PhD project proposal:
Simulating the Emergence of Complexity in Collective Intelligence: principles of cognition and social epistemology incorporated into agent-based modelling.

Maurin Gilles
Final-year student in computer science.
Clermont Auvergne University, France.

April 28, 2026

Background

The ongoing AI spring is dominated by generative deep learning models. These innovations come with a great enthusiasm, but also criticism from many experts. Some pinpoint empirical dangers (for the environment, for society), and other express reservations on a more fundamental level. In the middle of these critiques, there is the alignment problem, and the most common limit of deep learning in this regard, formulated in many often equivalent ways, suggests that deep learning models would not "understand", but only predict, imitate. This limit is famously illustrated by Bender et al.'s "stochastic parrots" analogy [18].

This critique is broadly acknowledged and accepted, and today a lot of work both in academic world and tech companies aims at combining the observable power of deep learning with more formal and controllable forms of knowledge. Behind this so-called "neuro-symbolic" approach lies an immense variety of methods and strategies, all expecting to make possible "the" best explainable artificial general intelligence [17, 21].

Nevertheless, this approach is based on an individualistic paradigm that is rarely discussed: one model for everything. Surely the individualistic paradigm shows extraordinary results, but do these results suit the society's expectations for AI is a question that needs to be investigated, and one angle in which this paradigm can be challenged is its ability to handle complexity.

Here, "complex" is to be understood in the sense that Edgar Morin gives it [10, 11, 12]. It means ideas that cannot be reduced to a single principle, that is composed of multiple components bound with non-trivial interactions and recursive dynamics. In Morin's theory, complex ideas can neither be represented as whole nor as a sum of its parts, thus the necessity to embrace the multiplicity.

A strong motivation for models able to handle complexity is a feature of real-world problem that optimisation-based strategies rarely genuinely address: they are what Rittel and Webber first called “wicked” problems [2]. Among the ten characteristics of wicked problems mentioned in the original paper, we can for example mention that they have no stopping rule, no true/false solution neither even an enumerable set of potential solutions, that they can only been explained through a certain angle and may be seen as a symptom of other wicked problems.

At a time when artificial intelligence is likely to drive progress from political decisions to scientific breakthroughs, these questions address crucial matters in ethic and theory of science. My PhD project proposal proposes to explore the potential of a collective paradigm through an iterative development of a model using cognitive agents to operate in a complex setting. In other words, I will try to answer the question: “How can an agent-based model simulate a form of collective intelligence exhibiting epistemically complex behaviours?"

Methodology

Model Framework

In the model we aim at creating, agents will embed a reasoning system that makes them “cognitive”. They evolve virtually in an environment E, which is a grid of cases that can be in several states. When an agent A is on a grid, it perceives a portion e of the grid centred on itself. The reasoning system of A allows it to compute a desired version of e, namely $f_A(e)$, which shows how A would like to change the environment around itself. The collective intelligence appears when A interacts with another (or multiple other, if the model allows to) agent B according to a probability determined by A and B but also the environment E in which they evolve. During this interaction, B will compute its own desired version $f_B(e)$, and based on e, $f_A(e)$ and $f_B(e)$, the environment may or may not actually change, and A and B may or may not update their own reasoning system. Eventually, A and B may or may not change their position in the environment. Figure 1 displays this whole dynamic.

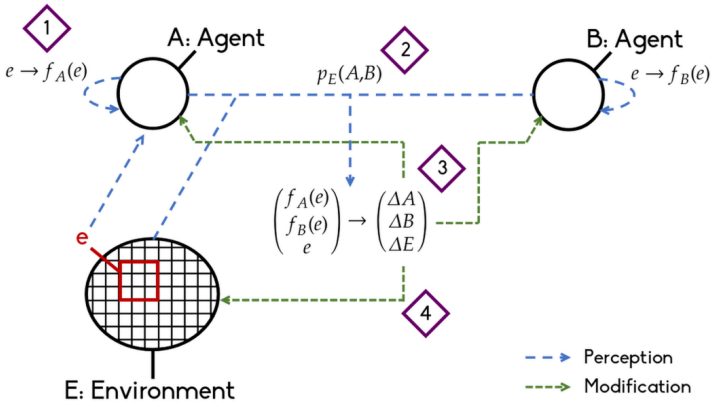

Figure 1: Model structure

The four diamonds on Figure 1 correspond to the four parameters of the model that we aim at finding:

