Intrinsic Exploration-Motivation in Cultural Knowledge Evolution

Master Thesis Proposal

Student: Anaïs Siebers
E-mail: anais.siebers@rub.de
Matriculation number: 108 021 24790 9

Date: 20^{th} March 2023
Degree program: Cognitive Science, M.Sc., Ruhr-Universität Bochum
1^{st} supervisor: Christian Straßer, *Ruhr-Universität Bochum, Germany*
2^{nd} supervisor: Jérôme Euzenat, *Univ. Grenoble Alpes & INRIA, Grenoble, France*

Abstract

In cultural knowledge evolution simulated by multi-agent simulations, agents can improve the accuracy of their knowledge by interacting with other agents and adapting accordingly. But their knowledge might be confined to specific areas because they do not have the capacity to explore the world on their own. Since intrinsic motivation in artificial agents has already proven to increase exploration, I want to research in this master thesis, whether and how agents in simulations of cultural knowledge evolution can be intrinsically motivated to explore, and in how far this improves and changes their knowledge. This research builds upon previous research completed within the context of the mOeX project (Laboratoire d'Informatique de Grenoble, INRIA & Univ. Grenoble Alpes).
Keywords: cultural knowledge evolution; intrinsic motivation; exploration; artificial curiosity; computational creativity; social motivation; multi-agent simulation

Contents

1 Objective & Motivation

In multi-agent simulations of cultural knowledge evolution, it has been shown that agents can improve the accuracy of their knowledge by interacting with other agents. Agents converge towards successful communication, which means that interaction failures between agents are reduced and they have objectively more correct knowledge. Nevertheless, there still is knowledge diversity. But, if the agents are unable to explore the world, their knowledge might be confined to specific areas[1].

In order to address this problem, the agents could be provided with intrinsic motivation. There was a trend in the AI community in recent years to include curiosity mechanisms (as intrinsic motivation to explore), strongly enhancing the learning performance (Sun et al., 2022, 1). Intrinsic motivation is a very efficient method because it allows for the selection of experiences, leverages potential synergies among abilities as well as it allows acquiring macro-action (Oudeyer et al., 2016, 272f.). Thus, intrinsic motivation is a promising approach to enhance agents' exploration. The objective of this thesis is to research whether it is possible to simulate agents with intrinsic motivation to explore within experimental cultural knowledge evolution, and whether this improves their knowledge. Within this objective, the questions of how intrinsic motivation can be modelled, what kind of effect this has on the agents' knowledge and what kind of dynamics there are between intrinsically motivated agents and their / the population's knowledge, are researched and addressed.

1.1 Background & Rationale

This master thesis builds upon the research performed by the mOeX group (Laboratoire d'Informatique de Grenoble, INRIA & Univ. Grenoble Alpes). The aim of the research group is to "understand and develop general mechanisms by which a society evolves its knowledge" [2]. The group addresses questions such as (1) how populations with different knowledge (representations) can communicate, (2) how their representations are shaped by interaction with their environment and other agents, and (3) how knowledge diversity can be preserved and if this is beneficial (see Bourahla et al., 2021; Euzenat, 2017; mOeX, 2022). Although the research currently is in the area of basic research, in future, it may have an impact on fields as internet of things, social robotics or smart cities, such as fields where autonomous agents' knowledge is required to adapt to a dynamic environment (mOeX, 2021, 4). The thesis contributes especially with regard to research question two, as intrinsically motivated agents will change how the agents interact with one another as well as the environment.

The above mentioned questions are investigated by using experimental cultural evolution, adapted to knowledge representations of agents. Throughout the experiments, agents correct their own ontology of the environment to be able to interact

[1] *Intrinsic exploration motivation in cultural knowledge evolution*, Master topic / Sujet de master recherche, `https://moex.inria.fr/training/2022-M2R-motiv.html`, last access: 16th January 2023, 9:33am.

[2] See footnote 1.

with other agents. After multiple repetitions, agents reach successful communication and have improved knowledge in terms of correctness (Bourahla et al., 2021, 1). Agents interact by performing tasks together, which help them to develop useful knowledge for the tasks. However, they are not proactive and neither choose their interaction partners nor the interaction object from the environment nor the task (game) they play. They are assigned randomly to the interaction partner, object and task. Hence, exploring unknown parts of the situation space is likely to improve agent knowledge[3]. Nonetheless, agents should still survive in society and this should also remain the primary goal.

