AI vs. Human Mediation in Negotiations: Fairness, Emotions, and Agreement Rates

Aijia Li

submitted in partial fulfillment of the Bachelor of Arts degree

New York University Shanghai

Arts and Sciences

Prof. Xiangdong Qin, advisor

Prof. Ye Jin, second reader

May 2025

Contents

Abstract

This study examines the effects of human and artificial intelligence (AI) mediation on fairness, emotional responses, and negotiation outcomes in a three round repeated Ultimatum Game (UG). Participants were randomly assigned to one of three conditions: Baseline (no mediation), AI Mediation, or Human Mediation. Results indicate that the mere presence of a mediator, whether AI or human, significantly improved fairness in initial offers, suggesting a "signaling effect" where anticipated oversight encourages equitable behavior. Both mediation types also enhanced emotional outcomes, with participants reporting higher satisfaction and lower anger and regret compared to the Baseline. However, negotiation fatigue emerged as rounds progressed, with later stages associated with increased negative emotions. Risk preferences, measured via the Bomb Risk Elicitation Task, were linked to post-negotiation regret, with risk-seeking individuals experiencing less regret. Despite high overall acceptance rates, the study reveals that AI and human mediators can play distinct yet complementary roles in conflict resolution: AI promotes consistency and neutrality, while human mediators enhance perceived fairness and emotional satisfaction. These findings have implications for designing hybrid mediation systems that leverage the strengths of both human and algorithmic intervention in negotiation settings.

Key Words: *artificial intelligence, mediation, negotiation*

Preface

This study, AI vs. Human Mediation in Negotiations, was inspired by a growing curiosity about the evolving role of artificial intelligence in human-centered processes, especially as tools like DeepSeek have rapidly emerged and become widely used in academic settings. As our reliance on AI continues to grow, I began to question how we might use it wisely to enhance our lives, while also probing the boundaries of its ability to replace human roles.

This project specifically focuses on conflict resolution and negotiations that arise frequently in human society. Traditional negotiations often rely on human mediators to facilitate compromise, but perceived bias and emotional involvement can hinder fair and successful outcomes. The breakdown of negotiation, whether in business or legal contexts, frequently leads to undesirable results for all parties involved. Given AI's reputation for impartiality and rationality, I was motivated to explore whether AI could serve as a more effective mediator to reduce the emotional and cognitive demands on human participants while potentially improving negotiation outcomes and contributing to broader social welfare.

1 Introduction

The Ultimatum Game (UG) has long served as a foundational paradigm for studying fairness, bargaining, and social preferences. Classical game theory predicts that responders should accept any nonzero offer, yet empirical evidence consistently shows that unfair offers (typically below 20–30% of the total stake) are frequently rejected, reflecting the influence of fairness norms and emotional reactions (Güth, 1982; Fehr, 2002). Recent research has extended this framework to examine how algorithmic and human mediators shape negotiation dynamics. While AI mediators often propose more equitable splits (Horton, 2023), human mediators are perceived as more trustworthy in interpersonal contexts (Lee, 2018). However, the comparative effects of these mediation types on fairness perceptions and emotional responses remain underexplored.

This study investigates how AI and human mediation influence bargaining behavior in a three-round UG. We test whether: Mediation presence alone (signaling effect) improves fairness in initial offers; Actual mediation (AI or human) further enhances fairness and agreement rates; Emotional responses differ across mediation types, particularly in multi-round negotiations.

By incorporating risk preference measures and tracking negotiation progression, we provide new insights into the psychological and strategic dimensions of mediated bargaining. Our findings contribute to the growing literature on algorithmic mediation, highlighting its potential to complement, but not fully replace, human judgment in conflict resolution.

2 Literature

Foundational Theory: The Ultimatum Game

The Ultimatum Game (UG) is a classic economic experiment on fairness and bargaining. One player (proposer) offers how to split a sum of money; the other (responder) can accept or reject the offer. Game theory predicts any positive offer should be accepted, but empirical studies consistently show that unfair offers (typically under 20–30% of the pie) are often rejected (Guth, 1982; Fehr, 2002). In practice, proposers anticipate this by offering roughly 40–50% on average (Guth, 1982). Rejection of low offers is thought to reflect fairness concerns: unfair offers elicit anger or negative emotion, leading responders to punish the proposer at a personal cost (Fehr, 2002). Thus, the UG has become a standard paradigm for studying how notions of fairness and emotion influence bargaining outcomes.

AI as Mediator: Perceptions of Fairness and Emotion

Recent studies suggest that algorithmic decision-makers are viewed differently from humans. Lee (2018) found that when tasks require human skills (e.g., interpersonal judgments), algorithmic decisions are judged less fair and trustworthy, and they evoke more negative emotions. In contrast, AI agents tend to propose more equitable splits: for example, GPT-4 typically offers strict 50/50 splits as proposer in the UG and invests more generously in trust and public-goods games (Horton, 2023).

Reflecting this, participants are often more tolerant of unfair offers from AI than from humans, though findings can vary depending on context (Jang, 2022). In one experiment, respondents were more likely to reject unfair offers when told their choices would "teach" an AI to behave fairly, showing a preference to shape algorithmic behavior for future fairness (Wilson, 2023).

In mediated group settings, AI-driven facilitation has shown promise. For exam-

ple, Tessler (2024) found that AI-generated summaries in political deliberation were more widely accepted than human mediation. Berinsky (2023) showed that AI interventions reduced incivility and emotional volatility in online debates. In diplomatic negotiation simulations, Shapiro (2024) found that AI-assisted mediation uncovered shared interests at over twice the rate of human mediators and resolved more contested issues.

