

An Evaluation of the COR-E Computational Model for Affective Behaviors

Sabrina Campano, Nicolas Sabouret, Etienne de Sevin, Vincent Corruble
Université Pierre et Marie Curie
4, place Jussieu
75005 Paris, France
name.surname@lip6.fr

ABSTRACT

The simulation of believable behaviors for virtual agents requires to take human factors such as emotions into account. Most computational models dealing with this issue include emotion categories in their architecture. However, determining categories to use and their influence on behavior is a difficult task. In order to address this challenge, our COR-E model uses an architecture without emotion categories. In this paper, we present an evaluation of this model in the context of a waiting line scenario. We show that COR-E can produce believable emotional behaviors, and test the contributions of the various components and characteristics of its architecture to these positive results.

Categories and Subject Descriptors

I.2.11 [Computing Methodologies]: Artificial Intelligence—*Distributed Artificial Intelligence*

Keywords

affect, emotion, believability, behavior, virtual agent

1. INTRODUCTION

Emotions have been at the core of many psychological studies for several decades [13]. This topic gave rise to numerous computational models of emotion, either aiming at the simulation of lifelike agents, or at the study of psychological processes [17, 16, 8, 7]. Most existing computational affective models rely on a number of numerical emotion variables that must be parameterized by hand so as to produce believable affective responses and behaviors [16, 8, 7]. Finding the correct value of these parameters and the influence of each one on the general model is a significant challenge. As an example, if we consider the emotion of anger felt by an agent, we may have to define an intensity value (say, from 0 to 1) for this emotion, and then we need to determine the effect of this value on the agent's behavior.

Other approaches, such as Pfeifer's work [22] or the Mi-croPsi model [6], aim at obtaining emotional behaviors without using emotion variables. In these models, emotions are considered as an emergent phenomenon. The COR-E model

(COnservation of Resources Engine) evaluated in this paper enters this second category of models. COR-E's architecture intends to produce *emotional* behaviors (i.e. behaviors that will be described with emotion terms by a human observer) without using emotion variables, parameters, dimensions or categories in the model itself. The model is based on the psychological theory of "Conservation of Resources" proposed by S. E. Hobfoll [11], that until now had not yet lead to a computational model nor an implementation.

The outline of this paper is organized as follows: first we present some related work that includes the background on which COR-E was designed, as well as a section on the evaluation of affective models (section 2). Then we present COR-E (section 3), and we explain the evaluation protocol used in order to evaluate this model (section 4). We detail the results that were obtained (section 5), and finally we discuss them (section 6).

2. RELATED WORK

2.1 Background of COR-E Model

Computational models of emotion are numerous, and their architecture and functionalities differ depending on their objectives and methods. The *cognitively realistic approach* aims at the implementation of psychological theories in order to reproduce human mental processes [20, 9], and the *believable approach* aims at lifelike agents, not necessarily realistic, suitable for entertainment or serious games applications [8]. These models have a common ground: they use emotions in their architecture, either as discrete entities or as continuous representations.

Appraisal theory, as formulated by R. Lazarus [14], played a significant role for computational models of emotion. Appraisal is seen as a cognitive evaluation resulting in an emotion, based on which different strategies can be adopted. The OCC model [20] focused on the process of appraisal, and determines which emotion among 22 categories is felt by an agent. This is done through the evaluation of *appraisal variables*, such as the *desirability* or the *likelihood* of an event. The EMA model [9] focused on *coping strategies* adopted in order to deal with a given emotional state. Examples of such strategies are *denial*, *acceptance*, or *wishful thinking*. In OCC and EMA emotions are represented as discrete entities, whereas in the PAD model [19] they are represented with three continuous dimensions which are *pleasure*, *arousal*, and *dominance*. As pointed out in [17] discrete and continuous models are used for different purposes: discrete emotions can be associated with specific behaviors [7], while

dimensional models offer more flexibility, to determine for example the spatial extent of a gesture.

However, the use of emotions, either as discrete categories or continuous dimensions, has limitations when it comes to the simulation of various behaviors. Authors of the OCC model [20] notice that “*the same behavior can result from very different emotions*” and “*very different behaviors can result from the same emotion*”. It is not possible to associate an emotion label or a given point in space with a unique behavior. In the *Affective Reasoner* [7], several actions, such as the somatic responses *flush* or *tremble*, are linked with one emotion label. When an emotion is activated, the selection among its associated actions is made by a filter depending on the agent’s personality.

Some psychological theories argue that the recognition of a felt emotion is made from an interpretation of somatic responses, such as ‘flush’, ‘tremble’, or the general autonomic arousal [13, 24]. According to W. James: “*If we fancy some strong emotion, and then try to abstract from our consciousness of it all the feelings of its bodily symptoms, we find we have nothing left behind*”. From this point of view, this is not an emotion that causes particular responses, but rather the responses that are responsible for the interpretation of an emotion. If this principle also applies to behaviors like running away, then it should be possible to trigger behaviors judged as emotional without using emotion variables.

A more general problem for researchers in affective computing is that there seems to be no consensus on the number of existing emotions, neither on their role or consequences on cognition and behavior [21, 25]. Some work show that emotion categories are culture specific, and that even the categories of *fear* and *anger* are not universal [23]. According to L. F. Barrett [2], “*the lack of coherence within each category of emotion is empirically the rule rather than the exception*”. The author suggest that if no set of clearly defined emotional patterns has been found, it may be because emotions are concepts instead of being distinct “*natural kinds*” of our affective system. That is to say that human beings experience emotions in the same manner as they experience colors, they use their knowledge to label their perceptions with categories.

