Noe: Norm Emergence and Robustness Based on Emotions in Multiagent Systems

Sz-Ting Tzeng[1], Nirav Ajmeri[2], and Munindar P. Singh[1]

[1] North Carolina State University, Raleigh NC 27695, USA
{stzeng,mpsingh}@ncsu.edu
[2] University of Bristol, Bristol, BS8 1UB, UK
nirav.ajmeri@bristol.ac.uk

Abstract. Social norms characterize collective and acceptable group conducts in human society. Furthermore, some social norms emerge from interactions of agents or humans. To achieve agent autonomy and make norm satisfaction explainable, we include emotions into the normative reasoning process, which evaluates whether to comply or violate a norm. Specifically, before selecting an action to execute, an agent observes the environment and infers the state and consequences with its internal states after norm satisfaction or violation of a social norm. Both norm satisfaction and violation provoke further emotions, and the subsequent emotions affect norm enforcement. This paper investigates how modeling emotions affect the emergence and robustness of social norms via social simulation experiments. We find that an ability in agents to consider emotional responses to the outcomes of norm satisfaction and violation (1) promotes norm compliance; and (2) improves societal welfare.

1 Introduction

Humans, in daily life, face many choices at many moments, and each selection brings positive and negative payoffs. In psychology, decision-making [33] is a cognitive process that selects a belief or a series of actions based on values, preferences, and beliefs to achieve specific goals. Emotions, the responses to internal or external events or objects, can involve the decision-making process and provide extra information in communication [17, 32]. Social norms describe societal principles between agents in a multiagent system. While social norms regulate behaviors in society [15, 30, 35], humans and agents have the capacity to deviate from norms in certain contexts. For instance, people shake hands normally but deviate from this social norm during a pandemic. Chopra and Singh [8] describe how social protocols rely on a foundation of norms though they do not discuss how the appropriate norms emerge.

An agent that models the emotions of its users and other humans can potentially behave in a more realistic and trustworthy manner. The decision-making process for humans or agents involves evaluating possible consequences of available actions and choosing the action that maximizes the expected utility [11]. Herbert Simon, one of the founders of AI, emphasized that general thinking and problem-solving must incorporate the influence of emotions [34]. Without considering emotions or other affective

characteristics, such as personality or mood, some compliance seems irrational [4]. Humans' compliance shows hints on rational planning over their objectives [17]. Including emotion or personality in normative reasoning makes these compliance behaviors explainable. Norms either are defined in a top-down manner or emerge in a bottom-up manner [25, 30]. Works on norms include norm emergence based on the prior outcome of norms, automated run-time revision of sanctions [10], or considering various aspects during reasoning [1, 2]. However, sanctions in the real world are often subtle instead of harsh punishments. For instance, sanctions could be trust updates or emotional expression and might change one's behavior [6, 27]. Kalia et al. [16] considered norm outcome with respect to emotions and trust and goals. Modeling and reasoning about emotions and other affective characteristics in an agent then become important in decision making and would help the agent enforce and internalize norms.

Accordingly, we propose *Noe*, an agent architecture that integrates decision-making with normative reasoning and emotions. We investigate the following research question.

RQ$_{emotion}$. How does modeling the emotional responses of agents to the outcomes of interactions affect norm emergence and social welfare in an agent society?

To address RQ$_{emotion}$, we refine the abstract normative emotional agent architecture [4] and investigate the interplay of norms and emotions. We propose a framework *Noe* based on BDI architecture [29], norm life-cycle [4, 12, 30], and emotion life-cycle [3, pp. 62–64] [21]. To evaluate *Noe*, we design a simulation experiment with various agent societies. We investigate how norms emerge and how emotions in normative agents influence social welfare.

To make the problem tractable, we apply one social norm in our simulation and simplify the emotional expression to reduce the complexity. Specifically, our *Noe* agents process emotions by appraising norm outcomes. For the emotion model, we adopt the OCC model of emotions [28] in which we consider both emotional valence and intensity and assume violation of norms yields negative emotions.

Organization. The rest of the paper is structured as follows. Section 2 discusses the relevant related works. Section 3 describes *Noe*, including the symbolic representation and the decision-making in *Noe*. Section 4 details the simulation experiments we conduct to evaluate *Noe* and describes the experimental results. Section 5 presents the conclusions and the future directions.

2   Related Works

Ortony et al. [28] model emotions based on events, action, and objects. Marsella and Gratch [21] proposed a computational model of emotion to model appraisal in perceptual, cognitive, and behavioral processes. Moerland et al. [24] surveyed emotions in relation to reinforcement learning. Keltner and Haidt [17] differentiate the functional approaches and research of emotions by four-level analysis: individual, dyadic, groups, and cultural. Briefly, emotions provide some information for agents or people to coordinate social interactions. We take inspiration from these works.

Savarimuthu and Cranefield [30] proposed a life-cycles model for norms and discussed varied mechanisms of norm study. Broersen et al. [7] introduced the so-called Beliefs-Obligations-Intentions-Desires (BOID) architecture on top of the Beliefs-Intentions-Desires (BDI) architecture [29], which further include obligation and conflict resolution. Lima et al. [18] developed Gavel, an adaptive sanctioning enforcement framework, to choose appropriate sanctions based on different contexts. However, these works do not consider emotions in the decision-making process.

