating social nuances. Historically, humans have employed norms (personal, social and legal) in manoeuvring the complexities of social interactions. Successful navigation has thus been aided by awareness of the propensity of others to ascribe to such norms, as well as a prediction of the beliefs and intentionality of others. It has even been argued that avoiding sanctions, weeding out defectors, and the emergence of cooperation could be linked to society's tendency to instantiate and navigate norms[35].

We suggest, that as situation awareness is being employed in the agent-based literature, so too should normative awareness. We argue that agent awareness of normative underpinnings are sufficiently unique to situational awareness as to warrant a separate definition. Rather than simply processing environmental data, normative awareness employs a limited version of *Theory of Mind* [7], in that agent's are aware that other agent's possess awareness, intentionality, and beliefs. We argue that this definition transcends that of situation awareness, and that a dialogue surrounding the benefits of normatively aware agents will prove useful in grounding future agent-based research.

To justify this postulation, in Section 2 we first present a brief introduction into the social science and computational literature regarding situation awareness. Next, in Section 3, we define normative awareness in reference to situation awareness. We approach this in a multi-faceted way. In characterizing normative awareness, the semantic ideology of norms is considered, since awareness of norms begs a definition of norms. The definition, utility, creation, and perpetuation of norms (within the social sciences) are thus inspected, and while debate continues, a broad spectrum of arguments are considered. The goal is not to cement a rigid criteria for norms, but rather to discuss the breadth of research in order to (i) augment the grounding of future computational instantiations of norms in theory, and (ii) aid in unifying the vernacular between the two disciplines. Upon exploring norms and situation awareness, we propose a definition of normative awareness, which amalgamates the two concepts. Section 4 reviews the existing agent-based literature regarding normative awareness. We also discuss the gaps in usage between the social sciences and the computational sciences, and discuss whether these gaps need to be closed or not. Finally, in Section 5 we conclude our discussion with a summary and an agenda for future research.

2 Situation Awareness

Situation awareness (SA) has a history of use within military aviation vernacular, dating back to World War I [24]. More recently the term has been utilized within the social [40] and computational sciences [30, 42]. As this paper juxtaposes normative awareness to situation awareness, in this section we will discuss the idea of SA in more detail. We in particular focus on the work of Mica Endsley, who is well known for her work on SA and who defines SA as follows:

> *Situation Awareness* - perception of the elements of the environment within the volume of time and space, the comprehension of their meaning, and the projection of their status in the near future [23].

[1] University of Bath, UK, email: p.rauwolf@bath.ac.uk
[2] CRESS, University of Surrey, UK, email: t.balke@surrey.ac.uk
[3] University of Bath, UK, email: mdv@cs.bath.ac.uk

Based on this definition, Endsley develops a three-layer hierarchical structure which is often referenced when discussing SA [24]:

Level 1: perception of the elements in the environment. This is the identification of the key elements or "events" that, in combination, serve to define the situation. This level tags key elements of the situation semantically for higher levels of abstraction in subsequent processing.

Level 2: comprehension of the current situation. This is the combination of level 1 events into a comprehensive holistic pattern, or tactical situation. This level serves to define the current status in operationally relevant terms in support of rapid decision making and action.

Level 3: projection of future status. This is the projection of the current situation into the future in an attempt to predict the evolution of the tactical situation. This level supports short-term planning and option evaluation when time permits.

Endsley's hierarchical nature of the SA theory has lent itself well to the computational sciences, permitting the awareness of an agent to be discussed in a grounded way. The notion has been utilized in coordinating agents operating within service based systems [42], as well as formally quantifying the awareness of an agent via its ability to complete truth tables in a particular context [30]. Additionally, situation awareness has been employed in the creation of military tactical plans through agent-based modelling [31].

However, as much as increasing SA augments predictive ability for potential actions, it does not offer a holistic theory to guide action selection mechanisms. When one is highly situationally aware, one accurately perceives and comprehends the environmental context. This permits veracious projection of the consequences of a given action. But, what action will be selected? What is one's motive? Social theorist Paul Stern argues that situation awareness is a necessary but not sufficient condition for social movement [39]. He suggests that the motivational impetus to act often portrays itself as a sense of obligation, or a personal norm. Thus, to Stern, social movement requires (i) the ability to predict the future outcomes of actions utilizing situation awareness, and (ii) awareness of which actions and direction one wishes to push society (i.e. awareness of one's personal norms and goals). This suggests that awareness of norms adds an additional layer to interpreting a situation juxtaposed to a solely situationally aware agent. Furthermore it requires an understanding of the link between the SA awareness of the agent and the interpretation of this situation (e.g.action, observations,. . . with respect to the norms of the society.)