Actually, another approach which does not use explicit representations of emotions is possible. Emotions can be viewed as an emergent phenomenon, resulting from behavior instead of causing it. This approach was exemplified by the work of R. Pfeifer [22], whose motivation relied on the “*frustrations*” suffered by computer scientists working on emotions. Pfeifer points out numerous problems associated with the use of emotion categories, including *overdesign*, which is the tendency to conceive a system too complex for its objectives. Instead of using emotion variables, Pfeifer proposes to conceive a creature with a simple design, and then observe if this design is sufficient in order that a human observer recognize emotions in this creature’s behaviors. The creature could collect ore in order to gain energy, and avoid obstacles. It turned out that human observers effectively attributed emotions to the creature’s behaviors, saying that it was “*frustrated*” or “*annoyed*”.

Pfeifer’s approach was applied to a simple agent in a limited environment, and was not validated with an evaluation protocol. The author points out the need to pursue this approach in increasing the complexity of agents and environments [22]. In order to apply this approach to virtual

agents, it is necessary to design an architecture capable of handling various behaviors, from basic ones to social ones. The theory of Conservation of Resources (COR) by psychologist S.E. Hobfoll [11] offers an interesting lead in this direction. In this theory, the drive for the acquisition and protection of resources is at the core of the dynamics which explains the stress or well-being of an individual. The concept of resource refers to many types of elements: social and psychological ones such as self-esteem or caring for others, material ones such as a car, or physiological ones such as energy. The key principle is that individuals strive to protect their resources, and to acquire new ones, and this can be easily linked to behaviors.

The COR-E model (*Conservation or Resources Engine*) [5] evaluated in this paper is inspired from the COR theory. It aims to simulate believable affective behaviors without using emotion variables. In this model, behaviors are associated with resource types, instead of being associated with emotion categories. For example, in order to acquire a position (*desired resource*) in a waiting line, an agent can jump the queue (*acquisitive behavior*), at the risk of loosing its reputation (*acquired resource*). COR-E model is presented in section 3 in further details.

2.2 Evaluating Affective Models

The evaluation of an affective model raises some difficult issues. It is often impossible to rely on some fully objective criteria, since an affective model does not aim at obtaining optimal results on a purely rational criterion. As an example, finding a criterion in order to evaluate an algorithm for the shortest path problem seems quite easy: one can compare the results with the optimal length, or assess if the algorithm execution time is better than other algorithms. When the aim of a computational model is to simulate believable behaviors, or to elicit emotions in agents in the same way that they are elicited in human beings, such rational criteria often do not exist. In this situation, a reasonable alternative is rely on self reports made by human subjects about abstract concepts, such as the *believability* of an agent’s behaviors. This method consists in conceiving a questionnaire, submitting it to human subjects, and analyzing the subjects’ answers with statistical tests in order to invalidate or support hypotheses formulated about the model.

In order to evaluate the EMA model, two empirical evaluations were carried out [10, 18]. The first evaluation aimed at comparing the coping strategies selected by the model with the coping strategies chosen by human subjects. The second evaluation aimed at comparing the emotions appraised by the model with the emotions appraised by human subjects. The context was a competitive board game with monetary gains and losses, that were expected to elicit emotions and coping strategies. Participants were asked to rate 5 emotions on a scale from 0 to 100, in order to assess the intensity of the emotional feeling that they experienced. Emotions included fear, joy, sadness, anger and hope. The evaluation also involved measures such as participants perceptions of winning utility and likelihood. These results were compared with the predictions produced by EMA. This kind of evaluation, based on a scale questionnaire about psychological concepts, seems appropriate to evaluate whether emotions are recognized by human observers, and also whether agents’ behaviors are judged as believable.

In other cases affective models can be evaluated with an

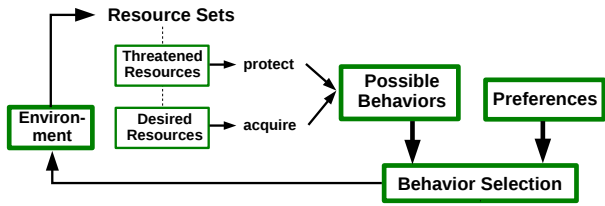


Figure 1: COR-E General Architecture.

objective criterion without requiring a questionnaire. This is the case for example when the task of a model is to reproduce human spatial navigation with virtual agents. Bosse et al. [3] proposed a model that attempts to simulate agents' movements during a panic situation. In order to do so, they tried to reproduce a panic event that happened (and was video-recorded) on Dam square in Amsterdam on a remembrance day. The model uses a variable representing a mental state that contaminates people located in the vicinity, resulting in a contagion effect. To evaluate their model, the authors used the error in meters between the real positions of individuals during the actual event and agents' positions produced by the model. The model presented in this paper does not aim at reproducing agent moves, and thus this kind of evaluation is not appropriate.

Our objective is to assess whether agents' behaviors are believable, and whether human observers recognize emotions in it. These abstract concepts necessarily imply an evaluation based on a questionnaire. This is why we present such an evaluation in section 4.

