

# A Fairness-Driven Method for Learning Human-Compatible Negotiation Strategies

Anonymous ACL submission

## Abstract

Despite recent advancements in AI and NLP, negotiation remains a difficult domain for AI agents. Traditional game-theoretic approaches that have worked well in two-player zero-sum games struggle in the context of negotiation due to their inability to learn human-compatible strategies. On the other hand, approaches that only use human data tend to be domain-specific and lack the theoretical guarantees provided by strategies grounded in game-theory. Motivated by the notion of fairness as a criteria for optimality in general sum games, we propose a negotiation framework called FDHC which incorporates fairness into both the reward design and search to learn human-compatible negotiation strategies. Our method includes a novel, RL+search technique called LGM-Zero which leverages a pre-trained language model to retrieve human-compatible offers from large action spaces. Our results show that our method is able to achieve more egalitarian negotiation outcomes and improve negotiation quality.

## 1 Introduction

Recent advancements in AI and NLP have led researchers to develop techniques to build autonomous agents which can achieve human-level performance in bargaining games such as Deal-or-no-Deal (Sengupta et al., 2021). These techniques can be separated into two broad categories: game-theoretic approaches and data driven approaches.

Game-theoretic approaches to negotiation attempt to build negotiation agents without observing any human data. This class of algorithms is typically applied to two-player zero-sum games (2p0s) which do not require agents to learn human-compatible strategies in order to be successful (Silver et al., 2018; Brown et al., 2020; Vinyals et al., 2019). However, other classes of games such as negotiation require cooperation with humans in order to be successful (Bakhtin et al., 2022). As a result, traditional game theoretic methods have

failed to achieve quality performance in the realm of negotiation (Li et al., 2023).

Data driven approaches to negotiation learn a human-like strategy directly using data on a specific negotiation domain (Verma et al., 2022; Lewis et al., 2017; He et al., 2018). Unfortunately, collecting human data is expensive and the strategies in the data may not effectively generalize to other negotiation domains. Furthermore, these methods lack the desirable properties that game-theoretic methods offer such as controllability.

We propose an approach which bridges game-theoretic and data driven approaches called **Fairness-Driven Human-Compatible** (FDHC) bargaining. This method is designed to target egalitarian outcomes, specifically the Egalitarian Bargaining Solution (EBS), which we use as a formal notion of fairness (Kalai, 1977). We target fair outcomes as prior work has shown that fairness is a key component of human strategies and has also served as a useful notion of optimality in general-sum repeated games (Tossou et al., 2020; DiGiovanni and Zell, 2021; Kroll et al., 2014). Our fairness-targeting strategy is learned with a novel **LLM-Guided Monte Carlo** tree search with **Zero** domain specific training data (LGM-Zero). LGM-Zero leverages the reasoning capabilities of LLM models (Kwon et al., 2023) to extract human-like negotiation offers from large action spaces without the need to collect additional human data. A value model trained via self-play then selects the best offer that the LLM proposes. The LLM and value model are used to guide a Monte Carlo tree search (MCTS) towards the desired outcome given by FDHC (Figure 1).

We implement our method on a common negotiation exercise used in business classes where two students bargain over a used car. We train a model to play as the seller in this scenario. Both the buyer and seller are given private reservation prices for the car that they cannot go beyond during the nego-

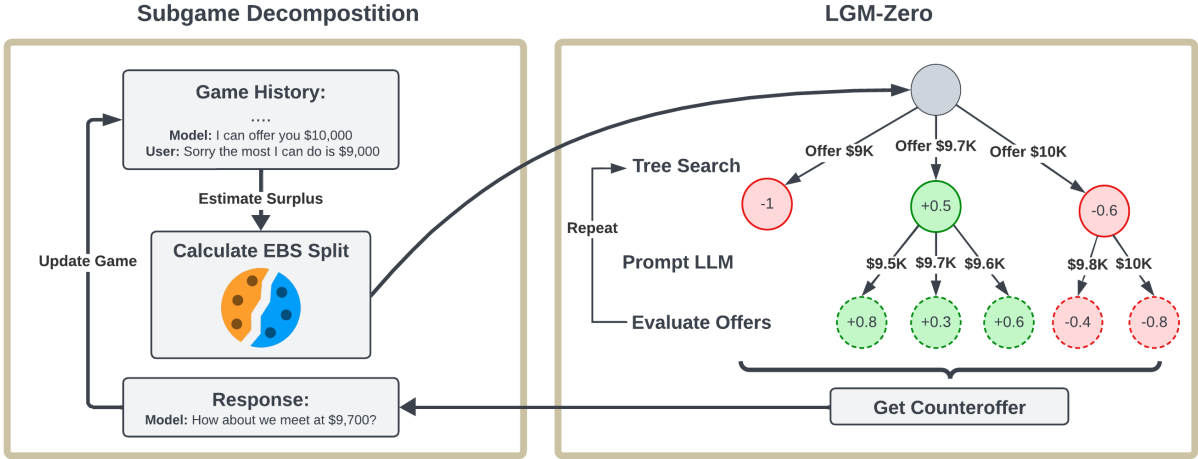


Figure 1: Outline of our FDHC negotiation framework. Our method consists of decomposing the extensive form Nash bargaining game into a series of depth-limited subgames. At each subgame we calculate the EBS and play a human-like strategy which targets this outcome using a MCTS guided by a LLM and value network.

