Exploring the Relationship Between Emotions and Norms in Decision-Making Processes of Intelligent Agents

Joaquin Taverner[1,2][a], Carmengelys Cordova[1,2][b], Elena Del Val[1,2][c], Soledad Valero[1,2][d] and Estefania Argente[1,2][e]

[1]*Valencian Research Institute for Artificial Intelligence (VRAIN), Universitat Politècnica de València, Valencia, Spain*

[2]*Valencian Graduate School and Research Network of Artificial Intelligence (valgrAI), Spain*

Keywords: Emotions, Norms, Intelligent Agents, Multiagent Systems.

Abstract: In this paper, we explore the relationship between norms and emotions, examining the fundamental implications that they entail in the development of future models of reasoning and decision-making for intelligent agents based on BDI. Our approach focuses on assessing the impact of anticipated emotions, self-image, and social-image projection as well as utility factors on the complex decision-making processes that an agent faces when deciding whether to comply with or violate a norm. To this end, we propose the use of two types of anticipated emotions, self-conscious, which shape the personal self-image, and social emotions, which are displayed by other agents in the environment and are used to estimate the social image. To represent the agent's emotional state, we present a new model based on the pleasure dimension in which we represent the self-conscious emotions of Pride and Guilt. Using the language for intelligent agents AgentSpeak, we propose a syntax for defining the norms in the agent's code. We show a new reasoning cycle based on the BDI model in which we add new functionalities to accommodate affective and normative processes. Affective processes support modifying the agent's emotional state as well as estimating anticipated emotions and computing self-image and social image. Normative processes allow the instantiation of active norms and normative reasoning.

1 INTRODUCTION

Traditionally, intelligent agent decision-making models have mainly focused on economic or practical reasoning. However, these approaches fall far short of achieving an accurate simulation of human behavior. Future simulation models will have to reach a higher level of abstraction, both in knowledge representation and in reasoning and decision-making processes, to encompass concepts such as emotions, ethics, values, or social norms (Zhang and Lu, 2021; Walton, 2019; Dorri et al., 2018; da Costa Pereira et al., 2017).

Norms play an essential role in individual decision-making and in structuring social behavior (Daci et al., 2010; O'Neill, 2017). Social norms facilitate cohesion and cooperation by establishing a shared set of rules that establish the boundaries between what is acceptable and inappropriate in different situations, contributing to collective identity and membership (Gross and Vostroknutov, 2022a; Andrighetto et al., 2015). Beyond guiding decisions, the internalization of these norms shapes personal values and how the consequences of actions are evaluated (Sterelny, 2019). Understanding the importance of norms is crucial for accurately analyzing human behavior and is a key element in any attempt to authentically simulate the complexities of social interactions in artificial intelligence.

On the other hand, emotions also play a fundamental role in human behavior and decision-making processes, influencing how we perceive and respond to the world around us. Affective states have a direct impact on the evaluation of situations, affecting our choices and actions (Lerner et al., 2015; Barnes and Thagard, 2019). In the context of social norms, emotions are intrinsically linked to conformity and compliance with established norms (Tangney et al., 2007; Tracy and Weidman, 2021). Norms not only act as ethical guides but also generate emotional responses associated with social acceptance or fear of

[a] https://orcid.org/0000-0002-5612-8033
[b] https://orcid.org/0009-0006-3130-3532
[c] https://orcid.org/0000-0002-1279-3429
[d] https://orcid.org/0000-0003-4565-326X
[e] https://orcid.org/0000-0002-5431-3868

Taverner, J., Cordova, C., Val, E., Valero, S. and Argente, E.
Exploring the Relationship Between Emotions and Norms in Decision-Making Processes of Intelligent Agents.
DOI: 10.5220/0012594700003636

In *Proceedings of the 16th International Conference on Agents and Artificial Intelligence (ICAART 2024) - Volume 1*, pages 471-479
ISBN: 978-989-758-680-4; ISSN: 2184-433X

rejection (Bagozzi et al., 2016; Demeulenaere, 2021). This connection between emotions and norms is bidirectional: emotions can drive compliance with norms, and, in turn, adherence to norms can elicit emotional responses (Argente et al., 2020). Understanding this dynamic link between emotions and norms is essential to simulate human social behavior and the behavior of human social organizations, both in individual decision-making and in broader social interactions.

In this paper we study the relationship between emotions and norms in depth, exploring the possible implications for future models of intelligent agent reasoning and decision making. The main research objective is to analyze the impact of anticipated emotions, self-image, and social image on the decision-making processes when an agent must decide whether or not to comply with a norm. To address these complex dynamics, we propose an innovative approach that incorporates emotional components into a normative agent model based on BDI model (Beliefs-Desires-Intentions) (Rao et al., 1995).

