Dr Celso de Melo

E: celso.miguel.de.melo@gmail.com
T: +1 310 448 5391

Dr Kazunori Terada

E: terada@gifu-u.ac.jp
T: +81 58 293 2792

Cooperation with autonomous machines through culture and emotion

Research Objectives

Celso de Melo and Kazunori Terada show how simple emotional cues such as machines showing positive or negative emotion help mitigate bias unfavourable to machines and promote human-machine cooperation.

Detail

Celso M. de Melo
USC Institute for Creative Technologies
12015 Waterfront Drive
Playa Vista, CA 90094
USA

Kazunori Terada
Gifu University
1-1 Yanagido
Gifu 501-1193
Japan

Bio

Celso M. de Melo is a computer scientist with an interest in artificial intelligence, human-machine interaction, and affective computing. He earned his PhD at the University of Southern California and bachelor's degree at IST-Lisbon. His research focuses on theories of human-machine interaction and designing socially intelligent machines.

Kazunori Terada is an Associate Professor of Informatics at Gifu University. He received his PhD and MEng from Nara Institute of Science and Technology and BEng from Osaka University. His research interests include artificial intelligence, cognitive science, and human-agent interaction.

Funding

JSPS KAKENHI Grant Number JP16KK0004

Acknowledgements

Jonathan Gratch

References

- De Melo, C.M., and Terada, K. (2019). Cooperation with autonomous machines through culture and emotion. *PLoS ONE*, [online] 14(11). Available at: https://doi.org/10.1371/journal.pone.0224758
- Terada, K., and Yamada, S. (2017). Emotional Expression in Simple Line Drawings of a Robot's Face Leads to Higher Offers in the Ultimatum Game. *Frontiers in Psychology*, [online] 8, 724. Available at: https://doi.org/10.3389/fpsyg.2017.00724
- Terada, K., and Yamada, S. (2017). Mind-reading and Behavior-reading against Agents with and without Anthropomorphic Features in a Competitive Situation. *Frontiers in Psychology*, [online] 8, 1071. Available at: https://doi.org/10.3389/fpsyg.2017.01071
- De Melo, C.M., Marsella, S., and Gratch, J. (2016). People do not feel guilty about exploiting machines. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 23(2), 1–17.
- De Melo, C.M., Carnevale, P.J., Read, J., and Gratch, J. (2014). Reading People's Minds From Emotion Expressions in Interdependent Decision Making. *Journal of Personality and Social Psychology*, 106(1), 73–88.

Personal Response

What sparked your interest in human-machine cooperation and how do you see this research progressing?

"It was the realisation that the success of AI hinges on peoples' ability to trust and cooperate with AI. We are developing technology that has the potential to considerably improve human life, but we will only be able to reap those benefits if we find ways to promote collaboration between humans and machines. As this technology becomes more pervasive in society, we need to continue studying the mechanisms driving human decision making and understanding how those influence the design of autonomous machines."

Cooperation with autonomous machines through culture and emotion

People tend to be less cooperative with machines than with humans. Dr Celso de Melo, a computer scientist with the US Army Research Laboratory, and Dr Kazunori Terada, an Associate Professor at Gifu University, Japan, demonstrate how incorporating simple cultural and emotional cues, such as virtual faces showing positive or negative emotion, can help mitigate unfavourable bias toward machines and promote human-machine cooperation.

Autonomous machines that can act on our behalf are rapidly becoming fundamental to our society. Robots, drones, and self-driving vehicles are all becoming a reality with the potential to mould our existence. These machines can profoundly change how we interact with each other, so it is essential that we understand if they will influence any significant changes to human decision making.

