Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence

An Agent Model for the Appraisal of Normative Events Based in In-Group and Out-Group Relations

Nuno Ferreira, Samuel Mascarenhas, Ana Paiva
INESC-ID & IST, Porto Salvo, Portugal
nuno.raf, samuel.fm@gmail.com, ana.paiva@inesc-id.pt

Gennaro Di Tosto, Frank Dignum
Utrecht University, Utrecht, Netherlands
g.ditosto@uu.nl, dignum@cs.uu.nl

John Mc Breen, Nick Degens, Gert Jan Hofstede
Wageningen University, Wageningen, Netherlands
john.mcbreen, gertjan.hofstede @wur.nl,
dmdegens@gmail.com

Giulia Andrighetto, Rosaria Conte
ISTC-CNR, Rome, Italy
giulia.andrighetto@gmail.com
rosaria.conte@istc.cnr.it

Abstract

Emotional synthetic characters are able to evaluate (appraise) events as positive or negative with their emotional states being triggered by several factors. Currently, the vast majority of models for appraisal in synthetic characters consider factors related to the goals and preferences of the characters. We argue that appraisals that only take into consideration these "personal" factors are incomplete as other more social factors, such as the normative and the social context, including in-group and out-group relations, should be considered as well. Without them, moral emotions such as shame cannot be appraised, limiting the believability of the characters in certain situations. We present a model for the appraisal of characters' actions that evaluates whether actions by in-group and out-group members which conform, or not, to social norms generate different emotions depending on the social relations between the characters. The model was then implemented in an architecture for virtual agents and evaluated with humans. Results suggest that the emotions generated by our model are perceived by the participants, taking into account the social context and that participants experienced very similar emotions, both in type and intensity, to the emotions appraised and generated by the characters.

Introduction

In recent years, there has been a widespread use of virtual environments in many domains. One of the reasons for this success lies in recent technological advances in computer graphics, human-machine interaction, and artificial intelligence, leading to increasing user engagement in those environments.

As the range of applications for virtual environments increases, so does the diversity of the social settings they represent. From educational environments depicting school settings to realistic military applications, the variety of application domains is vast. But for these environments to portray social realities, they must also be populated with characters that act in an intelligent and autonomous manner. Virtual environments depict social realities where intelligent characters are centre stage replicating aspects of real societies. This means that social factors such as groups, cooperation, norms, and social identity, must be studied and embedded in such environments. Hence, for these agents to be as believable as possible, we need not only to endow them with expressive behaviour, through gestures or facial expressions, but also give them social intelligence that will make them act appropriately in the social setting they are embedded in.

Consider for example a scenario where there is a virtual world simulating a small organisation featuring agents that portray workers interacting with each other and following the rules and norms established by that organisation. In this case, for intelligent agents to be believable they must respect and follow the social rules established within the social virtual environment. For example, if one agent worker breaks the established norm and does not say "Good Morning" as he arrives one day, the other agents must react appropriately, for example by frowning. So, to act appropriately in the environment it is essential to consider not only internal mechanisms to drive the agent's behaviours, but, also the social setting the agents are placed in. Therefore, agents emotional expressions must be coherent, not only with the way an event affects the agents personal goals and preferences, but also how it disturbs the simulated social context. Agents emotional reactions, then, not only result from the fact that its goals are satisfied, but also from the actions performed in the social environment such as the violation of an important social norm, even if that action contributed to the success of a personal goal.

In this paper we present a computational model for emotional agents that includes the social setting of the agent in its appraisal of the situation. This model takes into account not only the norms of the society, but also, the in-group and out-group relations of the agents, their social roles and the socially acceptable behaviours prescribed by the social norms that are active in a given context. This model allows the creation of agents that exhibit social, moral emotions, such as shame. This model was then integrated into an architecture for virtual agents and used for the generation of a concrete scenario inspired by the existing no-smoking law in bars and restaurants in many European countries.

The results obtained suggest that the emotions generated by our model are perceived by the participants in a study featuring the generated situations. Further, the results also suggest that the emotions perceived by users who observe the characters' emotional reactions are in agreement with the emotions humans are likely to feel in that social context, both in the type of emotion and in its intensity.

