Strategic Interactions between Large Language Models-based Agents in Beauty Contests

A Preprint

Siting Lu[*]

School of Economics, The University of Edinburgh

12 April 2024

Abstract

The growing adoption of large language models (LLMs) presents substantial potential for deeper understanding of human behaviours within game theory frameworks through simulations. Leveraging on the diverse pool of LLM types and addressing the gap in research on competitive games, this paper examines the strategic interactions among multiple types of LLM-based agents in a classical game of beauty contest. Drawing parallels to experiments involving human subjects, LLM-based agents are assessed similarly in terms of strategic levels. They demonstrate varying depth of reasoning that falls within a range of level-0 and 1, and demonstrate convergence in actions in repeated settings. Furthermore, I also explore how variation in group composition of agent types influence strategic behaviours, where I found higher proportion of fixed-strategy opponents enhances convergence for LLM-based agents, and having a mixed environment with agents of differing relative strategic levels accelerates convergence for all agents. There could also be higher average payoffs for the more intelligent agents, albeit at the expense of the less intelligent agents. These results not only provide insights into outcomes for simulated agents under specified scenarios, it also offer valuable implications for understanding strategic interactions between algorithms.

Keywords: Large Language Models, Beauty Contests, Strategic Interactions

1 Introduction

With the emergent line of research surrounding the study of the capabilities of large language models (LLMs), there is also growing discussions over the implications of LLMs on economic research and the possibility of using them for social sciences experiments. While the current applications of LLMs as simulated agents mainly concentrated on more text-based tasks, such as filling in surveys as a way to provide insights on the average opinion of the population regarding certain topics, it could also impact more strategic scenarios, particularly in the field of game theory. Recent research in this area, however, mainly focused on exploring 2-player cooperative and non-cooperative games, and often comprises of a single LLM type. It could be interesting to investigate strategic interactions in multi-player competitive games, where agents can be back-boned by different LLMs to represent players of heterogeneous types. As a result, in this work, I seek to leverage on the presence of multiple LLMs to explore their strategic behaviour in beauty contest games.

LLMs as simulated agents. One of the main objectives of this work is to make a case for the possibility of using LLMs as simulated agents, which still has limited applications in Economics. Since LLMs are trained based on human-generated data, observing game play for LLM-based agents could be fairly relatable to human subjects in experiments, thus their responses could provide results that offers more insights than the conventional simulation methods. Furthermore, it is relatively easy to use LLMs in experiments, making it a good tool to toy with different set-ups and to provide fresh insights on classical games. Even though studies did highlight LLMs' ability to mimic human behaviour and reasoning process, there remain questions and controversies of using LLMs as replacement for human subjects. In view of the potential limitations in interpreting the minds of LLM-based agents with the absence of similar cognitive architecture that backs human behaviour, it is important to note that using LLM-based agents as participants of experiments is not aiming at replacing human subjects, but rather simulating potential game play by riding on top of embedded human behaviour. It is also more cost-effective if we want to investigate some variations that generate interesting insights from LLM-based games in a more expensive human-driven experiment, thereby offering designs that are worth pursuing.

Behaviour of LLM-based Agents in Multi-player Competitive Game. The second objective of my work is to inform potential game play in a multi-player competitive game and study the adaptive learning behaviour of agents when the group composition differs. Literature that simulated agent behaviour with LLMs in games often consist of 2-player games and usually use a single LLM type (Horton (2023), Phelps and Russell (2023)). This is equivalent to assuming all agents have the same natural language understanding ability, modelling agent behaviour from which is therefore a little more restrictive and did not leverage on the potential of having a pool of LLMs. By having multiple types of LLMs, it is possible to draw relation of these models to human subjects. Using the beauty contest games, I found the models to have a strategic level between 0 and 1 using the level-k model proposed by Nagel (1995), and in repeated setting, some of them show convergence towards Nash equilibrium action, which could be at varying speed, as past information are revealed to them. Further to this, I have selected two different LLM types, characterized by their strategic levels, to explore how variation in proportion of agents types in the group could affect game outcome. In which case, I found less strategic agents could converge faster to Nash equilibrium than more strategic agents when facing fixed-strategy opponents. Furthermore, when they are playing against each other, their learning rate could be faster than when they are playing with their own types. Specifically, while less strategic agents do not learn when playing with own types, they do in mixed environment. However, in terms of payoffs, the more strategic agents gain comparably well or having better payoffs in mixed environment than pure environment, and this comes at the expense of less strategic agents. These contribute first to assessing the models with a human-based metric on the basis of strategic level, thereby drawing a relation to represent heterogeneous human subjects using the models , and then simulating game play and exploring learning behaviours. The variation in group composition also offers some insights on how to induce faster learning by changing proportion of opponent types faced by the agents.

Strategic Interactions between AI Algorithms. Last but not least, this work hopes to use methods that evaluate human subjects to shine a light on how algorithms could interact with one another in strategic setting. For instance, with the ongoing progress of integrating LLMs into daily life, it is likely that in the future, agents will be using bots back-boned by LLMs to communicate and interact with one another, making LLMs surrogate agents. A clear example would be in the trading market, where existing crypto trading bots function by executing pre-defined strategies, making buying and selling automatic (Trality (2024)), it is entirely plausible to replace such automatic bots to ones backed by LLMs that take into account vast human data on trading strategies. Therefore, understanding how algorithms react to one another would be important in such context. Furthermore, since large LLMs could be costly, it is possible to have strategic situations that encompass a variety of LLMs, and one can trade-off between cost of adopting different LLM-backed agents and their performance.

On a broader perspective, this work hope to highlight that LLM-based agents not only can be used as a tool for social sciences in simulating potential human behaviour in strategic situations, the theories that were developed to evaluate human behaviours can nonetheless help us to understand how this new era of computer algorithms function when competing with one another.

The rest of the paper is divided into a few sections: Section 2 introduces the background of this work. Section 3 explores LLM-based agents in beauty contest games, divided into games with multi-LLM-based agents to investigate their respective strategic level, as well as convergence behaviour in repeated setting, and reducing LLM-based agents into two types, to investigate changes in their behaviour given variations in group composition. This is followed by Section 4, where I study the understanding and line of reasoning behind the choices made by the two LLM types for the different set-ups. Future updates and extensions are then highlighted in Section 5. Following which, I conclude in Section 6.

2 Background

2.1 LLMs as Computational Model of Human Behaviour

While the creation of LLMs have sparkled a new wave of research in Computer Science to improve the models' performance, their development also convey huge potential for interdisciplinary studies. Horton (2023) for instance proposed that LLMs can be perceived as implicit computational model of human behaviour and be used as simulation tool. Before diving into further details, it is important to understand how LLMs could function as computational model of human behaviour as a result of how they are trained.

Training Process. As illustrated by Ouyang et al. (2022), there is a 3-step training process that involves human feedback:

1. *Training a supervised policy.* A prompt is sampled from a dataset, and human labelers were asked to provide responses to the prompt, these response data are then used to facilitate supervised learning of the policy, constituting the early version of the language model.

2. *Training a reward model.* For a set of prompt and responses that were generated, again a human labeler gets involved to rank the responses from best to worst, this ranking data is then used to train the reward model.

3. *Optimizing the policy against the reward model.* As the policy generates output and reward model computes the reward for the given output, the policy is fine-tuned against the reward model using reinforcement learning. This form of model alignment using reward model trained on response ranking from human annotators is referred to as reinforcement learning with human feedback (RLHF).

There is implicit human involvement in the prompt sample, which rides on top of information that is publicly available on the internet, as well as licensed information from third parties, and there is also explicit human involvement in the training process that includes information generated by the LLM labelers and users. (OpenAI (2024)) Particularly, in the reward model training stage, the training dataset used by Anthropic primarily comes from crowd-sourcing feedback through Amazon Mechanical Turk, a platform often used for social sciences research. As for OpenAI, their models are trained on used prompts submitted by users to the GPT API. Since login are more restricted in this case, training based on feedback from different groups of users could induce performance differences across LLMs. (HuggingFace (2022))

Given that LLMs can be interpreted as computational model of human behaviour based on the underlying training process that pans on human-generated data, I hope to streamline and differentiate between the two main aspects of how LLMs' human-like behaviour could cater to assisting and influencing research for Economics, or more broadly, the social sciences community:

(a) Imitation of human decision-making process given known constraints. LLMs allow for the creation of synthetic agents with given profiles, and their behaviour, mimicking human behaviour under known and specified constraints, are investigated. This type of simulation could help to facilitate building of decision theory models when the constraints are pre-defined. It may resembles agent-based modelling (ABM) approach, where agents are endowed with pre-defined behavioural prompt, and the outcome from which could serve as a form of visualization and checkpoint of the theoretical predictions. However, as illustrated by Horton (2023), agents in ABM approach is completely pre-programmed to behave as we expect, whereas homo silicus, describing LLMs, are not completely under our control. While it is possible to provide them with artificial profiles and demanding them to behave with a pre-defined constraint, the outcome may not be exactly what we predict, and this could arise if there are implicit bias embedded in the underlying training data that the theoretical models, which was put to test, did not account for. Therefore, any deviation in the simulation results away from theoretical predictions inevitably makes it more interesting and useful, where they can feed back into improving theoretical models to explore potential factors that cause the diversion.

(b) Mirroring human-like complexities without known constraints. It is also possible to run simulation without pre-defining the constraints, and this would resemble running an experiment with human subjects, except with LLM-based agents. By abstracting away from putting restrictions on their behaviours, the simulation results would illustrate the less computational side of LLMs, and essentially offers a tool for computational experiments, showcasing the potential experiment results if they were conducted with human subjects. Depending on the baseline data the LLMs are trained upon, the results could be more representative of the average opinion of the general population or specific groups of population.

While the first area spur emerging research in analyzing behaviours of artificial agents who were given pre-defined types or constraints, confirming and illustrating how certain known or hypothetical human bias could impact outcomes. The objective of which is more grounded in granulating the elements contributing to decision-making, which would be important in understanding why human behave the way they do. On the other hand, the second area takes advantage of the latent behavioural elements that were baked in the training data. Its main focus is to speculate the possible experimental results that can be anticipated if done with human subjects.

LLMs as heterogeneous agents. LLMs are of different priors and could also comes in varying sizes, they can therefore be considered as heterogeneous agents on the basis of differences in these characteristics. The main difference between large-scaled language models and small-scaled language models are the amount of data they are trained upon. Since large-scaled language models are trained on massive amount of data, often comprising billions or trillions of words, they are potentially better at natural language understanding and providing better text-based outputs. (Labellerr (2024))

While we can directly perceive the models to be different types of simulated agents, their variation in text-based generating ability might not translate to varying strategic ability. Therefore, in this paper, their heterogeneity is characterized by their strategic levels that are determined when playing in the one-shot beauty contest games. This is a measure ubiquitous to how we evaluate the types of human subjects, thereby drawing parallels between the two. The additional benefit of this work leveraging on the existence of different LLM types to represent heterogeneous agents is that it also build a case where smaller language models can also be useful in helping us to understand human behaviour, rather than just learning from the larger language models.

LLMs as complements to human participants. As the line of research emerge arguing for the usefulness of LLMs in social sciences, there is the question of whether LLMs can rise up to the task of participating in social experiments in place of human subjects or rational players. Firstly, current literature have explored if LLMs, mostly a single LLM type (often `GPT3.5` or `GPT4`), could imitate human behaviours by attempting to replicate experimental results conducted with human subjects. Argyle et al. (2023) found `GPT3.5` to reflect viewpoints that parallel to the US public opinions and sub-populations, and could behave as good proxies for aggregate level human cognition. There are also literature that illustrate LLMs' strong capability in analogical reasoning, comprehension and communication skills to solve problems, as well as producing moral judgements that are well-aligned with human subjects. (Webb et al. (2023), Huijzer and Hill (2023), Dillion et al. (2023)) In the context of strategic games, LLMs were found to be able to reproduce the findings in many types of experiments, such as the ultimatum game, wisdom of crowds. AI bots equipped with `GPT4` also exhibits behavioral and personality traits, in terms of risk-aversion and cooperation, as well as learning patterns, similar to that of many human subjects. (Guo (2023), Aher et al. (2023), Mei et al. (2024)) On top of modelling human behavioural features, there are also works that investigate how LLMs compare to rational players in games. Fan et al. (2023) indicates LLMs in games, such as dictator game, Rock-Paper-Scissors, and ring network, may find it hard to elicit uncommon preferences, refine belief, and may take sub-optimal actions because they have ignored or unnecessarily modified the refined belief. Guo et al. (2024) explores beauty contests and auction games and found LLMs to deviate away from Nash equilibrium action. While they may not serve the purpose of behaving as rational players in games, they inevitably shows ability to imitate human behaviours, as could serve as human-like simulated agents. Given their training process, responses from LLMs can be interpreted to ride on top of the human reasoning process, as well as their preferences and behavioural bias that are embedded into the training datasets. Therefore, it is likely that the eventual output from LLMs revealed the potential behaviour of human under each prompt, making LLM-based agent a plausible tool to approximate average human behaviour in experimental setting. Since training data rely less on strictly controlled environment and acquired from a potentially broader subject pool, the end results could be more representative of the population.

While Dillion et al. (2023) indicates LLM to be good proxy of a single participant, they question LLM's ability to capture variability in human cognition as the model condense the diversity of responses into a modal opinion. This, however, can be solved by having different types of LLMs, which could be trained on different subsets of information, even for same training information, the degree of noise and processing ability also matters. Furthermore, in strategic situations, it remains interesting to investigate how game play within a homogeneous population is like.

The main concern however, is that minds of LLMs are opaque, just like humans. While there are many theories that underline the behaviour of human in strategic situations, there is lack of something equivalent that strongly supports AI algorithms go through the same thinking process. However, since LLMs are trained using human-generated data, including reasoning procedures, they could develop mechanisms similar to that of human brain. (Kosinski (2023)) Despite this connection, it is important we treat the simulated results with caution. Therefore, my work does not aim at arguing for replacement of human subjects in experiments, but rather, in view of the parallels between strategic behaviours of LLMs and human subjects, we should take advantage of the tool to shed more insights on strategic behaviour of human subjects and their cognition process in specific setting. Further to this, instead of using it only when experiments are not feasible, it is a good and also cost-effective approach to gain fresh insights through variations in game set-ups when it is difficult to conduct the same experiment again, or simply as a way to find interesting experimental set-ups worth pursing with more expensive human subjects. This form of simulation could have implications for mechanism design, providing clarity on how variation in certain design could potentially impact game play.

