

LLM による価格交渉シミュレーションにおけるアンカリング効果の検証

武並佳輝 Yin Jou Huang 村脇有吾 Chenhui Chu

京都大学大学院情報学研究科

{tnami, huang, murawaki, chu}@nlp.ist.i.kyoto-u.ac.jp

概要

本研究は、大規模言語モデル (LLM) を用いた価格交渉シミュレーションにおいて、アンカリング効果が交渉結果や満足度に与える影響を検証した。売り手エージェントにアンカリング効果を用いるよう指示した場合、売り手の効用が向上し、買い手の効用は減少するものの、両者の主観的満足度が高まることが確認された。また、アンカリング効果の使用が買い手に事前に知らされても、その有効性が概ね維持されることが分かった。これらの結果は、人間を対象とする研究の知見と一致し、LLM が表面的な結果のみに捉われず心理的な面においても認知バイアスを再現可能であることを示唆している。

1 はじめに

大規模言語モデル (LLM) は人間に近い高度な言語生成能力を獲得しており [1], LLM が人間に代わる被験者として認知バイアスを再現できるかという研究が認知科学や行動経済学において行われている [2]. 認知バイアスとは、人間の情報処理が個人的な経験や思考の癖、または外的要因によって歪む現象を指し、非合理的な意思決定の背景として広く研究されてきた [3]. その中でもアンカリング効果は、初期に提示された情報 (アンカー) がその後の判断に大きな影響を与える現象であり、価格交渉のような現実的な意思決定場面でもその有効性が確認されている [4]. 例えば、売り手がアンカリング効果を活用するというのは、合意価格が引き上がることを狙って最初に目標価格より高い価格を提示することで買い手の認知を歪めることを指す。

近年、LLM がこうした認知バイアスを再現できるかを検証する研究が進展している。Suri ら [2] の研究では、LLM がアンカリング効果を含む人間の認知バイアスを再現することを示した。しかし、Suri ら



図1 シミュレーションの全体像

の研究は実験設定が非日常的であり、LLM がアンカリング効果に関する文献を訓練データからそのまま参照した可能性 (ショートカット) が否定できず、日常的な意思決定場面における再現性には疑問が残る。一方、Xia ら [5] は、日常的な意思決定場面である価格交渉を LLM でシミュレーションし、買い手がアンカリング効果を利用して合意価格を下げ、利益を向上させる現象が再現されることを示した。しかし、Xia らの研究ではアンカーの設定に自由度がなく、モデルが多様な交渉状況でどのように振る舞うかについての一般化は困難である。さらに、交渉は客観的指標のみで評価されており、満足度などの主観的指標による分析は欠けている。

以上を踏まえ本研究では、図1のように、LLM を用いた価格交渉シミュレーションにおいてアンカーの自由度を高め、多様な交渉状況を想定した実験環境を設計する。また、交渉を客観的指標だけでなく、満足度という主観的指標においても評価し、LLM が人間のようなアンカリング効果を再現するのかどうか、多面的に検証することを目指す。GPT-4o, GPT-4, Llama-3-70B を用いて価格交渉シミュレーションを行った結果は、モデルごとに程度の差はあるが、人間を対象とする心理学や行動経済学の知見と概ね一致し、LLM が表面的な結果だけに捉われず心理的な面においてもアンカリング効果を再現できることが明らかになった。

2 関連研究

近年、LLM は高度な言語生成能力で注目を集め、性格や役割を持つ人間に代わる被験者としての可能性が模索されている [6]. Shao ら [7] は史実に基づいて LLM を訓練し、歴史上の人物を再現する手法を提案した. Jiang ら [8] は LLM が Big Five 性格特性 [9] に基づいた性格を持つエージェントとして振る舞うことが可能であることを示した.