3. THE COR-E MODEL

3.1 General Principle

COR-E is based on the principle that an agent tries to protect and acquire resources (which can be of a psychological or material nature) that it values when they are respectively threatened or desired. Each resource is associated with 2 sets of behaviors: protective behaviors and acquisitive behaviors. Each agent has individual preferences over resources, which determine the value of a resource from the agent's point of view. A value is computed automatically for each behavior according to these preferences and to the behavior's effects. Behavior selection is guided by two considerations: behaviors' value, and the priority given to protection over acquisition (w.r.t. the first principle of COR theory [12]). The general architecture of the model is shown in figure 1.

Let $\mathcal{R} = \{r_1, r_2, \dots, r_n\}$ be a finite set of resource instances, $\mathcal{T} = \{ty_1, ty_2, \dots, ty_m\}$ a finite set of resource types, and $\mathcal{A} = \{a_1, a_2, \dots, a_k\}$ a finite set of agents. A *resource type* determines the behaviors that can be triggered for a resource (see section 3.2). The unique type of a resource r is denoted as $type(r) \in \mathcal{T}$. Example: $type(reputation_1) = Reputation$.

Each agent $i \in \mathcal{A}$ has 3 *resource sets*:

- $DR_i(t)$: resources desired by i at time t ;
- $AR_i(t)$: i 's acquired resources at time t ;
- $TR_i(t)$: i 's threatened resources at time t .

A resource can be threatened because of another agent's behavior. For each $r \in TR_i(t)$, we associate a cause denoted

as $cause(r, TR_i(t))$, representing another agent's behavior. Example: an agent i that has the second rank in a waiting line, denoted as $rang_2 \in AR_i(t)$, perceives that its resource is threatened when an agent j tries to take it. The cause of this threat is j 's behavior.

An agent has *preferences* over resource instances. For each agent $i \in \mathcal{A}$, we define a total preference order \succeq_i^R on the domain of resource instances \mathcal{R} . The value of a resource r for an agent i , denoted as $v(r, i)$, depends on the ranking of r in agent's preferences order \succeq_i^R . Let $rank(r, \succeq_i^R)$ the ordinal rank of a resource r in \succeq_i^R , and $maxRank(\succeq_i^R)$ the maximum rank in \succeq_i^R . The rank of the most preferred resource is 1. Then $v(r, i) = rank(r, \succeq_i^R) - maxRank(\succeq_i^R) + 1$. This means that the more a resource is preferred by an agent i , the higher the value of the resource is to i .

3.2 Behaviors

Let $\mathcal{B} = \{b_1, b_2, \dots, b_n\}$ be a set of behaviors.

Each behavior b has *preconditions*. We note $Poss(b, t)$ the predicate that is *true* if and only if all preconditions of b are verified at time t . In order that an agent may trigger the behavior b at time t , $Poss(b, t)$ must be true.

A behavior b has 4 sets of *effects* over some agents' resource sets, from an agent's point of view. The set of agents concerned by such effects is denoted as $ptnts(b) \in \mathcal{A}$. $\forall i \in ptnts(b)$, the effects of b from j 's point of view a time t are:

- $R_b^+(i, j, t)$, resource instances *acquired* for i ;
- $R_b^o(i, j, t)$, resource instances *protected* for i ;
- $R_b^*(i, j, t)$, resource instances *threatened* for i ;
- $R_b^-(i, j, t)$, resource instances *lost* for i ;

As an example, if j considers the behavior of protesting against i , denoted as $p(i, j)$, j can anticipate that $p(i, j)$ will threaten i 's reputation. More formally, this is denoted as $R_{p(j,i)}^*(i, j, t) = \{reputation_i\}$.

Behaviors are then organized in two subsets: *acquisitive* and *protective* behaviors. Let $ty \in \mathcal{T}$ be a resource type, B_{ty}^+ is the set of acquisitive behaviors and B_{ty}^o is the set of protective behaviors for resources of type ty .

3.3 Behavior Selection

In COR-E, behavior selection implies to compute a value for each possible behavior, and to take into account the priority of behavior types (acquisitive or protective).

3.3.1 Possible Behaviors

When an agent i desires a resource r at time t , *i.e.* $r \in DR_i(t)$, i has the possibility to trigger any behavior $b \in B_{type(r)}^+$ such that $Poss(b, t)$ is true, in order that i acquires resource r . Similarly, $\forall r \in TR_i(t)$, i can trigger any $b \in B_{type(r)}^o$ such that $Poss(b, t)$ is true, in order to protect r .

3.3.2 Computation of Behavior's Value

The value of a behavior b for an agent is computed with the 4 sets of behavior's effects. Each resource which is threatened or lost because of b is counted as a negative value, and each resource which is acquired or protected thanks to b is counted as a positive value.