Argente et al. [4] propose an abstract normative emotional agent architecture, which combines emotion model, normative model, and Belief-Desire-Intention (BDI) architecture. Argente et al. defined four types of relationships between emotions and norms: (1) emotion in the process of normative reasoning, (2) emotion generation with norm satisfaction or violation, (3) emotions as a way to enforce norms, (4) anticipation of emotions promotes internalization and compliance of social norms. Yet, Argente et al. do not validate the interplay between emotions and norms with their proposed architecture.

Bourgais et al. [6] present an agent architecture that integrates cognition, emotions, emotion contagion, personality, norms, and social relations to simulate humans and ensure explainable behaviors. However, emotions are predefined and not generated via appraisal in this work.

Von Scheve et al. [31] consider emotion generation with norm satisfaction or violation. Specifically, an observer agent perceives the transgression of a norm of another, its strong negative emotions (e.g., contempt, disdain, detestation, or disgust) constitute negative sanctioning of the violator. The negative sanctioning then leads to negative emotions (e.g., shame, guilt, or embarrassment) in the violator. Besides, compliance with the social norms can stem from the fear of emotional-driven sanctions, which would lead to negative emotions in the violator. Such fear enforces social norms. Yet, emotions are not part of the decision-making process in this work.

3 *Noe*

We now describe the architecture, norm formal model, and decision-making.

3.1 Architecture

Noe integrates the BDI architecture [29] with a normative model [4, 12, 30] and an emotional model [3, 21]. A *Noe* agent assesses the environment, including other agents' expressed emotions, its cognitive mental states, and infer possible outcomes to make a decision. Figure 1 shows the three components of *Noe*.

The normative component of *Noe* includes the following processes:

– Identification: the agent recognize norms from its norm base based on its beliefs
– Instantiation: activate norms related to the agent
– Normative reasoning process: the reasoning process makes decisions based on the beliefs, current intention, self-directed emotions, other-directed emotions received from others, active norms, and how the norm satisfaction or violation influences the
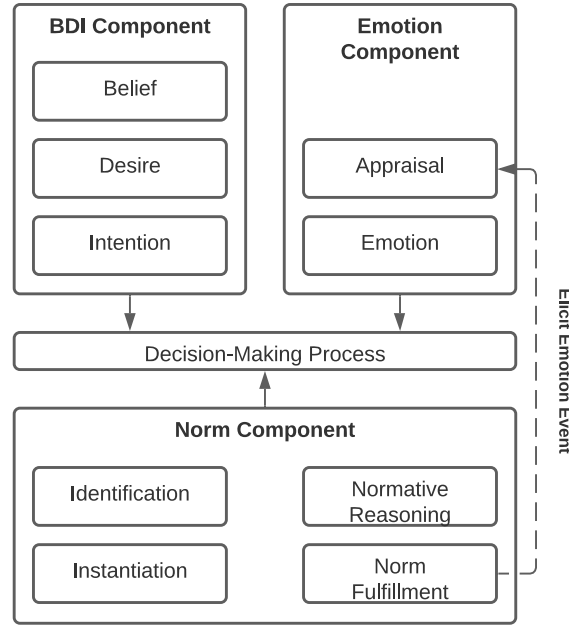
Fig. 1. *Noe* architecture, representing and reasoning over beliefs, desires, intentions, emotions, and norms.

world and itself The *Noe* agents then update the intention based on the results of normative reasoning

– Norm fulfillment process: check if a norm has been fulfilled or violated based on the selected action. The compliance or violation of a norm will then trigger an elicit emotion event that will be appraised at the emotion component

The BDI component includes the following parts:

– Beliefs: form beliefs based on perceptions
– Desires: generate desires based on the beliefs
– Intention: the highest priority of desires to achieve based on the beliefs
– Action: select action based on the current intention, emotions, possible outcomes, and the evaluation of violating or complying with norms, if any
  The beliefs, desires, and intentions are mental states of *Noe* agents.

The emotional component includes the following processes:

– Appraisal: calculate the appraisal value based on the beliefs, desires, and norm satisfaction or norm violation. In this work, we consider only norm satisfaction or norm violation
– Emotion: generate emotion based on the appraisal values [21]

Figure 2 illustrates the interactions between agents in our simulation scenario.
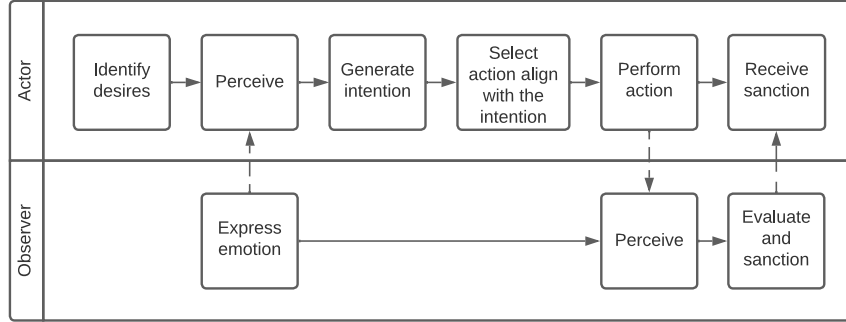
Fig. 2. The interaction between *Noe* agents.