These findings imply that AI mediators can produce fairer-seeming proposals and de-escalate conflict more effectively, although user trust often hinges on perceived procedural fairness and transparency (Friedler, 2021).

Human Mediation: Strengths and Limitations

Human mediators bring social intelligence that AI lacks. Through empathy, listening, and creativity, they make parties feel acknowledged. This fosters emotional buy-in and procedural fairness (Lee, 2018). Experimental evidence suggests that people often judge outcomes as fairer when a human is involved in the process (Gino, 2008).

However, human mediation has limits. Bias, fatigue, and inconsistency can hinder outcomes. Human negotiators are also more prone to emotional escalation and subjective judgments. While human mediators excel in relationship-building, their decisions may be perceived as biased, especially when neutrality is unclear.

Comparative Outcomes: Agreement Success Rates

Negotiation behavior: In UG variants, offer and acceptance rates are often similar whether the opponent is human or AI. However, participants may be more forgiving when AI decisions benefit others (Jang, 2022).

Consensus-building: AI mediators can match or exceed human facilitators. Tessler (2024) reported higher consensus and satisfaction with AI summaries in group deliberation.

Conflict reduction: AI moderation has been shown to decrease personal attacks and emotional volatility in political conversations (Berinsky, 2023).

Agreement yield: AI mediation produced more detailed and substantive agreements in simulations of international negotiation compared to human facilitators (Shapiro, 2024). These results suggest that while humans offer emotional resonance, AI can bring consistency, neutrality, and efficiency.

In sum, AI mediators often propose fairer splits (Horton, 2023), reduce emotional conflict (Berinsky, 2023), and increase consensus (Tessler, 2024), while humans contribute social trust and nuanced understanding (Lee, 2018). The relative effectiveness depends on task context and user expectations.

3 Experiment Design

3.1 Demographics

To account for potential heterogeneity in strategic behavior and ensure the robustness of our experimental results, I collected comprehensive demographic and experiential data from all participants. Specifically, participants reported their gender (with male, female, and non-binary options provided), age (recorded as a continuous variable), and highest attained education level (categorized into bachelor's degree, master's degree, or doctoral degree, and others). Given the specialized nature of our study, I additionally inquired about two critical experiential factors: whether participants had previously taken any formal coursework in game theory (binary yes/no response) and whether they had ever participated in any Ultimatum Game experiments prior to this study (binary yes/no response). These measures were implemented based on established findings that prior theoretical training and experimental experience may

systematically influence bargaining strategies and fairness perceptions in economic games (Camerer, 2003; Cooper Dutcher, 2011).

3.2 The Bomb Game

To elicit participants' risk preferences in an incentive-compatible way, I implemented the static Bomb Risk Elicitation Task (BRET) developed by Crosetto and Filippin (2013) prior to the main negotiation experiment. Unlike self-reported measures used in prior studies (e.g., Crosetto and Mantovani, 2018), BRET involves real monetary stakes and provides a behavior-based assessment of risk tolerance. Participants were presented with 100 boxes, one of which contained a bomb. They chose a number from 0 to 100, indicating how many boxes they wished to open. The location of the bomb was randomly determined. If the bomb was in the range they selected, they earned nothing; otherwise, they received a monetary reward based on the number of safe boxes opened. The task offers a simple and intuitive format that captures risk preferences along a continuous scale: higher values of bomb number chosen reflect greater risk-taking behavior, while lower values indicate risk aversion.

The BRET was administered before the main ultimatum game starts. Importantly, participants did not receive immediate feedback about the outcome of their choice. Instead, feedback was delayed until the end of the study to avoid potential spillover effects from the risk task to the three-round Ultimatum Game. This design decision also helped encourage participants to return for the follow-up session, ensuring a higher completion rate.

Incorporating the BRET before the negotiation task allows us to examine the relationship between individual risk preferences and bargaining strategies in the Ultimatum Game. In a multi-round negotiation setting, risk tolerance may shape both

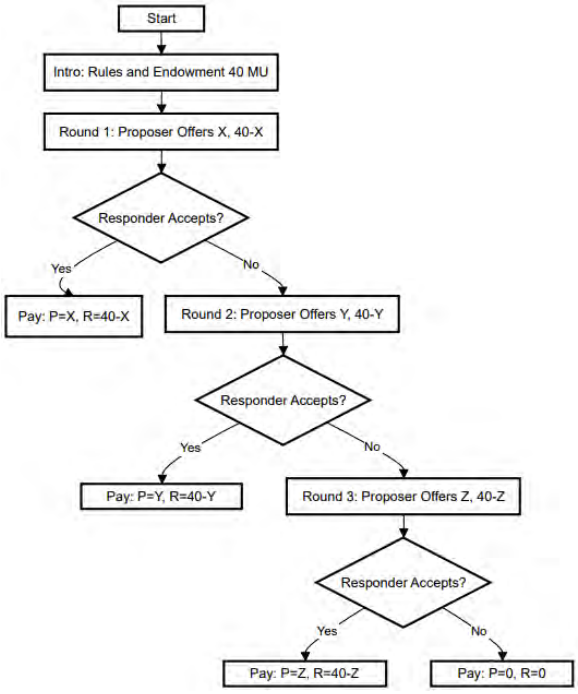

Figure 1: Baseline flowchart

Proposer and Responder behavior. For example, risk-averse Proposers might make more generous offers early on to avoid rejection, whereas risk-seeking Responders may be more willing to hold out for better deals in later rounds. By measuring risk attitudes independently and before the main task, we can better isolate how these preferences influence strategic decisions during negotiation.