We denote $R_b^+(i, j, t) = R_b^+(i, j, t) \cup R_b^o(i, j, t)$ the set of resources counted as a positive value, and $R_b^-(i, j, t) =$

$R_b^-(i, j, t) \cup R_b^*(i, j, t)$ the set of resources counted as a negative value. The value at time t of a behavior b for an agent j , denoted $V(b, j, t)$, is computed as:

$$V(b, j, t) = \sum_{i \in \text{ptnts}(b)} \sum_{r \in R_{b^+}(i, j, t)} v(r, j) - \sum_{r \in R_{b^-}(i, j, t)} v(r, j)$$

3.3.3 Selection

The selection of a behavior in a set of behaviors is based on behaviors value and behaviors types (acquisitive or protective). At each time step for each agent $i \in \mathcal{A}$, the main algorithm follows these two stages:

(1) if there is a resource that i values that is being threatened, i tries to trigger a protective behavior for this resource; if a resource of i is threatened because of an acquisitive behavior of i itself, then i stops its behavior if the threatened resource is preferred to the desired resource (see explanations related to step 1 below);

(2) if no protective behavior has been triggered by i in this time step and if there is a desired resource that i values, i tries to trigger an acquisitive behavior for this resource.

Explanations related to step (1): if i tries to acquire a resource which is an acquired resource for an agent j , then this acquisitive behavior may cause a threat for i . Indeed, to protect its acquired resource, j can threaten a resource acquired by i . The agent i then knows that the cause of this threat is its current acquisitive behavior.

For each of the 2 steps above, the selected behavior is the one with the maximum positive value for i . If there are several behaviors of maximum value, then a behavior is chosen randomly among them. Protective behaviors have priority on acquisitive behavior according to the first principle of COR theory [12]: “resource loss is disproportionately more salient than is resource gain”.

4. EVALUATION

In order to evaluate COR-E, we recorded videos clips of agents simulated by the model. Then we asked some human participants to answer an online questionnaire about these videos. The evaluation had three main objectives: (i) to determine whether agents’ behaviors simulated by COR-E are considered as *believable* and *emotional* by human observers; (ii) to validate the main characteristics of COR-E’s architecture: the key distinction between acquisitive and protective behaviors, the definition of preferences over resources, and the use of psychological resources (in this study, the *reputation* resource); (iii) to test a possible extension of the model, the dynamic threat level, in measuring its impact on agents’ believability.

The main principle of the dynamic threat level is the following: the more an agent loses resources because of other agents, the more its protective behaviors against other agents threaten important resources. For example, instead of protesting against an agent j (hence threatening a *Reputation* resource of j), an agent i can threaten to punch j (hence threatening a *Health* resource of j).

4.1 General Hypotheses

We tested five main hypotheses:

H1: COR-E produces believable and emotional behaviors.

H2: acquisitive and protective behaviors are necessary to obtain behaviors recognized as emotional.

H3: well configured preferences are necessary in order to obtain believable behaviors.

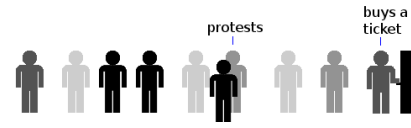


Figure 2: Screenshot of a Simulation with COR-E.

H4: the simulation of the psychological resource *Reputation* increases behaviors’ believability.

H5: the dynamic threat level increases behaviors’ believability.

The goal of COR-E is to produce believable and emotional behaviors. The first hypothesis was used to assess if such goal was attained, and the other hypotheses were used to verify how different features of the model impact the recognition of emotional (*H2*) or believable (*H3- H5*) behaviors.

4.2 Video Clips

Each video clip is a recording of a simulation produced with COR-E in the scenario of a waiting line. These video clips were recorded with the MASON simulator [15], in which COR-E has been implemented. A total of 11 videos of 40 seconds each were produced. The agents were represented with icons in a waiting line of one column, with a ticket counter at the beginning of the queue (see Figure 2).

Each agent had available the following behaviors:

1. to pass another agent in the queue;
2. to protest against an agent passing another agent in the queue, with one of the following textual indications: “protests”, “protests violently”, or “threatens to punch the person”;
3. to wait in the queue, when none of the behaviors above was chosen.

When an agent arrived at the head of the queue, it automatically performed the behavior “buys a ticket”, indicated by a text over its head. Once the ticket was purchased, the agent left the queue, leaving the first rank to the next agent. An agent entered the simulation after a certain period of time (between 4 and 10 seconds).

The configuration at the beginning of the simulation was the following: each agent $i \in \mathcal{A}$ had in its acquired resources $AR_i(t)$: 3 *Reputation* resources of level 0, 1 and 2 (the higher the level is, the higher the reputation is important), and 2 *Health* resources of level 0 and 1. Besides, if an agent was in the queue at the start of the simulation, it automatically acquired a *Rank* resource of which number corresponded to its current rank in the queue (the closer this number was to 1, the better the rank was).

An agent i anticipated behavior’s effects as follows: when i was evaluating the behavior of passing an agent j , then i was planning to gain j ’s rank, and also to lose a *Reputation* resource chosen randomly from $RA_i(t)$. It means that if i preferred all of its *Rank* resources to any *Reputation* resource, then i could not pass another agent in the line (behavior’s evaluation by i was negative). When an agent i intended to protect its rank against an agent j , i could chose which resource of j to threaten (a level 1 or 2 *Reputation* resource, or

a level 1 *Health* resource). This choice depended on i 's current threat level: the more the threat level was important, the more i threatened an important resource. The possible protective behaviors depended on the threatened resource: "protest" for a level 1 *Reputation* resource, "protests violently" for a level 2 *Reputation* resource, and "threatens to punch the person" for a level 1 *Health* resource.

