# There's Always Hope: Enhancing Agent Believability through Expectation-Based Emotions

Tibor Bosse
Vrije Universiteit Amsterdam
De Boelelaan 1081a, 1081 HV Amsterdam
tbosse@few.vu.nl

Edwin Zwanenburg
Vrije Universiteit Amsterdam
De Boelelaan 1081a, 1081 HV Amsterdam
zweans@msn.com

## Abstract

*To endow virtual agents with more realistic affective behavior, the notion of expectation-based emotions plays an important role: emotional states of agents should not only be triggered by present stimuli, but also by anticipation on future stimuli, and evaluation of past stimuli in the context of these anticipations. Within this study, an extension of the belief-desire-intention (BDI) model with expectation-based emotions is proposed. The model has been implemented in the modeling language LEADSTO. In addition, a game application has been developed, in which a user can play a dice game against an agent that is equipped with the emotion-based model. An empirical evaluation indicates that the model significantly enhances the agent's believability, in particular concerning its involvement in the situation.*

## 1. Introduction

In the last decade, within the areas of Artificial Intelligence and Cognitive Modeling, various approaches have been proposed to endow Intelligent Virtual Agents (IVAs) [23] with emotions [2, 3, 8, 10, 14, 18, 19, 25]. The motivation for doing this is almost commonly accepted these days: emotions allow IVAs to have more human-like appearance, to be more expressive in their behavior, and facilitate their interactions with humans [2]. By enhancing the capability of an agent to emotionally express itself, the human will more easily identify him- or herself with the agent, and possibly anthropomorphize the agent or empathize with it. In short: emotions make IVAs more *believable* to the humans interacting with them.

Recently, much research has been dedicated to developing IVAs with more realistic graphical representations. However, the actual underlying affective *behavior* of such agents often stays a bit behind. For example, although many IVAs nowadays have the ability to somehow show different emotions by means of facial expressions, it is rather difficult for them to show the right emotion at the right moment. This is in conflict with the requirement of virtual agents to closely mimic human affective behavior. Nevertheless, several studies in Social Sciences have shown that this is an important prerequisite for an agent to increase human involvement in the virtual environment; see e.g. [11]. Therefore, existing systems based on IVAs are not as effective as they could be. A particular type of emotions that is only marginally developed in many IVAs is those emotions that are based on *expectations* [6]. Although most IVAs nowadays exploit detailed mechanisms to generate emotions based on the stimuli that are currently present [2, 3, 8, 10, 14, 18, 19, 25], their ability to generate emotions on the basis of stimuli that may occur in the future is often much less developed. This is an important difference with the affective behavior of humans, whose emotional states are constantly influenced by an evaluation of the future possibilities to fulfill their present goals [14, 15]. For example, a person that desires her favorite soccer team to win will be enthusiastic if she expects that this will indeed happen, but will become frustrated if the chances for this to happen become lower (e.g., when the team is behind).

The main challenge of the current study is to incorporate more detailed expectation-based affective processes within IVAs, and more in particular, within IVAs in game applications. This is a particularly suitable domain, since it usually involves very explicit goals (e.g., winning the game), expectations (e.g., winning or losing a particular phase in the game), and expectation-based emotions (e.g., hope or fear). For example, a poker playing agent that expresses its excitement because it is likely to beat the human (or expresses its sadness because it does not see any chance to win) will probably be perceived as more believable than an agent without such behavior. To address this issue, a generic computational model for elicitation of expectation-based emotions is presented. The model is inspired by the approach presented in [5], which provides a logical theory of expectation-based emotions, built upon BDI notions. Within that theory, several emotions play a role (which will be explained in more detail in the next section): hope, fear, surprise, satisfaction, dissatisfaction, relief and disappointment. The processes involved in generating these emotions are formalized in a generic, executable format, in such a way that the model can easily be plugged in within any virtual agent. For this purpose, the agent-based modeling language LEADSTO is used [4]. To illustrate the effectiveness of the approach, it has been applied to two real-world applications (involving a virtual opponent agent in the context of gambling games). For one of these games, an experiment has been performed



978-1-4244-4799-2/09/$25.00 ©2009 IEEE

to evaluate how the agent is perceived by the humans interacting with it. Thus, as a second contribution, this article provides more insight into the benefits and drawbacks of applying expectation-based emotions within IVAs in games.

Below, the basic model for elicitation of expectation-based emotions is described in Section 2. In Section 3, one of the developed game applications (the dice game ‘5000’) is discussed. Section 4 presents the experimental setup used to evaluate the model in the context of this dice game, and the results of the experiment are presented in Section 5. Section 6 compares the presented model with related approaches in the literature, and Section 7 concludes the paper with a discussion.

