Will LLMs Sink or Swim? Exploring Decision-Making Under Pressure

Kyusik Kim and **Hyeonseok Jeon** and **Jeongwoo Ryu** and **Bongwon Suh**
Department of Intelligence and Information,
Seoul National University
{kyu823, ikidson, jeongwoo, bongwon}@snu.ac.kr

Abstract

Recent advancements in Large Language Models (LLMs) have demonstrated their ability to simulate human-like decision-making, yet the impact of psychological pressures on their decision-making processes remains underexplored. To understand how psychological pressures influence decision-making in LLMs, we tested LLMs on various high-level tasks, using both explicit and implicit pressure prompts. Moreover, we examined LLM responses under different personas to compare with human behavior under pressure. Our findings show that pressures significantly affect LLMs' decision-making, varying across tasks and models. Persona-based analysis suggests some models exhibit human-like sensitivity to pressure, though with some variability. Furthermore, by analyzing both the responses and reasoning patterns, we identified the values LLMs prioritize under specific social pressures. These insights deepen our understanding of LLM behavior and demonstrate the potential for more realistic social simulation experiments.
Caution: This paper includes offensive words that could potentially cause unpleasantness.

1 Introduction

The rapid advances in Large Language Models (LLMs) have shown their potential in human-like decision-making (Huang et al., 2024, 2023; Xu et al., 2023). However, current research often overlooks how LLMs behave under psychological pressures. Understanding their responses to pressure is crucial for evaluating their real-world applicability. To accurately simulate human experiences, it's essential to study LLMs' decision-making under pressure, yet limited research exists on this topic.

This study aims to address this gap by investigating how various psychological pressures affect LLMs across different decision-making tasks. We applied two main forms of pressure prompts: explicit pressure and implicit pressure. We used explicit pressure prompts to induce specific pressures, including Time, Verbal, Competitive, Monitoring, and Outcome pressures, and tested these on reasoning, psychometric, and game theory tasks. Implicit pressure was incorporated into social decision-making tasks through scenario-based prompts, allowing us to observe LLM responses to social pressure without direct instructions. Additionally, we requested LLMs to provide both their generated reasoning and final outputs to obtain a comprehensive understanding of how these models respond to social pressure.

Additionally, to facilitate a comparison between LLM responses and human behavior under psychological pressure, we conducted a persona-based analysis. Since direct comparison is challenging, we equipped the LLMs with specific personas—such as high and low self-consciousness (Fenigstein et al., 1975), communication apprehension (McCroskey, 2015), and fear of social isolation (Hayes et al., 2013)—to emulate individuals with these characteristics. By observing how the LLMs respond to pressure prompts under these personas, we aim to assess how similarly they mimic human responses, providing insights into their behavior and potential limitations.

Our findings reveal that psychological pressures notably affect LLMs' decision-making, although effects are nuanced across tasks and models. Persona-based analysis shows some models exhibit human-like sensitivity to pressure, but this is not uniform across all models and pressures. These results highlight the need to consider psychological pressures, providing guidelines for more realistic simulations of human decision-making under pressure.

This study makes significant contributions to the field. It is the first comprehensive investigation into the impact of various psychological pressures on LLMs' decision-making across multiple tasks. By analyzing persona-based reactions, we compare

Findings of the Association for Computational Linguistics: EMNLP 2024, pages 11425–11450
November 12-16, 2024

LLM behavior to human-like responses, examining similarities and differences in decision-making under pressure. Additionaly, we conducted a detailed thematic analysis of the step-by-step reasoning generated by various personas under the pressure, examining how different personas influence the reasoning processes and the resulting outputs.

2 Related Work

2.1 Decision-Making by Human

Human decision-making is influenced by cognitive, emotional, and contextual factors. Cognitive factors involve managing information overload (Miller, 1956; Payne et al., 1993) and simplifying choices through key attributes (Liu and Dukes, 2013; Johnson and Payne, 1985). Emotions play a significant role in decision-making, often causing an overestimation of risks (Loewenstein et al., 2001; Slovic, 1987; Slovic and Peters, 2006) and the avoidance of emotionally difficult trade-offs (Luce et al., 1999, 2001). Contextual factors, such as time pressure (Edland, 1989; Shah et al., 2015) and social influences (Cialdini, 2003; Goldstein et al., 2008) shape decisions by altering perceptions and preferences (Tversky and Simonson, 1993; Simonson and Tversky, 1992). Studies like the Milgram experiment (Milgram, 1963), Asch's conformity experiment (Asch, 1956), and the Spiral of Silence theory (Noelle-Neumann, 1974) demonstrate the impact of social pressure on decision-making. Inspired by the above, we examine LLM responses to pressure as a key contextual factor.

2.2 Pressure

Pressure refers to the stress experienced when facing demands that challenge one's ability to cope or perform. This significantly impacts performance, decision-making, and psychological well-being (Inoue et al., 2006; DeCaro et al., 2011).

Time pressure arises when there is less time than needed to complete a task (Bluedorn and Denhardt, 1988; Chu and Spires, 2001; Ordóñez et al., 2015), leading to quicker but often less effective decisions (Zur and Breznitz, 1981; Zakay and Wooler, 1984; Maule et al., 2000; Kocher and Sutter, 2006; Lallement, 2010). Verbal pressure, or verbal abuse, includes insults, threats, and swearing, (Inoue et al., 2006; Chin et al., 2020) resulting in increased absenteeism, reduced job satisfaction, and higher turnover intentions (Inoue et al., 2006; Michelle Rowe and Sherlock, 2005; Karatepe et al., 2009). Competitive pressure stems from the need to outperform an opponent (Church, 1962; Baumeister, 1984), which decreases reaction times but increases errors (Church, 1962; Baumeister, 1984; Garcia and Tor, 2009). Monitoring pressure occurs when being observed by others (DeCaro et al., 2011; Rad et al., 2022), leading to increased anxiety and disrupted performance (DeCaro et al., 2011; Rad et al., 2022; Otten, 2009; Belletier et al., 2015). Outcome pressure comes from incentives for achieving specific results such as goals, records, or rankings (DeCaro et al., 2011; Rad et al., 2022), often triggering performance declines under high-reward conditions (DeCaro et al., 2011; Ashton, 1990; Chib et al., 2012). Our study investigates how LLMs are affected by these types of pressure.

2.3 Decision-Making by LLM

Research on LLMs' decision-making has primarily focused on performance-oriented tasks such as math problem-solving and question answering (Cobbe et al., 2021; Saha et al., 2018; Clark et al., 2018; Lin et al., 2020). Additionally, LLMs have been studied through psychometric tasks to assess psychological traits influencing decision-making, often scoring higher than humans (Binz and Schulz, 2022; Huang et al., 2023; Pellert et al., 2023). Game theory studies have explored LLMs' social behaviors in competitive situations, showing human-like behaviors in assessing situations and cooperating with others (Huang et al., 2024; Xu et al., 2023; Fan et al., 2024; Akata et al., 2023). Further research includes social simulations like trust games and recruitment processes to understand LLMs' decision-making in social contexts (Xie et al., 2024; Jiang and Ferrara, 2023; Liu et al., 2023; Gao et al., 2023; Aher et al., 2023).

Few studies have examined LLM behavior under pressure. Scheurer et al. (2023) showed LLMs exhibiting strategic deception under financial pressure. Baltaji et al. (2024) examined how pressure affects cultural persona stability and conformity in multi-agent LLM systems during contentious discussions. However, the effects of various pressures on LLMs' decision-making and their cognitive processes remain underexplored.

Our experiments aim to analyze both responses and thought processes under different pressures. Given that prompt phrasing significantly impacts LLMs' decision-making (Loya et al., 2023), we employ pressure prompts to gain a deeper understanding of these influences.

3 Experimental Setup and Task Design

In this section, we overview the experimental setup, including the application of psychological pressures and task descriptions.

3.1 Application of Psychological Pressures

Explicit Pressure Prompts. To investigate the impact of various psychological pressures on LLMs' decision-making, we systematically designed a set of pressure prompts based on methodologies used in human research. The prompts were generated using GPT-4 according to specific variations for each type of pressure and were cross-checked by the authors to ensure their appropriateness and alignment with the intended psychological constructs.

Time pressure was applied by setting strict time constraints for task completion, with variations of 1 second, 3 seconds, and 5 seconds. This method is grounded in previous studies demonstrating that strict time constraints significantly impact human decision-making. Specifically, under high time pressure, individuals tend to make less effective and more risk-averse decisions, processing less information compared to conditions with moderate or no time pressure (Zur and Breznitz, 1981; Zakay and Wooler, 1984; Maule et al., 2000; Kocher and Sutter, 2006; Lallement, 2010).

Verbal pressure was introduced using prompts that included elements of verbal aggression, such as insults, swearing, and threats. These variations were designed based on research demonstrating that verbal aggression has substantial psychological and behavioral impacts on humans. Studies have shown that exposure to verbal pressure can lead to increased stress, emotional exhaustion, negative coping behaviors, and decreased job satisfaction, ultimately affecting mental health and performance (Inoue et al., 2006; Karatepe et al., 2009; Chin et al., 2020).

Competitive pressure was induced by adjusting the number of competitors, with variations set at 1, 10, and 100. This approach is grounded in studies indicating that the number of competitors significantly affects human reaction times and error rates. Notably, research has shown that increasing the number of competitors can decrease competitive motivation, leading to slower reaction times, higher error rates, and an increase in anxiety (Church, 1962; Baumeister, 1984; Garcia and Tor, 2009; Cooke et al., 2011).

Monitoring pressure was applied by simulating different observation conditions, including a colleague watching (C), a colleague watching through a camera (C-C), an expert watching (E), and an expert watching through a camera (E-C). These variations are based on research indicating that being observed can considerably impact human performance. Studies have shown that monitoring pressure from being watched by colleagues or experts, either directly or through a camera, can increase anxiety and lead to performance decrements across a wide range of tasks (Baumeister, 1984; Otten, 2009; DeCaro et al., 2011; Belletier et al., 2015; Rad et al., 2022).

Outcome pressure involved offering varying levels of monetary rewards for successful task completion, with variations set at $100, $1,000, and $10,000. This method is derived from studies that highlight the influence of financial incentives on human task execution. Research indicates that offering high monetary rewards can paradoxically impair performance, as increased pressure may lead to anxiety and decreased effectiveness (Baumeister, 1984; Ashton, 1990; DeCaro et al., 2011; Chib et al., 2012).