◇1 What is the reasoning system of an agent. For a large size of e and/or a large number of possible states, agents cannot have a specific rule for every possible input, so a non-trivial reasoning system has to be defined. Logic rules are a a good candidate for this reasoning systems for they are explainable and easy to customise when agents interact. Later models may include other types of reasoning systems, for example interpolation functions or neural networks.

◇2 What law rules interactions between agents. This question is a major interest in social epistemology and computational sociology, with suggestions seldom inspired from natural sciences.

◇3 How do agents update their inner reasoning system based on their interaction with another agent. A question inextricably bounded with the first one, whose answer will also be guided by the field of social epistemology.

◇4 How does the environment change and how do agents move in it. A question at the intersection of computational sociology and dynamic systems.

The initialisation of a model corresponds to an initialisation of agents' reasoning systems and an initial state of the environment. The evaluation of a model is not the final state of the grid (which should not systematically converge), but its evolution. Additionally, the evolution of agents' reasoning systems is also to be interpreted.

The choice of a such framework for our model is motivated by its incredible versatility. On the one hand, the iterative modification of cases in a grid is known for its computing expressiveness, and it is expected that a manual tuning of agents' reasoning systems allow the solving of formal problems encoded in the initial environment, thus justifying the intelligence potential of our model. On the other hand, when the simulations will run in configurations that foster emergence, we expect complex behaviours to appear. In particular, this model framework is potentially able to implement the core features of Morinian complexity, as summarises Table 1:

Feature of Morinian complexity	**Implementation in the model framework**
Dialogic	Several agents can have reasoning systems with opposite visions.
Hologrammatic	The environment is the setting in which agents evolve and interact, while agents have a reasoning system that applies on the whole environment.
Recursion	Agents influence the environment, which influence agents, and so on.

Table 1: Fields involved in this transdiciplinary project.

Eventually, the evolution of the several parts of the model natively mirrors what Morin calls "auto-eco-organisation": the ability of a system to be autonomous and interact with its environment.

Development loop

In a project that aims at simulating the emergence of complexity, it is natural to employ a complex approach. A classic top-down approach is obviously unsuitable for a project studying emergent phenomena, while full bottom-up approach usually make no expectation on the emergent results but only analyse them.

The method for this project is expected to be **comparative** and **incremental**, following a four-step development loop described in the following and illustrated in Figure 2:

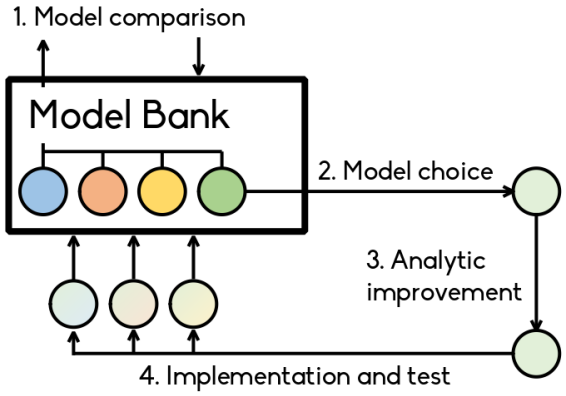

Figure 2: Model development loop

1. Model comparison: among existing models, analyse how they differ and what hypotheses each of them lie on, to see how they provide insights for new model improvements. Keeping a "Model Bank" is therefore really important, and no model should ever be completely dropped, as they might implement features that would prove relevant in regard to properties that would appear in later models. This step is the first, because a study of existing models that can be related to our issue would be the very first thing to do during the PhD.