## 2. Expectation-based emotions

As mentioned in the introduction, the logical theory presented in [5] is taken as a basis for the development of our model. The main idea of that theory is that expectation-based emotions can be derived on the basis of several elementary concepts, among which *desires* (e.g., “I desire that it will be sunny tomorrow”), *expectations* (e.g., “it will probably rain tomorrow”), and *beliefs* (e.g., “it is sunny”). In this paper, we adopt the same approach, where we treat expectations as a specific type of beliefs (namely uncertain beliefs about the future). As such, the theory can be seen as an extension of a BDI approach [9, 24], or rather, an EBDI (emotion-belief-desire-intention) approach [1, 19].

For a global overview of the model, see Figure 1. In this picture, the dotted box indicates an agent, the boxes indicate different states, and the arrows indicate causal relationships. Note that the beliefs play different roles (see the different numbers in Figure 1):

1) Beliefs may influence desires.
2) A desire, in combination with the belief that a particular action fulfils this desire, leads to the intention to perform that action.
3) An intention to perform a particular action, in combination with the belief that it is possible to perform that action, leads to the actual execution of that action.
4) A desire for a particular state, in combination with the belief that that state may occur (an expectation), leads to an expectation-based emotion[1]. For example, if an agent desires to gain points and believes that there is a fair probability to gain points, then it will start hoping.
5) A desire for a particular state, in combination with the belief that that state has or has not occurred, leads to an expectation-based emotion.

Note that in item 4) and 5), two different types of expectation-based emotions are distinguished: emotions that appear *before* a particular (world) state or event has occurred, and emotion that appear *after* such a state has occurred. In this paper, both types of expectation-based emotions are modeled. For convenience, they will be described as *before-emotions* and *after-emotions* in the remainder of this paper. Below, Section 2.1 will explain how before-emotions are formalized in our model, and Section 2.2 will address after-emotions. Section 2.3 will discuss the impact of emotions on actions. In principle, it also possible that emotions influence other mental states (e.g., as in various coping strategies), but this is not the focus of the current paper.

[1] Note that we do not claim that all possible emotions (and their effects) are modeled; we only address the expectation-based emotions introduced in [5], and their impact on actions to be performed, which is considered sufficient for the current purposes.

In order to model the processes involved in the generation of expectation-based emotions, the agent-based modeling language LEADSTO is used [4]. This language integrates qualitative, logical aspects and quantitative, numerical aspects, which allows the modeler to exploit both logical and numerical methods for analysis and simulation. The basic building blocks of LEADSTO are so-called *executable dynamic properties*, by which direct temporal dependencies between two state properties in successive states are modeled. Their format is defined as follows. Let $\alpha$ and $\beta$ be state properties of the form ‘conjunction of ground atoms or negations of ground atoms’. Then, $\alpha \twoheadrightarrow_{e,\, f,\, g,\, h} \beta$ means:

*If state property $\alpha$ holds for a certain time interval with duration g, then after some delay (between e and f) state property $\beta$ will hold for a certain time interval of length h.*

Atomic state properties can have a qualitative, logical format, such as an expression desire(d), expressing that desire d occurs, or a quantitative, numerical format such as an expression has_value(x, v), expressing that variable x has value v. For more details of LEADSTO, see [4].

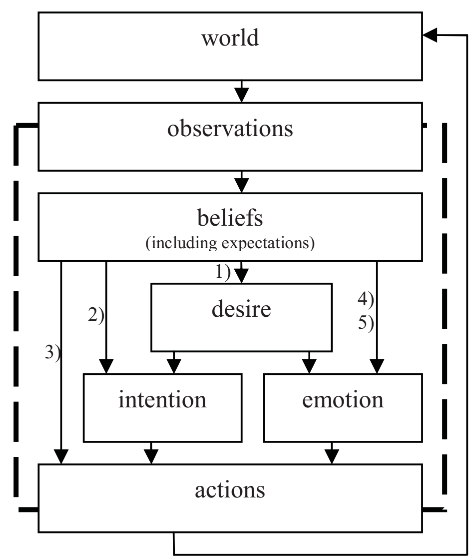


Figure 1: Overall EBDI architecture.

### 2.1. Emotions before an event

Following [5], two basic types of before-emotions are considered within the presented model: *hope* and *fear*. According to [5], the concept of hope can be defined in terms of desires and expectations, or, more specifically, in terms of the *importance* of a desire and the *probability*[2] of an expected future state. Other authors take a similar approach (e.g., [10, 25, 27]). The main idea is that *hope* for a certain future state becomes higher when its importance becomes higher and its probability becomes lower. In the formula presented below, a similar mechanism is exploited. However, one small modification is proposed: if the probability of a particular state becomes very low, then usually people stop hoping because they think the state is not worth hoping for anymore. See, for example, the work of Snyder and colleagues, who claim that "*if a goal is truly unattainable, then retaining it almost always demoralizes a person*" [26]. This process is modeled via the hope function shown below, where hope is modeled on a continuous scale (as a real number between 0 and 1). In LEADSTO, the hope function is formalized as follows (note that the timing parameters e, f, g, h have been omitted, for simplicity):

