

## MAP 2024: Combining Cognitively Inspired Computational Approaches with Reinforcement Learning (RL) Techniques

Thanks to the prestigious [Harris Faculty Fellowship](#), I was able to investigate deeper exciting research problems and propose a vibrant MAP 2024! It starts from the question:

*Can the coordination/cooperation dichotomy in MAS enlighten both, empathy investigation and the development of AI systems that cooperate purposefully with humans, expanding instead of replacing human intelligence?*

**MAP Trajectory.** In previous work, I successfully designed and developed a computational architecture that models moral reasoning and empathy in multi-agent tasks [3 - 5]. When I joined Grinnell College in 2020, I began my efforts to enhance and expand the architecture: I started modeling an **Empathy-Driven Architecture (EDA)** to examine the role of emotional processes and empathy in weighting goals for decision-making. I launched the ELBICA lab and outcomes from 2021's MAP led me reflect on the differences between moral reasoning and moral intuition and their influence on a group's reputation and moral behavior [1], and to test the robustness of cooperation against cascading failures in different network topologies [6].

In 2022, the ELBICA lab started investigating sense-making and clues of emotions and common sense in a set of images. I collected over 300 images that encode abstract flavor (such as subtle jokes or metaphors), and a subset of my team started describing and categorizing the images in a dataset whereas another subset (which does not have access to the images) started designing and translating the descriptions into 3-D scenes. My end goal is to disseminate this work using a Shiny app to enable people to interact with our descriptions, dataset, and 3-D scenes. In addition, my first-year Tutorial course inspired us to investigate how anthropomorphism impacts the way we interact with machines [7]. ELBICA lab's efforts got me intrigued about a dichotomy cooperation vs. coordination.

In 2023, outcomes from our 3-D project [8] led us to assistive technology tools, and I am deeply enthusiastic about it! Additionally, the ELBICA lab developed multi-agent systems (MAS) that implement traditional RL techniques to drive insights into the distinction cooperation/cooperation; our testbeds included distributed and networked agents (paper submissions under way).

**MAP 2024** (eight MAP students), is motivated by the role of concepts in the design and explainability of AI systems. AI's ubiquity highlights the importance of a well-thought-out use of concepts, which should be held even for systems that are too simple or not intended for human interaction. For instance, in the MAS literature, it is common to see a comparison of agents' performances across different settings. However, it is not as common a reflection on the meaning of applied concepts along with how/if meanings still hold in the experimental results, and this is a gap the ELBICA lab will address for coordination vs. cooperation. We will focus on two main areas, split into five tasks. The areas are:

*Investigating a dichotomy coordination vs. cooperation from the point of view of*

- a) *Interconnections between anthropomorphism and bio-inspiration in AI systems.*
- b) *Multi-Agent Systems (MAS) and independent reinforcement learners.*



## Project Details

Note: Students working on tasks I – III, in addition to investigating their own research problems, will help Dr. Elliott brainstorm on creating a visual dictionary of concepts.

- a) *Investigating a dichotomy coordination/cooperation from the point of view of: **Interconnections between anthropomorphism and bio-inspiration in AI systems. Three students, tasks I, II.***

Eligibility: students that took at least one course in the CS dept., enjoy a multidisciplinary approach, and learning/applying advanced topics in CS.

**Task I, two students:** will study anthropomorphism and bio-inspired AI systems. They will bridge our topics with the paper: "[The Scientific Study of Consciousness Cannot and Should Not Be Morally Neutral](#)". (Task inspired by my first-year Tutorial and a chat with Professor Andy Mobley.) This task focuses more on theory than coding. Students will:

- **Student 1.** Study the Escape System of Cockroaches and how/if they can inspire MAS tasks – in particular, tasks that require coordination or cooperation.
- **Student 2.** Start from [Malone and Crowston \(1990, 1994\)](#), and [Coordination Theory: A Ten-Year Retrospective](#) to investigate examples of coordination, what distinguishes it from cooperation, and the role of decision-making in defining a task as coordination or cooperation. The student will bridge examples from biology with MAS. Finally, the student will help to create 1. A framework for a mapping of ensembles, and 2. Approaches for visually depicting them.