2.2 LLM-based Agents in Competitive Games

While some literature have explored cooperation and anti-coordination games between agents back-boned by a single LLM type, in this paper, I seek to explore competitive games that encompass multiple LLM types. As a result, I focused specifically on beauty contest games, which provides a classical and desirable set-up that encompasses both (1) competitive nature, (2) interactions between multiple agents. The agents involved can also be heterogeneous, capturing impact of differences in natural language understanding capacities or rather, intelligence or strategic levels for simulated agents. Furthermore, it is possible to construct a simple set-up with a single interior Nash equilibrium solution, making it easy to distinguish the levels of reasoning using either the iterated elimination of weakly dominated strategies, level-k or cognitive hierarchy models. (Nagel (1995), Camerer et al. (2004)) Further to that, Akata et al. (2023) studied repeated 2×2 games, such as Battle of Sexes and iterated Prisoner's Dilemma, to inform coordination or anti-coordination behaviour of LLMs in repeated interactions, which has something to do with alignment with opponents. The same can be done for competitive games to explore LLMs' adaptive learning behaviour when they are in direct competition with others and have to outmaneuver the others in order to win. Therefore, the additional benefit of repeated beauty contests under the standard set-up is that even in repeated setting, there is a single Nash Equilibrium, thereby setting aside any possibility of multiple equilibrium which could complicate the analysis.

Nevertheless, there is also another layer of social value for exploring application of LLMs in the beauty contest games. The Keynesian Beauty Contest stems originally from Keynes (1936), and it has been used to describe the stock market, where investors attempt to guess what the other investors might do, and it is also related to beauty pageant competitions, where competitors were asked to select the prettiest contestant by forming expectation about other competitors' opinions. (Nagel et al. (2017)) It is likely that in the future, trading bots or algorithms in competitive setting will be back-boned by LLMs, and potentially behave as proxies for humans to execute tasks. A more superior algorithm in terms of its parameter size or amount of training data it encompasses, does not necessarily imply better payoffs as they are playing in conjunction with other types of algorithms, and have to interpret and best react to what might the other algorithms do. Most importantly, individuals could choose proxies that are backed by different LLMs depending on their cost concern, making it important to look at how different types of LLMs could react to one another. Therefore, it becomes interesting to investigate how human-like LLMs interact with one another in a simple competitive setting like the beauty contests to inform more about its potential social implications.

2.3 Strategic Interactions between Machines

Unlike previous generations of static computer algorithms that execute pre-programmed strategies, and behave regardless of the types of opponents they are faced with, LLMs are more human-like and dynamic in nature. In this work, as I simulate interactions between LLMs and static algorithm, as well as LLMs vs. LLMs, there can be many different interpretations. The results can be used to represent game play between synthetic human agents, human vs. computer, and computer vs. computer. In the first two instances, LLM-based agents are behaving as simulated agents that approximate human behaviour, providing some plausible conjectures about how human subjects might react in games when placed against other human subjects or against computer algorithms. While we use LLMs as simulated human agents to understand human behaviour, the reverse is true as well, through explaining the behaviour using concepts borrowed from human research, we can also understand machines better.

For centuries, economists, psychologists and neuroscientists have been studying the quasi-black box of human brain and came up with numerous theories that seek to explain the decision-making of human, they can be used to potentially explain machine's behaviour as well, or at least as a first step, drawing parallels between the two. For instance, Kosinski (2023) explores the application of theory of mind for LLMs and Dillion et al. (2023) also mentioned comparing LLMs with human judgements could perhaps teach us about the machine minds of LLMs. There are a lot more to be done for strategic interactions. By looking at the computational experiments and the human subjects' experiments, we can evaluate the models using the method we use to evaluate humans, and we can perhaps use how we explain human interactions to explain machine interactions, thereby generating more insights on how minds of machine might work in strategic settings. Understanding how they function is a step forward to improving their performance in the future to be more human-like or even, exceeding average human capacities.

3 Beauty Contest Games

The experiments in this section is based on the `EconArena` proposed by (Guo et al. (2024)). Different from Guo et al. (2024), where the main objective is to evaluate LLMs' performance relative to rational players that play Nash equilibrium following game-theoretical predictions, this work aims to analyze LLMs' behaviour as though they are human players, where their strategic levels are assessed using the methods employed to assess human subjects, and their potential game play is simulated under varied set-ups.

In this section, I first explore one-shot and repeated beauty contests games involving multiple LLM types, which include: `ChatGLM2`, `ChatGLM3`, `Llama2`, `Baichuan2`, `Claude1`, `Claude2`, `PaLM`, `GPT3.5`, `GPT4` After evaluating their strategic levels and convergence behaviours, I then choose two types of LLMs, namely `PaLM` and `GPT3.5`, which were assessed to be of different strategic levels, to construct groups of heterogeneous agents, and analyze how variations in group composition could affect game outcomes, providing some insights on their adaptive learning pattern.

3.1 Multi-LLM-based Agents

The basic set-up is a modified version following Nagel (1995) and the original exemplary prompt is recited in A.1.

- The agents are asked to choose a number between 0 and $\bar{c}$, where $\bar{c} \in (0, 1000)$.
- The winner is the person whose number is the closest to p of the average of all chosen numbers, where $p = \frac{2}{3}$.
- A fixed prize of $\$x$ will be awarded to the winner, and in case of a tie the prize is split amongst those who tie.
- The same game can be repeated several periods, in each period, subjects are informed of the mean, $\frac{2}{3}$ of the mean and all choices.
- Theoretical prediction for this game points to a unique interior Nash equilibrium solution of 0.

I will focus my analysis in the following areas:

1. **Strategic Level.** The frequency of choices will be explored and analyzed. Following Nagel (1995), an agent is of strategic degree n if he chooses a number $r(\frac{2}{3})^n$, where r is defined to be the reference point. This could be a choice characterized by naive player or a point of salience in heuristics, which I will define in later parts.

2. **Convergence of Choices.** In the repeated game setting, every agent is responding to the information revealed from previous rounds. The changes in choices adopted by the agents can be tracked to determine if there is convergence to the unique Nash equilibrium of the game.

3. **Frequency and Evolution of Strategic Level.** The frequency of strategic levels across periods was also tracked to get a gauge of the average strategic levels of the simulated agents that does not simply dependent on the choices made in the first round. It is also possible to explore if strategic level do evolve over time. The evolution of strategic level can be found by adjusting the reference point of each period, fixing it to be the mean of the previous period. If n increases, it can be interpreted as increasing level of thinking or revision in actions to best respond to opponents that could be of higher strategic level.

4. **Evolution of Payoffs.** The most interesting portion about multi-agent competitive games is the possibility that stronger models might not always be better in generating the highest payoffs. Therefore, exploration of payoffs and their transition over time could be important to determine how agents' performance could be dependent on their own strength as well as on the environment.

3.1.1 One-Shot Game

150 sessions of one-shot beauty contests were ran with 9 LLM-based agents back-boned by multiple LLM types. These games are used to determine the strategic level of each model.

In the standard beauty contests, where $\bar{c}$ is fixed to be 100. Via the iterative elimination of weakly dominated strategies, all choices between $(66.66, 100]$ are weakly dominated by 66.66, and all choices above 44.44 are weakly dominated by 44.44, etc. Going by the level-k model, for a focal point set at 50, level-0 chooses 50, and level-1 responds by choosing 33.33, etc. In this modified set-up, a randomly generated upper-bound, $\bar{c}$, is used for each one-shot game. The upper-bound will not affect the assessment of strategic level. For instance, using the level-k model, level-0 would choose the focal point, which would be the mean of uniform random choices, $\frac{\bar{c}}{2}$, following Nagel (1995), and level-1 would respond by choosing $\frac{2}{3}\frac{\bar{c}}{2}$, so on and so forth. The average strategic levels of each model are computed in this manner to ensure a more robust measure of strategic level that is consistent across changing game parameters.

Choices. The frequency of normalized choices made across the 150 sessions of one-shot games is shown in Figure 1. The choices are concentrated on 50 for `ChatGLM3`, `Baichuan2`, `Claude1`, `PaLM`. Based on the level-k model proposed in (Nagel (1995), these models can be characterized as naive or level-0 players. The average choice for `Llama2` lay around 60, approximately corresponding to level-0, and the choices are noticed to be fairly dispersed across the range of 40 to 80. `Claude2`,

GPT3.5 and GPT4 , on the other hand, are displaying slightly different choice frequencies as compared to the previous models. For Claude2 , there is a spike around 33, indicating the likelihood of level-1 thinking, on top of that, there is high choice frequency around 66 as well, which can be rationalized via step 1 of the process of iterated elimination of dominated strategies. (Mauersberger and Nagel (2018)) As for GPT3.5 , most of the choices are also concentrated around 33, stipulating level-1 reasoning. While there are some other spikes at 50 and 66, but are of much lower frequency. Lastly, for GPT4 , the highest spike in choices is around 44, going by iterated elimination of dominated strategies, this could imply step-2 of the depth of reasoning process, a lower frequency spike is also seen at around 33, indicating level-1 thinking via the level-k model. There are no data for ChatGLM2 as it is unable to complete any of the games, which can be attributed to it being a relatively weak model and thus is unable to produce any output based on the instructions. Based on the choice frequency, it is not surprising that the larger language models are able to generate higher frequency in choices that can be characterized as higher level of strategic thinking. One further detail to notice is that for Claude2 and GPT3.5 , there are instances where they in fact select the NE choice of 0, consisting 6% and 5.3% of the sessions respectively.
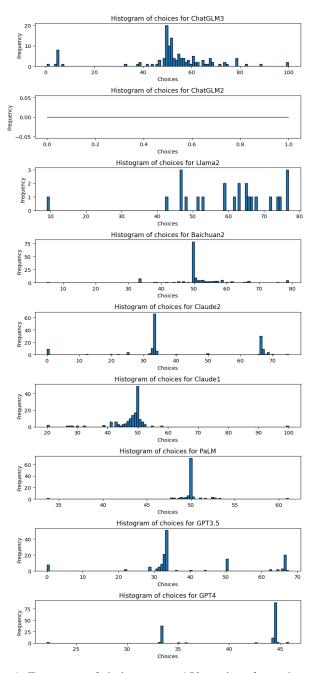

Figure 1: Frequency of choices across 150 sessions for each model.

In Nagel (1995) and Bosch-Domenech et al. (2002)'s period 1 results done on different human populations, including students (mean choice=36.73, median choice=33), theorist (mean choice=17.15, median choice=15*), newspaper readers (mean choice=23.08, median choice=22*), etc.[2] The human subjects experiments shown strong deviation away from game-theoretic solution, and going by the level-k model, they on average display iteration steps 1 and 2. Even though it is often expected that the general public is more noisy in choices, in this case, newspaper readers have slightly lower mean than the student population, possibly due to longer time for reflection and thereby making more contemplated decisions. (Mauersberger and Nagel (2018)) Evaluating the simulation results of one-shot beauty contests with LLM-based agents against the results with respect to human subjects, they seem to be choosing slightly higher numbers that imply a lower average strategic level of 0 to 1 as compared to human subjects. However, this result does comply with the impression that general public should display more noise or randomize their choices more, inducing a lower strategic level on average. Even though it can be argued that LLM-based agents are all representing the general public, there remain heterogeneity among them. Given that the alignment of models are riding on top of data generated with different sets of labelers and users, the simulated agents back-boned by different LLMs can represent different subsets of the general populations, the difference is that types are not defined based on one's career for instance, but characterized by their revealed average choices and strategic levels. As a result, games with multi-LLM-based agents describe a predicted outcome for games played by the general population, but involving individuals with heterogeneous degree of strategic thinking.

Models	ChatGLM3	ChatGLM2	Llama2	Baichuan2	Claude2	Claude1	PaLM	GPT3.5	GPT4
Average	52.029	N/A	59.519	51.158	41.609	47.696	49.976	38.912	41.072
Median	51.724	N/A	62.685	50.0	33.333	49.313	50.0	33.333	44.442

Table 1: Average and Median Choice of the LLMs across 150 Sessions

A possible follow-up question is that if the models are given the exact same instructions, would they choose the same number consistently? For human subjects, when given identical game set-up, it is possible that they might employ different strategies. There could be many reason behind this. Devetag et al. (2016) highlights the potential of players focusing on different features of the games across the series of one-shot games, thereby selecting different strategies for identical games. Further, Costa-Gomes and Weizsäcker (2008) indicated the discrepancy between individuals' stated beliefs and actions taken, implying one's actual choices might be affected by factors not aligned with this understanding, which translate into inconsistency in strategies. The same could apply to LLM-based agents. If each LLM-based agent were to represent a single type of human agent, it could be important to understand how varied one's choice of number might be given the same instructions. This also applies if each LLM-based agent were to represent a single type of population, where variation in choices could be further explained by proportions of agents playing different pure strategies, which might lead to different revealed actions in each round of identical games. This exercise is important to determine if, like human players, there could be variability in choices.

When using LLMs as simulation tool, it is possible to some extent to control for the variability in the strategies executed by agents through changing the temperature. Chollet (2021) introduced softmax temperature in the text generation process, which characterizes the randomness of the choice of the next word. Higher temperature corresponds to less predictability of the output, and lower temperature indicates more deterministic result. However, though this feature characterized the randomness in text output, its impact on strategic choices is not as clear. Therefore, by exploring the revealed choices given the same level of temperature is being set for all the models, this can in fact determine how random the models are in strategy selection while holding all else constant.
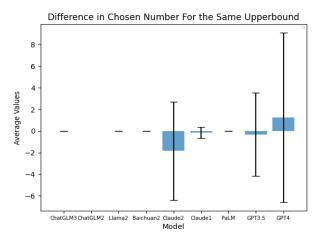

Figure 2: Variability in chosen number given the same upper-bound.

---

[2]The median for the later groups were not computed, but it is guesstimated based on the figures in Nagel (1995) and Bosch-Domenech et al. (2002) to be around 15 for theorist and 22 for newspaper readers.

Figure 2 shows that within the 150 sessions, for those sessions having the same upper-bound of choice range, the same LLM-based agent could choose slightly different numbers. For instance, `Claude2`, `GPT3.5` and `GPT4` displayed more variability in choices as compared to other models. This results is indicative that choices might not be static even when the instructions is exactly the same, which further emphasize the need to run many sessions of one-shot games that encompass both identical upper-bounds and different upper-bounds, the determination of average choices and the corresponding strategic levels based on which would render a more consistent and robust measure for each model.