The rest of the paper is organized as follows: Section 2, an initiation into the theoretical background is presented. Section 3 offers an overview of pertinent state-of-the-art research concerning affective and normative agents. The introduction of our novel normative affective agent model is detailed in Section 4. Finally, the concluding remarks and prospects for future work are presented in Section 6.

2 EMOTIONS AND NORMS

In human society, there are moral and social-conventional norms, but also legal, epistemic, aesthetic and personal norms (O'Neill, 2017). Nucci and Turiel identified moral norms as those relating to "justice, welfare, or rights of individuals or groups" (Nucci and Turiel, 1978). Conventional norms are defined as a set of rules and guidelines that are established by a society to regulate the behavior of its members (Killen et al., 2013). Norms prescribe the behavior that is considered acceptable or unacceptable in various social situations (Demeulenaere, 2021). From a sociological perspective, norms serve as a basis for maintaining order and social cohesion, providing a shared framework of reference that governs the behavior of individuals within a group or society and guides interactions and relationships between those individuals (Boytsun et al., 2011). The establishment of behavioral expectations facilitated by norms facilitates relationships, communication, and cooperation among members of society.

We can fundamentally categorize norms into five types (Argente et al., 2020):

- *Institutional norms*, which are established by competent authorities (e.g., the government) and are presumed to be followed by all individuals in a society.
- *Social norms* (or conventions), which emerge from repeated social interactions, establishing practices that are socially accepted.
- *Interaction norms*, which refer to formal agreements that affect specific groups for a limited period (e.g., a commercial contract).
- *Private norms*, which are internal rules self-imposed by individuals to guide their behavior.

Among the various theories on emotions, appraisal-based perspectives have gained wide acceptance in the field of emotional psychology (Lazarus, 1991). According to this approach, emotions are not simply considered as automatic responses to external stimuli but arise through a process of cognitive evaluation. Therefore, emotions can be considered as subjective and evaluative responses to events or situations, shaped by each individual's interpretation of the personal relevance and emotional significance of such events (Lewis et al., 2010).

Different types of emotional experiences affect individuals' decisions and behaviors in significant ways (Barnes and Thagard, 2019). For this paper, we focus on anticipated emotions (Bagozzi et al., 2016). These types of emotions represent to some degree emotional expectations regarding future events, either in terms of positive emotions (i.e., satisfaction), negative emotions (i.e., fear or anxiety), or neutral emotions (i.e., calm or indifference) (Carrera et al., 2011). Here we can distinguish two types of anticipated emotions depending on the target to which these emotions are directed. Thus, anticipated emotions can be directed towards oneself, reflecting how the individual internally perceives the emotional consequences of his/her intentions, which is part of one's self-image (Bodner and Prelec, 2003; Gross and Vostroknutov, 2022a). On the other hand, emotions can come from individuals in the social environment. The anticipation of social emotions is closely linked to concerns about how individuals will be perceived by other individuals in society (Simamora, 2021), which is known as social image (Gross and Vostroknutov, 2022b). Social emotions and social image provide guidance in decision-making permitting subjects to anticipate and evaluate future social outcomes (Grossman and Van der Weele, 2017; Gratch et al., 2006).

There is an evident influence between norms and emotions that acts in a bidirectional way. Emotions

often promote compliance with norms, while adherence to norms can trigger emotional responses (Argente et al., 2020). Self-conscious emotions, like Pride or Guilt, play a fundamental role in the self-evaluation of norm compliance (Tangney et al., 2007). Moreover, based on expectations, anticipated emotions are used to estimate the emotional impact of complying with a norm (Bagozzi et al., 2016).

In (Argente et al., 2020), four types of relationships between norms and emotions are defined (see Figure 1):

- *(i) Emotion is considered in the normative reasoning process*: emotions are a factor to be considered during the stage of reasoning in which it is determined whether to fulfill or violate a norm. Specifically, anticipated self- and social emotios play a crucial factor in the normative reasoning process by anticipating future emotional outcomes of fulfilling or violating a norm. For example, self- or social emotions (e.g., pride or disgust) can be motivators for complying with a norm.
- *(ii) Norm fulfillment/violation generates emotions*: norm fulfillment/violation becomes a triggering event for a cognitive-level appraisal process that results in the elicitation of emotions. Thus, the fulfillment of a norm can produce positive emotions in the individual (e.g., pride) and in the individuals that compose his/her social environment (e.g., gratitude). However, the violation of a norm can produce negative emotions (e.g., guilt) in the individual and generate social rejection.
- *(iii) Emotion enables compliance with social norms*: specifically, anticipated social emotions have been proposed as facilitators of compliance with social norms. When an individual anticipates the emotional impact of fulfilling or violating a norm (i.e., social image), he/she uses this information as a factor in determining whether to fulfill or violate a norm.
- *(iv) Emotion helps to internalize private norms*: the continuous observation of the emotional and social consequences of compliance or violation of behavior in a given social environment facilitates the internalization of private or internal norms. In addition, it allows individuals to improve their accuracy in establishing their own social emotions in the face of a given behavior.