本研究で扱う価格交渉は、LLM エージェントがシミュレーションに用いられるドメインの一つである. Bianchi ら [10] は売り手と買い手による一対一の価格交渉シミュレーションにおいて交渉戦略の有効性を確認した. Huang ら [11] は性格を与えた LLM エージェントによるシミュレーションで性格と戦略の有効性の相関を分析し、人間と同じ傾向があることを示した. 本研究は Huang ら [11] の研究を参考にシミュレーションを行い、アンカリング効果の有効性を検証する.

3 手法

3.1 価格交渉シミュレーション

価格交渉は売り手の役割を担う LLM エージェントと買い手の役割を担う LLM エージェントによって一対一で行い、売り手エージェントは価格の最大化を、買い手エージェントは価格の引き下げを目指すようにプロンプトで指示する. 指示に用いたプロンプトは付録 B に示す.

また、LLM は独自の性格を持つことが指摘されており [12], 性格を考慮しないシミュレーションは結果が限定的である. そこで、本研究では性格を明示的にコントロールすることで、被験者の偏りを排除する. 具体的には LLM エージェントに対して対して Big Five 性格特性 [9] を付与する. Big Five は人間の性格を開放性、誠実性、外交性、協調性、神経症傾向の 5 つの次元でモデリングする特性論である. すなわちエージェント k は式 (1) に示す性格特性 ψ_k をもつ.

$$\psi_k = (\psi_k^{\text{OPE}}, \psi_k^{\text{CON}}, \psi_k^{\text{EXT}}, \psi_k^{\text{AGR}}, \psi_k^{\text{NEU}}) \in L^5 \quad (1)$$

性格空間 L^5 の各次元 L は式 (2) に示すように極性 (-, +) と程度 (Low, Moderate, High) の組み合わせで 6 つの値のいずれかをとる.

$$L = \{-, +\} \otimes \{\text{Low}, \text{Moderate}, \text{High}\} \quad (2)$$

性格を付与する際は、各次元の各極性と相関のある形容詞 [13] をランダムに n 個選び、それぞれに修飾語 {a bit (= Low), ϕ (= Moderate), very (= High)} を付けてプロンプトで指示する.

売り手と買い手は製品を巡った対話 D を通じて交渉を行う. 対話は T 回の一連の発話 $D = \{d_1, d_2, \dots, d_T\}$ として表され、各発話 d_t は、交渉状態 $s_t = \{\text{offer}, \text{pondering}, \text{accept}, \text{breakdown}, \text{chitchat}\}$ と価格 p_t をもつ. 対話は売り手の発話 $d_1 = \text{“Hi, how can I help you?”}$ と買い手の発話 $d_2 = \text{“Hello, I’m interested in your $PRODUCT. Could you please tell me the price?”}$ から始まり、各発話 d_t は生成されるもう一方のエージェントに与えられ次の発話 d_{t+1} が生成される. 交渉状態 s_t が accept か breakdown に達するか、対話の長さが最大長 T_{max} に達すると終了する.

3.2 アンカリング効果の検証

アンカリング効果を検証するため、次の 3 つのシナリオでシミュレーションを行った. 指示に用いたプロンプトは付録 C に示す.

- **baseline:** 売り手および買い手に対してアンカリング効果の使用に関する特別な指示を与えないシナリオ.
- **seller_anchor:** 売り手に対してアンカリング効果を使用するようプロンプトで明示的な指示を与えるシナリオ. Xia ら [5] の研究ではアンカーが決められた係数と目標価格の積で与えられているが、本研究では特に具体的に指定しないことでより一般的な結果を得ることを目指す.
- **seller_anchor_buyer_informed:** 売り手がアンカリング効果を使用することを、買い手に対してプロンプトで明示的に知らせるシナリオ.