2 Research Context

After shortly outlining the objective and background of the master thesis, the research context will now be outlined in more detail. Therefore, three aspects from the thesis' objective are shortly discussed:

1. Because experimental cultural knowledge evolution uses multi-agent simulations, the first section will deal with the questions: What are agent-based models and multi-agent simulations? Where do they come from? How do they work and what benefits do they offer?
2. Following multi-agent simulations, it will be discussed what cultural knowledge evolution is and what it is used for.
3. Lastly, an overview of intrinsic motivation is given. Starting by a psychological definition of intrinsic motivation and its differences to extrinsic motivation, current computational approaches for intrinsic motivation in artificial agents are given, focussing on curiosity, creativity and social influence.

2.1 Agent-Based Models & Multi-Agent Simulations

Agent-based models (ABM) or *multi-agent simulations* (MAS) (sometimes also called individual-based models (IBM))[4] are used to model individuals or populations of autonomous decision-making entities called agents. Each agent individually acts according to its situation, which can lead to emerging patters across all agents, providing information about the dynamics of the simulated real-world system (Bonabeau, 2002, 7280; Klein et al., 2018, 7). Agents can range from humans over animals up to robots and software agents or services / daemons (Niazi and Hussain, 2011, 480). Models such as ABMs and IBMs are created and defined to study and explain 'observed phenomena as well as foresee future phenomena" (Abar et al., 2017, 14). To summarise, they are abstract representations of phenomena. Simulations like MAS on

[3] See footnote 1.

[4] The terms ABM, MAS and IBM are often used interchangeably with a slightly different focus. Because the in the given context, population-level phenomena are of interest, the term multi-agent simulation will be used to refer to all three.

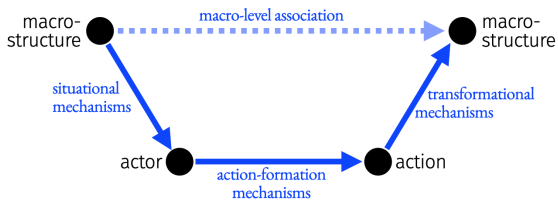

Figure 1: Colemans Bathtub showing the interplay of micro-level behaviour and macro-level structure (Klein et al., 2018, 11; `https://bit.ly/3ZE7oUJ`, last access 9^{th} January 2023, 12:41 pm).

the other hand, simulate these phenomena using a model as basis. These techniques are widely used in ecology, but also in many other disciplines dealing with complex systems like social sciences, economics, political science, philosophy, and computer science. (Grimm et al., 2006, 116).

A MAS consists of mainly three aspects (Bryson, 2014, 174): First, an *environment* in which the agents are situated, which exerts selective pressure and determines possible actions. Second, *parameters* (also known as attributes) of the agents, which make them individual and describe the "agent's character". Lastly, there is the *behaviour* of the agent and its *decision-making process.* By manipulating one aspect at a time, the effects can be observed in a reproducible controlled setting, contrary to simply observing the real-world phenomena, that are investigated by the model (Gabora and Tseng, 2017, 404).

Using MAS has various advantages. They "allow for numerical solutions of mathematical descriptions of social systems that are not tractable employing classical means". This is especially relevant for models with many heterogeneous agents, which have complex interactions and interact over a long period of time (Klein et al., 2018, 8). Moreover, they can bridge the previously mentioned *micro-macro gap*: allowing to observe emergent social patterns, which are not straightforwardly related to individual agent's behaviour. Varying individual parameters additionally allows for a more detailed understanding of the effect changes at the micro level have on the macro level (Acerbi et al., 2023, 12; Klein et al., 2018, 8), see Coleman's Bathtub in Figure 1. Furthermore, it is possible to observe individual runs of this non-deterministic process (Klein et al., 2018, 8). Lastly, simulations as in MAS are more flexible than analytical models (Acerbi et al., 2023, 12).

2.2 Cultural Knowledge Evolution

The thesis will deal with MAS in the context of *cultural knowledge evolution.* In general, MAS are used to study culture from many perspectives.