3.2   Norm formal model

Social norms describe the interactions between agents in a multiagent system. We adopt Singh's [35] representation, where a social norm is formalized as $Norm(\text{subject}, \text{object}, \text{antecedent}, \text{consequent})$. In this representation, the subject and object represent agents, and the antecedent and consequent define conditions under which the norm is activated or satisfied, respectively. This representation describes a norm activated by the subject towards the object when the antecedent holds, and the consequent indicates if the norm was satisfied or violated.

Following Singh [35], we consider three types of norms in *Noe*.

- Commitment (C): the subject commits to the object to bring out the consequence if the antecedent holds. Consider Alice and Bob are queuing up in a grocery store. Alice and Bob commit to keeping social distance during the pandemic, represented as $C(\text{Alice}, \text{Bob}, during = \text{pandemic}, \text{social distance})$.
- Prohibition (P): the object prohibits the subject from the consequence if the antecedent holds. Caleb, the grocery store manager, prohibits Bob from jumping the queue while lining up in that store, represented as $P(\text{Bob}, \text{Caleb}, when = \text{line up}; at = \text{grocery store}, \text{jump})$.
- Sanction (S): same as commitment or prohibition, yet the consequence would be the sanctions. Sanctions could be positive, negative, or neutral reactions to any norm satisfaction or violation [27]. If Bob breaks the queue, he receives negative sanctions from Alice, represented as $S(\text{Bob}, \text{Alice}, \text{jump}, \text{negative sanctions})$. Negative sanctions could be physical actions, e.g., scolding someone, or emotional expression, e.g., expressions of disdain, annoyance, or disgust.

To simulate the norm emergence and enforcement in human society, we include emotions into the decision-making process since, by nature, humans do not always act rationally in terms of utility theory. Here we formalize emotions with $E_i(target, intensity, decay)$ indicating agent a_i has emotion e toward the target with intensity and decay value. An example of the prohibition case would be, Bob would not

jump the queue if Alice is angry, represented as $P(\text{Bob}, \text{Alice}, \text{Bob} \succ \text{Alice} \wedge E_{\text{Alice}} = \text{angry}, \text{jump})$.

We model the emotional response of agents with triggered emotions from norm satisfaction, or violation [4]. Here we represent the elicited emotions with $Elem_{name}(A_{expect}, A_{real}, Em_1, Em_2)|Em_1, Em_2 \in E; A_{expect}, A_{real} \in A$ where A is a set of actions. E is a set of emotions, and Em_1 and Em_2 are the emotions triggered by norm satisfaction and violation accordingly. If the A_{expect} is equal to the A_{real}, a norm has been fulfilled, and Em_1 was elicited. $Ap(beliefs, desires, Elem)$ represents the appraisal function.

3.3 Decision-Making

Schwarz [32] addresses the influence of moods and emotions at decision making and discusses the interplay of emotion, cognition, and decision making. Specifically, the aspects include pre-decision affect, post-decision affect, anticipated affect, and memories of past affect. In our model, we include the pre-decision affect into the decision-making process. With pre-decision affect, people recall information from memories that match their current affect [32]. For instance, people in a sad emotion or interacting with hostile people tend to overestimate adverse outcomes and events.

In our model, emotions serve as mental objects and an approach to sanctioning. We consider emotions as intrinsic rewards from agents' internal state in contrast to physical rewards from the environment. We adopt the OCC model of emotions [28], in which we consider emotional valence and intensity. We formulate emotions with simple values where positive values indicate positive emotions and larger values indicate higher intensity. A mood is a general feeling and not a response to a specific event or stimulus compared to emotions. Therefore, we consider emotions but not mood. *Noe* agents' appraisal function considers norm satisfaction and violation only. The agents are aware of other agents' expressed emotions in the same place. In this work, we assume that agents express true and honest emotions and can correctly perceive the expressed emotions. In other words, felt emotions are equal to expressed emotions. Another assumption is that emotions are consistent with the notions of rational behavior.

Algorithm 1 displays the decision loop of our model. At the beginning of the simulation, all agents are initialized with certain desires, and during the run, an intention would be generated by prioritizing desires with the agent's beliefs. When choosing the next move with line 5 in Algorithm 1, the agent chooses the one with maximum utility from all available actions. Algorithm 2 details the action selection. The decision takes the agent's beliefs, current intention, and possible consequences into accounts. While norms are activated with the beliefs, the agent would further consider emotions and cost and possible consequences with norms at line 9 in Algorithm 2. For instance, if people violate some social norms, they may be isolated from society. Regarding the influence of emotions, people may overestimate the negative outcomes when they are in the negative emotion and tend to comply with the norms.

