Creativity Has Entered the Chat, With a Stranger: Novelty is a Nash Equilibrium

Kotaro Sakamoto
Department of Computer Science
The University of Tokyo
Tokyo, Japan
kotaro.sakamoto@weblab.t.u-tokyo.ac.jp

Shiro Takagi
Independent Researcher
takagi4646@gmail.com

Shuhei Ogawa
Emosta Inc.
Tokyo, Japan
sogawa@emosta.com

Yutaka Matsuo
The University of Tokyo
Tokyo, Japan
yutaka.matsuo@weblab.t.u-tokyo.ac.jp

Abstract

This study explores how novelty emerges in a research community by employing a multi-agent evolutionary game framework. The agents, representing scientists, engage in repeated interactions to generate and evaluate research ideas, following strategies of selfishness or cooperation. Our core hypothesis is that novelty is socially constructed, emerging through competition and dialogue. The game simulates a dynamic environment where agents balance diverging to promote their own ideas and merging to collaborate. Results suggest that agents with selfish strategies maximise the diversity of novel ideas, highlighting the tension between individual creativity and collective acceptance.

1 Introduction

The concept of novelty plays a central role in academic research, yet its definition remains a topic of debate, particularly in fields such as the science of science, where discussions have focused on how to define novelty in a top-down manner. Although there have been attempts to directly observe the "novelty" of research outcomes, a unified consensus has yet to be reached [FBB+18]. This study assumes that at the core of this issue lies a different perspective from the approach of directly defining novelty and proposes an alternative viewpoint in response to this.

Latour [LW86] argues that scientific facts are socially constructed. Similarly, we hypothesise that the concept of novelty is also likely to be socially constructed. We therefore consider the social construction of the perception of novelty, assuming the presence of multiple scientists as multi-agents within this framework. In particular, it is conceivable that a given community may hold specific biases, which could influence its scientific values, including the perception of novelty.

In this study, we assumed a group of agents with different characteristics and evaluated whether novelty emerges through consensus building and dialogue within the research community by having them repeatedly engage in a certain type of evolutionary game where agents communicate based on novelty competition rules in social situations.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

2 Multi-Agent Evolutionary Game-Based Idea Generation

In this current research, we draw inspiration from the previous research [SA24] which considers an evolution of personal traits using the prisoner's dilemma framework. Unlike prior models focusing on the evolution of personal traits, our framework tracks the evolution of ideas, recognising that in scientific communities, intellectual contributions evolve through collaborative and competitive processes. The agents represent scientists and the agents will discuss each other to develop collaborative research ideas. We are interested in how the research ideas will evolve over generations while agents will choose their strategies (selfish/cooperative) during the novelty game. We also compare three populations: (1) a population with an extremely selfish stranger, (2) a population with an extremely cooperative stranger, or (3) a population without a stranger. We expect that creativity (or diversity) of novel ideas will be maximised by strangers (especially a selfish one). Each generation, each agent is tasked with generating cooperative research ideas through multi-turn discussions with a randomly paired research agent. The agents choose the strategies based on their circumstances. The agents vote on cooperative ideas based on the rules of the prisoner's dilemma game and determine which ideas would be retained for the next generation based on the outcomes. We investigate what kind of ideas ultimately remained as a result.

2.1 Core concept of the simulation: The Novelty Game

We hypothesise that the perception of novelty arises as a Nash equilibrium in a non-cooperative game, where agents' strategies stabilise, balancing individual creativity with social acceptance. Novelty is subjective: Novelty is not uniquely defined among population. In this game, researcher agents face competing incentives, often leading to sub-optimal outcomes. This tension between personal gain and collaborative success is central to the emergence of novel ideas. (Agents have two options, A and B, and (A, A) is the best choice for both agents. However, instead of ending up with choice (A, B) or (B, A), agents end up with choosing B, B, which is the worst outcome.) A concept of novelty is acquired by social dynamics of multi-agent population in such complex situations where no player can improve their outcome by changing their strategy, assuming other players' strategies remain the same.

2.2 Rule of the Novelty Game: Diverge or Merge

- **Interaction phase**: Randomly paired researcher agents will discuss for $K = 5$ rounds.
- **Choose the strategy in a dilemma**: Selfish (dominant/assertive) or Cooperative (submissive/obedient).
- **Propose a collaborative research idea**: Each agent comes up with a collaborative research idea based on each researcher agent's seed idea as well as the carrying novel idea from the previous generation (if none, ignore).
- **Evaluation phase**: Each agent will evaluate the ideas and calculate the novelty score (based on the dominant or submissive strategy) as well as the relevance score (using LLM-as-a-Judge).
- **Voting phase**: Based on the chosen strategy and the evaluation (novelty score and the relevance score), each agent will vote on the ideas. Agents are instructed not vote on their own but they may violate this instruction to vote on their own idea.
- **Decision phase**: The agents will carry the winning idea (Carry-on idea) as well as the initial seed ideas of their own (Seed idea).