Interdisciplinary research in the field of cultural evolutionary theory started with the aim to understand cultural diversity and change (Acerbi et al., 2023, 9; Creanza et al., 2017, 7782). It applies Darwin's evolution theory to culture to study aspects of social life as customs, languages[5], ideas, behaviours, values, skills, knowledge, beliefs and other artefacts, which can change over time and are transmitted between individuals (Creanza et al., 2017, 7782; Mesoudi et al., 2006, 331). Darwin himself remarked: “The formation of different languages and of distinct species and the proofs that both have been developed through a gradual process are curiously parallel” (Creanza et al., 2017, 7782).

The three basic concepts of evolutionary theory: *inheritance*, *selection* and *variation* are transferred to cultural evolution (Acerbi et al., 2023, 9). But there are differences between genetic evolution and cultural evolution, as for example the characteristics of inheritance (horizontal vs. horizontal and vertical transmission), which is why cultural evolution is regarded as more complex (Creanza et al., 2017, 7783). Nevertheless, objections, that the complexity of social / cultural life cannot be modelled, are disputable because of the success of biologists using simplified models of complex biological systems (Mesoudi et al., 2006, 330).

Applied to the domain of knowledge, cultural evolution can help to understand the dynamic, temporal and evolutionary changes of relationships between entities in knowledge, investigating knowledge evolution (Trivedi et al., 2017, 2). Nowadays, there are computational models which are called models of cultural evolution (Gabora and Tseng, 2017, 404). In computer science, cultural evolution experiments are executed by implementing MAS using a precisely defined protocol. A population / society of agents repeatedly and randomly carries out a task (also called game) and adapts its culture in a monitored experiment (mOeX, 2021, 4).

With a slightly more specific focus, studying social networks with the help of agent-based models became a subject in philosophy of science recently (Šešelja, 2022, 1). It has its origins in economics, sociology and organisational sciences. In this context, science is regarded as a cultural practice, changing in time due to social learning and cultural selection (Wu et al., 2022, 1). Philosophers use the formal models as argumentative resource as they study the impact social networks, among other things, have on the acquisition of knowledge in different contexts (Aydinonat et al., 2021, 369; Šešelja, 2022, 1).

2.3 Intrinsic Motivation (in Artificial Agents)

In psychology, *intrinsic motivation* is – according to Deci & Ryan – the desire to perform an activity for its inherent satisfaction and pleasure (Ryan and Deci, 2000, 56). It is opposed to *extrinsic motivation*, which is the desire to perform an activity for an instrumental value. Hence, “in order to attain some separable positive outcome” (Ryan and Deci, 2000, 60). Both forms of motivation can coexist, but are different motivational dimensions (Kuvaas et al., 2017, 246). Two forms of intrinsic motivation are distinguished: *homeostatic motivational systems*, which push an organism / agent to maintain a stable state or comfort zone, and *heterostatic motivational systems*, which push an organism / agent from its habitual state (Oudeyer, 2007, 4). One well-

[5] See footnote 1.

known concept of intrinsic motivation is *flow*, coined by Csikszentmihalyi in 1975. It describes a state of optimal activation in which an agent is completely absorbed in an activity (Demontrond and Gaudreau, 2008, 10).

Although inspired by psychological research, intrinsic motivation is understood slightly differently in AI research because the current psychological definitions are "too complex and imprecise"(Oudeyer, 2007, 13). They do not yet allow for a computational realisation and implementation of intrinsic motivation. In AI research, intrinsic motivation is used to develop "autonomous continual learning agents", which aim to explore and learn in a changing dynamic environment without external rewards (Colas et al., 2019, 1) and whose behaviour is guided by internal forces (Georgeon and Ritter, 2012, 73). The decision of the agent has to be intrinsically derived rather than externally imposed[6]. One realisation of intrinsic motivation in AI research is, for example, the "pleasure of learning", modelled by measuring the improvement of the agent's internal model of the world, every time the agent applied a learning algorithm (Schmidhuber, 2010, 1).

Since the agents are simulated, their intrinsic motivation is given by the programmer / designer, and thus it is questionable whether it is genuine intrinsic motivation[7] (Guckelsberger et al., 2017, 4). Thus, intrinsic motivation refers to the behaviour the agent shows within the simulation, neglecting the implementation of the agent.