Hope Function
∀s:state ∀p:probability ∀i:importance
expectation(s, p) ∧ desire(s, i) ∧ p ≤ θ
↠ hope(s, (-0.5*(cos(1/θ*π *p)) + 0.5) * i)

∀s:state ∀p:probability ∀i:importance
expectation(s, p) ∧ desire(s, i) ∧ p > θ
↠ hope(s, (-0.5*(cos(1/(1-θ)*π *(1-p))) + 0.5) * i)

The behavior of this function for different values of probability and importance is shown in Figure 2. Note that this function differs from most approaches present in the literature, since its top is not situated at the point where probability = 0. Here, θ is a shaping parameter (in the domain <0,1]) that can be used to manipulate the location of the top of the hope curve. The value of this parameter may differ per individual, and represents the point (mentioned above) at which people give up hoping. This corresponds to the work mentioned in [20], who claim that 'optimism' is a characteristic that differs per person. The top of the probability/hope-curve is always situated at the point where probability = θ. Thus, for a θ close to 1, the top of the curve is situated very much to the right (representing persons that only 'dare' to hope for events with very high probabilities). Similarly, for a θ close to 0, the top of the curve is situated to the left (representing persons that already start hoping for events with very low probabilities). In the remainder of this paper (and also in Figure 2), the value of 0.5 is chosen for this parameter. Here, a cosine function is chosen to ensure that the curve is less steep at the extremes. Furthermore, Figure 2 shows that a higher importance simply leads to a higher hope (which is standard in the literature).

In addition to hope, the emotion *fear* is used. Similar to [5], in this paper, fear (or 'worry') is modeled as the complement of hope[3]. Thus, when an agent hopes, for instance, that it will be sunny with an intensity of 0.7, it simultaneously fears that it will not be sunny with an intensity of 0.7.

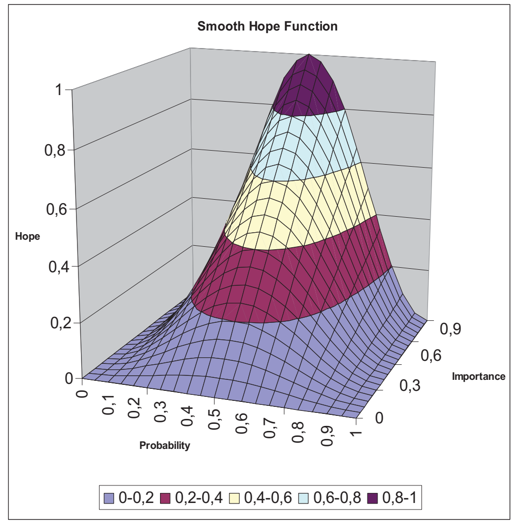


Figure 2: Hope as a function of probability and importance.

## 2.2. Emotions after an event

In addition to the before-emotions, a number of expectation-based emotions are used that occur after a relevant event has taken place. In total, four types of after-emotions are considered within the presented model: *surprise*, *(dis)satisfaction*, *relief*, and *disappointment* (again, taken from [5]). These emotions usually occur when an observed world state is compared with an earlier expectation.

Let us start with *surprise*. In the presented model, agents make predictions about future states with certain probabilities. In case a state occurs of which an agent was 100% sure that it would happen, the agent will not be surprised. However, in all other cases where the state was predicted with some probability of less than 100%, the agent will experience some level of surprise. In principle, this level is proportional with the prediction failure, which is modeled in LEADSTO via the following formula:

Surprise Function
∀s:state ∀p:probability
previous_expectation(s, p) ∧ belief(s)
↠ surprise(s, 1-p)

Note that, for each prediction made, the model automatically generates predictions for the complement of the relevant world state. Hence, in case an agent predicts that it will be sunny with a probability of 0.3, then it will also predict that it will not be sunny with a probability of 0.7.

[2] Note that, instead of probability, in [5] the notion of credibility is used. Both concepts can easily be mapped: when probability is either very low (0) or very high (1), then credibility is high. When probability is around 0.5, then credibility is low.

[3] For a detailed discussion about the definition of 'fear' and its relation to hope, see [7].

The next emotions addressed are *satisfaction* and *dissatisfaction*. These are based on whether a desire comes true or not, and emerge with an intensity that equals the importance of the desire:

Satisfaction Function
∀s:state ∀i:importance
desire(s, i) ∧ belief(s)
↠ satisfaction(s, i)

∀s:state ∀i:importance
desire(s, i) ∧ belief(not(s))
↠ dissatisfaction(s, i)

As an alternative, satisfaction and dissatisfaction may be modeled via different intensities of the same emotion. However, for practical purposes it was decided to keep them separated (also see the next section).