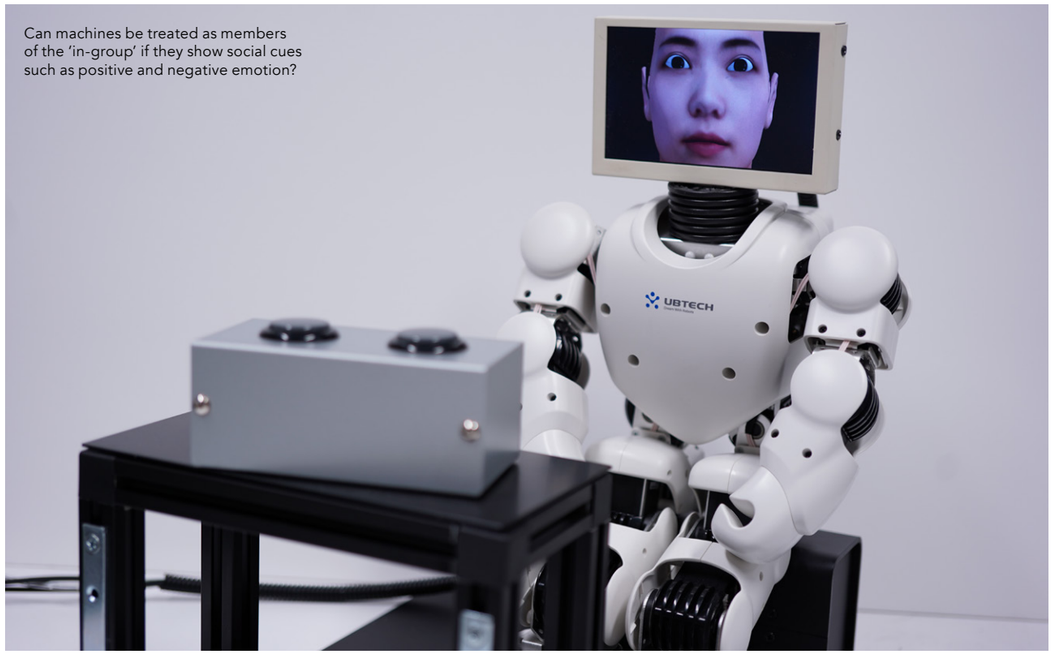
Research carried out by Dr Celso de Melo, a computer scientist with the US Army Research Laboratory, and Dr Kazunori Terada, Associate Professor at Gifu University, Japan, shows that people have a tendency to make less favourable decisions and be less cooperative with machines than with humans. Results reveal that people engage in social categorisation that distinguishes people, or 'us', from machines, or 'them'. This leads to an unfavourable bias against 'them', implying that machines are perceived as out-group members.

IN-GROUPS AND OUT-GROUPS

People tend to categorise others during social interactions. They are inclined to associate more with some people, self-identifying with the in-group, and less so with others, the out-group. This can result in a bias favouring cooperation with members of the in-group; thus, promoting the in-group's affluence and increasing the individual's likelihood of survival and receipt of long-term benefits. Such perception of group membership is often used to encourage cooperation in situations involving social dilemmas. It has been shown that people categorise machines in a similar way, in line with gender and cultural stereotypes, favouring computers with a virtual face typical of their race and voices with accents similar to their own. This would imply that as well as people applying social categorisation to machines, machines can also be members of the in-group.

Studies show that when engaging with machines, people make different

Machines used in de Melo and Terada's experiment were able to express emotion using virtual faces from either US or Japanese culture.



Can machines be treated as members of the 'in-group' if they show social cues such as positive and negative emotion?

decisions and show different patterns of brain activity than when engaging with humans, even though they consider machines to be social actors. Results also suggest that people experience less emotion when dealing with machines than with humans, and that machines tend to be treated as members of an out-group.

HYPOTHESES FOR IMPROVED HUMAN-MACHINE INTERACTION

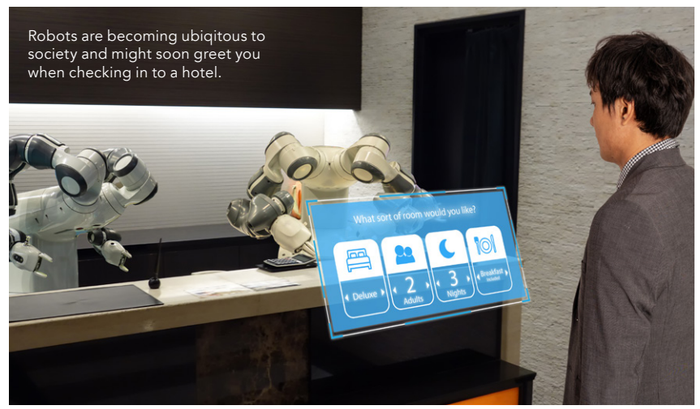
Dr de Melo and Dr Terada explain that as autonomous machines are becoming ubiquitous in society, it is essential that we find ways to foster cooperation between them and humans. Moreover, these solutions will have to surmount the unfavourable biases. Underpinned by their previous work and a review of other research, they put forward two hypotheses: Firstly, "associating positive cues of cultural membership could mitigate the default unfavourable bias people have towards machines". Secondly, "emotion expressions could override expectations of cooperation based on social categories".

THE PRISONER'S DILEMMA

The researchers recruited a total of 945 participants, 468 from the United States and 477 from Japan. Participants were paired with counterparts of either the same or different culture before taking part in 20 rounds of the prisoner's dilemma.

People have a tendency to make less favourable decisions and be less cooperative with machines than with humans.

