Distinguished Talks @

Ciências ULisboa | Informática

LASIGE driven by excellence

Thu 23 March 2023

Evolution of a Theory of Mind

Speaker

Tom Lenaerts

Affiliation

Université Libre de Bruxelles

When

March 23, 14:00

Where

Room C6.3.27

Abstract

Within AI research, proposals have been launched that argue for the creation of a new science of cooperative AI to address the problems that may arise from the diverse ecologies of AI systems that interact in complex ways with other AI and humans. One element of this new science would be the realization of AI that understand the motivations and ambitions of the other AI and humans in the mix. Such an understanding requires AI to have a Theory of Mind (ToM), i.e., they need to recognize that the other entities have cognitive and emotional states and reason about them.

The existence and characterization of a ToM in humans, upper primates and other species has been under investigation for decades now. Neurological disorders like autistic spectrum disorder have been linked to the impairment of ToM where affected individuals have difficulties in assigning internal states to, or recognise emotions expressed by, others. In AI, a variety of models of different complexity have been explored revealing advantages of introducing a form of ToM in artificial systems. What so far has remained underexplored is when a ToM may evolve if there is a choice between using it or not in the imagined complex ecology. In other words, what may be the conditions under which ToM will evolve? To answer this question, we developed an Evolutionary Game Theoretical model in which a finite population of individuals use strategies that incorporate (or not) a ToM, modelled using level-k recursive reasoning, to infer a best response to the anticipated behaviour of others within the context of the centipede game. We calibrated our results to the existing experimental knowledge on how humans behave in this game. We find that strategies incorporating a ToM evolve and prevail under natural selection, provided individuals make cognitive errors, and a temptation for higher future gains is in place. We found furthermore that such non-deterministic reasoning co-evolves with an optimism bias, favouring the selection of a new equilibrium configuration in the centipede game, which was not anticipated to date. Our work reveals not only a unique perspective on the evolution of bounded rationality but also a co-evolutionary link between the evolution of ToM and the emergence of misbeliefs.

Bio

Tom Lenaerts (PhD 2003) is Professor in the Computer Science department at the Université Libre de Bruxelles (ULB), where he is co-heading the Machine Learning Group (MLG). MLG targets machine learning, AI and behavioural intelligence research focusing on time series analysis, causal and network inference, collective decision-making, social AI and behavioural analysis with applications in mobility, medicine, finance and biology. He also holds a partial affiliation as research professor with the Artificial Intelligence Lab of the Vrije Universiteit Brussel and is affiliated researcher at the Center for Human-Compatible AI at UC Berkeley. He is chair of the Benelux Association for Artificial Intelligence, the main representing entity for academic AI research in the region and expert in the Global Partnership on Artificial Intelligence. He is working in a variety of interdisciplinary domains using AI and Machine Learning, involving topics like optimisation, multi-agent systems, collective intelligence, evolutionary game theory, computational biology and bioinformatics.