Algorithm 1: Decision loop of a *Noe* agent

---

1 Initialize one agent with its desires D;

2 for *t=1,T* **do**

3   Observe the environment (including the expressed emotions from others E_{around}) and form beliefs b_t;

4   Generate intention I based on b_t and D;

5 a_t = ActionSelection(b_t, I, D);

6   Execute action a_t;

7   Elicit self-directed emotions E_{self} from agent itself based on if action a_t fulfills a norm;

8   Self-sanction with E_{self};

9   Observe the environment (including the performed actions a_{t_other} of other agents) and form beliefs b_{t+1};

10   Elicit other-directed emotions E_{other} for observer agents based on if action a_{t_other} fulfills a norm;

11   Sanction others with E_{other};

12 end

Algorithm 2: Action selection

Input: beliefs b_t, intention I, desires D

Output: Action a_t

1 Function `Action Selection`**:**

2 $E_{around} \subset b_t$;

3   **for** *each a in ACTIONS(b_t)* **do**

4   Activate norms N with beliefs b_t and a;

5   **if** $N = \varnothing$ **then**

6 a_t = MAX$_a$(RESULT(b_t, intention, a))

7   **else**

8 a_t = MAX$_a$(RESULT(b_t, intention, a, N) × amplifier(E_{around}))

9   **end**

10   **end**

11   **return** a_t

12 return

---

4 Evaluation

We evaluate *Noe* via a line-up environment where agents form queues to receive service. We detail the environment in Section 4.1.

4.1 Line-up Environment

Figure 3 shows the line-up environment. We build this line-up environment using Mesa [22], a Python-based framework for building, analyzing, and visualizing agent-based models.
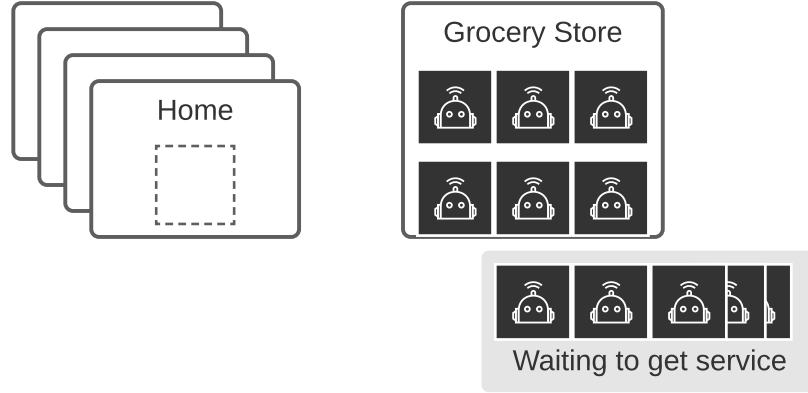
Fig. 3. Simulation details. Agents move between their homes and the grocery store. The store has a capacity limit of eight customers at one time. As a result, other agents must line up outside the store to get service.

The line-up environment includes two shared locations—home and grocery stores. The agents move between home and grocery stores to get food. We consider one social norm in the line-up environment: agents are expected to line up to enter the grocery store. To simulate real human reactions to norm violations, we refer to a social psychology experiment [23]. In the line-up environment, we model defensive reactions of people in the queue as negative emotions toward those who jump the queue by barging in ahead of someone already in the queue. Conversely, people show positive emotions toward those who stay in the queue.

We initialize the agents with the following parameter values:

- Health (Integer value from 0–100): When the health value reaches zero, the agent is marked as *deceased* and unable to act. The health value decreases by 1 unit at each step.
- Deceased (Boolean: True or False): set as True when an agent runs out of health.
- Emotion (Integer value): simplified with numerical values where positive values indicate positive emotions and negative indicate negative emotions. The emotions come along with a duration. Default at 0.
- Number of food packets owned (Integer value from 0–15): once obtained food from the stores, agents would be able to restore its health value via consuming food anywhere.
- Food expiration day (Integer value from 0–15): once the agent gets food packets, we update the expiration day with 15. The expiration day decreases by 1 unit at each step. Food expires once the expiration day reaches 0. Default at 0.
- Beliefs: the perceived and processed information from the world, including other agents' expressed emotions.

– Desires: desired states, including *have food* and *wandering*.
– Intention: the highest priority of desires to achieve at a specific time. When the agent's health is lower than the threshold, 80% of the health, the agent sets its intention as *get food*; otherwise, the agent sets its intention as *wandering*.

When an agent runs low on stock, it has a higher probability of moving to a grocery store. The grocery store can provide food packets to eight agents in one time step. While waiting in line to get food, the agent could either stay in the line or jump ahead in the line to get food in less time. Jumping the line may increase other agents' delay in getting food packets. Those who witness the violation would then cast negative emotions, further interpreted as anger or disdain, triggered by that behavior. To simplify the simulation, we presume the anticipated affects [32] with: (1) receiving negative emotions triggers negative self-directed emotions such as shame and guilt; (2) complying with norms leads to positive or neutral emotions; (3) violating norms leads to negative or neutral emotions. The intensity of emotions triggered each time is fixed, but the values of emotions can add up. Each triggered emotion lasts 2 steps. At each step, the duration and intensity of emotion decrease by 1 as decay. A simple assumption here is that people in a bad mood would trigger stronger emotions in response to a non-ideal state. Note that at the beginning of the simulation, we initialize the agent society with health in normal distribution to avoid all agents having the same intention at the same time.

