Shaping Cooperation between Humans and Agents with Emotion Expressions and Framing

Socially Interactive Agents Track

Celso M. de Melo, Peter Khooshabeh
US Army Research Laboratory, Playa Vista, CA 90094, United States
{celso.m.demelo.civ, peter.khooshabehadeh2.civ}@mail.mil

Ori Amir
UCSB Institute for Collaborative Biotchnologies, Santa Barbara, CA 93106-5100, United States
oriacadem@gmail.com

Jonathan Gratch
USC Institute for Creative Technologies, Playa Vista, CA 90094-2536, United States
gratch@usc.ict.edu

ABSTRACT

Emotion expressions can help solve social dilemmas where individual interest is pitted against the collective interest. Building on research that shows that emotions communicate intentions to others, we reinforce that people can infer whether emotionally expressive computer agents intend to cooperate or compete. We further show important distinctions between computer agents that are perceived to be driven by humans (i.e., avatars) vs. by algorithms (i.e., agents). Our results reveal that, when the emotion expression reflects an intention to cooperate, participants will cooperate more with avatars than with agents; however, when the emotion reflects an intention to compete, participants cooperate just as little with avatars as with agents. Finally, we present first evidence that the way the dilemma is described – or framed – can influence people's decision-making. We discuss implications for the design of autonomous agents that foster cooperation with humans, beyond what game theory predicts in social dilemmas.

CCS CONCEPTS

• **Computer methodologies** → **Artificial Intelligence**; • **Human-centered computing** → **Human computer interaction (HCI)**

KEYWORDS

Emotion Expression, Framing, Social Dilemmas

ACM Reference format:

C. de Melo, P. Khooshabeh, O. Amir, and J. Gratch. 2018. Shaping Cooperation between Humans and Agents with Emotion Expressions and Framing. In *Proc. of the 17th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2018), July 10-15, 2018, Stockholm, Sweden,* IFAAMAS, ACM, New York, NY, 3 pages.

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10-15, 2018, Stockholm, Sweden.

1 INTRODUCTION

Following increased interest in the interpersonal effects of emotion on people's decision making [1]-[4], there has been a growing body of research showing that emotion expressed by computer agents influences the decisions people make [4]-[8]. In a social dilemma, de Melo and colleagues [4]-[7] showed that emotion expressions could increase or decrease cooperation according to the type and context emotions were shown. Complementary, de Melo, Gratch, and Carnevale [7] showed that the strength of the effect tended to be higher when people believed the emotion expressed by a computer was being controlled by another person (i.e., an avatar) vs. a computer algorithm (i.e., an agent). However, despite these advances, much remains to be understood about the effect of emotion and here we extend this earlier literature with a study that compares people's behavior with cooperative, neutral, and competitive computer counterparts that are believed to be either agents or avatars.

Additionally, one area that has not received much attention, at least in the context of the study of emotion, is task framing. Research in the behavioral sciences demonstrates that the way a task is framed influences people's decisions [9]-[10]. In particular, Pruitt [10], [10] showed that the way the payoffs in the prisoner's dilemma were framed impacted cooperation (Fig. 1). When the game was decomposed, participants cooperated less when the game emphasized exploitation (Game II, Fig. 1-B) than cooperation (Game IV, Fig. 1-C), even though the payoffs are exactly the same as in the regular game (Fig. 1-A). Here we look at the role of framing on human-agent cooperation and, additionally, study whether it interacts with the effect of emotion.

A) Regular PD

		Other: Cooperate	Other: Defect
You	Cooperate	You: 12 Other: 12	You: 0 Other: 18
	Defect	You: 18 Other: 0	You: 6 Other: 6

B) Game II

		Other: Your Payoff	Other: Other Payoff
You	Cooperate	6	6
	Defect	12	-6

C) Game IV

		Other: Your Payoff	Other: Other Payoff
You	Cooperate	-6	18
	Defect	0	6

Figure 1: Different frames for the prisoner's dilemma.

EXPERIMENT

The experiment followed a 3 × 2 × 3 mixed factorial design: *emotion expression* (cooperative vs. neutral vs. competitive; within-participants) × *counterpart* (avatar vs. agent; within-participants) × *frame* (regular PD vs. Game II vs. Game IV; between-participants). The payoff matrices we used are shown in Fig. 1. Participants were informed that the points they earned would be exchanged for tickets to a $30 lottery. The experiment was fully anonymous, i.e., neither the counterparts, nor the experimenters could trace their decisions back to them.

The emotion patterns we used were the same as the ones used by de Melo et al. [4]. The cooperative pattern consisted of showing joy after mutual cooperation and regret following participant exploitation. The competitive pattern consisted of showing joy following exploitation and regret following mutual cooperation.

Participants engaged with 18 different counterparts (Fig. 3-A): three avatars and three agents for each of the three types of emotion expression patterns (cooperative, neutral, and competitive). With each counterpart, three rounds were played. In the first two rounds, the software made the decision for the participant to guarantee that the participant would experience the critical emotion expressions. Removing these decisions from the participants was justified as "being necessary to guarantee that the experiment would be completed in a timely manner". The critical decision, thus, was made in the third round. This procedure is summarized in Fig. 3-B. Finally, the timing for each round is shown in Fig. 3-C. We recruited 281 participants from Amazon Mechanical Turk for this experiment.

