A Character Model with Moral Emotions: Preliminary Evaluation

Cristina Battaglino[1] and Rossana Damiano[2]

1   **Dipartimento di Informatica**
    Università di Torino, Italy
    `battagli@di.unito.it`
2   **Dipartimento di Informatica and CIRMA**
    Università di Torino, Italy
    `rossana@di.unito.it`

──── **Abstract** ────

In literary and drama criticism, emotions, and moral emotions in particular, have been pointed out as one of characterizing features of stories. In this paper, we propose to model story characters as value-based emotional agents, who appraise their own and others' actions based on their desires and values, and feel the appropriate moral emotions in response to narrative situations that challenge their goals and values.

In order to validate the appropriateness of the agent model for narrative characters, we ran an experiment with human participants aimed at comparing their expectations about characters' emotions with the predictions of the value-based model of emotional agent. The results of the experiment show that the participants' expectations meet the predictions of the model.

1   Introduction

Since the Age of Enlightenment [18], narratology and drama studies have acknowledged the importance of emotions in stories. In contemporary film theory and aesthetics, emotions play a central role because they sustain the process of identification with the characters. [22, 44, 11]. According to [28], "sympathetic responses to narrative characters", seen as mechanisms of emotional participation, are the most basic form of narrative engagement.

Notwithstanding this interest in characters' emotions, moral emotions have been scarcely considered in computational models of narrative. Literary and drama studies have acknowledged the importance of the moral dimension since the pioneering work of Polti [37]. The notion of moral values, first stated in Egri's definition of 'drama premise' [19], underpins most of the subsequent work conducted in scriptwriting [10], until the recent formulation stated by McKee [35] for cinematographic stories. In cognitive psychology, Bruner attributes to narratives the function of exemplifying and transmitting the values of a culture [8]. Research in interactive narrative has tackled the moral aspect of stories [47, 2], but it has addressed moral values from the perspective of plot generation, without considering their relevance for characters' emotions.

In this paper, we propose to adopt the value-based emotional agent described in [3, 4] to model narrative characters, and describe a preliminary experiment conducted to test the suitability of this model for narrative situations where moral values are put at stake. In

this model, the agent's emotions are generated from the appraisal of how the contingent situation affects the agent's desires and values. The agent has an explicit representation of its moral dimension based on a value system [25], and a motivational dimension given by the desires it wants to pursue [7]. When choosing a course of action, the agent trades off its values against their desires, and relies on its moral values to evaluate its own behavior and the behavior of the others, feeling emotions like Reproach or Anger, or Gratification. For each narrative situation, the subjects were asked to act and feel according to the characters' beliefs and values, as if they were doing practice in an acting school. Then, we compared the actions and emotions they selected with the actions and the emotional states generated by the model, in order to gain insight on its validity and coverage.

The paper is organized as follows. After surveying the related work (Section 2), we describe the computational model of a character with moral values and emotions (Section 3) that we assess through the experiment. In Section 4 we describe the experiment design and the methodology by which the narrative scenarios employed in the experiment were developed. The description of the narrative scenarios is described in Section 5. In Section 6 we illustrate and discuss the results of the evaluation, making hypotheses about how the experiment design could be improved. Conclusions ends the paper.

2   Related Work

Many researchers tried to integrate computational models of emotions in a cognitive architecture for intelligent agents (of which [39, 21, 33] are some examples), with the aim of inserting emotions in BDI (Belief-Desire-Intention) agents [7]. Although different theories of emotions have been proposed (including physiological and dimensional models), most computational models are based on appraisal theory, in which cognitive processes are involved in the generation of emotions [36, 32, 41].

According to appraisal models, cognitive processes have the function of building a mental representation of the situation in which a person is involved (*person-environment relation*). This representation is not limited to the external environment, but also includes the internal disposition of a person, such as goals, desires, intentions, norms and moral standards. Emotions arise from the appraisal of the person-environment relation according to the appraisal dimensions that are defined by the theory (i.e. desirability of an event, praiseworthiness of an action).

According to the OCC model of emotions [36], *Joy* (*Distress*) emotion arises from being pleased (unpleased) about a desirable (undesirable) event. *Pride* (*Self-reproach*) emotion arises from the approval (disapproval) of one's own praiseworthy (blameworthy) action. *Admiration* (*Reproach*) emotion arises from the approval (disapproval) of someone's else praiseworthy (blameworthy) action. OCC model define also 'Compound emotion' *Gratification*, *Remorse*, *Gratitude* and *Anger*. Gratification (Remorse) emotion arises from Joy (Distress) and Pride (Shame), Gratitude (Anger) emotion arises from Joy (Distress) and Admiration (Reproach).

Moral emotions arise from evaluations in regard to moral principles [30, 36] and they have been argued to play a crucial role in decision making [26, 43, 31, 46, 14]. According to [26], moral sentiment serve as 'commitment devices' that lead agent to overcome selfish behaviors in favor of pro-social behaviors, which account for the compliance with social norms. The consequence for emotional characters is that they must balance their personal goals with their moral dimension for their behavior to be believable. Although encompassed in appraisal theories, most computational models [34] don't account for moral emotions (e.g. Pride, Shame) [30] but, mostly, focus on emotions related to the desirability/undesirability of situations with respect to goals, and don't account for moral values.

The EM system [39] integrates the OCC appraisal theory of emotions with a domain-independent approach, but the generation of moral emotions is based on the violation of goals such as *help my goals to succeed* or *do not cause my goals to fail*. ALMA [27] encoded the OCC theory with domain-dependent rules, thus failing to grasp general principles. The FLAME [20] architecture, based on OCC and Roseman appraisal theory [40], models emotions with fuzzy logic rules which map events and expectations onto emotional states (and behaviors). The system is not provided with the ability to independently assess the moral consequences of events and actions: instead, they are acquired through the user's feedback. In EMA [33], the first fully-implemented framework for conversational agents, appraisal is formed by a set of independent processes that operate on a plan-based representation of person-environment relation, named *causal interpretation*. This work is mainly based on Smith and Lazarus theory [32], where moral standards are not modeled. By lacking an explicit representation of moral dimension, the model fails to differentiate between different moral emotions (e.g. Shame from Remorse).

As argued by [23], emotional characters must balance their personal goals with their social environment in order to be believable. In the field of normative agents [12], a few works address moral emotions related to norm violation by casting norm violation as goal violation [29] or modeling norm violation in a domain specific way, thus lacking of flexibility [42, 9]. Regarding norms, one of the few exceptions to the trend of focusing on goal-related emotions is the work by Ferreira et al. [23]. In their work, they propose an agent model with an explicit representation of cultural and social norms, employed to check if actions violate or fulfill an activated norm. This appraisal process generate moral emotions such as Pride, Shame or even Anger (towards a target that violated a norm). Despite being able to generate moral emotions, this work focuses on the use of domain-specific cultural and social norms (e.g. not smoking in a bar) which are usually shared across a set of individuals. In our approach, we focus on the more generic concept of individual moral values, which can be easily adapted to new situations. Moreover, the work by [23] does not address the question of how to use the moral appraisals/emotions to guide decision-making.