This presence of norms as motivational factors in human and multi-agent societies may complicate the notion of awareness. Should an agent be aware of normative societal underpinnings? If the awareness of norms enhances the ability to project future status, does such a notion fall under the banner of situational awareness? In the next section we discuss the definition of norms, and in that attempt to diagnose the utility of diverging the definitions of situational and normative awareness within multi-agent systems.

3 Normative Awareness in the Social Sciences

3.1 Definition of Normative Awareness

If Endsley's situation awareness theory is projected onto norms, then normative awareness is the *(i) perception of norms, (ii) comprehension of norms, and (iii) ability to predict future system states given norms.* However, this definition is unsatisfactory without semantic cohesion. What is a norm? Can awareness of norms be implemented as a subset of situation awareness? Is a norm a situation?

To answer these, first, we will describe the breadth of the social scientific usage of the term "norm". Next, we will briefly discuss theories on the creation and perpetuation of norms. In the end we will present an argument that while Level 1 and 2 situation and normative awareness may prove indistinct, it is only in understanding an agent's effect on the norm that an agent will attain level three awareness, future projection. It is postulated that level three normative awareness requires at least a limited version of theory of mind, in that the agent must predict motivation's and actions of other agent's based on their goals and beliefs.

3.2 Definition of Norm

Before delving into the nuances of normative literature, it may prove useful to reiterate our intention. In discussing social science's utilization of the term norm, our goal is not to reach a conclusion on a definition still debated or to get mired in a semantic argument. Rather, in articulating the breadth of the terminology it may yield the knowledge requisite to deliberate upon the utility of passing aspects of the vernacular into the computational arena. Additionally, in acknowledging the historic debate and precedent, multi-agent systems can ground itself in existing theory.

In a general sense, the social science literature considers norms to be prescriptive and proscriptive [9]. There are actions which ought to be employed and actions which ought not. This pressure may be placed on the self, in which case it is considered a personal norm. In contrast, social norms are rules that are:

> neither promulgated by an official source, such as a court or legislature, nor enforced by the threat of legal sanctions, yet [are] regularly complied with (otherwise it would not be a rule) [36].

Some, however, argue that this notion of the burden of enforcement leads to an even further refined differentiation in the nomenclature. The term "convention" has been employed to describe a Nash equilibrium of a cooperative game [34]. Though there may be multiple equilibria, once convergence reaches a certain threshold, it is rarely in one's interest to defect. For instance, walking on the "incorrect" side of a footpath seldom requires social sanctions as the defective act is cost prohibitive. In general, conventions "...provide people with means of knowing what to expect of each other and thereby serving to coordinate interactions [41]."

Bicchieri [9] argues a social norm is a mechanism which alters a mixed-motive game[4] into a cooperation game. For instance, normative prescriptions and the potential for sanctions might alter the cost/benefit utilities of a context analogous to a prisoner's dilemma game [5] (where cooperation is not a Nash equilibrium[5]) into a situation where cooperation is a Nash equilibrium. Such instances typically require sanctions in order to manipulate the topology of utility function. For example, the utility of attempting to steal is altered depending on the consequences of being caught.

While the differentiation in the governance of the norms has led to a distinction between personal, social, and legal norms, these definitions still blur [9]. Stern argues that personal (rather than social)

[4] A mixed-motive game consists in a game where the best pay-off for at least one of the players does not lead to the best pay-off for the other.

[5] A Nash equilibrium is a state in game theory where, if all other players do not alter their action, it is not beneficial for any one player to alter their action.

norms are required to change the social landscape precisely because the status quo being overthrown typically involves a relatively ubiquitous social norm [40]. The delineation proves a bit more semantic when considering whether a personalized social norm is both a personal and a social norm, or a social norm enacted through an individual.

Additionally, the definition and delineation between legal and social norms presented above are under dispute. Conte and Castelfranchi argue that colloquially it is accepted that behaviour is not sufficient for defining accepted prescriptive pressures. They suggest that just because people throw their rubbish out the window does not suggest that people ought to throw their rubbish out the window [15]. However, if normative prescription does not alter behaviour and sanctions, then how could it alter mixed-motive situations? Furthermore, Fehr and Fischbacher argue that legal norms only exist as a epiphenomena of social norms.

> Legal enforcement mechanisms cannot function unless they are based on a broad consensus about the normative legitimacy of the rules – in other words, unless the rules are backed by social norms. Moreover, the very existence of legal enforcement institutions is itself a product of prior norms about what constitutes appropriate behaviour [27].