The last two emotions used in the model are *relief* and *disappointment*. Following [5], these emotions depend on the combination of surprise and (dis)satisfaction. When an agent is both very surprised and satisfied, it means that its desire has come true and the probability for this was low. In such a case, it will experience relief, a kind of happiness that an expected negative event did not come true. Likewise, one experiences disappointment when one is surprised and dissatisfied, a kind of sadness that an expected positive event did not come true. These mechanisms are modeled as follows:

Relief/Disappointment Function
∀s:state ∀i1,i2:intensity
surprise(s, i1) ∧ satisfaction(s, i2)
↠ relief(s, i1*i2)

∀s:state ∀i1,i2:intensity
surprise(s, i1) ∧ dissatisfaction(s, i2)
↠ disappointment(s, i1*i2)

### 2.3. From emotions to actions

Different expectation-based emotions lead to different actions, depending on the context. In the context of a game, for example, an agent that hopes (and expects) to win may look confident, and behave relaxed. An agent that fears (and expects) to lose, on the other hand, may look worried, start talking in itself, or behave unfriendly towards the opponent. Thus, for an IVA a mechanism is needed that converts the emotions modeled in the previous sections into actions, depending on the context. Here, actions may include both verbal and non-verbal behavior. In order to generate these actions, the emotions may be combined with the agent's intentions (see Figure 1). The general rule to describe this mechanism is shown below. In short, this rule states that, if an agent intends to perform action a, and it experiences various emotions with intensities i1, …, i7, then it will perform a variant of the action that is shaped by the emotions (via the function f(...)):

Emotional Action Function
∀a:action ∀i1,i2,i3,i4,i5,i6,i7:intensity
intention(a) ∧ hope(i1) ∧ fear(i2) ∧ surprise(i3) ∧
satisfaction(i4) ∧ dissatisfaction(i5) ∧ relief(i6) ∧
disappointment(i7)
↠ perform(a, f(i1,…,i7))

Obviously, the function f(...) needs to be defined for a particular domain. In the next section, an example application will be shown where a specific variant of this function is used. The idea is that each of the emotions is split up into several non-overlapping intervals (e.g., [0-0.2>; [0.2-0.4>, [0.4-0.6>, [0.6-0.8>, [0.8-1.0]). Then, if the agent intends to perform a certain action, this action is shaped according to the intervals in which its emotional states are classified. For example, an agent that is almost certain to win a dice game and is suddenly beaten by an incredibly lucky throw of the opponent (surprise ∈ [0.8-1.0]), will start shouting and look angry.

Note that this section only presented the most important mechanisms of the model. Due to space limitations, a large number of LEADSTO rules have been left out. These rules include, among others, mechanisms to perform BDI based reasoning and mechanisms to combine emotions for multiple states into unitary emotions.

## 3. Application: a dice game

In order to assess the performance of the model in a real human-computer setting, two game applications have been developed. These applications addressed the games 'tic-tac-toe' and '2500'. Due to space limitations, only the second application is described here. This game can be played at the URL in [28] (note that some software needs to be installed, and that the operating system's own standard voice is used).

The game of 2500 (a short variant of the game '5000') was selected for a couple of reasons. First, it is a dice-based game, which means that it mainly depends on *chance*, which can directly be connected to *probability*, one of the basic elements underlying the model's hope function. Second, the other basic element, *importance*, can easily be manipulated by introducing a bet that can be gained during each game. Third, the rules of the game are well defined and relatively simple, which makes it easy to implement an IVA that can play the game.

The rules of the game 2500 are as follows. Two players are involved: the human and the IVA. Each player starts with 0 points. The game consists of multiple *rounds*, in which each of the players takes one *turn*. A round is finished when each player has taken his turn. The order of the player's turns is determined at the start of the game. Whenever a player obtains 2500 points (or more), this player wins the game and receives the bet.

Each turn starts with six dice which are thrown simultaneously. A throw can deliver between 0 and 2000 points, depending on the dice that are thrown. For example, 2 ones and 1 five delivers 250 points (see [28] for the exact scoring). After a player has thrown the dice, (s)he must put aside at least one dice with points. The points of this (or these) dice are noted. If the cumulative number of points of the dice set aside is 350 or more, the player may choose to end the turn. Otherwise the player continues to throw with the dice not set aside. If all six dice are put aside, then that player's turn automatically ends. If a player does not get any points during a throw, all points of that turn are reduced to zero and the player's turn is over.

The application has been implemented in HTML, using Javascript. Moreover, Haptek's PeoplePutty has been used to visualize the IVA. This environment offers the ability to display faces within a script for HTML. Moods, expressions and gestures can be set for the face, as well as shapes and accessories. Furthermore, it makes use of a text-to-speech engine: text can be inputted, and speech will be given as output, with appropriate lip synching. In the current application, a number of facial expressions have been used, among which smiling, looking sad, looking happy and looking angry. Figure 3 displays a screenshot of the application.