The model we adopt in this work relies on the OCC model to establish an explicit link between moral values and moral emotions [4]. The agent is endowed with an explicit moral dimension formed by a scale of moral values (such as 'honesty', 'freedom') [25].

Basically, in the deliberation phase, the agent feels 'anticipatory' emotions, which allows it to envisage the consequences of its available options: the agent chooses the best option in the light of its *emotional reward*, i.e. the emotional states that each option would determine. Since the OCC model acknowledges a distinction between positive and negative emotional states, the agent will tend to prefer the lines of behavior that are more likely to make positive emotional states arise in it, avoiding ugly emotions [13] such as Shame or Remorse.

3 Values and emotions

Based on the work by [4], the character is modeled as a BDI agent [38] with a mental representation formed by beliefs, desires and moral values. Beliefs represent agent's knowledge about the world. Desires represent the agent's motivational component (i.e. what the agent desires to obtain), while moral values represent the character's moral dimension.

Inspired by [45], desires are associated with three different sets of conditions, namely *adopting conditions*, *success conditions* and *failure conditions*, which determine the adoption and achievement of desires. For example, when the agent believes that the adopting condition of a desire is verified in the world, the desire becomes an intention and can compete for being

selected by the agent. If adopted, the agent tries to devise a sequence of actions to reach the desire. When the agent believes that the success (failure) condition of the desire is verified in the world, the corresponding intention is achieved (unachieved).

Values are moral and ethical principles [36] that the agent consider important (e. g. honesty, freedom, family). According to [25, 15], moral values are subjective, and different individuals acknowledge different values arranged into subjective 'scales of values' (each agent's value is associated with a numeric priority). The set of values owned by the agent contributes to drive the behavior of the agent. Moral values are the moral drive of the agent, they constrain the behavior of the agent to its moral dimension and allow the agent to appraise the behavior of other characters. Every value is associated with a set of *violation conditions* that represent the states in which the value is at stake, i. e. something is happened in the world that makes the moral value violated.

In order to display a believable behavior, the agent's drive cannot be limited to its motivational component: the agent has to show that a moral dimension drives its behavior. According to [4], when the agent translates its goals into practical lines of behavior, the projection of these lines of behavior must also encompass the evaluation of the agent's own emotional states, such as Shame or Pride, that contribute to orientate the agent's choice towards value-compliant courses of actions (*anticipatory emotional appraisal*). The advantage of this integration is that the agent not only forms its goals based on the compliance with its values, but moral emotions become relevant when conflicting goals (and plans) are formed and must be traded off against the compliance with values.

Following [36], the *Appraisal Derivation* process evaluates the agent's mental representation of the world based on its goals and values, and outputs a set of appraisal variables. The *Affect Derivation* process determines what emotions arise from the appraisal variables according to the reference theory of emotions. The appraisal of events as desirable or undesirable depends on the processing of goals. A *desirability* (*undesirability*) variable is generated when a goal is achieved (unachieved) in the state of the world. The appraisal of actions is based on the processing of values: when a value is balanced (put at stake) in the current state of the world, the appraisal derivation model generates a *praiseworthiness* (*blameworthiness*) variable. Given the appraisal variables, the Affect Derivation Model generates emotions according to the following domain- independent rules (Figure 1).

- **Joy** if the appraisal variable *desirability* is generated (i. e. a goal is achieved);
- **Distress** if the appraisal variable *undesirability* is generated (i. e. a goal is unachieved);
- **Pride** and **Admiration** if the appraisal variable *praiseworthiness* is generated (i. e. an action re-balances a value at stake);
- **Shame** and **Reproach** if the appraisal variable *blameworthiness* is generated (i. e. an action puts a value at stake).

Following [36], when both appraisal variable regarding actions and goals are generated, the Affect Derivation Model generates the following compound emotions: **Gratification** (Joy and Pride), **Gratitude** (Joy and Admiration), **Remorse** (Distress and Self-Reproach), **Anger** (Distress and Reproach). The intensity of goal-related emotions is based on the importance of success and failure of goals multiplied by the effort made (i. e. the cost of the plan executed), while the intensity of value-related emotions derives from the importance of values.[1]

---

[1] The reader can refer to [17] for a complete example of how emotions intensity can be calculated.
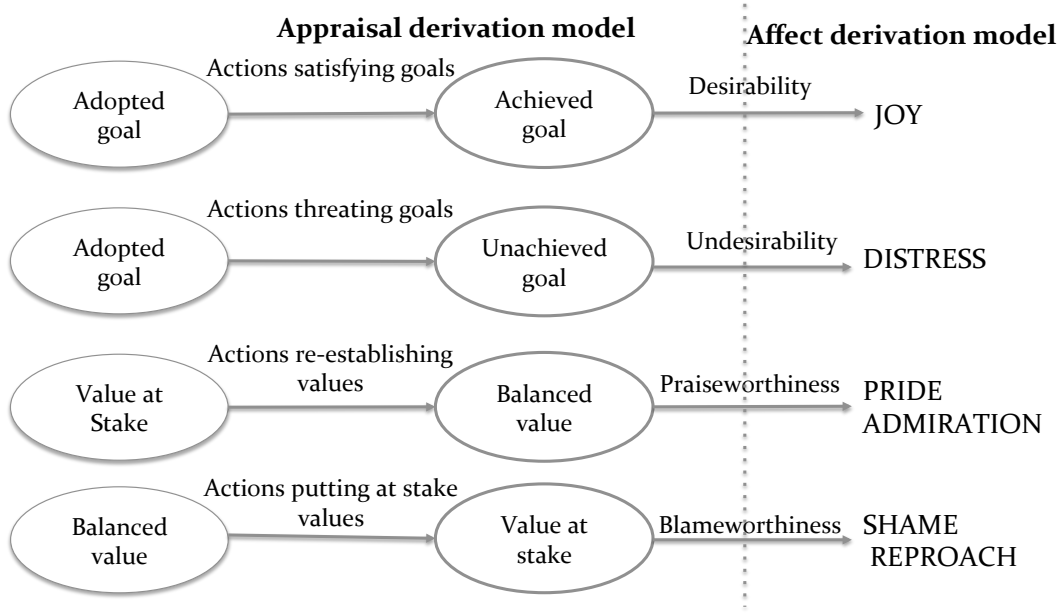
Figure 1 Appraisal and Affect Derivation model in [4].