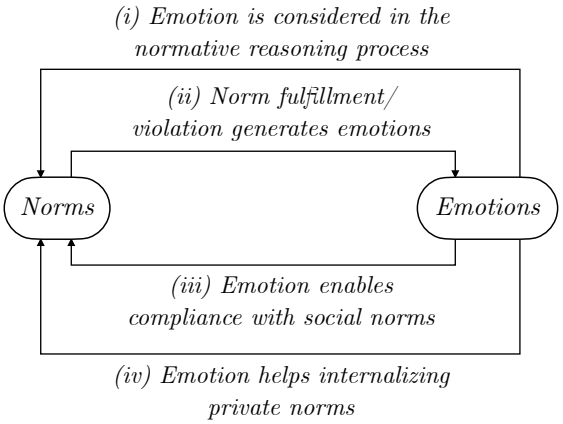

Figure 1: Relationships between emotions and norms.

3 EMOTIONS AND NORMS IN AGENTS

In the field of artificial intelligence, intelligent agents are defined as entities capable of perceiving their environment, making decisions autonomously, and performing actions to achieve specific objectives. One of the most widely used approaches for modeling agents is the Beliefs-Desires-Intentions (BDI) model (Rao et al., 1995). This model is based on the idea that agents act according to their beliefs about the world, their desires or goals to be achieved, and the intentions that derive from the combination of beliefs and desires.

In the literature, we can find different proposals that have covered to a greater or lesser extent the relationships between emotions and norms established in Section 2. For example, in (Staller et al., 2001) TABASCO$_{JAM}$, an architecture for normative BDI agents, is proposed. That architecture encompasses the relationship (i) (see Figure 1). In that architecture, the normative reasoning of agents is modeled by a set of conditional rules that include the intensity of emotion as a decision factor in determining whether or not to execute a given action.

In (Ahmad et al., 2012) the model for normative agents OP-RND-E is presented. The authors propose a model that uses the fulfillment or violation of norms as a motivating element for the elicitation of emotions. For this purpose, they use an emotional appraisal process capable of eliciting positive or negative emotions. These emotions are subsequently used by the agent to reason about the fulfillment of its objectives. Therefore, that proposal covers relationship (ii) through the use of elicited emotions.

Another interesting proposal can be found in (Fer-

reira et al., 2013). The authors use the architecture for affective agents BDI FAtiMA (Mascarenhas et al., 2022) to develop a normative agent model. That model makes it possible to define the relationships between agents and to establish acceptable social behaviors. Agents react emotionally, with self-conscious emotions such as pride or shame when they perceive that an agent in the environment has performed a behavior that is acceptable or non-acceptable. Similarly, in (Tzeng et al., 2021; Tzeng, 2022) a model is proposed that uses social emotions to determine compliance with norms. The agent has an affective appraisal process based on a finite set of rules that define predefined emotions related to the fulfillment/violation of each norm.

Anticipated emotions have also been used in normative agent models. For example, in (Kollmann et al., 2016), the ECABA architecture is presented. This architecture employs anticipated self-consciuos emotions in the normative reasoning process of agents to determine whether to fulfill or violate a norm.

As can be seen, there are different proposals of agents considering emotions in normative reasoning processes. However, these proposals generally explore only one of the relationships between emotions and norms described in Section 2. Also, to our knowledge, none of these proposals explore the combination of self-conscious and social anticipated emotions.

4 PROPOSAL

In this paper, we present an innovative approach to agent modeling using a BDI model and the AgentSpeak language (Bordini and Hübner, 2005). Our model incorporates emotions and norms comprehensively into the agent's decision-making process, addressing the complex relationships *(i)*, *(ii)*, and *(iii)* between these elements described in Section 2. We highlight the importance of emotional state and anticipated emotions in the agent's normative reasoning, especially when faced with decisions about norm compliance or violation.