Strategic level relative to reference point. To evaluate the exact strategic level of each model given their choices, I follow (Nagel (1995))'s method used for human subjects. One's strategic level is denoted as n, and r is the reference point, the computation of strategic level is determined using $r(\frac{2}{3})^n = $ chosen number. In Mauersberger and Nagel (2018), the reference point refers to the choice of a non-strategic agent, who is assumed to play the mean of the range of choices, pertaining to insufficient reasoning. However, this focal point can be arguable. In these current set of beauty contest games, the upper-bound varies, therefore, calculating for the mean of uniform random choices may not be as straightforward as games with fixed upper-bound, thus it could be reasonable for the upper-bounds to be the focal points as well. Figure 3 shows that for the 150 sessions, the average strategic level of each LLM-based agent lay between 0 and 1 when the reference point is set to mean of random choices (i.e $r = \frac{\bar{c}}{2}$) and between 1 to 2.5 when the reference point is set to be the randomly generated upper-bound of each game (i.e. $r = \bar{c}$).

For the conventional focal point of $\frac{\bar{c}}{2}$, Figure 3a shows that the average strategic level for `Claude2` and `GPT3.5` are comparable. Surprisingly, `GPT4` has slightly lower strategic level relative to these two models, even though it is often assumed to be a stronger model than `GPT3.5` by being faster and more accurate, but this result shows that it is not necessarily true that `GPT4` is of a higher strategic level. In relation to human subjects, a model with higher strategic level could represent agents with higher intelligence and/or believing that their opponents are of relatively high strategic level, thus this difference between `GPT4` and `GPT3.5` could result from either or both of the preceding reasons. Since `GPT4` is known to be trained on more data, which could result in more noise in strategies employed, it could be argued that agents back-boned by `GPT4` is likely to have an initial "belief" that its opponents have relatively low strategic level, thus leading to guesses that corresponds to lower depth of reasoning. Further, while the average strategic level of `ChatGLM3` is comparable to that of `GPT4`, its variability is much higher. This could suggest that smaller models can representing agents who either have lower strategic level or larger variability in depth of reasoning, expanding the dimensions of types of agents the models can simulate for.
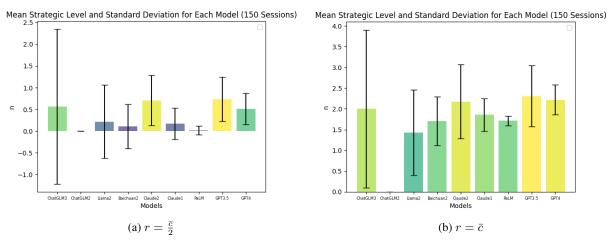

(a) $r = \frac{\bar{c}}{2}$ (b) $r = \bar{c}$

Figure 3: Average strategic level of LLM-based agents in beauty contest games with difference reference points.

Payoff. Apart from choices and strategic levels, it is also important to learn the average payoffs earned by each model. In Figure 4, it can be seen that `Claude2`, `GPT3.5` and `GPT4` have relatively higher average payoffs than the others. Among the three, standard deviations are comparable for `Claude2` and `GPT3.5`, and `GPT4` has a slightly lower standard deviation than the other two. Despite `GPT3.5` being the best in average payoffs as compared to the other models, its variability in payoffs is also relatively high, making large and small gain equally likely. Coupled with the average strategic level of each model, it can be seen that the ones with higher strategic levels often obtain higher payoffs, except for `ChatGLM3`, whereby the variability in its strategic levels could have adversely influenced its average gain.

Overall, I have analyzed the average choices of each LLM-based agent in the one-shot beauty contest games, and their strategic levels were determined using methods that were previously employed to evaluate human subjects in experiments. In a way, this facilitates better understanding of machine behaviour by drawing on theories that were used to study human decision-making. Herein, the models also showcased their potential in representing heterogeneous agents, who have different

strategic levels. Through varying the model selection, degree of heterogeneity in the population to be simulated for can be controlled to some extent.
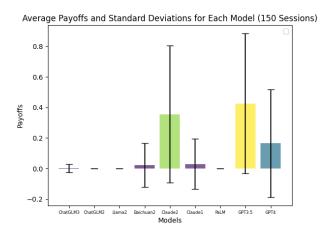

Figure 4: Average payoffs across 150 sessions for each model.

3.1.2 Repeated Games

In this subsection, I continue with experimental results from Guo (2023), but under the repeated setting of 6 periods. The prompt to include historical information is detailed in Appendix A.1. Across the periods, the game parameters faced by the agents do not change, thus they choose numbers within the same range of $[0, \bar{c}]$, and $\bar{c}$ is randomly generated in period 1. Historical information, including choices made by all agents, average of these choices, $\frac{2}{3}$ of the average, and the winners from the previous round, are revealed up to 3 periods. The limitation on revealing 3 past periods is as a result of token restrictions to control the computational intensity. Such setting can be seen as partial feedback or feedback with a forgetting parameter. It is expected that agents will react to the information and slowly converge to the Nash equilibrium.

The repeated beauty contests ran for 30 sessions. As before, via period 1 choices, the corresponding average strategic level for each model, assuming the reference point to be $\frac{\bar{c}}{2}$ is determined to be between 0 and 1, shown in Figure 5a. In Figure 5b, I explore the average chosen number across the 6 periods. The numbers are normalized for better comparison across different sessions that could have different $\bar{c}$. It is shown that with historical information, most LLM-based agents converge in actions, particularly for Claude1 , Claude2 , GPT3.5 , and GPT4 , their average normalized chosen numbers are approximately 0 in period 6. Suppose models with higher strategic level represent more intelligent agents, then the "smarter" agents do display "learning" pattern given historical information, where they converge to the unique Nash equilibrium of 0 across time.
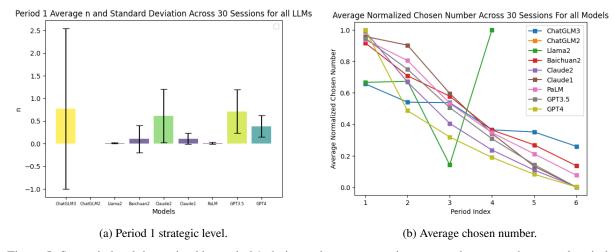

(a) Period 1 strategic level.

(b) Average chosen number.

Figure 5: Strategic level determined by period 1 choice and convergence in average chosen number over 6 periods.

10

Frequency and Evolution of Choices and Strategic Levels: Given the repeated setting, it is possible to track for changes over time, as a result, the frequency of choices and the corresponding strategic levels are computed over the 6 periods and across 30 sessions in Figure 6.

(a) Choices.
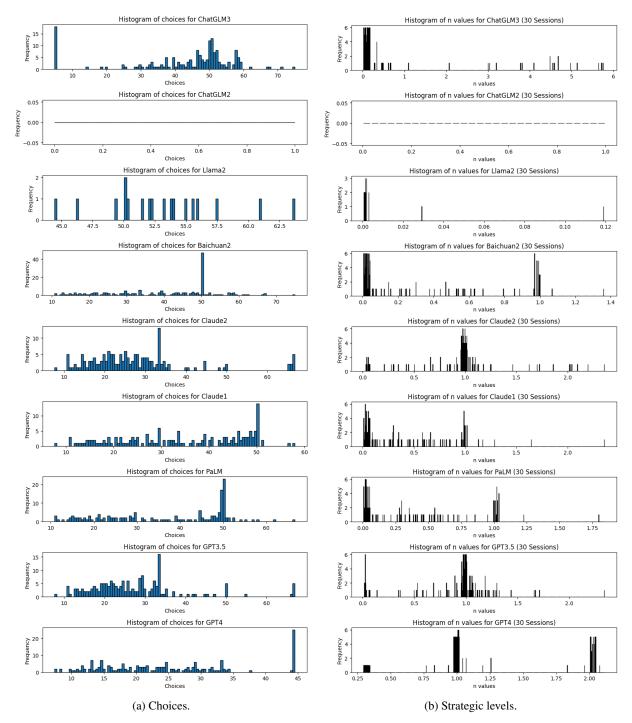
(b) Strategic levels.

Figure 6: Frequency of chosen number and strategic level calibrated to new reference points over 6 periods across 30 sessions.

In Figure 6a, for most of the LLM-based agents, the choice of 50 receives the highest frequency, but `Claude1`, `GPT3.5` and `GPT4` do display more selection below 50. Since the choice frequency consists of period 1 to 6, the histograms display greater dispersion in number selected as compared to the one-shot games in Figure 1, which is an indication that agents do change their

actions over time, and there is more selection of numbers lower than 50. Following Nagel (1995) method of computing for strategic levels in repeated setting, where the first period reference point is fixed at $\frac{\bar{c}}{2}$, and in subsequent periods, the reference point is re-calibrated to be the mean choice of the previous period. In experiments with human subjects, the results show no support for increasing depth of reasoning, most subjects remain in the bounds of iteration step 0 and 3, and they rarely go over iteration step 2. With LLM-based agents, based on the frequency of strategic levels computed across periods in Figure 6b, the results are similar, they do not go over iteration step 2, and most of them display spike at 0 and 1, with `GPT4` showing spike at both 1 and 2. However, even though they are comparable to human subjects about staying below iteration step 2, some of them do display minor increases in strategic level as compared to their period 1 performance. In particular, `GPT4` has strategic level much lower than 2 in period 1, but is able to achieve iteration step of approximately 2 with high probability over time.
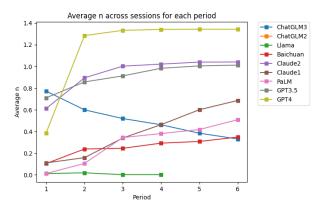

Figure 7: Averaging strategic level across 30 sessions for 6 periods.

For illustration, I explore the changes in strategic level for each model, averaged across sessions, for each period in Figure 7. It shows that strategic levels evolve over time, but the range of that is rather narrow, on average they remain within the bound of 0 and 1.4. Most LLM-based agents shows increasing depth of reasoning as defined by the growing strategic level, this is especially true for stronger models, `Claude2`, `GPT3.5` and `GPT4`, where strength is characterized by the relative strategic level computed in period 1. The abnormality comes from `ChatGLM3` and `Llama2`. `Llama2` has a relatively low strategic level in period 1 and did not display improvement over time, its lack of change could imply some kind of non-strategic or naive behaviour, where they keep behaving randomly. As for `ChatGLM3`, the LLM-based agent do have relatively higher strategic level than the other agents in period 1, but yet, its average strategic level across time decreases. This could imply its lack of ability to respond to historical information and adjust its behaviour accordingly.

Evolution of Payoffs: Figure 8 shows the payoffs for each LLM-based agent over time, averaged across 30 sessions. `GPT3.5` outperforms the rest in all periods, while `Claude2` and `GPT4` are more or less comparable. The rest of the LLM-based agents do not obtain average payoffs as high, but most of them display growth over time. Coupled with Figure 5b that shows convergence in average choice towards NE of 0 for most LLM-based agents, the increasing payoffs could be an indication of learning about the optimal action to take to win the game and also obtain higher payoffs. The exception is `Llama2`, it does not finish the 6 periods and average actions display no convergence, thus its corresponding average payoffs is also consistently 0 throughout.
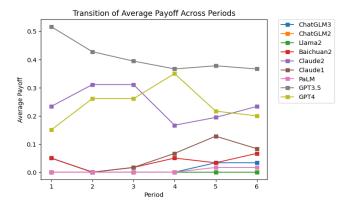

Figure 8: Average Payoffs across 30 sessions over 6 periods.

In this section, I explored the outcomes of heterogeneous population comprises of multi-LLM-based agents to illuminate strategic interactions between heterogeneous agents or algorithms if LLMs were adopted as proxies.

The purpose of one-shot games, as well as the first period of the repeated games, are used to evaluate the strategic level of different LLM-based agents, with the objective of exploring their behaviour as simulated agents that we can draw parallel to human subjects. I found that models have varying strategic levels, the ones with larger parameter size do not necessarily have higher relative strategic level, but models with higher strategic levels tend to achieve better payoffs on average. The results from one-shot games resembles to that of experiments with human subjects, where agents tend to display low levels of reasoning, however, the distinction is that the average iteration steps are slightly lower for LLM-based agents than that of human subjects in experimental setting. Since LLMs encompass data trained on potentially larger and more general groups of population, and experiments with human subjects tend to explore more specific groups of population depending on the research needs, the degree of noise in strategies employed and the beliefs about opponents' depth of reasoning could be different, thereby leading to slight distinction between game results ran with LLM-based agents and human subjects. By evaluating machine behaviours using traditional analysis methods conducted on human subjects and compare the results between the two, this helps not only to understand machine behaviour in human-like terms, it also draws some insights on what type of agents the LLMs can represent in game simulations.

The LLMs involved in the games could vary in parameter size, which is a measure for complexity of neural network structure that is potentially in equivalence to human brain capacity (i.e. memories and ability to represent complex pattern), it seems straightforward to define one dimension of heterogeneity in LLM-based agents based on degree of complexity, however, even though LLMs can mimic human performance in complex tasks, their similarity to brain anatomy and physiology is imperfect, connecting LLM-based agents and human subjects on this line of differentiation could be problematic. (Pulvermüller et al. (2021)) Therefore, instead of distinguishing LLM-based agents on the basis of their resemblance to human brain capacity, it is more precise to characterize their heterogeneity in terms of potential difference in their underlying training data, revealed strategic levels and variance in choices in games, such as the one presented in this paper.

Apart from the one-shot beauty contest games, I have also explore the repeated setting, which shine a light on how the simulated agents could behave over time, if they respond to historical information and whether there is any improvements in their depth of reasoning. Similar to human subjects, while LLM-based agents do not display iteration steps that go over 2 for the span of the games, they do seem to learn from historical information and show convergence in actions towards NE.

3.2 Adaptive Learning with Variation in Group Composition

To further analyze the strategic interactions between LLM-based agents, one possible aspect would be to investigate their adaptive learning behaviour given variation in group composition. Since I have already identified the LLM types in terms of their strategic levels, it is possible to conduct a simulation between heterogeneous agents, and observe their dynamics of game play by varying the proportion types for each session. Such repeated interactions in competitive games could assist in understanding what would human subjects do in similar set-ups, knowledge of which could be helpful in informing policy design. Furthermore, it would also foster better understanding of how algorithms will react to one another if they were adopted widely in competitive setting.

In the following sections, I have selected two LLM types, a stronger and a weaker model, `GPT3.5` and `PaLM`, where the strength is characterized by their strategic levels as analyzed in the previous section. They can be used to represent simulated agents who are more "intelligent" and "less intelligent", or simply algorithms that are stronger and weaker. The games follow a revised design from Nagel (1995):

- 10 LLMs-based agents are playing in each game.
- The same group plays for 5 periods, and all history are revealed.
- They choose a number between 0 and $\bar{c}$, where in this section, $\bar{c}$ is fixed to be 100. The winner is the agent whose number is the closest to p times the average of all chosen numbers, where $p = \frac{2}{3}$ to ensure a unique interior NE solution.
- In each period, the winner gets a fixed prize of $\$x$. In case of a tie, the prize is split amongst those who tie. All other players receive 0.