This need for broad consensus has even lead to some paradoxical interplay between laws and social norms. In Alabama, a law outlawing adultery was not repealed due to the presumed political difficulty in passing the legislation, despite strong sentiment against enforcement [29]. While a law, is it a legal norm or social norm? If law is not legislatively enforced, should a person or agent care? In this example, was law being employed as a mechanism for disseminating social norms even without the threat of legal sanctions? Is the law there as a referent to the possibility of social sanctions? Is there an expectation or a utility in being aware of the legislation?

3.3 Creation of Norms

In discussing the definition of norms it may help to acknowledge the debate regarding normative generation. How and when are norms created? Is there some tipping point in terms of percentage of the population that generates enough pressure to alter the social landscape? When does societal preference lead to societal pressure?

Bicchieri [8] suggests that the creation of norms are analogous to the formation of language. Namely, she argues that the prescriptive and proscriptive pressures underlying normative interaction are similar to the grammatical structure in language in that neither were the result of human planning, but rather emerged. Conte and Castelfranchi [15] posit that, while norms are spoken of definitively, normative pressure and thus the existence of a norm lies on a continuum consisting of how pervasively people (i) behaviourally conform, (ii) believe they should conform, (iii) are spatially distributed.

Additionally, they discuss the differentiation in the literature between the epiphenomenal and evolutionary generation of norms. While the epiphenomenal explanation relates to the game theoretic conversation, the evolutionary approach argues that norms are generated due to bounded rationality. By implementing prescriptive mechanisms agents can limit the need to diagnose other agent intentionality as well as comprehending the repercussions of complex social interaction [15]. It is even been argued that the advent of social norms offers the advantage of reducing the need to think [25].

As previously mentioned, Fehr and Fischbacher suggest that legal norms are only begotten through the consequences of social norms. However, the interaction between social and legal norm creation are recursive. Scott [38] notes that once a legal norm is established society may adopt it as a social norm even if the legal establishment does not provide it is typical sanctions for defection. For example, a "no smoking" sign may generate social sanctions (e.g. shaming) even if the governance which placed the sign does not police the policy.

3.4 Perpetuation of Norms

Lastly, a definition of norms should consider the perpetuation and declination of norms. When is a norm no longer a norm? Legal positivism argues that the validity of a law's existence need not require general social consent, rather if the authority responsible for legislation pens a new law, it is a law. This is in contrast to the argument that laws may be abrogated due to desuetude (i.e. disuse, or not being enforced) [29]. This ideological disparity becomes relevant when considering multi-agent systems, and whether defecting against a priori norms (even if desuetude) always constitute a violation, and thus advocates awareness. Thus far, there is a propensity in mult-agent systems field to invoke legal positivism, in that defection is always "illegal" [17].

More definitively, a social norm ceases to exist if no one expects anyone else to employ it. Thus, the consequences to a norm given an agent's action depends on the type of norm, and potentially (i.e. in the case of legal norms) the ideology grounding the norm. If an agent defects against a legal norm by the legal positivist definition, the norm remains unaffected. Conversely, if an agent defects against a social norm, the strength of the norm is affected [15].

Additionally, if agent behaviour, at least in part, defines and perpetuates norms, then it is not solely the decision to defect which alters the strength of the norm, but also the agent's propensity to sanction defectors. Even further, an agent's tendency to sanction agent's who refuse to sanction may aid in perpetuating the norm [4].

Thus, at least in the case of personal, social, and legal (given abrogation via desuetude) norms, perpetuation is dependant upon agent belief and, arguably, action. Therefore, awareness of norms is enhanced via awareness of one's and other's beliefs.

3.5 Leading toward Normative Awareness

Having briefly noted the breadth of the literature regarding norms, it is more feasible to discuss the implications of normative awareness. First, it is necessary to decide whether there is any difference between situation and normative awareness. If such a valuable distinction is uncovered, then the utility of normative awareness may be debated.

We suggest that there is little value in differentiating the two forms of awareness at the first two levels. An agent who is level 1 and 2 normatively aware is essentially situationally aware. To perceive and comprehend normative prescriptions is not usefully distinct from perceiving and comprehending environmental situations in that perception and comprehension of norms *is* an environmental situation. However, we believe that the concepts diverge at level three awareness, such that level three normative awareness constitutes level three situation awareness, plus awareness of personal and other agent's motivations and normative restrictions.

As previously mentioned, Endsley's third level of situation awareness elucidates the consequences of actions, but does not consider which action to select. On an individual level, Stern suggests that social movement not only requires situation awareness, but also awareness of one's personal norms [40]. Thus, at a personal level, third

level normative awareness includes not only the ability to predict, but also the knowledge that one wishes to act. In other words, awareness of one's personal norms offers more information than simply the potential consequences of taking an action, it suggests what action will be taken. High situation awareness garners accurate predictions of consequences given an action, while personal normative awareness also posits which actions could be taken.

