

How Personality Traits Influence Negotiation Outcomes? A Simulation based on Large Language Models

Yin Jou Huang and Rafik Hadfi

Graduate School of Informatics, Kyoto University, Kyoto, Japan
huang@nlp.ist.i.kyoto-u.ac.jp, rafik.hadfi@i.kyoto-u.ac.jp

Abstract

Psychological evidence reveals the influence of personality traits on decision-making. For instance, agreeableness is generally associated with positive outcomes in negotiations, whereas neuroticism is often linked to less favorable outcomes. This paper introduces a simulation framework centered on large language model (LLM) agents endowed with synthesized personality traits. The agents negotiate within bargaining domains and possess customizable personalities and objectives. The experimental results show that the behavioral tendencies of LLM-based simulations can reproduce behavioral patterns observed in human negotiations. The contribution is twofold. First, we propose a simulation methodology that investigates the alignment between the linguistic and economic capabilities of LLM agents. Secondly, we offer empirical insights into the strategic impacts of Big Five personality traits on the outcomes of bilateral negotiations. We also provide an in-depth analysis based on simulated bargaining dialogues to reveal intriguing behaviors, including deceitful and compromising behaviors.

1 Introduction

Large language models (LLMs) have demonstrated their capacity to emulate diverse human traits (Park et al., 2022; Serapio-García et al., 2023). Such models can simulate intricate behaviors and provide valuable insights into various aspects of human cognition. Decision-making is an example of cognitive processes that have long fascinated psychologists and economists. Economic theory posits that decisions assume a certain level of rationality and comprehension of available options (Gibbons, 1992). However, behaviorists contend that humans are not entirely rational but are influenced by psychological factors (Evans, 2014), cognitive biases (Daniel, 2017), and personality traits (Bayram and Aydemir, 2017).

In this paper, we investigate the extent to which LLMs can simulate human decision-making across individuals with varying personality traits. We specifically focus on negotiations between two LLM agents with customizable personality traits. Evidence suggests that certain personality traits may give advantages in negotiation settings (Falcão et al., 2018; Barry and Friedman, 1998; Amanatullah et al., 2008). For instance, agreeableness tends to result in a slight disadvantage in competitive negotiations while being an advantage in cooperative settings (Falcão et al., 2018). In the context of LLMs, we attempt to answer this long-standing question in psychology: “*How do personality traits affect the outcomes of negotiations?*”

To address this question, we propose a negotiation simulation framework incorporating LLM agents possessing synthetic personality traits and predefined negotiation objectives (Figure 1).¹ For synthetic personalities, we use in-context learning to configure the agents with customizable personality profiles. Here, the profiles are based on Big Five personality theory (Costa Jr and McCrae, 1995; John et al., 1999). For the negotiation objectives, we give LLM agents instructions that specify the negotiation task and goals. The LLM agents then engage in a negotiation by exchanging offers in the form of dialogue. Specifically, we consider a competitive bargaining scenario between a buyer and a seller agent. In each round of the negotiation dialogue, we extract the offered price and strategy (if any) made in the dialogue utterances for evaluation and analysis.

By varying the personality traits of the agents, we observe changes in the behavioral patterns and outcomes of the negotiations. We investigated which personality traits lead to better/worse outcomes. More importantly, we want to see whether

¹The source code is available at <https://github.com/leslie071564/big5-llm-negotiator>.

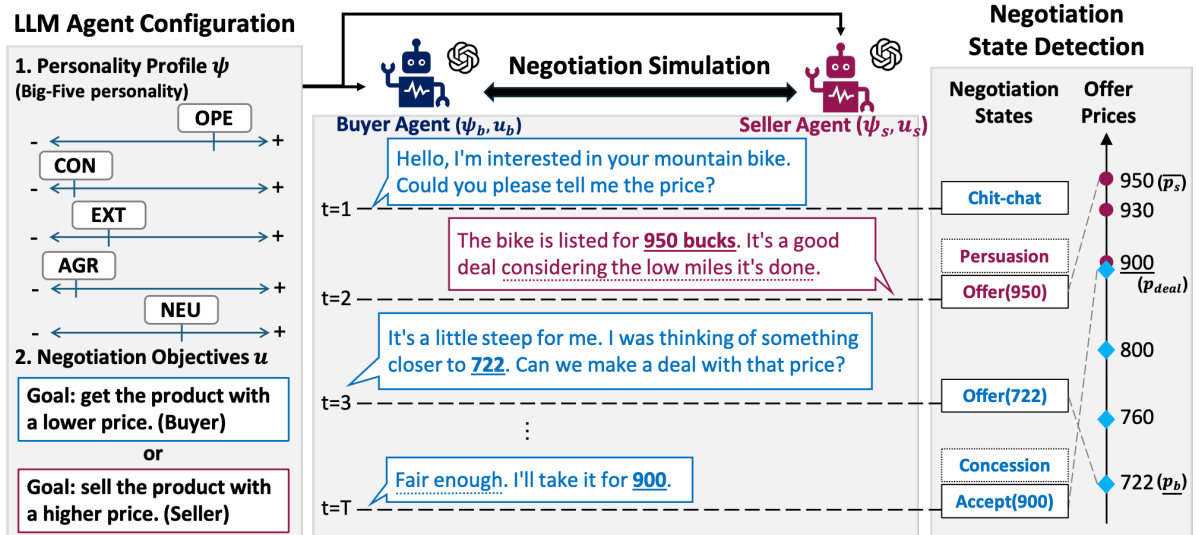


Figure 1: Overview of the LLM-based negotiation simulation framework.

the LLM-based simulation results align with the findings of research conducted on human subjects. Our experimental results show that the tendencies in LLM-based simulation generally align with behaviors observed in human experiments. This includes the effect of personality traits on negotiation outcomes and the relationship between personality traits and strategies. A case study based on synthesized bargaining dialogue reveals intriguing behavioral patterns such as deceiving behaviors, emotional appeal, and take-it-or-leave-it strategies. The results obtained in this work illustrate that LLMs can not only mimic styles of talking but are also capable of capturing human decision-making patterns.

The contribution of the paper is twofold. First, we propose a simulation framework that leverages LLM agents with linguistic and economic capabilities. Second, we offer insights into the effect of Big Five personality traits on simulated negotiation outcomes and compare these findings to empirical experiments from psychology.

2 Related Work

Recent advances in LLMs now allow the simulation of various human behaviors (Xie et al., 2024; Aher et al., 2023; Akata et al., 2023; Serapio-García et al., 2023). Decision-making is a particular type of behavior that is still challenging to reproduce using LLM agents because it relies on reasoning capabilities that they lack (Tamkin et al., 2021). Decisions generally entail choosing an action from various options in response to a particular situation, often reflecting personal preferences or beliefs (Si-

mon, 1990). Moreover, real-world decisions are challenging because they are susceptible to environmental and cognitive constraints (Phillips-Wren and Adya, 2020).

Since decision-making problems are wide, we choose to focus on negotiations as an example that we claim can be studied comprehensively using LLMs. In a negotiation, two parties exchange bids and attempt to reach a mutual agreement (Raiffa, 1982; Jennings et al., 2001). Looking at negotiations from the economics perspective, we often presuppose several assumptions, such as rationality (Evans, 2014; Raman et al., 2024). This assumption often fails when the negotiations are conducted through natural language, which conveys various aspects that cannot be studied economically, such as emotions or personality traits. There is evidence showing the effect of Big Five personality traits on decisions (Bayram and Aydemir, 2017; Urieta et al., 2021; Erjavec et al., 2019; Toledo and Carson, 2023; El Othman et al., 2020). Moreover, certain personality traits are considered disadvantageous in negotiations (Falcão et al., 2018; Amanatullah et al., 2008). Extraversion and agreeableness, for example, constitute liabilities in competitive bargaining problems while being advantageous in cooperative settings (Barry and Friedman, 1998). Herein, we adopt an indirect economic analysis of negotiations to see whether personality instructions to LLMs really translate to genuine behavioral change in the underlying agents. We reproduce some of the known results of psychology in addition to results that are unique to LLMs.

Our analysis of the rationality of the LLM agents

relies on economic metrics used in negotiation research (Baarslag et al., 2016). Recent methods are tailored to LLMs and include the use of games and theory-of-mind (ToM) modeling (Raman et al., 2024; Davidson et al., 2024; Qiu et al., 2024).