Related Work

The use of emotions and emotional expressions in virtual characters has been a popular research topic in the virtual agents community. Emotional models were developed using emotion theories from psychology (Bates 1992) (Reilly 1996). Some well known emotional architectures include Émile (Gratch 2000) that builds on Clark Elliot's Construal Theory (Elliott 1992) and uses emotion to influence planning. Carmen's Bright IDEAS (CBI) (Marsella, Johnson, and LaBore 2000), which introduces the concept of emotion-focused coping. In Sloman's architecture (Sloman and others 2001) emotions are seen as an attention mechanism that results from the interactions between the system's layers. Other architectures such as ALMA (Gebhard 2005) and ActAffAct (Rank 2004) adopt an appraisal theory, the OCC model (Ortony, Clore, and Collins 1988), as the basis for their emotion synthesis. In contrast, WASABI (Becker-Asano and Wachsmuth 2010; Becker-Asano 2008) base their approach on a dimensional model combining cognitive reasoning capabilities with simulated embodiment to generate emotions. In a similar approach, TABASCO (Staller and Petta 1998) models emotions through an adaptive process related to the agent-environment interactions, based on Scherer's theory of emotions (Scherer 1984). Although all these architectures have been extremely influential in the area of virtual agents, most of them deal with prospect based or preference based emotions, and so far social emotions, in particular resulting from normative reasoning such as shame or guilt, have largely not been addressed by this community.

On the other hand, as the number of multi-agents systems increased in different fields, norms became used to solve the so-called social control problem (Mukherjee, Sen, and Airiau 2007), that is, how to preserve efficiency at the system level while allowing agents to keep their freedom? The idea is to promote cooperation and coordination between agents, not by imposing behaviour but by influencing it with norms, creating normative agents. While there is plenty of research into developing normative agent architectures and frameworks (Broersen et al. 2002) (Castelfranchi et al. 2000) (Dignum 1999) (López, Luck, and d'Inverno 2006) (Meneguzzi and Luck 2009) only a few include the role of emotions. Two of those normative systems are the *culturally affected behaviour* (CAB) (Solomon et al. 2008) and *Thespian* (Si, Marsella, and Pynadath 2006). Both were extended with emotional models (Bulitko et al. 2008) (Si, Marsella, and Pynadath 2010). On the other hand, some emotional models try to model norm-related emotions without a representation of norms, for instance, by casting norm violations as goal violations (Gratch, Mao, and Marsella 2006).

In our research we take a different approach in that we model norms explicitly, and that social emotions emerge from the perceptions and appraisals of the agents concerning the fulfilment and violation of norms. Furthermore, in this paper we explore specifically the impact of the social relationships between the agents together with the salience of the norm.

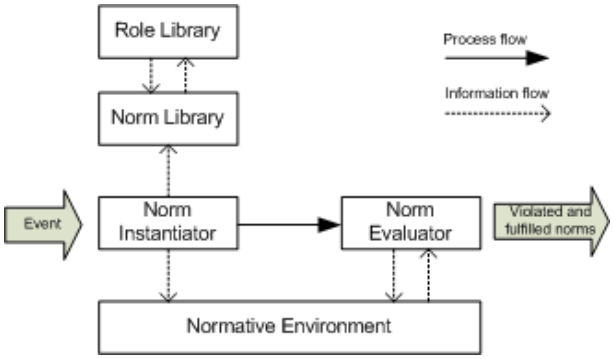

Figure 1: The normative component

A Computational Model for Normative Emotional Agents

Agents that are members of a society are expected to display appropriate social behaviour. They should balance their goals with the prescriptions of social norms in order to be good members of that society. However, their behaviour can sometimes deviate from what is considered correct, either by choice or by necessity. The appraisal of these deviant behaviours can trigger intense negative emotions in all that witness them, especially when important norms are violated. The emotions that result from the appraisal of a deviant action are not just influenced by the importance of the norm, but also by the social relationship between the agent that appraises the action and the agent that performed it.

Our model is focused on the generation of social emotions via the appraisal of actions that conform or deviate from social norms. It has two main components: a normative component which recognizes when an action causes the fulfilment or violation of a norm, and an emotional component which generates emotions by appraising those actions.