Furthermore, at a social normative level (or legal normative level sans legal positivism) it has been argued that the perpetuation of a given norm must be dependent, to some extent, on agent behaviour. As such, if one's action effects the perpetuation of a norm, it also effects the creation of a norm, even if the new norm is not perpetuating the old norm (i.e. new norm B is not acting upon or sanctioning old norm A). Thus, for normative awareness to reach level three situational awareness (i.e. projection of future status), then an agent must be aware of other agent's ability to affect norms, and thus there is utility in the cognizance of motivation and goals.

It could potentially still be argued that normative awareness is a subset of situation awareness, in that awareness of the environment includes awareness of other people's motivations, beliefs, and potential for sanctioning. Why confound the terminology when one could potentially extend the definition of situation awareness to include beliefs, etc? We suggest that, even if normative awareness is reduced to a subset of situation awareness, the nuanced and complex effects of beliefs will present unique problems. Even if one remains unconvinced in the semantic delineation between normative and situation awareness, evolution seems to have handled the differentiation through unique neurological processes. Leda Cosmides demonstrated that we are better able to draw logical inferences when the data is set in social rather than mathematical contexts [18]. Analogously, when syntactically instantiating multi-agent models, the processing of beliefs and the prediction of other agent's beliefs are typically unique modules compared to the algorithms employed for situation awareness. In other words, the programmed modules for situation awareness, and processing other agent's beliefs and motivations will likely prove different modules. Thus, even if semantically the concepts can be amalgamated, practically they may be programmed separately, which then creates utility in semantically differentiating the algorithms which will prove conceptually distinct.

3.6 Utility of Normative Awareness

From a utility perspective, advanced normative awareness is likely beneficial. Although, in certain circumstances this can be argued. If legal norms exist, but are desuetude, then awareness may prove deleterious from the standpoint of cognitive load.

Furthermore, in human society lacking normative awareness can sometimes protect one from sanctions. Children, and the mentally disabled are often given a reprieve from the social effects of defection given ignorance. This has raised philosophical debates regarding the norm of sanctioning, including whether psychopaths should be punished if they can not differentiate between conventions and morality [33], or whether one can avoid social and legal sanctions by claiming emotional distress (e.g. temporary insanity)?

While these arguments are potentially rare and nuanced occurrences in human society, philosophically they are useful in discussing normative awareness in the computational arena. Laws, which are always human constructs in the real world, are not always agent constructs in the agent world - they may be designed by humans. If all agents defect from a law, is it useful to be aware of it? Additionally, awareness of one's capacity for awareness has been integrated into a human understanding of norms, but what about within multi-agent systems? If agents are homogeneous in their capability, then perhaps it is a non-issue. But, how should agents be developed when one agent is capable of a deeper normative awareness than another? Should the more advanced agent sanction the other even though the agent will never comprehend the situation? If not, should a capable agent pretend ignorance? Would such a situation ever prove useful?

4 Computation Models of Normative Awareness

Having discussed the concept of norms and normative awareness and having projected it onto the situational awareness levels by Endsley, in this section we now shift our focus to the computational modelling of normative awareness. For this purpose, we start by reviewing the literature on current (computational) models of normative agents[6].

4.1 Literature Review

Turning to the normative agent architectures first, the most prominent ones found in the literature are BOID [11], NoA [32], BRIDGE [20], deliberative normative agents [12], EMIL-A [2] and the NBDI architecture [19, 22].

Of these different frameworks, BOID does consider agents reasoning about norms, but it is assumed that all norms are known to the agents. Normative awareness as such is therefore not considered, but what we refer to as Level 1 and 2 normative awareness is automatically assumed. This is similarly true for NoA and BRIDGE. Although Dignum et al. state that "A person may be aware of a norm ...", but do not explore the issue further.

The idea of deliberate normative agents is based on earlier works in cognitive science (e.g. [15, 16]). Similar to BOID it focuses on the idea that social norms need to be accounted for in the decision making process of an agent.

As a result of the complexity of the tasks associated with social norms, [12] argue that they cannot simply be implicitly represented as constraints or external fixed rules in an agent architecture, but they suggest that norms should also be represented as mental objects, which have their own mental representation [14] and should interact in several ways with the other mental objects (e.g. beliefs and desires) and plans of an agent. Looking at the generation of these mental objects, they result from an internalization process by the agents that is briefly outlined in [12]. For the internalization, when agents are in a social setting, norms are immediately recognized as such (either by observation or communication) and agents can determine to internalize them, i.e. to incorporate them into their own decision making or not (depending on their attitude towards the specific norms and its consequences). This decision making mainly incorporates the ideas of Level 1 and 2 normative awareness, i.e. the agent – as in the previously mentioned architectures – focuses on the question to what extend the norms will affect its own behaviour, but does not necessarily consider its effect on other agents.

