Other-Condemning Anger = Blaming Accountable Agents for Unattainable Desires

(Extended Abstract)

Mehdi Dastani
Utrecht University
Utrecht, The Netherlands
m.m.dastani@uu.nl

Emiliano Lorini
IRIT-CNRS
Toulouse, France
emiliano.lorini@irit.fr

John-Jules Meyer
Universiteit Utrecht
Utrecht, The Netherlands
j.j.c.meyer@uu.nl

Alexander Pankov
Universiteit Utrecht
Utrecht, The Netherlands
a.pankov@students.uu.nl

ABSTRACT

This paper focuses on the other-condemning anger emotion which is a social type of anger triggered by the behaviour of other agents. Other-condemning anger responds to frustration of committed goals by others, and motivates goal-congruent behavior towards the blameworthy agents. Understanding this type of anger is crucial for modelling human behavior in social settings as well as designing socially aware artificial systems. We summarize some existing psychological theories on other-condemning anger and advocate building logical frameworks to formally specify this emotion. We believe that a formalization should provide a precise conceptualization and characterization of other-condemning anger in terms of social and cognitive concepts such as beliefs, goals, intentions, controllability, accountability, and blameworthiness.

1. INTRODUCTION

Other-condemning anger is a reaction to the frustration of goals to which agents are committed, and motivates goal-congruent behavior towards the agents believed to be accountable for the goal frustration [3, 10, 7, 5]. Imagine a situation where autonomous robots commit themselves to transport containers from one place to another in some physical environment such as harbours or stockrooms. A robot R_1 that aims at picking up its container at a designated position may notice the container is removed by another robot R_2. A desirable response of the robot R_1 would be to send a request to the robot R_2, who is believed by R_1 to be accountable for the removal of the container, to make the container accessible to R_1 and/or to send a warning message to the manager of the environment to report this irregularity. We would like to emphasize that it is the general function of anger, i.e., specific type of response to specific type of situation, that we aim at integrating in the model of autonomous agents, rather than the physiological aspects of anger that is characteristic to the human body. For autonomous software agents that interact in social settings, the other-condemning anger emotion can be considered as a behavioural pattern or a heuristic that steers their behaviours.

Appears in: *Proc. of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017), S. Das, E. Durfee, K. Larson, M. Winikoff (eds.), May 8–12, 2017, São Paulo, Brazil.*

Although there have been many efforts in artificial intelligence to provide a precise specification of emotions in general [2, 8, 9, 14], there has not been, to our knowledge, a precise and adequate specification dedicated to the other-condemning anger emotion based on complex social constructs such as controllability, accountability and blameworthiness. These social concepts require an adequate formalization of notions such as actions, control, causality, and their relations with the agents' cognitive states. As the above robot example illustrates, the angry robot R_1 believes that its transportation goal is frustrated and that this is due to the removal action of robot R_2 who had control over its removal action (in the sense that R_2 could have chosen not to remove the container) and who is accountable for the caused consequences (R_1 cannot accomplish its transportation goal). The overtly social nature (being concerned with other agents) of this type of anger emotion and its potential to influence others' behavior, make them essential for modelling human-like social interaction and designing socially aware artificial systems, which can be used for example in entertainment and serious games, crowd simulations, and human-computer interaction.

2. MODELLING OTHER-CONDEMNING ANGER

In order to model other-condemning emotions, we propose to use a logical framework of multi-agent systems in which agents are specified by means of their knowledge, beliefs, desires, intentions, and actions. More specifically, we propose to use a tractable multi-agent extension of the DL-GA logic (dynamic logic of graded mental attitudes) developed by Dastani & Lorini in [2]. This model allows us to formally specify the appraisal and coping processes involved in other-condemning anger, thus providing a precise conceptualization of the other-condemning anger emotion and its logic. The logic supports reasoning about agents' knowledge, graded beliefs, graded desires and intentions as well as agents' future and past actions. One of its feature is that actions are modeled as propositional assignments whose effect is to toggle the truth values of atomic propositions.