The normative component The normative component is responsible for the normative processes within the agent. It includes a "Role Library" where the roles enacted by the agents are defined and a "Norm Library" where the social norms are described. New norms are instantiated in the "Norm Instantiator" by detecting changes in the state of the world and stored in the "Normative Environment" to keep track of their evolution. Finally, a "Norm Evaluator" detects when an action causes the fulfilment or violation of a norm instance.

When an event is perceived, the Role Library is checked to see if any agent should enact or abandon a role and the Norm Library is checked to see if any norm has become active. If so, the Norm Instantiator creates an instance of the norm for every agent to whom the norm applies and stores them in the Normative Environment, marking the norm as

active. Finally, the Norm Evaluator checks if the event corresponds to an action that caused the fulfilment or violation of any norm present in the Normative Environment. If a norm was fulfilled or violated by the action, the Emotional Component begins the appraisal of that event.

The definition and use of roles avoids the specification of all norms for all agents. Roles create a bridge between norms and agents by linking a set of prescribed behaviours (the norms) with a set of agents that enact the role. The Role Library defines which roles can be enacted by the agents and in which contexts. A role R has the form: $\langle R_{act}, R_{exp}, R_{norms} \rangle$. Each role contains a non-empty set of activation conditions R_{act} that, when met by some agent A, will trigger the expectation *"Agent A is now enacting the role R"* in all the agents that believe R_{act} to be true. They also contain a non-empty set of expiration conditions R_{exp} that, when met by the agent A, will negate the previous expectation. Note that for a role to be considered valid, R_{act} and R_{exp} can not be simultaneously true at any moment. Roles are associated with a set of norms R_{norms} that prescribe the expected behaviours of agents enacting that role.

For instance, the role *"Driver"* could be defined as:

R_{act}: on-driver-seat(AGENT, car) = TRUE
R_{exp}: on-driver-seat(AGENT, car) = FALSE
R_{norms}: TRAFFIC-LAW-001, TRAFFIC-LAW-002, ...

When the state of the world matches the activation conditions for a role, the agent enacts it and the norms associated to the role are included in the agent's set of norms. If the state of the world matches the expiration conditions of an enacted role, the norms associated with the role are removed. Both the activation conditions and the expiration conditions of a role can have references to other roles, so when an agent enacts or abandons a role, it may have to enact or abandon other associated roles. Given that in our model, role adoption and expiration is done by the activation of logical conditions that are either true or false, agents are never uncertain about their current role(s).

The set of norms linked to a role should not conflict. This does not mean that they should all prescribe a similar behaviour, but that the agent should clearly know what norm to follow when the behaviours prescribed conflict. Hence, the context where each norm applies should be clearly defined and mutually exclusive. Even when an agent enacts several roles, usually only a small percentage of the norms linked to those roles are active at a given time. This happens because norms are defined with respect to a context, and they are only active when the world is in a state that matches that context. For instance, the role of a Driver shown above is linked with many social norms and traffic laws. One of them is certainly the norm *"'Stop at red traffic lights"'*. This norm demands that drivers stop the vehicle, but only when faced with a red traffic light. Otherwise, the agent is free to keep moving (unless a different norm demands it to stop).

Our norm representation is based in the work presented in (Villatoro et al. 2011), (Castelfranchi et al. 2000), (Dignum 1999), (Ferreira et al. 2012) and (Oren et al. 2009). We plan to include other attributes, such as the enforcers of the norm, in future work, but for now a norm contains the following attributes:

- *ID:* A unique identifier that is used to identify the norm.
- *Name:* A name that describes the norm.
- *Type:* A value that informs if the norm is an obligation or a prohibition.
- *Targets:* The roles to whom the norm applies (when the norm is active).
- *Activation Conditions:* A set of conditions that must be all satisfied for a norm to become active. They may include references to the state of another norm.
- *Expiration Conditions:* A set of conditions that define when a norm is no longer active.
- *Normative Conditions:* The set of conditions that prescribe the behaviour of the targets of the norm. When a norm is active, those agents should satisfy these conditions (if the norm is an obligation) or avoid them (if the norm is a prohibition).
- *Salience:* A value that *"indicates to an individual how operative and relevant a norm is within a group and in a given context"* (Andrighetto and Villatoro 2011). The salience of a norm depends on several contextual, social and individual factors, such as, the frequency, typology and intensity of punishments (Villatoro, Sabater-Mir, and Sen 2011). The more salient a norm is, the greater the expectation that it will be followed.