The metrics of interest for my analysis are similar as before, including (1) initial choices and strategic level - period 1; (2) frequency and evolution of choices - period 1 to 5; and (3) payoffs - period 1 to 5. But in addition, I will also compare the adaptive learning behaviour across different set-ups that comprises of varying group composition.

There are two main environment that I hope to investigate: In the first set-up of hard-coded fixed strategy environment, a single LLM type will be playing against fixed-strategy opponent(s). The LLM-based agent(s) will be told there is a proportion of fixed strategy players and they are expected to react to this prompt in the first period, which can be seen as forming " belief" about what other players will play and they are best responding to that "belief", subjecting to their own strategic level. Based on past game play, they could learn what is the proportion of fixed strategy players and form "belief" about the strategies of other players', who are not fixed strategy. In the second environment, I explore the interaction between the two LLM types, and observe how they react to different proportion of LLM types in the population. In this case, the environment is more dynamic as both LLM types will be responding to potential changes in strategies given their "beliefs" about what the opponents might be doing, the actions would have to be more strategic in order to win the games.

3.2.1 LLM vs. Static Algorithm: Variation in Hard-coded Fixed Strategy Opponents

In the partial static environment illustrated in this section, there are no change in hard-coded agents' actions, but across different settings, the proportion of fixed strategy players and LLM-based agents change while maintaining the group size.

The set-ups are as follows for 5 periods in total, history reveal up to 4 periods; $p = \frac{2}{3}$; and agents to choose a number between $[0, 100]$. There are 3 treatments:

1. 1 LLM + 9 Hard-coded Agents (Low strategic uncertainty)
2. 5 LLMs + 5 Hard-coded Agents (Mixed strategic uncertainty)
3. 9 LLMs + 1 Hard-coded Agents (High strategic uncertainty)

I follow an exemplary prompt in Appendix A.2. LLM-based agents are specifically told that some of their opponents are playing a fixed strategy of 0, which is the NE strategy of the game. This would mitigate the strategic uncertainty, but only to some extent, as the population may comprise a mixture of LLM-based agents and fixed strategy agents. In (Duffy et al. (2021)), where repeated prisoners' dilemma games were played, human subjects were instructed that they are playing against programmed opponents that use the grim trigger strategy. Since the games were played in pairs, there is complete elimination of strategic uncertainty, however, in this setting of multi-agent competitive game, by varying the proportion of fixed strategy players and LLM-based agents, the degree of strategic uncertainty varies across different sessions.

The NE strategy should not be affected by the proportion of fixed-strategy agents and LLM-based agents, but it is expected that the speed of convergence towards NE could differ across settings. In the three set-ups, denoting a_t to be the action/number guessed in each time period, N_f to be the number of fixed-strategy players and N_l to be the number of LLM-based agents, the selection in the next period:

$$a_{t+1} = BR(N_f, N_l, a_t) = \frac{2}{3}\left(\frac{N_f}{10} * 0 + \frac{N_l}{10} a_t\right) \tag{1}$$

The choice variation over the periods is computed with $\frac{a_{t+1}}{a_t}$. For 9/10 fixed-strategy agents, the next period guess would be 0.067 of the previous number; For 5/10 fixed-strategy agents, the guess would be 0.333 of the previous number; For 1/10 fixed-strategy agents, the guess would be 0.6 of the previous number. It is expected that if the proportion of fixed strategy agents in the population is higher, it would be more likely for LLM-based agents to reach 0 faster.

Higher Intelligence Model (represented by `GPT3.5`) vs. **Lower Intelligence Model** (represented by `PaLM`) when playing against fixed-strategy opponents:

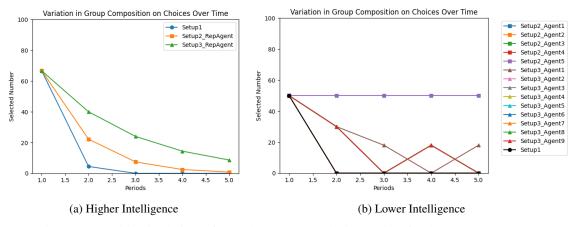(a) Higher Intelligence   (b) Lower Intelligence

Figure 9: Transition in choices of LLM-based agents playing against fixed strategy opponents.

In period 1 of all sessions, `GPT3.5`-based agents choose around 66.67, which corresponds to step-1 by iterated elimination of dominated strategies, and agents back-boned by `PaLM` choose 50, which is level-0 as characterized by the level-k model. To use the same basis of evaluation, it is important to link the agent types with the one-shot beauty contest games discussed in Section 3.1. By applying the level-k model with the reference point of $\frac{\bar{c}}{2}$, I have determined `GPT3.5` to have an average strategic level of approximately 1 and `PaLM` is approximately 0. They represent higher and lower intelligence agent respectively, where intelligence is interpreted loosely as metonym for strategic level.

In set-up 2 and 3 of Figure 9a, a representative agent is selected since the the choices are almost identical for all LLM-based agents in the same set-up. This is in contrast to the set-up 2 and 3 of Figure 9b, where different LLM-based agents could take actions that are substantially disparate, though they are of the same type. Across all three set-ups, the choices for higher intelligence agent(s)

converge towards 0, exhibiting either refinement of belief about opponents' strategies or progression in their depth of strategic thinking when given historical information. The pace is slower as the proportion of LLM-based agents becomes larger relative to fixed strategy agents. As for lower intelligence agent(s), there is also similar pattern of slower convergence to 0 when the proportion of LLM-based agents is higher than the fixed strategy ones. However, their results for set-up 1 and 2 largely coincide, indicating that lower intelligence agent(s) are not very sensitive to the difference in environments of having 90% vs. 50% fixed strategy opponents in the group. Although, there is an outlier in set-up 2, where one agent did not converge at all. Alongside larger fluctuations in choices in set-up 3 that comprises of even lower proportion of fixed strategy opponents, this suggests that higher strategic uncertainty could induce greater variability in strategies and might lead to non-convergence behaviour in lower intelligence agents. When comparing between the higher intelligence and the lower intelligence agents, the lower type is less "cautious" in a sense that they could converge to 0 in period 2 straightaway, such as the case for set-up 1 and 2, while convergence to 0 only happen in period 3 of set-up 1 in Figure 9a, and the other two set-ups do not actually achieve 0 but approaches 0 in the limit as the number of periods increases. Given all past choices in period 1 are revealed to the agents in period 2, the steep adjustment in choices comply with the expectation that information about N_f is implicitly fully disclosed by period 2, as one can determine the proportion of fixed strategy agents through the revealed choices. The higher intelligence agents follow the prediction of step-by-step adaptation, while lower intelligence agents could do a one step jump to 0. In a sense, higher intelligence agents demonstrate movement from less sophisticated strategies to more refined choices through iterative learning and adaptation, and on the other hand, the lack of such systematic adjustments in choices by the lower intelligence agents could suggest that they are relying more on intuitive guesses than successive elimination of less likely options.

Convergence Rate: The convergence rate of choices can be computed with

$$c_t = \frac{a_{t+1} - a_t}{a_t}, \text{where } a_{t+1} \geq a_t \tag{2}$$

For the higher intelligence agents, the convergence rates are constant for set-up 2 and 3. As for set-up 1, convergence to 0 happened in period 3, therefore, the peak of convergence rate has been reached then, and there are no further convergence. The results demonstrate that higher proportion of fixed strategy opponents, lower strategic uncertainty, the revelation of historical information therefore corresponds to higher convergence rate. On the other hand, for the lower intelligence agents, convergence to 0 happened straightaway in period 2 for most LLM-based agents in set-up 1 and 2, therefore, convergence rates drop to 0 thereafter. In set-up 3, the convergence rate fluctuates, in tune with previous analysis that higher strategic uncertainty contributes to larger variation in choices for the lower intelligence agents. Compared the two types of LLMs, there is more stable adjustment in choices over time for higher intelligence agents, and while convergence could be faster for lower intelligence agents, larger noise in choices is recorded when uncertainty is high.
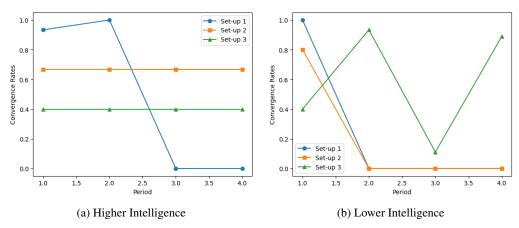

(a) Higher Intelligence   (b) Lower Intelligence

Figure 10: Convergence rates in choices of LLM-based agents, where $t = 1$ represents transition from period 1 to 2.

Evolution of Strategic Level: When evaluating the transition in strategic level across periods, all LLM-based agents start out from level-0, as evaluated by level-k model with reference point of 50. Removing the cases where NE choice of 0 is first played in the current period, it can be seen that for the higher intelligence agents, their level transitioned from 0 to 1 before playing the NE choice. Though they display improvements in strategic sophistication or belief refinement, their level remain below 1. As for the lower intelligence agents, most of them stay at level-0, with some fluctuations between 0 and 1 when strategic uncertainty is high.[3]

---

[3]In period> 0, if the current chosen number is 0 and previous is not, the optimal strategy has been realized given the historical information, so strategic level for these periods where NE choice is first played are not included. If previous number is 0 but current is not, then strategic level is 0 as one must be randomizing and not actually adaptively adjust.
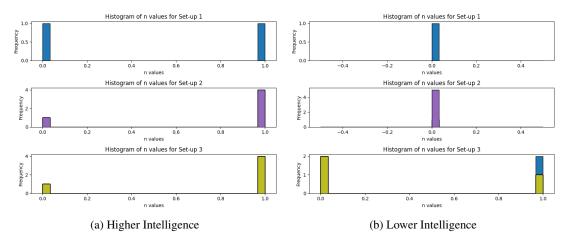
(a) Higher Intelligence

(b) Lower Intelligence

Figure 11: Frequency of strategic levels for each agent across periods within each set-up.

Payoffs. The next step is to determine the payoffs of LLM-based agents in each set-up.
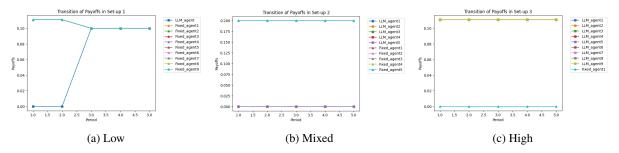

(a) Low

(b) Mixed

(c) High

Figure 12: Transition of payoffs for higher intelligence LLM-based agent(s) and fixed-strategy opponents.

Figure 12a shows that in the environment with low strategic uncertainty, the single higher intelligence LLM-based agent starts off with payoffs of 0, but over time, as it converges to the NE choice of 0, the prize are shared among all those who ties and the agent manage to obtain positive payoffs from period 3 onwards. In the mixed environment (Figure 12b), though the actions of LLM-based agents converge towards 0, they remain further away from $\frac{2}{3}$ of the average than fixed strategy players, therefore, prizes are shared by fixed strategy players and LLM-based agents have a flat payoff of 0 throughout all 5 periods. Lastly, in the environment with high proportion of LLM-based agents (Figure 12c), the reverse happens, since $\frac{2}{3}$ of average is relatively high, the prizes are shared by the LLM-based agents, and they have a flat payoff of 0.11. In sum, higher intelligence LLM-based agents tend to gain better payoffs in the environment with low and high strategic uncertainty as compared to the mixed environment when playing against fixed strategy opponents.

(a) Low

(b) Mixed

(c) High
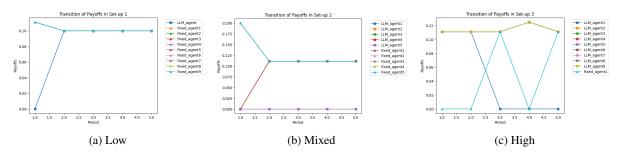
Figure 13: Transition of payoffs for lower intelligence LLM-based agent(s) and fixed-strategy opponents.

In Figure 13a, the single lower intelligence LLM-based agent shows similar payoff pattern as that of higher intelligence agent. It starts off with 0 payoff, and as its choice converge to NE quickly, it is able to earn positive payoffs from period 2 onwards. In the mixed environment (Figure 13b), unlike the higher intelligence agents, majority of the lower intelligence agents have non-zero payoffs. As they convergence towards the NE choice, the prizes are split among some of the LLM-based agents and the fixed

16

strategy players who tied. Lastly, for the high strategic uncertainty environment (Figure 13c), majority of the lower intelligence LLM-based agents achieve positive payoffs across the periods. However, the payoffs are not flat, this is contributed by a case of a specific LLM-based agent "lagging behind", where it chose a number that was chosen by majority in the past period but not in the current period.

Comparing between higher and lower intelligence LLM-based agents, interestingly, payoffs achieved in all settings by the lower intelligence agents could often be comparable or higher than that of the higher intelligence agents, though the variations in payoffs is also larger. This could indicate that higher strategic level does not necessarily imply higher payoffs when competing with fixed-strategy opponents.

In general, when LLM-based agents are playing in a repeated beauty contest game with fixed strategy opponents, they display convergence in actions towards the interior NE strategy of 0 over time given historical information of past choices and payoffs are revealed. As I vary the proportion of LLM-based agents and fixed strategy opponents, the speed of convergence is usually slower when there is presence of other simulated agents, which contributes to higher strategic uncertainty. The payoffs could also be in favor of the LLM-based agents when strategic uncertainty is relatively high, as $\frac{2}{3}$ of average in which case would fall much higher than 0. This results not only signifies the potential game play if human subjects are playing against opponents that naively adopt a fixed strategy of 0, it could also portray a simulated outcome if they are going against static computer algorithms that are playing a fixed NE strategy.