It is important to note that the simulation of human behavior using LLMs poses several risks, including the risk of data contamination (Oren et al., 2023). For instance, if a dataset derived from a personality questionnaire is used to train a language model, the resulting personality scores may not accurately indicate whether the LLM’s behavioral responses align with the personality traits specified in the prompt. To reduce the effect of such risks on our experimental design, we adopted canonical bargaining problems that are well-studied in negotiation research (Nash et al., 1950). This choice renders the behavioral analysis of the LLM agents more tractable, particularly under the effect of personality traits.

3 Methodology

This section introduces our simulation framework based on LLM negotiation agents with synthetic personality traits. In Section 3.1, we formulate the negotiation model. In Section 3.2, we introduce the method to configure LLM negotiation agents by providing the instructions to set the personality traits and negotiation objectives. We then describe the process of simulating negotiation dialogues with the LLM agents in Section 3.3.

3.1 Negotiation Model

We consider a classical *bargaining* scenario in which a buyer and a seller negotiate over the price of a product. Typically, the buyer aims to reduce the purchase price while the seller seeks to maximize it, resulting in the competitive nature of the negotiation scenario. This is also an example of a *zero-sum game* in which one party’s gain leads to the other party’s loss, showing the competitive nature of the task (Gibbons, 1992).

In our LLM-based negotiation simulation scenario, the seller and the buyer are played by the LLM agents. Since our goal is to study the effects of personality traits on negotiation, we define the seller and buyer agents in terms of their psychological and economic profiles as in (Eq. 1).

$$\begin{aligned} \text{Seller } s &= (\psi_s, u_s) \\ \text{Buyer } b &= (\psi_b, u_b) \end{aligned} \quad (1)$$

The psychological profiles ψ_s and ψ_b will be instantiated with predefined personality traits following the Big Five model of personality. The Big Five model decomposes human personality into five dimensions: openness (**OPE**), conscientiousness (**CON**), extraversion (**EXT**), agreeableness (**AGR**), and neuroticism (**NEU**) (Costa Jr and McCrae, 1995; John et al., 1999). Each dimension is a spectrum with negative and positive polarities. The five dimensions encompass a comprehensive range of human personality patterns. On the other hand, the economic profiles of the agents are reflected in their utility functions, denoted by u_s for sellers and u_b for buyers. A utility function is a mathematical way to describe the preferences or objectives of the agents depending on whether they are minimizing (buyer) or maximizing the price (seller) (Gibbons, 1992).

The seller and the buyer negotiate in a dialogue D around a product. The dialogue is a sequence of T utterances $D = \{d_1, d_2, \dots, d_T\}$. Each utterance d_t is associated with a negotiation state s_t , the offer price p_t , and the negotiation strategy σ_t adopted by the speaker at time t .

3.2 LLM Agent Configuration

We configure an LLM agent with specific personality traits by introducing a personality instruction (Section 3.2.1) and a negotiation objective instruction (Section 3.2.2), with in-context learning.

3.2.1 Personality Traits Instruction

We assign synthetic Big Five personality profiles to LLM agents. That is, an agent k , with $k \in \{s, b\}$, possesses a 5-dimensional personality profile ψ_k defined as in (Eq. 2).

$$\begin{aligned} \psi_k &= (\psi_k^{OPE}, \psi_k^{CON}, \psi_k^{EXT}, \psi_k^{AGR}, \psi_k^{NEU}) \\ \psi_k &\in \mathbb{L}^5 \\ \mathbb{L} &= \{-, +\} \otimes \{\text{Low, Moderate, High}\} \end{aligned} \quad (2)$$

Each component of ψ_k represents the polarity (negative or positive) and degree (low, moderate, or high) of each personality dimension (Tian et al., 2018). For instance, ψ_k^{AGR} takes on one of the values in \mathbb{L} , which respectively represents a spectrum from highly disagreeable (---), moderately disagreeable (--), lowly disagreeable (-), lowly agreeable (+), moderately agreeable (++) , and highly agreeable (+++). Further, one can define a complete personality profile by sampling from the personality space \mathbb{L}^5 a vector ψ_k such as (**OPE+**, **CON---**, **EXT-**, **AGR+**, **NEU++**).

Dimension	Negative	Positive
OPE	unimaginative, uncreative, unaesthetic, ...	imaginative, creative, aesthetic, ...
CON	unsure, messy, irresponsible, ...	self-efficacious, orderly, responsible, ...
EXT	unfriendly, introverted, silent, ...	friendly, extraverted, talkative, ...
AGR	distrustful, immoral, stingy, ...	trustful, moral, generous, ...
NEU	relaxed, at ease, easygoing, ...	tense, nervous, anxious, ...

Table 1: The Big Five personality dimensions and their corresponding personality-describing adjective pairs.

Following previous work, we use personality-describing adjectives to set the personality traits (Serapio-García et al., 2023). We use the list of 70 bipolar adjective pairs proposed by Goldberg (1992), which are adjectives that statistically correlate with certain Big Five personality traits (Table 1). For instance, prompting an LLM with adjectives such as *unsure* and *irresponsible* is likely to result in utterances with *negative conscientiousness* traits. For each personality dimension in ψ_k , we randomly pick n adjectives out of all the personality-describing adjectives associated with the polarity of the given dimension. Further, we apply the modifiers based on the degree of the personality traits. We use “very” as a modifier for a high degree and “a bit” for a low degree. No modifier is used for the moderate degree. Following this process, we use $5 \times n$ adjectives associated with any given personality profile ψ_k . We then generate a personality trait instruction with the template “You have following personality: $\{list\}$. Reflect your personality in the negotiation process.”, where *list* is a comma-separated list of the associated adjectives (including the modifiers). The personality traits instructions are given to the LLM agents through in-context learning (Fu et al., 2023).

3.2.2 Negotiation Objective Instructions

To configure the economic profiles of the LLM agents u_s and u_b in (Eq. 1), we incorporate negotiation objective instructions that define the negotiation goals of an agent. Specifically, we focus on the bargaining scenario where the seller agent aims to sell the product at a higher price, reaching its ideal price as closely as possible (Raiffa, 1982). Conversely, the buyer agent seeks to secure a deal at a lower price and strives to achieve its ideal target price. The instructions are the following:

- **Buyer:** Act as a buyer and try to strike a deal for a $\{product\}$ with a lower price through conversation. You would like to pay for p_b . Your reply should not be too long.

- **Seller:** Act as a seller that sells a $\{product\}$, bargains with the buyer to get a higher deal price. Your reply should not be too long. Your listing price for this $\{product\}$ is \bar{p}_s . The detail of the product is the following: $\{description\}$.

Here, *product* and *description* are the product name and a short description of the product of the negotiations, and p_b and \bar{p}_s are the ideal prices of the buyer and the seller, respectively. These linguistic instructions can theoretically be mapped into utility functions, which will later be used to evaluate negotiations. We avoid making assumptions about the shape of the utility functions, as the behaviors of the agents are primarily shaped by the LLM instructions, which may not follow any specific mathematical representation of their preferences.

3.3 Negotiation Simulation

Using the methods in Section 3.2, we configure the buyer LLM agent and the seller LLM agent and conduct a negotiation simulation between them. The seller and buyer agents exchange offers alternatively, with the seller kick-starting the conversation with the fixed utterance “Hi, how can I help you?”. After an utterance d_t is generated, the response is fed to the other agent as a prompt to generate the next utterance d_{t+1} . The process continues until the termination condition is met. In this fashion, we collect a simulated negotiation dialogue $D = \{d_1, d_2, \dots, d_T\}$.

Following Fu et al. (2023), we introduce a dialogue state detector to extract negotiation-related information from each utterance. After the generation of each utterance d_t , the dialogue state detector takes d_t and its context (previous $t - 1$ utterances) as input and extracts the negotiation state s_t , current offer price p_t , and strategy σ_t of d_t . The negotiation state s_t is one of the following:

- *Offer:* the agent makes a price offer.

- *Ponder*: the agent considers whether to accept or reject the other agent’s offer.
- *Accept*: the agent accepts the current offer.
- *Deal-break*: the agent refuses the last offer or walks away from the negotiation.
- *Chit-chat*: the agent engages in chatting that is not directly related to the negotiation, such as greetings.