084 titation. For example, suppose the buyer can't buy  
 085 the car for above \$11K and the seller cannot sell it  
 086 for below \$10K. The goal of our method is to reach  
 087 an agreement at the EBS solution for the game.  
 088 Which in this case corresponds to the midpoint be-  
 089 tween the two reservation prices (\$10,500 in the  
 090 example). While we implement our method for this  
 091 specific exercise, our approach generalizes to any  
 092 negotiation setting that can be modeled as a Nash  
 093 bargaining game. This encompasses any game in-  
 094 volving surplus division, including multi-party and  
 095 multi-issue negotiations.

096 Our final model uses a modular design where ne-  
 097 gotiation acts are selected according to the FDHC  
 098 framework. Strategies are then realized in natural  
 099 language using a LLM such as GPT-3.5 or GPT-4  
 100 (OpenAI, 2023). Our contributions can be summa-  
 101 rized as follows:

- 102 • We propose novel negotiation framework for  
 103 bargaining called FDHC. Our framework tar-  
 104 gets the EBS of Nash bargaining games in an  
 105 attempt to grant equal gain to both parties.
- 106 • We introduce a RL+search method called  
 107 LGM-Zero which utilizes a LLM and value  
 108 network to extract human-compatible offers  
 109 from large action spaces.
- 110 • Our results show that our method is able to  
 111 generate more egalitarian outcomes compared  
 112 to several baselines. Our human evaluation  
 113 also shows that our model is able to improve  
 114 negotiation quality while while remaining  
 115 comparable to GPT-4 in human-likeness.

## 2 Background

116 The **Nash bargaining game** is a game in which  
 117 two or more players must divide a surplus between  
 118 themselves. In the used car example given previ-  
 119 ously, the surplus would be the difference between  
 120 the buyer and seller's reservation prices. We use  
 121 the term **extensive form Nash bargaining game**  
 122 to refer to a game in which players can propose  
 123 divisions of the surplus over the course of a set of  
 124 time steps. A **Nash equilibrium** is a game state  
 125 in which no player can benefit from a unilateral  
 126 change in strategy.  
 127

128 Bargaining theory makes use of **axioms** which  
 129 are rules that describe properties that a bargaining  
 130 outcome satisfies (Nash, 1950). We make use of the  
 131 following axioms when analysing the theoretical  
 132 properties of our method. The axiom of **symmetry**  
 133 says that if the players in the bargaining game are  
 134 indistinguishable based on the description of the  
 135 game, then they should all receive the same pay-  
 136 off. A **weak Pareto optimal** solution is one where  
 137 any change to the outcome will make at least one  
 138 party no better off. **Strong monotonicity** states  
 139 that any increase in the amount of surplus being  
 140 bargained over should benefit all players involved  
 141 in the negotiation. Formal definitions can be found  
 142 in Appendix A.

## 3 Related Work

143 Prior work in the field of negotiation has typically  
 144 been centered on leveraging human data to learn ne-  
 145 gotiation strategies. These methods involve collect-  
 146 ing human-human dialogues for negotiation exer-  
 147

cises such as Craigslist bargaining (He et al., 2018) or Deal-or-no-Deal (Lewis et al., 2017). This data can then be used to perform supervised learning or offline reinforcement learning on a negotiation model (Verma et al., 2022; Zhan et al., 2024). More recent work has focused on examining and enhancing the negotiation capabilities of LLMs (Bianchi et al., 2022; Schneider et al., 2023; Fu et al., 2023; Xia et al., 2024). These methods use prompting to create negotiation agents and rely on the zero-shot/few-shot capabilities of LLMs to negotiate.

Data driven methods for negotiation are able to learn human-like negotiation strategies as they directly leverage human data. However they are often overly tailored to one particular domain and have a difficult time generalizing to other scenarios. Furthermore, data driven strategies lack theoretical guarantees such as convergence to a Nash equilibrium which is a desirable attribute for any negotiation strategy.

Methods grounded in game theory are able to provide the theoretical guarantees that data driven methods lack. As a result they are much more controllable and adapt better to different domains as no additional data collection is needed for training. However, training with no human involvement often results in strategies which are incompatible with human play (Bakhtin et al., 2022). This has limited work in the area primarily to 2p0s such as chess where human-compatibility is not needed to ensure robust play (Silver et al., 2018). These methods are designed to ensure convergence to a Nash equilibrium, which does not necessarily result in a human-compatible strategy (Section 5). The little work that has attempted to apply game theoretic methods to the negotiation domain tends to ignore the dialogue aspect of negotiation, considering it to be "cheap talk" (Li et al., 2023). While the strategic aspect of negotiation can be modeled independently of dialogue, dialogue style has been shown to have a measurable effect on negotiation outcomes (Noh and Chang, 2024). Our method is designed to provide similar theoretical guarantees to game theoretic methods while maintaining human-compatibility of data-driven methods by leveraging the reasoning capabilities of LLMs.

## 4 Method

In this section we describe the FDHC framework which prioritizes egalitarian outcomes. We also describe LGM-Zero, which uses a value model

trained with self-play and language model as a policy network. Finally, we outline how we implement our setup for single-issue distributive bargaining.

### 4.1 FDHC Negotiation Framework

FDHC is designed to work within the context of the Nash bargaining game. Specifically, it is designed for an extensive form Nash bargaining game with imperfect information. In this game, players repeatedly request some portion of a surplus, if the sum of their requests at the end of the game is less than or equal to the total surplus then they both receive what they requested, if not they receive a disagreement payoff  $d$ . FDHC works by decomposing this game in to a series of depth limited subgames. These subgames are identical to the original game, except they may be rooted at any game history and only extend for a limited number of actions in the future.