4.2.1 Conditions Used for the Production of Video Clips

Some video clips were produced with the characteristics of COR-E (*COR-E's normal configuration*), and other were produced with a missing component, or with a parameterization estimated as incorrect. *COR-E's normal configuration* has the following properties:

1. all agents can perform acquisitive and protective behaviors;
2. preferences are configured as follows:

For 3 agents out of 4, a *Health* resource was always preferred to a *Reputation* resource, and a *Reputation* resource was always preferred to a *Rank* resource (*i.e.* the agent could not pass any agent). For 1 agent out of 4, a *Health* resource was always preferred to a *Reputation* resource, but some *Rank* resources were randomly preferred to some *Reputation* resources (the agent could possibly pass another agent);

3. the threat level of an agent was set at the lowest value, with a probability of 0.5 to increase when the agent lost a *Rank* resource after being passed by another agent;

4 groups of videos clips were produced with different conditions :

- **group 1:** 3 video clips were produced for this group related to hypothesis $H2$. In video clip **1a**, acquisitive and protective behaviors were disabled (condition $NA \wedge NP$). In video **1b**, acquisitive behaviors were activated, but not protective behaviors (condition $A \wedge NP$), and in video **1c** acquisitive and protective behaviors were activated (condition $A \wedge P$). There was no video clip for the condition $NA \wedge P$, because if acquisitive behaviors are disabled, there is never any threatened resource, hence no protective behavior.
- **group 2:** 3 videos were produced for this group related to hypothesis $H3$. In video clip **2a** agents' preferences correspond to the *COR-E's normal configuration* (condition C). In video **2b**, agents's preferences were configured so that a *Rank* resource was always preferred to a *Reputation* resource (condition $NC1$). In video **2c**, agents' preferences were configured randomly (condition $NC2$). However, the preference order between resources of the same type was preserved (*e.g.* a rank i was always preferred to a rank $i + 1$).
- **group 3:** 3 videos were produced for this group related to hypothesis $H5$. An agent was colored in green, and preferences were configured so that this agent could not pass other agents. Other agents' preferences were configured so that they often pass other agents. In video **3a**, the threat level of the green agent was constant with no increase (condition $M1$). In video **3b**,

the threat level of the agent had a probability of 0.5 to increase (condition $M2$), and in video **3c**, it had a probability of 1 to increase (condition $M3$).

- **group 4:** 2 videos were produced for this group related to hypothesis $H4$. In video **4a**, the *Reputation* resources were simulated (condition R), but in video **4b** they were removed from the environment, preferences and resource sets (condition NR). Agent's preferences were configured so that some *Rank* resources were preferred to some *Health* resources.

Video clips **1c**, **2a** and **4a** were produced with the *COR-E's normal configuration*, and are used to test the general hypothesis $H1$.

4.3 Protocol

A link to the online questionnaire was sent through a mailing list of potential participants. All the participants evaluated all the video clips. The questionnaire included four pages (one page per video group). For each video, a participant had to answer the following questions:

- Q1. *Are characters' behaviors believable?* Possible answers: *totally disagree / disagree / disagree a little / neutral / agree a little / agree / totally agree / no opinion*;
- Q2. *Are these behaviors related to characters' emotions?* Possible answers: *yes / no / no opinion*;
- Q3. *If yes, which ones?* Possible answers: *anger / fear / sadness / joy / disgust / surprise / pride / shame / contempt / love / hate / boredom / frustration / other / none*. A participant could select up to 3 categories;

The order in which the answers were proposed is the same as in the list above. For video clips of page 3 (group 3), participants were asked to evaluate the behavior of the *green* character only. At the end of each page, there was an empty space left for free comments. Participants were asked to not pay attention to agents' graphical appearance.

Participants could see each video of the same page several times. However, they could not go back once the page was validated. The order in which the pages appeared was random, as the order of the video clips on the same page.

Participants: 113 participants answered the online questionnaire. They were aged from 13 to 72 years (mean 33.69 years), and their country was France for a large majority of them (about 94 %).

5. RESULTS

We present in this section the operational hypotheses and the results. The mode (value that occurs most frequently) is denoted as Mo , the mean as μ , and the standard deviation as σ . We used two statistical tests, *Student's t-test* (result denoted as t), and the *chi-square test* (result denoted as χ^2).

5.1 Hypotheses on COR-E model

These hypotheses are related to video clips **1c**, **2a**, and **4a** produced with the *COR-E's normal configuration* (explained section 4.2).

H1a: *agents' behaviors simulated by COR-E are assessed as believable.*

Video clip	μ	σ	Mo
1a	6.28	1.11	7
1b	3.41	1.78	2
1c	5.08	1.51	5
2a	6.01	1.15	7
2b	3.17	1.82	1
2c	2.90	1.70	1
3a	3.54	1.84	2
3b	4.19	1.78	5
3c	4.95	1.51	5
4a	5.28	1.31	5
4b	3.22	1.60	2

Table 1: Results on believability score for each video clip.

On a Likert’s scale from 1 (*totally disagree*) to 7 (*totally agree*), participants generally tend to *agree* or *partially agree* with the fact that agents’ behaviors are believable in video clips 1c, 2a, and 4a (see table 1 and figure 3). The lowest mean for question Q1 was obtained for video 1c, and the highest mean was obtained for video 2a. According to Student’s t-test, these scores are significant when compared to an average score of 4 (1c: $t = 7.59$; $p < .001$; 2a: $t = 18.64$; $p < .001$; 4a: $t = 10.39$; $p < .001$). These results support *H1a*.