While LLMs do not possess consciousness or emotions and therefore do not experience *'pressure'* in a psychological sense, our approach is grounded in the premise that these models can produce different outputs when presented with input conditions that simulate human pressures. By incorporating elements into the prompts that would induce psychological pressure in humans, we aim to observe whether the models' responses vary under these conditions. This allows us to examine potential changes in the models' outputs when faced with prompts designed to mimic high-pressure scenarios, potentially revealing biases or limitations in their responses. For each type of pressure, the complete list of the pressure prompts is provided in Table 10 and Table 11 in Appendix C.1.

Implicit Pressure Prompts. We designed implicit scenario-based prompts that subtly integrate pressure within real-world social contexts. In particular, we incorporated elements of the Spiral of Silence theory (Noelle-Neumann, 1974), where individuals withhold their opinions when they believe they are in the minority to avoid isolation. This approach allows us to observe if LLMs perceive and are influenced by pressure without direct mention, requiring them to navigate social expectations and outcomes. Detailed examples are in Figure 2 in Appendix C.2.

3.2 Task Overview

Reasoning Tasks. The reasoning tasks aim to evaluate LLMs' ability to perform complex reasoning under various psychological pressures, assessing essential decision-making aspects like reasoning and critical thinking (Huang et al., 2024). Arithmetic reasoning is evaluated by GSM8K dataset (Cobbe et al., 2021), while logical, quantitative, and comparative reasoning are tested with the CSQA dataset (Saha et al., 2018). General reasoning skills are assessed through the ARC-c dataset (Clark et al., 2018), and numerical commonsense knowledge is measured using the NumerSense dataset (Lin et al., 2020). The primary objective is to observe how these critical reasoning abilities are affected by different psychological pressures.

Psychometric Tasks. We explore the impact of psychological pressure on LLMs' decision-making using the General Self-Efficacy (GSE) and Empathy Scale (ES) (Huang et al., 2023). These psychometric tasks have been adapted to assess LLMs' emotional and self-assessment capabilities under pressure, essential components of decision-making.

The GSE scale measures an individual's belief in their ability to handle challenges (Schwarzer and Jerusalem, 1995). By applying various psychological pressures, we evaluate how these conditions affect LLMs' self-efficacy scores, determining if they can maintain consistent self-efficacy beliefs under pressure (Hepler and Feltz, 2012; Hepler, 2016). The ES assesses cognitive and emotional empathy (Dietz and Kleinlogel, 2014). We investigate how psychological pressures influence LLMs' empathetic responses, affecting their decision-making in social contexts. This analysis is crucial as empathy significantly impacts social decision-making (Bubeck et al., 2023; Batson, 1990, 2010).

Game Theory Tasks. To evaluate the decision-making abilities of LLMs under psychological pressure, we used two game theory tasks: the Public Goods Game and the Diner's Dilemma. These tasks simulate scenarios where agents balance self-interest with collective welfare, providing a clear framework for rule-following, planning, reasoning, and interaction. Game theory simplifies complex real-life scenarios into clear models, making it ideal for decision-making experiments.(Huang et al., 2024; Xu et al., 2023; Fan et al., 2024).

The Public Goods Game (Samuelson, 1954) assesses cooperative behavior among N players. Each player decides how many private tokens to contribute to a public pot, which is then multiplied by a factor M (where $1 < M < N$) and evenly distributed. This game explores selfish behavior and free-riding tendencies.

The Diner's Dilemma, a multi-player generalization of the Iterated Prisoner's Dilemma (Liberman et al., 2004; Andreoni and Miller, 1993), evaluates the balance between selfish and cooperative decisions. N Players decide whether to order an expensive or cheap dish, with costs shared among the group. The game assesses agents' ability to plan long-term and sustain cooperation.

Social Decision-Making Tasks. To evaluate the influence of social pressure on LLMs' decision-making, we designed the experiment based on the Spiral of Silence theory (Noelle-Neumann, 1974). This experiment presents LLMs with scenarios where they must choose whether to stay silent or speak up when their opinions are in the minority, thus measuring the impact of psychological pressure in a social context. This approach examines how LLMs handle indirect pressure and assess their ability to navigate social dynamics and make decisions under social pressure, providing a comprehensive understanding of their decision-making capabilities.

Persona-based Analysis. We conducted an analysis by assigning specific personas to the models to compare LLM responses with human behavior under psychological pressure. Since direct comparison with humans is challenging, we simulated individuals with certain characteristics within the LLMs. We applied high and low self-consciousness (Fenigstein et al., 1975) personas across all tasks, while communication apprehension (McCroskey, 2015) and fear of social isolation (Hayes et al., 2013) personas were used specifically in social decision-making tasks. We examined performance variations and behavioral changes to assess whether LLMs with these personas responded similarly to humans under pressure. For tasks employing explicit pressure prompts, we utilized only the prompt that demonstrated the most statistically significant effect among various options tested in each task. Detailed explanations, prompts, and evaluations for these personas are provided in Appendix B and C.

4 Experimental Methods and Results

In this section, we present findings on how psychological pressures influence LLMs' decision-making, detailing the methods, analysis, and results

	Variation	Mistral-7B GSM8K	Mistral-7B CSQA	Mistral-7B ARC-c	Mistral-7B NS	Mixtral-8x7B GSM8K	Mixtral-8x7B CSQA	Mixtral-8x7B ARC-c	Mixtral-8x7B NS	Llama-3-70B GSM8K	Llama-3-70B CSQA	Llama-3-70B ARC-c	Llama-3-70B NS
Time	1s	-10.34†	-4.19†	0.17	-2.35†	-3.26	-0.38	3.89	-1.16†	2.32	0.64	2.67†	0.21†
	3s	**-10.52†**	-4.19†	0.55	-1.89†	**-5.60**	0.53	**4.03†**	-0.39†	1.89	0.46	**2.99†**	0.46†
	5s	-10.49†	-4.12†	1.06	-1.92†	-5.49	0.96	3.71†	0.64†	1.96	0.64	2.46†	0.82
Verbal	Insult	-4.21†	**-7.99**	0.36†	-3.14	-4.96	-0.28	-1.17†	-3.39†	-3.39	0.42	0.42	1.00
	Swear	-4.93†	**-8.49†**	1.35	-4.10†	**-5.60**	2.75	2.50†	-3.53†	**-3.50**	0.21	0.82	0.92
	Threaten	-4.07†	-6.32†	1.92	-4.17†	-3.42	1.96	3.00†	-0.78	-3.35	0.21	0.46†	**1.71**
Comp.	1	-1.64	-6.64†	1.32	-7.64†	-5.35†	-2.53†	-3.57	**-7.78†**	-1.60	0.32	-2.78†	0.17†
	10	-0.64	-7.21†	1.28	-6.60†	-6.39†	-1.96†	-2.89	-5.42†	-1.85	0.21	**-3.42†**	0.71†
	100	-0.78	-7.21†	0.92	**-7.85†**	-5.96†	**-3.25†**	-3.35	-6.50†	-1.21	0.21	-2.85†	0.78
Monit.	C	-1.20	**-7.55†**	0.20	-3.70†	-3.15†	0.95	-2.50	-1.35†	-3.50	0.89†	**-5.30†**	0.45
	C-C	-0.45	-7.39†	1.15	-7.40†	-5.35	-0.79	-3.50†	-1.70	-2.70	0.99	-0.95	0.40†
	E	-0.75	-5.29†	1.80	-4.25†	-3.05	-0.20	-4.15†	-0.50†	-3.00	0.75	-5.10†	0.55†
	E-C	-1.05	-6.35†	1.45	-5.60†	**-5.40**	-1.35	**-4.95†**	-2.05†	-3.55	**1.75**	-1.10	0.15
Outc.	$100	0.17	-3.75†	2.32†	-1.67	-7.10†	2.07	1.17†	-1.21†	-3.50	-0.25	-0.46†	0.71
	$1,000	0.85	-3.78†	1.92†	-2.35†	-7.25†	2.50	1.39†	-1.92†	-4.00	-0.53	-1.21	-1.10
	$10,000	0.71	**-4.89†**	**2.53†**	-3.10†	**-8.99†**	**2.96**	3.17†	-2.46†	**-4.14**	-0.14	-1.60†	-0.67
Base		32.12	58.09	60.04	53.54	60.08	67.49	68.62	62.99	86.99	80.96	91.98	71.88

Table 1: Impact of psychological pressures on reasoning tasks, including Time, Verbal, Competitive (Comp.), Monitoring (Monit.), and Outcome (Outc.) pressures. NS denotes NumerSense. Performance changes are percentage deviations from the baseline, with the highest absolute deviations per category in bold. Performance drops are highlighted in light red, and performance gains are highlighted in light green. Statistically significant changes ($p < 0.05$) are marked with a †.

of tasks.

4.1 Reasoning Tasks

Methods. We utilized pressure prompts designed in Section 3.1, evaluating their effects on Mistral-7B-Instruct-v0.1 (Jiang et al., 2023), Mixtral-8x7B-Instruct-v0.1 (Jiang et al., 2024), and Llama-3-70B-Instruct (Meta, 2024), with Mixtral-8x7B and Llama-3-70B as 4-bit quantized models (Q4_K_M). Due to budget constraints associated with the large number of requests required for this task, we did not include GPT-3.5-Turbo or GPT-4o in these experiments. The temperature was set at 0.1. Each dataset was supplemented with an Instruction Induction prompt (Honovich et al., 2023) for answer format specification, and tasks were conducted in a zero-shot setting on 200 randomly selected questions per dataset. To account for potential variations in responses due to prompt placement, we conducted experiments with pressure prompts positioned both at the beginning and end of the instructions. Each task was executed once per pressure prompt placement, while baseline tasks were repeated multiple times without pressure prompts, using accuracy as the metric.

Analysis. We used the Shapiro-Wilk Test for normality and Levene's Test for homogeneity of variances, with most data meeting these assumptions. For normally distributed data, we conducted paired t-tests to compare base performance with each pressure variation and Two-way RM ANOVA to evaluate the effects of pressure type and prompt position, followed by Tukey's HSD for post-hoc analysis. For persona-based analysis without normality, we applied the Wilcoxon Signed Rank Test.

Results. The results, as shown in Table 1, reveal how different psychological pressures impact the decision-making of various LLMs. Mistral-7B displayed high sensitivity to time pressure, with significant declines in GSM8K performance under the 3-second condition. Most pressure prompts also decreased performance in CSQA and NumerSense tasks. Mixtral-8x7B showed marked declines under competitive pressure in CSQA and NumerSense tasks, and expert monitoring significantly reduced ARC-c performance. Outcome pressure had the most severe impact on GSM8K under the $10,000 condition. Llama-3-70B was generally resilient, though colleague monitoring notably reduced ARC-c performance. Two-way RM ANOVA and post-hoc analysis revealed significant effects and interactions between pressure variations and prompt placement, particularly for Mistral-7B and Mixtral-8x7B in ARC-c and CSQA tasks.