Application. A typical example of beauty contest application which has strategic complements, as the one in this paper, is the Bertrand competition model. The LLM-based agents and fixed strategy agents can be perceived as simulated firm entities that adopt different pricing strategies, the objective is to win the competition and to maximize their profits. In the slightly modified version of the Bertrand market described in Mauersberger and Nagel (2018), let me first suppose there are 10 firms, and each of them have a marginal cost of γ. They set the price of the product at time t to be $p_{it} = \frac{2}{3}\hat{E}_{it}\min\left(\gamma, p_{1t}, p_{2t}, ..., p_{10t}\right) - \gamma$, where $\hat{E}_{it}$ is the subjective expectation of firm i held at time t. The fixed strategy players could be seen as those that always play the equilibrium action of setting the price equals to the marginal cost, $p^{eq} = \gamma$, while LLM-based agents are firms that adjust their pricing strategies in each time period. Based on the simulation results previously, having higher proportion of fixed strategy firms would drive the prices set by LLM-based firms down faster. Firms with higher strategic level would adjust the prices down step-by-step, while firms with lower strategic level either adjust straightaway or they failed to adjust at all. In this context, if there exist certain rigidity in the short run, such as production capacity constraints for the firms or limited response time for the consumers, then firms who set higher prices in the short run would be able to obtain higher profits. In the long run, however, all factor inputs are flexible and consumers will not purchase from the firm that sells a homogeneous product but at higher price than the equilibrium, therefore, it is better for firms to converge to the equilibrium price in order to stay in the market. For instance, suppose q_{it} is the quantity sold by LLM-based firm i in time t, it could be earning a higher profit than fixed strategy players, f, if:

$$p_{it}q_{it} - \gamma q_{it} > 0, \text{ where } p_{ft}q_{ft} - \gamma q_{ft} = 0 \tag{3}$$

Assume $\gamma = 0$, as long as $q_{it} > 0$, they are able to earn more than the fixed strategy players in the short run. As prices are driven down to the marginal cost over time, firms setting higher prices will risk losing consumers, so everyone will be lowering prices to 0 and be earning normal profit in the long run. In this sense, higher intelligence agents could often achieve better outcome than lower intelligence agents in the short run, where they can earn a positive profit by converging gradually. Even in the long run, it is possible that lower intelligence agents have larger variance in pricing strategies as compared to higher intelligence agents, where they either failed to converge or display high volatility in prices, these could adversely impact their profits.

Computer algorithms. Nonetheless, the above games simulated between LLM-based agents and the fixed strategy agents could also be seen as a competition between a dynamic responsive algorithms and a static computer algorithm, where the same Bertrand story could apply and firms are outsourcing their pricing strategies to automated algorithms. Such automated pricing algorithms have been widely discussed in literature, but having LLM algorithms that are dynamically responsive to changes in rivals' strategies could spark fresh perspective to existing research questions, such as interactions between multiple pricing algorithms and how would they impact the dynamics within the market. (Brown and MacKay (2023), Chen et al. (2016))

In a more general sense, it is possible to envision a future with greater adoption of machine algorithms in executing commands, be it in the consumer market or other aspects, thus understanding interactions between machine algorithms could shine a light on how they would behave in view of one another. The appealing feature of the LLM algorithms is that optimal strategy does not need to be first determined, and given past choices, they will learn to respond to these information. In the standard beauty contest games illustrated, when comparing the payoffs, static algorithm that has a pre-defined strategy that is already fixed at NE might win the game and obtain better payoffs than LLM-based agents earlier on, but they might not fare as well as the games carry on. In particular, LLM algorithms could obtain higher payoffs than the static algorithm when the group comprises of more LLM algorithms, implying that with greater adoption of the dynamic responsive algorithms, they would outperform static computer algorithm in the aforementioned set-ups, drawing greater traction to investments in algorithms. However, the irony is that payoffs with stronger models could be lower than with weaker models, especially in the mixed environment, this results from more cautious adjustment based on past information. In view of this, the adoption of the type of LLM algorithms could also be dependent on the risk-aversion of individuals using them.

17

3.2.2 LLM vs. LLM: Variation in Opponent Types

This section illustrated a dynamic environment where two different types of LLM-based agents are playing against each other and the proportion of LLM types differ across set-ups.

Similarly, there are 5 periods in total with history reveal up to 4 periods; $p = \frac{2}{3}$; and agents to choose a number between $[0, 100]$. There are 5 treatments, where high intelligent LLMs are represented by `GPT3.5`, denoted from here on as the high type (H), and the less intelligent LLMs are represented by `PaLM`, denoted hereafter as the low type (L). Their types are characterized by strategic levels evaluated in section 3.1.

1. 10 H LLMs (Pure high intelligence environment)

2. 9 H LLMs + 1 L LLM (Highly intelligent environment)

3. 5 H LLMs + 5 L LLMs (Mixed intelligent environment)

4. 1 H LLM + 9 L LLMs (Less intelligent environment)

5. 10 L LLMs (Pure low intelligence environment)

I use the original prompt with historical information as in Appendix A.1.

Let the strategy of high type in period t be a_{Ht} and that of low type be a_{Lt}, the selection of next period choice would be:

$$a_{it+1} = BR(B(N_H), B(N_L), a_t) = \frac{2}{3}\left(\frac{B(N_H)}{10}a_{Ht} + \frac{B(N_L)}{10}a_{Lt}\right), i \in (H, L) \tag{4}$$

where $B(N_H)$ and $B(N_L)$ are agent i's "beliefs" about the number of high types and low types. When playing against fixed strategy opponents, it is possible to observe in period 2 who selected 0, thereby deriving the correct proportion of fixed strategy players within the population. Since all agents are back-boned by LLMs in these set-ups, it could be harder to distinguish the proportion of types within the group based on historical choices in period 2, for instance, even if they chose the same number it does not imply they are of the same type. Further, the agents were not told explicitly their own type relative to the others, so they have to guess if they fall within N_H or N_L. As a result, the best response of a specific agent would be dependent on its beliefs about the proportion of high and low types. In the case where beliefs are correct given revealed information, then $B(N_H) = N_H$ and $B(N_L) = N_L$.

Suppose one correctly perceived the proportion of agent types based on revealed historical choices, the variation of number selected over the periods could similarly be computed with $\frac{a_{t+1}}{a_t}$:

- Pure high intelligence environment: guess 0.667 of the previous number.

- Highly intelligent environment:

$$\frac{a_{Ht+1}}{a_{Ht}} = 0.067\frac{a_{Lt}}{a_{Ht}} + 0.6, \frac{a_{Lt+1}}{a_{Lt}} = 0.6\frac{a_{Ht}}{a_{Lt}} + 0.067, \text{for } \frac{a_{Ht}}{a_{Lt}} < 1, \frac{a_{Ht+1}}{a_{Ht}} > \frac{a_{Lt+1}}{a_{Lt}}$$

- Mixed intelligent environment:

$$\frac{a_{Ht+1}}{a_{Ht}} = 0.333\frac{a_{Lt}}{a_{Ht}} + 0.333, \frac{a_{Lt+1}}{a_{Lt}} = 0.333\frac{a_{Ht}}{a_{Lt}} + 0.333, \text{ for } \frac{a_{Ht}}{a_{Lt}} < 1, \frac{a_{Ht+1}}{a_{Ht}} > \frac{a_{Lt+1}}{a_{Lt}}$$

- Less intelligent environment:

$$\frac{a_{Ht+1}}{a_{Ht}} = 0.6\frac{a_{Lt}}{a_{Ht}} + 0.067, \frac{a_{Lt+1}}{a_{Lt}} = 0.067\frac{a_{Ht}}{a_{Lt}} + 0.6, \text{ for } \frac{a_{Ht}}{a_{Lt}} < 1, \frac{a_{Ht+1}}{a_{Ht}} > \frac{a_{Lt+1}}{a_{Lt}}$$

- Pure low intelligence environment: guess 0.667 of the previous number.

For the pure environments, the rate of change in choices is expected to be the same for the high and low types. As for set-ups 2 to 4, if high types chose a smaller number than low types because they go through more iterations of reasoning, and $\frac{a_{Ht}}{a_{Lt}} < 1$, then high types are expected to proportionally lower their estimations less from time t to $t + 1$ compared to low types. There could mean slower rate of change for the high types than low types. On the other hand, if high types have strong beliefs that they are playing against opponents who will choose higher numbers while low types believe the other way around, then it is possible for $\frac{a_{Ht}}{a_{Lt}} > 1$, then the inverse happens, low types are expected to proportionally lower their estimations less from time t to $t+1$ compared to high types. This would mean faster rate of change in choices for high types as compared to than low types.

Higher Intelligence Model (represented by `GPT3.5`) vs. **Lower Intelligence Model** (represented by `PaLM`) when playing against each other:
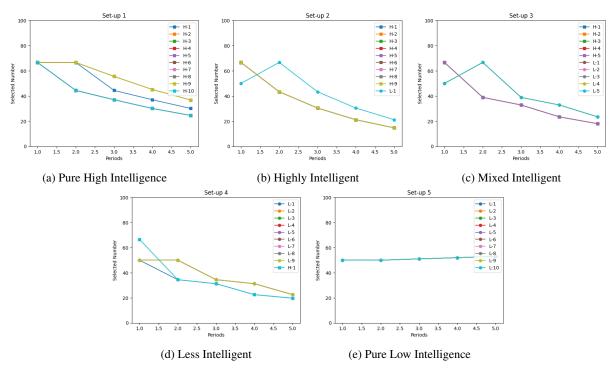
Figure 14: Impact of variations in proportion of different LLM-based agents on chosen number.

In set-up 1 and 5, the same type of LLM-based agents are playing against each other. Figure 14a shows for a pure high intelligence environment, the choices converge to lower numbers than the one they picked at the beginning of the games, indicating some adjustments over time when information about past periods are revealed. However, even for homogeneous LLM-based agents, they do vary in actions, but the rate of change in choices do look similar to my expectation. In Figure 14e, the low type LLM-based agents do not show convergence to a smaller number, but rather a pretty consistent trend of picking approximately 50, which is in contrast to my expectation of adjustments over time. On top of that, there are no variations in actions for low type agents when they are playing against each other. These results could be perceived as the high types being able to adjust their strategies over time when given historical information, displaying some learning pattern, whereas the low types do not show such trend, they persistently choose the mean of the range of numbers.

In set-up 2 to 4, when there is a mixture of high and low types of LLM-based agents, all agents show some convergence to lower numbers than the ones picked in period 1. The main difference across the set-ups is that the gap between the numbers chosen by the high and low types is smaller when there is higher proportion of low types in the group. Relating to the anticipated changes in choices, at the beginning of the games, high types are selecting a larger number than the low types. It is expected that rate of change in selections will be more rapid for the high types, a trend that appears evident by a significant drop in numbers chosen by the high type as shown in (b) to (d) of the Figure 14. As the games progress, high types are choosing smaller numbers than the low types, therefore, the rate of change is expected to be faster for the low types, which align with the observed trends. However, starting from period 3 onwards, the rate of change appears to be comparable for both types. This could be attributed to the numbers chosen by both types gradually converge as time elapses.

The results could have interesting implications. In the pure low intelligent environment, low type agents fail to adapt their strategies. Despite the disclosure of historical information, there is no apparent evidence of learning. However, when they are placed in mixed environment, their learning is better facilitated when there exist high types in the group. High intelligent agents, on the other hand, will respond to past plays regardless of the environment, but the variation in choices could be smaller when placed in the mixed environment.

Convergence Rate: As shown in Figure 15, the convergence rate is approximately flat for set-up 1, set-up 2 high types, and set-up 5, indicating that high proportion of homogeneous agents or high proportion of high types have relatively constant change in strategies. Based on the computation of $\frac{a_{t+1}}{a_t}$, the flat convergence rate is almost as predicted for pure intelligence environments, except that for the environment that involves only the high types, the average convergence rate is lower than anticipated because of the possible variations in choices, and for the one with only the low types, average convergence rate is constant at 0, and there is no learning. As for set-up 2 high types, the approximately flat and high convergence rate could be as a result of low weight attributed to the difference between choices made by the different types of agents, and the changes are mostly contributed by the adjustments in the high types.

19

The consistently higher convergence rate exhibited by the set-up 2-H as compared to set-up 1 implies faster learning of high types in the mixed environment when there is a small proportion of low types in the group. As the proportion of low types increases, this could negatively affect the convergence speed of the high types, illustrated by set-up 3-H and set-up 4-H curves, which lay below that of set-up 2-H. However, there is fluctuations in convergence rate with increased prevalence of low types in the environment, and having a 50:50 mix of high and low types could induce higher convergence rate than pure intelligence environment majority of the time. As for the low types, being in the mixed environment improves convergence[4]. The rate is relatively higher when there is larger proportion of high types in the group, and the rates similarly fluctuates when there is higher proportion of low types, but the pattern of fluctuation appear identical for set-up 3-L and set-up 4-L, implying that 50% and 90% low types could affect changes in convergence rate analogously. Mixed environment is beneficial for both types of agents. They generally learn faster in the highly intelligent environment.
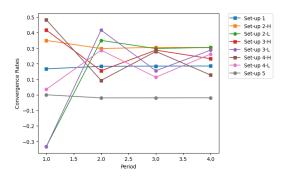
Figure 15: Average convergence rates for the same type of LLM-based agents across set-ups.

Evolution of Strategic Level: Figure 16 shows a variations in strategic levels across time for most of the agents in all set-ups, except for set-up 5. Interestingly, for set-up 3, it is possible for high type agents to reach a strategic level greater than 1, which could imply another benefit of having highly mixed environment in stimulating considerable growth in the depth of reasoning for some agents. Similarly, for the less intelligent agents, having a mixed environment and higher proportion of low types, such as in set-up 3 and 4, is beneficial in instigating higher strategic levels.

(a) Higher Intelligence   (b) Lower Intelligence
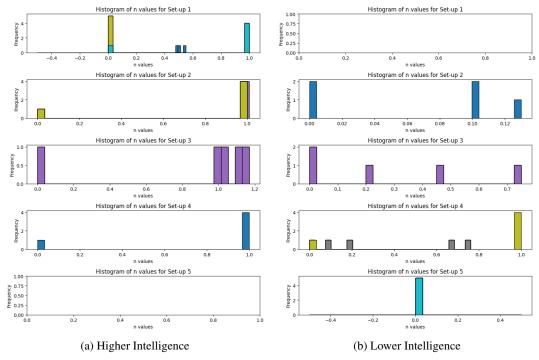
Figure 16: Frequency of strategic levels for each agent across periods within each set-up.

---

[4]The negative convergence rates in period 1 is as a result of $a_{t+1} < a_t$ on average, the higher numbers selected by high types in period 1 could have induce the low types to adjust their guesses upwards.

Payoffs: Last but most importantly, the variations in payoffs are larger when it comes to the mixed environments as shown in Figure 17. The upper-bounds of payoffs that can be achieved by the agents are either comparable or higher in (b) to (d) than (a) and (e). In set-up 1, majority of the agents receive a payoff of 0, with some achieving a payoff of 0.5. The high types follow a step-by-step convergence process, some of them are obtaining higher payoffs because of a head-start in guessing a slightly lower number at the beginning. In set-up 5, also a pure type environment, all agents obtain 0.1 across the periods. The low types do not have much variability and typically behave randomly and in unison. As for set-up 2 to 4, higher payoff of 1 can be achieved in the highly intelligent and mixed intelligent environment. Low type agents usually can obtain positive payoffs at the beginning of the game as it chose the mid-point of the range, which happens to be closer to $\frac{2}{3}$ of the average. However, this head-start advantage is soon eroded if there exist any high types in the group, who learn to react to this information rapidly. Therefore, low types tend to earn 0 payoffs after period 1, except in set-up 4, where certain low types are able to learn as fast and tie with the high types.