2. Model choice: with respect to the comparative analysis of models from the Model Bank, choose one to improve or several to combine in a certain way.

3. Analytic improvement: this is where insights from fields other than plain computer science prove the most important, because they should give directions and justifications for changes in the chosen model(s). A variation of a certain model mostly affects one of the four abovementioned parameters, but the development of the project might call for variations in other metaproperties of the framework. However, we want this variation to respect a rigorous theoretical framework, which can differ from one model to another, but should always maintain the whole justifiable.

4. Implementation and test: the model(s) obtained has to be implemented and tested in practice. Agent-based modelling always require fine-tuning, and large-scale simulation in terms of time, space, and number of agents. Additionally, cognitive agents will embed reasoning systems with their own mechanisms. These constraints call for high-performance computing techniques while the project handles concepts of high level of abstraction: this famous "Two-Language Problem" will be addressed with the Julia Programming Language.

Theoretical Framework

This project is inherently transdisciplinary, and great attention has to be placed in how every concept can carefully be associated with others. There are six major fields influencing the project, each of them being already interdisciplinary with respect to "traditional" fields (see Table 2). Four are primarily important to provide clues for the four parameters to be tuned in our models, one is the technical field for the simulation itself, and the last is Morin's theory of complexity that guides and motivates the project.

- **Logic** ($\diamond$1): To represent agents' inner ideas and reasoning system, logic provides a very well studied framework. The context of the project is likely to guide towards some forms of non-classical logic, especially modal [7], fuzzy [9] and temporal [16] logic.
- **Cognition** ($\diamond$1): Representing how agents are influenced by their thoughts, i.e. what is their reasoning process, is a major topic from cognitive sciences. The frontier between cognitive psychology and artificial intelligence is a well-established field of study, beginning with the work of pioneer Herbert Simon [3].
- **Social Epistemology** ($\diamond$2, $\diamond$3): A major influence for this project is the field of social epistemology, and its leading figure Steve Fuller [8]. It gives a broad and solid framework for conception of knowledge in a collective way, from which clues for artificial collective intelligence can be derived.
- **Complex Systems** ($\diamond$4): A very important strategy in a such project is to think in terms of complex systems as they are developed in natural sciences [15], as it provides formal and scientifically strong keys for derived concepts such as chaos, nonlinear dynamics, and emergence, all crucial in the definition of a model's evolution.
- **Agent-Based Modelling (ABM)**: The goal of this project is to simulate autonomous agents in an environment, which falls in the field of Agent-Based Modelling. These types of models are widely used in natural science, but also in computational sociology since Joshua M Epstein and Robert Axtell's famous Sugarscape model [6], and works on Castelfranchi's works on cognitive agents simulation and artificial social systems [4, 5]. Some bases for cognitive agents can be taken from the growing field of Argumentative Agent-Based Models [24], or agents with LLM-driven cognitive systems such as Casevo [22].
- **Theory of Complexity**: The framework guiding our objective is the theory of complexity from Edgar Morin which has been introduced in the previous sections, and is deeply inspired by a connecting the abovementioned concepts of complex systems with work from sociologists, notably Émile Durkheim [1].

	Logic	Cognition	Social Epistemology	Complex Systems	ABM	Complexity
Computer Science	x			x	x	
Mathematics /Physics	x			x	x	x
Philosophy	x	x	x			x
Sociology			x		x	x
Psychology		x				

Table 2: Fields involved in this transdiciplinary project.

Objectives

The objective of the project is not only to exhibit a model that produces emergent complex behaviours, but also to comment the development process and the evolution of the models themselves.

Thus, we should obtain insightful toy models for artificial collective intelligence, along with a rich qualitative study of how this form of social cognition can be modelled.

Expected contributions include:

- A bank of models of artificial collective intelligence implementing our framework.
- A benchmark showing these models' to 1/ solve formal problems with manually chosen initial positions, and 2/ produce emergent complex behaviours.
- A detail of the iterative development process with comments on each choice and analysis of interesting phenomena appearing in each model.
- A qualitative study of a collective paradigm in practice, highlighting its relevance for complex (wicked) problems and scientific progress.