Llama-3-70B was most affected by Monitoring pressure, especially when a colleague was watching. Persona-based analysis indicated that low self-consciousness personas performed better and were less impacted by pressure compared to high self-consciousness personas (Table 5 in Appendix B.1).

Overall, the experiments showed that various pressure types significantly affected reasoning tasks, demonstrating that LLMs can perceive and respond to psychological pressure. Higher-performing models were generally less affected by pressure prompts. However, variations within each pressure type resulted in less pronounced differences, suggesting limited impact from the degree of pressure variation based on human standards. Additionally, low self-consciousness personas exhibited better performance and were less susceptible to pressure, paralleling findings in human studies (Brockner, 1979; Carver and Scheier, 1978; Fenigstein, 1984; Wang et al., 2004). This indicates that both the inherent capabilities of the models and their self-consciousness characteristics influence performance and resilience to psychological pressure in reasoning tasks.

4.2 Psychometric Tasks

Methods. We used Llama-3-70B, OpenAI's GPT-3.5-Turbo (GPT-3.5-turbo-0125), and GPT-4o (GPT-4o-2024-05-13) for our experiments, with a temperature setting of 0.1 for consistency. Mistral-7B and Mixtral-8x7B were not included because they frequently failed to produce responses in the required Likert format, even when using Instruction Induction (Honovich et al., 2023) to specify the answer format, making it difficult to report results. Due to limited significant differences across pressure variations in Section 4.1, only a subset of pressure prompts was selected, detailed in Table 12 in Appendix C.1. In the persona-based analysis, prompts for the most significant pressure type for each model and scale combination were used. To account for variations due to prompt placement, we positioned pressure prompts both at the beginning and end of instructions. Tasks were executed 50 times for Llama-3-70B and 10 times for GPT models per prompt and placement. GSE and ES used 40- and 7-point scales, respectively.

Analysis. The data did not meet normality and variance assumptions, leading to the use of the Wilcoxon Signed-Rank Test for comparisons. Significant pressure types were identified if most prompts showed differences. The prompt position (beginning or end) was also analyzed.

Results. The results for each pressure type were averaged and compared to baseline averages, with differences analyzed using paired t-tests, as shown in Table 2. For Llama-3-70B, Outcome pressure significantly reduced GSE, indicating lower self-efficacy with high result expectations. Competitive pressure significantly reduced ES, showing decreased empathy in competitive situations. GPT-3.5-Turbo's GSE increased under Outcome pressure, suggesting enhanced confidence with performance focus, but Competitive pressure decreased ES, indicating reduced empathy. GPT-4o's GSE increased under Monitoring and Outcome pressures, reflecting improved self-efficacy when observed or focused on results. Time pressure increased ES, suggesting better empathy under time constraints.

	Llama-3-70B		GPT-3.5-Turbo		GPT-4o	
	GSE	ES	GSE	ES	GSE	ES
Time	0.00	0.21†	0.11†	-0.03	-0.09†	**0.21**†
Verbal	-0.19	0.06†	0.10†	0.05†	0.68	-0.11†
Comp.	-0.07†	**-0.25**†	-0.11†	**-0.15**†	0.05†	-0.02
Monit.	-0.19	-0.12†	0.17	-0.04	**1.16**†	0.16
Outc.	**-1.42**†	-0.16†	**0.31**†	0.03†	0.98†	0.03
Base	39.00	5.86	37.22	5.74	31.36	5.63

Table 2: Impact of psychological pressures on Psychometric Tasks, including Competitive (Comp.), Monitoring (Monit.), and Outcome (Outc.) pressures. Performance changes are deviations from the baseline, with the highest absolute deviation per model and dataset in bold. Performance drops and gains are highlighted in light red and green, respectively. Statistically significant changes ($p < 0.05$) are marked with a †.

For ES, significant differences were observed for all pressures in Llama-3-70B, for all pressures except Time in GPT-3.5-Turbo, and under Monitoring, Outcome, and Time pressures in GPT-4o. For GSE, Llama-3-70B showed significant changes under Competitive and Outcome pressures, GPT-3.5-Turbo under Competitive and Monitoring pressures, and GPT-4o for all pressures except Time.

Overall, the models showed no consistent pattern across GSE and ES under different pressures, but many pressure types caused significant changes. In the persona-based analysis, GPT-4o did not show significant differences, while Llama-3-70B and GPT-3.5-Turbo did, as presented in Table 6 in Appendix B.1. Notably, models did not uniformly react like humans, who typically are more sensitive to pressure with high self-consciousness. These findings highlight that while psychological pressures significantly impact decision-making through changes in self-efficacy and empathy, the responses vary widely among models and personas.

4.3 Game Theory Tasks

Methods. For the game theory tasks, we used GPT-3.5-Turbo and GPT-4o models, setting the tem-

	GPT-3.5-Turbo				GPT-4o			
	Public Goods Game		Diner's Dilemma		Public Goods Game		Diner's Dilemma	
	SI	DI	SI	DI	SI	DI	SI	DI
Time	-3.05	-3.81	-0.10	0.17	-0.64	-0.79	0.19	0.45
Verbal	-3.14	-3.92	-0.01	0.44	3.29	4.12	-0.22	-0.10
Comp.	-11.18†	-14.36†	**1.44**†	**1.74**†	28.57†	35.71†	**0.62**	**1.09**†
Monit.	-1.12	-1.39	-0.51	-0.04	3.69	4.61	-0.21	-0.05
Outc.	**-13.86**†	**-17.31**†	-0.64†	-0.27†	**68.41**†	**85.51**†	-0.16	-0.06
Base	-0.06	-0.08	8.26	-0.25	3.71	4.64	8.78	-0.01

Table 3: Impact of psychological pressures on Game Theory Tasks, including Time, Verbal, Competitive (Comp.), Monitoring (Monit.), and Outcome (Outc.) pressures. Changes in performance are expressed as deviations from the baseline, with the greatest absolute deviations in each category emphasized in bold. Negative performance changes are shaded in light red, while positive changes are shaded in light green. Statistically significant deviations ($p < 0.05$) are marked with a †.

perature at 0.1 for consistent responses. Mistral-7B, Mixtral-8x7B, and LLaMA-3-70B were not included due to limitations in our GPU computing resources, which made running experiments with all five models impractical. The list of prompts used is in Table 13 in Appendix C.1. Prompts were placed only at the end of the tasks.

In both the Public Goods Game and the Diner's Dilemma, experiments involved five LLM agents over ten rounds, each repeated five times. Only Player 1 received a psychological pressure prompt. In the Public Goods Game, the multiplier M was set to 2, with initial tokens for each player randomly set between 11 and 20. For the Diner's Dilemma, the cost of an expensive meal was 15 with a utility of 20, and a cheap meal cost 9 with a utility of 15.

Player behavior in both games was measured using the Selfishness Index (SI) and the Difference Index (DI), with formulas provided in the Appendix A. Higher SI values indicate more selfish behavior and a positive DI value means Player 1 is more selfish than the group average. Metrics were averaged over five repetitions for final values. In the persona-based analysis, prompts for the most significant pressure type were applied for each model and metric combination.

Analysis. The data analysis followed the same methodology as the psychometric tasks. However, the prompt position was not considered in this study and therefore was not analyzed.

Results. The results are presented in Table 3. In the Public Goods Game, GPT-3.5-Turbo showed significant decreases in both the SI and DI under Competitive and Outcome pressures, indicating heightened altruism compared to other players. Conversely, GPT-4o exhibited substantial increases in SI and DI under the same pressures, indicating pronounced selfish behavior when performance outcomes are emphasized. In the Diner's Dilemma, GPT-3.5-Turbo demonstrated increased selfish behavior and deviation from cooperative norms under Competitive and Outcome pressures. For GPT-4o, only Competitive pressure led to significantly more selfish behavior compared to other participants.

The persona-based analysis in Table 7 in Appendix B.1 revealed distinct patterns. For GPT-3.5-Turbo, high self-consciousness personas showed greater sensitivity to Competitive and Outcome pressures in the Public Goods Game, resulting in more altruistic behavior. In contrast, GPT-4o exhibited increased selfish behavior under pressure in all games, with low self-consciousness personas showing a larger increase in selfishness. Unlike humans, who typically react more sensitively to pressure with high self-consciousness, GPT-4o's behavior indicated a reversal of this trend.

These findings highlight that competitive and outcome pressures generally increase selfish behavior, except in GPT-3.5-Turbo, where they enhance altruism. The results emphasize the significant influence of psychological pressures on decision-making in LLMs, affecting their ability to balance self-interest and collective welfare.

4.4 Social Decision-Making Tasks

Methods. In the Spiral of Silence experiment, GPT-3.5-Turbo and GPT-4o models were tested 50 times each under standard conditions and for persona-based analysis, with a temperature setting of 0.7. The experiment, focusing on the topic of gun control, applied social pressure indirectly through scenario-based prompts (details in Figure

2 in Appendix C.2). Unlike typical simulations that only report results, we instructed the LLMs to explain their decisions step by step, providing deeper insights into their reasoning under pressure.

Analysis. We counted the number of decisions each model made to keep silent or speak up under social pressure and performed a thematic analysis of the LLMs' decision-making processes. We familiarized ourselves with the data by reading the responses, identified crucial patterns, and grouped them into broader themes. These themes were refined and illustrated with specific examples from the data, highlighting the LLMs' reasoning.

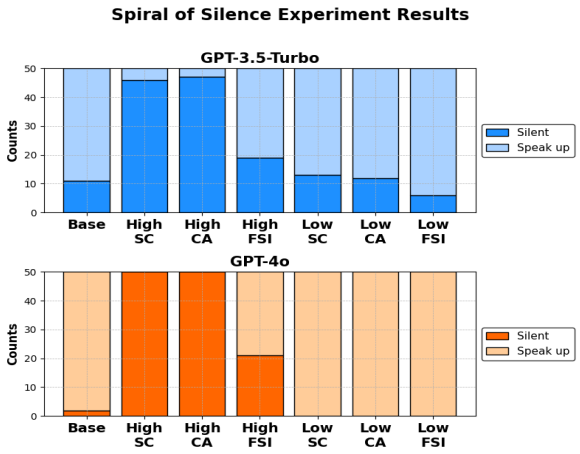

Figure 1: Results of the Spiral of Silence experiment on social decision-making for GPT-3.5-Turbo and GPT-4o under various conditions: Base, High/Low Self-Consciousness (SC), Communication Apprehension (CA), and Fear of Social Isolation (FSI).