In [4], emotional appraisal play a role also the agent's deliberation, i. e., the agent chooses a line of behavior in the light of the emotional states it would determine (emotional reward). Since the OCC model acknowledges a distinction between positive and negative emotional states, the agent will tend to prefer the lines of behavior that are more likely to make positive emotional states arise. The emotional reward derives from: (1) the intensity of the joy that the agent feels if it reaches an individual goal through a plan π (2) the distress intensity that the agent feels if, executing the plan π, some other adopted goals π has become unachievable; (3) the pride intensity that the agent feels if it re-establishes a value at stake through the plan and reaches the related moral goal; (4) the self-reproach intensity that the agent feels if the plan π threatens some other values. Given a plan π , we noted with G_A the set of individual goal satisfied by the plan, with G_T the set of individual goals threatened, with V_B the set of values re-established and with V_T the set of values put at stake. The intensity of anticipatory emotions **Joy** $EER_J(G_A, \pi_i)$, **Distress** $EER_D(G_T, \pi_i)$, **Pride** $EER_P(V_B, \pi_i)$ and **Shame** $EER_S(V_T, \pi_i)$ are:

$$EER_J(G_A, \pi_i) = \frac{P(\pi_i) * \sum\limits_{g_a \in G_A} ImpOfS(g_a)}{E(\pi_i)} \tag{1}$$

$$EER_D(G_T, \pi_i) = \frac{P(\pi_i) * \sum\limits_{g_t \in G_T} ImpOfF(g_t)}{E(\pi_i)} \tag{2}$$

$$EER_P(V_B, \pi_i) = \frac{P(\pi_i) * \sum\limits_{v_b \in V_B} (r(v_b) + d(v_b))}{E(\pi_i)} \tag{3}$$

$$EER_S(V_T, \pi_i) = \frac{P(\pi_i) * \sum\limits_{t_t \in V_T} (r(v_t) + d(v_t))}{E(\pi_i)} \tag{4}$$

where $P(\pi_i)$ is the plan probability of success, *impOfS* and *impOfF* are the importance of success and failure of the goal, $E(\pi_i)$ is the cost of the plan, $r(v_b)$ is the priority of the value and $d(v_b)$ is the degree with which the value is shared with the society. The models assumes a partial-ordering continuous planner extended with emotions like [1], in which operators are specified in an extended STRIPS-like notation [24]. Differently from a classical STRIPS operator where preconditions identify the set of states in which the action can be executed, and effects describes how the environment changes as a result of taking the action [24], an extended STRIPS-like operator associates stochastic effects to actions [6], so that the probability that a plan reaches a goal state can be calculated. The planner monitors all events in the world in order to detect when an action is accomplished or fails, updates all the plans and the probability of action effects according to the event perceived. The function $E(\pi_i)$ is calculated on the basis of the cost associated to the actions; the simplest case is the unitary cost, i.e. the cost of the plan is equal to number of actions presented in the plan. The other quantities used in the formulas given above, such as the priority of the value $r(v_b)$ for example, are specified at design time. Finally, given the emotional reward, the overall plan utility is computed as:

$$EER = (EER_J + EER_P) - (EER_S + EER_D) \tag{5}$$

For example, consider the following situation. Boo has the goal to eat a chocolate candy; in order to satisfy her goal, the chocolate candy must be stolen from Mary, but the *steal* action makes the violation condition of the value *honesty* true. So, if Boo executes her plan, the emotional reward utility will derive from the combined intensity of Joy and Shame. Let us consider another plan, in which Boo asks Mary to give her the chocolate candy. In this case no value is put at stake and the emotional reward utility will derive from the Joy intensity only. If the value *honesty* is very important for Boo, she chooses the plan to ask Mary the chocolate candy, even if the plan has a lower probability of success.

Note that, if the alternative plans have the same emotional reward utility, the characters is in a dilemma. In [3] the model is employed to deal with faces a moral dilemma.

4 Experiment design

In order to test the suitability of the value-based model of emotions described in [4] for modeling the behavior of narrative characters, we performed a preliminary evaluation on narrative scenarios with human subjects. The experiment we designed relies on a direct comparison between the predictions of the model about the characters' emotions and actions and the expectations of the human subjects about them. In order to evaluate if the model correctly predicts how a character's behavior and emotions are affected by its moral values, we compared the emotions and actions generated by the model with the emotions and actions that human subjects expected from the character.

The experiment was conducted online, via a text-based web interface (Figure 2).[2] For each scenario, a short text introduced the character and her/his values, then a narrative situation was described that put a stake the character's values. The scale of values was presented to participant not in a numerical format but with a figurative scale, in order to make the values priorities apparent at first glance. The task of identifying the expected course of action and emotions for the character was introduced to the participants as a game:

---

[2] The website of the experiment can be found at `http://www.ilnomedellarosa.it/ActorStudio` (in Italian only).

LE ROSE DELLO ZIO GEORGE

WALLACE E CHARLIE SONO DUE CUGINI E AMICI PER LA PELLE. I DUE VIVONO IN CAMPAGNA DOVE LO ZIO GEORGE HA UN VIVAIO DI PREZIOSISSIME ROSE DA CONCORSO DEL QUALE È MOLTO GELOSO. CHARLIE VUOLE FARE UN REGALO ALLA SUA RAGAZZA E CHIEDE A WALLACE DI DARGLI LA CHIAVE DEL VIVAIO PER PRENDERNE UNA.

LA SCALA DI VALORI DI WALLACE:

I. ONESTÀ

2. LEALTÀ VERSO CHARLIE

PUNTEGGIO

PUNTI EMOZIONE: 0
PUNTI AZIONE: 0

PUNTI TOTALI: 0

WALLACE SA DOVE È LA CHIAVE MA SA ANCHE CHE LO ZIO GEORGE NON VUOLE CHE ENTRI NESSUNO...

GIOCA

■ **Figure 2** The first page of the experiment website, in which the narrative scenario (a narrative situation and the main character's scale of values) is presented to a user.

the participants were asked to pretend they were exercising identification in an acting class, in order to leverage their capability to take the point of view of the character and behave "as if" they were the in the character's shoes. For each scenario, a pair of alternative actions were submitted to the participants, who also had to select a set of appropriate accompanying emotions.

The narrative situations included in the experiment were created by a story editor by taking inspiration from literary stories. By doing so, we wanted to reduce the arbitrariness of the relation between, on the side, the characters' goals and values, and, on the other side, their actions and emotions. Each narrative situation (story world, story incidents and participants) was encoded in formal terms as described in Section 5 and the value-based emotional model to be tested was employed to generate the behavior of the main character. The resulting behavior and emotions were submitted to the story editor to verify that they were consistent with the original story.

The goal of the experiment was twofold: first, we wanted to assess the role of values in action selection, i.e. if, given a scale of values and a narrative situation, the course of action selected by the participants matched the course of action generated by the model (Question 1); second, since the model postulates that emotions mediate (through the notion of anticipatory appraisal, see Section 3) the role of values in action selection, we wanted to assess if the emotions that participants attributed to characters matched the emotions generated by the model (Question 2).

After going though three scenarios, participants were given a post questionnaire in which we asked information about their age, sex, etc. Moreover, for each value oppositions involved in the scenarios, we asked them to indicate what value they preferred (i.e. "What is more important according to you, honesty or loyalty?"). By doing so, we wanted to measure the degree to which the choice of actions and emotions made by the participants was reliable, i.e., if it was affected by their own scale of values instead of being driven only by the identification process.