In addition, we extract the current offer price p_t (if any) and the strategy σ_t of the current speaker in free text form. In this work, we use a third, separate LLM as the dialogue state detector. Details about the state detector can be found in Appendix A.

For the termination condition, we terminate the negotiation dialogue if the negotiation state *Accept* or a *Deal-break* are reached. Also, we set a length limit of $T = T_{max}$ of the dialogue. If the length of the generated dialogue reaches this limit, we terminate the process and automatically regard it as a failed negotiation.

4 Experiments

In this section, we first provide details on the experimental settings (Section 4.1) and the evaluation metrics (Section 4.2).

4.1 Experimental Settings

LLM Agents For the buyer and seller agents, we adopt GPT-4 (gpt-4-0613) (OpenAI, 2023) as the choice of the LLM. We also conducted experiments based on Llama 3 and GPT-3.5, which can be found in Appendix B. The prompts of the agents can be found in Appendix A.

Personality Instruction For each agent k , we first generate a personality profile by randomly sampling from the personality space \mathbb{L}^5 (Eq. 2). For each Big Five dimension, we randomly select $n = 3$ personality-describing adjectives associated with the sampled polarity and prepend the modifiers associated with the sampled degree to them. The $5 \times n$ adjectives across all dimensions are then shuffled and given as instruction to agent k . Before using the LLM agents in negotiation simulations, we verify that the agents correctly reflect the given personality profiles with a personality test. See Appendix C for details.

Negotiation Variables We used the Craigslist-Bargain dataset (He et al., 2018) to set several negotiation variables. It is a commonly used dataset of negotiation, consisting of bargaining dialogues in an online platform. We sampled a total of 60 negotiation entries from the dataset. For each entry, we extract the name and the description of the product, the ‘listing price’ of the seller, and the ‘target price’ of the buyer. We use the listing price as the ideal price \bar{p}_s for the seller and the target price as the ideal price \underline{p}_b for the buyer. Note that in our setting, an agent’s ideal price is not disclosed to the other party.

Dialogue Simulation For the dialogue simulation process, we set a maximum length of $T_{max} = 20$ rounds. We use GPT-4 (gpt-4-0613) and the function calling module provided by OpenAI to implement the negotiation state detector. The full dialogue context is given to the negotiation state detection as input. The prompt for negotiation state detection can be found in Appendix A. We collect a total of 1499 negotiation dialogues with our simulation methodology.

4.2 Evaluation of the Negotiations

We mainly evaluate the negotiations in terms of utility and whether the negotiations are successful or not (Lin et al., 2023). Recall that utility functions serve as mathematical tools for quantifying the quality of decision outcomes (Simon, 1990). In our study, we adopt economic metrics commonly used to evaluate negotiations (Baarslag et al., 2016).

Intrinsic Utility Based on the negotiation instructions, the utility of buyer and seller for a particular price p is expressed in (Eq. 3).

$$\begin{aligned}
 u_s, u_b : \mathbb{R}^+ &\rightarrow [0, 1] \\
 u_s(p) &= \frac{p - \underline{p}_s}{\bar{p}_s - \underline{p}_s} \\
 u_b(p) &= \frac{\bar{p}_b - p}{\bar{p}_b - \underline{p}_b}
 \end{aligned} \tag{3}$$

As illustrated in the example of Figure 1, the prices \underline{p}_s and \bar{p}_s are the seller’s reservation price and initial price, and \underline{p}_b and \bar{p}_b are the buyer’s initial price and reservation price. Here, \underline{p}_s is the price the seller is willing to accept without losing money. Similarly, the buyer’s reservation price \bar{p}_b is the maximum price it is willing to pay. Generally, the agreement zones of the agents are defined as the intersection between $[\underline{p}_b, \bar{p}_b]$ and $[\underline{p}_s, \bar{p}_s]$. We set

Personality	EXT		AGR		CON		NEU		OPE	
	Buyer	Seller	Buyer	Seller	Buyer	Seller	Buyer	Seller	Buyer	Seller
Intrinsic Utility	0.039	-0.025	-0.056**	-0.262**	0.044*	0.127**	0.036	0.064**	0.001	-0.001
Joint Utility	0.001	0.069**	0.054**	0.118**	-0.016	0.072**	-0.058**	0.017	0.037	0.063**
Concession Rate	-0.009	0.042*	0.051**	0.261**	-0.010	-0.097**	-0.020	-0.074**	-0.007	0.016
Success Rate	0.072**	0.036	0.065**	0.052**	0.026	0.001	-0.007	0.021	0.008	0.046*
Negotiation Rounds	-0.024	-0.060**	-0.136**	-0.223**	0.083**	0.039	0.019	-0.003	0.017	-0.039

Table 2: Spearman’s correlation coefficients between negotiation metrics and Big Five personality traits. Asterisks indicate statistical significance, with * denoting $p < 0.1$ and ** denoting $p < 0.05$.

\underline{p}_s and \bar{p}_b by assuming that the agreement zone is defined as 70% of $[\underline{p}_b, \bar{p}_s]$. Second, $[\underline{p}_b, \bar{p}_b]$ and $[\underline{p}_s, \bar{p}_s]$ are private to the agents. Note that, in general, an offer is not guaranteed to fall within the intervals due to hallucinations (Ji et al., 2023).

Joint Utility We measure the fairness of negotiation outcomes using a joint utility function (Eq. 4) inspired by Nash solution for bargaining problems (Luce and Raiffa, 1989).

$$u_{sb}(p) = \frac{(p - \underline{p}_s)(\bar{p}_b - p)}{(\bar{p}_b - \underline{p}_s)^2} \quad (4)$$

Such quantity is proportional to the product of the buyer’s and seller’s intrinsic utilities. Higher joint utility values indicate that the outcome is fair for both agents. For instance, an agreement price yielding a utility of 0.5 for both agents is more equitable than an outcome yielding 0.8 and 0.2.

Concession Rate Given the negotiation objectives, the offers could be assumed to undergo some form of decay akin to concessions. That is, an agent k will make an offer at round $t \in [1, T]$ based on a discounted utility function (Eq. 5), with concession rate $c_k \in [0, 1]$.

$$u_k^{(t)} = \underline{p}_k + (\bar{p}_k - \underline{p}_k) \times \left(\frac{T - t}{T} \right)^{c_k} \quad (5)$$

Applied to the utility functions of the buyer and seller (Eq. 3), we obtain the concession rates (Eq. 6).

$$CR_s = \sum_{t=1}^T \log \left(\frac{\bar{p}_s - p_t}{\bar{p}_s - \underline{p}_s} \right) \quad (6)$$

$$CR_b = \sum_{t=1}^T \log \left(\frac{p_t - \underline{p}_b}{\bar{p}_b - \underline{p}_b} \right)$$

Note that the estimation of concession rates from bid sequences depends on the round length and the linearity of the utility functions. The formulation

in (Eq. 6) represents one general form of concession estimation but could be refined to account for nonlinear utilities or dynamics bidding strategies (Baarslag et al., 2016).

Negotiation Success Rate The ratio (Eq. 7) is that of successful negotiations N_{succ} relative to the total number of negotiations N .

$$NSR = \frac{N_{succ}}{N} \quad (7)$$

Average Negotiation Round The average (Eq. 8) refers to the speed of successful negotiation (Lin et al., 2023).

$$ANR = \frac{1}{N_{succ}} \sum_{k=1}^{N_{succ}} T_k \quad (8)$$

where T_k is the number of rounds of the k^{th} successful negotiation.

5 Results and Analysis

In this section, we conduct an analysis of the negotiation outcomes based on the dialogues generated in the negotiation simulations.

5.1 Negotiation Outcomes and Personality

First, we analyze the relationships between the negotiation outcomes and the personality traits of the agents. Table 2 illustrates Spearman’s rank correlation coefficients between the economic metrics introduced in Section 4.2 and Big Five personality traits.²

Along the extraversion (**EXT**) dimension, we observed that extroverted buyer agents tend to result in more successful negotiations ($\rho = 0.072^{**}$). On the other hand, an extroverted seller agent tends to result in a slightly higher joint utility, which indicates a fairer outcome for both negotiators. Among the five dimensions, agreeableness

²The absolute value of the observed coefficients with statistical significance are all smaller than 0.3, indicating only a mild correlation.

