Empathy and placebo for autonomous agents

Julien Saunier*, Hazaël Jones† and Domitile Lourdeaux†
*LEPSiS - UMR INRETS / LCPC
†Laboratoire Heudiasyc - UMR CNRS 6599 - Université de Technologie de Compiègne
Emails: saunier@inrets.fr, {joneshaz,dlourdea}@hds.utc.fr

Abstract

Computational modeling of emotion, physiology and personality is a major challenge in order to design believable virtual humans. These factors have an impact on both the individual behavior and the collective one. This requires to take into account the empathy phenomenon. Furthermore, in a crisis simulation context where the virtual humans can be contaminated by radiological or chemical substances, empathy may lead to placebo or nocebo effects. Stemming from works in the multi-agent systems (MAS) domain, we consider that a virtual human has two parts, its mind and its body. The agent is influenced by the mind, but controlled by the environment which manages the empathy and nocebo process. We describe these mechanisms and show the results of several experiments.

Keywords

Multi-agent architecture, Personality, Emotions, Empathy, Placebo

1. Introduction

The context of this research is a project of crisis simulation for the improvement of multi-institutional response to terrorist attacks. We study the case of Nuclear, Radiologic, Bacteriological and / or Chemical attacks in a supermarket. The goal is to propose a training tool for both security and rescue teams. Civilians are virtual autonomous humans / agents represented in a virtual environment. In order to improve the crisis simulation, designers have to take into account personality, emotion and physiology in decision making [1]. Being a training tool, the sequence of events has to be explainable.

Multi-agent systems (MAS) are a way to simulate virtual humans via cognitive agents. There are two main approaches to obtain an intelligent behavior : imitate the cognitive process [2], or manipulate a set of observed behaviors [3]. Our research follows a hybrid path that integrates high level interdependencies in behavior choice, in order both to explain the cognitive factors that lead to simulated situations and to keep the agents' complexity consistent with the simulation needs.

In this article, we focus on the social part of the agents. Collective behavior is not an aggregate of individual behaviors, in particular because of empathy [4] and placebo/nocebo effects [5]. Empathy is “the intellectual identification with or vicarious experiencing of the feelings, thoughts, or attitudes of another” (Dictionary). Basically, it means that the agents influence each other through some kind of affective interaction. A nocebo effect is an ill effect caused by the suggestion or belief that something is harmful.

In section 2, we show the motivations of our architecture. Section 3 describes the agents architecture, and section 4 the inter-agent empathy mechanism and how it can lead to nocebo. We show the results of several experiments in section 5.

2. State of the art

2.1. Architectures

Most of the architectures integrating emotions and personalities focus on facial modeling and/or user-virtual human interaction. Conversely, in our research we are interested in high-level decision process modeling, while biomechanics are managed through ad hoc external modules.

In cognitive modeling, BDI architecture [6] is often used for its intuitive representation of the agent’s reasoning. The reasoning is organised by modules for a clear structuring. However, in the original model, emotions, personality and physiology are not taken in account in the decision process. Noticing this limit, Jiang and al. developed the eBDI architecture [7] that introduce emotion in a BDI architecture, although this approach does not consider personality and physiological aspects.

In emotions modeling, several works have been proposed: Gratch [2] proposes the most accomplished current model for agent’s emotions representation. However, its formalism is complex and fully dedicated to the representation of emotions. As a consequence, this model is not easily adaptable and needs a complex calibration. It is adapted to domains where a subtle individual emotion representation is needed (facial expression representation, dialogue management, ...) for a limited number of agents. However, it cannot be used in our simulation since the objective is the modeling of individual and emergent collective emotions for many agents (up to several thousands). These objectives are the same as Silverman et al. [1]. They propose a complete architecture that considers agent’s emotions, physiology, personality and culture. This architecture is fully integrated and the functional separation of modules is static, in order to experiment unitary tests. This approach is complementary to ours. Firstly, Silverman et al. propose their own dedicated architecture while we evaluate emotion, personality and physiology in a well-known architecture (BDI). Secondly, we also explore how the environment can be exploited to provide empathy mechanisms instead of putting all the complexity in the agents.
DETT (Disposition, Emotion, Trigger, Tendency) agent architecture [8] deals with the link between personality and emotions in a straightforward way. It is based on properties defined in OCC model [9]. In particular, DETT defines tendencies, that is the inclination of an agent to feel and to update its emotions in time. However, the main limit to this approach is that DETT is not explanatory. There is a direct link between emotion and action, but no high-level decision.