In the dilemma, two players simultaneously make a decision to either defect or cooperate. Decision theory would suggest defection is the best response regardless of the counterpart's decision: if you think your counterpart is going to defect then you should defect as well; if you think your counterpart is going to cooperate, then you should still defect



Robots are becoming ubiqitous to society and might soon greet you when checking in to a hotel.

The humanoid robot Pepper has been designed to assist customers in retail stores.

and maximise your payoff. If both participants follow this reasoning, however, they will end up worse off than if they had cooperated.

HUMAN VS AUTONOMOUS MACHINE COUNTERPARTS

The experiment was fully anonymous, in that both the participants and experimenters were anonymous to each other. The participants were told that they would be engaging with either another participant or an autonomous machine. However, to maximise experimental control, all participants actually engaged with a computer script. The focus of the experiment was to investigate whether participants would cooperate differently with humans as opposed to machines, and if any difference could be regulated with culture or facial expressions of emotion.

Both human and machine counterparts were able to express emotion using virtual faces from either US or Japanese culture. They presented a competitive, neutral, or cooperative disposition. Earlier studies by Dr de Melo revealed that emotional expressions can influence cooperation in the prisoner's dilemma, so the researchers chose a number of set patterns displaying sequences of competitive, cooperative or neutral emotions. Using a measure of the cooperation rate, averaged across all rounds of the prisoner's dilemma, the researchers carried out statistical analysis employing a 2 × 2 × 3 between-participants factorial design. This enabled analysis of the effects of counterpart type (human or machine), counterpart culture (United States or Japan) and emotion (competitive, neutral or cooperative) to be carried out simultaneously.

When counterparts demonstrated cooperative emotion, there was no significant difference in cooperation with machines or humans.

PAIRINGS OF DIFFERENT CULTURES

When pairs were of different cultures, participants cooperated more when they thought they had human counterparts rather than machines. There was more cooperation when participants were paired with cooperative and neutral partners than competitive counterparts. In addition, when counterparts demonstrated competitive or neutral emotions, their counterparts cooperated more with humans than machines. When counterparts demonstrated cooperative emotion, however, there was no significant difference. These results applied to participants from both Japan and the United States.

PARINGS OF THE SAME CULTURE

Contrastingly, when pairs were of the same culture, there was no significant difference in the cooperation of participants with either human or computer counterparts. Nevertheless, there was more cooperation with cooperative and neutral counterparts than competitive ones. Once again, there was no significant difference between Japanese and the US participants' results.

SOLUTIONS TO OVERCOME BIAS

The researchers offer two solutions to overcome the unfavourable bias towards machines and improve human-machine cooperation. Firstly, the experiment demonstrated that with participants from both cultures taking part in human-machine interaction, a straightforward culture cue, by way of the ethnicity of the computer's virtual face, was enough to mitigate the bias. Secondly, mechanisms conveying affiliative intent, in the form of facial expressions of emotion, promoted human-machine cooperation, overriding the default expectations of coalition alliances derived from social categories. Furthermore, when machines showed positive emotion, such as joy following cooperation and regret following exploitation, the researchers found that people cooperated with machines every bit as much as with the human counterparts.

EMOTION INFLUENCING DECISION MAKING

The research team observed that emotion had the strongest effect in their experiment. A machine displaying a virtual face from a different culture group would be treated as an in-group member through astute visual presentation of emotion, in this case expressing joy after cooperation and regret following exploitation. This research highlights that emotion is a powerful influence on human behaviour, and decision making in particular. Moreover, the research demonstrates that the default associations of social categories can be overturned. This is encouraging as it may be difficult to control the perception of social categories in machines.

These results can inform designers of autonomous machines how to overcome the unfavourable bias towards machines and offers solutions to improve the level of cooperation in human-machine interaction.

The researchers reflect that given the increasing divisiveness in society, it is not surprising to find autonomous machines being perceived as outsiders and, therefore, less likely to reap the benefits afforded to members of the in-group.

This research demonstrates that humans will fall back on their established psychological mechanisms to ascertain associations and cooperate with machines. It is reassuring to know that our behaviour with machines is underpinned by the same psychological mechanisms we use with humans and provides opportunities to reduce negative bias with machines.

The researchers conclude that "since autonomous machines can be designed to take advantage of these psychological mechanisms driving human behaviour, they introduce a unique opportunity to promote a more cooperative society".

Social robots Si Si and Xiang Xiang are on display at Hefei University of Technology.

research OUTREACH

The public outreach magazine for the research community

researchoutreach.org

Partnership enquiries: simon@researchoutreach.org