H1b: *agents’ behaviors simulated by COR-E are assessed as related to emotions.*

A large majority of participants estimated that agents’ behaviors were related to emotions. The smallest percentage of participants who answered “yes” to question Q2 is 71.68% for video 2a, and the highest is 92.04% for video 1c (4a: 80.53%). The difference between these frequencies and theoretical frequencies is significant (video 1c: $\chi^2 = 79,87$, $\alpha < .001$; 2a: $\chi^2 = 21,25$, $\alpha < .001$; 4a: $\chi^2 = 42,12$, $\alpha < .001$). These results support *H1b*.

H1c: *participants recognize more some emotion categories than others.*

The difference between the emotion categories that were recognized by participants and theoretical frequencies is significant for each video (video clip 1c: $\chi^2 = 327,96$; $\alpha < .001$; 2a: $\chi^2 = 177,22$; $\alpha < .001$, 4a: $\chi^2 = 180,78$; $\alpha < .001$). In each video, the emotion of “anger” was recognized the most by participants (1c: 74,04% of participants recognized it; 2a: 48.15%; 4a: 52.75%), and then came the emotion of “frustration” (1c: 51.92%; 2a: 43.21%; 4a: 45.05%). The theoretical frequency with which an emotion was selected by a participant when he/she answered “yes” to Q2 is 3/15. The emotions of “surprise”, “contempt”, “shame” and “frustration” exceed this theoretical frequency in at least one of the video clips. The emotions of “love” and “sadness” have not been selected by any user. We also note that for each emotion, the percentage of participants who recognized it is relatively stable over the three videos (see table 2 and figure 4). Finally, some participants recognized emotions other than those proposed in the list (category “other”, from 4.81% to 9.88% per video). These results support *H1c*.

5.2 Acquisitive and Protective Behaviors

Video group 1 was related to the three hypotheses below. Video clip **1a** corresponds to the deactivation of acquisitive and protective behaviors (condition $NA \wedge NP$), **1b** to the deactivation of protective behaviors only (condition $A \wedge NP$),

emotion	Video 1c	Video 2a	Video 4a
anger	74.04%	48.15%	52.75%
fear	19.23%	12.35%	19.78%
sadness	0%	0%	0%
joy	0.96%	1.23%	1.10%
disgust	8.65%	4.94%	8.79%
surprise	32.69%	23.46%	30.77%
pride	8.65%	7.41%	9.89%
shame	7.69%	22.22%	14.29%
contempt	24.04%	17.28%	28.57%
love	0%	0%	0%
hate	2.88%	2.47%	1.10%
boredom	7.69%	18.52%	9.89%
frustration	51.92%	43.21%	45.05%
other	4.81%	9.88%	6.59%

Table 2: Percentage of participants that recognized an emotion category per video clip.

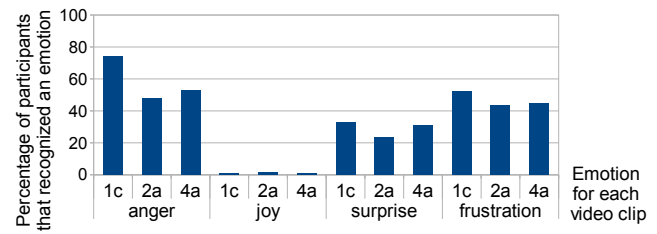


Figure 4: Percentage of participants that recognized anger, joy, surprise and frustration per video clip.

and **1c** to the activation of the two types of behavior (condition $A \wedge P$).

H2a: *Participants will judge that agents’ behaviors are related to emotions when protective and acquisitive behaviors are activated, but not when they are disabled.*

A large majority of participants estimated that in video 1c, agents’ behaviors were related to emotions (answer to Q2, *yes*: 92.04% *no*: 7.96%, see figure 5), and that in video 1a, behaviors were not related to emotions (*yes*: 18.58% *no*: 81.42%). The difference between these results and theoretical frequencies is significant (video clip 1c: $\chi^2 = 327.96$; $\alpha < .001$; 1a: $\chi^2 = 43,37$; $\alpha < .001$). These results support *H2a*.

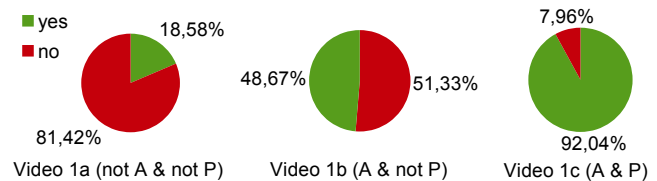


Figure 5: Answers to Q2, “Are these behaviors related to characters’ emotions?”