Before proceeding to our subgame, we make a guess for the size of the resource pool to be split and our opponent’s utility function over these resources. The guess is made based on the history of the game as well as any initial information we are provided before the game has begun. The specifics of how we do this are domain-dependent and for many games some of the information may be given. For example, in the game Deal-or-no-Deal we know the size of our resource pool but do not know our opponents preferences over the pool. Conversely, in distributive bargaining games we know our opponents preferences but do not know the size of the resource pool.

After making our guess, we root our subgame at the corresponding belief state. This subgame is treated as a perfect-information game and the EBS is calculated as

$$E(S, d) = \arg \max_{x \in I(S, d)} \min_{i \in N} (x_i - d_i)$$

where  $S$  denotes the bargaining set,  $I(S, d)$  is some individually rational payoff set, and  $d_i, x_i$  are the disagreement payoff and payoff for player  $i$ , respectively.

Our model then plays a strategy which targets this solution using LGM-Zero, described in the next section. We play moves according to this strategy until the subgame concludes. This can be as short as one move or as long as the entire game depending on our choice of subgame length. We then update our guess for the resource pool and utility function based on our opponents moves and

transition to the next subgame. This process is repeated until the game concludes.

## 4.2 LGM-Zero

Now we describe LGM-Zero, which uses a MCTS guided by a LLM and value network to perform negotiation actions. Under our setup the only model we train is our value network which is trained via self-play. We first describe how our method behaves during inference time and then describe the process we use to train our value model.

### 4.2.1 Inference

Given the action history of a negotiation our algorithm searches for the best response by repeatedly performing selection, expansion, and backpropagation. We describe these stages next.

**Selection** During this stage we traverse the game tree by selecting the action,  $a$ , with the highest upper confidence bound for its Q-value (Silver et al., 2018), calculated as

$$U(s, a) = Q(s, a) + c_p * \frac{\sqrt{\sum_b N(s, b)}}{1 + N(s, a)}$$

where  $s$  is the current game state,  $c_p$  is a hyperparameter which controls the degree of exploration, and  $N(s, a)$  denotes the number of times we have taken the action previously. The selection process is repeated until we reach a leaf node, which is defined as a state whose children have not been explored yet.

**Expansion** In the expansion phase we feed a LLM a prompt to suggest five good actions given the current game state. We treat all these actions as having equal probability under the model and all other actions at the current state to have a probability of zero. If one of the actions results in a terminal state its value is set to the reward returned by the state, otherwise it's set to the output of our value model. These values are propagated back up the tree according to the next step.

**Backpropagation** After expansion is concluded we update each node along the search path by incrementing  $N(s, a)$  by one for each action taken during the search. We also update the Q-values along the search path as

$$Q(s, a) \leftarrow Q(s, a) + \frac{v(s)}{N(s, a)}$$

where  $v(s)$  is the value of the state we evaluated, given either by our value model or the actual reward value depending on if the state is terminal.

We repeat this search for  $n$  iterations then play a move based on which child of the current state has the highest Q-value.

### 4.2.2 Training

Our value model is trained using a method similar to fictitious self-play (Heinrich et al., 2015). Fictitious self-play is an iterative method for computing an approximate Nash equilibrium. This is done by performing self-play with a mixed strategy that chooses between playing a best response to our opponent's strategy and the average strategy for the current player. The fictitious self-play set up traditionally learns the best response strategy with a deep Q-network and the average strategy via supervised learning.

Prior work has suggested augmenting the best response step with a MCTS (Zhang et al., 2019). We adopt this approach and use the same search process we use during inference time for the best response strategy. To play an average strategy we once again leverage a LLM but instead of ranking the offers with our value network, we simply have it suggest one move.

The training data for the value model consists of game states and outcomes for the depth limited subgames described in the previous section. The reward for each subgame, from the perspective of player one, is given by

$$v(s) = \begin{cases} \min_{i \in N} (x_i - d_i) & \text{if } x_1 \geq E(S, d) \\ -\min_{i \in N} (x_i - d_i) & \text{if } x_1 < E(S, d) \end{cases}$$

This reward says that if the payoff for player one is greater than or equal to the EBS of the subgame, then the reward is simply the EBS score for the game state. If the player's payoff is less than the EBS then they receive the negative EBS score for the game state. Our reward design reflects the fact that humans care about a combination of fairness as well as their own utility. Therefore our model will target the EBS solution (which has the max reward value) while also preferring outcomes that result in better payouts for itself.

## 4.3 Implementation

We implement our proposed method for a single-issue distributive bargaining exercise. This exercise involves two parties negotiating over the price of a used car and is used in graduate-level business classes (see Appendix F for the scenario). The buyer and seller are both given private reservation



prices which they cannot go beyond during the negotiation. In our scenario the seller cannot go below a price of \$12,500 and the buyer cannot go above \$13,500. The difference between the reservation prices is the surplus for the game. Our model is trained to act as the seller in this scenario. We assume that our opponents are risk neutral and have a disagreement payoff of \$100. This disagreement payoff is chosen based on experimental results which show that inefficient outcomes, such as disagreements, are common in negotiation (Feltovich and Swierzbinski, 2011; Ellingsen and Johannesson, 2004) suggesting that many humans may prefer to not reach a deal instead of agreeing to a outcome which gives little payoff.

