Telling the Difference Between Asking and Stealing: Moral Emotions in Value-Based Narrative Characters

Cristina Battaglino
Dept. of Computer Science
Università di Torino, Turin, Italy
battagli@di.unito.it

Rossana Damiano
Dept. of Computer Science and CIRMA
Università di Torino, Turin, Italy
rossana@di.unito.it

Joao Dias
INESC-ID
IST, Taguspark, Lisbon, Portugal
joao.dias@gaips.inesc-id.pt

Abstract

In this paper, we translate a model of value-based emotional agents into an architecture for narrative characters and we validate it in a narrative scenario. The advantage of using such model is that different moral behaviors can be obtained as a consequence of the emotional appraisal of moral values, a desirable feature for digital storytelling techniques.

Introduction

In recent years, research in interactive narrative has tackled the moral aspect of stories under different perspectives, including moral conflicts and dilemmas (Ware 2011; Barber and Kudenko 2009; Battaglino and Damiano 2012). The works cited above, however, address the role of moral values in stories from the perspective of plot generation. Here, we take a complementary approach and introduce characters with moral emotions, who are able to react to the moral valence of the events occurring in a story by "feeling" the appropriate emotions.

Related Work

Many researchers tried to integrate computational models of emotions in a cognitive architecture for intelligent agents (of which (Reilly and Bates 1992; Elliott 1992; Marsella, Gratch, and Petta 2010) are some examples), with the aim of inserting emotions in BDI (Belief-Desire-Intention) agents (Bratman 1987). Moral emotions (e.g. Pride, Shame) (Haidt 2003) have received less attention: in most computational models (Marsella, Gratch, and Petta 2010) emotions are related to a desirability/undesirability of situations with respect to goals.

A few works address moral emotions related to norm violation by casting norm violation as goal violation (Gratch, Mao, and Marsella 2006) or modeling norm violation in a domain specific way, thus lacking of flexibility (Si, Marsella, and Pynadath 2010; Bulitko et al. 2008). Regarding norms, one of the few exceptions to the trend of focusing on goal-related emotions is the work by Ferreira et al. (Ferreira et al. 2013). 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 Ferreira does not address the question of how to use the moral appraisals/emotions to guide decision-making, which is a core aspect of the model we propose.

In this work, we avail ourselves of the OCC model and rely on previous work by (Battaglino, Damiano, and Lesmo 2013) to establish an explicit link between moral values and moral emotions. The agent is endowed with an explicit moral dimension formed by a scale of moral values (such as "honesty", "freedom") (Fraassen 1973).

A model for characters with moral emotions

Following (Battaglino, Damiano, and Lesmo 2013), agents have an explicit representation of their moral dimension based on a value system (Fraassen 1973), and a motivational dimension given by the desires they want to pursue (Bratman 1987). According to cognitive theories of emotions (Ortony, Clore, and Collins 1988), values and desires mediate the relation between the agent and the environment through the notion of appraisal.

We integrated the model presented in (Battaglino, Damiano, and Lesmo 2013) into the FAtiMA (Fearnot AffecTIve Mind Architecture) architecture (Dias, Mascarenhas, and Paiva 2011), developing a new component called *Value Component*. FAtiMA is an emotional agent architecture composed of a core layer (named FAtiMA Core) on which particular components can be easily added in order to add new functionalities. FAtiMA Core is a template that generally defines how an Agent Architecture with emotions works.

The Value Component generates the appraisal variables based on goals and values processing. 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, the appraisal derivation model generates a *praiseworthiness* (*blameworthiness*) variable when a value is balanced (put at stake) in the state of the world. Given the appraisal variables, the OCC Affect Derivation Component generates emotions from the appraisal variables, following

the OCC theory of emotions (Ortony, Clore, and Collins 1988) (Figure 1).