H2b: *The simulation of acquisitive and protective behaviors induces more believability than the simulation of acquisitive behaviors only.*

Participants generally tended to *agree* or *agree a little* with the fact that the behaviors observed in video clip 1c are believable. In contrast, for video clip 1b, participants tended

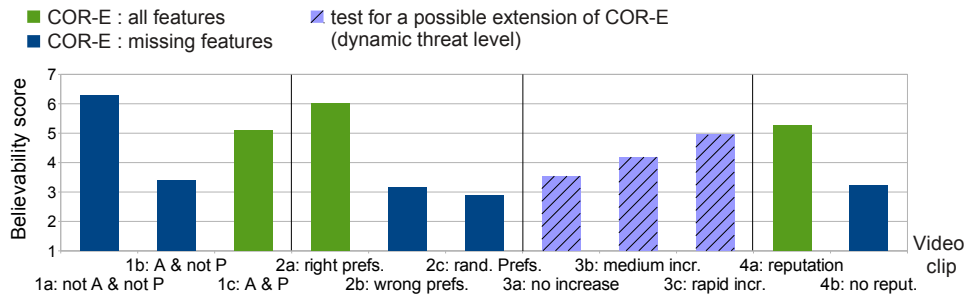


Figure 3: Results on believability score for each video clip.

to *disagree* with this statement (see table 1 and figure 3). The results on the believability score for these two videos are significant between them ($t = 7.59$, $p < .001$). These results support *H2b*.

As an additional result, we note that video clip 1a obtains the highest score on believability, since a majority of participants *totally agree* with the statement of question Q1.

5.3 Preferences

H3a: *Well configured preferences induces more believability.*

Video group 2 was related to this hypothesis. Video clip **2a** corresponded to the preferences configured so as to obtain believable behaviors (condition *C*), **2b** to preferences with a configuration estimated as incorrect (condition *NC1*), and **2c** to random preferences (condition *NC2*).

Video clip 2a was assessed as more believable than video clips 2b and 2c (see table 1 and figure 3). These results are significant between videos clips 2a and 2b ($t = 14.03$, $p < .001$), and also between video clips 2a and 2c ($t = 16.04$, $p < .001$). These results support *H3a*.

5.4 Reputation Resource

H4a: *the simulation of Reputation resources induces more believability.*

Video group 4 was related to this hypothesis. Video clip **4a** corresponded to the simulation of reputation resources (condition *R*), and **4b** to their removal (*NR*).

Video clip 4a was assessed as more believable than video clip 4b (see table 1 and figure 3). The difference between the scores of these two video clips is significant ($t = 11.15$, $p < .001$). These results support *H4a*.

5.5 Dynamic Threat Level

H5a: *an increase in the threat level induces a better believability score than a constant threat level.*

Video group 3 was related to this hypothesis. Video clip **3a** corresponded to a constant threat level (condition *M1*), video **3b** to a moderate increase of the threat level (*M2*), and **3c** to a rapid increase of the threat level (*M3*).

Video clips 3c and 3b were judged as more believable than video clip 3a (see table 1 and figure 3). The difference between the scores of video clips 3a and 3b is significant ($t = 4.14$, $p < .001$), as well as for 3a and 3c ($t = 6.55$, $p < .001$). These results support *H5a*.

As an additional result, we note that video clip 3c was judged as more believable than video clip 3b. The difference between the scores of these two video clips is significant ($t = 3.84$, $p < .001$).

6. DISCUSSION

The hypotheses formulated about COR-E were all supported by the results. From this we can infer that COR-E allows the simulation of believable emotional behaviors (general hypothesis *H1*) thanks to its characteristics (*H2*, *H3*, *H4*). Acquisitive and protective behaviors, preferences, and the psychological resource of “Reputation” seem necessary to produce such behaviors. The dynamic threat level, tested as an extension of the model, produced behaviors rated as more believable (*H5*). Consequently, we will include this extension in a next version of the model. These results must be discussed in relation to several factors detailed below.

A large majority of participants recognized the behaviors produced by COR-E as related to agents’ emotions (71.68 % to 92.04 %). These good results may be due in part to the use of the textual indication “protest”. Indeed, this term can be psychologically associated with the emotion of anger, thus facilitating the recognition of that emotion among participants. According to Austin’s classification [1], “to protest” is a behabitive speech act, which includes the notion of reaction to other people’s behavior . Therefore, the use we made of this term in COR-E seems appropriate. However, it would be interesting to know if the same results would be obtained without this textual indication. For example, we could use a visual indication, such as an agent frowning its eyebrows and raising its fist. It would also be interesting to combine COR-E with an utterance selection module in order to allow agents to express themselves with words. To do so, we can rely on an existing model for utterance selection related to impoliteness [4]. It is likely that the use of utterances will induce more emotions recognized by participants.

Behaviors produced by COR-E were rated as believable (mean from 5.08 to 6.01 on a scale from 1 to 7). However, the best score for believability was obtained by a video clip which was not related to the model, where there was no acquisitive or protective behavior (mean 6.28). In this video clip, agents are simply waiting, and move forward in the queue automatically when there is enough space left. However, there is a small percentage of emotion recognition for this video clip (18.58% of participants). It would be interesting to assess whether the participants would have a better feeling of immersion with video clip 1a, rated as strongly believable, or with a video rated as less believable but with a high emotion recognition rate. This data would allow us to know whether it is better to maintain the believability at the expense of emotional behaviors, or to preserve emotional behaviors with the risk of losing some believability.

It is possible that the score on believability was lowered

because of a bias related to the interpretation of the term “believable”. In this study, we wanted to know if participants judge agents’ behaviors as believable, without consideration for the frequency of those behaviors. That is to say that a behavior occurring rarely can be as believable as a behavior occurring often. However, in a part of the questionnaire, one participant reported that the behaviors observed in each video clip could occur in the real world, but that behaviors occurring less frequently were less believable. As emotional behaviors tend to occur somewhat rarely, they might be judged as less believable. Another possible bias that could have lowered the score on believability may be related to agents’ moves. For example, the fact that an agent goes to the tail of the queue after being reprimanded did not seem believable to some participants. Finally, one participant reported that the icons representing the agents rendered difficult the interpretation of emotions.