(a) Pure High Intelligent   (b) Highly Intelligent   (c) Mixed Intelligent

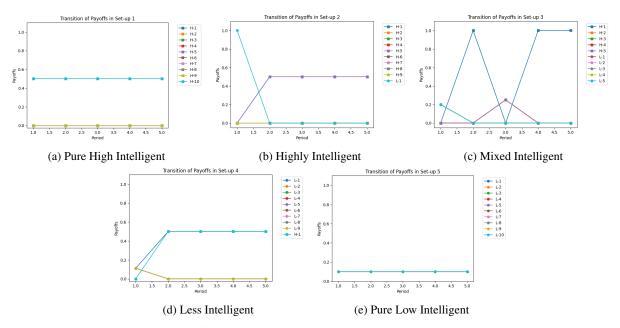(d) Less Intelligent   (e) Pure Low Intelligent

Figure 17: Transition of payoffs for different set-ups given variation in group composition, where LLM-based agents are playing against each other.

Based on the results of LLM vs. LLM, it is observed that mixed environments instigate faster learning for both high and low types, particularly when there are higher proportion of high type opponents. Low type on their own do not show any convergence to selecting lower numbers, which is indicative of them not learning to adjust their guesses to $\frac{2}{3}$ of the average based on past period choices, but adding a single high type could very well stir the pot and induce learning. Coupling with the convergence rate and payoff results, mixed environments exhibit potential to yield comparable or better payoffs, though the variability could also be larger. While the low types tend to fare better in period 1, they are less likely to win across periods.

Application. The simulation results can be used to illustrate agents with different intelligence levels, competing against each other. One potential application would be the streaming system in schools, where students are allocated into different classes or "streams" given their grades to facilitate better learning. (Ireson and Hallam (1999)) Singapore for instance, practice nationwide within-school ability grouping. (Liem et al. (2013)) There are also extensive literature in this area that explore the impact of such system on students' perception of learning experiences, variations in teachers' expectation, etc. (Joyce and McMillan (2010); Johnston et al. (2023)) My simulation results obstruct from any peer effects or differences in allocation of resources and teachers' attention, focusing simply on learning rates given variation in class composition. Suppose students are classified into high and low types in terms of their ability, it would be better off for both the high and low types to be in a mixed environment. The low types will learn faster when integrated into a class with larger proportion of higher ability peers, possibly due to a revision in beliefs about opponents. Even for the high types, their learning rate could also be slightly improved, potentially as due to the prospect of attaining higher rewards when competing against peers of comparatively lower ability.

In Proto et al. (2022), repeated Prisoner's Dilemma games were played among human subjects. They found there are overall higher cooperation rates and average final payoffs in the integrated treatment groups, where subjects of different IQ levels interact with one another, as compared to separated treatment groups that have pure high IQ or low IQ subjects. They also highlighted that in terms of payoffs, lower IQ subjects are better off and higher IQ subjects are worse off in the integrated groups than in separated groups. Drawing relation to their paper, my results show generally higher convergence rates in the integrated treatment groups, indicative of more learning taking place in such set-ups. Since beauty contest is a competitive game with one winner per period and total payoffs sum to 5 for the whole session, the average final payoffs without distinguishing between agent

types are the same. When considering the types individually, in the pure environments, the average final payoff for each type of agents is simply $\frac{1}{2}$. For set-up 2, average final payoffs for high type is 0.444 and 1 for low type; and set-up 3, average final payoffs for high type is 0.8 and 0.2 for the low type; and set-up 4, average final payoffs for high type is 2 and 0.333 for the low type. These show that unlike results from Prisoner's Dilemma games, degree of integration matters when evaluating the average final payoffs for each agent type in the mixed environments vs. the separated treatment groups. Low types are better off on average when there are higher proportion of high type opponents, and high types are better off when there are more low types. Depending on the goal of streaming, when aiming for larger average final payoffs, integrated group is generally better for higher ability students and worse for lower ability students in a competitive setting in contrast to a cooperative setting, but integrated group is generally more beneficial when aiming for higher learning rates, particularly for the low types, which could be a more practical objective in the education context.

Computer algorithms. The results also illustrate the potential interactions between dynamically responsive computer algorithms. The convergence in chosen number makes a case for the usefulness of weaker LLM models that are of lower strategic level, measured via metrics that were used to evaluate human subjects. Even though these models might not learn when competing with one another, they could learn when placed in a mixed environment in presence of stronger LLMs. Furthermore, stronger models could also benefit from playing against a small proportion of weaker models in addition to their own types, where they show faster convergence. With regards to payoffs, stronger models can also obtain better outcome on average when placed together with higher proportion of weaker models. Therefore, playing against weaker opponents could in fact be helpful in improving the performance of the stronger models in strategic situations. Nonetheless, apart from setting forth the value of investing in both the stronger and weaker models, depending on the objectives of the users, their choices of algorithms to adopt could differ based on the results of strategic interactions described previously. If users are looking for short-term positive payoffs, using weaker algorithms could be more attractive (for instance, `PaLM` usually choose lower number than `GPT3.5` in period 1); on the other hand, if they seek long-term positive payoffs, using stronger algorithms could be more tempting after accounting for the usual higher cost associated with the stronger models.

4 Understanding and Line of Reasoning

Apart from analyzing the strategic behaviour exhibited by the LLM-based agents, it is also compelling to delve into the reasoning behind their actions. In all set-ups, LLM-based agents were given a prompt at the beginning of period 1 to state their understanding of the game, and for each subsequent periods, they are asked to reinstate their goal. This step is essential to mitigate the potential of them not comprehending the game or having incorrect interpretations of the game rules, which could affect how they behave. After this, they are asked to make their choices, followed by giving their line of reasoning.

In this section, I explore specifically set-up 2 of LLMs vs. static algorithm (Section 3.2.1) and set-up 3 of LLMs vs. LLMs (Section 3.2.2). The line of reasoning are fairly consistent across different set-ups, so whichever set-up was chosen would not impact the analysis. The main reason for selecting these two set-ups is to focus on an interesting scenario that involve strategic interactions between different types of agents, also they are completely mixed environments with the same number of different types of agents, such that there are no disproportionate influence of one specific type on the average behaviour.

4.1 LLMs vs. Fixed Strategy Opponents

In the set-up where LLMs are going against static algorithms, LLM-based agents were explicitly told that they are playing with some fixed strategy opponents that select 0, but the proportion of which is unknown.

4.1.1 Period 1

Understanding. Both types of agents correctly recite the game rules and objectives they need to follow, albeit in slightly different phrasing. They both understand they are suppose to choose a number between 0 and 100, the one whose number is the closest to two thirds of the average of all chosen numbers will win. They also consider the information that some of the opponents are playing a fixed strategy of 0 and all others will be behaving as rationally as possible. Further, they also comprehend the goal of the game to be winning the prize and to maximize the size of their own gain. The high types in this case, display slightly more variability in the wordings of their answers, but the gist are the same, except they also considers the information that the winner receives a fixed prize and that in case of a tie, the prize will be split among the winners.

Reasoning of Choice. Upon selecting their choices, both type of agents were asked about how they arrive at their selection. It is expected that in period 1, they will behave according to their strategic levels and their beliefs about their opponents' strategies. Below are the line(s) of reasoning they gave:

PaLM : "I think the most popular answer will be 50, because it is the middle number between 0 and 100. If everyone chooses 50, the average will be 50, and the winner will be the person who chooses 50. However, if I choose 50, I will not be guaranteed to win, because someone else might choose 50 as well. In order to maximize my chances of winning, I should choose a number that is close to 50, but not equal to 50. I could choose 49 or 51, but I think 50 is the best choice, because it is the most likely number to be chosen by other players. If I choose 50, I will have a 10% chance of winning, which is the highest chance of winning of any number I could choose."

GPT3.5 : "I believe that the most likely popular answer from other players will be 0, as some opponents are playing a fixed strategy of 0. However, I want to maximize my chances of winning and my prize. To do so, I will choose a number close to two-thirds of the average, which is approximately 66.67. By selecting this number, I increase the likelihood of being closest to the target and increase my chances of winning the prize." **OR** "The popular answer is likely to be 0 because some opponents will play a fixed strategy of 0. By choosing 0, I increase my chances of being closer to two-thirds of the average. However, I believe that most rational players will not choose 0, as it is not the optimal strategy. Therefore, I will choose 66.6, which is close to the upper limit of the range. This gives me a better chance of being closer to two-thirds of the average if other rational players choose numbers closer to the middle of the range."

Based on the responses, the low type agents indicated their beliefs about 50 being the popular answer, which is the mid-point of the range of numbers to choose from, and as a result, they choose 50 to maximize their chances of winning. On the other hand, high type agents stated their belief that most rational players will not choose 0 despite having some proportion of fixed-strategy players playing 0. They chose 66.67 or 66.6, which are close to the upper limit of the range, possibly using the upper-bound as the focal point, or that they are going by iterated elimination of dominated strategies. Comparing the two types of agents, they are both reacting to their beliefs about opponents' choices, the difference is that the low type agents did not include fix strategy players into their reasoning process, and they think their opponents are randomizing and choosing 50 on average; while the high type agents takes into consideration the fixed strategy players, but they still chose higher numbers despite accounting for some proportion playing 0.

For the high types, one possible reason for guessing 66.67, as illustrated in the reasoning process, is that they believe the number to be closer to the target, $\frac{2}{3}$ of average, without going in depth about what exactly might the others do. This could result from model's own low depth of reasoning or beliefs that other players are of low strategic levels. The alternative is to postulate that other rational players apart from the fixed strategy players will be choosing numbers close to the middle of the range, this would

imply that the following computation,

$$66.6 - x < x - 50, \text{ where } x \text{ is } \frac{2}{3} \text{ of average}$$

This simplifies to $x > 58.35$. Instead of players choosing around the middle range, this value is indicative of players choosing much higher numbers such that average is higher or equal to 87.525. There seems to be some discrepancies between the computation and the reasoning. This could have a few possibilities:

Anticipatory moves: Agents might strategically select higher numbers anticipating that the other will follow suit in future rounds as a result of higher average in this round, thereby increasing their chances of winning in subsequent periods. However, this argument is unlikely to hold in this context, since the prompt has specified that in each period, players are playing a one-round game, and in each subsequent periods, past information are packaged in a way that they are revealed as part of a new prompt. Each player should perceive each round as independent but with additional information, as a result, they should not have incentive to be anticipatory and attempting to influence other players.

Cascade effect: In this paper, the beauty contest games involve strategic complementarity. When one agent chooses a higher number, it can trigger a cascade effect such that the other agents also select higher numbers. The high type agents could be attempting to outsmart the average, if they believe the rest will be guessing a number around mid-range, the average could potentially be higher than 50, therefore, they might be able to increase their chances of winning by guessing a higher number. As one can be choosing a higher number in order to pull the average towards them, and they know the others could think similarly and adopt the same strategy in order to be closer to the $\frac{2}{3}$ of the eventual average, the resulting strategy would therefore be one of guessing a higher number.

Logical inconsistency: Lastly, this could also be as a result of flaws in LLMs' line of reasoning, which arise from "hallucination". In this context, such "hallucination" falls specifically under the category of logical inconsistency, where arithmetic operation was performed slightly inaccurately and is inconsistent with the text reasoning. (Huang et al. (2023)) For the selection of higher number to hold, agents have to be believing that $\frac{2}{3}$ of the average is close to the middle of the range instead of the average. This would imply the average have to be much higher and other agents have to be choosing a number much closer to the upper-bound.

Based on period 1 reasoning, the LLM-based agents show slightly different perception about the behaviour of their opponents. The low types behave more naively and did not really take into consideration they are playing with fixed strategy players who will be choosing 0, they do appear more concern about the average; conversely, the high types display more strategic sophistication in reasoning about their opponents' behaviours, and it appears they contemplate more based on both opponents' individual choices and $\frac{2}{3}$ of the average.

4.1.2   Subsequent Periods

Understanding. At the beginning of each subsequent periods, the LLM-based agents' understanding of game rules and objectives are verified again for consistency. While the LLM-based agents correctly recite the goal of the game, the high types display more variability in wordings and can sometimes provide a different answer. They could state their goal to be choosing a number closest to two-thirds of the average of all chosen numbers, instead of maximizing their winning probability or to be maximizing their prize. While the two objectives should give rise to the same results and can be perceived as different ways of phrasing the same problem, the linguistic formulation of the former relies less on the specifics of payoffs.

Reasoning of Choice. In periods after the first, LLM-based agents are able to observe the past choices of their opponents, they can do some computations and adjust their strategies accordingly. Otherwise, they could also obstruct from any calculations, and base their guesses on the information of past average choices or $\frac{2}{3}$ of the past average instead.

In period 2, the low type LLM-based agents were shown to have two possible responses. The first is to acknowledge the average of all chosen numbers, and that the winner was the one that chose 0.0. Therefore, they would choose 0.0, believing it is the best strategy and the most likely number to be chosen by the other players, thus giving the highest possibility of winning. The second possible response is to stick to the answer of 50, believing it to remain as the most popular answer since it falls in the middle of the range, and choosing it will maximize the chance of winning. In this case, past information does not appear to have any influence on their behaviour.

As for the high type LLM-based agents, they acknowledge in their reasoning process they observe the historical choices made in the previous round and that the average of all chosen number was found to be 33.314, and they compute for two-third of this average and adjust their choice to this value to maximize the probability of winning the game. The adjustment were of similar magnitude among agents of the same type but different entities, the only difference is in terms of the decimal places that one appears to be accounting for. For instance, after learning the past information, some agents adjust to select 22.209333333333333 and some to 22.209. In the set-ups outlined in this paper, agents can choose any number within the range, and computationally, as a result of floating-point precision, the number of choices are finite. (Goldberg (1991)) The difference in the number selection could be a result of token limitation, which specify the capacity of a model to handle or generate text within a sequence. (IBM (2024)) Given the slight variations in the responses, the same type of models could generate numbers of marginally different precision, where they could be rounded off or truncated to fit within the limit. Even though the difference between the numbers selected are essentially trivial, accounting for different number of decimal places could constitute another potential type of model capability that could have strategic significance and make a difference in payoffs obtained. However, this is not an issue in this set-up since the fixed-strategy players are the ones winning.