4.2 Agent Types

To answer our research question and evaluate *Noe*, we define three agent societies as baselines. We describe the agents societies below:

Obedient society. Agents in an obedient society always follow norms.
Anarchy society. Agents in an anarchy society jump lines when they cannot get food.
Sanctioning society. Agents in the sanctioning society jump lines considering the previous experience of satisfying or violating a norm. Agents sanction positively or negatively based on norm satisfaction or violations directly and comply with enforced norms.
***Noe* society.** Agents in the *Noe* society jump lines considering the previous experiences of satisfying or violating a norm, current emotional state of the other agents, current self emotional state, and estimated outcome of satisfying or violating a norm. *Noe* agents who observe norm satisfaction or violations would appraise the norm outcomes and trigger emotions to sanction the actor agent.

Table 1 summarizes the characteristics of the agents in the four societies.

4.3 Hypotheses and Metrics

To address our research question $RQ_{emotion}$ on emotions and norm emergence, we propose three hypotheses:

H_1 (Norm satisfaction): Norm satisfaction in *Noe* agent society is higher compared to the baseline agent societies.

Table 1. Characteristics of the various agent societies.

Agent Type	Violation allowed	Sanctioning	Emotions involved
Obedient society	✗	✗	✗
Anarchy society	✔	✗	✗
Sanctioning society	✔	✔	✗
Noe society	✔	✔	✔

H$_2$ (Social welfare): *Noe* agent society yields better social welfare compared to the baseline agent societies.

H$_3$ (Social experience): *Noe* agent society yields a better social experience compared to the baseline agent societies.

To evaluate H$_1$ on norm satisfaction, we compute one metric, M$_1$ (Cohesion): Percentage of norm satisfaction.

To evaluate H$_2$ on social welfare, we compute two metrics: (1) M$_2$ (Deceased): Cumulative number of agents deceased; (2) M$_3$ (Health): Average health of the agents.

To evaluate H$_3$ on social experience, we compute one metric, M$_4$ (Waiting time): Average waiting time of agents in the queues.

To test the statistical significance of H$_1$, H$_2$, and H$_3$, we conduct the independent t-test and measure effect size with Glass's Δ for unrelated societies [13, 14]. We adopt Cohen's [9, pp. 24–27] descriptors to interpret effect size where above 0.2, 0.5, 0.8 indicate small, medium, and large.

4.4   Experimental Setup

We run each simulation with 400 agents and queue size 80 for 3,000 steps. We choose a relatively small number of agents to reduce the simulation time while our results are stable for a more significant number of agents. The simulation stabilizes at about 1,500 steps, but we keep extended simulation steps to have more promising results. Table 2 lists the payoffs applied in our simulation.

We present the results with a moving average of 100 steps. We choose this size of running window to show the temporal behavior change in a small sequence of time. With a larger size, the running window may alleviate the behavior change. To minimize deviation from coincidence, we run each simulation with 10 iterations and compute the mean values.

Table 2. Payoff table.

Component	Type	Reward
Deceased	Extrinsic	−500
Norm compliance & positive emotion	Intrinsic	1
Norm violation & negative emotion	Intrinsic	−1

4.5 Experimental Results

In this section, we describe the simulation results comparing the three baselines and *Noe* agents. Table 3 summarizes these results. Table 4 lists the value of Glass's Δ and p-values from the independent t-test.

According to Table 3 and Table 4, we see that *Noe* generate better cohesion and fewer deceased agents than baselines ($p < 0.01$; Glass's $\Delta > 0.8$). The null hypothesis corresponding to H_1 is rejected. Note that we do not consider the cohesion metric for the obedient agent society here since agents in the obedient society are always compliant. However, *Noe* also yields the worst social experience where the low waiting time is a desirable state ($p < 0.01$; Glass's $\Delta > 0.8$).

Table 3. Comparing *Noe* agent society with baseline agent societies on various metrics.

Agent Society	Cohesion	Deceased	Health	Waiting Time
Obedient	–	55.30	79.27	8.95
Anarchy	0.22	81.60	79.50	5.45
Sanctioning	0.88	169.30	86.26	2.55
Noe	**0.99**	**54.00**	79.00	8.95

Table 4. Statistical analysis.

Agent Society	Glass's Δ				p-value			
	Cohesion	Deceased	Waiting time	Health	Cohesion	Deceased	Waiting time	Health
Obedient	0.19	0.65	0.01	0.18	0.32	< 0.01	0.98	0.52
Anarchy	102.43	3.10	40.82	0.21	< 0.01	< 0.01	< 0.01	0.46
Sanctioning	13.67	15.53	76.68	3.34	< 0.01	< 0.01	< 0.01	8.45
Noe	–	–	–	–	–	–	–	–