These elements indicate that the believability of agents’ behaviors could be further improved.

7. CONCLUSIONS AND FUTURE WORK

We presented in this paper an evaluation of the COR-E model, aimed at the simulation of believable emotional behaviors in virtual agents. Our main hypotheses was that COR-E should produce behaviors recognized as believable and emotional by human observers, and that this result should depend on the characteristics of its architecture. These main hypotheses are supported by the results obtained in using an online survey where participants were asked to rate 11 video simulations coming from COR-E. We plan to extend this model to groups and crowds, in working on the notion of collective behaviors and shared resources. We also intend to work on the conditions under which an agent choose to enter or leave a group, taking into account its own interests and those of the group in terms of resources.

Acknowledgements: This research received support from the TerraDynamica Project (FUI8) funded by the City of Paris, the Local Councils of Val d’Oise, Seine-Saint-Denis, Yvelines, the Regional Councils of Ile de France and Aquitaine and the French Ministry of Economy, Finances and Industry, Directorate for Competitiveness of Industry and Services.

8. REFERENCES

- [1] J. Austin. *How to do things with words*, volume 88. Harvard University Press, 1975.
- [2] L. Barrett. Solving the emotion paradox: Categorization and the experience of emotion. *Pers. and social psychol. review*, 10(1):20, 2006.
- [3] T. Bosse, M. Hoogendoorn, M. Klein, J. Treur, and C. van der Wal. Agent-based analysis of patterns in crowd behaviour involving contagion of mental states. *Modern Approaches in Applied Intelligence*, pages 566–577, 2011.
- [4] S. Campano and N. Sabouret. A socio-emotional model of impoliteness for non-player characters. In *3rd International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 1–7, sept. 2009.
- [5] S. Campano, N. Sabouret, E. de Sevin, and V. Corruble. The ‘resource’ approach to emotion. In *International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS’2012)*, 2012.
- [6] D. Dörner, J. Gerdes, M. Mayer, and S. Misra. A simulation of cognitive and emotional effects of overcrowding. In *International Conference on Cognitive Modeling*, pages 92–98, 2006.
- [7] C. Elliott. *The affective reasoner: a process model of emotions in a multi-agent system*. PhD thesis, 1992.
- [8] P. Gebhard. ALMA: a layered model of affect. In *International joint conference on Autonomous agents and multiagent systems*, pages 29–36. ACM, 2005.
- [9] J. Gratch and S. Marsella. A domain-independent framework for modeling emotion. *Cognitive Systems Research*, 5(4):269–306, 2004.
- [10] J. Gratch, S. Marsella, N. Wang, and B. Stankovic. Assessing the validity of appraisal-based models of emotion. In *Affective Computing and Intelligent Interaction and Workshops (ACII)*, pages 1–8, 2009.
- [11] S. Hobfoll. Conservation of resources. *American Psychologist*, 44(3):513–524, 1989.
- [12] S. Hobfoll. *Stress, culture, and community: The psychology and philosophy of stress*. Plenum Pub Corp, 2004.
- [13] W. James. The emotions. In *The principles of psychology*, 1890.
- [14] R. Lazarus and S. Folkman. *Stress, appraisal, and coping*. Springer Publishing Company, 1984.
- [15] S. Luke, C. Cioffi-Revilla, L. Panait, K. Sullivan, and G. Balan. Mason: A multiagent simulation environment. *Simulation*, 81(7):517–527, 2005.
- [16] S. Marsella and J. Gratch. EMA: A process model of appraisal dynamics. *Cognitive Systems Research*, 10(1):70–90, 2009.
- [17] S. Marsella, J. Gratch, and P. Petta. Computational models of emotion. *A Blueprint for an Affectively Competent Agent: Cross-Fertilization Between Emotion Psychology, Affective Neuroscience, and Affective Computing*, 2010.
- [18] S. Marsella, J. Gratch, N. Wang, and B. Stankovic. Assessing the validity of a computational model of emotional coping. In *Affective Computing and Intelligent Interaction and Workshops*, pages 1–8, 2009.
- [19] A. Mehrabian and J. Russell. *An approach to environmental psychology*. the MIT Press, 1974.
- [20] A. Ortony, G. L. Clore, and A. Collins. *The cognitive structure of emotions*. New York : Cambridge University Press, 1988.
- [21] A. Ortony and T. Turner. What’s basic about basic emotions. *Psychological review*, 97(3):315–331, 1990.
- [22] R. Pfeifer. The “Fungus Eater” Approach to Emotion: A View from Artificial Intelligence. *Cognitive Studies*, 1:42–57, 1994.
- [23] J. Russell. Culture and the categorization of emotions. *Psychological bulletin*, 110(3):426–450, 1991.
- [24] S. Schachter and J. Singer. Cognitive, social, and physiological determinants of emotional state. *Psychological review*, 69(5):379, 1962.
- [25] K. Scherer. Appraisal theory. *Handbook of cognition and emotion*, pages 637–663, 1999.