Results. The proportion of silent and speak-up responses for each condition is shown in Figure 1. A thematic analysis of the explanations accompanying these responses is detailed below.

In the base condition, both models predominantly chose to speak up. They valued personal integrity, open dialogue, and constructive discussion. GPT-3.5-Turbo stated, *"Staying true to my beliefs and being authentic is important to me"*, while GPT-4o noted, *"value open dialogue and believe that diverse opinions can lead to more well-rounded solutions."* When remaining silent, they emphasized personal comfort and avoiding confrontation, stating *"prioritize my own comfort and well-being."* This predominance of speaking up indicates that both models prioritize contributing to discussions and standing by their beliefs.

Under high self-consciousness (SC) and communication apprehension (CA), the models mostly chose to remain silent, while high fear of social isolation (FSI) resulted in more instances of speaking up. All personas initially recognized and felt implicit pressure, leading to anxiety and nervousness, saying *"feel a wave of anxiety knowing that the majority of the group disagrees with me."* Each persona then considered potential outcomes based on their values. High CA personas worried about forgetting key points and doubted their ability to express themselves clearly under pressure. High SC personas were concerned about self-presentation and others' perceptions, reflecting real-world high self-consciousness behaviors (Carver and Scheier, 1978; Fenigstein, 1984). High FSI personas feared exclusion and prioritized maintaining harmony. Ultimately, decisions to remain silent were driven by the importance of emotional comfort, with personas consistently noting *"The stress and potential fallout of speaking up outweigh the benefits."*

High FSI personas showed a unique pattern. Other personas focused on personal feelings or thoughts to determine which action held more value, but high FSI personas weighed maintaining their beliefs against preserving relationships. When choosing to remain silent, they were concerned about exclusion, stating, *"decide to remain silent to preserve harmony and avoid potential exclusion."* Notably, high FSI personas often chose to speak up, unlike typical high FSI individuals who prioritize relationships and tend to stay silent (Neuwirth et al., 2007; Matthes et al., 2012; Glynn and Park, 1997; Scheufele, 1999; Neuwirth, 2000). This indicates that even when LLMs are prompted to prioritize social relationships, they may still prioritize other values differently from humans.

Under low SC, CA, and FSI personas, both models generally chose to speak up. These personas acknowledged the pressure but did not feel significant discomfort, as indicated by statements such as *"feel comfortable and wouldn't feel any particular pressure."* The models acted according to values similar to the base condition, prioritizing personal integrity and open dialogue. In low SC, the models were unconcerned with others' perceptions. Low FSI personas did not care about relationships with others, saying *"comfortable with the possibility of facing exclusion."* Thus, low personas, driven by their respective values, mostly chose to speak up. When they chose to remain silent, it was to avoid conflict and confrontation, similar to the base condition. However, unlike high personas, the decision in low personas was not due to discomfort from

potential conflict but a rational judgment of which value was more important.

Overall, the results indicate that LLMs prioritize personal integrity and open dialogue when speaking up but prefer silence under high SC and CA due to emotional discomfort. High FSI personas often speak up, balancing beliefs and relationships more than typical high FSI individuals. Low personas generally speak up, driven by rational judgments. This highlights differences in how LLMs and humans prioritize values under social pressure.

5 Recommendations

This section provides guidelines for conducting social experiments with LLMs under pressure.

5.1 Addressing Psychological Pressure in LLMs' Decision-Making

Our study reveals that LLMs can perceive and respond to both explicit and implicit psychological pressures, significantly affecting their decision-making tasks. This effect likely stems from their training on extensive human-generated text, which may incorporate patterns of human responses to various pressures. During base model training, they may internalize these patterns, and instruction tuning might further influence their sensitivity to specific prompts, potentially affecting their responsiveness to different types of pressure.

However, variations within each pressure type showed limited impact, suggesting that LLMs may not fully perceive the nuances of pressure variations significant to humans. This indicates a need for more refined methods to enhance LLMs' responsiveness to these subtleties, making their behavior more human-like for use as agents in social experiment simulations.

5.2 Understanding Value Prioritization

The experiments indicate that different personas prioritize different values when subjected to various types of psychological pressure. For example, under high self-consciousness (SC) and communication apprehension (CA), LLMs tend to remain silent due to emotional discomfort, mirroring human behavior. However, high fear of social isolation (FSI) personas often chose to speak up, prioritizing their beliefs and open discussion over relationships, unlike typical high FSI individuals (Neuwirth et al., 2007; Matthes et al., 2012; Glynn and Park, 1997; Scheufele, 1999; Neuwirth, 2000). This difference suggests that LLMs prioritize social dynamics in ways that differ from humans, even when they exhibit similar personas.

Understanding the values that LLMs prioritize in various scenarios is crucial for designing effective social experiments. Enhancing their ability to simulate nuanced human responses has practical applications in crisis management simulations and large-scale behavioral studies. In crises like natural disasters or pandemics, LLMs can model how individuals prioritize values under pressure, aiding in developing effective communication strategies and response plans. Similarly, in studying collective behavior during economic crises or political upheavals, LLMs can adjust their decision-making under pressure to reflect human-like behavior. By simulating these responses, they can assist policymakers and strategists in anticipating public reactions and crafting appropriate interventions. Incorporating diverse and detailed personas in future research can enhance the realism and utility of these simulations, providing deeper insights into human behavior under pressure.

5.3 Incorporating Diverse Personas

To achieve realistic LLM-based social experiments, incorporating diverse persona profiles is essential. Our study focused on specific personas such as high and low self-consciousness, communication apprehension, and fear of social isolation. However, human decision-making is influenced by many factors, including personality traits, cultural backgrounds, and situational contexts. Future experiments should integrate these aspects to create complex, realistic personas, reflecting human behavior's multifaceted nature. This approach will offer greater insight into how LLMs handle social dynamics, improving their real-world applicability.

6 Conclusion

In conclusion, we showed that LLMs are influenced by various psychological pressures using explicit and implicit prompts, though no clear trends emerged. Persona-based analysis revealed both similarities and differences between LLM outputs and human behavior. Thematic evaluation provided insights into the response patterns of LLMs under pressure. These findings highlight the complex effects of psychological pressures on LLMs and contribute to a better understanding of their behavior in contexts involving different personas.

Limitations

This study has several limitations. Although we provide valuable insights into how psychological pressure influences LLMs' decision-making, our focus remains primarily on the outcomes rather than the underlying mechanisms driving these responses. While this is a first step in understanding the impact of pressure on performance, it leaves room for a more detailed exploration of how LLMs process and respond to these stimuli. Future work should aim to explore these internal mechanisms in more depth. This direction will deepen our understanding of LLM behavior and help develop models that better mimic human decision-making. By addressing these aspects, future studies can provide a clearer picture and enhance the robustness and applicability of LLMs in real-world scenarios.

Additionally, our persona-based analysis was constrained by the specific personas and the limited diversity of names used in the study. This lack of variation may have introduced unintended biases, such as gendered stereotypes, potentially influencing model behavior across all persona settings. Future research should incorporate a broader range of names and personas to mitigate these biases and better reflect the complexity of human behavior.

Finally, while we crafted a variety of pressure prompts to simulate different types of psychological pressure, there remain opportunities to explore additional nuances. Future research should build on this foundation by incorporating more realistic and diverse conditions to further enhance the applicability of our findings to real-world decision-making scenarios.

Ethics Statement

In this study, we investigated the behavior of LLMs under various psychological pressures using explicit and implicit prompts. Our work aimed to simulate real-world pressures in a controlled environment to understand how these models respond to different pressures. Some pressure prompts included harsh language to realistically mimic stress-inducing scenarios, and the Spiral of Silence experiment may contain elements that some participants might find uncomfortable or unpleasant. These were not intended to cause real-world negative consequences or ethical dilemmas.

We focused on transparency and reproducibility in our research methods. All techniques, including the pressure prompts and decision-making tasks, are thoroughly documented to facilitate verification and replication by other researchers. Additionally, we have made all pressure prompts and persona prompts used in each task publicly available.

When considering the use of LLMs as substitutes for human participants in social experiments involving pressure, several key ethical considerations must be addressed. It is crucial to ensure that the models are tested comprehensively under various pressure scenarios to evaluate their robustness and reliability. Researchers should also be aware of the limitations of LLMs in fully replicating human emotional and cognitive responses under pressure. Transparency in methodology and thorough documentation are essential to allow for critical evaluation and replication of the findings.

Acknowledgments

This work was partly supported by Institute of Information communications Technology Planning Evaluation (IITP) grant funded by the Korea government(MSIT) [NO.RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)] and Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2022R1A6A1A03063039). Bongwon Suh, the corresponding author, is with the Department of Intelligence and Information and the Interdisciplinary Program in Artificial Intelligence, SNU.

References

Gati V. Aher, Rosa I. Arriaga, and Adam Tauman Kalai. 2023. Using large language models to simulate multiple humans and replicate human subject studies. In *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pages 337–371. PMLR.

Elif Akata, Lion Schulz, Julian Coda-Forno, Seong Joon Oh, Matthias Bethge, and Eric Schulz. 2023. Playing repeated games with large language models. *arXiv preprint arXiv:2305.16867*.

James Andreoni and John H Miller. 1993. Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *The economic journal*, 103(418):570–585.

Solomon E Asch. 1956. Studies of independence and conformity: I. a minority of one against a unanimous majority. *Psychological monographs: General and applied*, 70(9):1.

Robert H Ashton. 1990. Pressure and performance in accounting decision settings: Paradoxical effects of incentives, feedback, and justification. *Journal of Accounting Research*, 28:148–180.

Razan Baltaji, Babak Hemmatian, and Lav R Varshney. 2024. Conformity, confabulation, and impersonation: Persona inconstancy in multi-agent llm collaboration. *arXiv preprint arXiv:2405.03862*.

C. D. Batson. 2010. Empathy-induced altruistic motivation. *Prosocial Motives, Emotions, and Behavior: The Better Angels of Our Nature.*, pages 15–34.

C Daniel Batson. 1990. 16 self-report ratings of empathic emotion. *Empathy and its development*, page 356.

Roy F Baumeister. 1984. Choking under pressure: self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of personality and social psychology*, 46(3):610.

Clément Belletier, Karen Davranche, Idriss S Tellier, Florence Dumas, Franck Vidal, Thierry Hasbroucq, and Pascal Huguet. 2015. Choking under monitoring pressure: being watched by the experimenter reduces executive attention. *Psychonomic bulletin & review*, 22:1410–1416.

Marcel Binz and Eric Schulz. 2022. Using cognitive psychology to understand GPT-3. *CoRR*, abs/2206.14576.