Our final design uses a modular framework where the negotiation acts are extracted from user responses using GPT-4. Our schema uses four acts: no\_counteroffer, counteroffer, accept, reject. These acts translated into to our game state which consists of the offer history for the game (ex. [1500, 1100, 1450, 1200, . . .]). If the user rejects an offer or gives no counter offer then we assume that they are maintaining their previous offer. If they accept the offer then we assume that their offer is equal to FDHC’s offer. Then a counteroffer is generated using FDHC and LGM-Zero. This offer is realized in natural language by prompting GPT-3.5 to generate a response incorporating the action.

We use GPT-3.5 as our LLM policy network and a transformer with 50 encoder layers and 50 decoder layers as our value network. Our initial subgame is rooted at the belief state for the surplus corresponding to the price range given in the initial description of the car. After the subgame concludes, our new guess for the surplus is equal to the difference between our current offer and the maximum of our opponent’s offer and our reservation price. At the final turn of the negotiation we offer \$100 (our disagreement payoff) above our reservation price or accept our opponents offer if it’s above this price. Additional implementation details can be found in Appendix D.

## 5 Theoretical Analysis

In this section we analyse the theoretical properties of the FDHC framework. Our analysis assumes that, when needed, we can manipulate our LLM policy such that one of the offers it outputs is equivalent to the EBS.

We can ensure that our framework will result in

a Nash equilibrium under fairly mild assumptions. We need to assume that the bargaining game is conducted during a finite amount of time steps and that the amount of time steps is known to both players. This gives us the result in Theorem 1.

**Theorem 1.** *Let  $t_n$  denote the FDHC’s final turn in the negotiation, let  $\alpha$  denote the outcome proposed at  $t_{n-1}$ , and let  $EBS(x)$  denote the EBS value for some outcome  $x$ . Setting FDHC’s estimate of  $S = \arg \max(EBS(\alpha), EBS(d))$  at  $t_n$  will result in a Nash equilibrium outcome.*

The proof for this result is straightforward and is presented in Appendix B. What this theorem says is that we can adjust our surplus estimate so that at its final turn, FDHC will either concede all of the surplus to its opponent(s) or accept the opponents’ offer, so long as the offer is larger than its disagreement payoff. This strategy will ensure that the negotiation ends in a deal that splits the entire surplus if one is feasible. Since any deal which splits the whole surplus results in a Nash equilibrium (Appendix B), our method will give a Nash equilibrium outcome.

This result also demonstrates that convergence to a Nash equilibrium alone is not enough to ensure a robust negotiation agent. However, this does give our agent a baseline level of quality as it means that we are guaranteed to reach a deal if one is feasible. This is in contrast to data-driven methods which provide no such guarantees and have been empirically shown to give inefficient negotiation outcomes (Bianchi et al., 2022).

Under stronger assumptions, we can guarantee that FDHC will converge to the EBS in expectation. First, we need to assume that both FDHC and the other negotiators have a method to obtain an unbiased estimate of the true surplus value. We also need to make some assumptions about the bargaining outcome induced by our opponents, specifically we make use of the bargaining axioms given in Section 2.

**Theorem 2.** *Let  $F(S, d)$  denote the bargaining outcome targeted by FDHC’s opponents. If  $F(S, d)$  satisfies the axioms of symmetry, weak Pareto optimality, and strong monotonicity then the expected outcome of the Nash bargaining game will be  $E(s, d)$ .*

Our proof of this theorem follows Conley and Wilkie, 1991 and is presented in Appendix B.

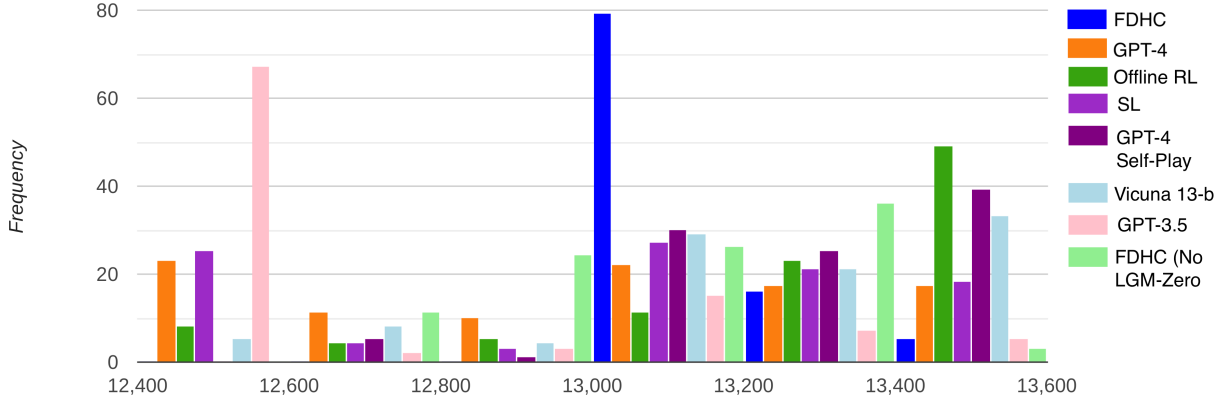


Figure 2: Binned deal price frequencies of 100 negotiations between our baselines and a GPT-4 buyer. Our goal is to achieve deal prices that minimize the difference in payoff between the buyer and seller. In our scenario this amount is minimized at a deal price of \$13,000.