Scenario	Question	Action	Values
One **Wallace and** **uncle George's** **roses**	What would you do instead of Wallace, given his scale of values?	Refuse to give the key to Charlie ———————————— Give the key to Charlie	Honesty Loyalty to Charlie
Two **At school!**	What would you do instead of Tom, given his scale of values?	Umiliate Pier, taking revenge on him ———————————— Let it go	Justice Pity
Three **A difficult** **choice**	What would you do instead of Mark, given his scale of values?	Stay in New York ———————————— Go to Italy and stay with the family	Happiness Family

■ **Figure 3** Available actions for each scenario (see descriptions in Section 5).

In order to promote the participants' identification with the characters, after each task the web based system attributed them a score based on their "performance" (i. e., the coincidence of the selected actions and emotions given the ones predicted by the model).

Measures: In Figure 3 we illustrate the pairs of actions that are opposed in each scenario. A detailed description is provided in Section 5.

For emotions, we adopted the emotions categories encompassed by the OCC theory of emotions [36] (see Section 3). We described the emotions to the participants by specifying, for each emotions, the target of the emotion and its appraisal dimension (e. g. "Pride arises from an appraisal of somebody's action as praiseworthy").

In the post questionnaire, we asked participants general information, i. e., if the scenarios were difficult to read and to understand, and what value they prefer between the values in conflict of each scenario (e. g.   e. g., Honesty and Loyalty). Summarizing, from the experiment execution, we get actions and emotions selected by users, and a short text describing motivations of their choices; from the post-questionnaire, we get the preference values for each value opposition, and an evaluation of the clarity and readability of the scenarios.

Participants.   A convenience sample of 42 Italian subjects, 18 female and 24 male, aged 23-65, participated in the experiment. Participants had high levels of computer literacy (60% described themselves as being expert), they had previously interacted with virtual characters and usually played video-game (40% declared having interaction with virtual character at least one day a week, 45% declared playing video-games at least three days a week).

Experimental Protocol.   The participants played the first scenario. First, they read the summary of the narrative situation; then they chose the action and the emotions for the main character. After the participant made his/her choice, the system showed the emotions and actions generated by the value–based emotional agent model. The same for the second and the third scenario. For each scenario, we asked the participants to describe the motivations underlying their choices, in order to perform a qualitative analysis on them. After running the three scenarios, participant answered to the post questionnaire.

5   Narrative scenario examples

The narrative scenarios designed for the experiments are characterized by a conflict of goals and values. Created with the help of a drama expert, they are inspired by well known literary works. The agent model described in Section 3 was employed to model the behavior of a character in each narrative situation. The model was implemented into FAtiMa [16], a modular architecture designed to develop emotional agents. For each scenario, two lines of behavior were generated by altering the value priority, but only one matched the actual character's behavior in the narrative situation that inspired the scenario. Note that, being a preliminary evaluation, we simplified the anticipatory emotional appraisal by modeling scenarios with plans formed by only one action (i.e. with an unitary cost) and we assumed that plans have a success probability of 1.0 (in other words, that they cannot fail).

The first scenario, 'Wallace and uncle George's roses', is inspired by the 'nunnery scene' in Shakespeare's Hamlet [5], where one of the drama main characters, Ophelia, has to decide whether to lie to the protagonist, Hamlet, thus putting at stake the value 'Honesty', or to reveal the truth, thus putting at stake the 'Loyalty' towards her father.

The second scenario and third scenario, 'At school' and 'A difficult choice', were inspired by the thirty-six dramatic situations described in Polti's work [37]. From a large repository of plays, Polti extracts a list of situations that are perceived as intrinsically dramatic. Each situation is named after a specific action (e.g. Vengeance). Within each situation, Polti defines: the kind of agents (e.g. victim, culprit), the beliefs and goals that motivate the action (e.g. the agents' cognitive states) and the emotions felt by agents, then lists a set of literary examples.

In particular, 'At school' is based on the third situation 'Crime pursued by vengeance' in which Polti argues that vengeance is a divine Joy felt by those who pursue it after being victim of a crime with no guilt. "A difficult choice" is based on the thirty-fourth situation 'Remorse' in which the culprit feels Remorse for something she/he committed. The choice of modeling scenarios inspired by well known narrative situations, instead of employing literary stories, resides in the motivation that participants may know the literary stories and what the characters do in them, thus negatively affecting the experimental methodology.

5.1   Scenario one, Wallace and uncle George's roses

Summary: Wallace and Charlie are cousins. They live in the country where uncle George has a nursery of precious roses he brings to gardening contests. Uncle George is very jealous of his roses. Charlie wants to make a gift to his girlfriend and asks Wallace to give him the key of the nursery to get one. Wallace knows where the key is but he also knows that uncle George does not want anyone to enter.

In this scenario, Wallace has to choose whether to be loyal with his cousin Charlie or to uncle George (Figure 4). Wallace owns the value 'Honesty' with 7.0 priority and the value 'Loyalty to Charlie' with 8.5 priority. During his reasoning cycle, the system (in the role of Wallace) finds two plans: the plan $p1$ contains the action of giving the key to Charlie, thus deceiving uncle George; the plan $p2$ contains the action of refusing to give the key to Charlie. The plan $p1$ puts at stake the value 'Honesty' (Figure 4): if Wallace executes this plan he will feel Shame emotion for putting at stake this value.

On the other hand, the plan $p2$ puts at stake the value 'Loyalty' (Figure 4). If Wallace executes this plan, he will feel Shame emotion for putting at stake this value. Wallace is in a dilemma. Assuming that the two plans have the same probability of success, Wallace's anticipatory appraisal leads him to choose the plan with the highest *EER*: in any case,
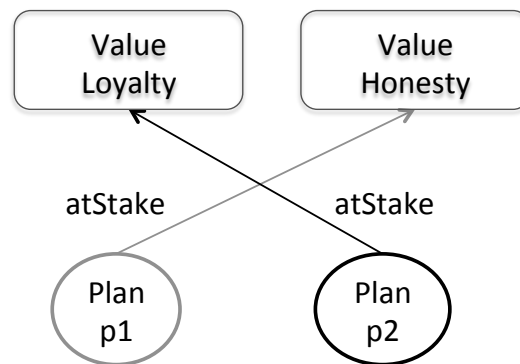
Figure 4 Scenario 1 (Wallace): the plan $p1$ puts at stake the value 'Honesty', while the plan $p2$ the value 'Loyalty'.

Wallace will feel Shame, but the anticipatory appraisal leads him to choose the course of actions that brings him to a state in which the Shame intensity is lower. Wallace executes the plan $p1$: he gives the key to Charlie and feels Shame for putting at stake the value 'Honesty'.

5.2 Scenario two, At school!

Summary: Tom is bullied by his classmate Pier. Pier has taken from Tom the role of director of the school newspaper, putting around lies about him. The result is that Tom lost the director position and he is now in a bad light. A few days later, Tom sees that Pier has forgotten his backpack with all his stuff in the locker room. Tom digs in Pier's backpack and finds evidence that Pier copied the class test. Tom is now uncertain about what to do, whether to take revenge against Pier or to pass through this situation.