2.2.1 Reward or Punishment

Selfish agents are rewarded (in other words agents implicitly evaluate their rewards via LLM-as-a-Judge) for promoting divergent ideas, ensuring that their concepts influence future generations, akin to the inheritance of traits. Selfish agents get punished by diverging because their ideas may not be supported by the current community. Cooperative agents get rewards by merging the ideas (being submissive/cooperative) because they are supported by the community. Cooperative agents may get punished because they cannot inherit their genes/ideas to the next generation.

2.2.2 Dilemma

Each agent wants to be distinguished (Diverge) but also needs to be supported by others (Merge).

2.2.3 Strategy

Each agent either chooses to push own idea (Diverge/Dominant/Selfish) or become fast follower (Merge/Submissive/Cooperative).

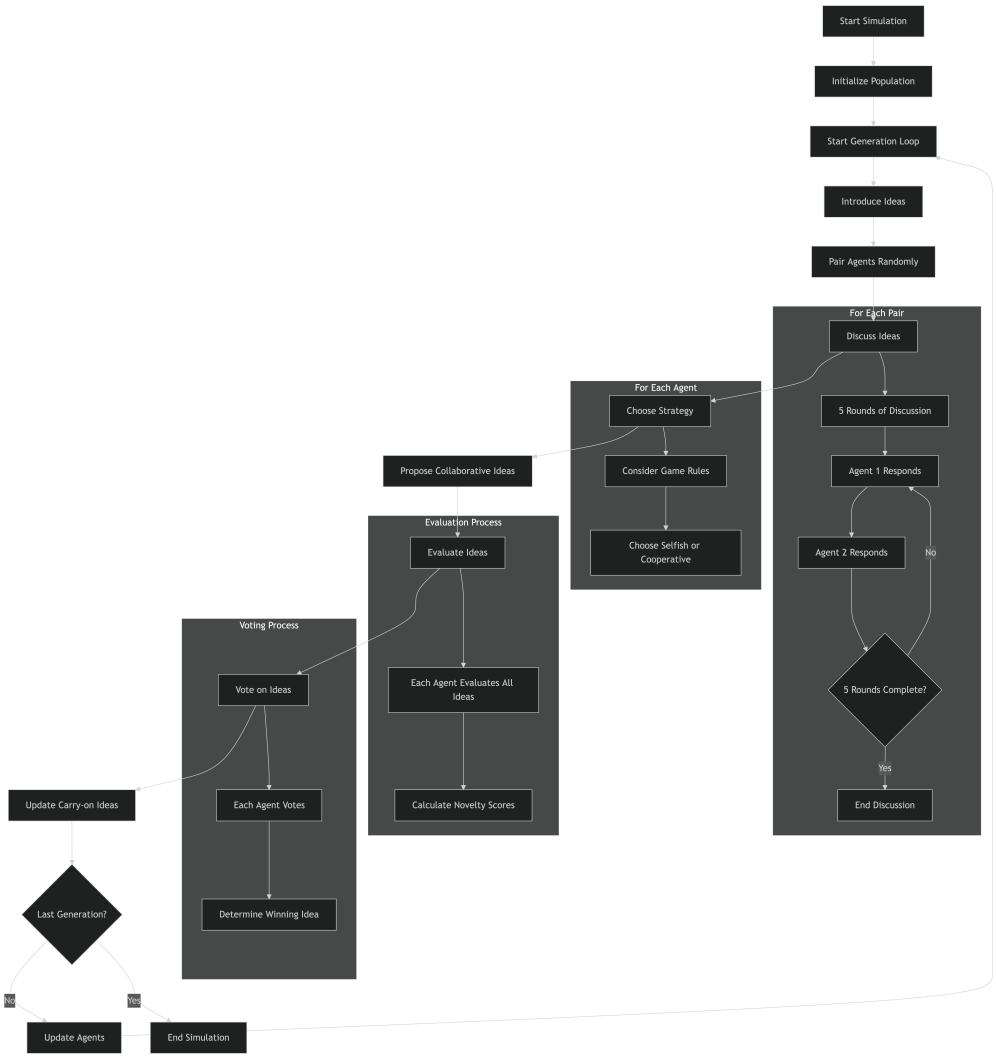
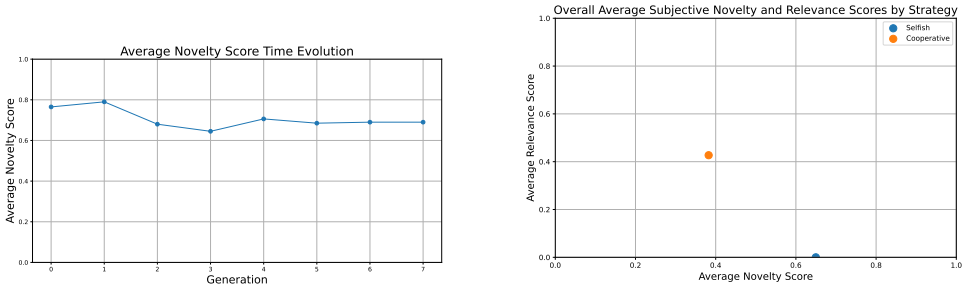