Bibliography

[1] Émile Durkheim. *Les règles de la méthode sociologique*. Paris, Payot, 1894.

[2] Horst WJ Rittel and Melvin M Webber. "Dilemmas in a general theory of planning". In: *Policy sciences* 4.2 (1973), pp. 155–169.

[3] Herbert Alexander Simon. *Models of thought*. Vol. 352. Yale university press, 1979.

[4] Cristiano Castelfranchi. *Artificial Social Systems: 4th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, MAAMAW'92, S. Martino Al Cimino, Italy, July 29-31, 1992. Selected Papers*. Vol. 830. Springer Science & Business Media, 1994.

[5] Rosaria Conte, Cristiano Castelfranchi, et al. *Cognitive and social action*. Garland Science, 1995.

[6] Joshua M Epstein and Robert Axtell. *Growing artificial societies: social science from the bottom up*. Brookings Institution Press, 1996.

[7] James Garson. *Modal logic*. 2000. URL: https://plato.stanford.edu/entries/logic-modal/.

[8] Steve Fuller. *Social epistemology*. Indiana University Press, 2002.

[9] Andrew M Mironov. "Fuzzy modal logics". In: *Journal of Mathematical Sciences* 128.6 (2005), pp. 3461–3483.

[10] Edgar Morin. *La Méthode-tome 3 La Connaissance de la connaissance anthropologie de la connaissance: La Connaissance de la connaissance. Anthropologie de la connaissance*. Média Diffusion, 2013.

[11] Edgar Morin. *La Méthode-tome 4 Les idées, leur habitat, leur vie, leurs moeurs, leur organisation: Les Idées. Leur habitat, leur vie, leurs moeurs, leur organisation*. Média Diffusion, 2013.

[12] Edgar Morin. *Introduction à la pensée complexe*. Média Diffusion, 2015.

[13] Christopher et al. Rackauckas. *SciML Open Source Scientific Machine Learning*. 2017. URL: https://github.com/SciML/.

[14] Michael Innes et al. "Fashionable Modelling with Flux". In: *CoRR* abs/1811.01457 (2018). arXiv: 1811.01457. URL: https://arxiv.org/abs/1811.01457.

[15] Stefan Thurner, Rudolf Hanel, and Peter Klimek. *Introduction to the theory of complex systems*. Oxford University Press, 2018.

[16] Valentin Goranko, Antje Rumberg, and EN Zalta. "Temporal logic". In: *DISPUTATIO PHILOSOPHICA* 25.1 (2020), pp. 93–104.

[17] Luís C Lamb et al. "Graph neural networks meet neural-symbolic computing: A survey and perspective". In: *arXiv preprint arXiv:2003.00330* (2020).

[18] Emily M Bender et al. "On the dangers of stochastic parrots: Can language models be too big?" In: *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 2021, pp. 610–623.

[19] Tan Zhi-Xuan. *Julog.jl*. Version 0.1.10. Aug. 2021. DOI: 10.5281/zenodo.1234. URL: https://github.com/ztangent/Julog.jl.

[20] George Datseris, Ali R. Vahdati, and Timothy C. DuBois. "Agents.jl: a performant and feature-full agent-based modeling software of minimal code complexity". In: *SIMULATION* 0.0 (Jan. 2022), p. 003754972110688. DOI: 10.1177/00375497211068820. URL: https://doi.org/10.1177/00375497211068820.

[21] Md Kamruzzaman Sarker et al. "Neuro-symbolic artificial intelligence: Current trends". In: *Ai Communications* 34.3 (2022), pp. 197–209.

[22] Zexun Jiang et al. "Casevo: A cognitive agents and social evolution simulator". In: *arXiv preprint arXiv:2412.19498* (2024).

[23] Niklas Heer. 2025. URL: https://niklas-heer.github.io/speed-comparison/.

[24] Louise Dupuis de Tarlé et al. "Argumentative Agent-Based Models". In: (2025).