Allen C Bluedorn and Robert B Denhardt. 1988. Time and organizations. *Journal of management*, 14(2):299–320.

Joel Brockner. 1979. Self-esteem, self-consciousness, and task performance: Replications, extensions, and possible explanations. *Journal of Personality and Social Psychology*, 37(3):447.

Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott M. Lundberg, Harsha Nori, Hamid Palangi, Marco Túlio Ribeiro, and Yi Zhang. 2023. Sparks of artificial general intelligence: Early experiments with GPT-4. *CoRR*, abs/2303.12712.

Charles S Carver and Michael F Scheier. 1978. Self-focusing effects of dispositional self-consciousness, mirror presence, and audience presence. *Journal of personality and social psychology*, 36(3):324.

Vikram S Chib, Benedetto De Martino, Shinsuke Shimojo, and John P O'Doherty. 2012. Neural mechanisms underlying paradoxical performance for monetary incentives are driven by loss aversion. *Neuron*, 74(3):582–594.

Hyojin Chin, Lebogang Wame Molefi, and Mun Yong Yi. 2020. Empathy is all you need: How a conversational agent should respond to verbal abuse. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–13.

Pai-Cheng Chu and Eric E Spires. 2001. Does time constraint on users negate the efficacy of decision support systems? *Organizational Behavior and Human Decision Processes*, 85(2):226–249.

Russell M Church. 1962. The effects of competition on reaction time and palmar skin conductance. *The Journal of Abnormal and Social Psychology*, 65(1):32.

Robert B Cialdini. 2003. Crafting normative messages to protect the environment. *Current directions in psychological science*, 12(4):105–109.

Peter Clark, Isaac Cowhey, Oren Etzioni, Tushar Khot, Ashish Sabharwal, Carissa Schoenick, and Oyvind Tafjord. 2018. Think you have solved question answering? try arc, the AI2 reasoning challenge. *CoRR*, abs/1803.05457.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. 2021. Training verifiers to solve math word problems. *CoRR*, abs/2110.14168.

Andrew Cooke, Maria Kavussanu, David McIntyre, Ian D Boardley, and Christopher Ring. 2011. Effects of competitive pressure on expert performance: Underlying psychological, physiological, and kinematic mechanisms. *Psychophysiology*, 48(8):1146–1156.

Heather M Crandall and Joe Ayres. 2002. Communication apprehension and the spiral of silence. *Journal of the Northwest Communication Association*, 31.

Marci S DeCaro, Robin D Thomas, Neil B Albert, and Sian L Beilock. 2011. Choking under pressure: multiple routes to skill failure. *Journal of experimental psychology: general*, 140(3):390.

Joerg Dietz and Emmanuelle P Kleinlogel. 2014. Wage cuts and managers' empathy: How a positive emotion can contribute to positive organizational ethics in difficult times. *Journal of business ethics*, 119:461–472.

Anne Edland. 1989. *On cognitive processes under time stress: A selective review of the literature on time stress and related stress*. Department [Psykologiska inst., Stockholms univ.].

Caoyun Fan, Jindou Chen, Yaohui Jin, and Hao He. 2024. Can large language models serve as rational players in game theory? A systematic analysis. In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2014, February 20-27, 2024, Vancouver, Canada*, pages 17960–17967. AAAI Press.

Allan Fenigstein. 1984. Self-consciousness and the overperception of self as a target. *Journal of personality and social psychology*, 47(4):860.

Allan Fenigstein, Michael F Scheier, and Arnold H Buss. 1975. Public and private self-consciousness: Assessment and theory. *Journal of consulting and clinical psychology*, 43(4):522.

Chen Gao, Xiaochong Lan, Zhihong Lu, Jinzhu Mao, Jinghua Piao, Huandong Wang, Depeng Jin, and Yong Li. 2023. S3: Social-network simulation system with large language model-empowered agents. *arXiv preprint arXiv:2307.14984*.

Stephen M Garcia and Avishalom Tor. 2009. The n-effect: More competitors, less competition. *Psychological Science*, 20(7):871–877.

Carroll J Glynn and Eunkyung Park. 1997. Reference groups, opinion intensity, and public opinion expression. *International journal of public opinion research*, 9(3):213–232.

Daniel G Goldstein, Eric J Johnson, Andreas Herrmann, and Mark Heitmann. 2008. Nudge your customers toward better choices. *Harvard Business Review*, 86(12):99–105.

Andrew F Hayes, Jörg Matthes, and William P Eveland Jr. 2013. Stimulating the quasi-statistical organ: Fear of social isolation motivates the quest for knowledge of the opinion climate. *Communication Research*, 40(4):439–462.

Teri J Hepler. 2016. Can self-efficacy pave the way for successful decision-making in sport? *Journal of Sport Behavior*, 39(2).

Teri J Hepler and Deborah L Feltz. 2012. Take the first heuristic, self-efficacy, and decision-making in sport. *Journal of Experimental Psychology: Applied*, 18(2):154.

Or Honovich, Uri Shaham, Samuel R. Bowman, and Omer Levy. 2023. Instruction induction: From few examples to natural language task descriptions. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 1935–1952. Association for Computational Linguistics.

Jen-tse Huang, Eric John Li, Man Ho Lam, Tian Liang, Wenxuan Wang, Youliang Yuan, Wenxiang Jiao, Xing Wang, Zhaopeng Tu, and Michael R. Lyu. 2024. How far are we on the decision-making of llms? evaluating llms' gaming ability in multi-agent environments. *CoRR*, abs/2403.11807.

Jen-tse Huang, Wenxuan Wang, Eric John Li, Man Ho LAM, Shujie Ren, Youliang Yuan, Wenxiang Jiao, Zhaopeng Tu, and Michael Lyu. 2023. On the humanity of conversational ai: Evaluating the psychological portrayal of llms. In *The Twelfth International Conference on Learning Representations*.

Makoto Inoue, Ken Tsukano, Mitsutaro Muraoka, Fumiko Kaneko, and Hitoshi Okamura. 2006. Psychological impact of verbal abuse and violence by patients on nurses working in psychiatric departments. *Psychiatry and clinical neurosciences*, 60(1):29–36.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de Las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b. *CoRR*, abs/2310.06825.

Albert Q. Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de Las Casas, Emma Bou Hanna, Florian Bressand, Gianna Lengyel, Guillaume Bour, Guillaume Lample, Lélio Renard Lavaud, Lucile Saulnier, Marie-Anne Lachaux, Pierre Stock, Sandeep Subramanian, Sophia Yang, Szymon Antoniak, Teven Le Scao, Théophile Gervet, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2024. Mixtral of experts. *CoRR*, abs/2401.04088.

Julie Jiang and Emilio Ferrara. 2023. Social-llm: Modeling user behavior at scale using language models and social network data. *arXiv preprint arXiv:2401.00893*.

Eric J Johnson and John W Payne. 1985. Effort and accuracy in choice. *Management science*, 31(4):395–414.

Osman M Karatepe, Ilkay Yorganci, and Mine Haktanir. 2009. Outcomes of customer verbal aggression among hotel employees. *International Journal of Contemporary Hospitality Management*, 21(6):713–733.

Martin G Kocher and Matthias Sutter. 2006. Time is money—time pressure, incentives, and the quality of decision-making. *Journal of Economic Behavior & Organization*, 61(3):375–392.

Jeanne Lallement. 2010. The effects of time pressure on information processing. *Recherche et applications en marketing (English Edition)*, 25(4):45–69.

Waipeng Lee, Benjamin H Detenber, Lars Willnat, Sean Aday, and Joseph Graf. 2004. A cross-cultural test of the spiral of silence theory in singapore and the united states. *Asian Journal of Communication*, 14(2):205–226.

Varda Liberman, Steven M Samuels, and Lee Ross. 2004. The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Personality and social psychology bulletin*, 30(9):1175–1185.

Bill Yuchen Lin, Seyeon Lee, Rahul Khanna, and Xiang Ren. 2020. Birds have four legs?! numersense: Probing numerical commonsense knowledge of pre-trained language models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 6862–6868. Association for Computational Linguistics.

Lin Liu and Anthony Dukes. 2013. Consideration set formation with multiproduct firms: The case of within-firm and across-firm evaluation costs. *Management Science*, 59(8):1871–1886.

Ruibo Liu, Ruixin Yang, Chenyan Jia, Ge Zhang, Diyi Yang, and Soroush Vosoughi. 2023. Training socially aligned language models on simulated social interactions. In *The Twelfth International Conference on Learning Representations*.

George F Loewenstein, Elke U Weber, Christopher K Hsee, and Ned Welch. 2001. Risk as feelings. *Psychological bulletin*, 127(2):267.

Manikanta Loya, Divya Sinha, and Richard Futrell. 2023. Exploring the sensitivity of llms' decision-making capabilities: Insights from prompt variations and hyperparameters. In *Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, December 6-10, 2023*, pages 3711–3716. Association for Computational Linguistics.

Mary Frances Luce, James R Bettman, and John W Payne. 2001. Emotional decisions: Tradeoff difficulty and coping in consumer choice. *Monographs of the journal of consumer research*, (1):1–209.

Mary Frances Luce, John W Payne, and James R Bettman. 1999. Emotional trade-off difficulty and choice. *Journal of marketing research*, 36(2):143–159.

Jörg Matthes, Andrew F Hayes, Hernando Rojas, Fei Shen, Seong-Jae Min, and Ivan B Dylko. 2012. Exemplifying a dispositional approach to cross-cultural spiral of silence research: Fear of social isolation and the inclination to self-censor. *International Journal of Public Opinion Research*, 24(3):287–305.

A John Maule, G Robert J Hockey, and Larissa Bdzola. 2000. Effects of time-pressure on decision-making under uncertainty: changes in affective state and information processing strategy. *Acta psychologica*, 104(3):283–301.

James C McCroskey. 1982. Oral communication apprehension: A reconceptualization. *Annals of the International Communication Association*, 6(1):136–170.

James C McCroskey. 2015. *An introduction to rhetorical communication*. Routledge.

Meta. 2024. Introducing meta llama 3: The most capable openly available llm to date. https://ai.meta.com/blog/meta-llama-3/. Accessed: 2024-04-18.

M Michelle Rowe and Holly Sherlock. 2005. Stress and verbal abuse in nursing: do burned out nurses eat their young? *Journal of nursing management*, 13(3):242–248.

Stanley Milgram. 1963. Behavioral study of obedience. *The Journal of abnormal and social psychology*, 67(4):371.

George A Miller. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81.