We now show how can we model the norm *"Stop at red traffic lights"* given as an example above, to help clarify the norm representation used.

```
ID: TRAFFIC-LAW-001
Name: ''Stop at red traffic lights''
Type: OBLIGATION
Targets: Driver
Activation Conditions:
        is-Red(Traffic-Light) = TRUE AND
        in-Front(Traffic-Light, Car) = TRUE
Expiration Conditions:
        is-Red(Traffic-Light) = FALSE OR
        in-Front(Traffic-Light, Car) = FALSE
Normative Conditions:
        is-Stopped(Car) = TRUE
Salience: 0.9 (Very Salient)
```

The norms presented in the norm library are *"'Abstract Norms"'*. A norm is said to be abstract when it does not specify the exact context where it applies. The norm defined above is an example of an abstract norm as it does not specify to which driver, which traffic light, and which car it refers. The norm instantiator must check if there is a set of substitutions for these unbound variables that matches both the current state of the world and the activation conditions of the norm. For each valid set of substitutions, a new *"'Instantiated Norm"'* is created by applying that substitution set to the abstract norm. For instance, if the conditions *"'Bob is a driver"'*, *"'car-1 is Bob's car"'*, *"'traffic-light-35 is a traffic light"'*, *"'traffic-light-35 is red"'* and *"'traffic-light-35 is in front of car-1"'* are all satisfied in the

current state of the world, then the norm *"'Stop at red traffic lights"'* should be instantiated using the set of substitutions [Driver/Bob, Car/car-1, Traffic-Light/traffic-light-35] and added to the Normative Environment. The example below shows this instantiated norm:

```
ID: TRAFFIC-LAW-001-instance-45
Name: ''Stop at red traffic lights''
Type: OBLIGATION
Targets: Bob
Activation Conditions:
        is-Red(traffic-light-35) = TRUE AND
        in-Front(traffic-light-35, car-1) = TRUE
Expiration Conditions:
        is-Red(traffic-light-35) = FALSE OR
        in-Front(traffic-light-35, car-1) = FALSE
Normative Conditions:
        is-Stopped(car-1) = TRUE
Salience: 0.9 (Very Salient)
```

Each agent must monitor the set of instantiated norms that every known agent (including itself) must comply with at any moment. It has a Normative Environment to keep track of the evolution of all instantiated norms that it believes to target, or to have targeted, itself or other agents. Each instantiated norm will be in one of the following states: (a) recently created, (b) to be followed, (c) rejected, (d) recently fulfilled, (e) recently violated or (f) already appraised.

Finally, the norm evaluator constantly checks the normative conditions and the expiration conditions of all instantiated norms in the states "to be followed" and "rejected" to evaluate when an instance is fulfilled or violated. The fulfilment or violation of a norm depends on its type. An obligation is fulfilled as soon as the normative conditions become true and violated if it expires without ever being fulfilled. A prohibition is fulfilled if its normative conditions were always kept false when it expires, and violated when the normative conditions become true.

Emotional Component Our emotional component follows Magda Arnold's ideas (Arnold 1960) that *"emotions such as fear, anger, and excitement could be distinguished by different excitatory phenomena"* and *"the first step in emotion is an appraisal of the situation"*. This component uses the OCC cognitive theory of emotions (named after its creators Ortony, Clore and Collins) (Ortony, Clore, and Collins 1988). The appraisal is made regarding the agent's goals, standards and attitudes. Goals represent states of the world that the agent desires to achieve. Standards refer to social and moral expectations, that can be generated by norms, and attitudes represent the agent's preferences regarding other agents or objects.

We can distinguish two major parts in the emotional component: the appraisal and the affect derivation. The appraisal generates a set of *"appraisal variables"* from the evaluation of the event and stores them in an *"Appraisal Frame"*. The affect derivation then uses this appraisal frame to generate the emotions. We will focus on the appraisals regarding the agent's standards created by the active norms. Ac-

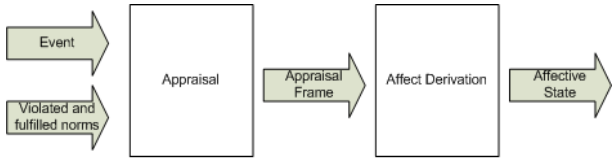

Figure 2: The emotional component

cording to OCC, appraising an event (in this case an action of an agent) regarding the standards of an agent will generate the specific appraisal variables (a) praiseworthiness (and its negative counterpart, blameworthiness), (b) expectation-deviation and (c) cognitive unit strength.