To develop a model of normative-affective agents, we propose a normative agent model that considers anticipated emotions at both the individual and social levels. This article focuses on the deontic operator of prohibition, but in the future, the proposal could be extended to other operators. In our model, norms are represented using the tuple $\langle id, A, P, R, S \rangle$, where *id* serves to identify the norm, *A* represents the set of conditions activating the norm, *P* is the set of actions affected by the norm, and *R* and *S* denote the reward and sanction, respectively. For instance, to express the prohibition of exceeding 60 kilometers per hour, and following the prolog-style syntax commonly used in AgentSpeak, the syntax of the norm is defined as:

```
prohibition(ms, so_60, accelerate, 0, 1)
```

where, `ms` (max speed) is the norm's identifier, `so_60` (speed over 60) denotes the activation condition, in this example, it is activated when the agent is driving in a zone limited to 60 kilometers per hour; `accelerate` is the action of accelerating the vehicle, and the reward for adhering to the speed limit is 0, while the sanction is 1.

In this system, all agents are capable of perceiving the actions performed by other agents in their environment with a certain degree of probability. When agents perceive that an action complies with or violates a norm, they can react emotionally.

Table 1 summarizes the proposed emotions for our normative-affective agent model. We define two social emotions, Gratitude, and Disgust, along with three internal emotions, Pride, Guilt, and Calm. Social emotions are triggered, with a certain probability, when agents perceive that another agent has complied with or violated a norm. These emotions have been carefully chosen to encompass a broad spectrum of emotional responses within both social and internal contexts. Gratitude stands out for reinforcing positive social interactions, fostering connection, and acknowledging the beneficial actions of others (Bartlett et al., 2012). Meanwhile, disgust plays a crucial role in signaling aversion to undesirable social behaviors, thus contributing to the regulation of social norms and values (Inbar and Pizarro, 2022). On the other hand, pride serves as a mechanism of positive reinforcement, providing a sense of achievement and boosting self-esteem. In contrast, guilt functions as a self-regulation mechanism, prompting reflection on past actions and contributing to more conscientious and socially responsible behavior (Onwezen et al., 2013). Finally, calm is presented as an essential element for establishing emotional balance, offering a serene state that facilitates thoughtful decision-making and effective stress management (Meshulam et al., 2012).

Internal emotions are defined through the pleasure dimension (see Figure 2). Through this representation, the agent can have different intensities for each emotion, providing a more nuanced perspective in its emotional response.

During the normative reasoning process, both self-emotions and social emotions are anticipated to estimate self-image and social image. Subsequently, self-image and social image are used by the normative reasoning process to decide whether to fulfill or violate the norm.

Table 1: Emotions that can be elicited in our agent model.

Pleasure	Emotion	
	Social	Self
Positive	Gratitude	Pride
Negative	Disgust	Guilt
Neutral		Calm

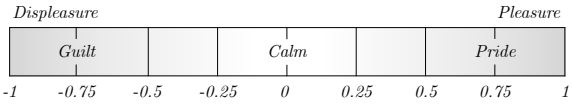

Figure 2: Pleasure dimension.

The Algorithm 1 illustrates the agent's behavior (see Figure 3). The agent can initially have a set of initial beliefs $B^0 = \{b_1^0, b_2^0, ...\}$ stored in the agent's belief set $B = \{b_1, b_2, ...\}$; a set of initial intentions $I^0 = \{\iota_1^0, \iota_2^0, ...\}$ saved in the agent's intention set $I = \{\iota_1, \iota_2, ...\}$; a set of norms $N^0 = \{n_1^0, n_2^0, ...\}$ stored in the norm set $N = \{n_1, n_2, ...\}$, where $n_i = \langle id, A, P, R, S \rangle$; and the initial emotional knowledge E_0 stored in the variable E. Emotional knowledge consists of a tuple $\langle \sigma, \Delta^+, \Delta^-, \rho^+, \rho^- \rangle$, where σ represents the agent's pleasure level; Δ^+ and Δ^- denote the agent's tendency to experience positive (e.g., Pride) or negative (e.g., Guilt) emotions, respectively; and ρ^+ and ρ^- are vectors recording the number of positive (e.g., Gratitude) and negative (e.g., Disgust) emotions perceived in agents in the environment as a result of certain actions.

Once the variables are initialized, the process begins by obtaining the set of environment perceptions $\beta \leftarrow \{\theta_1, \theta_2, ...\}$ using the `percept` method. These perceptions include both the agent's observations in the environment and any messages received from other agents. Subsequently, the `belief_revision` method takes the agent's beliefs B along with the set of perceptions β as input and updates both the belief set B and the emotional information E, i.e., the emotions perceived in the environment.

Next, the agent's desires D are estimated using the `options` method, and intentions I are calculated through the `filter` method. Once these estimated belief, desire, and intention sets are in place, the agent devises a plan π using the `plan` method. This method, utilizing B, D, and I, along with the available action catalog A, establishes the plan to be followed.

