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


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.
Informatics & Technology | Celso de Melo & Kazunori Terada

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.

A utonomous 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 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. 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.

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

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

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

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".
Partnership enquiries: simon@researchoutreach.org

researchoutreach.org

Partnership enquiries: simon@researchoutreach.org