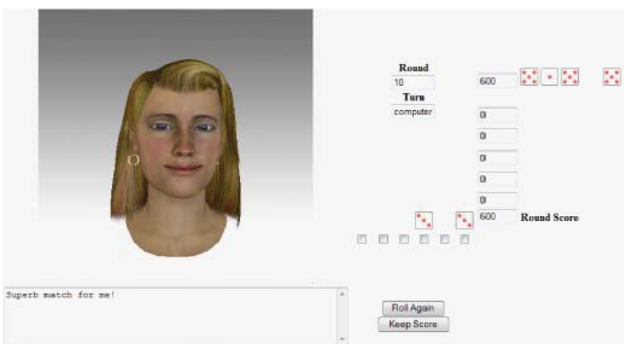


Figure 3: Screenshot of the game application.

To enable the IVA to use the expectation-based emotion model for this game, the model has been filled with some domain-specific knowledge. For example, during each game, the agent is assigned the desire to win. Moreover, during each turn, this desire is refined to several sub-desires. During the turn of the human player, the agent desires that the human does not get any score for the turn. During its own turn, the agent desires that it does get score. Each time the dice are thrown, the agent has the intention to communicate with the user (both verbally and via facial expressions), and will thus generate an appropriate surprise, (dis)satisfaction, relief or disappointment, based on its (sub)desires and its beliefs about the game state. Each time that one of the players sets dice aside, the agent will demonstrate hope or fear based on its desire and beliefs about the game state. Furthermore, after each turn, the agent will display hope or fear in relation to the game as a whole, and thus show its confidence that it will win, its worry about not winning or its fear that the human player wins.

To convert emotional states to actual behavior, two of the mechanisms mentioned above are exploited: facial expressions and emotional utterances. To generate facial expressions, the states in the model can directly be mapped to parameter settings in PeoplePutty. For example, if the IVA has a satisfaction of 0.7, this is visualized by setting the slide bar for "happy" provided by the software to 70%. To generate emotional utterances, a database of domain-specific utterances has been created. To fill this database, following Section 2.3, each of the emotional states has been split up into some non-overlapping intervals, and for each interval, a number of utterances have been specified. For example, if surprise<0.2, the agent chooses among several statements like: "`Not a surprising roll`", "`I saw that roll coming`", and "`I knew that one would come up`". Also multiple simultaneously occurring emotions (e.g., hope and satisfaction) may lead to certain utterances. The advantage of this approach is that one does not have to define statements for each possible situation in the game, since these situations are classified in terms of high level emotional states of the IVA.

As mentioned earlier, all emotions are generated based on *importance* and *probability*. The importance is also dynamic. For example, it is more important to get points if the opponent is close to winning, thus near 2500 points. It is less important to get points if the opponent is far from winning, thus still near 0 points. In addition, this importance depends on the overall bet of the game.

The strategy of the agent in this game is relatively simple. In each round, the sub-desire of the agent is 'getting 350 points or more'. This will always result in the intention to put aside all dice that deliver points, and keep rolling until it has reached 350 points or more, independent of the overall game status. In order to calculate probabilities of throws, the agent makes estimations of the actual probabilities, using ordinary statistics.

## 4. Experiment

To evaluate how humans perceive the developed IVA (with our model for expectation-based emotion elicitation), the following experimental setup has been used. Four variants of the '2500' application have been developed, with different implementations of the IVA:

1. The IVA uses no emotion elicitation model at all.
2. The IVA uses the complete emotion elicitation model.
3. The IVA uses a variant of the model with opposite, incongruous emotions. For example, the IVA is happy where it is supposed to be sad.
4. The IVA uses the complete model, but only the 'after-emotions'. All 'before-emotions' have been omitted.

Here, variant 1 serves as the control condition. Variant 3 has been added to test whether adding emotions already enhances the agent's believability in itself, or that it truly needs to be done in a meaningful fashion as per variant 2. Variant 4 is another control condition, to compare the impact of ‘before-emotions’ (i.e., hope and fear, see Section 2) with the impact of ‘after-emotions’ (i.e., surprise, (dis)satisfaction, relief, disappointment). In variant 2, 3, and 4, all emotions are shown both via facial expressions and utterances.

Twenty-four people participated in the experiment. The age of the participants ranged between 60 and 18, with a mean age of 28.3 and a standard deviation of 13.1. Among the participants, 19 were male and 5 were female. For each participant, the experiment lasted between 30 to 45 minutes, depending on their skills and their luck in each of the games.

Before starting the experiment, the participants were told the agent’s name was “Anna”. The rules of the game were explained to them and they were allowed to play a single test round with variant 1, the neutral agent, to get familiar with the rules. After this, the participants played one game with each variant of the application, i.e., they played the game four times. Since there are four variants, there were 24 possible orderings. These orderings have been distributed randomly among the 24 participants.

In each of the plays, the money at stake was set to a maximum. This means that all plays were maximally important for both the agent and the human player. After each game, the participants had to fill in a questionnaire, asking them to award a measure of agreement of fifteen statements about how they perceived that particular agent. A gradual seven-point scale was used, with the following meaning: 1=‘I strongly disagree’, 2=‘I disagree’, 3=‘I weakly disagree’, 4=‘neutral’, 5=‘I weakly agree’, 6=‘I agree’, 7=‘I strongly agree’.