We distinguish two types of anger emotions. The first type of anger, called *plain anger*, involves two agents and captures the setting where an agent's committed goals are frustrated by another agent. The second type of anger, called *social anger*, involves three agents and captures the situation where the first agent gets

angry at the second agent because the second agent harms a third agent who is in some social relation with the first agent. For social anger, we assume some social rules the existence of which are due to (or depend on) some norms or organisation that governing the multi-agent environment. These assumed social rules may relate the goals of the first and the third agents such that the frustration of the third agent's goals by the second agent indirectly frustrates the goals of the first agent. For example, consider an extension of the robot example with a new manager agent that is responsible for the distribution and accomplishment of the transportation goals of all transport robots, including robot R_1. In this setting, the manager agent and robot R_1 are in an organisational setting where the achievement of the transportation goals of R_1 may contribute to the achievement of the goals of the manager agent. If R_2 frustrates the goal of R_1, then R_2 will indirectly and through the existence of the social rule frustrate the goal of the manager agent and therefore make this agent angry.

The theoretic and empirical support for our modelling proposal is derived from cognitive and social psychology, in particular the emotion theories [3, 10, 7, 12, 13, 5]. Other-condemning anger is commonly viewed as a negatively valenced reaction to the actions of other agents [10]. It is an instance of the *other-condemning* emotions [5], and triggered by frustration of a goal commitment [10, 7]. In our robot example, the goal that the transport robot is committed to, i.e., the goal to have the container at its designated position, is frustrated. This broad view of other-condemning anger has been refined by emotion theories to distinguish it from other negative emotions such as sadness, guilt and remorse that also can arise from *goal incongruence*.

Most emotion theories distinguish other-condemning anger from other negative emotions by attributing *blame* for goal incongruence to other agents [7, 3]. As a result, blame towards someone else becomes a necessary condition for other-condemning anger, for without the attribution of blame we can expect an emotion such as sadness. What does it mean, however, to blame someone for goal incongruence? According to [7], blame is an appraisal based on *accountability* and imputed *control*. To attribute accountability is to know who caused the relevant goal-frustrating event, and to attribute control is to believe that the accountable agent could have acted differently without causing the goal-incongruence. In our example, robot R_1 believes that robot R_2 is accountable for removing the container and that R_2 has the choice not to remove the container. According to Lazarus, anger is triggered if, in addition to above conditions, the *coping potential* (the evaluation of the possible responses) is viable. The prototypical *coping strategy* of other-condemning anger generally involves attack, or other means of getting back at the blameworthy agent, with the intention of restoring a goal-congruent state of affairs [6, 3, 7]. In our running example, the robot R_1 can send a request to R_2 to make the container accessible to R_1 and/or to report this irregularity to the environment manager.

The second type of anger, i.e., social anger, is similar to what is often called *moral anger*, where a first agent is morally angry at a second agent because the second agent harms a third agent by violating some moral norm (for a review on the literature from social psychology see [16]; for a more philosophical treatment see [11]). In such cases, an agent can rightfully be angry without any of his own goals being directly frustrated. In our extended example, the manager agent, which may be a software agent as well, may get angry at robot R_2, because R_2 has frustrated the goal of R_1. The actual reason for an agent to get angry at the third agent is the existence of a social rule that prescribes and promotes cooperation. For example, in case of human agents the reason for being angry can be the violation of a moral rule that prescribes agents not to harm the autonomy of each other. As argued in [12], autonomy is seen as a right (i.e., a moral norm) pertaining to harm against persons. The typical coping strategy for social anger is similar to the coping strategy for the plain anger and promotes socially-congruent behavior. Combining this aspect of social anger with the elicitation conditions of plain anger allows us to informally describe other-condemning anger in psychological terms as follows: Displeasure from thwarting of a personal goal, or a social rule aimed at preserving the goal commitment of other agents, combined with attribution of blame for the goal-thwarting state of affairs to another agent, and an estimate of one's own coping potential as favouring attack towards the blameworthy agent.