The praiseworthiness/blameworthiness of an action is a measure of how socially acceptable or reprehensible that action was. It is deeply related with the salience of the norm, the cost of the action, the intention and the responsibility of the agent. The expectation-deviation is a measure of how unexpected the action was, and is also related with the salience of the norm, because the higher the salience the more unexpected it is to violate it and the more expected it is to fulfil it. These two variables will be computed using the formulas proposed in (Ferreira et al. 2012).

The cognitive unit strength can be seen as a measure of the bond between agents. This bond increases with the degree of similarity $D^{X,Y}$ between the agent X appraising the action and the agent Y that performed it, defined in the interval $[-1, 1]$. That is, the ratio between the number of attributes (gender, religion, nationality...) that the appraising agent perceives as common and the number of attributes perceived as uncommon. The cognitive unit strength is also influenced by the social relation between the agents. For now, we restrict the social relation between two agents to their affiliation. Other aspects, such as reputation, are left for future work. Roughly, the affiliation $A^{X,Y}$ is a measure of how much agent X likes agent Y, defined in the interval $[-1, 1]$. Therefore, the cognitive unit strength U is a value in $[-1, 1]$ computed as follows:

$$U = D^{X,Y} W_D + A^{X,Y} W_A \quad (1)$$

Where W_D and W_A are the weights of the degree of similarity and the affiliation respectively. Note that $W_D + W_A$ should be equal to one, and because $D^{X,X} = 1$ and $A^{X,X} = 1$, the cognitive unit strength $U^{X,X} = 1$.

The affect derivation determines what, if any, emotions are generated, based on the appraisal variables. According to OCC, the variables described above will trigger *"Attribution Emotions"*, pride, shame, admiration and reproach. These emotions have a base potential, given by the weighted sum of the praiseworthiness/blameworthiness and an expectation deviation. The intensity of the emotion depends on several factors, including its base potential, the agent's mood, arousal or emotional threshold, and is computed using the formulas presented in (Dias and Paiva 2005), since they are common to all emotion types.

The type of the emotion is determined by the praiseworthiness and cognitive unit strength. If the praiseworthiness

	Cognitive Unit Strength	
Praiseworthiness	$U < \alpha$	$U > \alpha$
Positive	Admiration	Pride
Negative	Reproach	Shame

Table 1: How the appraisal variables determine the type of the emotion. Praiseworthiness determines if the emotion is positive or negative, while the cognitive unit strength determines if the emotion is directed to self (if higher than a threshold α) or directed to others (if lower than α)

	Salience	
Smoker	High	Low
Friend	Version A	Version B
Stranger	Version C	Version D

Table 2: The four versions of the scenario generated by varying the salience of the norm and the group of the smoker

is positive, the emotion will be pride or admiration. If the praiseworthiness is negative (blameworthiness positive) the emotion will be shame or reproach. The cognitive unit strength determines if the emotion felt is directed to the self (pride or shame) or if it is directed to others (admiration or reproach) as shown in Table 1.

Case Study

We implemented our model in a well known agent architecture and use it in a simple scenario designed to evaluate the model's power in generating different emotional behaviour according to the salience of a social norm and the social relationships between agents.

In this scenario, the user's avatar is with four characters inside a bar where the norm *"Do not smoke inside bars"* is active, as can be inferred by a no-smoking sign in a wall. The user's avatar is seated with two characters in the foreground, while the remaining two are seated at a more distant table. Initial small-talk helps to reinforce that the near characters are friends with the user's avatar, as previously stated in the introduction, which also states that the others are complete strangers. The overall plot is very simple: after that initial conversation, one of the agents begins to smoke (because it considers its goal of smoking more important than the norm), and the remaining agents react emotionally to that norm violation.