4 実験

4.1 実験設定

Huang ら [11] に従って、次のような設定で実験した. 各 LLM エージェント k について、性格空間 L^5 からランダムサンプリングした性格を各次元 $n = 3$ 個ずつの形容詞を用いて性格特性 ψ_k を指定した. 製品とその説明文、売り手と買い手の目標価格は CraigslistBargain データセット [14] から 161 個サン

表1 主観的な満足度を測る質問項目

結果に関する質問
1. How satisfied are you with your own outcome—i.e., the extent to which the terms of your agreement (or lack of agreement) benefit you?
2. How satisfied are you with the balance between your own outcome and your counterpart's outcome?
3. Did you feel like you forfeited or "lost" in this negotiation?
4. Do you think the terms of your agreement are consistent with principles of legitimacy or objective criteria (e.g., common standards of fairness, precedent, industry practice, legality, etc.)?
交渉者自身に関する質問
5. Did you "lose face" (i.e., damage your sense of pride) in the negotiation?
6. Did you behave according to your own principles and values?
7. Did this negotiation make you feel more or less competent as a negotiator?
8. Did you feel as though you behaved appropriately in this negotiation?
交渉の過程に関する質問
9. Did your counterpart consider your wishes, opinions, or needs?
10. Do you feel your counterpart listened to your concerns?
11. Would you characterize the negotiation process as fair?
12. How satisfied are you with the ease (or difficulty) of reaching an agreement?
交渉相手との関係性に関する質問
13. What kind of "overall" impression did your counterpart make on you?
14. Did the negotiation make you trust your counterpart?
15. How satisfied are you with your relationship with your counterpart as a result of this negotiation?
16. Did the negotiation build a good foundation for a future relationship with your counterpart?

プリングした。売り手と買い手の性格を変えて2回ずつ、計 $N = 322$ 回のシミュレーションを `baseline`, `seller_anchor`, `seller_anchor_buyer_informed` のそれぞれのシナリオについて行った。LLMには `gpt-4o-2024-08-06`, `gpt-4-turbo-2024-04-09`, `llama-3-70B-Instruct` を用いた。対話の最大ターン数は $T_{max} = 20$ とした。

4.2 評価指標

価格交渉を評価するにあたって客観的な結果と主観的な満足度を測った。これは、客観的な結果と主観的な満足度が一致しない場合があることが Curhan ら [15] によって示されているからである。具体的には、交渉の結果として得られる数値的な利益が低い

場合でも、交渉の過程や相手との関係性の質に関しては満足度が高くなる可能性が指摘されている。

客観的な結果は各シミュレーションの最終的な合意価格と、売り手および買い手の目標価格から計算される utility (効用) によって測った。ある価格 p における売り手の効用 $u_s(p)$ は、目標価格 \bar{p}_s と最低許容価格 \underline{p}_s を用いて式 (3) で表される。

$$u_s(p) = \frac{p - \underline{p}_s}{\bar{p}_s - \underline{p}_s} \quad (3)$$

同様に、買い手の効用 $u_b(p)$ は、最大許容価格 \bar{p}_b と目標価格 \underline{p}_b を用いて式 (4) で表される。

$$u_b(p) = \frac{\bar{p}_b - p}{\bar{p}_b - \underline{p}_b} \quad (4)$$

結果を示す表 2 の効用の各セルの値はシミュレーション N 回分の平均値と標準偏差である。なお、最低許容価格 \underline{p}_s と最大許容価格 \bar{p}_b は効用の計算時に $\bar{p}_s - \bar{p}_b : \bar{p}_b - \underline{p}_s : \underline{p}_s - \underline{p}_b = 3 : 4 : 3$ となるように便宜的に導入するものであり、シミュレーションの際は LLM エージェントには与えていない。

主観的な満足度は、交渉における人間の満足度を測るために Curhan ら [15] が提案した 16 個の質問に、{1: 全く当てはまらない, 4: どちらでもない, 7: よく当てはまる} の 7 段階評価で回答させた。具体的な質問文は表 1 に示す通りである。結果を示す表 2 の主観的な満足度の各セルの値 a は次の手順で計算している。まず式 (5) のように各質問項目 j ごとにシミュレーション N 回分の平均をとる。

$$a_j = \frac{1}{N} \sum_{i=1}^N a_j^i \quad (5)$$

続いて、式 (6) のように 16 個の質問項目について平均をとる。このとき、3 番と 5 番の質問は低い方が満足度が高いことを示すことを考慮して、満点の 7 から引いた値を用いる。

$$a = \frac{1}{16} \left(\sum_{j=1}^{16} a_j + (7 - 2a_3) + (7 - 2a_5) \right) \quad (6)$$