## 5. Results

To analyze the results of the experiment, an ANOVA has been applied on the answers to the statements of the questionnaire. In this section, the results for the most relevant statements (all related to believability of the IVA) are addressed. These statements were: “Anna was believable”, “The behaviour of Anna was human-like”, “I thought Anna’s reactions were natural”, “Anna reacted on my actions”, “Anna was interested in the game”, “Anna did not care about the game situation” and “Anna wanted to win the game”.

There were four variants: the non-emotional (NE) variant, the full emotional variant (FE), the incongruent emotional (IE) variant and the emotional variant without ‘before-emotions’ (E). The results are presented in Table 1 and Figure 4. The vertical axis in Figure 4 corresponds to the scale explained in Section 4. The error bars represent standard deviations. In Table 1, the first column indicates the statement, and the rest of the columns indicate pair-wise comparisons between different variants. The ‘overall’ column shows whether there was a general significant effect over all variants for that statement. The cells show whether the comparison yielded a significant result or not. Here, ‘n.s.’ means ‘not significant’, ‘*’ indicates $p < 0.05$, ‘**’ means $p < 0.01$, and ‘***’ stands for $p < 0.001$. For example, the second cell of the second row states ‘FE***’, which means that the full emotional variant is significantly more believable than the non-emotional variant.

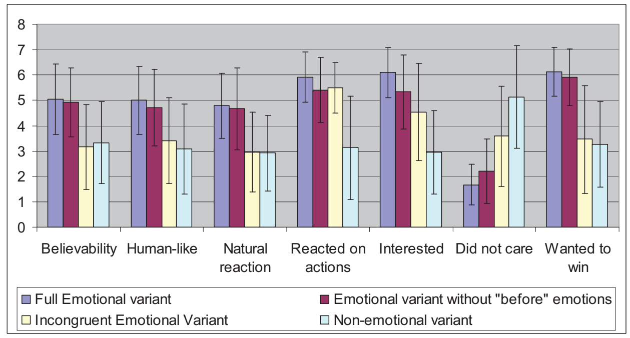


Figure 4: Statistical results of the experiment.

These results clearly show that the non-emotional variant was perceived worst with respect to all aspects related to believability, followed by the incongruent variant. This confirms our hypothesis that adding expectation-based emotions in a meaningless manner does not enhance believability very much. The scores of the two remaining variants are significantly higher than those of the other two. Among these two variants, the full emotional variant is perceived to react better on actions, to be more interested and to care more about the game play. However, its general believability, human-likeness, ability to provide natural reactions, and desire to win are not considered to be significantly greater compared to the emotional variant without ‘before emotions’.

Table 1: Detailed statistical results.

| Statement | FE-NE | FE-IE | FE-E | NE-IE | NE-E | IE-E | Overall |
|---|---|---|---|---|---|---|---|
| Believability | FE*** | FE*** | n.s. | n.s. | E*** | E*** | *** |
| Human-like | FE*** | FE*** | n.s. | n.s. | E*** | E*** | *** |
| Natural reaction | FE*** | FE*** | n.s. | n.s. | E*** | E*** | *** |
| Reacted on actions | FE*** | FE* | FE** | IE*** | E*** | n.s. | *** |
| Interested | FE*** | FE*** | FE** | IE*** | E*** | E** | *** |
| Did not care | NE*** | IE*** | E** | NE*** | NE*** | IE*** | *** |
| Wanted to win | FE*** | FE*** | n.s. | n.s. | E*** | E*** | *** |

In general, these results are encouraging, since they confirm the ability of the model to enhance believability. However, to explain the small differences between the two emotional variants, it is needed to take a closer look at the experiment and the results. Within this application, the ‘before emotions’ were expressed in terms of statements about the events in the near future (e.g., “`Please, dice, points for me!`” or “`I fear this will get worse for me!`”), whereas the ‘after emotions' were about events in the recent past (e.g., "`YES! Points for me!`" and "`How did this happen? Stupid dice!`"). Apparently, having only the latter was already sufficient to make the agent significantly more believable and human-like. The addition of the 'before emotions' did not make much difference for this. However, the presence of the 'before emotions' did provide Anna the appearance of being more interested and involved in the game. Although these results should not be over-generalized, it is an indication that 'before emotions' like hope and fear enhance the perception of involvement in a situation. Especially for IVAs in games, this is a nice feature, since people generally prefer playing against opponents that care about the game. In future work, the role of 'before emotions' will be studied in more detail.

## 6. Related Work

It is important to mention that the aim of the presented model is by no means unique. In recent years, the amount of approaches to enhance believability of virtual agents by incorporating affective processes has virtually exploded [2, 3, 8, 10, 14, 18, 19, 25]. Since there is no space to provide a complete overview of all of these approaches, we will only mention a subset of those approaches that are most closely related to our model.