An iterative process then begins in which each action α constituting the plan π is chosen one by one. For each action α, the agent calculates the possible active norms using the `norm_instantiation` method. This method evaluates the activation conditions of the norms in N, considering the agent's beliefs B. Once the active norms are determined,

Algorithm 1: Agent's reasoning cycle.

1: $B \leftarrow B_0$
2: $I \leftarrow I_0$
3: $N \leftarrow N_0$
4: $E \leftarrow E_0$
5: **while** True **do**
6: $\beta \leftarrow$ *percept()*
7: $B, E \leftarrow$ *belief_revision*(B, β)
8: $D \leftarrow$ *options(B,I)*
9: $I \leftarrow$ *filter(B,D,I)*
10: $\pi \leftarrow$ *plan(B,D,A)*
11: **while** not $\pi = \{\}$ **do**
12: $\alpha \leftarrow$ first element of π
13: $\pi \leftarrow$ tail of π
14: $\mu \leftarrow$ *norm_instantiation(B,N)*
15: $\sigma', \rho' \leftarrow$*anticipated_emotions(B,E,*μ*)*
16: $\theta \leftarrow$ *norm_reasoning(B,E,*$\mu, \sigma', \rho', \alpha$*)*
17: **if** $\theta =$ violate **then**
18: *execute(*α*)*
19: **else**
20: $\pi \leftarrow \{\}$
21: **end if**
22: $\beta \leftarrow$ *percept()*
23: $B, E \leftarrow$ *belief_revision*(B, β)
24: $D \leftarrow options(B, I)$
25: $I \leftarrow filter(B, D, I)$
26: *update_mood(B,E,*θ*)*
27: **if** *succeeded(I,B)* or *impossible(I,B)* **then**
28: **Break**
29: **end if**
30: **end while**
31: **end while**

the agent estimates anticipated emotions using the `anticipated_emotions` method. This method utilizes B, E, and the active norms μ to calculate the parameters σ' and ρ', representing self-image and social image, respectively. σ' is a tuple $\left\langle \sigma'_f, \sigma'_v \right\rangle$ indicating the anticipated emotions when complying with or violating the norm, respectively. These parameters are estimated considering the current emotional state of the agent by the equations:

$$\sigma'_f = (\sigma + \Delta^+)^2 \tag{1}$$

$$\sigma'_v = 1 - (\sigma + \Delta^-)^2 \tag{2}$$

Similarly, ρ' is a tuple $\left\langle \rho'_f, \rho'_v \right\rangle$ representing the anticipated social emotions when complying with or violating the norm. These factors are estimated as a sum of the positive and negative social emotions (i.e., Gratitude and Disgust) perceived on the environment in each case, taking into account the number of agents (nA) inside the environment:

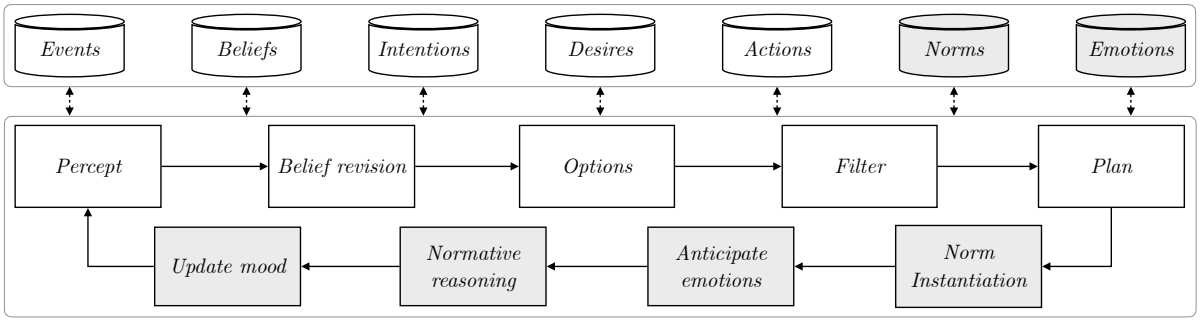

Figure 3: Process diagram of the proposed normative-affective BDI agent.

$$\rho_f = \frac{\sum_{i=1}^{|\rho^+|} \rho_i^+}{|\rho^+| \cdot (nA-1)} \tag{3}$$

$$\rho_v = \frac{\sum_{i=1}^{|\rho^-|} \rho_i^-}{|\rho^-| \cdot (nA-1)} \tag{4}$$