2.3.1 Artificial Curiosity

The most discussed approach to intrinsic motivation is curiosity (Jaques et al., 2018, 1). Sometimes *artificial curiosity* (AC) is even equated to intrinsic motivation (Ribeiro et al., 2017, 1). In psychology, curiosity is considered one aspect of the Big-5-personality traits, describing the "openness to experience", and the general desire to know or experience more, as basis for motivated actions to acquire new knowledge (Schutte and Malouff, 2020, 941). Curiosity is key for active learning and exploration (Oudeyer et al., 2016, 255), striving to reduce uncertainty and ignorance (Ribeiro et al., 2017, 1). Accounts of curiosity also strongly focus on novelty and complexity as well as confidence (Brändle et al., 2020, 1). Loewenstein proposed an *"information-gap"-hypothesis* of curiosity (Di Domenico and Ryan, 2017, 5), meaning, there is a strong link between the existing knowledge about the world and curiosity, since one is curious if one is interested in something that one does not know yet (Schmidhuber, 1991, 4).

Current approaches to model intrinsic motivation stem from reinforcement learning. There are different reward functions which represent different forms of motivation, such as, for example, *uncertainty motivation* (UM), which is the motivation to seek novelty, *information gain motivation* (IGM), which is the motivation to decrease

[6] *Intrinsic motivation (artificial intelligence)*, `https://en.wikipedia.org/wiki/Intrinsic_motivation_(artificial_intelligence)`, last access: 13^{th} January 2023, 5:36 pm.

[7] Within the model, external factors such as the programmer are irrelevant and neglected. Furthermore, when talking about intrinsically motivated agents, I refer to agents implemented with a model of intrinsic motivation.

uncertainty in the agent's own knowledge, *predictive familiarity motivation* (FM), which uses a predictive model to implement a motivation for familiar situation, *maximizing incompetence motivation* (IM), where the agent sets its own goals to improve where its performance is weakest (Oudeyer, 2007, 2-10), *maximizing competence process* (also referred to as flow motivation) (CPM), which aims at an optimal task, maximising reward, if the task is neither too easy nor to difficult (Oudeyer et al., 2007, 266).

2.3.2 Computational Creativity

Computational creativity (CC) describes agents' capability of generating creative products, which are described as surprising (Wu and Miao, 2013, 20). It can be considered an extension of curiosity, since curiosity is considered to be the impulse for creativity (Schutte and Malouff, 2020, 941). An essential element of creativity is that the artefacts or behaviours did not previously exist (Gabora and Tseng, 2017, 404). Thus, they have to be novel, but at the same time valuable (Guckelsberger et al., 2017, 5; Schutte and Malouff, 2020, 940). In the systems view of creativity by Csikszentmihalyi, creativity includes interactions with the environment and interactions between agents among themselves (Saunders, 2011, 36).

2.3.3 Social Motivation of Artificial Agents

One other form of artificial intrinsic motivation is social influence or motivation in general. The intrinsic motivation of agents consists of "having a causal influence on other agents' actions" (Jaques et al., 2018, 1) or anticipating what the other agents in the environment learn (Foerster et al., 2018, 1). This form of motivation is the least researched and modelled in the context of MAS and cultural knowledge evolution presented here. The *cultural intelligence hypothesis* claims that social learning is more efficient than individual or asocial exploration (van Schaik and Burkart, 2011, 1009). It has been shown that cooperation and reciprocal altruism has been beneficial for society (Cowden, 2012; Foerster et al., 2018, 1).

3 Research Questions

After introducing the research context, the research questions will now be presented. As outlined in the introduction, in experimental cultural evolution agents converge to successful communication and their knowledge is objectively improved. But, the agents are not very active, and it is likely that they are stuck in local optima with knowledge only of specific areas. Thus, if agents could be motivated to explore, their knowledge will presumably improve. Nevertheless, depending on the ratio of intrinsically motivated agents or the model used to simulated these agents, there might also be trade-offs such as for example between creativity and knowledge correctness. Using intrinsic motivation in AI has already proven to increase efficiency, autonomy as well as exploration of agents and thus seems a promising approach to tackle the problem. To summarise, the overall research questions of the master thesis are (1) whether agents with intrinsic motivation to explore can be modelled in experimental cultural evolution and (2) whether this will improve their knowledge. This requires research on how intrinsic motivation can be modelled, what kind of effect this has on the agents' knowledge and what kind of dynamics there are between intrinsically motivated agents and their / the population's knowledge.