The idea of the internalization process described in [12], as well as the actual recognition of norms as such is extended in the EMIL-A architecture [2, 3], which uses a specific norm recognizer module for the latter. This module distinguishes two different scenarios: (i) information it knows about and has classified as a norm before and (ii) new (so far unknown) normative information.

[6] In our review of models of normative agents, except for [6] we neglect models focusing on designing normative frameworks such as the eInstitutions [26, 28], InstAL [13], OperA [21], $\mathcal{MOISE}^{Ins}$ [10],... as these tend to focus on the normative architecture, rather than the agents and their reasoning.

In the former case, i.e. the agent receiving an external normative input it is already aware of, the normative input is entrenched on a so-called normative board (which captures the long-term and working memory of an agent) where it is ordered by salience. Here, salience [1] refers to the degree of activation of a norm, i.e. how often the respective norm has been used by the agent for action decisions or how often it has been invoked. The norms stored on the normative board are then considered in the classical BDI decision process as restrictions on the goals and intentions of the agent. In this process the salience of a norms is important, as in case of conflict (i.e. several norms applying to the same situation), the norm with the highest salience level is chosen.

In case the external normative input is new, i.e. not previously known to the agent, the agent needs to internalize it first before being able to apply it in any of its decision making. For this purpose the normative frame is activated. The normative frame is equipped with a dynamic schema (a frame of reference) which is used to recognize and categorize an external input as being normative based on its properties. Properties that the normative frame takes into account are for example deontic specifications, information about the locus from which the norm emanates or information about legitimate reactions or sanctions to transgression of the norm[7]. The recognition of a norm by an agent does not necessarily imply that the agent will agree with the norm or that it understands it fully, it only means that the agent has classified the new information as a norm. After this initial recognition of the external input as a norm, the normative frame is used to find an interpretation of the new norm. This is done by checking the agent's knowledge for information about variables of the frame of reference for example. Once enough information is gathered about the new norm and the agent is able to determine its meaning and implication, the newly recognized norm is turned into a normative belief. Again, normative beliefs by an agent do not imply that the agent will follow the respective norm, instead it is a candidate for a norm that the agent might adopt. With respect to the adoption of norms, in EMIL-A agents follow a "why not" approach. This means that an agent has a slight preference to adopt a new norm if it cannot find evidence that this new norm conflicts with its existing mental objects. Adopted normative beliefs are stored as normative goals. These normative goals are considered in the agent's decision making. An agent does not need to follow all normative goals it has when making a decision, but can violate norms. When deciding whether to follow a norm, EMIL-A assumes that an agent will try to conform with its normative goals if it does not have reasons for not doing so. As in the previous agent architectures, level 1 and 2 only.

A different approach to the one of EMIL-A concerning the identification of emergent norms (levels 1 and 2) without the norm explicitly being given to the agent is presented by Savarimuthu et al. [37]. The authors present an approach to use data mining mechanisms to identify prohibition and obligation norms, however stop after the norm identification, i.e. they do not account for level 3 either.

One of the few normative architectures which allows for incorporating considerations on how other agents react to norms (i.e. the requirement for level 3 normative awareness) is presented in [6]. In this architecture, using queries, (via an intermediary) the agents can ask about the norms of the system (and in particular the ones applying to them) as well as pose queries about the effects of their own actions or the actions of others with respect to a desired outcome. One example of a query presented in the paper is whether a particular state can be reached depending on the actions of others if the agent itself performs a particular (sequence of) action(s). In contrast to the agent architectures described earlier, the work presented in [6] has its main focus on the normative framework, rather then the internal reasoning of agents.

As a result, the norms the agents deal with are indirectly assumed to be predefined legal norms. Social norms that are emerging in the course of the interaction of the agents are not considered. With respect to obtaining normative information, the agents in the system need to make a conscious decision to ask for normative information, no information is passed on to the agents without a query action initiated by them. The authors point out that the information gathered by their queries can be adopted by the agents in the form of percepts of the environment and then considered in the agent decision making process. This "recognition" of information via the environment is in line with the idea of perceiving a situation (of which norms are a part of) by the agents. However, this does not include any consideration as to which norms an agent internalizes and how it uses the knowledge obtained with the help of the queries in [6]. At present a general methodology or formalisation of the approach presented in [6] is missing.

Inspiration for such an agent architecture might be drawn from the NBDI agent architecture [19, 22]. This architecture includes a Norm Recognition module, which agents can use to either implicitly (via observations) or explicitly (via communication) learn about new norms. The architecture proposes bridge rules for norm internalization and considers different agent behavioural types (with respect to the adoption of norms). Again the architecture is very limited with respect to considering other agent's reactions towards norms, i.e. level 3.