### 2.2. Social phenomena

Studies from the psychology field, such as [10], highlight the individual and social dimensions of crisis. At the individual level, disasters engender high levels of stress, which can either be adapted stress or overwhelming stress. Adapted stress mobilizes the mental and physiological capacities, while overwhelming stress exhausts the energetic reserves through one of its four modalities: stuporous inhibition, uncontrolled agitation, individual panic flight, and automatic behavior. Symmetrically, at the collective level, behaviors can be adjusted or maladjusted. Maladjusted behaviors such as panic and violence may arise when the rescue teams are weakly organized and when the victims believe they are poorly informed or treated.

Mutual awareness [11] and empathy are necessary for collective behavior to appear. Empathy is the low-level mechanism which enables the agents to perceive each others’ physical and emotional state. At a higher level, mutual awareness involves a symbolic representation of the activities of the others.

Stemming from works in the multi-agent systems domain, our virtual human decision process is designed as an autonomous agent. The environment has been recently put forward as a first-order abstraction [12] which can encapsulate the responsibility of spreading a part of the agent’s state. Following this principle, the agent has two parts: its mind and its body [13]. The mind contains the decision process and is the autonomous part of the agent. The body is influenced by the mind, but controlled by the environment. One may try to fly, it does not mean one can. Practically, the agent’s state is observable and its accessibility is regulated by the environment.

This model is useful in terms of representation (reactive body, uncertainty of an agent on its contamination state, . . .). A part of the computation generally realized in the agent is delegated to the environment. Hence, the functionalities are clearly separate and the agent architecture is centered on the decision process. In terms of computation costs, encapsulating services in the environment does not increase the global load. In fact, an advantage is to allow the environment to share a part of the computation (e.g. distances) for several agents instead of doing it in each agent, however it may create a bottleneck.

### 3. Agent architecture

We introduce briefly the architecture PEP → BDI which is an extension of the eBDI framework [7]. The goal of this model is to consider emotion, personality and physiology of an agent in its decision process. More details may be found in [14].

Algorithm 1 details steps of perception to action cycle.

**Algorithm 1 : PEP → BDI main loop**

**Inputs:**
- $E_0$ initial emotions, $B_0$ initial beliefs, $I_0$ initial intentions, $P h_0$ initial physiology, $P h_0$ physiological state, $P e_0$ initial personality.
- Personality $P e E$ emotional tendencies, $P e P$ percept tendencies and $P e D$ action tendencies.

1. $E ← E_0$, $B ← B_0$, $I ← I_0$, $P h ← P h_0$, $P e ← P e_0$  
2. While true do:  
3. $B_p ∪ B_e ∪ B_h ← Sense(Env, PeP) ∪ Msg(Env, PeP)$ $\cup Body(Env, PeP)$  
4. $E ← \text{primary emotion update}(E, I, B_e, Ph, PeE)$  
5. $B ← belief\_revision(B, E, I, B_e)$  
6. $Ph ← \text{physical\_state\_revision}(B, E, I, B_e)$  
7. $D ← options(B, I, Ph, PeD)$  
8. $I ← filter(E, B, D, I, Ph)$  
9. $E' ← E$  
10. $E ← \text{secondary emotion update}(E, I, B, Pe)$  
11. If $E' \neq E$ then  
12. $B ← belief\_revision(B, E, I, B_e)$  
13. $Ph ← \text{physical\_state\_revision}(B, E, I, B_e)$  
14. $D ← options(B, I, Ph, PeD)$  
15. $I ← filter(E, B, D, I, Ph)$  
16. $\pi ← \text{plan}(I, actions)$  
17. execute($\pi$)  

Step 1 is agent initialization. Line 2 is the life cycle loop of an agent. Then the agent takes new information (perception, message and body) from the environment (line 3). This new information generates immediate emotions (4), and the agent changes its beliefs (5) in function of its emotions. Physiological informations are updated in the same way as beliefs (6). Then, the selection of desire and intentions (7-8) is similar to the classical BDI scheme except for emotion and physiological influence. Once intentions are selected, the agent updates its emotions again (9-10). If new emotions are different (11), it updates again its beliefs, physiology, desires and intentions (12 to 15). Then, it plans its actions (16) and executes its new plan (17).