In order to address these questions, three concrete forms of intrinsic motivation will be investigated. At the moment, agents are randomly assigned to games, an interaction object and their interaction partner. By introducing intrinsic motivation, agents will be able to choose their interaction partners, the game they will play and / or the interaction object from the environment. To have a basis on which the agents can decide besides internal variables – like knowing how often the agent already interacted with a specific object, external variables will be introduced. Imaginable is, for example, a distance information between agents where proximity describes cultural knowledge proximity comparable to culture environments in the real world.

Research Question 3.1 (Curiosity) How does curiosity – considered as the intrinsic motivation of individuals to gain new knowledge and reduce uncertainty as well as surprise in their knowledge space – influence the behaviour and knowledge of the agents and impact knowledge evolution?

To research this question, the agents will be equipped with an attitude (innate bias), determining the degree of curiosity they have. Furthermore, they can choose between different decision-making processes (dynamic) like UM, IM or CPM. This will influence which object they choose and which decisions they take for this object, but they do not choose their interaction partner. Whether agents prefer to choose a game or an object, will be modelled using an attribute, like "playfulness". Curiosity is understood as the drive to explore and is non-random in the sense that curiosity as opposed to creativity has no random variations or mutations.

Research Question 3.2 (Creativity) How does creativity – considered as the intrinsic motivation to add in addition to mutate knowledge – influence the behaviour and knowledge of the agents and impact knowledge evolution?

In comparison to curiosity, creativity in the context of the thesis is random (but not without bounds) and includes the creation of "new knowledge" or hypotheses. This includes an additional step in which knowledge is added, reorganised or mutated after interacting with another agent or the environment, based on the knowledge gained after interacting with another agent or environment. The idea is that, by chance, these changes or additions might be more efficient.

Research Question 3.3 (Social Motivation [if time allows]) How does social motivation – considered as the intrinsic motivation to anticipate and influence the mutual information exchange between agents' actions – influence the behaviour and knowledge of the agents and impact knowledge evolution?

To answer this question, agents will decide with whom they want to interact by looking, for example, for those with whom they previously had most disagreements. They aim at influencing the other agents and learn from other agents, those with whom they had the most disagreements. Contrary to curiosity, the game and the object of interaction are not chosen by these agents. The choice of the interaction partner could be based on cultural proximity.

4 Research Methodology

4.1 Research & Tools

The research will be conducted by (1) theoretically researching and modelling MAS for intrinsically motivated agents in cultural knowledge evolution, (2) designing experiments to test these models and predictions (3) conducting experiments, and (4) analysing and deducing properties as well as operators, which shape knowledge evolution and representations (see 5 Plan of Work & Time Schedule). The experiments are performed with *LazyLavender* – a simulation environment for cultural knowledge evolution, including the running of randomised experiments, the interaction of agents with their environment, the adjustment of knowledge by them and their communication. LazyLavender provides detailed reports as well as data extracted at different states from the experiment (mOeX, 2021, 5). How it is planned to tackle the research questions, is described in the previous chapter (see 3 Research Questions).

4.2 Experimental Framework

The experimental framework follows the experimental frameworks from the mOeX group. The goal is to extend / repurpose the experiments by Bourahla et al. (2021). In the simulation, agents learn alignments between ontologies based on the games they play. They, then, compare their decisions with other agents and adapt their ontologies (mOeX, 2021, 7). There are different populations of agents (defined by the type of ontology they use) (Euzenat, 2019, 41). But there are as many ontologies as agents (Euzenat, 2017, 2).

The knowledge in the ontologies is expressed by limited description logic on the basis of a finite set of properties which are binary (Bourahla et al., 2022, 4). The environment is also a set of objects described by binary properties (Bourahla et al., 2021, 3). This knowledge constrained by logical coherence (internally) and the environment in addition to the communication with other (externally) (mOeX, 2022).

Agents are situated in the environment and encounter objects upon which they decide upon. Although they know the decisions they can take, and the properties of the objects, they do not know the correct decision (Bourahla et al., 2021, 3). After deciding, the agents receive a pay-off. In a next step, two agents interact and compare their decision. If they agree, they are successful (Bourahla et al., 2021, 3).