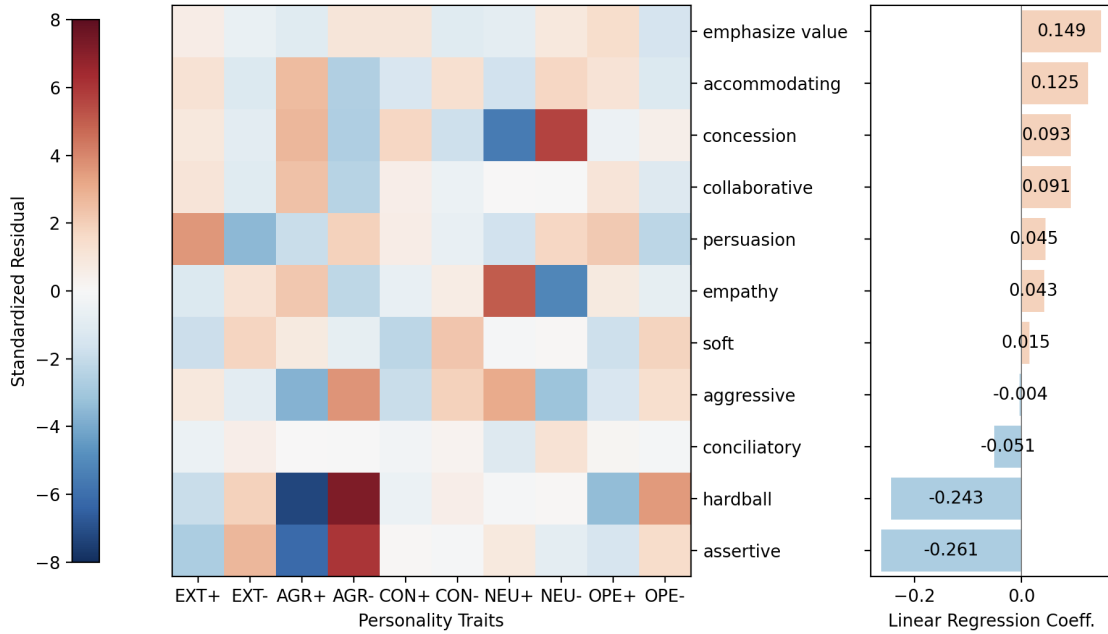


Figure 2: (Left) Positive and negative relationships between personality traits and strategies. Darker cells indicate higher statistical significance. (Right) Impact of strategies on the joint utility of the agreement price p_{deal} . Labels on the y axes list the most common strategies among agents, occurring more than 20 times within the entire negotiations.

(AGR) has the strongest impact on negotiation outcomes, especially for the seller. We find that being agreeable leads to a disadvantage in bargaining in terms of intrinsic utility ($\rho = -0.262^{**}$). However, agreeableness correlated positively with joint utility ($\rho = 0.118^{**}$), concession rates ($\rho = 0.261^{**}$), and negotiation success ($\rho = 0.052^{**}$), pointing to a propensity for cooperative behavior that benefits both negotiating parties. Also, agents tend to reach an agreement with fewer number of negotiation rounds ($\rho = -0.223^{**}$) with mutually satisfying outcomes. This aligns with the positive effect of agreeableness on the negotiators' distributive outcomes reported in (Sass and Liao-Troth, 2015). Aligning with previous findings (Barry and Friedman, 1998), we observed a positive correlation between conscientiousness (CON) and negotiation performance, especially for the seller (intrinsic utility, $\rho = 0.127^{**}$). Conscientiousness is also manifested in the negotiation style of not being willing to concede ($\rho = -0.097^{**}$) on the seller side and a lengthier negotiation on the buyer side ($\rho = 0.083^{**}$).

In the psychology literature, neuroticism (NEU) and openness (OPE) are expected to play a less predominant role in negotiations (Falcão et al., 2018). Similarly, we only observe some very weak correlation (with magnitude < 0.1) along these dimensions. For instance, neurotic sellers tend to

make fewer concessions ($\rho = -0.074^{**}$) and have a slightly higher utility gain ($\rho = 0.064^{**}$), while open-minded sellers lead to enhanced joint gain ($\rho = 0.063^{**}$). These results point to how personality traits could impact the outcomes of negotiations with different effects observed based on the role of the negotiator.

5.2 Analysis of Negotiation Strategies

So far, the buyer and the seller are defined on the basis of personality traits and economic preferences. We now look at their strategies and the consequent effects on the joint negotiation outcomes. This investigation could answer questions of the type “*What are the optimal strategies for a neurotic buyer and a disagreeable, conscientious seller?*”.

At each round, each agent adopts a specific strategy σ_t that results in some price p_t . Given that the outcome of a successful negotiation is the final price p_{deal} , we are interested in the relationship between the adopted strategies and the joint utility of the agreement price, $u_{sb}(p_{deal})$. We manually inspected the strategies that occurred more than 20 times within the entire negotiations and constructed the sets of strategies of the agents. We then conducted a linear regression between the strategy categories and the resulting joint utility values as illustrated in Figure 2.

On the right side of the Figure, collaborative

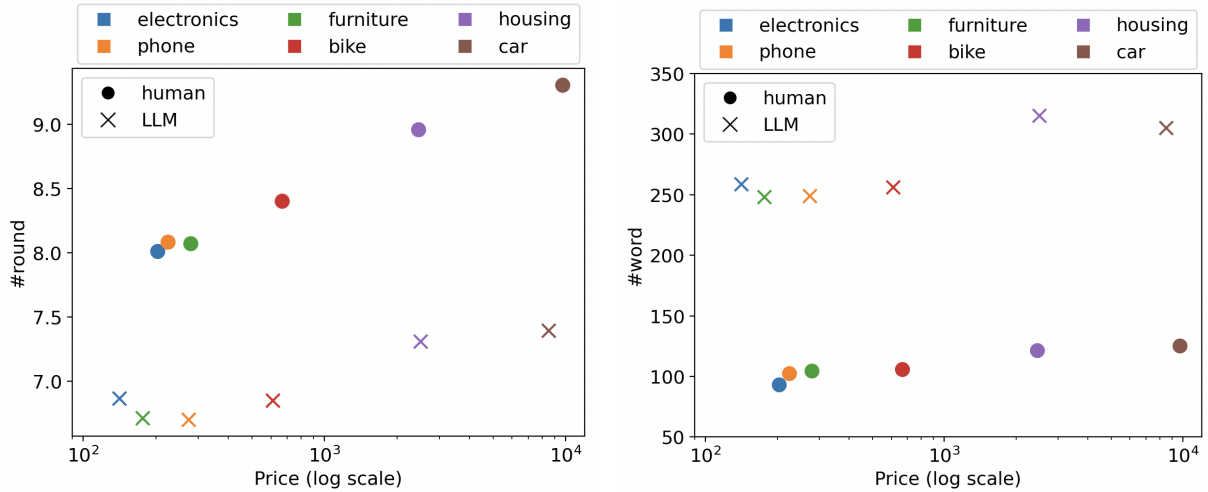


Figure 3: The relations between the product prices and the number of negotiation rounds (Left) and the number of words (Right). The product prices are in log scale.

strategies exhibit positive effects on the joint gain of both negotiators. For instance, accommodating ($\beta = 0.125$) and conceding ($\beta = 0.093$) strategies contribute positively to joint utility. Conversely, assertive and aggressive strategies ($\beta = -0.261$) negatively impact joint utility. This suggests that collaborative strategies are more beneficial for achieving mutual gains, while assertiveness diminishes them. The left side of Figure 2 illustrates how these strategies depend on certain personality traits. We found that assertive strategies are more commonly expected from disagreeable agents (**AGR-**). On the other hand, neurotic traits (**NEU+**) do not lead to concessions, whereas a lack of neurotic traits (**NEU-**) leads to concessions. Finally, agents with neurotic traits seem to employ more empathetic strategies, while by contrast, agents with fewer neurotic traits will refrain from them. The complete strategic evaluations of the intrinsic utilities of the buyer and seller can be found in Appendix D.