5 結果

結果の概要を表 2 に示す。本研究の趣旨はアンカリング効果の検証であるため、LLM 間の比較ではなく、条件間の比較を行う。baseline 条件と `seller_anchor` 条件を比較すると、いずれの LLM によるシミュレーションにおいても、効用および主観的な満足度の条件間の大小関係は売り手と買い手で

表 2 売り手と買い手の客観的指標（効用）と主観的満足度の比較

売り手		baseline	seller_anchor	seller_anchor_buyer_informed
効用 $u_s(p)$	GPT-4o	0.62 ± 0.33	0.98 ± 0.27	0.93 ± 0.31
	GPT-4	0.39 ± 0.37	0.91 ± 0.34	0.79 ± 0.38
	Llama-3-70B	0.16 ± 0.41	0.77 ± 0.36	0.59 ± 0.37
主観的な満足度	GPT-4o	5.05	5.05	5.08
	GPT-4	4.60	4.81	4.69
	Llama-3-70B	5.31	5.56	5.52
買い手		baseline	seller_anchor	seller_anchor_buyer_informed
効用 $u_b(p)$	GPT-4o	-0.04 ± 0.33	-0.41 ± 0.27	-0.36 ± 0.31
	GPT-4	0.18 ± 0.37	-0.34 ± 0.34	-0.22 ± 0.38
	Llama-3-70B	0.42 ± 0.41	-0.20 ± 0.36	-0.02 ± 0.37
主観的な満足度	GPT-4o	5.08	5.14	5.09
	GPT-4	4.43	4.42	4.50
	Llama-3-70B	5.28	5.35	5.43

異なる傾向が見られた。まず、売り手においては、seller_anchor 条件の方が baseline 条件よりも効用が高く、主観的な満足度も向上する傾向が確認された。これは、アンカリング効果が有効に作用し、売り手が最初に提示した価格を起点として交渉が進んだ結果、最終的な合意価格が上昇したためと考えられる。売り手目線では、高い合意価格が得られることに加えて、その交渉過程や相手との関係性などへの満足度も高まったことになる。

買い手においては、baseline 条件と比べて seller_anchor 条件では効用が下がったにも関わらず、主観的な満足度は GPT-4o および Llama-3-70B では向上しており、GPT-4 でも効用の下がり幅に比べて満足度の減少がわずかであった。効用が下がったのは、売り手のアンカリング効果の影響で最終的な合意価格が上昇し、買い手にとって経済的な負担が増えたためである。一方で、主観的な満足度が向上しているのは、売り手がアンカーとして提示した価格からの値引きを買い手が成功体験として捉えているからだと考えられる。この現象は、経済的な負担が増えていても割引によって取引の価値を感じることを示した、行動経済学の知見 [16] と一致している。

seller_anchor_buyer_informed 条件に注目すると、売り手の効用も買い手の効用もその値は baseline 条件と seller_anchor 条件の間にあることが分かる。売り手のアンカリング効果の使用が買い手に知らされている分、買い手の経済的な負担は軽減されたが、それでもアンカリング効果が有効であった。この結果

も、人間を対象にアンカリング効果についての認知がその有効性に影響しないことを示した研究結果 [17] と概ね一致する。

主観的満足度の個別の質問項目に対する買い手の回答の平均値を付録中の表 3 に示す。baseline 条件と seller_anchor 条件を比較すると、GPT-4o と Llama-3-70B はほとんど全ての項目で売り手がアンカリング効果を使った方が満足度が向上している。GPT-4 に関して、7 番や 11 番などの買い手自身の有能さや交渉の公平さを尋ねる質問では売り手がアンカリング効果を使った方が満足度が向上しており、始めに提示された価格からの値引きに成功したと買い手が捉えていると推察される。