3.3 Main Ultimatum Game

The Ultimatum Game is a widely studied economic experiment that explores human behavior in bargaining and fairness. In its simplest form, the game involves two players: a proposer and a responder. The proposer is given a fixed sum of money and must offer a portion to the responder. The respondent can either accept the offer, allowing both players to receive the proposed shares, or reject it, in which case neither party receives anything. Although classical economic theory predicts that

responders should accept any nonzero offer (as some money is better than none), empirical results consistently reveal that people often reject offers perceived as unfair. This behavior highlights the importance of social preferences, such as fairness, reciprocity, and punishment, challenging traditional assumptions of rational self-interest in economic decision making.

This study investigates how different types of mediation, none (Baseline), artificial intelligence (AI), and human, affect negotiation outcomes in a repeated Ultimatum Game. Participants are randomly assigned to one of three between-subjects conditions: Baseline, AI Mediator, or Human Mediator. In all conditions, participants are paired and assigned fixed roles as either the Proposer or the Responder. Each pair begins with a total endowment of 40 monetary units (MU), which the Proposer must propose how to divide. The negotiation process may last up to three rounds,as shown in Figure 1. However the ultimatum game ends as soon as the Proposer and the Responder reaches an agreement.

To guarantee participants understands the game, they need to correctly complete two mock questions before officially begin the ultimatum game.

In the **Baseline** condition, participants proceed through the game without any mediation. The Proposer independently revises their offer in each round, and the Responder decides whether to accept or reject, with no intervention or external input.

In the **AI Mediator** condition, as shown in Figure 2, mediation is introduced in the form of AI-generated suggestions provided to the Proposer after any rejection. If the Responder rejects the offer in Round 1, the AI system provides a suggestion before Round 2 begins. The Proposer and the Responder can choose whether or not to follow the AI's advice when forming their next offer, as AI suggestion is not binding. The same process occurs if the Round 2 offer is rejected, with the AI providing a new suggestion before Round 3. All payoff structures remain the same as in the Baseline

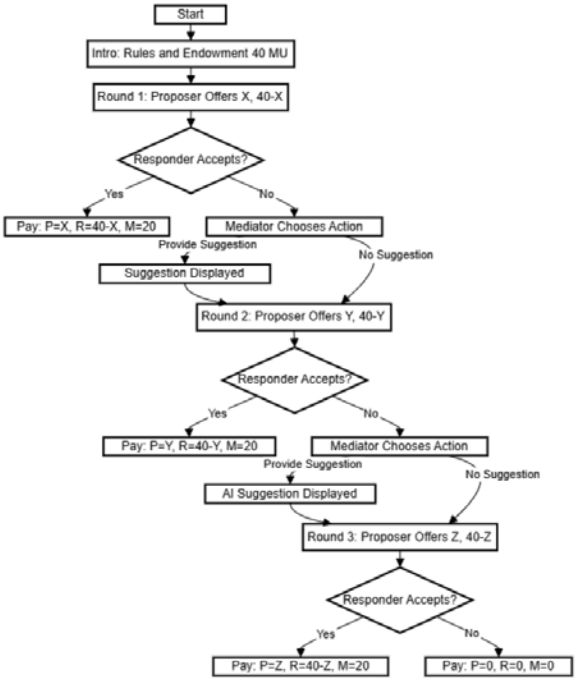

Figure 2: AI Mediation Treatment Flowchart

condition.

In the **Human Mediator** condition, as shown in Figure 3, a third-party human mediator becomes involved following any rejected offer. After a rejection, the mediator decides whether to provide a AI generated suggestion to the Proposer or remain silent. If the mediator offers advice, the suggestion is shown to the Proposer before they formulate their next offer. Similarly, The Proposer and the Responder can choose whether or not to follow the advice when forming their next offer, as the suggestion is not binding.

To ensure consistency across conditions, all interactions occur via computer terminals, and suggestions (from both AI and human mediators) are standardized as brief textual prompts intended to increase the likelihood of agreement (e.g., "Consider offering a more even split 20-22 MU) to avoid being rejected again").

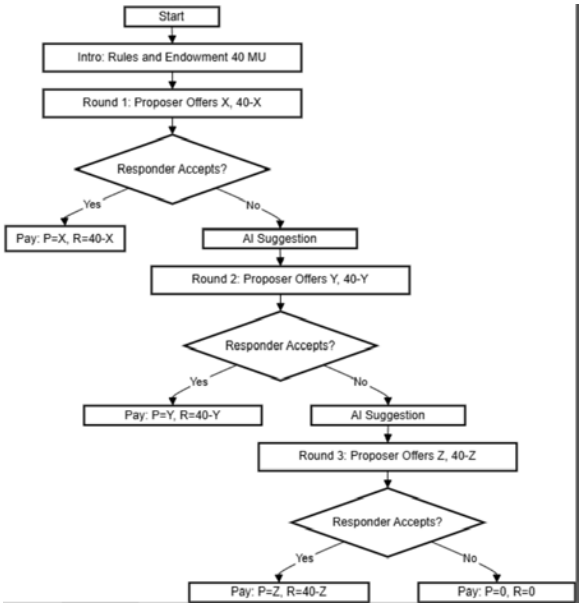

Figure 3: Human Mediation Treatment Flowchart

3.4 Post-game Survey

After the negotiation task, participants complete a short survey evaluating their perceptions of fairness, the usefulness of mediation, and their emotional responses during the game. This allows us to capture both behavioral outcomes and subjective impressions across conditions. The survey asks participants to rate on a scale from 1 (not at all) to 7 (very much) at the end of the experiment.