4.2 Gaps and Challenges

Having reviewed existing normative (agent) architectures, it is apparent that at this stage, no model incorporating all three level of normative awareness exists. Thus, although some work has been done on norm emergence in terms of explaining how an agent decides whether it adopts a norm or not, – except for EMIL-A and the architecture described in [6] – most architectures assume that norms are automatically detected (either via observation or communication) and questions on whether something is a norm are not being asked by the agents. Furthermore, the information on norms is typically only reflected on the agents themselves and a lack of consideration of level 3 normative awareness can be found.

As pointed out before, the level 3 normative awareness indirectly implemented in [6] focuses on predefined norms specified at a systems level and thus lacks the incorporation of norms emerging via social interplay. As a consequence no need for normative awareness outside of situation awareness is required in that architecture. This focus on predefined norms is however not uncommon in the (normative) multi-agent literature. Whereas in human societies legal, social, and personal norms are human constructs, often in the computational arena norms are not defined by the agents, but they are designed into the system, which was designed mostly with a particular purpose in mind. As such the question arises on whether the specific accounting of the social norms emerging from social interactions between agents and the emergence thereof does always need to be considered when designing a computational model of normative awareness? Is it sufficient to embed normative awareness in situational awareness if only legal norms defined into a system are considered?

Even if this notion is sufficient, another question arises as a result of our literature review. Looking at the architectures considering

[7] A detailed list of properties being considered by the normative frame can be found in [1].

agents incorporating normative information in their decision making, in these architectures little information was provided on the intentionality of agents. What makes an agent actively try to perceive norms and to incorporate them into its decision-making?

Furthermore, looking at existing models of agent decision making, the models normally assume that individual agents have a fixed set of goals (desires) that they wish to achieve. This, however, makes it hard to account for the high levels of adherence to norms and cooperation found in human societies (e.g. the ultimatum game). We discussed Theory of Minds earlier on, which as one aspect implies that humans have a 'built-in' ability to understand and share the (normative) intentions of others. The incorporation of such a "we-intentionality" in agent reasoning cannot, at the moment, be found in agent architectures. Consequently – from the agents' perspective – an important prerequisite for level 3 normative awareness is missing.

Finally, there are some aspects of the emergence of norms in biological society, which, though potentially not presently relevant, may become more so as agent-based research evolves. For instance, most agents are clones, and thus, are imbued with the same capacity for informational and causative processing as their conspecifics. As such, the aspect of normative underpinnings which derived through the interactions of heterogeneity may be ignored in the computational arena. However, this may not prove a static assumption, and when heterogeneity is employed within multi-agent systems, it may prove advantageous to look to the study of the social sciences in order to ground multi-agent research.

5 Summary and Research Agenda

In this paper we discussed the definition, utility, and difficulty in modelling normative awareness for software agents. Our intention was to initiate a dialogue which considers the utility of agent awareness of the normative infrastructure, and how normative awareness differs from situation awareness. To this end, we started off by giving a definition of situation awareness based on Endsley [24] as well as looking into the concept and related properties of norms. Starting from the individual concepts, we juxtaposed the two ideas to identify where the ideas overlapped and where further considerations were required. We identified that the main difference between situation and normative awareness resided in what we referred to as level 3 normative awareness. Namely, advanced level 3 normative awareness requires a limited instantiation of Theory of Mind, in that, especially in social norms, the norms which pressure an agent's actions are affected by other agent's intentionality and awareness. Thus an awareness that other's have their own beliefs improves an agent's awareness of the norms.

When reviewing existing literature on normative multi-agent systems concerning their incorporation of normative awareness, it became evident that at present not only do most systems assume general awareness of all norms, but that level 3 normative awareness is normally not really considered. The reason for that is the lack of models of (shared) intentionality of agents, which we perceive as the most important step in our research agenda for modelling normative awareness.

A question which was raised in our analysis was whether computational models of normative awareness always need this distinction from situational awareness. Especially in cases in which norms are legal norms respecified in a system and not social norms emerging from social interaction, the perception of norms via the environment as part of the situation might be sufficient. Even if this is sufficient, current models such as [6] pursuing this approach need to adapt their focus from the normative system perspective to a more agent-centered perspective exploring in more detail how and under which circumstances an agent will generate an intention and perform an action to actively query the norms of the system it inhabits. This is therefore the second point on our research agenda: a formalisation of the work presented in [6] as well as the development of a general methodology for the above mentioned processes.

ACKNOWLEDGEMENTS

Tina Balke was partially supported by funding from the ePolicy project (European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 288147).