### Task II, one student.

- **Student 3.** This MAP derives from applications of our work with concepts and 3D scenes. The student will investigate connections between **assistive technologies**, 3D modeling, concepts related to homeostatic regulation, decision-making, environment, task, and affordances. No coding is involved, and first-year students are welcome to apply!

- b) *Investigating dichotomy coordination/cooperation from the point of view of: **MAS and independent reinforcement learners, five students, tasks III – V.***

Eligibility: CS students that have successfully concluded CSC207, who enjoy programming in Java, and learning about advanced topics in CS.

All students must read: [Multiagent Learning: Basics, Challenges, and Prospects](#) and take a look at [Economics and the Theory of Games by Vega-Redondo, Fernando](#)

### Task III, one student.

- **Student 4.** The student will scrutinize various MAS environments and tasks to create a catalogue of environments in the MAS literature focusing on independent reinforcement learning tasks and challenges. This task is theory-heavy (little to no coding): the student will study reinforcement learning techniques, game theory, and challenges in multi-agent systems. "[Multi-agent deep reinforcement learning: a survey](#)" and "[Look! It's a Computer](#)



[Program! It's an Algorithm! It's AI!": Does Terminology Affect Human Perceptions and Evaluations of Algorithmic Decision-Making Systems?"](#) will be used as a starting point.

#### Task IV, two students.

- **Students 5 and 6.** will implement and test a chosen experiment in MAS to investigate cooperation vs. coordination. Possible experiments are: multi-turn games, the pursuit domain, or the Michigan Intelligent Coordination Experiment (MICE) testbed. (Task heavy on both: theory and coding; students must feel comfortable with the Java programming language.)

#### Task V, two students.

- **Students 7 and 8.** Will investigate two papers to identify connections, if any, in the context of [distributed representations](#). The papers are [CaMeRa: A Computational Model of Multiple Representations](#) and [Self-Improving Reactive Agents Based on Reinforcement Learning, Planning and Teaching by Long-Ji Lin](#). (Task heavy on both: theory and coding; students must feel comfortable with the Java programming language.)

**How student(s) will fulfill the goals of the MAP.** As usual, ELBICA students will have a vibrant and profound multidisciplinary research experience! They will read a lot, hold various research-focused meetings, and write a manuscript to be submitted to a suitable peer-review conference.

### Selected References

- [1] Yu, X., Morri, R. and Elliott, F.M. EDA, [An Empathy-Driven Computational Architecture](#). 9<sup>th</sup> [Workshop on Goal Reasoning](#), ACS 2021.
- [2] Elliott, F.M., Stassun, K, and Kunda, M. Iaci - a human-inspired computational architecture to help us understand visual data exploration. In *Procs. of the Sixth Annual Conference on Advances in Cognitive Systems*. Stanford, California: ACS, 2018a.
- [3] Elliott, F.M., Ribeiro, C. H. Emergence of cooperation through simulation of moral behavior. In *Hybrid Artificial Intelligent Systems (HAIS 2015)*. Procs. of the 10th I. Conf. on Hybrid Artificial Intelligence Systems, Bilbao, Spain. *Lecture Notes in Artificial Intelligence*, volume 9121, pages 200–212. Springer International Pub., 2015.
- [4] Elliott, F.M., Ribeiro, C. H. Moral behavior and empathy modeling through the premise of reciprocity. In *Procs. of the 1st International Conference on Human and Social Analytics*. St. Julians, Malta: Huso, 2015.
- [5] Elliott, F.M., Ribeiro, C. H. A computational model for simulation of moral behavior. In *Procs. of the I. Conf. on Neural Computation Theory and Applications (NCTA-2014)*, pages 282–287. SCITEPRESS (Science and Technology Publications), 2014.
- [6] M. He, M. Gao, Y. Gao, and F. Elliott. Cascading failures and the robustness of cooperation in a unified scale-free network model. In *I. Conf. on Complex Networks & Their Applications*, pages 365–376. Springer, 2021.
- [7] Swaim, E. and Elliott, F. Complex Behavior Vs. Design - Interpreting AI: Reminders from Synthetic Psychology. The Ninth International Conference on Human and Social Analytics HUSO 2023, Barcelona, Spain.
- [8] How do Abstraction and Emotions Travel Different Spaces? To appear at: The Tenth International Conference on Human and Social Analytics HUSO 2024.