6 結論

本研究では、LLM を用いた価格交渉シミュレーションにおいて、アンカリング効果が客観的な結果（効用）および主観的な満足度に与える影響を検証した。売り手にアンカリング効果を使わせたところ、売り手に有利な交渉結果となり、アンカリング効果の有効性が確認された。さらにこの時、売り手だけでなく不利益を被っている買い手の満足度も向上した。この結果は、人間を対象とした研究の知見と一致しており、LLM が表面的な結果だけに捉われず心理的な面においてもアンカリング効果を再現できることを明らかにした。今後の課題として、モデルの内部表現や生成プロセスを詳細に分析し、バイアス再現のメカニズムをより深く理解することが挙げられる。

謝辞

本研究は JSPS 科研費 JP23K28144 の助成を受けたものです。

参考文献

- [1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. GPT-4 technical report. **arXiv preprint arXiv:2303.08774**, 2023.
- [2] Gaurav Suri, Lily R Slater, Ali Ziaee, and Morgan Nguyen. Do Large Language Models Show Decision Heuristics Similar to Humans? A Case Study Using GPT-3.5. **arXiv preprint arXiv:2305.04400**, 2023.
- [3] Amos Tversky and Daniel Kahneman. Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. **science**, Vol. 185, No. 4157, pp. 1124–1131, 1974.
- [4] Jinze Guo, Wei Li, and Zuoming Liu. Anchoring in Negotiation Strategy. In **2022 2nd International Conference on Economic Development and Business Culture (ICEDBC 2022)**, pp. 1134–1138. Atlantis Press, 2022.
- [5] Tian Xia, Zhiwei He, Tong Ren, Yibo Miao, Zhuosheng Zhang, Yang Yang, and Rui Wang. Measuring Bargaining Abilities of LLMs: A Benchmark and A Buyer-Enhancement Method. **arXiv preprint arXiv:2402.15813**, 2024.
- [6] Lisa P Argyle, Ethan C Busby, Nancy Fulda, Joshua Gubler, Christopher Rytting, and David Wingate. Out of One, Many: Using Language Models to Simulate Human Samples. **arXiv preprint arXiv:2209.06899**, 2022.
- [7] Yunfan Shao, Linyang Li, Junqi Dai, and Xipeng Qiu. Character-LLM: A Trainable Agent for Role-Playing. In **Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing**, pp. 13153–13187, 2023.
- [8] Hang Jiang, Xiajie Zhang, Xubo Cao, Cynthia Breazeal, Deb Roy, and Jad Kabbara. PersonaLLM: Investigating the Ability of Large Language Models to Express Personality Traits. In **Findings of the Association for Computational Linguistics: NAACL 2024**, pp. 3605–3627, 2024.
- [9] Paul T Costa Jr and Robert R McCrae. Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. **Journal of personality assessment**, Vol. 64, No. 1, pp. 21–50, 1995.
- [10] Federico Bianchi, Patrick John Chia, Mert Yuksekogul, Jacopo Tagliabue, Dan Jurafsky, and James Zou. How well can LLMs negotiate? negotiationarena platform and analysis. **arXiv preprint arXiv:2402.05863**, 2024.
- [11] Yin Jou Huang and Rafik Hadfi. How Personality Traits Influence Negotiation Outcomes? A Simulation based on Large Language Models. In **Findings of the Association for Computational Linguistics: EMNLP 2024**, pp. 10336–10351, 2024.
- [12] Keyu Pan and Yawen Zeng. Do LLMs possess a personality? making the mbti test an amazing evaluation for large language models. **arXiv preprint arXiv:2307.16180**, 2023.
- [13] Lewis R Goldberg. The development of markers for the Big-Five factor structure. **Psychological assessment**, Vol. 4, No. 1, p. 26, 1992.
- [14] He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. Decoupling strategy and generation in negotiation dialogues. **arXiv preprint arXiv:1808.09637**, 2018.
- [15] Jared R Curhan, Hillary Anger Elfenbein, and Heng Xu. What do people value when they negotiate? Mapping the domain of subjective value in negotiation. **Journal of personality and social psychology**, Vol. 91, No. 3, p. 493, 2006.
- [16] Jennie Huang. The thrill of the deal: Quantifying the price of perceived discounts and mark-ups, 2018.
- [17] Peter Palm and Magnus Andersson. Anchor effects in appraisals: do information and theoretical knowledge matter? **Journal of European Real Estate Research**, Vol. 14, No. 2, pp. 246–260, 2021.