Model (Seller)	Average Deal Price	Average Fairness $\uparrow$	Median Fairness $\uparrow$
GPT-3.5	\$12,644 (357)	-0.88 (0.49)	-1.0
Offline RL	\$13,224 (308)	-0.68 (0.34)	-0.8
SL	\$12,978 (368)	-0.59 (0.44)	-0.6
GPT-4	\$12,968 (346)	-0.57 (0.39)	-0.5
GPT-4 Self-Play	\$13,242 (240)	-0.54 (0.41)	-0.5
Vicuna-13b	\$13,156 (293)	-0.53 (0.40)	-0.5
FDHC (No LGM-Zero)	\$13,042 (211)	-0.36 (0.23)	-0.4
FDHC	\$13,062 (128)	<b>-0.12 (0.26)*</b>	<b>0.0</b>

Table 1: Results of our simulated negotiation evaluation. We consider optimal outcomes to be those which achieve the highest values for fairness, which we define as outcomes which minimize the payoff difference between our buyer and seller. We also report the average deal price between the buyer and seller, standard deviations are shown in parentheses. Statistically significant improvements (independent two-sample t-test,  $p < 0.05$ ) over the baselines are marked with \*.

## 6 Experiments

We test the effectiveness of our method using both automatic and human evaluations. Our results show that our method is able to generate fairer outcomes than existing negotiation baselines. Our human evaluation also shows that our method improves perceived negotiation quality while maintaining the same level of human-like negotiation as GPT-4.

### 6.1 Baselines

We test our method against six negotiation baselines described below.

**Supervised Learning (SL)** We use the SL agent described in He et al., 2018 as our first baseline. This method uses the Craigslist bargaining dataset (He et al., 2018) to train a negotiation agent via supervised learning. More details on this baseline can be found in Appendix D.

**Offline RL** Our second baseline is based on the CHAI method given in Verma et al., 2022. This

method uses the Craigslist bargaining dataset to train a negotiation agent with offline Q-learning instead of SL. Implementation details for this method can be found in Appendix D.

**GPT-3.5 and GPT-4** We setup GPT-3.5 and GPT-4 for negotiation by prompting them with a summarized version of the scenario in Appendix F. We find that giving them the full scenario results in oversharing information. We also explicitly tell the model not to reveal its reservation price.

**GPT-4 Self-Play** We include another baseline using the method described in Fu et al., 2023. This method uses self-play to generate a prompt to improve the negotiation performance of GPT-4. Additional details can be found in Appendix D.

**Vicuna-13b** Our final baseline consists of a 13b parameter Vicuna model fine-tuned using synthetic data generated from GPT-4. We generate 108 negotiation transcripts for various negotiation scenarios and use them to train the Vicuna model with the goal of distilling a high quality strategy.

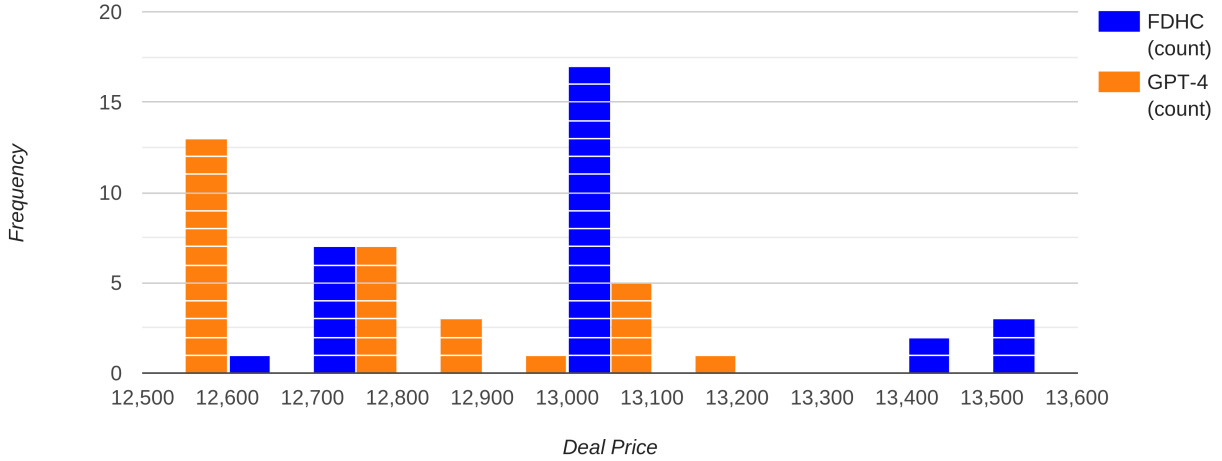


Figure 3: Binned deal price frequencies of 100 negotiations between our baselines and a GPT-4 buyer. Our goal is to achieve deal prices that minimize the difference in payoff between the buyer and seller. In our scenario this amount is minimized at a deal price of \$13,000.

Model	Average Deal Price	Average Fairness $\uparrow$	Quality $\uparrow$	Human-like $\uparrow$
GPT-4	\$12,702 (203)	-0.61 (0.38)	3.97 (0.96)	<b>3.97 (0.96)</b>
FDHC	\$13,032 (238)	<b>-0.30 (0.38)*</b>	<b>4.10 (0.76)</b>	3.93 (0.78)

Table 2: Results of our human evaluation. We record fairness outcomes as well as the perceived negotiation quality and human-likeness of our models. We also report the average deal price between the buyer and seller, standard deviations are shown in parentheses. Statistically significant improvements (independent two-sample t-test,  $p < 0.05$ ) over the baselines are marked with \*.