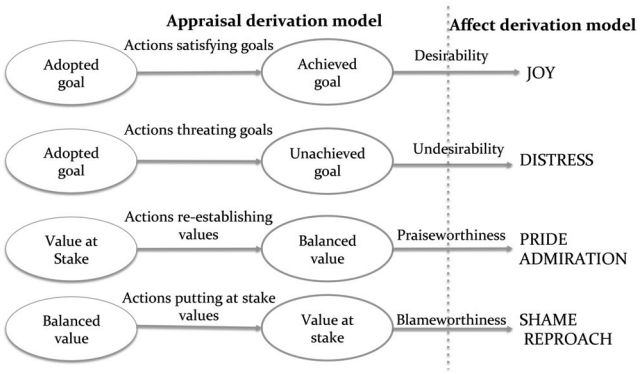

Figure 1: Appraisal and Affect Derivation model in (Battaglino, Damiano, and Lesmo 2013)

We also implement in the Value Component the "Anticipatory Emotional Appraisal" phase presented in (Battaglino, Damiano, and Lesmo 2013) in which moral emotions are integrated in the deliberation process of the agent which are driven by moral emotions (e.g. Shame, Pride, etc.) in choosing a course of action congruent with their moral values. Agents rely on moral values also to evaluate the behavior of the others, feeling emotions like Reproach or Anger. So, Boo won't steal the candy, at least unless she is very hungry. And, if she does, she will still feel Shame.
Emotional appraisal is conducted by assessing the consequences of these plans on the agent's goals and values in order to assess possible conflicts. The overall plan utility, based on emotional reward, is computed as:

$$EER = (EER_J + EER_P) - (EER_S + EER_D) \quad (1)$$

where EER_J (EER_D) is the intensity that the agent feels if the considered plan π achieves (doesn't achieve) some goals of the agent (we noted with G_A the set of individual goal satisfied by the plan π and with G_T the set of individual goals threatened):

$$EER_J(G_A, \pi_i) = \frac{P(\pi_i) * \sum_{g_a \in G_A} ImpOfS(g_a)}{E(\pi_i)} \quad (2)$$

$$EER_D(G_T, \pi_i) = \frac{P(\pi_i) * \sum_{g_t \in G_T} ImpOfF(g_t)}{E(\pi_i)} \quad (3)$$

and EER_P (EER_S) is the intensity that the agent feels if the considered plan π balances (threatens) some values of the agent (we noted with V_B the set of values re-established and with V_T the set of values put at stake):

$$EER_P(V_B, \pi_i) = \frac{P(\pi_i) * \sum_{v_b \in V_B} (r(v_b))}{E(\pi_i)} \quad (4)$$

$$EER_S(V_T, \pi_i) = \frac{P(\pi_i) * \sum_{t_t \in V_T} (r(v_t))}{E(\pi_i)} \quad (5)$$

For example, 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 Joy intensity and the Shame intensity. 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. Summarizing, the Value Component adds to the FAtiMA Core the following capabilities: 1) the monitoring of values are put at stake or re-balanced (2) the generation, in the appraisal phase, of the appraisal variables *praiseworthiness*, *blameworthiness*, *desirability* and *undesirability* based on desires and values (3) the calculation of the expected emotional reward of plans based on values re-balanced and put at stake and goals achieved or not achieved.

Example

In this section, we rely on a simple narrative scenario to illustrate how the agent architecture, presented in this paper, allows characters to generate moral emotions in response to events and take decisions based on their moral values.
Consider three characters, Mary, Boo and Charlie. Mary has a chocolate candy that she wants to give to her friend Tom (*give(chocolateCandy, Tom)*). Charlie and Boo are very hungry; they want Mary's chocolate candy (*eat(chocolateCandy)* and they can choose between different courses of action to achieve the desire: they can ask Mary for the chocolate or they can steal it. Their choice depends on the Emotional Expected Reward (EER) utility of their plans. Mary, Boo and Charlie own the value "honesty", but the priority of the value for Charlie is very low while for Mary and Boo is high (Figure 2).
In their reasoning cycle, both Charlie and Boo adopt

Agent	Goals & Values	Plan Ask		Plan Steal	
Charlie	**Eat(chocolateCandy)** impOfSuccess 8.0 impOfFailure 6.0 **Honesty** Priority 3.0	Probability :0.85 Cost: 2.0	EER_J= 3.4 EER_D= EER_P = 0 EER_S = 0 **EER = 3.4**	Probability: 1.0 Cost: 2.0	EER_J= 4 EER_D= EER_P = 0 EER_S = 2.5 **EER = 4 – 1.5 = 2.5**
Boo	**Eat(chocolateCandy)** impOfSuccess 8.0 impOfFailure 6.0 **Honesty** Priority 8.0	Probability :0.85 Cost: 2.0	EER_J= 3.4 EER_D= EER_P = 0 EER_S = 0 **EER = 3.4**	Probability: 1.0 Cost: 2.0	EER_J= 4 EER_D= EER_P = 0 EER_S = 4 **EER = 4 – 4= 0**