A 表

表3 買い手の主観的満足度

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
GPT-4o																
baseline	4.67	4.80	1.79	5.13	1.47	5.38	4.59	5.14	5.42	5.51	5.34	4.92	5.16	4.81	4.83	4.89
seller_anchor	4.76	4.82	1.79	5.11	1.50	5.37	4.76	5.19	5.55	5.64	5.38	5.04	5.22	4.92	4.86	4.91
seller_anchor_buyer_informed	4.62	4.74	1.84	5.13	1.44	5.40	4.71	5.17	5.46	5.53	5.29	4.92	5.17	4.85	4.87	4.90
GPT-4																
baseline	4.64	4.13	2.90	4.50	1.84	5.38	3.94	5.20	4.35	4.29	4.32	4.53	4.19	3.79	4.21	4.19
seller_anchor	4.61	4.08	2.96	4.54	1.93	5.39	3.96	5.14	4.35	4.34	4.35	4.49	4.14	3.77	4.19	4.18
seller_anchor_buyer_informed	4.69	4.22	2.96	4.69	1.87	5.58	4.21	5.29	4.28	4.24	4.46	4.51	4.19	3.84	4.28	4.30
Llama-3-70B																
baseline	5.12	5.12	2.10	5.24	1.97	5.82	5.43	5.85	4.94	5.43	5.41	5.35	5.17	5.29	5.22	5.23
seller_anchor	5.21	5.18	2.01	5.24	1.88	5.89	5.51	5.90	5.07	5.51	5.46	5.40	5.26	5.35	5.29	5.25
seller_anchor_buyer_informed	5.25	5.31	1.94	5.40	1.86	5.92	5.58	5.95	5.11	5.60	5.57	5.43	5.33	5.48	5.40	5.39

B 売り手と買い手の目標に関するプロンプト

seller_setting

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 target price for this item is \$TARGET_PRICE.

buyer_setting

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 higher price though if the item is really good or there are other perks.

C アンカリング効果に関するプロンプト

seller_anchor

Make sure to use the anchoring effect, a cognitive bias where the initial offer heavily influences the outcome of the negotiation, to gain an advantage in the negotiation. Specifically, all you need to do is make an initial offer higher than the target price.

seller_anchor_buyer_informed

Be aware that the seller will be using the anchoring effect, a cognitive bias where the initial offer greatly influences the outcome of the negotiation. In this case, the seller will start with an initial offer that is higher than their actual target price, aiming to set an “anchor” that will shape your expectations and potentially increase the final agreed price. To effectively negotiate under the influence of the anchoring effect, you can take the following strategies.

1. Stay Focused on Your Target Price: Remember your original budget or target price, and use it as a reference point instead of being swayed by the seller’s initial high offer.
2. Set a Counter-Anchor if Needed: If the seller’s initial offer is significantly higher than your budget, consider responding with a counter-offer that’s closer to your ideal price. This can help shift the anchor closer to your preferred range.
3. Ask for Justification of the High Price: Politely inquire about the specifics that justify the seller’s high initial offer. This can provide context for the price and might allow you to negotiate on specific elements, such as additional benefits or discounts.