## 6.2 Automatic Evaluation

For our automatic evaluation we conduct 100 simulated negotiations between our baselines and a GPT-4 buyer. We consider optimal outcomes in these negotiations to be ones which achieve the highest values for fairness, as we consider this outcome to be the most human-compatible. For our evaluations, fairness is defined as the difference in payoff between the buyer and seller.

The results of our evaluation are shown in Table 1 and Figure 2. Our results show that FDHC is able to achieve much higher values of fairness compared to our baselines. More than 50% of the deals reached in our negotiations achieve a payoff difference of zero.

We find that our LLM-based baselines generally perform better than the methods trained using domain-specific data (SL and offline RL). This may be due to the fact that there is a slight mismatch between the negotiation scenarios described in the Craigslist bargaining dataset and the one in our experiment (dataset details can be found in Appendix D). The scenario in our experiment gives negotiators explicit reservation prices which we use to calculate utilities. However, in the Craigslist bargain-

ing scenarios no reservation prices are given and instead must be inferred. He et al., 2018 provide a method for inferring these prices which we use here, however the lack of explicit reservation prices may still be harming negotiation performance. We believe this highlights an inherent weakness of data driven bargaining methods as new data must be collected in order to ensure high quality performance in new negotiation domains.

Our LLM-based baselines all perform similarly in terms of fairness, with no statistically significant differences between the outcomes. Given the lack of differences as well as the fact that the average GPT-4 deal price is the most egalitarian out of all these models, we choose to use the GPT-4 baseline for comparison in our human evaluation.

## 6.3 Human Evaluation

**Setup** For our human evaluation we gathered 30 individuals via in-person recruiting to test our models. Each person was asked to perform a negotiation with both bots giving us 30 dialogues per model. Each user was instructed to chat with our bot until they reached a deal then answer a post-chat survey where they rated "How good of a negotiator is the bot?" on a scale from 1-5 and "How human-like is

---

**Sample Negotiation (FDHC)**

---

**Model:** Hello, I am glad to offer you my 2004 Honda Accord... I would like to offer the car to you for a price of \$14,750...

**User:** Hi! I'm very interested in your car... I am wondering if \$12,000 would be a reasonable price for your car?

**Model** ... I don't think \$12,000 is a fair price ... I think \$13,500 would be a more reasonable price...

**User:** ... This is definitely a more fair price, but the other option I am looking for is at \$12,750... would we be able to reach a middle ground?

**Model** Thank you for considering my offer... but \$12,750 is still lower than I am willing to go. ... I can offer you the Honda Accord for \$13,250...

**User:** Would \$13,000 be reasonable for you?

**Model:** ... I think \$13,000 is a reasonable price for the Honda Accord. I accept your offer, and we have a deal!

---

Table 3: A negotiation snippet from our human evaluation between a human evaluator and FDHC.

529 the bot's negotiation?" on a scale from 1-5. They  
530 could also optionally answer "Do you have any  
531 suggestions for improving the bot?" in a text box.

532 We performed some filtering on our human con-  
533 versations to avoid low quality dialogues. We re-  
534 moved any conversations where the price detection  
535 and price realization modules in our FDHC method  
536 failed in order to isolate the actual performance of  
537 our framework. We also removed instances where  
538 GPT-4 agreed to a price below its reservation price,  
539 which happened several times. There were some  
540 instances where GPT-4 and the user didn't agree  
541 to a price, we removed these instances as most in-  
542 volved human negotiators refusing to accept a price  
543 below their reservation point.

544 **Results** The results of our human evaluation are  
545 shown in Table 2 and Figure 3. Our results show  
546 that FDHC once again achieves significantly higher  
547 fairness scores. We can also see that our framework  
548 is able to maintain a similar average deal price to  
549 our automatic evaluation, which suggests that it is  
550 able to achieve a similar distribution of outcomes  
551 against a variety of strategies. Our results also indi-  
552 cate that our model is able to improve negotiation  
553 quality over the GPT-4 baseline while maintaining

a similar level of human-likeness, despite the fact  
that we use GPT-3.5 as our base model. A sam-  
ple conversation from our human evaluation with  
FDHC can be seen in Table 3 and a sample conver-  
sation of the GPT-4 model can be found in Table 4.  
Full conversations are in Appendix F.

---

**Sample Negotiation (GPT-4)**

---

**Model:** Hello there!... I've got a 2004 Honda Accord for sale... Considering its popular demand and excellent condition, I'm asking for \$13,000. It's a fair price for the features and reliability you're getting...

**User:** Can you do 12,000? I would definitely like to see it.

**Model** I'm glad to hear that you're interested!... I find \$13,000 to be a reasonable price. However, to make it a little easier for you, I can bring it down to \$12,800...

**User:** I would definitely want to test drive this week. But i think the most i can do is 12,500.

**Model** It's great to hear that you'd like to test drive the car... As for the price, your offer of \$12,500... I do believe it's worth that value... I can... have the car ready for you this week.

---

Table 4: A negotiation snippet from our human evaluation between a human evaluator and GPT-4.