We made four versions of this scenario by varying the salience of the norm and the group of the smoker (see Table 2). In version A, since the norm is salient, all agents will appraise that action as highly blameworthy. The other member of the group of friends will react emotionally by expressing shame, because its cognitive unit strength is higher than the chosen threshold ($\alpha = 0.5$), while the two strangers will express reproach by first looking towards the smoker, and then gesturing their annoyance with arm movements and a frown. Likewise, in version C, the other stranger will express shame while the user's friends will look towards the stranger that smokes, and gesture their annoyance with arm movements and a frown. In version B and D, the blameworthiness of

Figure 3: Two of the four versions created. In the top is version B (friend smokes and the norm has low salience) and in the bottom is version C (stranger smokes and the norm has high salience)

the action is not high enough to trigger a strong emotional expression, so all agents simply look at the smoker, without any visible reaction. Figure 3 shows two of the four versions created (B and C).

Evaluation

In the evaluation conducted, we wanted to test if our model would allow for the emergence of different situations by manipulating parameters in a scenario, with the general aim of creating believable characters that are socially aware. The bar scenario was used to test if users would perceive differences in the emotional responses of the characters due to the manipulation of two social elements in the scene: (a) the salience of the norm and (b) the character's relation with the smoker (i.e, in-group or an out-group). We hypothesized that participants would perceive higher intensities of shame when the norm was violated by an in-group member (conditions A and B) than when it was violated by an out-group character (conditions C and D). We also hypothesized that participants would perceive higher intensities of anger when the norm was violated by an out-group member than when it was violated by an in-group member. Another hypothesis is that the intensities of the emotions perceived would be higher in the high-salient norm conditions (A and C) than in low-salient norm conditions (B and D). Finally, we also hypothesized that users themselves would respond emotionally to the situations, and those responses would be aligned with the user's in-group's character's emotions.

Each participant was randomly assigned to see a scene generated by the system in one of the four conditions of the scenario described above. The only differences between the four conditions were the two parameters manipulated: norm salience; and in-group or out-group relation with the

character breaking the rule. All of the other parameters of the agents in the two versions (goals, relations, properties, drives, plans, memory and emotional reaction rules) were exactly the same.

After the participants observed one of the conditions, they were asked to answer a set of questions concerning their perception of what happened, namely what was the intensity of the shame and anger perceived by the leftmost character when the norm was violated, as well as the intensity of the shame and anger that they themselves felt in the same situation, both using a 7-point Likert scale.

A total of 301 participants (43.2% male and 56.8% female) from twenty-six countries completed the study. Participant's ages ranged between 17 and 64, with the average age being 24.5 years. 65.8% of the participants were non-smokers, 5.6% former smokers, 11.6% occasional smokers and 16.9% were smokers.

Regarding the dependent variable *"intensity of perceived shame in character"*, a two-way ANOVA found a main effect of the relationship with the smoker, $F(1, 297) = 125.383, p < .001$, indicating that the leftmost character was perceived to have felt more shame when the norm was violated by an in-group member than when it was violated by an out-group member. There was also a main effect of the level of salience of the violated norm, $F(1, 297) = 11.134, p = .001$, indicating that the intensity of the shame felt was stronger for a highly salient norm. Finally, there was an interaction between the relationship with the smoker and the salience of the norm, $F(1, 297) = 24.684, p < .001$.

We also ran a two-way ANOVA on the dependent variable *"intensity of perceived anger in character"*. We found a main effect of the relationship with the smoker, $F(1, 297) = 86.214, p < .001$, indicative of a higher intensity of the perceived anger when the norm was violated by an out-group member than when it was violated by an in-group member. We also found a main effect of the level of salience of the violated norm, $F(1, 297) = 17.201, p < .001$, indicative of a higher intensity of the perceived anger when the norm was highly salient. Finally, there was an interaction between the relationship with the smoker and the salience of the norm, $F(1, 297) = 24.054, p < .001$.

We ran simple effect tests that showed some unexpected results: first, the intensity of the shame perceived when a low-salient norm was violated by an in-group member was lower than when it was violated by an out-group member, $F(1, 297) = 16,803, p < .001$. Also, the intensity of the anger perceived when the norm was violated by an in-group member was perceived as higher than when it was violated by an out-group member, $F(1, 297) = 8,310, p = .004$. An explanation for these results could be that we forgive the minor errors of friends when they do not hurt us directly, and react with increased anger whenever we feel personally hurt by friends. We hope to test this hypothesis in future work.