The normative reasoning process, represented by the `norm_reasoning` method, uses the calculated information to determine the decision to comply or violate the norm through an equation:

$$\theta = \begin{cases} \text{fulfill}, & \text{if } \theta_f \geq \theta_v \\ \text{violate}, & \text{otherwise} \end{cases} \tag{5}$$

$$\theta_f = \omega_\sigma \cdot \sigma'_f + \omega_\rho \cdot \rho'_f + \omega_\varphi \cdot \varphi_f \tag{6}$$

$$\theta_v = \omega_\sigma \cdot \sigma'_v + \omega_\rho \cdot \rho'_v + \omega_\varphi \cdot \varphi_v \tag{7}$$

where ω_i represents the importance that the agent gives to each factor and φ_f and φ_v are the expected utility if the norm is fulfilled or violated, respectively.

Once the agent makes the decision θ, its emotional state is updated using the `update_mood` function. If the agent decides to violate the norm (using the previous example, exceeding 60 kilometers per hour), the action α is executed through the `execute` method. If it chooses to comply with the prohibition, the plan is discarded, and no actions are executed.

Subsequently, perceptions, beliefs, social emotions, desires, and intentions are rechecked. Following that, the agent's emotional state is updated using the `update_mood` function. Finally, if the plan has achieved the goal (`succeeded`) or is unattainable (`impossible`), the plan execution is halted.

5 EXAMPLE

Consider an environment with five coexisting agents ($nA = 5$). After applying Algorithm 1 for t cycles, one agent, named *AgentA*, is in a state where B and I have the current set of beliefs and intentions; N includes the previous exceeding 60 kilometers per hour norm (named *maxSpeed*); its emotional knowdlege $E_t = \langle \sigma_t, \Delta_t^+, \Delta_t^-, \rho_t^+, \rho_t^- \rangle$, where $\sigma_t = 0$ (i.e, calm emotional state), $\Delta_t^+ = 1$ and $\Delta_t^- = 1$ (strong inclination for both positive and negative emotions, that is, it tends to take great pride in its "good" actions but also feels deep remorse for its "bad" actions); $\rho_t^+ = [1,2,1]$ and $\rho_t^- = [2,2,1]$, where each component indicates the number of positive (e.g., Gratitude) or negative (e.g., Guilt) emotions, respectively, perceived in the environment in each cycle, as a result of its previous actions. Moreover, *AgentA* values its social image ($\omega_\rho = 0.9$), has high importance on its self-image ($\omega_\sigma = 0.75$) and relative importance on the norm utility ($\omega_\varphi = 0.5$).

Suppose `accelerate` is the action to be evaluated now, *maxSpeed* is active, and the expected utility of the norm is higher if violated than fulfilled (e.g. $\varphi_f = 0.25$ and $\varphi_v = 0.75$).

In the `anticipated_emotions` method, *AgentA* calculates:

$$\sigma'_f = (0.5+1)^2 = 2.25$$

$$\sigma'_v = 1-(0.5+1)^2 = -1,25$$

$$\rho'_f = \frac{(1+2+1)}{3 \cdot (5-1)} = 0.333$$

$$\rho'_v = \frac{(2+2+1)}{3 \cdot (5-1)} = 0.417$$

In the `norm_reasoning` method, *AgentA* calculates:

$$\theta_f = 0.75 \cdot 2.25 + 0.9 \cdot 0.333 + 0.5 \cdot 0.25 = 2.11$$

$$\theta_v = 0.75 \cdot -1.25 + 0.9 \cdot 0.417 + 0.5 \cdot 0.75 = -0.19$$

AgentA fulfills the *maxSpeed* norm, discarding the `accelerate` action. Despite the preference for violating the norm ($\varphi_v > \varphi_f$), its preference for social-image and self-image makes it fulfill the norm.

Now, consider an antagonist, *AgentB*, with high pride and low guilt of its actions ($\Delta_t^+ = 1$ and $\Delta_t^- = 0.1$), low importance on its self-image ($\omega_\sigma = 0.1$), low importance on its social image ($\omega_\rho = 0.1$) and relative importance on the norm utility ($\omega_\varphi = 0.5$). In a similar situation as *AgentA*, the `anticipated_emotions` method of this agent *AgentB* would have calculated:

$$\sigma'_f = (0.5+1)^2 = 2.25$$

$$\sigma'_v = 1-(0.5+0.1)^2 = 0.64$$

$$\rho'_f = \frac{(1+2+1)}{3\cdot(5-1)} = 0.333$$

$$\rho'_v = \frac{(2+2+1)}{3\cdot(5-1)} = 0.417$$

And the `norm_reasoning` method of agent *AgentB* would calculate:

$$\theta_f = 0.1\cdot 2.25 + 0.1\cdot 0.333 + 0.5\cdot 0.25 = 0.383$$

$$\theta_v = 0.1\cdot 0.64 + 0.1\cdot 0.417 + 0.5\cdot 0.75 = 0.48$$

In this case, *AgentB* violates the *maxSpeed* norm, executing the `accelerate` action, as it does not care about others' feelings and its social image and prefers violating the norm.