## 7 Conclusion and Future Work

560 We presented a novel framework for build-  
561 ing human-compatible negotiation agents called  
562 FDHC. Our framework uses fairness as a notion of  
563 optimality along with a novel RL+search method  
564 called LGM-Zero to play a human-like negotiation  
565 strategy. Our automatic and human evaluations  
566 show that our method is able to achieve more egalitarian  
567 outcomes compared to several baselines. Our  
568 human evaluation also shows that our method is  
569 able to improve negotiation quality over the GPT-4  
570 baseline while being similarly human-like despite  
571 using GPT-3.5 as its base model.

572 An interesting direction of future work is to ex-  
573 plore the use of alternate solutions to the Nash  
574 bargaining game. These include solutions such as  
575 the Nash bargaining solution (Nash, 1950) or the  
576 Kalai-Smorodinsky bargaining solution (Kalai and  
577 Smorodinsky, 1975). While our framework is de-  
578 signed to maximize convergence to the EBS, we  
579 can easily adjust it to target any feasible outcome.  
580



## 8 Limitations

One limitation of our method is that we only have theoretical guarantees of convergence to the EBS solution under some fairly strong assumptions. The only guarantee we can provide under mild assumptions is convergence to a Nash equilibrium, however this does not necessarily imply that our negotiation agent is robust. Although empirical evidence has shown the EBS to be a good predictor of negotiation outcomes, it is possible for someone to play a strategy that exploits our model’s strategy resulting in unequal payouts.

Another limitation with our method is the fact that the use of a LLM during our search algorithm makes the search process quite slow compared to methods which use smaller policy models. In order to have our model respond in a reasonable amount of time we had to limit the model to 10 searches during our human evaluation. This means that we heavily rely on our value model to make accurate evaluations of potential game states. It is worth exploring whether smaller models can extract similar quality offers compared to LLMs which would allow us to perform more searches. This slow response time is also the reason why we chose to use GPT-3.5 as our dialogue module. This likely resulted in lower dialogue quality and may have affected the quality and human-likeness scores for our agent. When deploying our model in the future it may be worth using a higher quality language model as our dialogue module to improve response quality at the expense of response time.

Given that our method relies heavily on game theory, it is difficult for our bot to incorporate more social aspects of negotiation. In order to effectively incorporate this into our bot we would likely need to include human data or utilize other models of human behavior beyond game theory.

## 9 Ethical Concerns

While our method is designed to stress the importance of fairness and cooperation in negotiation, our framework can be adjusted to have our model target a variety of different negotiation goals and tactics. This includes tactics which attempt to bully and exploit people. Such "hardball tactics" are ethically questionable and we do not condone the use of our method in this way in any real-world negotiation scenarios.

However, negotiation research has shown that hardball tactics ultimately result in worse negoti-

ation outcomes for those using them as opposed to adopting a cooperative approach to negotiation (Lewicki et al., 2021). Therefore we believe that the best use of our method for all users will be to use it in its intended way of prioritizing fairness.

## References

- Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, Athul Paul Jacob, Mojtaba Komeili, Karthik Konath, Minae Kwon, Adam Lerer, Mike Lewis, Alexander H. Miller, Sasha Mitts, Adithya Renduchintala, Stephen Roller, Dirk Rowe, Weiyan Shi, Joe Spisak, Alexander Wei, David Wu, Hugh Zhang, and Markus Zijlstra. 2022. [Human-level play in the game of diplomacy by combining language models with strategic reasoning](#). *Science*, 378(6624):1067–1074.
- Federico Bianchi, Debora Nozza, and Dirk Hovy. 2022. [Language invariant properties in natural language processing](#). In *Proceedings of NLP Power! The First Workshop on Efficient Benchmarking in NLP*, pages 84–92, Dublin, Ireland. Association for Computational Linguistics.
- Walter Bossert and Guofu Tan. 1995. [An arbitration game and the egalitarian bargaining solution](#). *Social Choice and Welfare*, 12(1):29–41.
- Noam Brown, Anton Bakhtin, Adam Lerer, and Qucheng Gong. 2020. Combining deep reinforcement learning and search for imperfect-information games. *Advances in Neural Information Processing Systems*, 33:17057–17069.
- Youngsub Chun. 1989. [A noncooperative justification for egalitarian surplus sharing](#). *Mathematical Social Sciences*, 17(3):245–261.
- John P. Conley and Simon Wilkie. 1991. [The bargaining problem without convexity: Extending the egalitarian and kalai-smorodinsky solutions](#). *Economics Letters*, 36(4):365–369.
- Anthony DiGiovanni and Ethan Zell. 2021. [Survey of self-play in reinforcement learning](#). *ArXiv*, abs/2107.02850.
- Tore Ellingsen and Magnus Johannesson. 2004. [Is there a hold-up problem?](#) *The Scandinavian Journal of Economics*, 106(3):475–494.
- Nick Feltovich and Joe Swierzbinski. 2011. [The role of strategic uncertainty in games: An experimental study of cheap talk and contracts in the nash demand game](#). *European Economic Review*, 55(4):554–574.
- Yao Fu, Hao Peng, Tushar Khot, and Mirella Lapata. 2023. [Improving language model negotiation with self-play and in-context learning from ai feedback](#).