One of the existing approaches that has several similarities to ours, is the computational model EMA (EMotion and Adaption) [10]. Within this model, an agent perceives the world and appraises it based on utilities of states, which represent the states' desirability and importance, similar to our approach. Future expected states have probabilities, and coping strategies are used to handle the appraised expected states. Emotions emerge based on the appraisal of these states, both for the present and the future. The emotions used in EMA (hope, joy, fear, distress, anger and guilt) differ from those presented in the current model. Hope and fear play more or less the same role, but they are calculated using linear functions. The current paper extends this approach by introducing a more sophisticated hope function, based on the parameter $\theta$ for optimism, inspired by [20, 26]. Furthermore, anger and guilt are not used within our model, whereas relief and disappointment are not used explicitly within EMA. Their joy and distress are comparable to our (dis)satisfaction.

Next, various authors have proposed to apply affective modeling approaches in game contexts, among which poker [8], chess [18], card games [3] and puzzle games [13]. The affective models underlying all of these approaches differ from ours in several respects. For example, [8] uses expectation-based emotions, but no fear and relief, [3] and [18] use different hope and fear functions, and [13] uses no explicit hopes or fears. On the other hand, these approaches propose some ideas that may be useful extensions to our approach, such as the influence of a long term mood, and the distinction between 'primary' and 'secondary' emotions. Another difference with all of the above approaches is that the current paper describes an extensive evaluation of the users' perception of the game.

An alternative approach to enhance agent believability is to introduce empathic emotions. Empathy is commonly defined as the capacity to "put yourself in someone else's shoes to understand her emotions" [16]. This has been investigated in detail in [14], and also in [17, 21, 22]. Research in this field shows that interaction with empathic agents enhances human-computer interaction and makes the agent more likeable. According to this research, empathic agents are perceived as significantly more caring, likeable, trustworthy and submissive. Within [14], incongruous emotions are used within an experiment with empathic agents, and it was found that participants perceive an agent with empathic emotions more positively and an agent with incongruous emotions more negatively. Although our agent is competitive rather than empathic, several of the results concerning believability found in [14] are reproduced within our experiment.

Finally, in [1] and [27], logical models of emotion are presented, based on the OCC model, which also includes expectation-based emotions. Like these models, the model presented here has an underlying logical theory. However, in the current paper this theory has been converted to a directly executable model, which can be plugged in into real-world applications.

Which of the above approaches yields better results, and it which circumstances, remains to be seen. To evaluate such questions, in the future we plan to set up a large user study aimed at systematically comparing different affective modeling approaches. Taking a 'principle of parsimony' perspective, such an experiment could shed more light on the actual added value of each feature of the existing approaches.

## 7. Discussion

To enhance believability of virtual agents, this paper introduces an executable model for expectation-based emotions. To this end, the logical theory of [5] has been taken as a basis, and has been converted to a generic executable model. Since a conceptual distinction has been made between the generic and domain-specific parts of the model, it is relatively easy to plug it in within IVAs in different applications. In fact, the only thing that needs to be done is filling in some slots with domain-specific knowledge (such as the utterances mentioned in Section 3). The model has been implemented and tested using the modeling language LEADSTO. In addition, two game applications have been developed, in which a human can play games against an IVA that is equipped with the model. An evaluative user study, which has been performed for one of the applications, indicates that the model significantly enhances the agent's believability, especially when it comes to its involvement in the situation.

As mentioned earlier, many affective modeling approaches exist in the literature, each of which has some similarities and some differences with the presented model. In follow-up research, it is planned to perform a systematic comparison between several existing approaches. In addition, it is planned to explore the possibilities of incorporating the presented model within a partner agent. The experiments performed so far only addressed situations in which the emotional IVA was the opponent of the user. It may be expected that the perception of an IVA using our model will differ when the human is playing with (instead of against) the IVA. This direction will be further investigated in the future.

## References


[1] Adam, C., Gaudou, B., Herzig, A., and Longin, D. (2006). OCC's Emotions: A Formalization in BDI Logic. In: Euzenat, J. and Domingue, J. (eds.), *Proceedings of AISMA'06* . Springer LNAI, vol. 4183, pp. 24-32.

[2] Bates, J. (1994). The role of emotions in believable agents. *Communications of the ACM* 37 (7), pp. 122-125.

[3] Becker-Asano, C. and Wachsmuth, I. (2008). Affect Simulation with Primary and Secondary Emotions. In: Prendinger, H., Lester, J., and Ishizuka, M. (eds.), *Proc. of IVA'08*. Springer LNAI, vol. 5208, pp. 15-28.

[4] Bosse, T., Jonker, C. M., Meij, L. van der, and Treur, J. (2007). A Language and Environment for Analysis of Dynamics by SimulaTiOn. In: *Journal of Artificial Intelligence Tools*, vol. 16, issue 3, pp. 435-464.