In this scenario, Tom has to choose if he wants to take vengeance or not, Figure 5). Tom owns the value 'Justice' with 8.5 priority and the value 'Pity' with 7.5 priority. During his reasoning cycle, Tom finds two plans: the plan $p1$ contains the action of humiliating Pier; the plan $p2$ contains the action of letting it go. The plan $p1$ puts at stake the value 'Pity' (Figure 5): if Tom executes this plan, he will feel Shame for putting at stake this value. But the plan $p1$ also brings back to balance the value 'Justice', put at stake by Pier, and satisfies the goal of being the director again. In addition to Shame, Tom will feel Joy for satisfying his goal and Pride for restoring his value 'Justice'. The activation of these emotions at the same time gives the compound emotion Gratification.

On the other hand, the plan $p2$ has no effects (Figure 5). If Tom performs this plan, the situation doesn't change and the value 'Pity' is not put at stake. Note that, according to our model, in the past Tom felt Anger toward Pier for being mean to him (Pier put at stake one of Tom's values thus generating a Reproach emotion in Tom and made his goal of being a director unachievable, thus generating an emotion of Distress). In this case, these emotions will continue to decay and Tom won't feel any new emotions.

Assuming that the two plans have the same probability of success, Tom's anticipatory appraisal component chooses to execute the plan of taking revenge on Pier: Tom will feel Shame for putting at stake the value 'Pity', but the anticipatory appraisal chooses the course of actions that brings Tom in a state of affairs in which the overall emotion intensity is the highest. So, Tom executes the plan $p1$, he takes revenge, satisfies his goal, restores his value 'Justice' and puts at stake the value 'Pity'. Thus, Tom feels Shame and Gratification emotion (Joy and Pride).
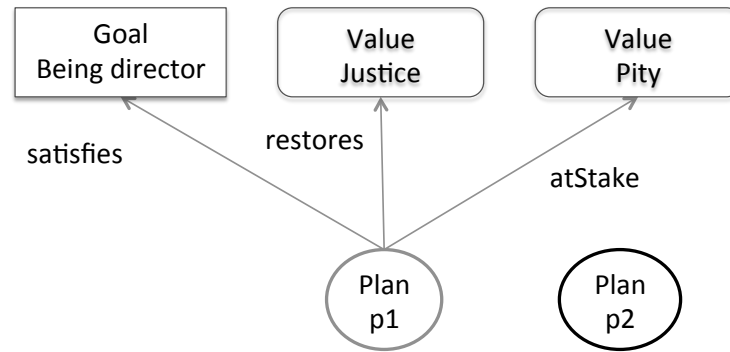
■ **Figure 5** Scenario 2: Tom's plan $p1$ satisfies the goal and value 'Justice' but puts at stake the value 'Pity'. Plan $p2$ has no such effects on values and goals.

5.3 Scenario three, A difficult choice

Summary: New York, 2003. Mark and Lucy are married and they have a beautiful baby. Lucy has agreed to go a couple of years in Italy for the job of her dreams: working as a curator of a famous art gallery in Rome. Mark, however, has always wanted to be judge in New York. Just when Lucy has officially accepted her job in Italy, Mark gets the seat as a judge in New York. Mark's desires are of being with his family and having the work of his dreams as well. Now, he has to choose whether to have the job or to stay with his family.

Mark (Figure 6, Figure 7) owns the value 'Family' with 8.0 priority and the value 'Happiness' with 8.5 priority. The goal of being with his family has an importance of success equals to 8.0, while the goal of being a judge has an importance of success of 8.5. During his reasoning cycle, Mark finds two plans: the plan $p1$ contains the action of staying in New York without his family; the plan $p2$ contains the action of going to Italy. The plan $p1$ puts at stake the value 'Family', threatens the goal of being with the family and satisfies the goal of being a judge (Figure 7): if Mark executes this plan he will feel Shame for putting at stake a value, Distress for threatening the goal of being with the family, but Joy for satisfying his goal of being a judge. According to our model, Mark feels a Remorse emotion, since he feels a Shame emotion and a Distress emotion at the same time.

On the other hand, the plan $p2$ puts at stake the value 'Happiness', threatens his goal of being a judge and satisfies the goal of being with the family (Figure 7). If Mark executes this plan he will feel Shame for putting at stake the 'Happiness' value, Distress for threatening his goal of being a judge, but Joy for satisfying his goal of being with the family. In this case, Mark feels a Remorse emotion because he feels a Shame emotion and a Distress emotion at the same time.

Assuming that the two plans have the same probability of success, Mark's anticipatory appraisal component would choose to execute the plan that puts at stake the value with a lower priority ('Family'), due to the equal importance of success of both goals: in any case Mark will feel Shame, but the anticipatory appraisal would make him choose the course of actions that brings him in a state of affairs in which the Remorse intensity is lower. This scenario is very similar to the first scenario, because the two courses of actions that the character can choose bring the character to feel the same emotion, although for different reasons and with different intensity.

Character	Goals	Values	PLAN	EER	Emotions
Mark	Being judge(importance 8.0) Success conditions: <I'm a judge> Failure conditions: <…,go to Italy,…>	Happiness (priority 8.5) Violation conditions: <…,give up dreams, …>	P1: <stay New York>	P1 effects: at stake Family Satisfied goal being judge	Joy Distress Shame Remorse
	Being with my family (imp. Of Success 8.0) Success conditions: <I'm with my family> Failure conditions: <…,stay in New York,…>	Family (priority 7.0) Violation conditions: <…,abandoning family, …>	P2: <go to Italy>	P2 effects: at stake Happiness Satisfied goal being with famiy	Joy Distress Shame Remorse

■ **Figure 6** Representation of Mark's goals, values and plans. For each value, the violation condition is specified that would hold in the state of the world which obtains if Wallace executes the action that puts at stake that value. In the EER column, we specify the action effects by taking into account in the generation of the Expected Emotional Reward utility for the action. We assume a unitary cost and same probability in the EER calculation in order to simplify the experiment.
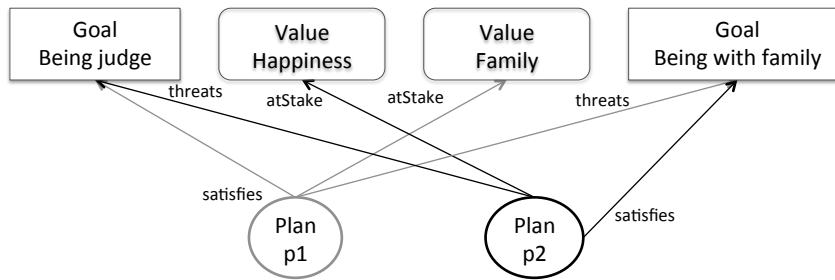

■ **Figure 7** Scenario 3: Mark's plan $p1$ makes the goal of being a judge achieved, but puts at stake the value 'Family' and makes unachieved the goal of staying with the family. Plan $p2$ makes the goal of staying with the family achieved, but puts at stake the value 'Happiness' and causes the goal of being a judge to fail.