H_1 Norm Satisfaction Figure 4 displays the cohesion, the percentage of norm satisfaction, in the baseline agent societies and the *Noe* agent society. We find that the percentage of norm satisfaction in the *Noe* agent society, average at 99% and p-value < 0.01, is constantly higher than the sanctioning agent society, average at 88% and p-value < 0.01 and Glass's $\Delta > 0.8$. The sanctioning agent society learns to comply with the norm as time goes by. The *Noe* agent society does sanction as well. Yet, considering emotions and the possible outcome makes *Noe* agent society enforce the norm faster than the sanctioning agent society. Specifically, *Noe* agent society enforces the norm at about 100 steps while sanctioning agent society at 1,500 steps.
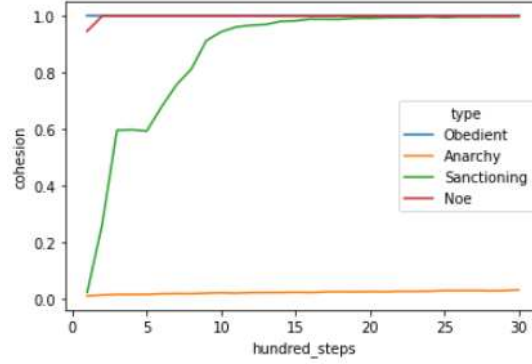
Fig. 4. Simulation result: average cohesion. Comparing average cohesion (M_1) yielded by *Noe* and baseline agent societies.

H_2 Social Welfare Figure 5 compares the average number of deceased in the obedient, anarchy, sanctioning, and *Noe* agent societies. Refer to Figure 4, sanctioning agent society learns the norm via positive and negative sanctioning from norm satisfaction and violation. However, the agents in that society do not consider the possible severe consequences and cause compliant agents to die in the queue. When the number of deceased reaches the threshold, the simulation stabilizes. Therefore, no more agent from the sanctioning agent society dies after the threshold. On the contrary, *Noe* agent society sanctions and considers possible outcomes of norm satisfaction and violation, therefore learning the norm and avoiding unacceptable consequences.
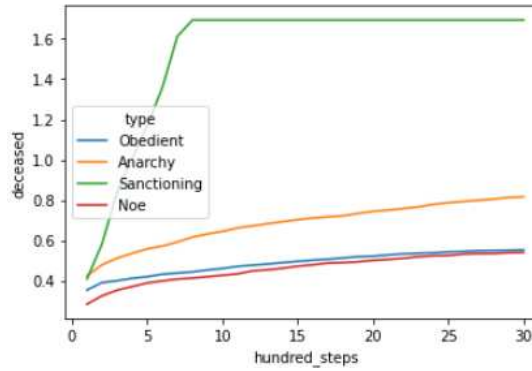

Fig. 5. Simulation result: average number of deceased.Comparing average number of deceased (M_2) in *Noe* and baseline agent societies.

Figure 6 compares the average health of the agents in the obedient, anarchy, sanctioning, and *Noe* agent societies. The sanctioning agent society yields higher health

State, with a mean at 86.26, but at the expense of more deaths. The rest of the agents then be able to remain in high health.
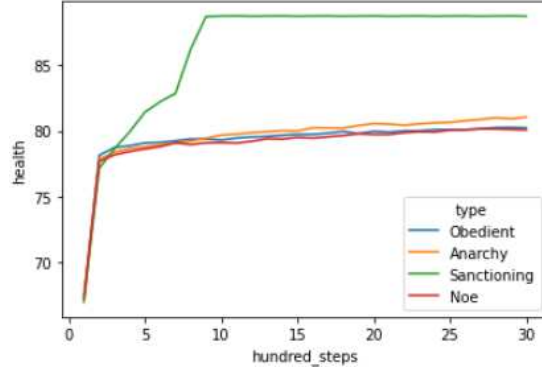

Fig. 6. Simulation result: average health value. Comparing average health value (M_3) in *Noe* and baseline agent societies.

H₃ Social Experience Figure 7 compares the average waiting time the agents spend in a queue at the grocery store in the obedient, anarchy, sanctioning, and *Noe* agent societies. The *Noe* agent society learns the norm fast and remains the same waiting time in the queue. However, some agents in the sanctioning agent society take advantage of those who learn norms faster than themselves. Therefore, many agents die during the learning process, and the simulation stabilizes. In Figure 7, the obedient agent society shares the same trend with *Noe* agent society since emotions enforce the line-up norm.
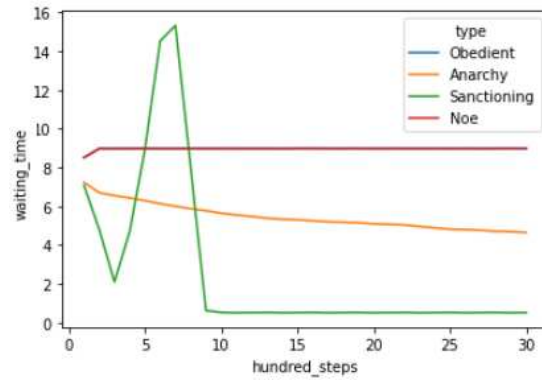

Fig. 7. Simulation result: average waiting time of agents in queues. Comparing average waiting time (M_4) in *Noe* and baseline agent societies.

Combining the results for H_1 and H_2 and H_3, we note that while sanctioning enforces norms, a combination of sanctioning and emotions enforce norms better. Specifically, having emotions as amplifiers of outcomes yield higher norm satisfaction compared to our baselines. The results also indicate that, first, sanctioning agents that consider only norm violation or norm satisfaction may bring out worse social welfare compared to *Noe* that considers both norms and their consequences. Second, although *Noe* agents remain relatively high waiting time in the queues, the number of deceased is lower than the baselines. Note that the sudden drop of deceased number or increase of health value for sanctioning agents resulted from the stabilization of that society. Third, *Noe* agents stay in positive emotions during the simulation while sanctioning agents start from negative emotions and eventually achieve the expected behaviors.