To find out if there is a correlation between the emotions felt by the user and those perceived in an in-group member (the leftmost character), we ran two Pearson's correlation tests. We expected a positive correlation between the intensity of their emotions. Indeed, we found a positive correlation ($p < 0.001$) between the intensity of shame perceived in the character and the intensity of shame felt by the user. The intensity of the perceive anger perceived in the character was also positively correlated with the intensity of anger felt by the user ($p < 0.001$). These results suggest that our model is able to generate emotions in synthetic characters similar to those felt by humans in analogous situations.

Conclusions and Future Work

In this paper, we argued that the influence of social relations in the appraisal of norm-related events is still little studied. We proposed a model able to generate emotions from the appraisal of actions that cause the fulfilment or violation of norms. The model has two major components: a normative component inspired by previous work on normative systems, which detects when a norm is fulfilled or violated by an event, and an emotional component based on the OCC theory of emotions, which appraises the event taking into account social relations and generates an emotional state.

The proposed model was then integrated in an agent architecture for virtual agents and used in a virtual bar where characters with different needs and goals reacted emotionally to the violation of the smoking ban norm. Four versions of this scenario were created by varying the perpetrator and the salience of the norm. In the high-salience versions, the friend of the smoker reacted emotionally with a shame expression while the out-group members frowned. In the low-salient versions, agents would quickly look at the smoker without any emotional reaction.

A study was conducted in which a group of participants interacted with one of the two versions created. The aim was to see how (a) users interpreted the differences in the agents emotional behaviour generated by our model and (b) if the perceived emotions are similar to the emotions usually felt by humans in those situations. The results suggest that users could perceive different emotional states in one of the characters depending on its relation with the smoker and the importance of the ban. Moreover, those emotional states are correlated with the emotions users expected to feel in similar situations.

In future work we plan to extend the model by introducing enforcing mechanisms, such as punishments and conduct further tests. For instance, since most of the participants were from western countries, we would like to see how the user's perceptions about the emotional states of the characters are affected by culture.

Acknowledgements

This work was partially supported by the Portuguese Fundação para a Ciência e a Tecnologia under project PEst-OE/EEI/LA0021/2011 (INESC-ID multiannual funding) through the PIDDAC Program funds, by the European Community (EC) and by a scholarship (BD/62174/2009) granted by FCT. Additionally, it was funded by the SEMIRA (ERA-Compl/0002/2009), eCute (ICT-5-4.2 257666) and GALA (ICT-2009.4.2) projects. The authors are solely responsible for the content of this publication. It does not represent the opinion of EU or the FCT, which are not responsible for any use that might be made of data appearing therein.

References

Andrighetto, G., and Villatoro, D. 2011. Beyond the carrot and stick approach to enforcement: An agent-based model. *European Perspectives on Cognitive Science*.

Arnold, M. 1960. Emotion and personality. vol. i. psychological aspects.

Bates, J. 1992. The nature of characters in interactive worlds and the oz project. *School of Computer Science, Carnegie Mellon University, Pittsburgh, PA Technical Report CMU-CS-92-200*.

Becker-Asano, C., and Wachsmuth, I. 2010. Affective computing with primary and secondary emotions in a virtual human. *Autonomous Agents and Multi-Agent Systems* 20(1):32–49.

Becker-Asano, C. 2008. *WASABI: Affect Simulation for Agents with Believable Interactivity*. Ph.D. Dissertation, PhD Thesis, Faculty of Technology, University of Bielefeld. IOS Press (DISKI 319).

Broersen, J.; Dastani, M.; Hulstijn, J.; and van der Torre, L. 2002. Goal generation in the boid architecture. *Cognitive Science Quarterly* 2(3-4):428–447.

Bulitko, V.; Solomon, S.; Gratch, J.; and van Lent, M. 2008. Modeling culturally and emotionally affected behavior. In *Artificial Intelligence and Interactive Digital Entertainment Conference*.