Basically, emotions are based on OCC [9]. Emotions are grouped by pairs of opposites, for example, pride and shame. We take into account emotions relevant to our crisis simulation: fear and hope, anger and gratefulness, shame and pride, reproach and trust. In this article, the focus is put on empathy in a terrorist attack simulation. Hence, relevant factors are fear, anger and stress (physiological parameters). The personality is formed by parameters that indicate personality traits [15]. In a crisis situation, the agents evolve on a short time period (a few hours) where only some prominent behavior elements are expressed. As a consequence, the personality of an agent does not evolve during the simulation. We also chose to
simplify the personality model in order to use only personality traits relevant to the simulation: emotional tendencies such as the tendency to repel fear (braveness), social traits such as leadership or trust, and individual traits such as caution. Table 1 sums up emotions and personality traits that we considered relevant to the simulation.

<table>
<thead>
<tr>
<th>Personality</th>
<th>Definition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empathy, Affective link, Altruism</td>
<td>Links to others</td>
<td></td>
</tr>
<tr>
<td>Curiosity, Cautiousness, Bravery</td>
<td>Influence on the agent’s decision-making</td>
<td></td>
</tr>
<tr>
<td>Leadership, Docility, Normativity</td>
<td>Ability to follow/give orders</td>
<td></td>
</tr>
<tr>
<td>Stressability, Nervosity</td>
<td>Resistance to pressure</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Emotions</th>
<th>Definition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joy / Sadness</td>
<td>Occurrence of a desirable/undesirable event</td>
<td></td>
</tr>
<tr>
<td>Fear / Hope</td>
<td>Occurrence of an unconfirmed desirable/undesirable event</td>
<td></td>
</tr>
<tr>
<td>Shame / Pride</td>
<td>Action done by the agent approved/not approved by standards</td>
<td></td>
</tr>
<tr>
<td>Reproach / Trust</td>
<td>Action done by the other approved/not approved by the agents standards</td>
<td></td>
</tr>
<tr>
<td>Anger / Gratefulness</td>
<td>Stemming from sadness and reproach</td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 1 Relevant personality traits and emotions in a crisis situation

4. State spreading and its influence on agents

4.1. Empathy

As stated in the introduction, empathy enables agents to be impacted by other agents’ states. Spatial and/or psychological proximity are requirements for empathy to take place. A basic illustration is given in figure 1. At the beginning, \(a_1\) is unhappy, \(a_2\) and \(a_4\) are happy and \(a_3\) is neutral. In this case, \(a_1\) influences and is influenced by \(a_2\) and \(a_3\). Similarly, \(a_3\) influences and is influenced by \(a_1\) and \(a_4\). While \(a_2\) becomes angry, sharing \(a_1\)’s mood, the state of \(a_3\) is not modified since it is subjected to contrary emotions from \(a_1\) and \(a_4\).

![Fig. 1](image)

Let us note that the order used to calculate the effects of empathy modifies them. In the example, calculating the new state of \(a_1\) before the new state of \(a_2\) leads \(a_1\) to become happy, hence modifying all the chain of calculation. In practice, since we simulate several thousands of agents, we assume that the global equilibrium is not modified substantially by the local effects of the order of calculation.

The functional separation of mind and body means that the agent knows its physical / emotional state, but can only influence it. In fact, its emotions will be updated through three mechanisms:

- **Internal dynamics**: emotions and physiology evolve as time goes by in function of the agents’ personality traits towards an equilibrium.
- **Event dynamics**: emotions and physiology evolve in function of stimuli, *i.e.* the agent perception of the situation, and of the agents’ personality traits.
- **External dynamics**: emotions and physiology vary in function of the other agents and of the sensitivity of the agent.