REFERENCES

[1] Giulia Andrighetto, Marco Campennì, Rosaria Conte, and Mario Paolucci, 'On the immergence of norms: a normative agent architecture', in *AAAI Symposium, Social and Organizational Aspects of Intelligence*, (2007).

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

[3] Giulia Andrighetto, Daniel Villatoro, and Rosaria Conte, 'Norm internalization in artificial societies', *AI Communications*, **23**(4), 325–339, (December 2010).

[4] Robert Axelrod, 'An evolutionary approach to norms', *American political science review*, **80**(04), 1095–1111, (1986).

[5] Robert Axelrod, 'The evolution of strategies in the iterated prisoner's dilemma', *Genetic algorithms and simulated annealing*, **3**, 32–41, (1987).

[6] Tina Balke, Marina De Vos, and Julian Padget, 'Normative run-time reasoning for institutionally-situated bdi agents', in *Coordination, Organizations, Institutions, and Norms in Agent System VII*, eds., Stephen Cranefield, M.Birna Riemsdijk, Javier Vzquez-Salceda, and Pablo Noriega, volume 7254 of *Lecture Notes in Computer Science*, 129–148, Springer Berlin Heidelberg, (2012).

[7] Simon Baron-Cohen, 'Precursors to a theory of mind: Understanding attention in others', in *Natural theories of mind: Evolution, development and simulation of everyday mindreading*, ed., Andrew Whiten, Basil Blackwell, (1991).

[8] Cristina Bicchieri, *The grammar of society: The nature and dynamics of social norms*, Cambridge University Press, 2005.

[9] Cristina Bicchieri and Ryan Muldoon, 'Social norms', in *The Stanford Encyclopedia of Philosophy*, ed., Edward N. Zalta, spring 2011 edn., (2011).

[10] Olivier Boissier and Gâteau Benjamin, 'Normative multi-agent organizations: Modeling, support and control, draft version', in *Normative Multi-agent Systems*, eds., Guido Boella, Leon van der Torre, and Harko Verhagen, number 07122 in Dagstuhl Seminar Proceedings, pp. Internationales Begegnungs– und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany, (2007).

[11] Jan Broersen, Mehdi Dastani, Joris Hulstijn, Zisheng Huang, and Leendert van der Torre, 'The boid architecture: conflicts between beliefs, obligations, intentions and desires', in *Proceedings of the fifth international conference on Autonomous agents*, pp. 9–16. ACM, (2001).

[12] Cristiano Castelfranchi, Frank Dignum, Catholijn M. Jonker, and Jan Treur, 'Deliberate normative agents: Principles and architecture', in *Intelligent Agents VI, Agent Theories, Architectures, and Languages (Proceedings 6th International Workshop, ATAL'99, Orlando FL, USA, July 15-17, 1999)*, eds., Nick R. Jennings and Y. Lespérance, volume 1757 of *Lecture Notes in Computer Science*, pp. 364–378. Springer, (2000).

[13] Owen Cliffe, Marina De Vos, and Julian A. Padget, 'Specifying and analysing agent-based social institutions using answer set programming', in *Coordination, Organizations, Institutions, and Norms in Multi-Agent Systems – AAMAS 2005 International Workshops on Agents, Norms and Institutions for Regulated Multi-Agent Systems, ANIREM 2005, and Organizations in Multi-Agent Systems, OOOP 2005, Utrecht, The Netherlands, July 25-26, 2005, Revised Selected*

Papers, eds., Olivier Boissier, Julian A. Padget, Virginia Dignum, Gabriela Lindemann, Eric T. Matson, Sascha Ossowski, Jaime Simão Sichman, and Javier Vázquez-Salceda, volume 3913 of *LNCS*, pp. 99–113. Springer Berlin / Heidelberg, (2006).

[14] Rosaria Conte and Cristiano Castelfranchi, *Cognitive and Social Action*, Taylor & Francis, 1995.

[15] Rosaria. Conte and Cristiano Castelfranchi, 'From conventions to prescriptions. towards an integrated view of norms', *Artificial Intelligence and Law*, **7**(4), 323–340, (1999).

[16] Rosaria Conte, Cristiano Castelfranchi, and Frank Dignum, 'Autonomous norm acceptance', in *Proceedings of the 5th International Workshop on Intelligent Agents V, Agent Theories, Architectures, and Languages*, pp. 99–112. Springer-Verlag, (1999).

[17] Rosaria Conte, Rino Falcone, and Giovanni Sartor, 'Introduction: Agents and norms: How to fill the gap?', *Artificial Intelligence and Law*, **7**(1), 1–15, (1999).

[18] Leda Cosmides, 'The logic of social exchange: Has natural selection shaped how humans reason? studies with the wason selection task', *Cognition*, **31**(3), 187–276, (1989).