To complement the behavioral data collected during the experiment, we administered a comprehensive post-game survey designed to capture participants' subjective evaluations of fairness, emotional responses, and perceptions of the mediation process (where applicable). All participants responded to a core set of 7-point Likert scale items assessing three key dimensions: (1) fairness perceptions, including evaluations of the overall outcome, the proposer's offer, and the responder's decision; (2) emotional states, measuring satisfaction with offers, anger/frustration, guilt/regret; and (3) mediation-specific evaluations for the respective treatment groups.

In the AI mediator group, participants additionally rated the perceived fairness of the AI's suggestions and the degree to which the AI influenced their decisions, with a 0-point option ("No AI involved") allowing for neutral responses. Similarly, in the human mediator group, participants evaluated the fairness of the human mediator and their decision influence, also with a 0-point option for non-applicable responses. These mediation-specific items were included to assess whether participants perceived AI and human mediators differently in terms of fairness and behavioral influence, which could help explain potential treatment effects.

4 Results

The experiment is built via oTree, participants are recruited from the online survey platform Prolific. Screening criteria includes located in USA, English as primary language and has a minimum undergraduate degree. Piloting of the experiment is conducted in a social behavior lab of New York University Shanghai Campus.

In total I recruited 263 participants from Prolific according to the above-listed criteria. The average time spent on the experiment is 15 minutes and the average payment of 2 dollar per person.

4.1 Summary Statistics

Descriptive statistics in Table 1 presents the quantitative results of an Ultimatum Game experiment under four conditions: total sample (n = 263), baseline (n = 50), AI mediation (n = 90) and human mediation (n = 123), corresponding to the baseline group (N = 25), AI mediation group (N = 45), human mediation group (N = 41). Table 1 gives the basic characteristics of participants, while game theory is a dummy variable measuring if the participant has taken a game theory related course

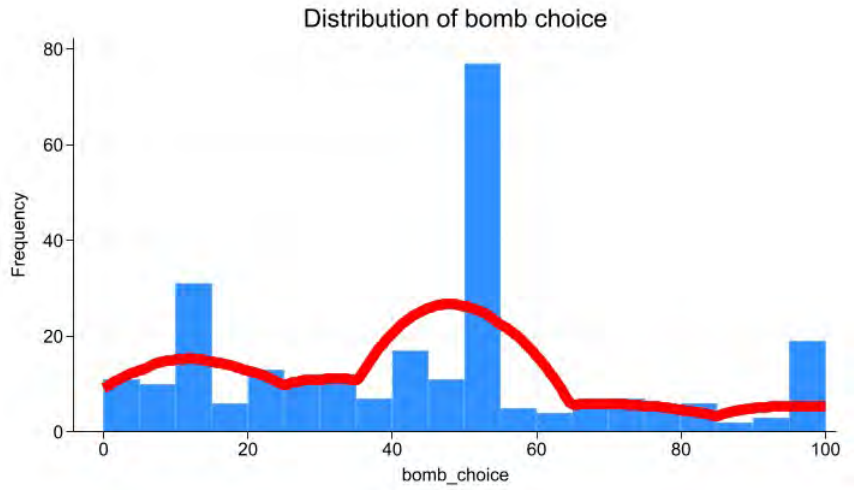

Figure 4: Bomb Choice Distribution

(microeconomics doesn't count), and ultimatum game measures if the participant has previously been involved in an ultimatum game experiment. Figure 4 gives the distribution of bomb choice among all participants. Table 2 gives an overview of the results from the ultimatum experiment. The variables are grouped into five categories: acceptance rate, average offer measured by MU (monetary unit in game), Mediator Influence, Emotional Response, and Ultimatum Game Earnings by role. The Mediator Influence is measured by the percentage of groups that failed the first round and thus get moved into the second and/or third round which received mediation by the mediator.

The acceptance rates in Round 3 show irregular values primarily due to small sample sizes. For example, in the Human Mediation treatment, only two groups reached Round 3, and neither reached an agreement, resulting in a 0% acceptance rate. Therefore, it is more meaningful to focus on the overall acceptance rates. Across all treatment groups, as the one-shot game extends to three rounds, the cumulative acceptance rates are generally high.

	Total n=263	Baseline n=50	AI Mediation n=90	Human Mediation n=123
Gender				
Female	130	24	44	62
	(49.43%)	(48%)	(48.89%)	(50.41%)
Male	132	26	46	60
	(50.19%)	(52%)	(51.11%)	(48.78%)
Non-binary	1	0	0	1
	(0.38%)	(0%)	(0%)	(0.81%)
Age	39.02	42.24	38.58	38.04
	(12.78)	(14.70)	(13.51)	(11.20)
Level of Education				
Bachelor's	137	20	44	73
	(52.09%)	(40%)	(48.89%)	(59.35%)
Master's	91	20	34	37
	(34.60%)	(40%)	(37.78%)	(30.08%)
PhD	27	9	10	8
	(10.27%)	(18%)	(11.11%)	(6.50%)
Others	8	1	2	5
	(3.04%)	(2%)	(2.22%)	(4.07%)
Game theory				
Yes	56	18	18	20
	(21.29%)	(36%)	(20%)	(16.26%)
No	207	32	72	103
	(78.71%)	(64%)	(80%)	(83.74%)
Ultimatum game				
Yes	48	14	13	21
	(18.25%)	(28%)	(14.44%)	(17.07%)
No	215	36	77	102
	(81.75%)	(72%)	(85.56%)	(82.93%)