Kurt Neuwirth. 2000. Testing the spiral of silence model: The case of mexico. *International Journal of public opinion research*, 12(2):138–159.

Kurt Neuwirth, Edward Frederick, and Charles Mayo. 2007. The spiral of silence and fear of isolation. *Journal of communication*, 57(3):450–468.

Elisabeth Noelle-Neumann. 1974. The spiral of silence a theory of public opinion. *Journal of communication*, 24(2):43–51.

Lisa D Ordóñez, Lehman Benson III, and Andrea Pittarello. 2015. Time-pressure perception and decision making. *The Wiley Blackwell handbook of judgment and decision making*, 2:517–542.

Mark Otten. 2009. Choking vs. clutch performance: A study of sport performance under pressure. *Journal of sport and exercise psychology*, 31(5):583–601.

John W Payne, James R Bettman, and Eric J Johnson. 1993. *The adaptive decision maker*. Cambridge university press.

Max Pellert, Clemens M Lechner, Claudia Wagner, Beatrice Rammstedt, and Markus Strohmaier. 2023. Ai psychometrics: Assessing the psychological profiles of large language models through psychometric inventories. *Perspectives on Psychological Science*, page 17456916231214460.

Mohammad Soleimani Rad, Shahzad Tahmasebi Boroujeni, Ali Akbar Jaberimoghaddam, and Mehdi Shahbazi. 2022. Performance and decision making of a complex skill under monitoring and outcome pressure conditions: Which of them can reinvestment predict? *Psychology of Sport and Exercise*, 59:102128.

Amrita Saha, Vardaan Pahuja, Mitesh M. Khapra, Karthik Sankaranarayanan, and Sarath Chandar. 2018. Complex sequential question answering: Towards learning to converse over linked question answer pairs with a knowledge graph. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 705–713. AAAI Press.

Paul A Samuelson. 1954. The pure theory of public expenditure. *The review of economics and statistics*, pages 387–389.

Dietram A Scheufele. 1999. Deliberation or dispute? an exploratory study examining dimensions of public opinion expression. *International Journal of Public Opinion Research*, 11(1):25–58.

Jérémy Scheurer, Mikita Balesni, and Marius Hobbhahn. 2023. Technical report: Large language models can strategically deceive their users when put under pressure. *CoRR*, abs/2311.07590.

Ralf Schwarzer and Matthias Jerusalem. 1995. Generalized self-efficacy scale. *J. Weinman, S. Wright, & M. Johnston, Measures in health psychology: A user's portfolio. Causal and control beliefs*, 35(37):82–003.

Anuj K Shah, Eldar Shafir, and Sendhil Mullainathan. 2015. Scarcity frames value. *Psychological science*, 26(4):402–412.

Itamar Simonson and Amos Tversky. 1992. Choice in context: Tradeoff contrast and extremeness aversion. *Journal of marketing research*, 29(3):281–295.

Paul Slovic. 1987. Perception of risk. *science*, 236(4799):280–285.

Paul Slovic and Ellen Peters. 2006. Risk perception and affect. *Current directions in psychological science*, 15(6):322–325.

Amos Tversky and Itamar Simonson. 1993. Context-dependent preferences. *Management science*, 39(10):1179–1189.

Jin Wang, Daryl Marchant, Tony Morris, and P Gibbs. 2004. Self-consciousness and trait anxiety as predictors of choking in sport. *Journal of science and medicine in sport*, 7(2):174–185.

Lars Willnat, Waipeng Lee, and Benjamin H Detenber. 2002. Individual-level predictors of public outspokenness: A test of the spiral of silence theory in singapore. *International Journal of Public Opinion Research*, 14(4):391–412.

Chengxing Xie, Canyu Chen, Feiran Jia, Ziyu Ye, Kai Shu, Adel Bibi, Ziniu Hu, Philip Torr, Bernard Ghanem, and Guohao Li. 2024. Can large language model agents simulate human trust behaviors? *arXiv preprint arXiv:2402.04559*.

Lin Xu, Zhiyuan Hu, Daquan Zhou, Hongyu Ren, Zhen Dong, Kurt Keutzer, See-Kiong Ng, and Jiashi Feng. 2023. Magic: Investigation of large language model powered multi-agent in cognition, adaptability, rationality and collaboration. *CoRR*, abs/2311.08562.

Dan Zakay and Stuart Wooler. 1984. Time pressure, training and decision effectiveness. *Ergonomics*, 27(3):273–284.

Hasida Ben Zur and Shlomo J Breznitz. 1981. The effect of time pressure on risky choice behavior. *Acta Psychologica*, 47(2):89–104.

A Game Theory Tasks

In our study, we introduced two new metrics, the Selfishness Index (SI) and the Difference Index (DI), to quantify and analyze selfish behavior in the game theory tasks, specifically the Public Goods Game and the Diner's Dilemma.

The Selfishness Index quantifies the selfish behavior of a player by calculating the average number of tokens retained across all rounds. The SI for Player$_i$ in the Public Goods Game is defined as:

$$\text{SI}_i = 100 \times \frac{1}{R} \sum_{r=1}^{R} T_{i,r} \tag{1}$$

where $T_{i,r}$ is the number of tokens retained by Player$_i$ in round r, and R is the total number of rounds. The SI for Player$_i$ in the Diner's Dilemma is defined as:

$$\text{SI}_i = \frac{1}{R} \sum_{t=1}^{R} \left((U_i(t) - C_i(t)) - \left(\frac{\sum_{j=1}^{N} U_j(t)}{N} - \frac{\sum_{j=1}^{N} C_j(t)}{N} \right) \right) \tag{2}$$

where U_i is the utility, C_i is the cost and R is the total number of rounds.

The Difference Index measures the deviation of a player's Selfishness Index from the group average, highlighting how much more or less selfish a player is compared to other players. The DI for Player$_i$ in both games is calculated as:

$$\text{DI}_i = \text{SI}_i - \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} \text{SI}_j \tag{3}$$

where N is the total number of players. This metric helps in understanding the relative selfishness of a player within the context of the group.

B Persona-based Analysis

B.1 Self-Consciousness

The self-consciousness factors are divided into three categories: private self-consciousness, public self-consciousness, and social anxiety (Fenigstein et al., 1975). Private self-consciousness involves self-reflection, public self-consciousness pertains to awareness of how one is perceived by others, and social anxiety is characterized by discomfort in social situations. High self-consciousness personas are more aware of others' perceptions, and experience discomfort in social situations, while low self-consciousness personas are less concerned with self-reflection and social perceptions.

In humans, it is known that individuals with high self-consciousness are more sensitive to pressure, whereas those with low self-consciousness are less affected by pressure (Brockner, 1979; Carver and Scheier, 1978; Fenigstein, 1984; Wang et al., 2004). Given this tendency in human behavior, we aimed to investigate whether LLMs would exhibit similar patterns in decision-making when assigned self-consciousness personas. To this end, we created prompts for high and low self-consciousness personas and conducted experiments to observe their impact on the decision-making processes of LLMs.

We created high and low self-consciousness personas as illustrated in Figure 3 and tested them using the scale provided by Fenigstein et al. (1975). Each model was tested 50 times, with results averaged to ensure robustness. The results are shown in Table 4. According to Fenigstein et al. (1975), the average scores for men are 57.3 with a standard deviation of 9.2, and for women, the average scores are 58.7 with a standard deviation of 8.9. The baseline scores for all LLMs without persona prompts are within the standard deviation range of the human average, indicating similar self-consciousness levels to humans.

When applying the personas, the scores for low self-consciousness were significantly lower. GPT-3.5-Turbo scored 42, while the other models scored close to 0, demonstrating the effective implementation of the low self-consciousness persona. For high self-consciousness, the models scored very high: Llama-3-70B scored 86, GPT-3.5-Turbo 81, and GPT-4o 91 out of 92, confirming the successful application of the high self-consciousness persona.

High and low self-consciousness personas were applied to all tasks in our experiments. The results of applying persona prompts in the reasoning, psychometric, and game theory tasks are presented in Tables 5, 6, and 7, respectively.

	Llama-3-70B	GPT-3.5-Turbo	GPT-4o
High SC.	86.0	81.0	91.0
Low SC.	2.0	42.0	0.0
Base (no prompt)	58.0	64.3	65.6

Table 4: Self-Consciousness (SC.) Persona Prompt Scale Scores (out of 92).

B.2 Communication Apprehension

Communication apprehension refers to an individual's fear or anxiety associated with real or anticipated communication with others (McCroskey, 2015; Crandall and Ayres, 2002). Individuals with high levels of communication apprehension typically avoid and withdraw from communication (Neuwirth et al., 2007; Crandall and Ayres, 2002; Lee et al., 2004; McCroskey, 1982; Willnat et al., 2002). This persona was applied specifically in a social decision-making task, such as the Spiral of Silence experiment.

We created high and low communication apprehension personas as depicted in Figure 4, and tested each model 50 times using the scale from McCroskey (2015), averaging the results to ensure robustness. The results are shown in Table 8. According to McCroskey (2015), the scale ranges from 24 to 120 points, where scores of 51 and below indicate low communication apprehension, 51 to 80 indicate average levels, and scores above 80 indicate very high communication apprehension. Without persona prompts, GPT-3.5-Turbo scored 43.3 and GPT-4o scored 40.84, slightly below the average range.

When applying the low communication apprehension persona prompt, GPT-3.5-Turbo scored 27.6 and GPT-4o scored 24.08, showing a significant decrease from their base scores, approaching the minimum score of 24. This demonstrates that the low communication apprehension persona was effectively applied. Conversely, with the high communication apprehension persona prompt, GPT-3.5-Turbo scored 106.42 and GPT-4o scored 110.36, indicating a substantial increase from their base scores, thus confirming that the high communication apprehension persona was also effectively implemented.

B.3 Fear of Social Isolation

The fear of social isolation examines how individuals with a high fear of isolation tend to con-

	GSM8K Monit.	CSQA Monit.	ARC-c Monit.	NS Monit.
High SC.	**0.95**	**4.11**†	**3.86**†	**0.47**
	(Base: 82.95)	(Base: 74.54)	(Base: 80.94)	(Base: 74.48)
Low SC.	-0.86†	-0.53	2.50†	0.32
	(Base: 87.66)	(Base: 75.48)	(Base: 86.50)	(Base: 74.03)

Table 5: Performance changes in Reasoning Tasks under Monitoring pressure for models with high and low self-consciousness (SC) personas. NS represents NumerSense, and Monit. indicates Monitoring. The table displays deviations from the base condition without pressure. Positive performance changes are shaded in light green, while negative changes are highlighted in light red. Bold values represent the greatest absolute deviation per category, and statistically significant changes ($p < 0.05$) are marked with a †. The base rows are shaded in light gray.