6 CONCLUSION

Considering emotional consequences (anticipated emotions) in the cognitive reasoning of agents when determining whether to comply with or violate a norm is a step towards improving the simulation of human social behavior. This paper has analyzed the interplay between norms and emotions, paying attention to their implications in the decision-making process of intelligent agents in normative environments. A novel perspective has been introduced by considering the influence of anticipated emotions as well as the utility impact on the processes associated with decision-making. We have differentiated two categories of anticipated emotions: self-emotions, instrumental in shaping the agent's self-image, and social emotions, emanating from other agents within the environment. The emotional state of the agent is encapsulated in a model grounded in the pleasure dimension, specifically incorporating the self-conscious emotions of Pride and Guilt. Furthermore, we proposed a syntax within the framework of AgentSpeak for the definition of norms in the agent's code. We have also introduced a novel reasoning cycle, extending the BDI model to accommodate additional functionalities tailored for affective and normative processes. Affective processes have been designed to facilitate alterations in the agent's emotional state, compute anticipated emotions, and assess self-image and social image. Concurrently, normative processes empower the instantiation of active norms and the execution of normative reasoning, thus enriching the landscape of intelligent agent decision-making. This proposal enhances the agent's decision-making capabilities, enabling them to make more informed and socially acceptable decisions. This approach lays the foundation for the development of more intelligent and socially aligned agents capable of navigating complex social environments.

In this work, anticipated emotions are estimated from default values that do not differentiate between different types of norms. In the future, it would be interesting to evaluate anticipated emotions when complying with or violating a norm by considering factors such as the importance of the norm, both at the individual level (e.g., evaluating personal ethical values) and at the societal level (e.g., evaluating the ethical behavior of society), or previous emotional experiences when complying with or violating a particular norm. In this way, agents could infer which types of norms have a higher emotional and social cost.

ACKNOWLEDGEMENTS

Work partially supported by Generalitat Valenciana CIPROM/2021/077, Spanish Government projects PID2020-113416RB-I00 and TED2021-131295B-C32; TAILOR project funded by EU Horizon 2020 under GA No 952215.

REFERENCES

Ahmad, A., Ahmad, M. S., Yusoff, M. Z. M., and Ahmed, M. (2012). Formulating agent's emotions in a normative environment. In *Knowledge Technology*, pages 82–92. Springer.

Andrighetto, G., Grieco, D., and Tummolini, L. (2015). Perceived legitimacy of normative expectations motivates compliance with social norms when nobody is watching. *Frontiers in psychology*, 6:1413.

Argente, E., Del Val, E., Perez-Garcia, D., and Botti, V. (2020). Normative emotional agents: a viewpoint paper. *IEEE Transactions on Affective Computing*, 13(3):1254–1273.

Bagozzi, R. P., Belanche, D., Casaló, L. V., and Flavián, C. (2016). The role of anticipated emotions in purchase intentions. *Psychology & Marketing*, 33(8):629–645.

Barnes, A. and Thagard, P. (2019). Emotional decisions. In *Proceedings of the eighteenth annual conference of the cognitive science society*, pages 426–429. Routledge.

Bartlett, M. Y., Condon, P., Cruz, J., Baumann, J., and Desteno, D. (2012). Gratitude: Prompting behaviours that build relationships. *Cognition & emotion*, 26(1):2–13.

Bodner, R. and Prelec, D. (2003). Self-signaling and diagnostic utility in everyday decision making. *The psychology of economic decisions*, 1(105):26.

Bordini, R. H. and Hübner, J. F. (2005). Bdi agent programming in agentspeak using jason. In *International workshop on computational logic in multi-agent systems*, pages 143–164. Springer.

Boytsun, A., Deloof, M., and Matthyssens, P. (2011). Social norms, social cohesion, and corporate governance. *Corporate Governance: An International Review*, 19(1):41–60.

Carrera, P., Caballero, A., Muñoz, D., and Oceja, L. (2011). Anticipated emotions and personal experience for predicting behavioral intentions and behavioral expectations. *The Spanish journal of psychology*, 14(2):535–547.

da Costa Pereira, C., Tettamanzi, A. G., Villata, S., Liao, B., Malerba, A., Rotolo, A., and van Der Torre, L. (2017). Handling norms in multi-agent system by means of formal argumentation. *Journal of Applied Logics-IfCoLoG Journal of Logics and their Applications*, 4(9):1–35.