3 RESULTS & DISCUSSION

The results reinforced that emotion expressed by agents shaped cooperation with humans, $F(2, 556) = 17.17$, $p < .001$, partial $\eta^2 = .058$: people cooperated more with agents that showed cooperative rather than neutral emotions; and, in turn,

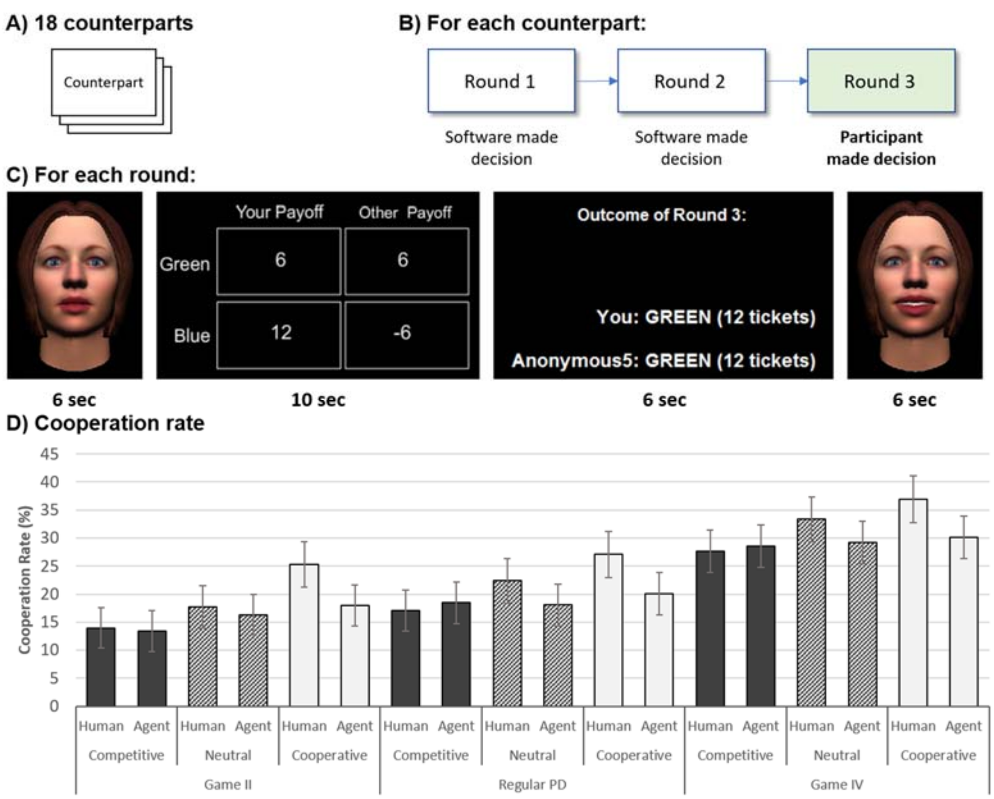

Figure 2: Experimental procedure and results. The bars show standard errors.

more with agents that showed neutral rather than competitive emotions.

Secondly, the results further reinforced the existence of a bias that led people to cooperate more with avatars than with agents, $F(1, 556) = 15.22$, $p < .001$, partial $\eta^2 = .052$. However, they also revealed that this bias is moderated by the emotion expressions, $F(2, 556) = 6.11$, $p = .002$, partial $\eta^2 = .022$: participants cooperated more with avatars than agents for cooperative or neutral displays; but, for competitive displays, there was no difference in cooperation rate.

Thirdly, the results showed that the way the task is framed can influence cooperation, $F(2, 278) = 3.88$, $p = .022$, partial $\eta^2 = .027$: people cooperated more in Game IV than the regular PD, and more in the regular PD than Game II. However, this effect was independent to the effect of emotion expressions.

Finally, the paper makes a methodological contribution consisting of a simpler non-contingent experimental procedure for expression of emotion that requires fewer rounds than in earlier work. The main advantage is that it, thus, requires less time to replicate the effect of emotion in practical applications.

ACKNOWLEDGMENTS

This research was supported by mission funding to the U.S. Army Research Laboratory. The content does not necessarily reflect the position or the policy of any Government, and no official endorsement should be inferred.

REFERENCES

[1] N. Frijda and B. Mesquita. 1994. The social roles and functions of emotions. In S. Kitayama&H. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 51–87). Washington, DC: American Psychological Association.

[2] D. Keltner and J. Haidt. 1999. Social functions of emotions at four levels of analysis. *Cogn. Emot.*, 13, 505–521.

[3] G. Van Kleef, C. De Dreu and A. Manstead. 2010. An interpersonal approach to emotion in social decision making: The emotions as social information model. *Adv. Exp. Soc. Psychol.*, 42, 45–96.

[4] C. de Melo, P. Carnevale, S. Read and J. Gratch. 2014. Reading people's minds from emotion expressions in interdependent decision making. *J. Pers. Soc. Psychol.*, 106, 73-88.

[5] C. de Melo, L. Zheng and J. Gratch. 2009. Expression of moral emotions in cooperating agents. In *Proceedings of Intelligent Virtual Agents (IVA'09)*.

[6] C. de Melo, P. Carnevale and J. Gratch. 2012. The impact of emotion displays in embodied agents on emergence of cooperation with people. *Presence*, 20, 449-465.

[7] C. de Melo, J. Gratch and P. Carnevale. 2014. Humans vs. Computers: Impact of emotion expressions on people's decision making. *IEEE T. Affect. Comput.*, 6, 127-136.

[8] C. de Melo, P. Carnevale and J. Gratch. 2011. The effect of expression of anger and happiness in computer agents on negotiations with humans. In *Proceedings of Autonomous Agents and Multiagent Systems (AAMAS'11)*.

[9] A. Tversky and D. Kahneman. 1981. The framing of decisions and the psychology of choice. *Science*, 211, 453-458.

[10] D. Pruitt. 1967. Reward structure and cooperation: The decomposed prisoner's dilemma game. *J. Pers. Soc. Psychol.*, 7, 21-27.

[11] D. Pruitt. 1970. Motivational processes in the decomposed prisoner's dilemma game. *J. Pers. Soc. Psychol.*, 14, 227-238.