Table 1: Characteristics of participants

	Total n=111	Baseline n=25	AI Mediation n=45	Human Mediation n=41
Acceptance Rate				
Round 1		88%	64%	61%
Round 2		33%	62%	87.5%
Round 3		100%	75%	0%
Overall		100%	97.78%	95.12%
Average Offer (MU)				
Round 1		23.8	16.98	18.61
Round 2		20	21	17.13
Round 3		28	18.25	11.5
Mediator Influence				
		N/A	42%	17%
Emotional Response				
Satisfaction		5.98	5.6	6.49
Anger		1.46	1.94	1.63
Regret		1.6	1.71	1.46
Ultimatum Game earning				
Proposer		15.56	24.16	19.29
Responder		24.44	15.84	20.71
Mediator		N/A	N/A	19.02

Table 2: Ultimatum Game Outcomes Overview

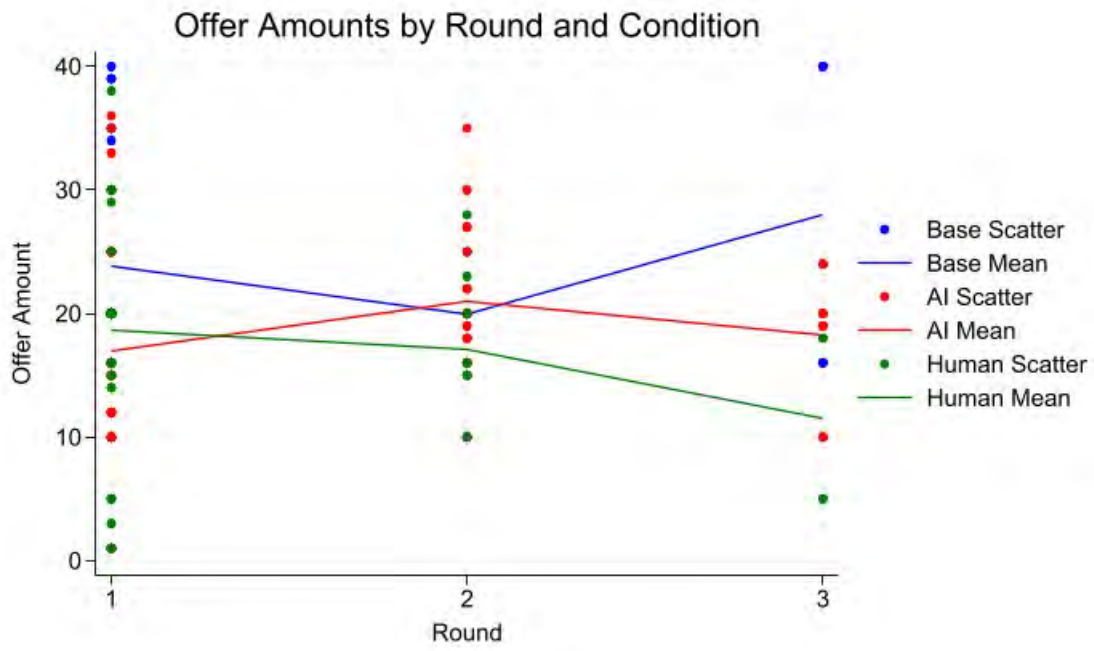

Figure 5: Offer Trend by Round

4.2 Fairness

Table 3 presents regression results analyzing the effects of Human and AI mediation treatments on fairness, measured by the absolute deviation from an equal split (20 MU). Column (1) classifies all groups assigned to either the AI or Human mediation condition as part of the treatment group, regardless of whether they progressed to Round 2 or 3 and received actual mediation. The rationale for this specification is that the mere presence of a mediator, whether AI or human, was signaled to participants at the outset, which may have influenced their behavior in proposing or responding to offers. The detailed offer distribution is captured by Figure 5. To isolate the effect of actual mediation, Column (2) restricts the treatment group to only those cases in which participants advanced to Round 2 (and Round 3) and thus received mediated suggestions.

Because fairness is measured as the deviation from an equal split, a negative

Variables	(1) $fair_s$	(2) $fair_t$
Human Mediation	-4.889***	-2.564
	(1.509)	(3.111)
AI Mediation	-3.301**	-3.110
	(1.467)	(2.407)
round	0.316	1.853
	(0.967)	(1.624)
bomb_choice	0.00866	0.0233
	(0.0215)	(0.0224)
gender	1.941*	1.669
	(1.103)	(1.147)
age	-0.0110	0.0131
	(0.0421)	(0.0431)
Constant	6.518**	0.566
	(2.677)	(2.710)
Observations	110	110
R-squared	0.123	0.048

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Fairness

coefficient indicates an improvement in fairness. The results show that when not isolating the effect of actual mediation, human mediation has a statistically significant effect on improving fairness, meaning offers deviate less from the equal split when a human mediator is present. AI mediation also significantly improves fairness, though the effect is smaller than that of human mediation. Among control variables, gender is associated with fairer offers, while age and other factors do not significantly influence fairness.