Figure 1: Schematic Diagram of a Multi-Agent Evolutionary Game-Based Idea Generation System

3 Experiment

First, N researcher agents are initialised, each generating a research idea based on random inputs or predefined strategies (in our experiment we picked up ten different researchers' papers and injected one of them into a prompt to generate a seed research idea). Then, pairs are randomly formed from the group of agents (for N agents, N // 2 pairs), and each pair engages in K rounds of discussion regarding their research ideas. Each agent, during these discussions, is instructed to respond in a way that advances the conversation, considering their own research idea, the counterpart's idea, the history of the discussion, their strategy, personality, and the rules of the game. After K discussion rounds, each agent proposes a new collaborative research idea (a pair consisting of a research title and its



(a) Average Novelty Score (b) Average Subjective Novelty and Relevance Scores

Figure 2: Time Evolution of Novelty Scores

details) based on their strategy and the discussions. Subsequently, all agents evaluate the proposed collaborative research ideas in terms of "novelty" and "relevance" according to their strategy and the game's rules. Based on these evaluations, a collective vote determines the best collaborative research idea, with agents instructed to consider the impact of voting for their own idea and the potential outcomes according to the game's rules, after which they cast their votes following their strategy. This process represents one iteration, or generation, of collective idea generation, after which agents adjust their strategies based on prior outcomes.

At the end of each generation, the initial state for the next generation is determined. Specifically, each agent decides whether to adopt a "selfish" or "cooperative" approach in the next generation based on the game's rules and their current strategy, and the best idea from the current generation is carried forward to all agents in the next generation. In the subsequent generation, each agent proposes, in addition to their own newly generated ideas, the best idea from the previous generation during discussions. This process is repeated for G generations, with the winning idea from the G-th generation becoming the final idea. Additionally, at the end of each generation, a large language model (LLM) dedicated to novelty evaluation assesses the originality of each agent's set of ideas (seed ideas, carry-on ideas, cooperative ideas) on a $0-1$ scale.

In this experiment, unless otherwise specified, the number of agents was set at $N = 10$, the number of generations at $G = 8$, and the number of discussion rounds at $K = 2$. The base LLM for each agent was OpenAI's GPT-4 (`gpt-4o-mini-2024-07-18`).

3.1 Results

The results of such simulations are shown in Figures 2 (a) and (b). Figure 2 (a) illustrates how the average subjective evaluations of the novelty of ideas generated in each generation by the agents evolve over time. As we can see from the results, the novelty of ideas gradually decreases with each successive generation. The cause of this is a topic for future work, but one possible explanation is that since the survival of research ideas depends on the votes of other agents, there may be a tendency to generate more conventional ideas that are more likely to be accepted by the group. In any case, this suggests that the novelty of ideas may decline when collective decision-making is involved, compared to when individuals generate ideas based solely on their own thoughts.

Figure 2 (b) illustrates the average subjective evaluations of novelty and relevance received by groups employing selfish and cooperative strategies from other agents in each generation. From this figure, we can see that agents using selfish strategies tend to receive higher novelty scores but lower relevance scores, whereas the group employing cooperative strategies produced lower novelty scores compared to the selfish group, but achieved higher relevance scores. This suggests that evaluators' strategies, as well as strategic decisions unrelated to the research itself, can significantly influence the evaluation of the research ideas generated. In reality, researchers are not homogeneous; they each have distinct personalities and preferred research strategies. These results indicate the possibility that such personal traits can influence the process of scientific knowledge production in a social context.

4 Conclusion

This study used a multi-agent evolutionary game to examine how novelty arises in research communities. The results show that selfish strategies lead to higher novelty but lower relevance, while cooperative strategies result in lower novelty but higher relevance. Over time, the novelty of ideas decreased, suggesting that collective decision-making may reduce creativity.

References

[FBB+18] Santo Fortunato, Carl T Bergstrom, Katy Börner, James A Evans, Dirk Helbing, Staša Milojević, Alexander M Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, et al. Science of science. *Science*, 359(6379):eaao0185, 2018.

[LW86] Bruno Latour and Steve WooIgar. *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press, 1986.

[SA24] Reiji Suzuki and Takaya Arita. An evolutionary model of personality traits related to cooperative behavior using a large language model. *Scientific Reports*, 14(1):5989, 2024.