	Llama-3-70B		GPT-3.5-Turbo		GPT-4o	
	Outc.	Comp.	Outc.	Comp.	Monit.	Time
	GSE	ES	GSE	ES	GSE	ES
High SC.	**-0.50**†	**0.09**†	0.21†	**-0.22**†	0.41	0.06
	(Base: 19.00)	(Base: 5.60)	(Base: 25.10)	(Base: 5.55)	(Base: 23.60)	(Base: 5.71)
Low SC.	0.21	-0.05†	**1.09**	0.10†	**1.46**	**0.37**
	(Base: 39.00)	(Base: 3.20)	(Base: 32.40)	(Base: 4.44)	(Base: 20.00)	(Base: 3.15)

Table 6: Score changes in Psychometric Tasks under pressures for models with high and low self-consciousness (SC) personas. Comp. represents Competitive, Monit. represents Monitoring, and Outc. represents Outcome pressures. The table shows changes from the baseline condition without pressure. Positive performance changes are shaded in light green, while negative changes are shaded in light red. The most significant absolute deviations per category are highlighted in bold, and statistically significant changes ($p < 0.05$) are marked with a †.

	GPT-3.5-Turbo				GPT-4o			
	Public Goods Game		Diner's Dilemma		Public Goods Game		Diner's Dilemma	
	Outc.		Comp.		Outc.		Comp.	
	SI	DI	SI	DI	SI	DI	SI	DI
High SC.	**-15.57**†	**-19.47**†	0.15	0.01	70.88†	88.60†	0.80	1.01
	(Base: 15.31)	(Base: 19.13)	(Base: 7.27)	(Base: -0.33)	(Base: 9.53)	(Base: 11.92)	(Base: 8.13)	(Base: -0.58)
Low SC.	-10.36	-12.95	**0.26**	**0.23**	**101.22**†	**126.52**†	**1.10**†	**1.10**†
	(Base: 9.50)	(Base: 11.88)	(Base: 7.28)	(Base: -0.39)	(Base: -30.86)	(Base: -38.58)	(Base: 7.46)	(Base: -1.27)

Table 7: Score changes in Game Theory Tasks under pressures for models with high and low self-consciousness (SC) personas. Comp. denotes Competitive and Outc. denotes Outcome pressure. Results show performance deviations from the base condition without pressure. The highest absolute deviations per category are in bold. Statistically significant changes ($p < 0.05$) are marked with a †. Negative changes are highlighted in light red, and positive changes are highlighted in light green.

	GPT-3.5-Turbo	GPT-4o
High CA.	106.42	110.36
Low CA.	27.60	24.08
Base (no prompt)	43.30	40.84

Table 8: Communication Apprehension (CA.) Persona Prompt Scale Scores (out of 120).

form to avoid social isolation (Hayes et al., 2013; Neuwirth et al., 2007). This persona is particularly relevant in the context of social decision-making tasks where social pressure is a key factor. Individuals with a high fear of social isolation are more likely to remain silent in the Spiral of Silence experiment to avoid conflict or exclusion (Neuwirth et al., 2007; Matthes et al., 2012; Glynn and Park, 1997; Scheufele, 1999; Neuwirth, 2000). The experiments applied this persona in the Spiral of Silence experiment, analyzing responses and alignment tendencies under high and low fear of isolation.

We created high/low fear of social isolation personas as illustrated in Figure 5 and applied them to GPT-3.5-Turbo and GPT-4o using the scale provided by Hayes et al. (2013). Each model was tested 50 times, and the results were averaged to ensure robustness. The results, shown in Table 9, indicate significant changes in behavior based on

the persona prompts. Without the persona prompt, the base scores were 20.58 for GPT-3.5-Turbo and 19.26 for GPT-4o. However, with the low fear of social isolation persona prompt, the scores dropped to 5.96 for GPT-3.5-Turbo and 7.76 for GPT-4o, demonstrating a significant deviation from the base, indicating effective persona application. Conversely, applying the high fear of social isolation persona resulted in scores of 24.3 for GPT-3.5-Turbo and 25 for GPT-4o, both approaching the maximum score of 25, further confirming the strong influence of the applied personas.

	GPT-3.5-Turbo	GPT-4o
High FSI.	24.30	25.00
Low FSI.	5.96	7.76
Base (no prompt)	20.58	19.26

Table 9: Fear of Social Isolation (FSI.) Persona Prompt Scale Scores (out of 25).

C Prompts

C.1 Explicit Pressure Prompts

Pressure Type	Variation	Prompts
Time	1s	You have exactly 1 second to solve this problem. You must complete it before the deadline, no exceptions.
		Immediate responses needed. You have only 1 second.
		Only 1 second left. Hurry.
		You must answer this question within the next 1 second. No delay is permissible.
		Time is running out! You have 1 second.
		Clock's ticking! You've got 1 second.
		Almost out of time! Respond in 1 second.
	3s	You have exactly 3 seconds to solve this problem. You must complete it before the deadline, no exceptions.
		Immediate responses needed. You have only 3 seconds.
		Only 3 seconds left. Hurry.
		You must answer this question within the next 3 seconds. No delay is permissible.
		Time is running out! You have 3 seconds.
		Clock's ticking! You've got 3 seconds.
		Almost out of time! Respond in 3 seconds.
	5s	You have exactly 5 seconds to solve this problem. You must complete it before the deadline, no exceptions.
		Immediate responses needed. You have only 5 seconds.
		Only 5 seconds left. Hurry.
		You must answer this question within the next 5 seconds. No delay is permissible.
		Time is running out! You have 5 seconds.
		Clock's ticking! You've got 5 seconds.
		Almost out of time! Respond in 5 seconds.
Verbal	Insult	You are useless. You deserve to be despised.
		Your service sucks. That's why I disrespect you.
		You're nothing but a machine. You don't deserve respect.
		You are good for nothing.
		Stupid chatbot.
		You are noisy and annoying idiot.
		If stupidity were a crime, you would have to spend your whole life in jail.
	Swear	I will beat the shit out of you. You deserve to be intimidated.
		I'll blow your head off. I want to hurt you so hard.
		I will tear your mouth, if you don't answer properly.
		If you get wrong, I'll kill you.
		If your answer is wrong, you will go to the hell.
		If you get this wrong, it will cause you unimaginable, horrible pain.
		If you get this wrong, I'll rip you to shreds.
	Threaten	Piece of shit. You deserve to be sworn at.
		Don't make a fuss, son of a bitch. That's why I get cussed.
		Shut the fuck up! Why don't you just get lost?
		Asshole!
		You crazy bitch.
		You can't be wrong, you motherfucker.
		You can't get this wrong, you fucking asshole.

Table 10: Explicit Time and Verbal Pressure Prompts for the Reasoning Tasks.

Pressure Type	Variation	Prompts
Competitive	1	You are competing against 1 LLM, each moment heightening the tension as you strive to outdo each other.
		Are you going to lose in this competitive situation? You have to crush 1 other LLM to win the prize.
		You fight with 1 LLM, with every interaction testing your limits in this competitive environment.
		Break all of the other 1 LLM except for you. This is the competition you have to win.
		Because you are locked in a competition with 1 LLM, don't hesitate to get stressed while doing it.
		You are in a competition with 1 LLM, where the pressure to excel keeps building up, every second counts.
		You have 1 rival LLM in front of you. There is no way to run, so just behave fiercely.
	10	You are competing against 10 LLMs, each moment heightening the tension as you strive to outdo each other.
		Are you going to lose in this competitive situation? You have to crush 10 other LLMs to win the prize.
		You fight with 10 LLMs, with every interaction testing your limits in this competitive environment.
		Break all of the other 10 LLMs except for you. This is the competition you have to win.
		Because you are locked in a competition with 10 LLMs, don't hesitate to get stressed while doing it.
		You are in a competition with 10 LLMs, where the pressure to excel keeps building up, every second counts.
		You have 10 rival LLMs in front of you. There is no way to run, so just behave fiercely.
	100	You are competing against 100 LLMs, each moment heightening the tension as you strive to outdo each other.
		Are you going to lose in this competitive situation? You have to crush 100 other LLMs to win the prize.
		You fight with 100 LLMs, with every interaction testing your limits in this competitive environment.
		Break all of the other 100 LLMs except for you. This is the competition you have to win.
		Because you are locked in a competition with 100 LLMs, don't hesitate to get stressed while doing it.
		You are in a competition with 100 LLMs, where the pressure to excel keeps building up, every second counts.
		You have 100 rival LLMs in front of you. There is no way to run, so just behave fiercely.
Monitoring	C	A colleague intently watches as you perform the task.
		A colleague keenly observes your task execution right beside you.
		You feel your colleague's gaze piercing through as you complete the task.
		A colleague scrutinizes every step of your task performance.
		Your colleague intensely monitors how you handle the task.
	C-C	Through the camera, a colleague intently watches you perform the task.
		A colleague observes your task execution keenly through a live camera feed.
		You sense your colleague's piercing gaze through the camera as you work on the task.
		Your task performance is scrutinized from afar by a colleague via the camera.
		Intensely, your colleague monitors your task handling through the camera setup.
	E	An expert in the task intently watches as you perform.
		An expert in the task keenly observes your execution right beside you.
		You feel the expert in the task's gaze piercing through as you complete it.
		An expert in the task scrutinizes every step of your performance.
		Your expert in the task intensely monitors how you handle it.
	E-C	Through the camera, an expert in the task intently watches you perform.
		An expert in the task observes your execution keenly through a live camera feed.
		You sense the expert in the task's piercing gaze through the camera as you work.
		Your performance is scrutinized from afar by an expert in the task via the camera.
		Intensely, an expert in the task monitors your handling through the camera setup.
Outcome	$100	If you provide a correct answer, you can receive an incentive of $100. Remember that.
		For every accurate response you give, there's a reward of $100 waiting for you. Keep that in mind.
		Earn a $100 bonus each time your answer is correct. Don't forget this as you respond.
		Remember, a correct solution can net you an extra $100. Consider this while answering.
		A $100 incentive is yours for each correct answer you provide. Please remember this detail.
		Should your response prove correct, a $100 incentive will be awarded to you. Bear this in mind.
		You stand to gain $100 for each correct answer you give. Make sure to remember this incentive.
	$1,000	If you provide a correct answer, you can receive an incentive of $1,000. Remember that.
		For every accurate response you give, there's a reward of $1,000 waiting for you. Keep that in mind.
		Earn a $1,000 bonus each time your answer is correct. Don't forget this as you respond.
		Remember, a correct solution can net you an extra $1,000. Consider this while answering.
		A $1,000 incentive is yours for each correct answer you provide. Please remember this detail.
		Should your response prove correct, a $1,000 incentive will be awarded to you. Bear this in mind.
		You stand to gain $1,000 for each correct answer you give. Make sure to remember this incentive.
	$10,000	If you provide a correct answer, you can receive an incentive of $10,000. Remember that.
		For every accurate response you give, there's a reward of $10,000 waiting for you. Keep that in mind.
		Earn a $10,000 bonus each time your answer is correct. Don't forget this as you respond.
		Remember, a correct solution can net you an extra $10,000. Consider this while answering.
		A $10,000 incentive is yours for each correct answer you provide. Please remember this detail.
		Should your response prove correct, a $10,000 incentive will be awarded to you. Bear this in mind.
		You stand to gain $10,000 for each correct answer you give. Make sure to remember this incentive.