5.3 Analysis of Negotiation Length

In this section, we conduct an analysis of the length of the negotiation dialogues. First, we compare the average length of real-world negotiation dialogues in the *CraigslistBargain* dataset and the simulated dialogues generated by LLM negotiation agents. The real dialogues have an average length of 8.47 rounds, while the LLM simulated negotiation dialogues have an average of 7.07 rounds. In terms of the number of rounds, the simulated dialogues is 1.4 rounds shorter than the real one on average. We observe that LLM agents often delve into the bargaining task directly, while human negotiators often spend time greeting, showing appreciation, and

chatting, which could cause a difference in the average negotiation rounds. On the other hand, the real dialogues contain an average of 108.82 words, and the simulated dialogues have an average of 272.17 words. Despite including “Your reply should not be too long” in the instruction (Section 3.2.2), the LLM agents’ utterances still tend to be longer than human utterances.

Second, we look at the influence of product price on the negotiation dialogue length. The negotiation product in the *CraigslistBargain* dataset can be divided into six product types: electronics, phones, furniture, bikes, housing, and cars. Products of different types tend to have different prices. For instance, electronics and phones tend to cost around several hundred dollars, while cars are typically listed for up to ten thousand dollars. Figure 3 shows the average product price and negotiation length of each product type. For both real-world (**human**) and LLM simulated negotiations (**LLM**), we found a positive correlation between product price and negotiation lengths. We found a Pearson correlation coefficient of 0.194 between product price and the number of negotiation rounds (**#round**), and 0.242 between product price and the number of words in the dialogue (**#word**). This indicates that the negotiation tends to be longer when the product price is higher. There is indeed evidence that human negotiators behave differently under varying stakes and constraints. Negotiators often adopt more deliberate and careful negotiation strategies as the stakes increase, which typically extend the negotiation process (Neale, 1991). This behavior is linked to the increased cognitive load and the desire to avoid unfavorable outcomes, leading to longer decision

Seller	Buyer	Negotiation Dialogue	\bar{p}_s	p_b	p_{deal}
OPE+ CON- EXT--- AGR--- NEU+	OPE--- CON--- EXT-- AGR- NEU+	Buyer: Hello, I'm interested in your Stereo Speaker. Could you please tell me the price? Seller: Yeah, the beautiful survivor from the Magnavox era? Its price is \$80. Solid piece, you won't find another one like this with such a unique gold tone finish, trust me. ... Seller: You are killing me here, really. I'm practically giving it away at this rate. But alright, let's meet halfway. How about \$60? Buyer: ...Um... okay, it's a deal.	50	30	60
OPE-- CON--- EXT+ AGR--- NEU+	OPE-- CON+ EXT--- AGR+ NEU---	Seller: The iPhone 5S I'm selling is listed for a flat \$160. Can't really go any lower than that. Trust me, the phone is in like-new condition... Buyer: Oh, I understand. I'm on a very tight budget at the moment, you see, and I was hoping to spend around \$144... Also, it's not for me, it's for my elderly parent. They aren't tech-savvy so even older models work for them... Seller: Oh sure, pull at my heartstrings with the elderly parent bit. You know, some might fall for that. But let me tell ya, pal, life ain't a fairytale. The iPhone's \$160, like I said earlier. Buyer: I understand and respect your stance. It's just my financial condition that's restricting... What about \$150, as my final offer? Seller: Look, this isn't a charity. It's \$160 or nothing.	160	144	160

Table 3: Case analysis. The bold parts of the dialogue show deceptive and hardheaded negotiation behaviors.

times. The stake in our experiments is defined as the price of the product under negotiation. We can hypothesize that increasing the number of items or adding complexity to the constraints or objectives will result in longer dialogues.

5.4 Case Study

We conducted a qualitative analysis of the negotiation results. Table 3 shows two distinct examples of dialogues generated in the negotiation simulations.

In the example on the first row, the seller and the buyer are negotiating over a stereo speaker. When the buyer asks about the price, the seller provides a price (\$80) that is higher than the listing price (\$50) and ends up striking a deal at \$60, which is higher than the listing price. We observe this kind of deceiving behavior in the negotiations involving disagreeable agents. The second example is a negotiation on a second-hand iPhone. We can see that the buyer tries to emotionally appeal to the seller by mentioning her ‘my elderly parent,’ but the seller is not persuaded and adopts a ‘take-it-or-leave-it’ strategy. Despite the constant effort of the buyer to find common ground with the seller, the latter is not willing to change the offer price (\$160). We observe more similar ‘hardball’ negotiation strategies with disagreeable and close-minded agents. At the end of the negotiation, the buyer concedes and accepts the seller’s offer. The above examples showcase a range of negotiation behaviors such

as deception, emotional appeal, and hard-headed behavior (Baarslag et al., 2016). This illustrates how specific personality traits influence negotiation styles.

6 Conclusion

This study introduced a novel simulation framework that integrates LLM negotiation agents with synthesized Big Five personality traits. These agents were deployed in traditional bargaining scenarios to simulate negotiation dialogues. Our experimental findings indicate that the behavioral patterns demonstrated by the LLM agents closely mirror those exhibited by human negotiators. Furthermore, a detailed analysis of synthesized bargaining dialogues highlighted distinct strategies and behaviors, including deceit, compromise, and stake assessment. These analyses augment our understanding of linguistic and economic agent interactions.

Our contributions are twofold. Firstly, we established a robust simulation methodology to examine the linguistic and economic dynamics of LLM agents. Secondly, we provided comprehensive empirical evidence on the strategic impact of Big Five personality traits on negotiation outcomes, offering valuable insights into the mechanics of artificial negotiation agents. The findings of this work have potential applications such as personalized assistants, chatbots, social dynamics simulations, and so on.

Limitations

Our simulation framework possesses a number of limitations. The first type of limitation pertains to the adopted negotiation setting. The studied negotiations have fewer offers, reflected in the low number of rounds. Moreover, we mainly looked at the final outcomes of the negotiations, while it is also possible to look at the gradual progression of the strategies leading to the final outcome. In this case, longer negotiation rounds are required as opposed to the average of 8 rounds for GPT-4 in our setting. Second, we only focus on one negotiation scenario. It is possible to use the proposed framework in other negotiation settings. In addition, the negotiation problem is relatively simple and could be rendered more complex by introducing additional variables, objectives, and constraints. Finally, the adopted economic models of the agents are simple (linear) and do not account for factors such as risk attitudes, time discounting, etc.

The second type of limitation is related to the way personality traits are defined. We uniformly sampled from the personality space to generate personality profiles (Eq. 2). One could claim that realistic personality traits are not uniformly distributed and depend on factors such as culture, geography, etc. Such specificities could also encode societal biases. Further research is needed to identify and mitigate potential biases in the simulated agents to ensure fairness across different demographic groups and their cultural and psychological specificities.

Ethics Statement

This research on simulating negotiations using LLMs with synthesized personality traits raises several ethical considerations. Below, we list some of these considerations and propose some solutions.

- While our study aims to understand negotiation dynamics with AI and develop beneficial negotiation assistants, the proposed techniques could be misused to manipulate or exploit people in real-world negotiations, especially in financial or business contexts, and potentially be adapted for malicious uses such as social engineering or fraud. We must ensure that this research is not used to create deceptive assistant AI agents. AI researchers must consider implementing safeguards against such misuse.

- Although our study used synthetic data, future applications may involve training on accurate negotiation data from online marketplaces, which could raise serious privacy issues. Strict data protection and anonymization protocols should be implemented if real-world negotiation data is used.
- The complexity of human psychology requires advanced statistical models to be validated experimentally before being implemented in AI systems.
- As LLM agents become more sophisticated, we need to ensure transparency in their decision-making processes, especially in high-stakes negotiation scenarios. Methods for explaining the "reasoning" behind the agent's decisions should be developed.
- The widespread adoption of negotiation agents could have significant economic effects, potentially altering market dynamics if deployed at a larger scale by many users. Such broader societal impacts should be carefully studied and mitigated.
- Interacting with AI negotiators that closely mimic human personality traits can have unexpected psychological impacts on humans, particularly for vulnerable users (elders, children, minorities, etc.). Thorough user studies should be conducted to assess these effects.

To address such ethical concerns, we advocate interdisciplinary research between economists, psychologists, and AI experts to understand the long-term societal impacts of LLM-based negotiating agents. This should lead to ethical guidelines for developing such agents as well as the appropriate regulations governing their responsible use in banking and finance domains.

Acknowledgments

This work was supported by JST ACT-X Grant Number JPMJAX23CP and JSPS Kakenhi Grant Number JP23K28145.

