

Narrative Scenarios as a Testbed for Moral Agents

Cristina Battaglino¹ and Rossana Damiano²

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. Moral emotions play an important role in the design and implementation of virtual agents, since they have implications not only for the agent's deliberation process but also for its interactional and social functions. In this paper, we propose an empirical methodology that leverages narrative situations to evaluate the emotional state felt by an intelligent agent.

1 Introduction

Moral emotions [9,13,8] deserve special attention in the design of virtual agents, since they play a crucial role in many agent's functions, ranging from self regulation [4] to the handling of social and interactional aspects [5]. Moral emotions have been integrated into a number of computational models of emotions, with approaches that range from the hard coding of moral standards in the emotional component [12,10], to the design of high level moral values to drive the generation of emotions [3]. In this paper, we propose a methodology for evaluating the generation of moral emotions that leverages the capability of narratives of exemplifying the connection between moral values and emotions.

2 Values and Emotions

In order to test the feasibility of the proposed methodology, in this paper we apply it to the emotional agent model described by [3], where a BDI agent is extended with emotions and moral values. In [3], agents have an explicit representation of their moral dimension based on a value system [6], and a motivational dimension given by the desires they want to pursue [2]. According to [1], the emotional state of the agent is the result of the appraisal and affect generation processes. The appraisal process is based on goal and value processing and outputs a set of appraisal variables that are the input to the affect generation process. In particular, the appraisal process generates a *desirability* or (*undesirability*) variable when an agent's goal is achieved (or unachieved) in the state of the world as an effect of some action or event; it generates a *probability* variable depending on the probability that an agent's plan succeeds; finally, it generates a *praiseworthiness* (or *blameworthiness*) variable when an agent's value is balanced (or put at stake) by the execution of some action.

3 Experimental Protocol

The experiments we conducted to evaluate the moral emotional agent relied on narrative scenarios that we presented to participants in a text-based form through a web site¹. Emotions were described through text labels (e.g. “Hamlet feels *Reproach*, *Shame* and *Anger* towards Ophelia), without any colors. The scenarios were selected from famous works of fiction with the help of a drama expert, who identified them based on availability of well established critical interpretations of the emotions felt by the characters [11,7]. The selected scenes are taken from: *Hamlet’s* (Shakespeare), *The count of Monte Cristo* and *The Vicomte of Bragelonne: Ten years later* by Alexandre Dumas, *Thérèse Raquin* (Émile Zola). The methodology we followed to create the narrative scenarios is the following: first, we analyzed the structure of the narrative, then, with the help of the expert, we identified the cognitive states of the characters and we characterized their mental attitudes (i.e. beliefs, desires and values). By doing so, we also verified that the agent model under evaluation had the necessary expressiveness to model the selected narrative situations.

A convenience sample of twenty subjects, 9 female and 11 male, aged 24-35, participated in the study. In the web experiment, we assigned participants randomly to the test conditions (V+) and (V-). The (V+) group evaluated scenarios in which emotions are associated to characters following the model in [3], in which moral values and moral emotions are presented. The (V-) group evaluated scenarios with emotions related to goals, such as Joy and Distress. The primary hypotheses of the study are:

- H1: emotions in the V- condition are less believable than V+ condition (*believability*);
- H2: emotions in the V- condition are less complete than V+ condition (*completeness*);

For example, in the scenario of Hamlet, Hamlet feels Anger towards Ophelia because she lied, putting at stake Hamlet’s value honesty (*blameworthy action*) and threatening Hamlet’s goal of saving her from the corruption of the court (*undesirable event*). In the V+ group, we associated Anger, Reproach and Distress emotions to Hamlet because the V+ condition contemplates values. In the V- condition, we associate only Distress emotion because the V- condition doesn’t provide values but only goals, so emotions related to values are not taken into account.

Given a scenario, participants first read the scenario and then he/she received the post questionnaire about the scenario. The believability and completeness measures of the affective states of characters involved in the scenario were assessed by direct questions in the post questionnaire, with a 5-item Likert scale response. We also asked to the participants to list the emotions they would feel if they were to identify with the story characters, using multiple-choice questions based on OCC categories of emotions [1].

¹ The web experiment is online at www.ilnomedellarosa.it/Valutazione - in Italian only.