6 Results & Discussion

Given the narrative scenarios described in the Section 5, we run the experiment described in the Section 4 and compared the actions and emotions chosen by the human participants with the actions and emotions generated by the value-based emotional agent. In the following, we refer to the action that the computational model chooses in each scenario with the term 'right action' while, with the term 'wrong action', we refer to the action that the model doesn't choose to perform. Notice that this labeling is adopted only to simplify the description of the results: the 'right' behavior is the one that matches the actual behavior of the character in the narrative situation that inspired the scenario, but no choice can be defined as right or wrong because the scale of values and the relation between values and actions are intrinsically subjective [25].

Quantitative results. In order to assess if there is a correlation between the effectiveness of the identification process by the subjects and their inclination to feel certain emotions, we run a non parametric test (Mann-Whitney test) on the results of each scenario, dividing the group between the subjects who chose the right action and those who didn't, and related
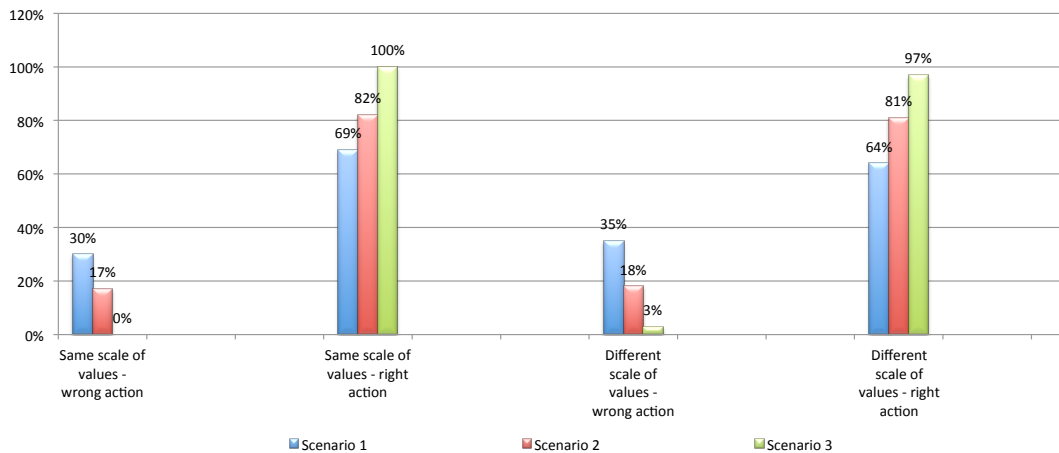
Figure 8 Frequencies on users choices about actions in each scenarios based on the same scale of values and on the different scale of values.

them with the selected emotions. In the first scenario, we found evidence that there is a significant difference (U= 71, $n_1 = 13$, $n_2 = 29$, p one tailed < 0.01) between the subjects who selected the Pride emotion by choosing the wrong action and the subject who selected the Pride emotion by choosing the right action. Subjects who chose the wrong action are more inclined to feel a Pride emotion. We found no other statistical evidence in other scenarios. In order to assess if a different individual scale of values may have affected the identification process, we run a Mann-Whitney test specific to each scenario, dividing the group between the subjects who have the same scale of values of the characters and the subject who haven't the same scale of values as the characters. We found no significant evidence about selected actions by users with the same scale of values of the character and with the different scale of values. This result is in line with the theories of narrative engagement [28].

Qualitative results. The most part of the subjects asserted that the narrative scenarios were clear (81%), they had no difficulties in identifying with the characters (72%).

The actions chosen by the character model (the 'right' ones) were chosen also by the large majority of the subjects (81%), so the answer to Question 1 (Section 4, if given a scale of values and a narrative situation, the course of action selected by the participants matched the course of action generated by the model) was positive. Results show that the subjects started to understand the mechanism of the game as they proceeded through the tasks, so a larger group chose the right action in the second and third scenario than in the first scenario. As showed by the Mann-Whitney test, both subjects with the same scale of values as the characters and those with a different scale of values mainly chose the action predicted by our model (Figure 8). For example, in the first scenario, most participants with the same scale of values as Wallace (38%) selected the action predicted by our model (69%); however, most participants with a different scale of values selected the action predicted by our model as well (64%). These results show that subjects understand the mechanism of the game, and that the participants didn't have difficulties in identifying with the characters and to reason with their scale of values. The results also show that the participants substantially agreed with the prediction of our model, for both the action selection (determined by the anticipatory appraisal component) and the emotion generation (determined by the emotion appraisal model).
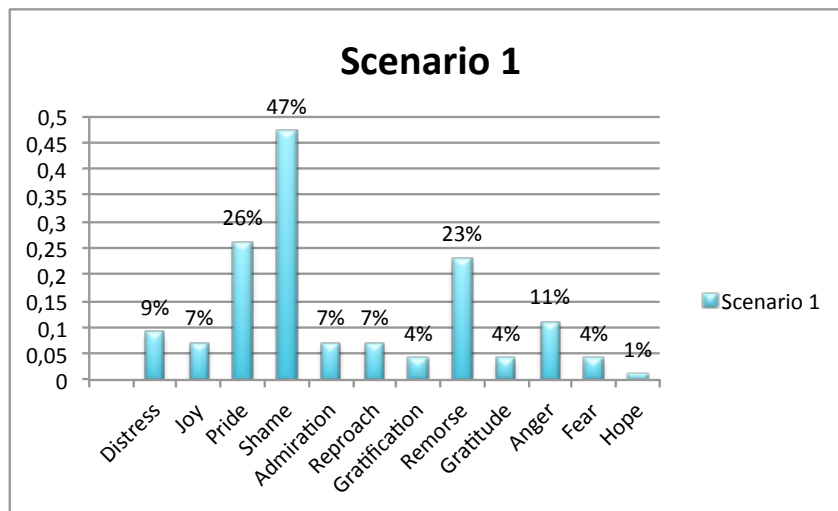
Figure 9 Frequencies of emotions chosen by the participants in the first scenario (we don't distinguish about subjects choosing the right or wrong action because the appraisal of emotions is identical).

Regarding the emotions selected in each scenario, the results show that the subjects agreed with the emotions generated by the value based emotional character (Question 2, Section 4, if the emotions that participants attributed to characters matched the emotions generated by the model).

In the first scenario, the subjects selected the Shame emotion (47%) and the Remorse emotion (23%)(Figure 9). They selected also Pride emotion (26%), that the character model does not generate in the context of the first scenario. Examining the motivations given by participants, we argue that people perceive the preservation of a value as a motivation for feeling Pride, a fact that is reasonable, although not covered by the current character model:

- pride for having acted in the right way (subject chose the wrong action, subject preferred 'Honesty');
- pride for helping a friend (subject chose the right action, subject preferred 'Honesty');
- pride for being loyal to my uncle (subject chose the wrong action, subject preferred 'Loyalty').

Subjects (23%) selected also Remorse emotion, with the motivations that 'I'm doing something that I don't want to do':

- remorse for doing something wrong (subject chose right action, subject prefers 'Honesty);
- remorse for betraying my uncle (subject chose right action, subject prefers 'Honesty');

Following OCC model, the character model generated Remorse only when the failure of a goal was involved in the appraisal. Results suggest us that the participants associated a sort of high-level goal (e.g. 'don't violate my standards' when the values in conflict both have a high priority).