Daci, J. et al. (2010). Legal principles, legal values and legal norms: are they the same or different? *Academicus International Scientific Journal*, 1(02):109–115.

Demeulenaere, P. (2021). Norms. In *Research Handbook on Analytical Sociology*, pages 233–248. Edward Elgar Publishing.

Dorri, A., Kanhere, S. S., and Jurdak, R. (2018). Multi-agent systems: A survey. *Ieee Access*, 6:28573–28593.

Ferreira, N., Mascarenhas, S., Paiva, A., Di Tosto, G., Dignum, F., McBreen, J., Degens, N., Hofstede, G. J., Andrighetto, G., and Conte, R. (2013). An agent model for the appraisal of normative events based in in-group and out-group relations. In *AAAI*.

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

Gross, J. and Vostroknutov, A. (2022a). Why do people follow social norms? *Current Opinion in Psychology*, 44:1–6.

Gross, J. and Vostroknutov, A. (2022b). Why do people follow social norms? *Current Opinion in Psychology*, 44:1–6.

Grossman, Z. and Van der Weele, J. J. (2017). Self-image and willful ignorance in social decisions. *Journal of the European Economic Association*, 15(1):173–217.

Inbar, Y. and Pizarro, D. A. (2022). How disgust affects social judgments. In *Advances in Experimental Social Psychology*, volume 65, pages 109–166. Elsevier.

Killen, M., Rutland, A., Abrams, D., Mulvey, K. L., and Hitti, A. (2013). Development of intra-and intergroup judgments in the context of moral and social-conventional norms. *Child Development*, 84(3):1063–1080.

Kollmann, S., Siafara, L. C., Schaat, S., and Wendt, A. (2016). Towards a cognitive multi-agent system for building control. *Procedia Computer Science*, 88:191–197.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.

Lerner, J. S., Li, Y., Valdesolo, P., and Kassam, K. S. (2015). Emotion and decision making. *Annual review of psychology*, 66:799–823.

Lewis, M., Haviland-Jones, J. M., and Barrett, L. F. (2010). *Handbook of emotions*. Guilford Press.

Mascarenhas, S., Guimarães, M., Prada, R., Santos, P. A., Dias, J., and Paiva, A. (2022). Fatima toolkit: Toward an accessible tool for the development of socio-emotional agents. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 12(1):1–30.

Meshulam, M., Winter, E., Ben-Shakhar, G., and Aharon, I. (2012). Rational emotions. *Social Neuroscience*, 7(1):11–17.

Nucci, L. P. and Turiel, E. (1978). Social interactions and the development of social concepts in preschool children. *Child development*, pages 400–407.

O'Neill, E. (2017). Kinds of norms. *Philosophy Compass*, 12(5):e12416.

Onwezen, M. C., Antonides, G., and Bartels, J. (2013). The norm activation model: An exploration of the functions of anticipated pride and guilt in pro-environmental behaviour. *Journal of economic psychology*, 39:141–153.

Rao, A. S., Georgeff, M. P., et al. (1995). Bdi agents: from theory to practice. In *Icmas*, volume 95, pages 312–319.

Simamora, B. (2021). How proponents and opponents influence achievement motivation: The role of the anticipated emotions of other people. *Gadjah Mada International Journal of Business*, 23(1):1–35.

Staller, A., Petta, P., et al. (2001). Introducing emotions into the computational study of social norms: A first evaluation. *Journal of artificial societies and social simulation*, 4(1):U27–U60.

Sterelny, K. (2019). Norms and their evolution. In *Handbook of Cognitive Archaeology*, pages 375–397. Routledge.

Tangney, J. P., Stuewig, J., and Mashek, D. J. (2007). What's moral about the self-conscious emotions. *The self-conscious emotions: Theory and research*, pages 21–37.

Tracy, J. L. and Weidman, A. C. (2021). *The self-conscious and social emotions: A personality and social functionalist account.* The Guilford Press.

Tzeng, S.-T. (2022). Engineering normative and cognitive agents with emotions and values. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, pages 1878–1880.

Tzeng, S.-T., Ajmeri, N., and Singh, M. P. (2021). Noe: Norm emergence and robustness based on emotions in multiagent systems. In *International Workshop on Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems*, pages 62–77. Springer.

Walton, D. (2019). Practical reasoning in the deliberations of an intelligent autonomous agent. *Blended Cognition: The Robotic Challenge*, pages 53–84.

Zhang, C. and Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration*, 23:100224.