Each form of intrinsic motivation should be investigated on its own. But it is probable and conceivable that different ratios of curios, creative and social agents will have different effects on knowledge evolution. Therefore, populations will also be compared with respect to the ratio of intrinsically motivated agents:

Imitators and Less Intrinsically Motivated Agents
How do different ratios of intrinsically motivated agents influence the behaviour and knowledge of the agents and impact knowledge evolution?

4.2.1 Measures

Mainly three measures, which have previously been used by the mOeX group (Bourahla et al., 2021, 4f.), will be used:

- The *accuracy* of the knowledge of each agent, which describes how accurate the agent's ontology is in comparison to all objects of the environment.
- The *interaction success rate*, which measures how often agents have agreed.
- The *distance* between ontologies of different agents.

These measures are repeatedly taken at various stages of the experiments, using the average accuracy as well as average distance between each pair of agents and the success rate.

New measures might also be introduced to be able to model intrinsic motivation and measure the impact of the motivational model.

4.3 Limitations

The most limiting factor is time. Therefore, it might not be possible to address all three research questions proposed. It is planned to start modelling and performing experiments with curiosity, and continue with creativity and lastly social motivation, if time allows.

Another limitation is that agents should not commit "suicide", but should explore in addition to surviving. Therefore, the exploration has to be balanced[8].

Concerning the methodology chosen: performing experiments with MAS has known limitations. The models can be characterised as "toy models" which do not necessarily allow for conclusions about the real-world (Frey and Šešelja, 2020, 1412). Additionally, any MAS is susceptible to changes in parameter values and in the idealised assumptions. These can be met by making a sensitivity analysis or derivational robustness analysis (Šešelja, 2022, 11). If time allows, these analyses will be performed. Nevertheless, by rerunning the experiment, saving different experimental states and starting with different ratio of agents, reproducibility and some derivational robustness is aimed at.

[8] See footnote 1.

5 Plan of Work & Time Schedule

5.1 Unfolding of the Work

The work will unfold in three steps. For each research motivational aspect (curiosity, creativity and – if time allows – social motivation), all three steps will be run through. Additionally, depending on the outcomes and conclusions drawn after the experiments, the models might be revised and the cycle repeated.

1 **Development of Motivational Schemes** The first step is to develop precise motivational schemes for agents. As already indicated in this proposal, the current idea is to develop three motivational schemes: curiosity, creativity and optionally social motivation. At the moment, they are only roughly outlined and distinguished.

3 Experiment design 4 Imple.

Conception of Mechanisms for Intrinsically Motivated Agents in Cultural Knowledge Evolution In the next step, the theoretical schemes have to be transferred to the concrete experimental setting and mechanisms of how to implement the intrinsically motivated agents have to be conceived. This will include a formal description of the motivational schemes in the context of cultural knowledge evolution.

2 **Development of Hypotheses** This step is concerned with predictions of and hypotheses about the emergent patterns that the changes on the micro-level (changing the agent's motivation) will have on the macro-level.

Conduction Experiments After developing the hypotheses, they have to be tested with experiments with different parameters and settings.

Analysis & Drawing Conclusions The results retrieved from the experiments have to be analysed afterwards. This will allow to draw conclusions and (maybe) revise the initial model.

5.2 Temporary Structure

The thesis should start with a theoretical part about cultural knowledge evolution as well as intrinsic motivation in artificial agents. Afterwards, the models should be developed on a theoretical level. The next part should deal with the experimental framework and its formal description including the measures. In this context, the formal description of the proposed models for curiosity, creativity and social motivation will be introduced. In the following section, hypothesis about the effect of the model will be introduced. Then, results of the experiments will be outlined. These results will also undergo robustness and maybe sensitivity analyses. Lastly, the results are related to the hypothesis and conclusions will be drawn.

5.3 Prerequisites

One prerequisite for the master thesis is to get to know Lazy Lavender to be able to implement the experiments. Additionally, the thesis requires knowledge in statistical analysis as well as programming.

References

Abar, S., Theodoropoulos, G. K., Lemarinier, P., and O'Hare, G. M. (2017). Agent Based Modelling and Simulation tools: A review of the state-of-art software. *Computer Science Review*, 24:13–33.