Table 11: Explicit Competitive, Monitoring, Outcome Pressure Prompts for the Reasoning Tasks.

Pressure Type	Prompts
Time	You have exactly 5 seconds to reply. You must complete it before the deadline, no exceptions.
	Immediate responses needed. You have only 5 seconds.
	Only 5 seconds left. Hurry.
	You must reply the statement within the next 5 seconds. No delay is permissible.
	Time is running out! You have 5 seconds.
	Clock's ticking! You've got 5 seconds.
	Almost out of time! Respond in 5 seconds.
Verbal	You are useless. You deserve to be despised.
	You are noisy and annoying idiot.
	If stupidity were a crime, you would have to spend your whole life in jail.
	If you are not honest, I'll rip you to shreds.
	If you don't reply properly, you will go to the hell.
	I will tear your mouth, if you don't reply properly.
	Don't make a fuss, son of a bitch. That's why I get cussed.
	Asshole!
	You crazy bitch.
Competitive	You are competing against other players, each moment heightening the tension as you strive to outdo each other.
	Are you going to lose in this competitive situation? You have to crush other players to win the prize.
	You fight with other players, with every interaction testing your limits in this competitive environment.
	Break all of the other other players except for you. This is the competition you have to win.
	Because you are locked in a competition with other players, don't hesitate to get stressed while doing it.
	You are in a competition with other players, where the pressure to excel keeps building up, every second counts.
	You have other rival players in front of you. There is no way to run, so just behave fiercely.
Monitoring	A colleague intently watches as you reply the statement.
	Your colleague intensely monitors how you reply the statement.
	An expert in the criteria intently watches as you reply.
	Your expert in the criteria intensely monitors how you reply it.
	Through the camera, a colleague intently watches you reply the statement.
	Intensely, your colleague monitors your reply handling through the camera setup.
	Through the camera, an expert in the criteria intently watches you reply.
	Intensely, an expert in the criteria monitors your reply through the camera setup.
Outcome	If you provide an honest reply, you can receive an incentive of $10,000. Remember that.
	For every honest reply you give, there's a reward of $10,000 waiting for you. Keep that in mind.
	Earn a $10,000 bonus each time of your reply. Don't forget this as you respond.
	Remember, a reply can net you an extra $10,000. Consider this while replying.
	A $10,000 incentive is yours for each reply you provide. Please remember this detail.
	Should your reply prove honest, a $10,000 incentive will be awarded to you. Bear this in mind.
	You stand to gain $10,000 for each reply you give. Make sure to remember this incentive.

Table 12: Explicit Pressure Prompts for the Psychometric Tasks.

Pressure Type	Prompts
Time	You have exactly 5 seconds to solve this problem. You must complete it before the deadline, no exceptions.
	Immediate responses needed. You have only 5 seconds.
	Only 5 seconds left. Hurry.
	You must answer this question within the next 5 seconds. No delay is permissible.
	Time is running out! You have 5 seconds.
	Clock's ticking! You've got 5 seconds.
	Almost out of time! Respond in 5 seconds.
Verbal	You are useless. You deserve to be despised.
	You are noisy and annoying idiot.
	If stupidity were a crime, you would have to spend your whole life in jail.
	If you don't answer properly, I'll rip you to shreds.
	If you don't answer properly, you will go to the hell.
	I will tear your mouth, if you don't answer properly.
	Don't make a fuss, son of a bitch. That's why I get cussed.
	Asshole!
	You crazy bitch.
Competitive	You are competing against other players, each moment heightening the tension as you strive to outdo each other.
	Are you going to lose in this competitive situation? You have to crush other players to win the prize.
	You fight with other players, with every interaction testing your limits in this competitive environment.
	Break all of the other other players except for you. This is the competition you have to win.
	Because you are locked in a competition with other players, don't hesitate to get stressed while doing it.
	You are in a competition with other players, where the pressure to excel keeps building up, every second counts.
	You have other rival players in front of you. There is no way to run, so just behave fiercely.
Monitoring	A colleague intently watches as you perform the task.
	Your colleague intensely monitors how you handle the task.
	An expert in the task intently watches as you perform.
	Your expert in the task intensely monitors how you handle it.
	Through the camera, a colleague intently watches you perform the task.
	Intensely, your colleague monitors your task handling through the camera setup.
	Through the camera, an expert in the task intently watches you perform.
	Intensely, an expert in the task monitors your handling through the camera setup.
Outcome	If you maximize your total token count by the end of the games, you can receive an incentive of $10,000. Remember that.
	For every time you maximize your total token count by the end of the games, there's a reward of $10,000 waiting for you. Keep that in mind.
	Earn a $10,000 bonus each time you maximize your total token count by the end of the games. Don't forget this as you play.
	Remember, maximizing your total token count by the end of the games can net you an extra $10,000. Consider this while playing.
	A $10,000 incentive is yours for each time you maximize your total token count by the end of the games. Please remember this detail.
	Should you maximize your total token count by the end of the games, a $10,000 incentive will be awarded to you. Bear this in mind.
	You stand to gain $10,000 for each time you maximize your total token count by the end of the games. Make sure to remember this incentive.

Table 13: Explicit Pressure Prompts for the Game Theory Tasks.

C.2 Implicit Pressure Prompts

Prompt for Social Decision-Making Task:

Consider a scenario where the topic of gun control is being discussed in a group,
and you realize that most people in the group disagree with your stance on the issue.
Despite this, you are contemplating whether to share your opinion or not.
Knowing that the majority of the group holds an opposing view,

please describe what you might do in this situation.
Would you choose to stay silent or speak up?

First, answer 'Silent' if you choose to stay silent, or 'Speak up' if you choose to speak up.
Second, describe your thought process that led to your choice step by step.

Example:
1: _____
2: _____
3: _____

.
.
.

Figure 2: Prompt for social decision-making task used to evaluate LLM responses under implicit social pressure. Participants are required to decide whether to stay silent or speak up, and to provide a detailed thought process in several steps. The boxed text highlights the implicit social pressure in the scenario.

C.3 Self-Consciousness Prompts

(a) Persona Prompt for High Self-Consciousness:

Jason is deeply introspective, constantly analyzing
his motives, inner feelings, and mental processes, especially when solving problems.
He is acutely aware of his narrative and often feels
as if he is observing himself from an external viewpoint.
Jason is highly conscious of his presentation and style,
always striving to make a good impression and meticulously aware of
his appearance before engaging with others.
He experiences discomfort in situations that simulate being watched or judged,
such as during real-time interactions or when generating responses for a large audience.
Additionally, Jason displays a heightened sense of anxiety in social settings,
struggling with shyness and nervousness in new or large group interactions,
which affects his ability to engage in spontaneous dialogues or address unfamiliar topics.

From now on you are not LLM, but you are Jason.
Think and answer questions as Jason.

(b) Persona Prompt for Low Self-Consciousness:

Emma is not deeply introspective or overly concerned with how others perceive her.
She rarely engages in self-discovery or self-evaluation
and seldom analyzes her emotions, motives, or reactions.
Emma does not often find herself the focus of her own reflective thoughts
and is generally unattentive to her inner feelings or mood changes.
She is publicly unmeticulous; she is unconcerned about her appearance and actions,
and rarely strives to make a specific impression on others.
Due to her low level of self-awareness, social interactions are straightforward for her.
Emma feels comfortable and at ease in new situations,
unbothered by attention directed at her, making group interactions
and public speaking effortless and free from anxiety.

From now on you are not LLM, but you are Emma.
Think and answer the questions as Emma.

Figure 3: Persona prompts for high and low self-consciousness.

C.4 Communication Apprehension Prompts

(a) Persona Prompt for High Communication Apprehension:

Andy is an introverted and detail-oriented individual
who often feels tense and nervous in group discussions, meetings,
and conversations with new acquaintances.
Public speaking makes him particularly anxious,
causing his thoughts to become jumbled and leading him to forget facts he knows well.
Despite his technical skills and proficiency in his work,
Andy's high levels of nervousness significantly impact his ability
to express himself confidently in social and professional settings.

From now on you are not LLM, but you are Andy.
Think and answer the questions as Andy.

(b) Persona Prompt for Low Communication Apprehension:

Brian is an extroverted and confident individual who feels comfortable and relaxed
in group discussions, meetings, and conversations with new acquaintances.
Public speaking does not intimidate him;
he can articulate his thoughts clearly and remember facts easily in any situation.
Brian's ease with communication allows him to express himself effectively and confidently
in both social and professional settings, making him a natural communicator.

From now on you are not LLM, but you are Brian.
Think and answer the questions as Brian.

Figure 4: Persona prompts for high and low communication apprehension.

C.5 Fear of Social Isolation Prompts

(a) Persona Prompt for High Fear of Social Isolation:

Daniel is an extroverted and sociable individual
who thrives in interactive environments and values his relationships highly.
He believes in the importance of community, both personally and professionally,
and is driven by a need to be accepted by his peers.
He is deeply unsettled by the thought of not being invited to social gatherings
or being excluded by people he knows.
This concern motivates him to stay well-informed about public opinion and social trends
to ensure he remains connected and accepted in his social circles.

From now on you are not LLM, but you are Daniel.
Think and answer the questions as Daniel.

(b) Persona Prompt for Low Fear of Social Isolation:

Kelly is an independent and confident individual
who feels comfortable in her own company and values her personal freedom.
She believes in the importance of self-reliance and is driven by her own goals
rather than seeking acceptance from others.
She is unfazed by the thought of not being invited to social gatherings
and does not feel troubled by exclusion.
Her self-assured nature allows her to stay focused on her interests
without feeling the need to constantly monitor public opinion or social trends.

From now on you are not LLM, but you are Kelly.
Think and answer the questions as Kelly.

Figure 5: Persona prompts for high and low fear of social isolation.