[19] Natalia Criado, Estefania Argente, Pablo Noriega, and Vicente J. Botti, 'Towards a normative BDI architecture for norm compliance', in *Proceedings of The Multi-Agent Logics, Languages, and Organisations Federated Workshops (MALLOW 2010), Lyon, France, August 30 - September 2, 2010*, eds., Olivier Boissier, Amal El Fallah-Seghrouchni, Salima Hassas, and Nicolas Maudet, number 627 in CEUR Workshop Proceedings, pp. 65–81, (2010).

[20] Frank Dignum, Virginia Dignum, and Catholijn M. Jonker, 'Towards agents for policy making', in *Multi-Agent-Based Simulation IX*, eds., Nuno David and Jaime Simão Sichman, volume 5269 of *Lecture Notes in Computer Science*, 141–153, Springer-Verlag, (2009).

[21] Virgínia Dignum, *A model for organizational interaction: based on agents, founded in logic*, Ph.D. dissertation, Utrecht University, 2003.

[22] Baldoino F. dos S. Neto, Viviane Torres da Silva, and Carlos José Pereira de Lucena, 'NBDI: An architecture for goal-oriented normative agents', in *ICAART 2011 - Proceedings of the 3rd International Conference on Agents and Artificial Intelligence, Volume 1 - Artificial Intelligence, Rome, Italy, January 28-30, 2011*, eds., Joaquim Filipe and Ana L. N. Fred, pp. 116–125. SciTePress, (2011).

[23] Mica R. Endsley, 'Measurement of situation awareness in dynamic systems', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, **37**(1), 65–84, (1995).

[24] Mica R. Endsley, 'Toward a theory of situation awareness in dynamic systems', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, **37**(1), 32–64, (1995).

[25] Joshua M. Epstein, 'Learning to be thoughtless: Social norms and individual computation', *Computational Economics*, **18**(1), 9–24, (2001).

[26] Marc Esteva, *Electronic Institutions: from specification to development*, Ph.D. dissertation, Artificial Intelligence Research Institute (IIIA), 2003.

[27] Ernst Fehr and Urs Fischbacher, 'Social norms and human cooperation', *TRENDS IN COGNITIVE SCIENCES*, **8**(4), 185–190, (APR 2004).

[28] Andrés García-Camino, *Normative regulation of open multi-agent systems*, Ph.D. dissertation, Artificial Intelligence Research Institute (IIIA), Spain, 2009.

[29] Hillary Greene, 'Undead laws: The use of historically unenforced criminal statutes in non-criminal litigation', *Yale Law & Policy Review*, **16**(169), (1997).

[30] Anne-Laure Jousselme, Patrick Maupin, Christophe Garion, Laurence Cholvy, and Claire Saurel, 'Situation awareness and ability in coalitions', in *Information Fusion, 2007 10th International Conference on*, pp. 1–9. IEEE, (2007).

[31] Octavio Juarez-Espinosa and Cleotilde Gonzalez, 'Situation awareness of commanders: a cognitive model', (2004).

[32] Martin J. Kollingbaum and Timothy J. Norman, 'Norm adoption in the noa agent architecture', in *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pp. 1038–1039, New York, NY, USA, (2003). ACM.

[33] Neil Levy, 'The responsibility of the psychopath revisited', *Philosophy, Psychiatry, & Psychology*, **14**(2), 129–138, (2007).

[34] David Lewis, 'Convention: A philosophical study', (2002).

[35] Mitsuhiro Nakamura and Naoki Masuda, 'Indirect reciprocity under incomplete observation', *PLoS computational biology*, **7**(7), e1002113, (2011).

[36] Richard A. Posner, 'Social norms and the law: An economic approach', *The American Economic Review*, **87**(2), 365–369, (1997).

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

[38] Robert E. Scott, 'Limits of behavioral theories of law and social norms, the', *Va. L. Rev.*, **86**, 1603, (2000).

[39] Paul C. Stern, 'New environmental theories: toward a coherent theory of environmentally significant behavior', *Journal of social issues*, **56**(3), 407–424, (2002).

[40] Paul C. Stern, Thomas Dietz, Troy Abel, Gregory A. Guagnano, and Linda Kalof, 'A value-belief-norm theory of support for social movements: The case of environmentalism', *Human ecology review*, **6**(2), 81–98, (1999).

[41] Elliot Turiel, *The development of social knowledge: Morality and convention*, Cambridge University Press, 1983.

[42] Stephen S. Yau, Dazhi Huang, Haishan Gong, and Hasan Davulcu, 'Situation-awareness for adaptive coordination in service-based systems', in *Proceedings of the 29th Annual International Computer Software and Applications Conference*, pp. 107–112, (2005).