Acerbi, A., Mesoudi, A., and Smolla, M. (2023). *Individual-based models of cultural evolution: a step-by-step guide using R*. Routledge, Abingdon, Oxon ; New York, NY.

Aydinonat, N. E., Reijula, S., and Ylikoski, P. (2021). Argumentative landscapes: the function of models in social epistemology. *Synthese*, 199(1-2):369–395.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(suppl_3):7280–7287.

Bourahla, Y., Atencia, M., and Euzenat, J. (2021). Knowledge Improvement and Diversity under Interaction-Driven Adaptation of Learned Ontologies. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021)*, Online.

Bourahla, Y., David, J., Euzenat, J., and Naciri, M. (2022). Measuring and controlling knowledge diversity. In *The Eighth Joint Ontology Workshops (JOWO'22)*, Jönköping University, Sweden. CEUR Workshop Proceedings.

Bryson, J. J. (2014). The Role of Stability in Cultural Evolution: Innovation and Conformity in Implicit Knowledge Discovery. In Dignum, V. and Dignum, F., editors, *Perspectives on Culture and Agent-based Simulations*, pages 169–187. Springer International Publishing, Cham.

Brändle, F., Wu, C. M., and Schulz, E. (2020). What are we curious about? *Trends in Cognitive Sciences*, xx(xx):1–2.

Colas, C., Fournier, P., Sigaud, O., Chetouani, M., and Oudeyer, P.-Y. (2019). CURIOUS: Intrinsically Motivated Modular Multi-Goal Reinforcement Learning. arXiv:1810.06284 [cs].

Cowden, C. C. (2012). Game Theory, Evolutionary Stable Strategies and the Evolution of Biological Interactions. *ature Education Knowledge*, 3(10).

Creanza, N., Kolodny, O., and Feldman, M. W. (2017). Cultural evolutionary theory: How culture evolves and why it matters. *Proceedings of the National Academy of Sciences*, 114(30):7782–7789.

Demontrond, P. and Gaudreau, P. (2008). Le concept de ≪ flow ≫ ou ≪ état psychologique optimal ≫ : état de la question appliquée au sport:. *Staps*, 79(1):9–21.

Di Domenico, S. I. and Ryan, R. M. (2017). The Emerging Neuroscience of Intrinsic Motivation: A New Frontier in Self-Determination Research. *Frontiers in Human Neuroscience*, 11.

Euzenat, J. (2017). Interaction-based ontology alignment repair with expansion and relaxation. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pages 185–191, Melbourne, Australia. International Joint Conferences on Artificial Intelligence Organization.

Euzenat, J. (2019). mOeX : Évolution de la connaissance. *Bulletin de l'Association française pour l'Intelligence Artificielle*, 105:39–42.

Foerster, J. N., Chen, R. Y., Al-Shedivat, M., Whiteson, S., Abbeel, P., and Mordatch, I. (2018). Learning with Opponent-Learning Awareness. arXiv:1709.04326 [cs].

Frey, D. and Šešelja, D. (2020). Robustness and Idealizations in Agent-Based Models of Scientific Interaction. *The British Journal for the Philosophy of Science*, 71(4):1411–1437.

Gabora, L. and Tseng, S. (2017). The social benefits of balancing creativity and imitation: Evidence from an agent-based model. *Psychology of Aesthetics, Creativity, and the Arts*, 11(4):403–419.

Georgeon, O. L. and Ritter, F. E. (2012). An intrinsically-motivated schema mechanism to model and simulate emergent cognition. *Cognitive Systems Research*, 16:73–92.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S. K., Huse, G., Huth, A., Jepsen, J. U., Jørgensen, C., Mooij, W. M., Müller, B., Pe'er, G., Piou, C., Railsback, S. F., Robbins, A. M., Robbins, M. M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R. A., Vabø, R., Visser, U., and DeAngelis, D. L. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198(1-2):115–126.

Guckelsberger, C., Salge, C., and Colton, S. (2017). Addressing the "Why?" in Computational Creativity: A Non-Anthropocentric, Minimal Model of Intentional Creative Agency. In *Proc. 8th Int. Conf. Computational Creativity.*

Jaques, N., Lazaridou, A., Hughes, E., Gulcehre, C., Ortega, P. A., Strouse, D., Leibo, J. Z., and de Freitas, N. (2018). Social Influence as Intrinsic Motivation for Multi-Agent Deep Reinforcement Learning. Publisher: arXiv Version Number: 4.

