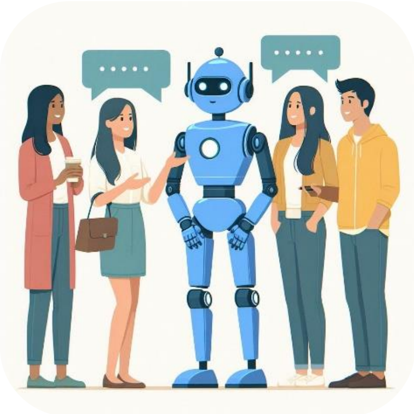
MERC PhD Project Proposal 2023/2024

Title of the research project

The Bot in the Group:
Steering Collective Behavior in Complex Human Networks



Keywords

Complex networks; Human-in-the-loop; Model identification; Reinforcement learning

Supervisors

Supervisor 1: **Marco Coraggio** (marco.coraggio@unina.it)
http://marco-coraggio.com
Expertise: Control theory, Complex networks, Data-driven control

Supervisor 2: **Mario di Bernardo** (mario.dibernardo@unina.it)
https://sites.google.com/site/dibernardogroup
Expertise: Control theory, Complex systems, Dynamical systems

Supervisor 3: **Francesco De Lellis** (francesco.delellis@unina.it)
Expertise: Machine learning, control systems

Project description

Introduction

Context and motivation. As human beings, we are hardwired to crave interaction with other people and have built our societies around it: this inclination is exemplified by people conversating while sharing a meal, playing sports together, and by the large economic and political organizations in our world. For interpersonal communication to be effective, subtle yet nuanced interactions must take place, such as the display of emotions and of willingness to lead or follow. In virtual/augmented reality environments or even in physical ones, it is possible to steer the collective behavior of interacting people, e.g., via the addition of one or more so-called *autonomous avatars*, driven by computers and adaptive algorithms. The goal of such interventions is to foster positive aspects of the group interaction, such as inclusion, altruism, and coordination, which has been linked to wellbeing [Bieńkiewicz, 2021]. To achieve this though, it is necessary that the avatars learn to convincingly interact with people, portraying the whole spectrum of subtle signals that humans use for effective communication, while abiding by ethical guidelines. One important difficulty lies in modeling human behavior to the required level of complexity, to then use this knowledge to inform the behavior of the autonomous avatar.

In this project, we aim to develop such methodologies to steer collective behavior in complex human networks, to promote positive features such as coordination and inclusion.

State of the art. In 2011, to study motor interaction between two people, Noy and colleagues introduced the *mirror game*, a “collaborative 2 player game whose purpose is to enjoy creating motion together that is synchronized and interesting” [Noy, 2011]. The game was then extended to a multiplayer scenario to study group interaction; in particular, in [Alderisio, 2017a], it was shown that a group of people performing the exercise with the aim to synchronize the oscillating motion of their arms could be modeled as a simple network of *Kuramoto phase oscillators*. Although the natural frequency of motion of the participant was captured by the model, more detailed aspects such as the tendency to adapt individual frequency or to follow others remained unidentified. Subsequent studies described phenomena such as the tendency of people to slow down when attempting to synchronize [Calabrese, 2022] and the strategies that people put into place to lead the group [Calabrese, 2021]. In [Shalal, 2020], the coordination capability of a real group of violinists was thoroughly studied, observing peculiarly human behavior, such as ignoring frustrating auditory input to reach a stable solution.

The insertion of so-called *cognitive architectures* (algorithms and control laws driving autonomous avatars) in human groups to steer collective motor behavior was preliminary investigated in [Lombardi, 2021], using a deep reinforcement learning model, trained to exhibit specific desired kinematic features, while playing the multiplayer mirror game. In [Mckee, 2023], considering networks of real people playing monetary games, a reinforcement learning algorithm was used to rewire interaction edges, encouraging altruism more effectively than the best methods available in the literature.

Objectives

The workplan and objectives are flexible and will be adapted depending on the inclination of the student and the results obtained in the early phases of the project.

O1. Develop methods to identify personalized models of interaction.
O2. Develop cognitive architectures that exploit these models online to foster positive interaction.
O3. Extend the methodologies to more complex movements through phase reduction.
O4. Validate the methodologies on groups of real people performing joint motor tasks.

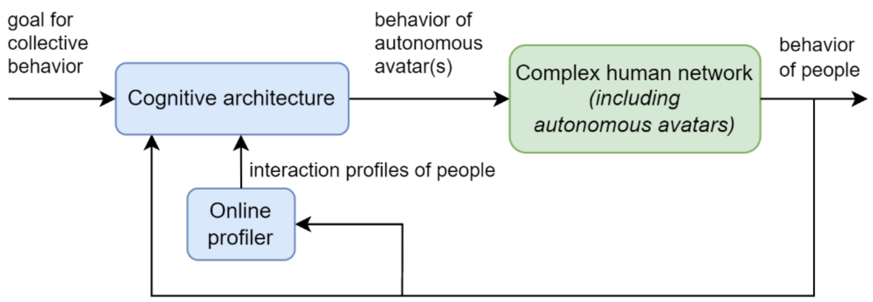


Figure 1: Block scheme of a complex human network steered by a cognitive architecture.

Methodology

Inspired by [McKee, 2023], we aim to achieve Objective O1 by developing a methodology to identify online task-dependent, accurate, and personalized models of human interaction. In particular, given a task, informed by real data, we will identify simple yet expressive models, where different parameter sets are associated to different *profiles*, each related to a specific way of interacting with others. A supervised or possibly unsupervised (if the number of profiles is initially undecided) machine learning technique will be used to associate the sets of parameters to the different profiles. The first benchmark application considered will be the multiplayer mirror game, where the classic Kuramoto oscillators will be extended with time-varying natural frequency and person-dependent coupling gains, to model individual disposition to lead/follow and the attention given to others. Real interaction data will be gathered both through the use of CHRONOS [Alderisio, 2017b]—a computer-based platform to play the multiplayer mirror game—and by experiments with real people, performed at the University of Montpellier, in the context of the European Research Project Sharespace (sharespace.eu).