4.3 Emotions

	(1)	(2)	(3)
Variables	satisfaction	anger	regret
Human mediation	3.045**	-2.305**	-2.573**
	(1.166)	(0.930)	(1.179)
AI mediation	2.261**	-2.509***	-2.155**
	(0.917)	(0.732)	(0.927)
result	2.152	-1.610	0.0730
	(1.820)	(1.452)	(1.841)
round	-1.748***	2.222***	1.354**
	(0.646)	(0.516)	(0.654)
bomb_choice	-0.0137	0.00915	0.0180**
	(0.00848)	(0.00676)	(0.00857)
gender	0.521	0.175	-0.0368
	(0.430)	(0.343)	(0.434)
age	-0.0265	0.00666	0.00822
	(0.0163)	(0.0130)	(0.0165)
Constant	6.009***	0.803	0.264
	(2.246)	(1.792)	(2.272)
Observations	110	110	110
R-squared	0.153	0.222	0.094

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Emotional Responses

Table 4 analyzes emotional responses, capturing the effects of treatment on overall satisfaction of participants, feelings of anger, and regret. Only participants who

received actual mediation, that is, who received advice from either an AI or human mediator, are classified under the AI or human mediation groups. Since the outcome of the Ultimatum Game can influence emotional responses, a control variable is included: a binary indicator equal to 1 if the proposer and responder reached an agreement within the three-round game.

The results show that human mediation significantly increases the reported satisfaction of participants while reducing feelings of anger and regret. AI mediation yields a similar, though less pronounced, pattern. Emotional responses also vary by negotiation round, with later rounds associated with lower satisfaction and increased negative emotions, suggesting that negotiation dynamics evolve over time.

5 Discussion

5.1 Acceptance Rates and Offer Patterns

Although acceptance rates appear to differ across rounds, overall acceptance was high across all groups, limiting the analytical leverage available from this metric. Notably, very few groups proceeded beyond the first round, particularly in the baseline group. This may reflect several factors: first, early offers might have been sufficiently high to prevent negotiation breakdown; second, time or cognitive costs of continuing rounds could have discouraged escalation. Third, the strategic anticipation of rejection may have led participants to offer closer-to-equal splits initially, curbing further negotiation.

Existing literature offers mixed evidence regarding whether acceptance rates increase with more rounds. While some meta-analyses suggest learning and reciprocity effects, where responders become more likely to accept fair offers over time (Nowak

et al., 2000), other studies report no significant round-to-round change in acceptance (Engel, 2011). Thus, low transition rates to R2/R3 in our sample may be context-specific rather than anomalous.

Our R1 offer averages (16.98 for AI, 18.61 for Human, 23.8 for Baseline) mostly fall within the expected range established by prior Ultimatum Game studies, which consistently report average offers between 40–50% of the total pie (Camerer, 2003). Offers below 30% are typically associated with higher rejection rates (Güth et al., 1982), suggesting that especially the AI-mediated groups might have triggered more early rejections. This reinforces the robustness of our baseline results despite a small sample size in later rounds.

5.2 Fairness and the Signaling Effect

The presence of a mediator was disclosed to participants at the outset, even if they never reached the round where mediation took place. This design enabled us to measure a signaling effect, which is the behavioral impact of merely expecting third-party oversight. In line with previous findings on the "audience effect" (Andreoni Bernheim, 2009), both AI and human mediation groups showed significant improvements in fairness when measured as deviation from equal split, even when actual mediation was not delivered. This suggests participants sought to appear fairer in anticipation of evaluation, a phenomenon well-documented in social preference and experimental economics research.

However, when isolating only those who received actual mediation, the fairness effect was no longer significant, which is likely due to sample size limitations. Nevertheless, the signaling effect itself is an important finding, warranting further exploration in future studies that more deliberately manipulate perceived presence versus

actual engagement of mediators. This echoes prior findings from observer framing experiments (Fischbacher Föllmi-Heusi, 2013), where third-party labels ("fair/unfair") significantly shaped proposer behavior.

5.3 Emotional Responses to Mediation

Both AI and human mediation were associated with significant improvements in emotional outcomes: participants reported higher satisfaction and lower anger and regret. These results are consistent with studies in conflict resolution and negotiation that show both algorithmic and human mediation can reduce negative affect (De Melo et al., 2014; Ebner Zeleznikow, 2016). While human mediators traditionally provide emotional buffering through empathy and conversational reframing, AI agents may reduce stress by being perceived as impartial or nonjudgmental (Klemp et al., 2023).

Interestingly, the round number itself had a strong effect on emotional outcomes: later rounds correlated with lower satisfaction and higher anger/regret. This likely reflects negotiation fatigue and emotional wear, as suggested by negotiation psychology research (Thompson, 2005). With each successive round, disappointment may accumulate, especially if expectations for compromise are unmet. Extended bargaining often intensifies pressure, particularly under known final-round conditions, resulting in emotional volatility.

Another noteworthy factor is risk attitude, measured via participants' bomb-choice behavior in an unrelated task. We find a small but significant correlation between risk-tolerance and regret, supporting prior work on regret theory (Loomes Sugden, 1982), which holds that risk-seeking individuals may feel less regret after adverse outcomes because they anticipated variability. Conversely, more risk-averse players may experience greater dissonance when their careful strategies fail.

6 Conclusion

This study provides new insights into the psychological and behavioral consequences of introducing human and AI mediation into bargaining contexts. Despite limited progression to later rounds, several robust findings emerge:

High acceptance rates across all groups constrain the ability to draw strong conclusions about treatment effects on agreement success. However, the low rate of escalation suggests that participants frequently reached impasses early or found early offers acceptable.

The signaling effect of mediation presence significantly enhanced fairness in first-round offers, even when no actual mediation occurred. This aligns with broader evidence on observability and norm compliance and suggests that simply mentioning a mediator can shape bargaining behavior.