In this article, we focus on internal and external dynamics. Internal dynamics are managed by the agent itself. It is calculated as:

\[
\text{new_emotion} = \text{old_emotion} \times \text{emotion_tendency}
\]

The same formula manages the internal dynamics of stress. Emotions are stored as numerals with lower and upper bounds, *e.g.* no stress is encoded as 0 and maximal stress is encoded as 10. These formulae mean that, depending on the personality of the virtual human, and especially of its tendencies, it will naturally tend towards either the lower or upper bound, except if the initial emotion is 0. This mechanism is balanced by the event and external dynamics.

External dynamics are managed by the perception function of the MAS environment, in order to give the right information to the right agent(s). Concerning the empathy mechanism specifically, the environment updates regularly the agents’ body state (figure 2).

**Fig. 2** Environment empathy module and agents’ interactions

![Fig. 2](image)
The empathy manager is a module of the MAS environment. It gets (1) the current state of the agent, here \(a_1\). It updates (2) accordingly its state of the world. The state of the world contains the body properties of all the agents. Then, the empathy manager calculates the effects of empathy on the agents’ neighbors in function of their previous state and of their tendency to empathy. Finally, the MAS environment spreads (3) these into the concerned agents’ bodies, \(a_2\) in our example.

The formula is designed to take into account the proximity between agents:

\[
new\_stress = \frac{1}{dist(origin, target)old\_stress} \times t_e
\]

with \(dist(a, b)\) the distance between \(a\) and \(b\), and \(t_e\) the empathy tendency of the target.

We use the same formula to modify the individual emotions, for example fear and anger.

Compared to a computation inside each agent, this modeling through the environment offers two advantages:

- The agent architecture is focused on high-level decision, while the environment takes a part of the agent complexity which concerns low-level mechanisms.
- The environment can re-use its computation results for several agents, and in particular the distances, instead of having each agent compute its distance to the other agents it perceives.

### 4.2. Placebo

In our crisis scenario, part of the civilians inhale and/or touch chemical and radiological substances. These civilians quickly develop symptoms, which range from itching to suffocation. A major difficulty for the rescuers is to isolate the contaminated victims. However, a part of the population will mimic reactions to the toxics without real exposure [16]. Four factors are involved in the nocebo effect: personality, physiology, emotions and beliefs.

Studies show that in clinical experiments 30 to 55 % of the patients in hospital settings can feel placebo. Concerning the nocebo, Harrington [17] estimates that up to 55% of the patients in hospital settings can feel nocebo. These events are:

- Information or rumor about the presence of a contaminating substance
- Addition of a new belief evaluating an agent to be contaminated
- Survival prediction belief crossing a lower threshold

Stress and emotions are numerals, emotion tendencies belong to \([1 - t, 1 + t]\). The calculation is as follows:

\[
threshold(contam(itself)) = \left(\sum_{ag\in\text{agents}} \frac{1}{dist(itself, ag)contam(itself) + stress \times pess} \right) \times t_e \times p
\]

with \(contam(itself)\) the observable contamination of an agent through its symptoms, \(pess\) its pessimism and \(p\) the probability to be subjected to nocebo.

When an event triggers this calculation, a random number is generated and compared with the threshold. If \(threshold(contam(itself))\) is crossed, the agent believes it is contaminated and starts to mimic the symptoms of nearby seemingly contaminated agents. Doing so, it triggers the same calculation for its neighbors while increasing the probability their result is positive.

### 5. Experiments

We have run experiments using the MadKit\(^1\) platform. MadKit is a general-purpose multi-agent system platform written in Java.

To the best of our knowledge, there is no real data available in order to validate numerically our empathy model. Studies on crisis situations describe the phenomenon, but do not quantify it. Concerning the nocebo, Harrington [17] estimates that up to 55% of the patients in hospital settings can feel placebo.

1. http://www.madkit.org
effects for pain relief. However, we do not know how this figure relates to panic situations.