Next, we will focus on Objective O2 and on the design of *cognitive architectures* (CAs), that is the programs and algorithms that drive the autonomous avatars used to steer the group's collective behavior. Initially, a formal analysis of the control design process will be carried out, both to formalize the assumptions required for the CA, and to show the feasibility of the operation. Then, the implementation of the CA will be based on reinforcement learning algorithms; since training them with real people would require an unfeasible number of interactive experiments, we will train them with data generated from the models derived in fulfilment of Objective O1. The CA will need to simultaneously identify the profile of the people it is interacting with and determine the best action to effectively steer the group towards the desired behavior (see Figure 1). In principle, this influence can be put into place both by the actions of an autonomous avatar, or even by modifying people's interaction, e.g., adding/removing edges in the communication graph. On the initial benchmark problem of multiplayer mirror game, the goal will be that of increasing coordination, as measured by the *order parameter*, even in the presence of people with motor impairments, or even uncooperative group members. The control design will be supported by a formal study of convergence and performance, where possible. The analytical study will also cover the development of advanced reinforcement learning techniques, potentially also adaptable to other application domains.

Once the personalized identification system and the advanced cognitive architectures are validated on low-dimensional models, to achieve Objective O3, we will investigate phase-reduction (e.g., autoencoders, based on supervised machine learning [Starke 2022]) and dimensionality-reduction techniques to apply the devised methodologies to steer collective human behavior in more complex tasks, where more degrees of freedom are necessary.

Finally, to complete Object O4, we will apply the developed methodologies, in conjunction with the dimensionality-reduction technologies to steer the motion of people performing more complex movements, such as physiotherapy exercises and sport training sessions. The data for these tasks will be acquired in the context of the European Research Project Sharespace.

Relevance to the MERC PhD Program

Relevance and beneficiaries

The technologies developed in this project will be used to populate future virtual reality and augmented reality environments with autonomous avatars, with possible application in healthcare (e.g., rehabilitation exercises), sports training, and artistic/recreational activities such as dance or videogames. Being on the forefront of this research field will also allow us to develop cognitive architectures designed through the so-called *ethics by design* approach, imbuing ethical principles such as privacy, accessibility, and inclusion directly in the operation of the algorithms.

Relevance to the MERC PhD program

This project is focused on the analysis and steering of complex human networks, both using model-based analytical tools (e.g., phase oscillators network models) and empirical ones (e.g., reinforcement learning). Therefore, the project has an important interdisciplinary component, combining complexity science, dynamical systems, statistics and machine learning.

Skills. During the project, both tutored by the supervisors and through self-study, the student will develop skills in several fields, including:

- Dynamical systems, with a focus on model identification,
- Network science,
- Machine learning,
- Data analysis.

Additionally, tutoring will also focus on sharpening the student's technical writing and presentation skills, as well as developing the ability to study scientific literature swiftly and effectively.

References

Key references

- F. Alderisio, G. Fiore, R. N. Salesse, B. G. Bardy, M. di Bernardo, "Interaction patterns and individual dynamics shape the way we move in synchrony," Scientific Reports, 7(1):6846, 2017a.
- M. Lombardi, D. Liuzza, M. di Bernardo, "Dynamic input deep learning control of artificial avatars in a multi-agent joint motor task," Frontiers in Robotics and AI, 8, 2021.
- K. R. McKee et al., "Scaffolding cooperation in human groups with deep reinforcement learning," Nature Human Behavior, 7(10), 2023.
- S. Shahal et al., "Synchronization of complex human networks," Nature Communication, 11(1), 2020.

Additional references

- F. Alderisio, M. Lombardi, G. Fiore, M. di Bernardo, "A novel computer-based set-up to study movement coordination in human ensembles," Frontiers in Psychology, 8, 2017b.
- M. M. N. Bieńkiewicz et al., "Bridging the gap between emotion and joint action," Neuroscience & Biobehavioral Reviews, 131:806–833, 2021.
- C. Calabrese, M. Lombardi, E. Bollt, P. De Lellis, B. G. Bardy, and M. di Bernardo, "Spontaneous emergence of leadership patterns drives synchronization in complex human networks," Scientific Reports, 11(1), 2021.
- C. Calabrese, B. G. Bardy, P. De Lellis, M. di Bernardo, "Modeling frequency reduction in human groups performing a joint oscillatory task," Frontiers in Psychology, 12, 2022.
- L. Noy, E. Dekel, U. Alon, "The mirror game as a paradigm for studying the dynamics of two people improvising motion together," Proceedings of the National Academy of Sciences, 108(52):20947–20952, 2011.
- S. Starke, I. Mason, T. Komura, "DeepPhase: periodic autoencoders for learning motion phase manifolds," ACM Transactions on Graphics, 41(4):1–13, 2022.

Joint supervision arrangements

The student will meet at least weekly with at least one of the supervisors. The whole team will meet at least once every 1 or 2 months for a progress update.

Location and length of the study period abroad (min 12 months)

The student will be able to spend a research period (or research periods) at the lab of one of the 13 international partner in the European research project Sharespace (see sharespace.eu/partners for a list) and at the University of Sydney under the supervision of prof Michael Richardson.

Any other useful information

The project will involve collaboration with the University of Montpellier, where experiments on real people performing joint motor tasks can be performed, and participation to the European Research Project Sharespace, tasked with designing next-generation shared hybrid spaces in virtual and augmented reality.