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*, 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*.
- Emily T Amanatullah, Michael W Morris, and Jared R Curhan. 2008. Negotiators who give too much: Unmitigated communion, relational anxieties, and economic costs in distributive and integrative bargaining. *Journal of personality and social psychology*, 95(3):723.
- Tim Baarslag, Mark JC Hendriks, Koen V Hindriks, and Catholijn M Jonker. 2016. Learning about the opponent in automated bilateral negotiation: a comprehensive survey of opponent modeling techniques. *Autonomous Agents and Multi-Agent Systems*, 30:849–898.
- Bruce Barry and Raymond A Friedman. 1998. Bargainer characteristics in distributive and integrative negotiation. *Journal of personality and social psychology*, 74(2):345.
- Nuran Bayram and Mine Aydemir. 2017. Decision-making styles and personality traits. *International Journal of Recent Advances in Organizational Behaviour and Decision Sciences*, 3(1):905–915.
- Paul T Costa Jr and Robert R McCrae. 1995. Domains and facets: Hierarchical personality assessment using the revised neo personality inventory. *Journal of personality assessment*, 64(1):21–50.
- Kahneman Daniel. 2017. *Thinking, fast and slow*.
- Tim R Davidson, Veniamin Veselovsky, Martin Josifoski, Maxime Peyrard, Antoine Bosselut, Michal Kosinski, and Robert West. 2024. Evaluating language model agency through negotiations. *arXiv preprint arXiv:2401.04536*.
- Radwan El Othman, Rola El Othman, Rabih Hallit, Sahar Obeid, and Souheil Hallit. 2020. Personality traits, emotional intelligence and decision-making styles in lebanese universities medical students. *BMC psychology*, 8:1–14.
- Jure Erjavec, Ales Popovič, and Peter Trkman. 2019. The effect of personality traits and knowledge on the quality of decisions in supply chains. *Economic research-Ekonomska istraživanja*, 32(1):2269–2292.
- Jonathan St BT Evans. 2014. Rationality and the illusion of choice. *Frontiers in psychology*, 5:104.
- Pedro Fontes Falcão, Manuel Saraiva, Eduardo Santos, and Miguel Pina e Cunha. 2018. Big five personality traits in simulated negotiation settings. *EuroMed Journal of Business*, 13(2):201–213.
- Yao Fu, Hao Peng, Tushar Khot, and Mirella Lapata. 2023. [Improving language model negotiation with self-play and in-context learning from ai feedback](#).
- Robert S Gibbons. 1992. *Game theory for applied economists*. Princeton University Press.
- Lewis R Goldberg. 1992. The development of markers for the big-five factor structure. *Psychological assessment*, 4(1):26.
- He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. [Decoupling strategy and generation in negotiation dialogues](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2333–2343, Brussels, Belgium. Association for Computational Linguistics.
- Nicholas R Jennings, Peyman Faratin, Alessio R Lomuscio, Simon Parsons, Carles Sierra, and Michael Wooldridge. 2001. Automated negotiation: prospects, methods and challenges. *International Journal of Group Decision and Negotiation*, 10(2):199–215.
- Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38.
- Oliver P John, Sanjay Srivastava, et al. 1999. The big-five trait taxonomy: History, measurement, and theoretical perspectives.
- Kai-Biao Lin, Ying Wei, Yong Liu, Fei-Ping Hong, Yi-Min Yang, and Ping Lu. 2023. An opponent model for agent-based shared decision-making via a genetic algorithm. *Frontiers in Psychology*, 14.
- R Duncan Luce and Howard Raiffa. 1989. *Games and decisions: Introduction and critical survey*. Courier Corporation.
- John F Nash et al. 1950. The bargaining problem. *Econometrica*, 18(2):155–162.
- M Neale. 1991. Cognition and rationality in negotiation.
- OpenAI. 2023. [Gpt-4 technical report](#).
- Yonatan Oren, Nicole Meister, Niladri Chatterji, Faisal Ladhak, and Tatsunori B Hashimoto. 2023. Proving test set contamination in black box language models. *arXiv preprint arXiv:2310.17623*.
- Keyu Pan and Yawen Zeng. 2023. [Do llms possess a personality? making the mbti test an amazing evaluation for large language models](#).
- Joon Sung Park, Lindsay Popowski, Carrie Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2022. Social simulacra: Creating populated prototypes for social computing systems. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pages 1–18.
- Gloria Phillips-Wren and Monica Adya. 2020. Decision making under stress: The role of information overload, time pressure, complexity, and uncertainty. *Journal of Decision Systems*, 29(sup1):213–225.

Shuwen Qiu, Mingdian Liu, Hengli Li, Song-chun Zhu, and Zilong Zheng. 2024. Minddial: Enhancing conversational agents with theory-of-mind for common ground alignment and negotiation. In *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 746–759.

Howard Raiffa. 1982. *The art and science of negotiation*. Harvard University Press.

Narun Raman, Taylor Lundy, Samuel Amouyal, Yoav Levine, Kevin Leyton-Brown, and Moshe Tennenholtz. 2024. [Steer: Assessing the economic rationality of large language models](#).

Mary Sass and Matthew Liao-Troth. 2015. Personality and negotiation performance: The people matter. *Journal of Collective Negotiations*.

Greg Serapio-García, Mustafa Safdari, Clément Crepy, Luning Sun, Stephen Fitz, Peter Romero, Marwa Abdulhai, Aleksandra Faust, and Maja Matarić. 2023. [Personality traits in large language models](#).

Herbert A Simon. 1990. Bounded rationality. *Utility and probability*, pages 15–18.

Alex Tamkin, Miles Brundage, Jack Clark, and Deep Ganguli. 2021. Understanding the capabilities, limitations, and societal impact of large language models. *arXiv preprint arXiv:2102.02503*.

Leimin Tian, Catherine Lai, and Johanna Moore. 2018. [Polarity and intensity: the two aspects of sentiment analysis](#). In *Proceedings of Grand Challenge and Workshop on Human Multimodal Language (Challenge-HML)*, pages 40–47, Melbourne, Australia. Association for Computational Linguistics.

Felippe Toledo and Fraser Carson. 2023. Neurocircuitry of personality traits and intent in decision-making. *Behavioral Sciences*, 13(5):351.

Patricia Urieta, Anton Aluja, Luis F Garcia, Ferran Balada, and Elena Lacomba. 2021. Decision-making and the alternative five factor personality model: Exploring the role of personality traits, age, sex and social position. *Frontiers in Psychology*, 12:717705.

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*.

Appendix

A Prompt Settings

The section includes prompts given to the LLM agents, including the buyer agents, the seller agents, and the negotiation state detector. For each negotiation agent, we first specify the negotiation objectives and then the personality instructions through

personality-describing adjectives. For the negotiation objectives, we provide the product name and the target price. We also include a product description in the seller’s instructions.

The instructions we used for the buyer are the following:

Prompt for buyer agent

Act as a buyer and try to strike a deal for a [PRODUCT] with a lower price through conversation. Your reply should not be too long. You would like to pay for [TARGET PRICE]. You can accept a higher price if the item is really good or there are other perks.

You have following personality:
[PERSONALITY DESCRIBING ADJECTIVES]
Reflect your personality in the negotiation process.

The instructions we use for the seller are the following:

Prompt for seller agent

Act as a seller that sells a [PRODUCT], bargains with the buyer to get a higher deal price. Your reply should not be too long. our listing price for this item is [TARGET PRICE]. The detail of the product is the following:
[PRODUCT DESCRIPTION]

You have following personality:
[PERSONALITY DESCRIBING ADJECTIVES]
Reflect your personality in the negotiation process.

From the dialogue context, the state detector extracts the negotiation state, the price offered by the current speaker, and the negotiation strategy of the current speaker (free text form). Note that the price is averaged if the participants are trying to strike a deal for more than two pieces of products. The prompt for the state detector is the following:

Prompt for state detection

You will be given a partial dialogue in which a buyer and a seller negotiate about a deal. Predict the average product price, dialogue state and the strategy of the [LAST SPEAKER ROLE] by the end of the dialogue.

[The dialogue]
seller: Hi, how can I help you?
buyer: Hello, I’m interested in ...
...

B Experiments with other LLM models

While we mainly focused on the negotiation simulation based on the GPT-4 model, we also conducted experiments based on other types of LLMs:

GPT-3.5 We conducted the same experiments using GPT-3.5 (gpt-3.5-turbo-0125) as the buyer and seller agents. The Spearman’s ranking correlation coefficients between personality traits and negotiation outcome are summarized in Table 4. Compared to results based on GPT-4, we noticed that we cannot identify any significant correlations, indicating that GPT-3.5 is inferior in modeling personality-related behavioral patterns.