3. CONCLUDING REMARKS

Although the focus of our work is other-condemning anger, we believe that a logical framework could be used to model various other-condemning social emotions such as disgust and contempt. The characteristic features of the other-condemning emotions are its multi-agent flavor and the inclusion of emotion intensity. Although the importance of emotion intensity has been stressed by appraisal theorist, most of the formal models in the literature have ignored at least one of them. For example, [1, 9, 15] ignores emotion intensity and [2] does not have multi-agent flavor. Although our proposal is inspired by [2], we believe the model should be modified significantly in order to accommodate the characteristic feature of other-condemning emotions. In particular, we believe other-condemning and socially oriented anger requires extending the single agent framework proposed in [2] to a multi-agent framework. Moreover, the framework needs to be extended with the converse of physical actions to reason about the state of the world before the execution of an action. Such feature is of crucial importance to some components of anger, e.g., responsibility and blame.

Another influencing work on the topic has been the work of Steunebrink, Dastani and Meyer [14]. Unlike our proposal, [14] takes emotion intensity as primitive, without explaining how it depends on belief and goal strengths. Furthermore, we believe that the developed logical framework should be rigorously specified and analyzed. The work presented in [14] does not provide any decidability results or axiomatization, which is required to investigate the decidability of the logical framework. Finally, [4] propose a formal model of emotions which incorporates both emotion intensities and coping. However, the authors do not provide any details on the underlying logic, which makes comparing the two approaches difficult.

The contribution of our proposal should be viewed as twofold. First, it advocates a precise conceptualization and characterization of other-condemning social anger and its integration in a cognitive model of agency. Second, it advocates formal understanding of human-like social behavior which should pave the way towards designing socially aware software systems. We intend to extend the set of other-condemning emotions in future work and provide an analysis on the relation between various social and moral emotions. We believe that the dynamic nature of any logic of other-condemning emotions should be powerful enough to allow complex actions such that the accountability notion can involve actions that have been performed in some state in the past.

REFERENCES

[1] Carole Adam, Andreas Herzig, and Dominique Longin. A logical formalization of the occ theory of emotions. *Synthese*, 168:201–248, 2009.

[2] M. Dastani and E. Lorini. A logic of emotions: from appraisal to coping. *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pages 1133–1140, 2012.

[3] N. H. Frijda. *The emotions*. Cambridge Univ Pr, 1986.

[4] J. Gratch and S. Marsella. A domain-independent framework for modeling emotion. *Cognitive Systems Research*, pages 269–306, 2004.

[5] Jonathan Haidt. The moral emotions. *Handbook of affective sciences*, 11:852–870, 2003.

[6] E Izard Carroll. Human emotions. *New York7 Plenum*, 1977.

[7] R. S. Lazarus. *Emotion and adaptation*. Oxford University Press, USA, 1991.

[8] E. Lorini. A dynamic logic of knowledge, graded beliefs and graded goals and its application to emotion modelling. *Logic, Rationality, and Interaction*, pages 165–178, 2011.

[9] E. Lorini and F. Schwarzentruber. A logic for reasoning about counterfactual emotions. *Artificial Intelligence*, 2010.

[10] A. Ortony, G. L. Clore, and A. Collins. *The cognitive structure of emotions*. Cambridge Univ Pr, 1990.

[11] Jesse Prinz. *The emotional construction of morals*. Oxford University Press, 2007.

[12] Paul Rozin, Laura Lowery, Sumio Imada, Jonathan Haidt, et al. The cad triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of personality and social psychology*, 76:574–586, 1999.

[13] K. R. Scherer. Appraisal considered as a process of multilevel sequential checking: A component process approach. *Appraisal processes in emotion: Theory, methods, research*, 92:120, 2001.

[14] B. Steunebrink, M. Dastani, and J. J. Meyer. A formal model of emotion-based action tendency for intelligent agents. *Progress in Artificial Intelligence*, pages 174–186, 2009.

[15] P. Turrini, J. J. C. Meyer, and C. Castelfranchi. Coping with shame and sense of guilt: a dynamic logic account. *Autonomous Agents and Multi-Agent Systems*, 20(3):401–420, 2010.

[16] E. Alicia Vélez García and Feggy Ostrosky Solís. From morality to moral emotions. *International Journal of Psychology*, 41:348–354, 2006.