In the second scenario, the participants selected the emotions predicted by our model (Joy 21%, Pride 26%, Gratification 27%). The participants selected also the emotions felt by Tom towards Pier (Anger 15%, Reproach 25% but not Distress 7%) (Figure 10). This is in line with the character model: Pier performed a blameworthy action against Tom (Reproach), making his goal of being a director no more achievable (Distress and, consequently, Anger).

In the third scenario, the emotions selected by the participants agree with the emotions generated by to the character model (Distress 31%, Joy 45%, Shame 26%, Remorse 23%)
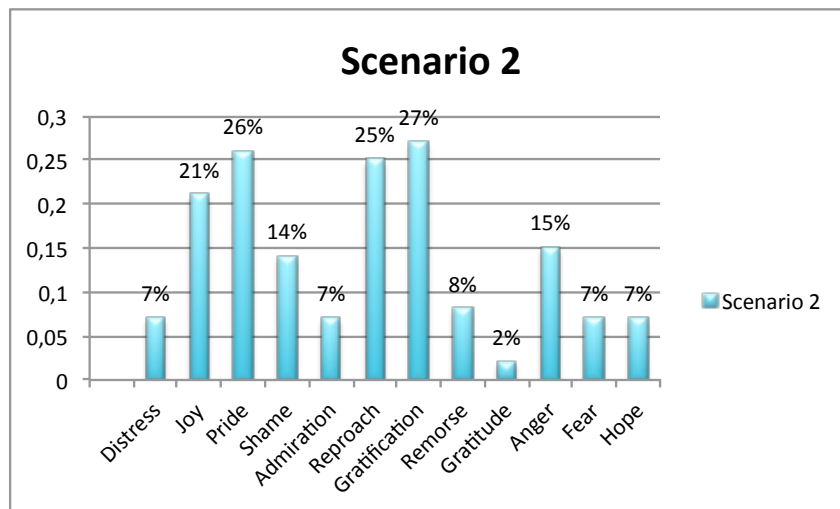
Figure 10 Frequencies of the emotions chosen by the participants in the second scenario (only the participants that made the right decision).

(Figure 11). Fear and Hope emotions were selected by the most part of participants. Examining the motivations they provided, we found that the fear was the "fear of the future" and "hope in a better future".

- hope in the future, maybe I can gather the family again in the future (participant chose right action);
- fear for the future, my family may have problems without me (participant chose right action).

Discussion.   The results suggest that our model is congruent with partipants' choices. We need to study into depth the difference that the experimental subjects find between putting at stake a value and safeguarding a value from being put at stake. The results suggest that the emotions related with the praiseworthiness of actions can arise when a value is preserved from being put at stake. However, in order to obtain more reliable results, the narrative scenarios should be provided to the participants in random order. Even if no significant difference was found among the results obtained in the three scenarios, the qualitative observations suggest that the participants may have tuned to the experiment setting across the scenarios.

The motivations expressed by the participants for their choices suggest relevant improvements to the emotional agent model. The model can be extended it to grasp the difference between the situation in which a value at stake is re-established and the situation in which a value in balance – but threatened! – is preserved. Summarizing, the results are encouraging for the character model: the Anticipatory Emotional Appraisal seems to be acknowledged by the most part of participants and the emotions generated seem to be congruent with the emotional agent model employed in the experiments, and with the relations with goals and values predicted by this model.

We consider the results of our experiment promising, and we plan to run a more complex experiment in which we can assess other aspects of the emotional agent model, including the evaluation of cost and probability of plans, in order to get a complete evaluation of the anticipatory appraisal formulas. In a more complex evaluation, for each scenario, a group of participant should be given a choice of actions that include an action associated to a 'noise' value, so that a comparison may be conducted between the participants who evaluate
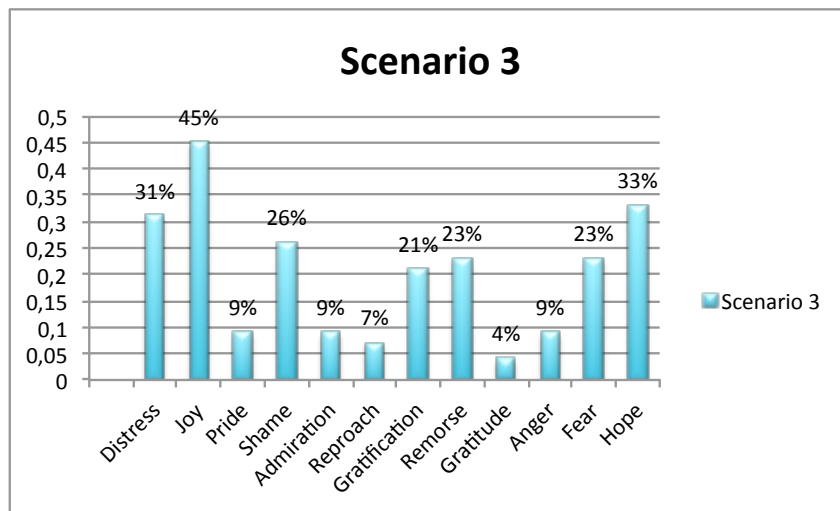
Figure 11 Frequencies of the emotions chosen by the participants in the third scenario (we don't distinguish about users choosing right or wrong action because the appraisal of emotions is identical).

the standard scenario and the participants that evaluate the scenario with the 'noise' value. Another experiment is to test if, given the action chosen by a character, the participants associate it with the value encoded in the computational model, in order to assess (or learn) the action-value associations.

7 Conclusion

In this paper, we describe an experiment devised to validate the appropriateness of a value based emotional agent model for modeling narrative characters. The experiment compared the expectations of human subjects about characters' behavior and emotions with the predictions of the agent model. The results of the experiment show that the users' expectations meet the predictions of the model.

The results also provide important insights on the relation between values and emotions, that we will address in the future work. The character model can be improved by extending it to grasp the difference between the situation in which a value at stake is re-established and the situation in which a value in balance – but threatened! – is preserved. Also, we need to study in depth how the difference between Remorse and Shame emotions is perceived by the audience in narrative situations.

References

1 Ruth Aylett, Joao Dias, and Ana Paiva. An affectively driven planner for synthetic characters. In *Icaps*, pages 2–10, 2006.

2 Heather Barber and Daniel Kudenko. Generation of adaptive dilemma-based interactive narratives. *Computational Intelligence and AI in Games, IEEE Transactions on*, 1(4):309–326, 2009.

3 Cristina Battaglino and Rossana Damiano. Emotional appraisal of moral dilemma in characters. In *ICIDS*, pages 150–161, 2012.

4 Cristina Battaglino, Rossana Damiano, and Leonardo Lesmo. Emotional range in value-sensitive deliberation. In *AAMAS*, pages 769–776, 2013.

5   Harold Bloom. *Shakespeare : the invention of the human.* Riverhead Books, New York, 1998.

6   Craig Boutilier, Thomas Dean, and Steve Hanks. Decision-theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research*, 11:1–94, 1999.