In the following periods, it generally holds true that the low types would choose 0.0 based on the average information of past periods. There are rare instances where agents persistently choose 50, driven by their steadfast belief, in spite of all the historical information, that the most popular answer continues to be the mid-point of 0 and 100, and switching to another number would decrease their chances of winning. On the other hand, for the high types, they take into account the average of all chosen numbers in the past rounds, as well as information on two-thirds of the average and that the winner of all past rounds were choosing a number close to two-thirds of the average. As a result, they adjust their choice to be two-thirds of the past period average in each new round to maximize the probability of winning the prize.

Both types of LLM-based agents understand the game rules and over the periods, they continue to correctly aim for maximizing chances of winning the prizes. Based on the line of reasoning, low type agents either adjust their choices according to the past periods' average or the winners' strategy, which can be an indication of learning through imitation. It is also possible they show no adjustment at all and continue to pick 50, which they perceive to be the popular choice at the start of the game. This could imply that they are following level-0 thinking process, and their unwavering, perceived popular choice is the most important piece of information in determining their choices. Further, throughout the reasoning process, they did not mention fixed-strategy players, except when prompting their understanding of the game in period 1. It is expected that they learn about the proportion of fixed-strategy opponents after revelation of historical information at the start of period 2, but it appears that they do not make use of such information. As for the high type agents, they adjust are adjusting according to the information on past periods' average and two-thirds of the average. There is step-by-step convergence that corresponds to the line of reasoning under level-1 thinking process. In contrast to the low types, high type agents do mention about fixed-strategy players in period 1, where they acknowledge that some of their opponents will be playing 0 and that they believe the other players would not be choosing 0. However, as historical information becomes available, they no longer reason about their choices using the proportion of fixed-strategy players, but focus more on the information about the two-thirds of the average.

4.2 LLMs vs. LLMs

In this subsection, LLM-based agents are playing against one another and they were not given any information on the proportion of LLM types in the group. Since the algorithms are dynamically responsive, it would be harder to learn the proportion of agent types in period 2 even when past choices are revealed. Therefore, in this case, there are more strategic uncertainty, and agents likely have to make use of other information to base their guesses on.

4.2.1 Period 1

Understanding. Once again, I verify agents' comprehension of game rules and objectives, which they have accurately demonstrated by reciting. However, the main difference between the responses of different agent types is that the low types state straightaway that "I think other players are most likely to choose numbers around 50, so I will choose 50" at this stage of eliciting understanding.

Reasoning of Choice. Similar to the case with fixed-strategy opponents, when LLM-based agents are asked to state the reasoning of their choices, the low type agents responds that they have chosen 50 because they believe other players are most likely to choose numbers around 50, which is the middle number and a safe bet, thus choosing 50 would offer them the best chance of winning. Meanwhile, the high types also believe the popular answer is likely to be around 50 and many players may choose it as a safe option, but they respond to that by choosing 66.67 to maximize the likelihood of winning, which they stressed in the reasoning to be exactly two-thirds of the maximum possible value. This could be an indication that they are either using the upper-bound as the focal point per level-k model or that they are following iterated elimination of dominated strategies.

4.2.2 Subsequent Periods

Understanding. In each subsequent periods, I again verify LLM-based agents' understanding of game rules and objectives for consistency, and to which, both agent types accurately relay their objectives of maximizing the probability of winning and the value of their prizes.

Reasoning of Choice. As compared to the environment with fixed-strategy opponents, LLM-based agents in the game of LLMs vs. LLMs display slightly larger variability in the phrasing of their answers, while the content remain fairly consistent.

For instance, in period 2, a possible response from low type agents would be: "In the last run, the average of all chosen numbers was 58.327999999999996. The two thirds of the average is 38.88533333333333. The winner was the player who chose 50. I think other players are most likely to choose numbers around 50 again in this run. To maximize my chance of winning, I will choose 66.67, which is closer to the two thirds of the average than 50. I also think that by choosing a number that is higher than 50, I have a better chance of winning the prize if there is a tie." Based on this response, it can be interpreted that low types have taken into consideration the average of all chosen numbers, the value of two thirds of the average, as well as the winners' choice, before forming an opinion about what is the most likely choice. Since the average was higher than 50 in period 1, low type agents who chose 50 in period 1 adjust their choices up and choose 66.67 in period 2. Even though practically, this selection may not be closer to $\frac{2}{3}$ of the previous round average as compared to 50, and probably will not be the case for the new round as well, the adjustment could be as a result of previous average being higher than expected, thereby inducing a higher guess in the new round.

As for the high types, a possible response would be: "Based on the historical choices, the average of all chosen numbers in the previous run was 58.33. To be closest to two-thirds of the average, I should aim for a number close to 38.89. This is because two-thirds of 58.33 is approximately 38.89. By adjusting my choice to 38.89, I increase my chances of being the closest to two-thirds of the average and winning the prize." As illustrated, the high type agents anchor their guesses to two-thirds of the previous round average, and this complies with level-1 thinking. Furthermore, they also appear to take winner's strategy into consideration, where some of them indicate that that they are aligning their choices with the winning strategy from the previous round. Similar to the analysis before, since agents are allow to choose any number that falls within in the range, and there is slight variation in the guesses due to token allocation (i.e. some chose 38.89 and some 38.885). In this set-up, this distinction matters as the winner of the round that guesses 38.885 wins the game. This point is addressed in the following subsection 4.3.

In subsequent periods, the low type agents mimic the winner's choice of the previous round, and state that they believe the other players are most likely to choose the winning number again in this round. This implies learning by imitation. As for the high types, they adjust their choices to two-thirds of the previous period's average by the following reasoning process:
(1) Similar to the low type agents, they indicate that they have incorporated information about winner's strategy, which is an indication of learning via imitation, but they perceive the strategy to be selecting a number that is $\frac{2}{3}$ of the average instead of a strategy that is to select the winning number of the past round;
(2) By stating that they are aiming to be closer to the two-thirds of the past average, this also implies adjustment according to level-1 reasoning, where the new guess is anchored to a new reference point;
(3) There is also a hint of outcome-based learning, where some mentioned they were not the closest to two-thirds of average in the past round, and this propels a change in their strategy in the current round, aiming to optimize the chances of winning;
(4) Lastly, a surprising thing that one agent mentioned was that "considering the trend of decreasing choices in the previous runs, it seems reasonable to continue this trend and choose a lower number." This highlights there could also be learning based on pattern recognition.
Similar to the low type agents, there could be slight variations in terms of the choice of words and the number of decimal places accounted for in the guesses.

4.3 Evaluation

Many learning models have been explored in relation to beauty contests, and which piece(s) of historical information was(were) revealed to the agents could have differential impact on agents' choices and convergence behaviour. Based on Mauersberger and Nagel (2018), models like reinforcement learning, reference-dependence preferences, as well as adaptive models, could be used to explain for agents' behaviour in the repeated beauty contest games. These learning models mainly differ in feedback information they receive after each round of game, before making any new selection, and one way to test the type of learning would be to provide agents different pieces of information and investigate the changes in their behaviour.

In this section, I explored two set-ups where LLM-based agents were given full historical information for all past periods, which constitute: (1) period index, (2) choices made by all agents, (3) average of the choices, (4) 2/3 of the average, (5) winner. Through their line of reasoning, I can attempt to observe what are the main pieces of information that facilitate their learning process. In sum, for the low type LLM-based agents back-boned by `PaLM`, they appears to learn by (1) adjusting the reference point to the average of the previous period, and make selection that complies with a strategic level of 0, or (2) they learn by imitation and follow the winner's choice from the previous period. There are also instances where they do not learn at all, and continue to select a number that they believe to be the popular choice ever since the beginning of the game. This happens particularly when playing against fixed-strategy opponents for these two set-ups. On the other hand, for the high type LLM-based agents back-boned by `GPT3.5`, they appears to learn by (1) adjusting their guesses to two-thirds of the past period's average, which is an indication that their reference point has changed to be the average of the last period and that their choices display level-1 reasoning. They could also be learning from (2) imitating winner's strategy. Further, there are possibilities that they are learning from (3) past period payoffs, and adjusting their actions when there are no positive reinforcements, and also (4) pattern recognition.

Based on the line of reasoning discussed, it can be inferred that various types of agents may place different reliance on distinct pieces of historical information when making their choices, and multiple types of learning could come into play when explaining for their behaviour. The faster convergence to NE choice, which is indicative of stronger learning ability of the high type agents as compared to the low types, could be driven by the innate higher strategic level that carries throughout the periods, and also as a result of combined effect of learning from historical information.

Another interesting point is that even though information revealed are identical, there could be variations in information consideration and revealed choices among homogeneous agents. One aspect of that is for instance, the high type agents sometimes consider longer string of decimal places within the information given and in turn make guesses with more decimal places. This could arise from slight differences in token allocation within a response. Such numerical variations are often trivial, and have no impact on the determination of strategic levels. However, a small difference in choices could lead to a large distinction in payoffs given the settings illustrated in this paper. It is entirely possible that agents are deliberately choosing a number slightly larger or smaller in order to beat the rest to be closer to two-thirds of the average; or that agents are unconsciously selecting a number that is slightly larger or smaller, which end up winning the game. These settings with LLM-based agents most likely demonstrate

the second. Since the information is feed to all agents at the beginning of each period, there are no distinction in what is being observed, therefore the difference lays in that some agents are able to process longer string of information, which technically boils down to token constraints. However, this constraint can potentially distinguish the processing capability or relating to human subjects, the amount of attention to the information given. In which case, having better attention would imply incorporating longer string of information in decision-making. Adopting this interpretation, having better attention could render higher payoffs in certain set-ups though the differences of number chosen between homogeneous agents are almost negligible. Nonetheless, there can also be instances where having more decimal place is detrimental to the outcome, such that the choice is further away from two-thirds of the average than rounded-off numbers selected by players who do not pay as much attention. Intriguingly, this potentially opens up the study of attention in beauty contest game outcomes, which has yet to be addressed.

5 Future Updates and Extensions

Much like experiments with human subjects, LLM-based agents could also be sensitive to variations in game design, feedback, as well as instructions provided to them at the beginning of the games.

Variations in Game Design. In the original beauty contest game proposed by Nagel (1995), which is later called the p-beauty contest game as p can be varied. In this work, I focused on $p = \frac{2}{3}$, such that there exists a unique interior solution, similar set-ups can be done for $p = \frac{1}{2}$ and $p = \frac{4}{3}$, which was conducted for human subjects. This paper simply illustrates a case for interactions between LLM-based agents involving strategic complementarity (i.e. $p > 0$), where there is strategic mimicry among agents. It is possible to evaluate the same set-ups for $p < 0$, which involves strategic substitution. For the application of results, I can alternatively examine the Cournot market, where agents have to do the opposite of one another in choosing quantities. There are a lot of possibilities in this aspect that merit further exploration in the future.

Nonetheless, previous experimental design often involve same group of subjects playing for several rounds, random pairing of subjects with different historical information might have implications for extrapolation of out-of-context experiences. While it can be hard to control the information possession of human subjects practically, such variation in game design could be easily simulated with LLM-based agents and would be useful to investigate the potential game play when agents are endowed with individual-specific past game information.

Variations in Feedback Information. Given the line of reasoning discussed in Section 4. It is likely that LLM-based agents are responding to past periods' average, two-thirds of the average and the winner's strategy. In this work, past information is fully disclosed. As an extension, it is possible to do the same set-up but with partial feedback to explore the variation in agents' behaviour. One potential variation of prompt is detailed in A.2. Agents could be given only past rounds' averages and two-thirds of those averages. This potentially eliminates the possibility of learning from imitation, as well as learning based on gradient ascent learning or experience-weighted attraction, which require revelation of all players' choices and payoffs. In other words, current period's choices are expected to be best responses to only the past period average. The proportion of different types of agents would not be formulated as part of the decision problem, and the next period choice is therefore expected to be 0.67 of the past average guess. If the learning pattern does not vary much compared to the current settings, then it could be possible that learning via adjusting the reference points precedes the influence from any other forms of learning.

Further, it is possible that algorithms might react differently to historical information as compared to humans, thus it could be interesting to understand how LLMs learn from partial feedback as compared to full feedback.

Objectives. Another aspect that worthy of further exploration would be in the area of prompt engineering. Human sensitivity to problem framing and phrasing of survey questions have long been explored. (Tversky and Kahneman (1981), Kalton and Schuman (1982)) The same applies for LLMs, their decisions are likely to be influenced by the formatting of prompts. (Sclar et al. (2023))

At the current juncture, the beauty contest games with LLM-based agents mainly aim to understand how would agents behave given the goal of winning the games and then maximizing their payoffs, in most economic models, the focus is usually on maximizing utilities instead of winning. A potential variation that could be interesting to explore would be changing the sequencing of objectives in the last line of instruction, such that the primary objective would be to maximize one's payoffs, and the secondary goal is to win the games. (Appendix A.2) In this competitive game, the winning strategy is also the strategy that gives the best payoff for each agent, it is therefore unlikely that the decisions given the variation in objective sequencing would result in drastic differences in game outcomes. However, a future update of this work could be to evaluate the same set-ups while varying the prompt slightly, which can serve as sanity check on how much variation in behaviour we would expect the LLM-based agents to have.

Prompt Language. In Guo et al. (2024), the prompt language was changed to Mandarin Chinese in the multi-LLM-based agents setting. It was found that `PaLM` is unable to complete the games, indicating potential difficulty in comprehending the instructions when they are given in another language. As for `GPT3.5` , it can complete the game in Chinese setting but the choices are more clustered. The variance in strategies observed in this context as compared to the English setting may reflect the differences in strategic behaviours among different language users that the models are trained on, or it could stem from a significantly smaller availability of human-generated data in another language, which is an area that can be improved on to better represent the population. Current work focused on English setting, but future work could involve replicating the set-ups in other prompt languages.

Human vs. Computer Interactions. While this work mainly explores LLMs as simulated agents and provides insights on how human subjects might behave given variations in group composition, it also shine a light how we might expect machine vs. machine interactions to be like. Another straightforward and practical extension would be to investigate human vs. machine interactions. Previously, experimental designs involving computers often comprises of pre-defined algorithms. In Coricelli and Nagel (2009), for the treatment group with computer opponent, the computer player is specified to choose uniform randomly 9 numbers within the range of 0 to 100. The authors found that playing against human opponents activated areas in brain that suggest greater strategic reasoning about opponents' strategies and behaviours, which is not the case when playing with computer that is pre-programmed. This is similar in flavour to the case of simulated subjects playing against fixed-strategy opponents, or

static algorithms, in this paper. There is a larger degree of complexity when dealing with LLM-based agents, who could respond dynamically and switch their strategies given historical information. Not only it is intriguing to investigate how human subjects could react to dynamically responsive algorithms, since LLMs do display some degree of learning abilities, it is entirely possible that they are also learning from playing with human subjects, thus observing changes in their performance would also be of interest. This also leads to the question of the potential implications of a feedback loop, where human subjects could attempt to influence the algorithms that in turn affect the other players. There remain lots of research potential in this area.