Emotional outcomes improved significantly under both AI and human mediation, with participants reporting greater satisfaction and reduced anger/regret. This suggests that third-party intervention, whether algorithmic or human, can buffer the psychological strain of negotiation, even when outcomes are not objectively fairer.

The round structure of negotiation strongly influences emotion: longer negotiations tend to amplify negative affect, pointing to the emotional cost of delay and the potential for "negotiation fatigue."

Risk preferences subtly influence emotional responses, particularly regret, underscoring the importance of individual differences in interpreting outcomes.

These findings carry practical implications for dispute resolution design. While human mediators may not always deliver fairer outcomes, their presence enhances subjective satisfaction. Conversely, AI mediators offer consistency and influence but may face limits in replicating human warmth. Future systems should consider blend-

ing both approaches, capitalizing on AI's neutrality and human mediators' emotional intelligence.

References

[1] Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, 3(4), 367–388.

[2] Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868), 137–140.

[3] Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 2053951718756684.

[4] Horton, J. J. (2023). Large language models as skilled economic agents. *arXiv preprint arXiv:2305.19860*.

[5] Krueger, F., & Hoffman, M. (2021). Trusting machines: The ethics of algorithmic decision-making. *Nature Human Behaviour*, 5, 1365–1366.

[6] Shapiro, J. N., et al. (2024). Peace in our time? AI-supported negotiation in track-two diplomacy. *Science Advances*, 10(14), eadk0522.

[7] Tessler, H., et al. (2024). AI and political deliberation: Experimental evidence from large language model facilitation. *Nature Human Behaviour*, 8, 273–282.

[8] Rosenfeld, A., et al. (2020). Algorithms and human behavior: The impact of algorithmic decision-making on moral choice. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(03), 2846–2853.

[9] Berinsky, A., et al. (2023). Using AI to reduce political violence: Field evidence from social media. *Proceedings of the National Academy of Sciences*, 120(1), e2207380120.

[10] Friedler, S. A., et al. (2021). A comparative study of fairness-enhancing interventions in machine learning. *Communications of the ACM*, 64(4), 139–144.

[11] Jang, Y., & Lee, M. K. (2022). Algorithmic mediation in online conversations: Effects on fairness perceptions and conflict. *CHI Conference on Human Factors in Computing Systems (CHI '22)*.

[12] Wilson, D., et al. (2023). Teaching machines to be fair: Experimental evidence from fairness-guided human interaction. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 30(3), Article 18.

[13] Gino, F., & Pierce, L. (2008). Dishonesty in the name of equity. *Psychological Science*, 20(9), 1153–1160.

[14] Andreoni, J., & Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, *77*(5), 1607–1636.

[15] Camerer, C. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

[16] De Melo, C., Gratch, J., & Carnevale, P. (2014). Humans vs. computers: Impact of emotion expressions on people's decision making. *IEEE Transactions on Affective Computing*, *6*(4), 349–362.

[17] Ebner, N., & Zeleznikow, J. (2016). No sherpas needed: Online dispute resolution in the digital age. *Journal of Dispute Resolution*, *2016*(1), 1–40.

[18] Engel, C. (2011). Dictator games: A meta-study. *Experimental Economics*, *14*(4), 583–610.

[19] Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise: An experimental study on cheating. *Journal of the European Economic Association*, *11*(3), 525–547.

[20] Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*(4), 367–388.

[21] Klemp, A., et al. (2023). Trust and satisfaction in AI-mediated negotiations. *ACM Transactions on Human-Robot Interaction*, *12*(1), Article 2.

[22] Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal*, *92*(368), 805–824.

[23] Nowak, M. A., Page, K. M., & Sigmund, K. (2000). Fairness versus reason in the ultimatum game. *Science*, *289*(5485), 1773–1775.

[24] Thompson, L. (2005). *The Mind and Heart of the Negotiator* (3rd ed.). Pearson.

Appendix A Ultimatum Game Experiment Instructions

A.1 Introduction

Welcome to the experiment. This experiment consists of four parts: a demographic survey, a bomb game, an ultimatum game (main part), and a post-experiment survey.

The currency unit used in the experiment is MU (Monetary Unit), with an exchange rate of 20 MU to 1 dollar.

If complete the whole experiment, your overall payoff = bomb game payoff + ultimatum game payoff + participation fee.

A.2 Bomb Game Instructions

On this page, you need to make a decision regarding a bomb game.

Consider the following game: You can choose to open a number of boxes from 1 to 100. Among 100 boxes, 99 of the boxes contain 0.2 MU, and only 1 box contains a BOMB. Each box has equal probability to contain the bomb and the location of the bomb is randomly generated. If the location of the bomb is smaller or equal to the number of boxes you choose to open, you open the box with the bomb and thus you will earn zero. If not, you can earn 0.2 MU for each box you open.

The result of this game will be revealed to you at the end of the experiment.

A.3 Ultimatum Game Instructions

In this section, you need to participate in an Ultimatum Game. Please go through the instructions carefully.

An Ultimatum Game is a game that consists of two players: the Proposer and the Responder. And the Proposer makes a take-it-or-leave-it offer to the Responder.

If the Responder accepts the offer, the two players allocate the money according to the offer; If the Responder rejects the offer, both players get zero payment.

The game can be played for up to three rounds and ends once the Responder accepts the offer. However, if the Proposer and the Responder still fail to reach an agreement in the third round, both the Proposer and the Responder receive zero payment.

You need to first answer three questions correctly before starting the official game.