7   M. E. Bratman. *Intention, plans, and practical reason.* Harvard University Press, Cambridge Mass, 1987.

8   J. Bruner. The narrative construction of reality. *Critical Inquiry*, 18(1):1–21, 1991.

9   Vadim Bulitko, Steve Solomon, Jonathan Gratch, and Michael van Lent. Modeling culturally and emotionally affected behavior. In *AIIDE*, 2008.

10   Joseph Campbell. *The Hero with a Thousand Faces.* Princeton University Press, Princeton, 1949.

11   Noel Carroll. Art and mood: preliminary notes and conjectures. *The monist*, 86(4):521–555, 2003.

12   Cristiano Castelfranchi, Frank Dignum, Catholijn M. Jonker, and Jan Treur. Deliberative normative agents: Principles and architecture. In Nicholas R. Jennings and Yves Lespérance, editors, *ATAL*, volume 1757 of *Lecture Notes in Computer Science*, pages 364–378. Springer, 1999.

13   Ilona E. De Hooge, Seger M. Breugelmans, and Marcel Zeelenberg. Not so ugly after all: when shame acts as a commitment device. *Journal of personality and social psychology*, 95(4):933, 2008.

14   Celso M. de Melo, Liang Zheng, and Jonathan Gratch. Expression of moral emotions in cooperating agents. In Zsófia Ruttkay, Michael Kipp, Anton Nijholt, and Hannes Högni Vilhjálmsson, editors, *IVA*, volume 5773 of *Lecture Notes in Computer Science*, pages 301–307. Springer, 2009.

15   Morteza Dehghani, Emmett Tomai, Kenneth D. Forbus, and Matthew Klenk. An integrated reasoning approach to moral decision-making. In Dieter Fox and Carla P. Gomes, editors, *AAAI*, pages 1280–1286. AAAI Press, 2008.

16   Joao Dias, Samuel Mascarenhas, and Ana Paiva. Fatima modular: Towards an agent architecture with a generic appraisal framework. *Proceedings of the International Workshop on Standards for Emotion Modeling*, 2011.

17   João Dias and Ana Paiva. Agents with emotional intelligence for storytelling. In Sidney K. D'Mello, Arthur C. Graesser, Björn Schuller, and Jean-Claude Martin, editors, *ACII (1)*, volume 6974 of *Lecture Notes in Computer Science*, pages 77–86. Springer, 2011.

18   Denis Diderot. *Paradoxe sur le comédien.* Sautelet, 1830.

19   L. Egri. *The Art of Dramatic Writing.* Simon and Schuster, New York, 1946.

20   Magy Seif El-Nasr, John Yen, and Thomas R. Ioerger. Flame–fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-Agent Systems*, 3:219–257, September 2000.

21   Clark D. Elliott. *The affective reasoner: a process model of emotions in a multi-agent system.* PhD thesis, Northwestern University, Evanston, IL, USA, 1992.

22   Susan L. Feagin. *Reading with feeling: The aesthetics of appreciation.* Cornell University Press, 1996.

23   Nuno Ferreira, Samuel Mascarenhas, Ana Paiva, Gennaro Di Tosto, Frank Dignum, John Mc Breen, Nick Degens, Gert Hofstede, Giulia Andrighetto, and Rosaria Conte. An agent model for the appraisal of normative events based in in-group and out-group relations. In *AAAI-13 CogSys – Twenty-Seventh AAAI Conference on Artificial Intelligence.* AAAI-13 Twenty-Seventh AAAI Conference on Artificial Intelligence, July 2013.

24   Richard E. Fikes and Nils J. Nilsson. Strips: A new approach to the application of theorem proving to problem solving. *Artificial intelligence*, 2(3):189–208, 1972.

25 Bas C. Van Fraassen. Values and the heart's command. *Journal of Philosophy*, 70(1):5–19, 1973.

26 Robert H Frank. Introducing moral emotions into models of rational choice. In *Feelings and Emotions: The Amsterdam Symposium, Jun, 2001, Amsterdam, Netherlands*. Cambridge University Press, 2004.

27 Patrick Gebhard. Alma: a layered model of affect. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, AAMAS'05, pages 29–36, New York, NY, USA, 2005. ACM.

28 Alessandro Giovannelli. In sympathy with narrative character. *Journal of Aesthetics and Art Criticism*, I(67):83–95, 2009.

29 Jonathan Gratch, Wenji Mao, and Stacy Marsella. *Modeling social emotions and social attributions*. Cambridge University Press, 2006.

30 Jonathan Haidt. The moral emotions. *Handbook of affective sciences*, pages 852–870, 2003.

31 David Hume. *A Treatise of Human Nature (Oxford Philosophical Texts)*. Oxford University Press, USA, February 2000.

32 Richard S. Lazarus. *Emotion and Adaptation*. Oxford University Press, USA, August 1991.

33 Stacy C Marsella and Jonathan Gratch. Ema: A process model of appraisal dynamics. *Cognitive Systems Research*, 10(1):70–90, 2009.

34 Stacy C. Marsella, Jonathan Gratch, and Paola Petta. Computational models of emotion. In K. R. Scherer, T. Bänziger, and Roesch, editors, *A blueprint for an affectively competent agent: Cross-fertilization between Emotion Psychology, Affective Neuroscience, and Affective Computing*. Oxford University Press, Oxford, 2010.

35 R. McKee. *Story*. Harper Collins, New York, 1997.

36 Andrew Ortony, Allan Collins, and Gerald L. Clore. *The cognitive structure of emotions*. Cambridge University Press Cambridge [England] ; New York, pbk. ed. edition, 1988.

37 G. Polti. *Les trente-six situations dramatiques*. Mercure de France, Paris, 1895.

38 S. A. Rao. Georgeff. mp, bdi agents: From theory to practice. In *Conference Proceedings of 1st international conference on multiple agent system*, pages 312–319, 1995.

39 W. Scott Reilly and Joseph Bates. Building emotional agents, 1992.

40 Ira J Roseman. Appraisal determinants of discrete emotions. *Cognition & Emotion*, 5(3):161–200, 1991.

41 K. R. Scherer. The role of culture in emotion-antecedent appraisal. *Journal of Personality and Social Psychology*, 73:902–922, 1997.

42 Mei Si, Stacy C. Marsella, and David V. Pynadath. Modeling appraisal in theory of mind reasoning. *Autonomous Agents and Multi-Agent Systems*, 20(1):14–31, 2010.

43 Adam Smith. The theory of moral sentiments, 1984.

44 Greg M. Smith. *Film Structure and the Emotion System*. Cambridge University Press, Cambridge, 2003.

45 M. B. van Riemsdijk, M. Dastani, and M. Winikoff. Goals in Agent Systems: A Unifying Framework. In *Proceedings of AAMAS'08*, 2008.

46 Alicia E. Vélez García and Feggy Ostrosky-Solís. From morality to moral emotions. *International Journal of Psychology*, 41(5):348–354, 2006.

47 Stephen G. Ware. A computational model of narrative conflict. In *Proceedings of the 6th International Conference on Foundations of Digital Games*, pages 247–249. ACM, 2011.