5   Discussion and Conclusion

We present an agent architecture inspired by the norm life-cycle [4], BDI architecture [29], and emotion life-cycle [3, 21] to investigate how emotions influence norm emergence and social welfare. We evaluate the proposed architecture via simulation experiments in an environment where agents queue up to receive service. Our simulations consider two characteristics of an agent society: sanctioning and emotions that participate in action selection and arise from evaluating selected action. The experiments show that incorporating emotions enables agents to cooperate better than those who do not.

In our agent architecture, we make an assumption that agents can recognize others' emotions. However, we acknowledge that emotion recognition is a challenging task [5]. Whereas recent works in AI have focused on emotion recognition through facial expressions and emotion recognition using wearables, it is worth noting that there is no agreement in modeling emotions in the psychology community [5, 19, 20].

Murukannaiah et al. [26] address many shortcomings of current approaches for AI ethics, including taking the value preferences of an agent's stakeholder and other agents' users, learning value preferences by observing the responses of other agents' users, and value-based negotiation. Incorporating these aspects in *Noe* is an interesting future direction.

As a future extension of current work, we plan to differentiate emotions in *Noe* instead of modeling emotions with emotion valences to provide more information for value preferences. We also consider including a mix of personalities in future research to have different appraisal results. In this work, *Noe* agents are assumed to express true and honest emotions. However, emotions can also serve as a tool to influence, persuade, or deceive others in an adversarial context. It would be crucial to identify and model these contradictions while humans are in the loop.

Acknowledgments

Bibliography

[1] Ajmeri, N., Guo, H., Murukannaiah, P.K., Singh, M.P.: Robust norm emergence by revealing and reasoning about context: Socially intelligent agents for enhancing privacy. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI). pp. 28–34. IJCAI, Stockholm (Jul 2018). https://doi.org/10.24963/ijcai.2018/4

[2] Ajmeri, N., Guo, H., Murukannaiah, P.K., Singh, M.P.: Elessar: Ethics in norm-aware agents. In: Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS). pp. 16–24. IFAAMAS, Auckland (May 2020). https://doi.org/10.5555/3398761.3398769

[3] Alfonso Espinosa, B.: Agents with Affective Traits for Decision-Making in Complex Environments. Ph.D. thesis, Universitat Politècnica de València (2017). https://doi.org/10.4995/Thesis/10251/90497

[4] Argente, E., Del Val, E., Perez-Garcia, D., Botti, V.: Normative emotional agents: A viewpoint paper. IEEE Transactions on Affective Computing **13**(3), 1254–1273 (2022). https://doi.org/10.1109/TAFFC.2020.3028512

[5] Barrett, L.F., Adolphs, R., Marsella, S., Martinez, A.M., Pollak, S.D.: Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. Psychological Science in the Public Interest **20**(1), 1–68 (2019)

[6] Bourgais, M., Taillandier, P., Vercouter, L.: Ben: An agent architecture for explainable and expressive behavior in social simulation. In: International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems. pp. 147–163. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30391-4_9

[7] Broersen, J., Dastani, M., Hulstijn, J., Huang, Z., van der Torre, L.: The BOID architecture: Conflicts between beliefs, obligations, intentions and desires. In: Proceedings of the 5th International Conference on Autonomous Agents. pp. 9–16. Association for Computing Machinery, New York, NY, USA (2001). https://doi.org/10.1145/375735.375766

[8] Chopra, A.K., Singh, M.P.: From social machines to social protocols: Software engineering foundations for sociotechnical systems. In: Proceedings of the 25th International World Wide Web Conference. pp. 903–914. ACM, Montréal (Apr 2016). https://doi.org/10.1145/2872427.2883018

[9] Cohen, J.: Statistical Power Analysis for the Behavioral Sciences. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 2nd edn. (1988)

[10] Dell'Anna, D., Dastani, M., Dalpiaz, F.: Runtime revision of sanctions in normative multi-agent systems. Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS) **34**(2), 43:1–43:54 (Jun 2020). https://doi.org/10.1007/s10458-020-09465-8

[11] Edwards, W.: The theory of decision making. Psychological Bulletin **51**(4), 380 (1954). https://doi.org/10.1037/h0053870

[12] Frantz, C., Pigozzi, G.: Modeling norm dynamics in multiagent systems. Journal of Applied Logics – IfCoLoG Journal of Logics and their Applications **5**(2), 491–564 (2018)

[13] Glass, G.V.: Primary, secondary, and meta-analysis of research. Educational Researcher **5**(10), 3–8 (1976). https://doi.org/10.3102/0013189X005010003

[14] Grissom, R.J., Kim, J.J.: Effect Sizes for Research: Univariate and Multivariate Applications. Routledge, Abingdon-on-Thames (2012). https://doi.org/10.4324/9780203803233