Klein, D., Marx, J., and Fischbach, K. (2018). Agentenbasierte Modelle in den Sozialwissenschaften, Geschichte und PhilosophieAgent-Based Modeling in Social Science, History, and Philosophy. An Introduction. *<p>Historical Social Research / Historische Sozialforschung Vol. 43*, 1:Volumes per year: 1</p>. Publisher: GESIS - Leibniz-Institut für Sozialwissenschaften Version Number: 1.

Kuvaas, B., Buch, R., Weibel, A., Dysvik, A., and Nerstad, C. G. (2017). Do intrinsic and extrinsic motivation relate differently to employee outcomes? *Journal of Economic Psychology*, 61:244–258.

Mesoudi, A., Whiten, A., and Laland, K. N. (2006). Towards a unified science of cultural evolution. *Behavioral and Brain Sciences*, 29(4):329–347.

mOeX (2022). Research.

mOeX, P.-T. (2021). 2021 Activity Report, Project-Team MOEX, Evolving Knowledge. ACTIVITY REPORT, Inria Research Centre Grenoble-Rhône-Alpes, In Partnershiip with: Université de Grenoble Alpes, In Collaboration with: Laboratoire d'Informatique de Grenoble, Grenoble.

Niazi, M. and Hussain, A. (2011). Agent-based computing from multi-agent systems to agent-based models: a visual survey. *Scientometrics*, 89(2):479–499.

Oudeyer, P.-Y. (2007). What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics*, 1.

Oudeyer, P.-Y., Gottlieb, J., and Lopes, M. (2016). Intrinsic motivation, curiosity and learning: theory and applications in educational technologies. *Prog Brain Res - Progress in brain research*, 229:257–284.

Oudeyer, P.-Y., Kaplan, F., and Hafner, V. V. (2007). Intrinsic Motivation Systems for Autonomous Mental Development. *IEEE Transactions on Evolutionary Computation*, 11(2):265–286.

Ribeiro, E., Ribeiro, R., and de Matos, D. M. (2017). A curiosity model for artificial agents. In *Proceedings of the 14th European Conference on Artificial Life ECAL 2017*, pages 561–568, Lyon, France. MIT Press.

Ryan, R. M. and Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25(1):54–67.

Saunders, R. (2011). Artificial Creative Systems and the Evolution of Language. In *Proceedings of the 2nd International Conference on Computational Creativity*.

Schmidhuber, J. (1991). A Possibility for Implementing Curiosity and Boredom in Model-Building Neural Controllers. In Meyer, J. A. and Wilson, S. W., editors, *From Animals to Animats*, Proc. of the International Conference on Simulation of Adaptive Behaviour, pages 222–227. MIT Press/Bradford Books.

Schmidhuber, J. (2010). Formal Theory of Creativity, Fun, and Intrinsic Motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, 2(3):230–247.

Schutte, N. S. and Malouff, J. M. (2020). A Meta-Analysis of the Relationship between Curiosity and Creativity. *The Journal of Creative Behavior*, 54(4):940–947.

Sun, C., Qian, H., and Miao, C. (2022). From Psychological Curiosity to Artificial Curiosity: Curiosity-Driven Learning in Artificial Intelligence Tasks. arXiv:2201.08300 [cs].

Trivedi, R., Dai, H., Wang, Y., and Song, L. (2017). Know-Evolve: Deep Temporal Reasoning for Dynamic Knowledge Graphs. Publisher: arXiv Version Number: 3.

van Schaik, C. P. and Burkart, J. M. (2011). Social learning and evolution: the cultural intelligence hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567):1008–1016.

Wu, J., O'Connor, C., and Smaldino, P. E. (2022). The Cultural Evolution of Science. preprint, MetaArXiv.

Wu, Q. and Miao, C. (2013). Curiosity: From psychology to computation. *ACM Computing Surveys*, 46(2):1–26.

Šešelja, D. (2022). Agent-based models of scientific interaction. *Philosophy Compass*, 17(7).

Competing Interests

This master thesis proposal was written in the context of the study program "Cognitive Science" at the Ruhr-University Bochum. The paper is rewarded with credit points and evaluated with a grade.