Hence, the experiments are designed to validate qualitatively the model and show its properties in terms of appearance of consistent behavior and of stability / evolution of the model in function of its parameters.

In these simulations, the agents are pseudo-randomly located in a two-dimensional space. Except when mentioned otherwise, there are $12 \times 12$ agents and the simulations are run 100 times. Half of the agents have a tendency to get stressed over 1. The darker the point is, the higher the agent’s stress is.

5.1. Empathy

Figure 3 shows the simulation evolution over one minute. It starts just after a stressful event occurred (such as an explosion). Consequently, the agents have various levels of stress, depending on their perception, their personality and how they assessed the situation. After one minute, the agents’ stress is stabilized. The top right screenshot shows that clusters of agents with the same stress level are forming, because the agents are sensible to the state of their neighbors. The bottom right screenshot shows what happens if no empathy mechanism is implemented: only the stress tendency of the agent is taken into account, and no collective phenomenon can take place.

This experiment shows that the integration of the empathy mechanism enables the emergence of homogenous groups of agents despite heterogeneous personalities, which is typical of crowd behavior.

Then, we evaluate the behavior of the mechanism over time. Figure 4 shows the proportion of stressed agents in function of time, for 4225 agents. An agent is counted as stressed if its stress value is over 1. The proportion of stressed agents quickly decreases over the first minute to 35%, and then reaches an equilibrium around 32%. The mechanism tends quickly towards its equilibrium and is stable over time when there are no events.

We also study the impact of the initial proportion of agents stressed by the events. Figure 4 represents the stabilized proportion of stressed agents in function of the initial proportion of stressed agents. The experiments show that under 30% or over 80%, the empathy mechanism tends to unify the stress level of all the agents. Between these thresholds, clusters of agents (such as shown in Fig. 3) are forming.

For an initial proportion of 50%, there is on average 32.7% of the agents whose stress level is impacted (positively or negatively) by the empathy mechanism. The stress deviates by more than 30% in 2/3 of these impacted agents, while the last third is only slightly impacted.

Finally, we have tested the stability of our results in function of the number of agents (from 100 to 3000). It shows that the number of agents does not have a major impact on the mechanism because the proportion fluctuates by less than 1%.

We also study the impact of the initial proportion of agents stressed by the events. Figure 4 represents the stabilized proportion of stressed agents in function of the initial proportion of stressed agents. The experiments show that under 30% or over 80%, the empathy mechanism tends to unify the stress level of all the agents. Between these thresholds, clusters of agents (such as shown in Fig. 3) are forming.

5.2. Contamination / nocebo effect

Figure 5 shows the simulation evolution when nocebo effects are added. The points are shadowed when the agents show symptoms of contamination.

Only the agents showing symptoms at the beginning of the simulation are really contaminated ($\frac{1}{10}$ of the agents). In these experiments, the agents do not move. Hence, events triggering the calculation are only (i) after the initialization, once the agents perceive their surroundings (and therefore the contaminated agents in their area), and (ii) the agents showing nocebo-induced symptoms.

The results of our experiments are consistent with the principles discussed in section 4.2. Nearly one half of the agents exhibiting symptoms are not contaminated. Immediate proximity is a major factor, combined with stress and pessimism. Finally, we note the high correlation between symptoms spreading and clusters of agents whose stress is important.

6. Conclusion and perspectives

In this article, we have proposed mechanisms to model empathy and nocebo. The emotions of the agents are calculated thanks to a conjunction of three dynamics: internal, event-driven and external. External dynamics represent the empathy
and enable the agents’ internal states (stress, emotions) to influence each other. The nocebo effect calculation builds on the empathy mechanism to enable the emergence of false beliefs, which in turn influence the body state.

An originality of this work is to delegate to the environment the task of spreading the states of the agents. The advantages are (i) an agent architecture fully dedicated to the decision process and (ii) the reuse of a part of the computation. Combined with the agents’ decision process, the empathy mechanism enables the representation of complex social behavior.

Experiments have shown the soundness of our environment for empathy and nocebo mechanisms. Further works should include validation of simulated behaviors. Because of the lack of real-world data, this work of calibration will require expert feedbacks.

References