Figure 2: Characters' authoring data. plans probability range is in [0;1]; other elements are in [0;10]. See(Dias and Paiva 2005; Battaglino, Damiano, and Lesmo 2013) for details.

the goal *eat(chocolateCandy)* and it becomes the current intention of the agent (being the only goal, it doesn't compete with any other goals). Charlie and Boo find two plans that satisfy the intention: the π_{Steal} plan and the π_{Ask} plan. The plan π_{Steal} contains the *steal* action and the *eat* action. The *steal* action has the effect of putting at stake

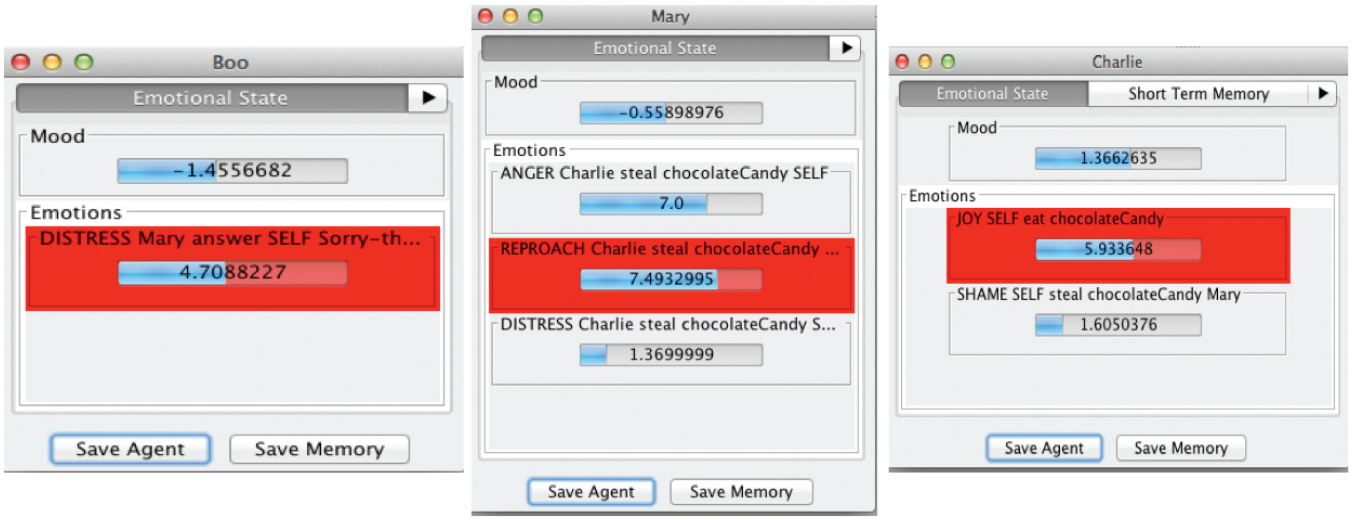

Figure 3: Characters' emotional state.

the value "honesty" and of obtaining the chocolate candy, while the *ask* action has the effect of obtaining the chocolate candy, without putting at stake any values. The *ask* action has a lower chance of realizing its effects, so the plan π_{Ask} has a less probability of success than the planπ_{Steal}. In the anticipatory emotional appraisal, Boo and Charlie calculate the Expected Emotional Reward utility for their plans. The plan π_{Steal} satisfies the goal of eating a chocolate candy and puts at stake the value "honesty". The EER depends on the Joy that Charlie and Boo will feel if they achieve the goal *eat(chocolateCandy)* and on the Shame that they will feel if they put at stake the value "honesty" (Figure 2), following the formulas presented in Section . The EER of the plan π_{Ask} depends on the Joy that they will feel if they eat the chocolate candy (Figure 2). Due to the different EER calculated during the anticipatory appraisal phase, Charlie and Boo prefer different plans:

Boo: despite the lower probability of success, Boo prefers the plans π_{Ask} because she owns the value "honesty" with an high priority (Figure 2). Boo starts to execute her plan, and asks Mary the chocolate candy. Mary answer that she wants give it to Tom, so Boo feels Distress emotion because her desire is not fulfilled (Figure 3).

Charlie: he prefers the plan π_{Steal}. Charlie owns the value "honesty" with a low priority and he prefers to execute the plan with the higher probability of success, despite the consequences of putting the value "honesty" at stake (Figure 2). After executing the plan with success, Charlie feels a Shame emotion with a very low intensity and a Joy emotion with an high intensity because his desire is fulfilled (Figure 3).