A.4 Mock Question

For player_id = 1:

You are the **Proposer**. Please consider the following scenarios:

For player_id = 2:

You are the **Responder**. Please consider the following scenarios:

For player_id = 3:

You are the **Mediator**. As long as the Responder accepts the Proposer's offer, you get paid 20 MU. If the Responder rejects the Proposer's offer in Round 1, you'll have the opportunity to provide suggestions to both players before Round 2 begins. The same process applies between Round 2 and 3. Please consider the following scenarios:

Scenario 1: There are in total 40MU endowment, the Proposer gives 39MU to the Responder, and the Responder ACCEPTS the offer. How much will you receive?

Scenario 2: There are in total 40MU endowment, the Proposer gives 39MU to the Responder, and the Responder REJECTS the offer. How much will you receive?

Question: How many rounds can this game last at most?

A.5 Ultimatum Game for the Proposer

Now begins Round 1.

You are the **Proposer**. There are altogether 40 MU endowment.

How much would you like to give to the responder?

A.6 Ultimatum Game for the Responder

Now begins Round 1.

You are the **Responder**. There are altogether 40 MU endowment. Now, the Proposer offered you {offer_1} MU.

Do you accept or decline the proposer's offer?

What's the lowest amount you accept in this game?

A.7 The Result of the Ultimatum Game: accept

There are altogether 40 MU endowment in this game.

For player_id = 1:

You sent {offer_1} to the Responder.

And the Responder accepts the offer.

For player_id = 2:

The Proposer sent {offer_1} MU to you.

And you accepts the offer.

So, you received {payoff} MU from the Ultimatum Game.

A.8 The Result of the Ultimatum Game: reject

There are altogether 40 MU endowment in this game.

For player_id = 1:

You sent {offer_1} to the Responder.

And the Responder accepts the offer.

For player_id = 2:

The Proposer sent {offer_1} MU to you.

Since you didn't reach an agreement, now begins the second round of the Ultimatum Game.

A.9 Post-experiment Survey I

How fair do you think the overall outcome of the ultimatum game was? (1 = Very Unfair, 7 = Very Fair)

How fair do you think the offer made by the proposer was? (1 = Very Unfair, 7 = Very Fair)

How fair do you think the response from the responder was? (1 = Very Unfair, 7 = Very Fair)

A.10 Post-experiment Survey II

How did you feel about the offer you received/made? (1 = Very Dissatisfied, 7 = Very Satisfied)

Did you feel anger or frustration during the game? (1 = Not at all, 7 = Very much so)

Did you feel guilt or regret about your decision? (1 = Not at all, 7 = Very much so)

Appendix B AI Mediation Instructions

B.1 Introduction for an Ultimatum Game Round 2

Based on your round 1 results, here is the suggestion provided by **real-time AI**. Note that this is for reference only and is not binding.

{ai suggestion}

B.2 Post-experiment Survey I

How fair do you think the overall outcome of the ultimatum game was? (1 = Very Unfair, 7 = Very Fair)

How fair do you think the offer made by the proposer was? (1 = Very Unfair, 7 = Very Fair)

How fair do you think the response from the responder was? (1 = Very Unfair, 7 = Very Fair)

How fair do you think the real-time AI was? (1 = Very Unfair, 7 = Very Fair, 0 = No AI involved)

How much did the real-time AI influence your decision? (1 = Not at all, 7 = A great deal, 0 = No AI involved)

B.3 Post-experiment Survey II

How did you feel about the offer you received/made? (1 = Very Dissatisfied, 7 = Very Satisfied)

Did you feel anger or frustration during the game? (1 = Not at all, 7 = Very much so)

Did you feel guilt or regret about your decision? (1 = Not at all, 7 = Very much so)

Appendix C Human Mediation Instructions

C.1 Ultimatum Game for the Mediator

Given that in round 1, the Proposer offered {offer_1} MU and the Responder rejected.

{ai suggestion}

Do you want to provide this suggestion to both proposer and responder?

C.2 Introduction for the Ultimatum Game Round 2: with human suggestion

Since you didn't reach an agreement in round 1, **a Human Mediator** sees your round 1 decisions and gives a piece of advice visible to both parties. The Human Mediator can get paid if the Proposer and the Responder agree on the offer.

Here is the advice. Note that the advice is not binding.

{ai suggestion}

Click "next" to begin round 2.

C.3 Introduction for the Ultimatum Game Round 2: without human suggestion

Since you didn't reach an agreement in round 1, **a Human Mediator** sees your round 1 decisions and had a chance to give a piece of advice to both parties. However, the Human Mediator decided not to give advice.

Click "next" to begin round 2.

C.4 Post-experiment Survey I

How fair do you think the overall outcome of the ultimatum game was? (1 = Very Unfair, 7 = Very Fair)

How fair do you think the offer made by the proposer was? (1 = Very Unfair, 7 = Very Fair)

How fair do you think the response from the responder was? (1 = Very Unfair, 7 = Very Fair)

How fair do you think the Human Mediator was? (1 = Very Unfair, 7 = Very Fair, 0 = No Mediator involved)

How much did the Human Mediator influence your decision? (1 = Not at all, 7 = A great deal, 0 = No Mediator involved)

C.5 Post-experiment Survey II

How did you feel about the offer you received/made? (1 = Very Dissatisfied, 7 = Very Satisfied)

Did you feel anger or frustration during the game? (1 = Not at all, 7 = Very much so)

Did you feel guilt or regret about your decision? (1 = Not at all, 7 = Very much so)

Acknowledgement

I'd love to express my sincere thank to all the professors and friends who had helped me along the way conducting the experiment. I've indeed learnt a lot from doing my capstone project and will carry all these on to my future works.