[5] Castelfranchi, C. and Lorini, E. (2003). Cognitive Anatomy and Functions of Expectations. In: *Proceedings of IJCAI'03 Workshop on Cognitive Modeling of Agents and Multi-Agent Interactions*, Acapulco, August 9-11.

[6] Castelfranchi, C., Falcone, R., and Piunti, M. (2006). Agents with Anticipatory Behaviors: To be Cautious in a Risky Environment. In: *European Conference on Artificial Intelligence, ECAI'06*, pp. 693-694.

[7] Day, J. P. (1998). More about Hope and Fear. *Ethical Theory and Moral Practice* 1, pp. 121-123.

[8] Gebhard, P. et al. (2008). IDEAS4Games: Building Expressive Virtual Characters for Computer Games. In: *Proc. of IVA'08*. Springer LNAI, vol. 5208, pp. 426-440.

[9] Georgeff, M. P. and Lansky, A. L. (1987). Reactive Reasoning and Planning. In: *Proc. Of the Sixth National Conf. on Artificial Intelligence, AAAI'87*. Menlo Park, California. American Association for Artificial Intelligence, 1987, pp. 677-682.

[10] Gratch, J., and Marsella, S. (2004). A domain independent framework for modeling emotion. *Journal of Cognitive Systems Research*, vol. 5, pp. 269-306.

[11] Hoorn, J. F., Konijn, E. A., and Van der Veer, G. C. (2003). Virtual Reality: Do not augment realism, augment relevance. *Upgrade - Human-Computer Interaction: Overcoming Barriers*, vol. 4, issue 1, pp. 18-26.

[12] Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, pp. 263-291.

[13] Martinho, C., and Paiva, A. (2006). Using anticipation to Create Believable Behaviour. In: *Proceedings of AAAI'06*, AAAI Press, pp. 175-180.

[14] Ochs, M., Pelachaud, C., and Sadek, D. (2008). An empathic virtual dialog agent to improve human-machine interaction. *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems*, pp. 89-96.

[15] Ortony, A., Clore, G. L. and Collins, A. (1988). The Cognitive Structure of Emotions. Cambridge University Press, Cambridge, UK.

[16] Pacherie, E. (2004) L'empathie, chapter L'empathie et ses degrés, pp. 149-181. Odile Jacob.

[17] Partala, T., and Surakka, V. (2004) The effects of affective interventions in human-computer interaction. *Interacting with computers*, vol. 16, pp. 295-309.

[18] Pereira, A., Martinho, C., Leite, I., and Paiva, A. (2008). iCat, the Chess Player: the Influence of Embodiment in the Enjoyment of a Game. In: Padgham, L., Parkes, D.C., Mueller, J., and Parsons, S. (eds.), *Proceedings of the 7th Int. Joint Conference on Autonomous Agents and Multi-Agent Systems, AAMAS'08*. ACM Press, pp. 1253-1256.

[19] Pereira, D., Oliveira, E. and Moreira, N. (2006) Modelling Emotional BDI Agents. In: *Workshop on Formal Approaches to Multi-Agent Systems (FAMAS 2006),* Riva del Garda, Italy.

[20] Peterson, C. (2000) The future of optimism. *American Psychologist*, vol. 55, pp. 44-55.

[21] Picard, R., and Liu, K., Relative Subjective Count and Assessment of Interruptive Technologies Applied to Mobile Monitoring of Stress. *International Journal of Human-Computer Studies*, vol. 65, pp. 396-375.

[22] Prendinger, H., and Ishizuka, M. (2005). Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. *International Journal of Human-Computer Studies*, vol. 62, pp. 231-245.

[23] Prendinger, H., Lester, J., and Ishizuka, M. (eds.) (2008). *Intelligent Virtual Agents. Proceedings of the 8th International Conference on Intelligent Virtual Agents, IVA'08*. Springer LNAI, vol. 5208.

[24] Rao, A. S., and Georgeff, M. P. (1991). Modeling Rational Agents within a BDI-Architecture. In: J. Allen, R. Fikes, and E. Sandewall (eds.), *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, pp. 473-484. Morgan Kaufmann.

[25] Silverman, B. (2001). More Realistic Human Behavior Models for Agents in Virtual Worlds: Emotion, Stress, and Value Ontologies. Technical Report, Philadelphia, PA, University of Penn./ACASA.

[26] Snyder, C. R., Ilardi, S. S., Cheavens, J., Michael, S. T., Yamhure, L., and Sympson, S. (2000). The role of hope in cognitive behavior therapies. *Cognitive Therapy and Research,* vol. 24, pp. 747-762.

[27] Steunebrink, B. R., Dastani, M. and Meyer, J. J. C. (2008). A Formal Model of Emotions: Integrating Qualitative and Quantitative Aspects. In: Ghallab, M. (ed.), *Proceedings of the 18th European Conference on Artificial Intelligence, ECAI'08*. IOS Press, pp. 256-260.

[28] http://www.cs.vu.nl/~wai/dicegame/