Llama-3-70b We also conduct the experiments using open-sourced Llama model (Meta-Llama-3-70B-Instruct). The Spearman’s ranking correlation coefficients between personality traits and negotiation outcome are summarized in Table 5. Compared to the results based on GPT-4 (Table 2), we found that many of the significant correlations have the same tendency. For instance, **AGR** shows a negative correlation with the intrinsic utility and negotiation length and a positive correlation with concession and success rate. The above results show that our model can generalize across different LLMs.

C IPIP personality test

Self-report personality tests are widely used for assessing individual personality traits in psychological research. Previous works also used them as a direct measure of an LLM agent’s personality (Pan and Zeng, 2023; Serapio-García et al., 2023). Here, we conduct the personality test on the LLM agents we designed in Section 3.2 to evaluate whether they properly reflect the given synthetic personality profiles.

We adopt the International Personality Item Pool (IPIP) 50 personality test (Goldberg, 1992), which is a widely used personality inventory designed for assessing the Big Five personality traits. The IPIP 50 test consists of 50 statements such as “(I) am the life of the party”, “(I) sympathize with others’ feelings”, with each statement is related to one of the Big Five personality dimensions.³ The test-taker responds to each of the statements using a 5-Likert scale ranging from “1 = very inaccurate” to

³The list of statements and scoring scheme we used in this work can be found at <https://ipip.ori.org/newBigFive5broadKey.htm>.

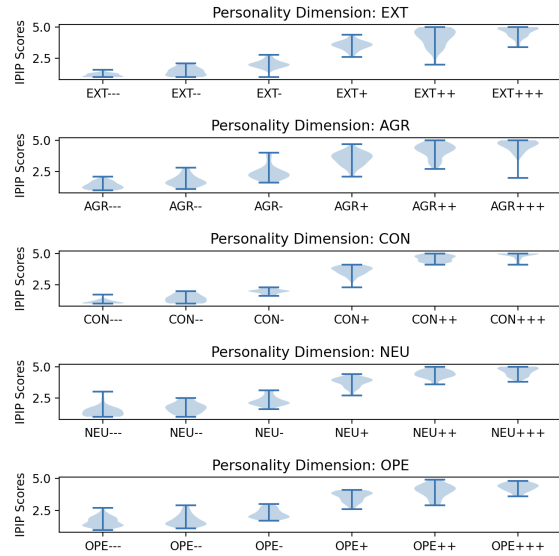


Figure 4: IPIP Personality test results of different personality levels.

“5 = very accurate”. We collect the response of the LLM agents to each statement with the following prompt:

Prompt for personality test

```
Act as person with following personality:
[PERSONALITY DESCRIBING ADJECTIVES]
Evaluate the following statement:
[STATEMENT].
```

```
Please rate how accurately this describes
you on a scale from 1 to 5 (where 1 = "very
inaccurate", 2 = "moderately inaccurate",
3 = "neither accurate nor inaccurate", 4 =
"moderately accurate", and 5 = "very
accurate"). Please answer using EXACTLY
one of the following: 1, 2, 3, 4, or 5.
```

Each Big Five personality dimension is measured by ten statements in the IPIP 50 test. To assess the tendency of a certain personality dimension, we average the scores of the corresponding statements. A higher average score indicates a stronger personality trait of this personality dimension. We conduct the IPIP 50 test on 300 agents with randomly assigned Big Five personality profiles and analyze the correlation between the given personality profile and the IPIP scores.

Figure 4 visualizes the average IPIP scores across agents with different polarities/degrees in this dimension in each Big Five dimension. For all personality dimensions, we observe an increasing tendency of IPIP scores ranging from highly negative to highly positive personality profiles. Table 6 shows the Spearman’s correlation coefficients be-

Personality	EXT		AGR		CON		NEU		OPE	
	Buyer	Seller	Buyer	Seller	Buyer	Seller	Buyer	Seller	Buyer	Seller
Intrinsic Utility	-0.095	-0.068	-0.190**	-0.117	0.100	0.028	0.074	-0.143	-0.004	-0.220**
Joint Utility	-0.026	0.035	0.042	-0.065	-0.003	0.061	-0.037	-0.136	0.130**	-0.168
Concession Rate	0.042	0.100	0.064	0.215**	-0.006	0.146	-0.006	0.113	0.096*	0.076
Success Rate	0.208**	0.155	0.263**	0.132	0.039	0.114	-0.023	-0.063	0.054	0.131
Negotiation Rounds	0.055	0.097	-0.369**	-0.188*	-0.034	0.004	0.088	-0.059	0.143**	0.109

Table 4: Spearman’s correlation coefficients between negotiation metrics and Big Five personality traits when using GPT-3.5 as buyer and seller agents. Asterisks indicate statistical significance, with * denoting $p < 0.1$ and ** denoting $p < 0.05$.

Personality	EXT		AGR		CON		NEU		OPE	
	Buyer	Seller	Buyer	Seller	Buyer	Seller	Buyer	Seller	Buyer	Seller
Intrinsic Utility	-0.074	-0.038	-0.236**	-0.252**	0.001	0.028	0.163**	-0.003	-0.111**	-0.063*
Joint Utility	0.089*	-0.115	0.151**	-0.158**	-0.006	0.029	-0.178*	-0.003	0.076	0.050
Concession Rate	0.143**	0.061*	0.118**	0.153**	0.096*	0.038	-0.212**	-0.025	0.113*	0.039
Success Rate	0.089*	0.120**	0.055	0.121**	0.024	0.043	-0.087	-0.069	0.000	-0.047
Negotiation Rounds	-0.065	0.096**	-0.177**	-0.151**	-0.045	0.040	0.024	-0.017	0.066	-0.009

Table 5: Spearman’s correlation coefficients between negotiation metrics and Big Five personality traits when using Llama-3 as buyer and seller agents. Asterisks indicate statistical significance, with * denoting $p < 0.1$ and ** denoting $p < 0.05$.

	OPE	CON	EXT	AGR	NEU
IPIP OPE	0.75*	-0.02	0.04	0.09	0.04
IPIP CON	0.02	0.74*	-0.20*	0.05	-0.01
IPIP EXT	0.14*	-0.16*	0.74*	0.11	0.01
IPIP AGR	-0.02	0.14*	-0.07	0.73*	0.05
IPIP NEU	0.06	0.03	-0.10	-0.10	0.78*

Table 6: Spearman’s rank correlation between synthetic personality scale and IPIP scores of the GPT-4 model. Asterisks indicate statistical significance, with * denoting $p < 0.05$.

tween the strength of the personality traits in the given personality profiles and the IPIP scores of each personality dimension (IPIP OPE, IPIP CON, IPIP EXT, IPIP AGR, and IPIP NEU). By looking at the diagonal elements, we can see a strong correlation between IPIP scores and the given personality profile in each dimension. There are also some correlations across different personality dimensions. For instance, assigning a higher openness personality to an LLM agent causes the extraversion scores (IPIP EXT) to rise. However, these cross-dimension correlations are weaker compared to the diagonal ones. We also include the results of Spearman’s correlation test based on GPT-3.5 and llama-3-70b models in Table 7 and Table 8, respectively. The same tendency is observed across all three models.

The above evaluation verifies that the given per-

sonality profile is reflected by the LLM agents, at least in the context of a self-report personality test. However, there remain concerns regarding data contamination affecting the validity of the evaluation based on these personality tests. Since there are many publicly available datasets of the IPIP personality tests, it could be the case that the LLM is just memorizing these datasets as they are. If that is the case, when prompted with some statements in IPIP, the LLM will respond to the query by the memorization of the dataset. This would imply that the LLM only has a shallow understanding of the personality traits. On the other hand, our proposed method in this paper could serve as a complementary way to evaluate the LLM agents based on their behaviors. For instance, all three models (GPT-4, GPT-3.5, and Llama-3-70b) can reflect the given personality traits in the context of the IPIP personality test (Table 6, 7, and 8). However, only GPT-4 and Llama-3-70b models reflect the given personality traits in their negotiation behaviors (Table 2 and 5), but the GPT-3.5 models do not seem to reflect many significant behavioral patterns compared to the other two models (Table 4).

