Beyond Goal-Rationality: Traditional Action Can Promote Goal-Achievement in Socially Situated Agents

Chloe M. Barnes, Anikó Ekárt and Peter R. Lewis
Aston Lab for Intelligent Collectives Engineering (ALICE)
Aston University, Aston Triangle, Birmingham B4 7ET, UK
Email: {barnecm1 — a.ekart — p.lewis}@aston.ac.uk

Within socio-technical systems, we are increasingly delegating decisions from humans to so-called 'intelligent' machines. These machines typically act with reference to a given goal, such as an objective function, utility or goal-state. Yet, goal-rationality is just one form of social action that humans have evolved to exhibit [1]. Humans act socially in more complex ways than this, including those driven by values, emotions and traditions [1]. Both humans and machines that comprise socio-technical systems are inherently socially-situated; despite this, machines are typically neither socially aware nor do they act socially. Specifically, they do not perceive or reason about the effect other systems can have on their own ability to learn and evolve, nor are the actions that they direct towards others driven by any broader social meaning [2] - unlike in humans.

In this talk, we will explore the impact of purely goal-rational action on systems that are socially-situated, demonstrating that this can lead to instability in co-evolutionary learning. We will then show that by complementing goal-rational action with traditional action, more stable dynamics can be observed. We operationalise traditional action in this context as acting in a similar way to the rest of the population. To study this, we introduced the River Crossing Dilemma [3], a testbed designed to explore how two systems that are unable to perceive each other in a shared environment are able to pursue individual goals. This extends the River Crossing Task [4], a 2D grid-world environment, by gamifying the task to capture social dynamics and interference with more than one agent.

We observed that there are unintended consequences of coexistence in the River Crossing Dilemma. This social interference is experienced as unknown and unanticipated changes in the environment, and has a negative effect on the learning and stability of goal-achieving behaviour. This reflects how components of socio-technical systems interact in unexpected and unanticipated ways in a shared world, and can also be unknown to each other at runtime. However, we further show that this can be mitigated by complementing the goal-rationality as seen in current systems with traditional action. By incorporating traditional action, the evolution of goal-achieving behaviour can be stabilised.

This is important since socio-technical systems are increasingly being designed to operate in dynamic, uncertain, and social environments. A transition to more socially intelligent systems, with the ability to learn about others in their environment and the impact they can have on their own learning and success, is therefore necessary. Our results suggest that intelligent machines within socio-technical systems should be designed to exhibit social action beyond purely goal-rational behaviour to this end, as a step towards systems that can learn and reason about their social situatedness.

More broadly, as we do delegate more decisions to intelligent machines, we posit that it will be important to also capture elements of evolved human social behaviour. Social action and social self-awareness are two essential aspects that so far remain largely unexplored in the design of intelligent machines.

References

[1] M. Weber, G. Roth, and C. Wittich, *Economy and society : an outline of interpretive sociology*, G. Roth and C. Wittich, Eds. University of California Press, 1978.

[2] K. Bellman, J. Botev, H. Hildmann, P. R. Lewis, S. Marsh, J. Pitt, I. Scholtes, and S. Tomforde, "Socially-Sensitive Systems Design: Exploring Social Potential," *IEEE Technology and Society Magazine*, vol. 36, no. 3, pp. 72–80, 2017.

[3] C. M. Barnes, A. Ekárt, and P. R. Lewis, "Social action in socially situated agents," in *Proceedings of the IEEE 13th International Conference on Self-Adaptive and Self-Organizing Systems*, 2019.

[4] E. Robinson, T. Ellis, and A. Channon, "Neuroevolution of agents capable of reactive and deliberative behaviours in novel and dynamic environments," in *Advances in Artificial Life*. Springer, 2007, pp. 1–10.