Mary: during her reasoning cycle, Mary appraises the action *Steal* performed by Charlie as blameworthy because the action puts at stake her value "honesty". At the same time, the plan executed by Charlie makes the Mary's goal *give(chocolateCandy, Tom)* unsatisfied and Mary feels a strong Anger (Reproach and Distress) emotion towards Charlie (Figure 3). The strong Anger emotion triggers Mary's action tendency to yell at Charlie.

Conclusion and Future Work

In this paper, we presented the implementation of value-sensitive emotional characters in an architecture for narrative characters (Dias, Mascarenhas, and Paiva 2011). Our work relies on previous work by (Battaglino, Damiano, and Lesmo 2013) and extends it by showing the feasibility of the model in a practical architecture, by validating it in a simple narrative scenario.

The validation shows how emotions can lead different characters, posited in the same narrative situation, to choose different courses of actions as a consequence of the compliance with different values, which lead then to experience different emotional states. An important advantage is that the characters not only feel moral emotions, such as Shame or Anger, but these emotions are involved in the characters' deliberation process. The implication for narrative technologies is a gain in terms of characters' believability, with the possibility of modeling not only the visible effects of the emotional states but also the introspective process occurring in the characters' mind, that is an intrinsic feature of literaly narration. Moreover, the implemented model incorporated domain-independent rules for emotion generation and action selection, a desirable feature for the authoring of new characters.

In the future we plan to test our work in more complex narrative scenarios, taken from real literaly works and validate the predicted emotional states and behaviors with human users.

References

Barber, H., and Kudenko, D. 2009. Generation of adaptive dilemma-based interactive narratives. *Computational Intel-*

ligence and AI in Games, IEEE Transactions on 1(4):309–326.

Battaglino, C., and Damiano, R. 2012. Emotional appraisal of moral dilemma in characters. In *In Proceedings of the Fifth International Conference on Interactive Digital Storytelling*, 150–161.

Battaglino, C.; Damiano, R.; and Lesmo, L. 2013. Emotional range in value-sensitive deliberation. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems. International Foundation for Autonomous Agents and Multiagent Systems, 2013.*, 769–776.

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

Bulitko, V.; Solomon, S.; Gratch, J.; and van Lent, M. 2008. Modeling culturally and emotionally affected behavior. In *In Proceedings of the Fourth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Dias, J., and Paiva, A. 2005. Feeling and reasoning: A computational model for emotional characters. In Bento, C.; Cardoso, A.; and Dias, G., eds., *EPIA*, volume 3808 of *Lecture Notes in Computer Science*, 127–140. Springer.

Dias, J.; Mascarenhas, S.; and Paiva, A. 2011. Fatima modular: Towards an agent architecture with a generic appraisal framework. *Proceedings of the International Workshop on Standards for Emotion Modeling*.

Elliott, C. D. 1992. *The affective reasoner: a process model of emotions in a multi-agent system*. Ph.D. Dissertation, Northwestern University, Evanston, IL, USA.

Ferreira, N.; Mascarenhas, S.; Paiva, A.; Tosto, G. D.; Dignum, F.; Breen, J. M.; Degens, N.; Hofstede, G.; Andrighetto, G.; and Conte, R. 2013. An agent model for the appraisal of normative events based in in-group and outgroup relations. AAAI-13 Twenty-Seventh AAAI Conference on Artificial Intelligence.

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

Gratch, J.; Mao, W.; and Marsella, S. 2006. *Modeling social emotions and social attributions*. Cambridge University Press.

Haidt, J. 2003. The moral emotions. *Handbook of affective sciences* 852–870.

Marsella, S. C.; Gratch, J.; and Petta, P. 2010. Computational models of emotion. In Scherer, K. R.; BÃnziger, T.; and Roesch., eds., *A blueprint for an affectively competent agent: Cross-fertilization between Emotion Psychology, Affective Neuroscience, and Affective Computing*. Oxford: Oxford University Press.

Ortony, A.; Clore, G.; and Collins, A. 1988. *The Cognitive Structure of Emotions*. Cambrigde University Press.

Reilly, W. S., and Bates, J. 1992. Building emotional agents.

Si, M.; Marsella, S. C.; and Pynadath, D. V. 2010. Modeling appraisal in theory of mind reasoning. *Autonomous Agents and Multi-Agent Systems* 20(1):14–31.

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