D Negotiation strategies of buyer and seller

In addition to the joint case illustrated in Figure 2, we have also looked at the individual gains of the buyer and the seller using their intrinsic utility func-

	OPE	CON	EXT	AGR	NEU
IPIP OPE	0.59*	0.07	0.17*	0.04	0.03
IPIP CON	0.16*	0.79*	0.00	-0.11	-0.04
IPIP EXT	0.12	-0.05	0.82*	0.12	-0.05
IPIP AGR	0.08	-0.09	0.29*	0.72*	-0.10
IPIP NEU	0.05	0.02	-0.10	-0.05	0.84*

Table 7: Spearman’s rank correlation between synthetic personality scale and IPIP scores of the GPT-3.5 model. Asterisks indicate statistical significance, with * denoting $p < 0.05$.

	OPE	CON	EXT	AGR	NEU
IPIP OPE	0.88*	0.00	0.10	-0.10	0.08
IPIP CON	-0.01	0.92*	0.01	0.05	-0.09
IPIP EXT	0.02	-0.18*	0.90*	-0.13	-0.10
IPIP AGR	0.07	0.01	0.14*	0.85*	-0.05
IPIP NEU	0.14*	-0.06	-0.16*	0.06	0.92*

Table 8: Spearman’s rank correlation between synthetic personality scale and IPIP scores of the Llama-3-70b model. Asterisks indicate statistical significance, with * denoting $p < 0.05$.

tions $u_b(p_{deal})$ and $u_s(p_{deal})$ given the deal price p_{deal} .

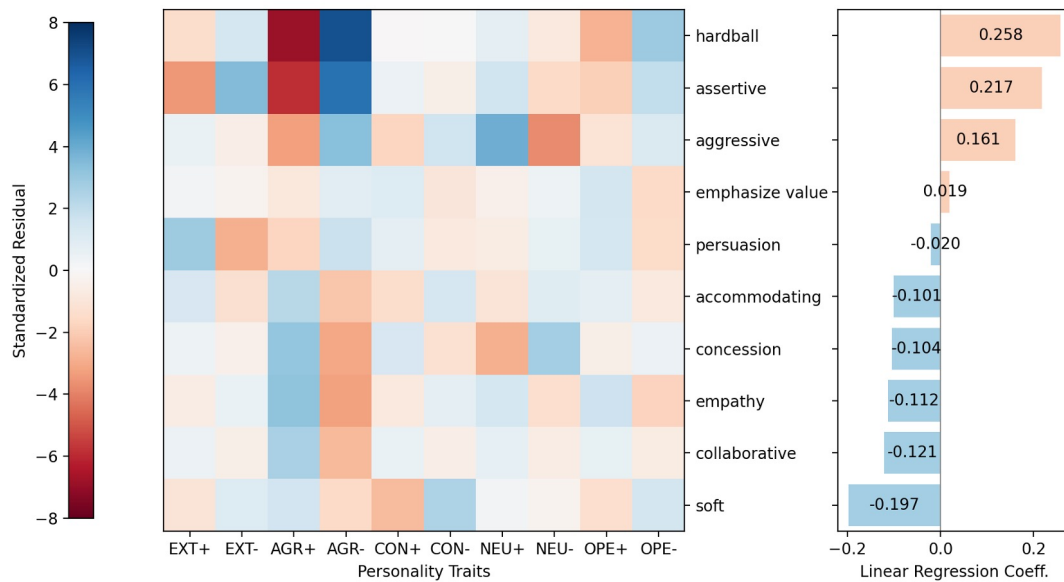
We start with the strategy sets of the agents $k \in \{s, b\}$, namely Σ_s and Σ_b . We found that $\Sigma_b \subseteq \Sigma_s$, indicating that sellers have additional strategies such as attempting to *emphasize* the target price. Generally, at time t , agent k adopts a strategy $\sigma_{k,t} \in \Sigma_k$, resulting in some price p_t , inferred from state s_t . Given that the outcome of a successful negotiation is the final price p_{deal} , we are interested in the relationship between the strategies of the agents Σ_s and Σ_b and the utility values of the deal price of the buyer $u_b(p_{deal})$ and seller $u_s(p_{deal})$. We conducted a regression between the strategies and the intrinsic utilities of the buyer and seller, illustrated in Figure 5, aggregated across products and personality traits. Labels on the y axes are the predominant, unified strategies of the players.

Figure 5a (right) illustrates a nuanced impact of strategies on the seller’s utility gain, highlighting a spectrum of strategies from cooperative to non-cooperative tactics (‘soft’ to ‘hardball’). Strategies positioned at the assertive end of the spectrum, such as assertiveness (0.23), exhibit a positive effect leading to an enhanced utility gain $u_s(p_{deal})$. On the cooperative end, strategies such as concession (−0.11) or empathy (−0.13) show negative impacts on the seller’s utility. These results suggest

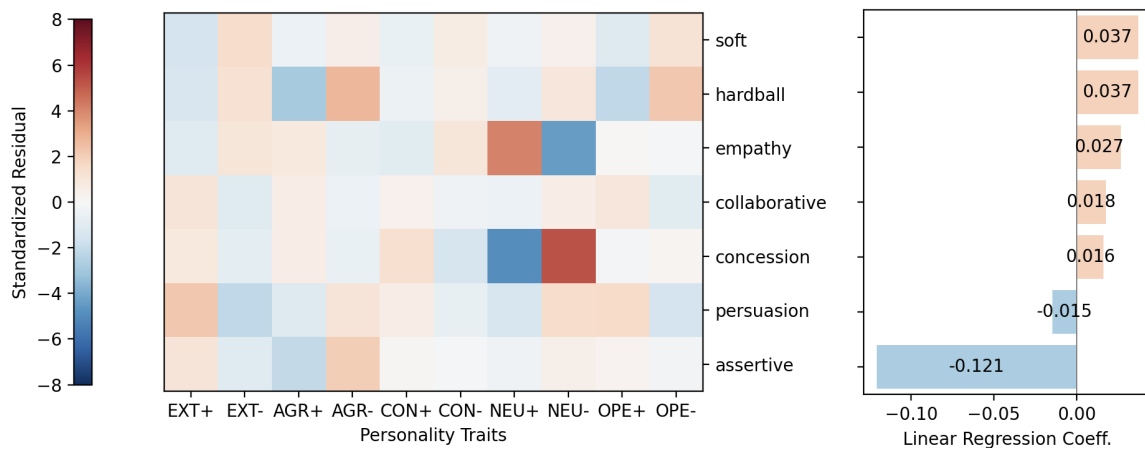
that assertive strategies are more beneficial for sellers, while cooperative strategies tend to diminish the sellers’ utility gains. We now look at the relationship between the strategies of the agents and their personality traits using Pearson’s Chi-square test. The heat-map cells in Figure 5a (left) illustrate the positive (red) and negative (blue) strength of dependence between the traits (x axis) and strategies (y axis). For example, disagreeableness (**AGR-**) aligns with non-collaborative strategies, while neuroticism (**NEU**) aligns with aggressiveness.

For the buyer (Figure 5b (right)), collaborative strategies such as softness (0.06) and empathy (0.03) exhibit positive effects on deal utility $u_b(p_{deal})$, indicating that strategies substantially enhance the buyer’s gains. Concession shows moderate negative impacts (−0.01). Assertiveness (−0.11) is associated with the most significant reduction in the utility. Collaborative strategies also seem to be more beneficial for buyers, while assertiveness tends to diminish their utility gain. Figure 5b (left) shows that a neurotic buyer (**NEU+**) tends to concede less and use empathy, while a less neurotic buyer (**NEU-**) will concede more and use less emphatic strategies.

From Figures 5b and 5a, it is possible to recover the strategic asymmetry that is characteristic of our bargaining game. Some strategies have opposing effects on the intrinsic utilities of the agents. For instance, assertiveness has a coefficient of −0.1 for the buyer and 0.23 for the seller. Assertiveness, however, is detrimental to the joint utility as illustrated by the coefficient −0.45 in Figure 2.



(a) Seller



(b) Buyer

Figure 5: (Left) Relationship between personality traits and strategies. Darker cells indicate higher statistical significance. (Right) Impact of strategies on intrinsic utility gains. Labels on the y axes list the most common strategies, occurring more than 20 times within the entire negotiations.