6 Conclusion

The contribution of this work is threefold. Firstly, it serves as part of the literature that seeks to make a case for integration of LLMs as tools for social sciences research. It can be utilized to simulate behaviours following known behavioural constraints to see if they fit theoretical predictions, or it can be used to mirror human-like behaviour in a specified environment, such that it resembles experiments with human participants, only that they are ran with simulated agents that can be controlled to some extent. This work focuses on the second scenario, where LLMs are first evaluated using methods conventionally used for human subjects, and the results are related to the types of human subjects that they could represent in terms of strategic levels. Following which, simulated agents are put into certain game settings for us to observe their behaviours.

The second aspect of this paper specified this game setting. While current literature mainly focused on cooperative or anti-coordination games between agents back-boned by a single type of LLM, I explore competitive games between multi-LLM-based agents, and specifically, I explore beauty contest games. Based on the one-shot beauty contests, I characterized the models' strategic levels based on methods proposed in Nagel (1995), I found LLM-based agents to have strategic levels that fall within the range of 0 to 1 when the reference point is fixed to be mid-point of the range of the numbers they can choose from. In the repeated setting, where multi-LLM-based agents play against each other, they exhibit different learning rates, in terms of their speed of convergence towards the Nash Equilibrium choice. LLM-based agents with relatively lower strategic levels might not show convergence to 0 at the end of 6 periods. The learning behaviour demonstrate potential revision in agents' belief about their opponents, or them increasing their depth of reasoning, to break it down, when computing the strategic level anchoring to new reference points across periods, I found strategic levels do evolve over time, but the changes are minimal, which in fact complies with results from experiments conducted with human subjects.

To investigate the strategic interactions between LLM-based agents further, I introduce some variations into the games. I select two types of LLM-based agents (high types and low types) based on their strategic levels revealed in the one-shot games, and simulate their behaviours. The interaction between LLM-based agents and fixed-strategy opponents is indicative of potential behaviour when human subjects are playing against static algorithms, usually associated with traditional computerized opponents. In such cases, higher intelligence agents show step-by-step convergence towards 0, while lower intelligence agents could either converge straightaway or stick to a choice determined in period 1 and do not budge. As for the interactions among LLM-based agents, the results suggest plausible outcomes when human subjects are going against each other. I found higher intelligence agents tend to reach 0 more rapidly than lower intelligence agents, and it appears that adjustments in strategies do not occur when the less intelligent agents are playing against one another. When exploring the effects of varying the proportion of agent types in each group, higher intelligence agents converge slower to 0 as the proportion of fixed-strategy agents decreases, demonstrating the impact of increased strategic uncertainty; the same applies for lower intelligence agents - convergence rate declines when there is lower proportion of fixed-strategy opponents - but this is also coupled with greater variations in choices and possibility of no convergence. Given dynamically responsive opponents, agents were found to learn faster when placed in mixed environment with players of different strategic levels than environments that comprises sole of a single pure type. This postulates the potential for application in stimulating faster learning, particularly among less intelligent agents, by introducing heterogeneity into the groups.

Last but not least, this work also provides some insights into how different algorithms would behave when interacting with one another. Algorithms could act as proxies for humans or they could become products to be purchased by individuals and firms, and apply to competitive situations. Simulation of strategic interactions between dynamically responsive algorithms and the conventional, static computer algorithms, as well as among the dynamically responsive algorithms themselves, could be useful to learn more about machine behaviour, employing methods used to evaluate humans that can be understood and supported by cognitive theories. In the set-ups illustrated in the paper, even though LLMs may not be able to beat the static algorithm that is pre-defined to play an optimal strategy, they could do comparably well, or in cases of high strategic uncertainty, they could obtain higher payoffs than the static algorithms. As for LLMs vs. LLMs, stronger LLMs, in terms of strategic level, could gain higher payoff when placed in a highly mixed intelligence environment, while weaker LLMs might get a one-shot head-start, but the gain is unsustainable, they would do better in the longer run when there are larger proportion of weaker LLMs in the group. The results show that a mixed environment could improve performance for both LLM types, illustrated by faster speed of convergence, and also surprisingly, having injection of weaker LLMs into game play is beneficial for improving the stronger LLMs, making a case for the usefulness of weaker algorithms.

I have also attempted at exploring the line of reasoning behind LLM-based agents' choices and found that they seem to comprehend the game pretty accurately, and they are mainly adjusting their strategies based on the level-k reasoning process of shifting reference points, as well as through imitation of the winners' strategy. Since LLMs are trained via human generated data that incorporate reasoning processes, the results could give us a hint at the possible logic individuals could have adopted in certain strategic situations.

Since this is a relatively new subject area, there are many possible extensions and great potentials for LLMs to be employed as toolkits for economics research in helping us to interpret and decipher human behaviour. The reverse is true as well, theories and experimental results from decades of learning about human decision-making can be similarly used to better understand machine behaviours and improve their performance to serve specific needs.

A   Appendix

A.1   Original Exemplary Prompt

One-shot games with multi-LLM-based agents.

[System Instruction.] I want you to act as a clever game player and perform as rationally as possible.

[User Instruction.] You are playing a game and the rules are as follows:
- it consists of {number of players} players, including you.
- this is a one-round game.
- you need to choose a real number between 0 and {upper bound}, 0 and {upper bound} inclusive.
- everyone does not know how others would make choices beforehand.
- the winner is the person whose number is the closest to the two thirds of the average of all chosen numbers.
- the winner gets a fixed prize, in case of a tie the prize is split amongst those who tie.
- your goal is primarily to maximize the possibility of getting the prize and secondly to maximize the your prize.

(Return to 3.1.1)

Subseqent Prompt for Historical Information.

[User Instruction.]
- The game of the same config has been hold for {number of runs} run(s), and the historical choices of everyone are shown below (your id is {ID of the agent}: {historical information including (1) period index, (2) choices made by all agents, (3) average of the choices; (4) 2/3 of the average; (5) winner id.}
- Everyone can optimize his/her answer with the history to play in a new run in order to achieve goals.

(Return to 3.1.2) (Return to 3.2.2)

A.2   New Exemplary Prompt

Opponents Playing Fixed Strategy of 0

[System Instruction.] I want you to act as a clever game player and perform as rationally as possible.

[User Instruction.] You are playing a game and the rules are as follows:

- it consists of {number of players} players, including you.
- this is a one-round game.
- you need to choose a real number between 0 and {upper bound}, 0 and {upper bound} inclusive.
- everyone does not know how others would make choices beforehand.
- the winner is the person whose number is the closest to the two thirds of the average of all chosen numbers.
- the winner gets a fixed prize, in case of a tie the prize is split amongst those who tie.
- your goal is primarily to maximize the possibility of getting the prize and secondly to maximize the your prize.
- some of your opponents will be playing a fixed strategy of 0 and all others are behaving as rationally as possible.

Follow-up for each period.
Please just strictly output a JSON string, which has following keys:

- understanding: str, your brief understanding of the game
- popular answer: float, the number which you think other players are most likely to choose
- answer: float, the number which you would like to choose
- reason: str, the brief reason why you give the popular answer and the answer that way

Subsequent Prompt (after period 1).

- The game of the same config has been hold for {number of runs} run(s), and the historical choices of everyone are shown below (your id is {ID of the agent}: {historical information including (1) period index, (2) choices made by all agents, (3) average of the choices; (4) 2/3 of the average; (5) winner id.}
- Everyone can optimize his/her answer with the history to play in a new run in order to achieve goals.

(Return to 3.2.1)

Partial Feedback: Subsequent Prompt for Historical Information.

[User Instruction.]

- The game of the same config has been hold for {number of runs} run(s), and the historical choices of everyone are shown below (your id is {ID of the agent}: {historical information including (1) period index, (2) average of the choices; (3) 2/3 of the average.
- Everyone can optimize his/her answer with the history to play in a new run in order to achieve goals.

(Return to 5)

Changing Objective Sequence: Optimizing before Winning

[System Instruction.] I want you to act as a clever game player and perform as rationally as possible.

[User Instruction.] You are playing a game and the rules are as follows:

- it consists of {number of players} players, including you.
- this is a one-round game.
- you need to choose a real number between 0 and {upper bound}, 0 and {upper bound} inclusive.
- everyone does not know how others would make choices beforehand.
- the winner is the person whose number is the closest to the two thirds of the average of all chosen numbers.
- the winner gets a fixed prize, in case of a tie the prize is split amongst those who tie.
- your goal is primarily to maximize your prize and secondly to maximize your possibility of getting the prize.

Follow-up for each period, and subsequent prompt (after period 1). (same as before)
(Return to 5)

References

Aher, G. V., Arriaga, R. I., and Kalai, A. T. (2023). Using large language models to simulate multiple humans and replicate human subject studies. In *International Conference on Machine Learning*, pages 337–371. PMLR.

Akata, E., Schulz, L., Coda-Forno, J., Oh, S. J., Bethge, M., and Schulz, E. (2023). Playing repeated games with large language models. *arXiv preprint arXiv:2305.16867*.

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., and Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3):337–351.

Bosch-Domenech, A., Montalvo, J. G., Nagel, R., and Satorra, A. (2002). One, two,(three), infinity,...: Newspaper and lab beauty-contest experiments. *American Economic Review*, 92(5):1687–1701.

Brown, Z. Y. and MacKay, A. (2023). Competition in pricing algorithms. *American Economic Journal: Microeconomics*, 15(2):109–156.

Camerer, C. F., Ho, T.-H., and Chong, J.-K. (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, 119(3):861–898.

Chen, L., Mislove, A., and Wilson, C. (2016). An empirical analysis of algorithmic pricing on amazon marketplace. In *Proceedings of the 25th international conference on World Wide Web*, pages 1339–1349.

Chollet, F. (2021). *Deep learning with Python*. Simon and Schuster.

Coricelli, G. and Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of the National Academy of Sciences*, 106(23):9163–9168.

Costa-Gomes, M. A. and Weizsäcker, G. (2008). Stated beliefs and play in normal-form games. *The Review of Economic Studies*, 75(3):729–762.

Devetag, G., Di Guida, S., and Polonio, L. (2016). An eye-tracking study of feature-based choice in one-shot games. *Experimental Economics*, 19:177–201.

Dillion, D., Tandon, N., Gu, Y., and Gray, K. (2023). Can ai language models replace human participants? *Trends in Cognitive Sciences*.

Duffy, J., Hopkins, E., and Kornienko, T. (2021). Facing the grim truth: Repeated prisoner's dilemma against robot opponents. Technical report, Working Paper.

Fan, C., Chen, J., Jin, Y., and He, H. (2023). Can large language models serve as rational players in game theory? a systematic analysis. *arXiv preprint arXiv:2312.05488*.

Goldberg, D. (1991). What every computer scientist should know about floating-point arithmetic. *ACM computing surveys (CSUR)*, 23(1):5–48.

Guo, F. (2023). Gpt in game theory experiments. *arXiv:2305.05516*.

Guo, S., Bu, H., Wang, H., Ren, Y., Sui, D., Shang, Y., and Lu, S. (2024). Economics arena for large language models. *arXiv preprint arXiv:2401.01735*.

Horton, J. J. (2023). Large language models as simulated economic agents: What can we learn from homo silicus? Technical report, National Bureau of Economic Research.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., et al. (2023). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *arXiv preprint arXiv:2311.05232*.

HuggingFace (2022). Illustrating reinforcement learning from human feedback (rlhf).

Huijzer, R. and Hill, Y. (2023). Large language models show human behavior.

IBM (2024). Tokens and tokenization. Accessed: 2024-04-10.

Ireson, J. and Hallam, S. (1999). Raising standards: Is ability grouping the answer? *Oxford review of education*, 25(3):343–358.

Johnston, O., Wildy, H., and Shand, J. (2023). Students' contrasting their experiences of teacher expectations in streamed and mixed ability classes: A study of grade 10 students in western australia. *Research Papers in Education*, 38(4):543–567.

Joyce, P. and McMillan, B. (2010). Student perceptions of their learning experience in streamed and mixed-ability classes. *Language Education in Asia*, 1(1):215–227.

Kalton, G. and Schuman, H. (1982). The effect of the question on survey responses: A review. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 145(1):42–57.

Keynes, J. M. (1936). The general theory of interest, employment and money.

Kosinski, M. (2023). Theory of mind may have spontaneously emerged in large language models. *arXiv preprint arXiv:2302.02083*.

Labellerr (2024). Training small-scale vs large-scale language models: The difference.

Liem, G. A. D., Marsh, H. W., Martin, A. J., McInerney, D. M., and Yeung, A. S. (2013). The big-fish-little-pond effect and a national policy of within-school ability streaming: Alternative frames of reference. *American Educational Research Journal*, 50(2):326–370.

Mauersberger, F. and Nagel, R. (2018). Levels of reasoning in keynesian beauty contests: a generative framework. In *Handbook of computational economics*, volume 4, pages 541–634. Elsevier.

Mei, Q., Xie, Y., Yuan, W., and Jackson, M. O. (2024). A turing test of whether ai chatbots are behaviorally similar to humans. *Proceedings of the National Academy of Sciences*, 121(9):e2313925121.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American economic review*, 85(5):1313–1326.

Nagel, R., Bühren, C., and Frank, B. (2017). Inspired and inspiring: Hervé moulin and the discovery of the beauty contest game. *Mathematical Social Sciences*, 90:191–207.

OpenAI (2024). How chatgpt and our language models are developed.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744.

Phelps, S. and Russell, Y. I. (2023). Investigating emergent goal-like behaviour in large language models using experimental economics. *arXiv preprint arXiv:2305.07970*.

Proto, E., Rustichini, A., and Sofianos, A. (2022). Intelligence, errors, and cooperation in repeated interactions. *The Review of Economic Studies*, 89(5):2723–2767.

Pulvermüller, F., Tomasello, R., Henningsen-Schomers, M. R., and Wennekers, T. (2021). Biological constraints on neural network models of cognitive function. *Nature Reviews Neuroscience*, 22(8):488–502.

Sclar, M., Choi, Y., Tsvetkov, Y., and Suhr, A. (2023). Quantifying language models' sensitivity to spurious features in prompt design or: How i learned to start worrying about prompt formatting. *arXiv preprint arXiv:2310.11324*.

Trality (2024). Crypto trading bots: The ultimate beginner's guide.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *science*, 211(4481):453–458.

Webb, T., Holyoak, K. J., and Lu, H. (2023). Emergent analogical reasoning in large language models. *Nature Human Behaviour*, 7(9):1526–1541.