4 Evaluation

In order to determine if there are differences in the variables under study (believability and completeness of the set of characters' emotions, H1 and H2 respectively) between the groups V+ and V-, we conducted the Mann-Whitney U test on the completeness and believability Likert scores. We run a series of tests considering all the scenarios, then we run the same tests on every scenario. In (Table 2, Table 1) we summarize the results of the Mann-Whitney U tests on the Likert scores.

Table 1. Mann Whitney U statistics for *complete* metric

Results	U	p (two tailed)	p (one tailed)
Hamlet	19	.01	.005
Count of Monte Cristo	13	.003	.001
Thérèse Raquin	33	.185	.09
Vicomte Bragelonne	6	.001	.0
All	277	.0	.0

Table 2. Mann Whitney U statistics for *believable* metric

Results	U	p (two tailed)	p (one tailed)
Hamlet	6	.0	.0
Count of Monte Cristo	19	.012	.006
Thérèse Raquin	32	.133	.133
Vicomte Bragelonne	45	.6	.3
All	453	.0	.0

Quantitative Results. Considering all scenarios, results show that the perceived level of believability and completeness of emotions in the V- condition were significantly lower than the V+ condition at $p < 0.01$. Regarding the specific scenarios, in the *Hamlet* and *Count of Monte Cristo* scenarios, the perceived level of believability of emotions in the V- condition was significantly lower than the scores in the V+ condition, while, in the *Hamlet*, *Count of Monte Cristo* and *Vicomte the Bragelonne*, the perceived level of completeness of emotions in the V- condition was significantly lower than the scores in the V+ condition (with significance at $p < 0.05$ in general). Results don't show statistical significance for the third scenario.

Discussion. The analysis of the results confirms the hypothesis H2 but not H1. However, the emotions associated with the V- condition were a subset of emotions in the V+ condition, so the lack of some of them may have been not

perceived. We also observed that participants attributed to the characters the emotions predicted by the moral emotional agent model. The third scenario was not evaluated as complete, possibly because the characters' values were perceived by the participants.

5 Conclusion

In this paper, we proposed a methodology to evaluate the generation of moral emotions that relies on narrative scenarios, considered as paradigmatic cases of the interplay between emotions and moral values. By modeling the characters in each scenario as value-based emotional agents [3], we asked a set of testers to assess the adequacy of the characters' emotions to the given scenarios. The results suggest that the participants perceive as more believable the emotional states that encompass moral values and emotions, and that they agree with the generated emotions. This suggests that the users perceive the relevance of moral values – and of moral emotions – in narrative situations. As future work, we plan to design and run further tests to verify this hypothesis.

References

1. Collins, A., Ortony, A., Clore, G.L.: The cognitive structure of emotions. Cambridge University Press, Cambridge (1988)
2. Bratman, M.E.: Intention, plans, and practical reason. Harvard University Press, Cambridge (1987)
3. Lesmo, L., Battaglini, C., Damiano, R.: Emotional range in value-sensitive deliberation. In: AAMAS, pp. 769–776 (2013)
4. de Melo, C.M., Zheng, L., Gratch, J.: Expression of moral emotions in cooperating agents. In: Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjálmsson, H.H. (eds.) IVA 2009. LNCS, vol. 5773, pp. 301–307. Springer, Heidelberg (2009)
5. de Melo, C.M., Carnevale, P., Gratch, J.: The influence of emotions in embodied agents on human decision-making. In: Safonova, A. (ed.) IVA 2010. LNCS (LNAI), vol. 6356, pp. 357–370. Springer, Heidelberg (2010)
6. Van Fraassen, B.C.: Values and the heart's command. *Journal of Philosophy* 70(1), 5–19 (1973)
7. Freytag, G.: Technique of the drama, an exposition of dramatic composition and art. S.C. Griggs and Company, Chicago (1985)
8. Garcia, A.E., Ostrosky-Solís, F.: From morality to moral emotions. *International Journal of Psychology* 41(5), 348–354 (2006)
9. Haidt, J.: The moral emotions. In: *Handbook of Affective Sciences*, pp. 852–870 (2003)
10. Marsella, S., Gratch, J.: Ema: A process model of appraisal dynamics. *Cognitive Systems Research* 10(1), 70–90 (2009)
11. Polti, G.: Les trente-six situations dramatiques. Mercure de France, Paris (1895)
12. Reilly, W.S., Bates, J.: Building emotional agents (1992)
13. Stark, S.: Emotions and the ontology of moral value. *The Journal of Value Inquiry* 38, 355–374 (2004), doi:10.1007/s10790-005-1341-y