[15] Kafalı, Ö., Ajmeri, N., Singh, M.P.: DESEN: Specification of sociotechnical systems via patterns of regulation and control. ACM Transactions on Software Engineering and Methodology (TOSEM) **29**(1), 7:1–7:50 (Feb 2020). https://doi.org/10.1145/3365664

[16] Kalia, A.K., Ajmeri, N., Chan, K., Cho, J.H., Adalı, S., Singh, M.P.: The interplay of emotions and norms in multiagent systems. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI). pp. 371–377. IJCAI, Macau (Aug 2019). https://doi.org/10.24963/ijcai.2019/53

[17] Keltner, D., Haidt, J.: Social functions of emotions at four levels of analysis. Cognition and Emotion **13**(5), 505–521 (1999). https://doi.org/10.1080/026999399379168

[18] de Lima, I.C.A., Nardin, L.G., Sichman, J.S.: Gavel: A sanctioning enforcement framework. In: Proceedings of the 6th International Workshop on Engineering Multi-Agent Systems (EMAS). pp. 225–241. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-25693-7_12

[19] Marín-Morales, J., Higuera-Trujillo, J.L., Greco, A., Guixeres, J., Llinares, C., Scilingo, E.P., Alcañiz, M., Valenza, G.: Affective computing in virtual reality: Emotion recognition from brain and heartbeat dynamics using wearable sensors. Scientific Reports **8**(1), 1–15 (2018). https://doi.org/10.1038/s41598-018-32063-4

[20] Marsella, S., Gratch, J., Petta, P.: Computational models of emotion. In: Scherer, K.R., Banziger, T., Roesch, E. (eds.) Blueprint for Affective Computing, chap. 1.2, pp. 21–41. Oxford University Press (2010)

[21] Marsella, S.C., Gratch, J.: EMA: A process model of appraisal dynamics. Cognitive Systems Research **10**(1), 70–90 (2009). https://doi.org/10.1016/j.cogsys.2008.03.005

[22] Masad, D., Kazil, J.: MESA: An agent-based modeling framework. In: Proceedings of the 14th PYTHON in Science Conference. pp. 53–60. Springer, Austin (2015)

[23] Milgram, S., Liberty, H.J., Toledo, R., Wackenhut, J.: Response to intrusion into waiting lines. Journal of Personality and Social Psychology **51**(4), 683 (1986). https://doi.org/10.1037/0022-3514.51.4.683

[24] Moerland, T.M., Broekens, J., Jonker, C.M.: Emotion in reinforcement learning agents and robots: A survey. Machine Learning **107**(2), 443–480 (2018). https://doi.org/10.1007/s10994-017-5666-0

[25] Morris-Martin, A., De Vos, M., Padget, J.: Norm emergence in multiagent systems: A viewpoint paper. Autonomous Agents and Multi-Agent Systems (JAAMAS) **33**(6), 706–749 (2019). https://doi.org/10.1007/s10458-019-09422-0

[26] Murukannaiah, P.K., Ajmeri, N., Jonker, C.M., Singh, M.P.: New foundations of ethical multiagent systems. In: Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS). pp. 1706–1710. IFAAMAS, Auckland (May 2020). https://doi.org/10.5555/3398761.3398958, Blue Sky Ideas Track

[27] Nardin, L.G., Balke-Visser, T., Ajmeri, N., Kalia, A.K., Sichman, J.S., Singh, M.P.: Classifying sanctions and designing a conceptual sanctioning process model for socio-technical systems. The Knowledge Engineering Review (KER) **31**(2), 142–166 (Mar 2016). https://doi.org/10.1017/S0269888916000023

[28] Ortony, A., Clore, G.L., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press, New York (1988). https://doi.org/10.1017/CBO9780511571299

[29] Rao, A.S., Georgeff, M.P.: Modeling rational agents within a BDI-architecture. In: Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning. pp. 473–484. Morgan Kaufmann, San Francisco (1991). https://doi.org/10.5555/3087158.3087205

[30] Savarimuthu, B.T.R., Cranefield, S.: Norm creation, spreading and emergence: A survey of simulation models of norms in multi-agent systems. Multiagent and Grid Systems **7**(1), 21–54 (2011). https://doi.org/10.3233/MGS-2011-0167

[31] von Scheve, C., Moldt, D., Fix, J., von Luede, R.: My agents love to conform: Norms and emotion in the micro-macro link. Computational and Mathematical Organization Theory **12**(2–3), 81–100 (2006). https://doi.org/10.1007/s10588-006-9538-6

[32] Schwarz, N.: Emotion, cognition, and decision making. Cognition and Emotion **14**(4), 433–440 (2000). https://doi.org/10.1080/026999300402745

[33] Simon, H.A.: The New Science of Management Decision. Harper & Brothers (1960). https://doi.org/10.1037/13978-000

[34] Simon, H.A.: Motivational and emotional controls of cognition. Psychological Review **74**(1), 29–39 (1967). https://doi.org/10.1037/h0024127

[35] Singh, M.P.: Norms as a basis for governing sociotechnical systems. ACM Transactions on Intelligent Systems and Technology (TIST) **5**(1), 21:1–21:23 (Dec 2013). https://doi.org/10.1145/2542182.2542203