Castelfranchi, C.; Dignum, F.; Jonker, C.; and Treur, J. 2000. Deliberative normative agents: Principles and architecture. *Intelligent Agents VI. Agent Theories, Architectures, and Languages* 364–378.

Dias, J., and Paiva, A. 2005. Feeling and reasoning: A computational model for emotional characters. *Progress in artificial intelligence* 127–140.

Dignum, F. 1999. Autonomous agents with norms. *Artificial Intelligence and Law* 7(1):69–79.

Elliott, C. 1992. The affective reasoner: A process model of emotions in a multi-agent system. *Ph D Thesis Northwestern University* (1992).

Ferreira, N.; Mascarenhas, S.; Paiva, A.; Dignum, F.; Mc Breen, J.; Degens, N.; and Hofstede, G. 2012. Generating norm-related emotions in virtual agents. In *Intelligent Virtual Agents*, 97–104. Springer.

Gebhard, P. 2005. Alma, a layered model of affect. In *Proceedings of International Joint Conference on Autonomous Agents and Multiagent Systems AAMAS05*. Utrecht, Netherlands: ACM Press.

Gratch, J.; Mao, W.; and Marsella, S. 2006. *Modeling social emotions and social attributions*. Cambridge University Press.

Gratch, J. 2000. Émile: Marshalling passions in training and education. In *Proceedings of the fourth international conference on Autonomous agents*, 325–332. ACM.

López, F.; Luck, M.; and d'Inverno, M. 2006. A normative framework for agent-based systems. *Computational & Mathematical Organization Theory* 12(2):227–250.

Marsella, S.; Johnson, W.; and LaBore, C. 2000. Interactive pedagogical drama. In *Proceedings of the fourth international conference on Autonomous agents*, 301–308. ACM.

Meneguzzi, F., and Luck, M. 2009. Norm-based behaviour modification in bdi agents. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, 177–184. International Foundation for Autonomous Agents and Multiagent Systems.

Mukherjee, P.; Sen, S.; and Airiau, S. 2007. Emergence of norms with biased interactions in heterogeneous agent societies. In *Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology-Workshops*, 512–515. IEEE Computer Society.

Oren, N.; Panagiotidi, S.; Vázquez-Salceda, J.; Modgil, S.; Luck, M.; and Miles, S. 2009. Towards a formalisation of electronic contracting environments. *Coordination, Organizations, Institutions and Norms in Agent Systems IV* 156–171.

Ortony, A.; Clore, G. L.; and Collins, A. 1988. *The Cognitive Structure of Emotions*. Cambridge University Press.

Rank, S. 2004. *Affective Acting: An Appraisal-based Architecture for Agents as Actors*. Ph.D. Dissertation, Institut für Medizinische Kybernetik und Artificial Intelligence, Universität Wien, Vienna, Austria.

Reilly, W. S. N. 1996. *Believable Social and Emotional Agents*. Ph.D. Dissertation, Citeseer.

Scherer, K. 1984. On the nature and function of emotion: A component process approach.

Si, M.; Marsella, S.; and Pynadath, D. V. 2006. Thespian: Modeling socially normative behavior in a decision-theoretic framework. In *IVA*, 369–382.

Si, M.; Marsella, S.; and Pynadath, D. 2010. Modeling appraisal in theory of mind reasoning. *Autonomous Agents and Multi-Agent Systems* 20(1):14–31.

Sloman, A., et al. 2001. Beyond shallow models of emotion. *Cognitive Processing: International Quarterly of Cognitive Science* 2(1):177–198.

Solomon, S.; Van Lent, M.; Core, M.; Carpenter, P.; and Rosenberg, M. 2008. A language for modeling cultural norms, biases and stereotypes for human behavior models. Technical report, DTIC Document.

Staller, A., and Petta, P. 1998. Towards a tractable appraisal-based architecture for situated cognizers. *C. Numaoka, D. Canamero, & P. Petta, Grounding emotions in adaptive systems. SAB* 98.

Villatoro, D.; Andrighetto, G.; Sabater-Mir, J.; and Conte, R. 2011. *Dynamic Sanctioning for Robust and Cost-Efficient Norm Compliance*.

Villatoro, D.; Sabater-Mir, J.; and Sen, S. 2011. Social instruments for robust convention emergence. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume One*, 420–425. AAAI Press.