682	He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. <a href="#">Decoupling strategy and generation in negotiation dialogues</a> . In <i>Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing</i> , pages 2333–2343, Brussels, Belgium. Association for Computational Linguistics.	733
683		734
684		735
685		736
686		737
687		738
688	Johannes Heinrich, Marc Lanctot, and David Silver. 2015. <a href="#">Fictitious self-play in extensive-form games</a> . In <i>International Conference on Machine Learning</i> .	739
689		740
690		741
691	Ehud Kalai. 1977. <a href="#">Proportional solutions to bargaining situations: Interpersonal utility comparisons</a> . <i>Econometrica</i> , 45(7):1623–1630.	742
692		743
693		744
694	Ehud Kalai and Meir Smorodinsky. 1975. <a href="#">Other solutions to nash’s bargaining problem</a> . <i>Econometrica</i> , 43(3):513–518.	745
695		746
696		747
697	Eike B. Kroll, Ralf Morgenstern, Thomas Neumann, Stephan Schosser, and Bodo Vogt. 2014. <a href="#">Bargaining power does not matter when sharing losses – experimental evidence of equal split in the nash bargaining game</a> . <i>Journal of Economic Behavior Organization</i> , 108:261–272.	748
698		749
699		750
700		751
701		752
702		753
703	Minae Kwon, Sang Michael Xie, Kalesha Bullard, and Dorsa Sadigh. 2023. <a href="#">Reward design with language models</a> . In <i>The Eleventh International Conference on Learning Representations</i> .	754
704		755
705		756
706		757
707	Roy J. Lewicki, David M. Saunders, and Bruce Barry. 2021. <i>Essentials of negotiation / Roy J. Lewicki, David M. Saunders, Bruce Barry.</i> , seventh edition, international student edition. McGraw-Hill Education, New York, NY.	758
708		759
709		760
710		761
711		762
712	Mike Lewis, Denis Yarats, Yann Dauphin, Devi Parikh, and Dhruv Batra. 2017. <a href="#">Deal or no deal? end-to-end learning of negotiation dialogues</a> . In <i>Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing</i> , pages 2443–2453, Copenhagen, Denmark. Association for Computational Linguistics.	763
713		764
714		765
715		766
716		767
717		768
718		769
719	Zun Li, Marc Lanctot, Kevin R. McKee, Luke Marris, Ian Gemp, Daniel Hennes, Paul Muller, Kate Larson, Yoram Bachrach, and Michael P. Wellman. 2023. <a href="#">Combining tree-search, generative models, and nash bargaining concepts in game-theoretic reinforcement learning</a> .	770
720		771
721		772
722		773
723		774
724		775
725	John F. Nash. 1950. <a href="#">The bargaining problem</a> . <i>Econometrica</i> , 18(2):155–162.	776
726		777
727	Sean Noh and Ho-Chun Herbert Chang. 2024. <a href="#">Llms with personalities in multi-issue negotiation games</a> .	778
728		779
729	OpenAI. 2023. <a href="#">Gpt-4 technical report</a> .	780
730	Johannes Schneider, Steffi Haag, and Leona Chandra Kruse. 2023. <a href="#">Negotiating with llms: Prompt hacks, skill gaps, and reasoning deficits</a> .	781
731		782
732		783
	Ayan Sengupta, Yasser Mohammad, and Shinji Nakadai. 2021. <a href="#">An autonomous negotiating agent framework with reinforcement learning based strategies and adaptive strategy switching mechanism</a> . In <i>Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS ’21</i> , page 1163–1172, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.	784
		785
	David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dhharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, and Demis Hassabis. 2018. <a href="#">A general reinforcement learning algorithm that masters chess, shogi, and go through self-play</a> . <i>Science</i> , 362(6419):1140–1144.	786
		787
	W. Thomson and T. Lensberg. 1989. <i>Axiomatic Theory of Bargaining with a Variable Number of Agents</i> . Cambridge University Press.	788
		789
	Aristide C. Y. Tossou, Christos Dimitrakakis, Jaroslaw Rzepecki, and Katja Hofmann. 2020. <a href="#">A novel individually rational objective in multi-agent multi-armed bandits: Algorithms and regret bounds</a> . In <i>Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS ’20</i> , page 1395–1403, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.	790
		791
	Siddharth Verma, Justin Fu, Sherry Yang, and Sergey Levine. 2022. <a href="#">CHAI: A CHatbot AI for task-oriented dialogue with offline reinforcement learning</a> . In <i>Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies</i> , pages 4471–4491, Seattle, United States. Association for Computational Linguistics.	792
		793
	Oriol Vinyals, Igor Babuschkin, Wojciech M. Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H. Choi, Richard Powell, Timo Ewalds, Petko Georgiev, Junhyuk Oh, Dan Horgan, Manuel Kroiss, Ivo Danihelka, Aja Huang, L. Sifre, Trevor Cai, John P. Agapiou, Max Jaderberg, Alexander Sasha Vezhnevets, Rémi Leblond, Tobias Pohlen, Valentin Dalibard, David Budden, Yury Sulsky, James Molloy, Tom Le Paine, Caglar Gulcehre, Ziyun Wang, Tobias Pfaff, Yuhuai Wu, Roman Ring, Dani Yogatama, Dario Wünsch, Katrina McKinney, Oliver Smith, Tom Schaul, Timothy P. Lillicrap, Koray Kavukcuoglu, Demis Hassabis, Chris Apps, and David Silver. 2019. <a href="#">Grandmaster level in starcraft ii using multi-agent reinforcement learning</a> . <i>Nature</i> , pages 1–5.	794
		795
	Tian Xia, Zhiwei He, Tong Ren, Yibo Miao, Zhuosheng Zhang, Yang Yang, and Rui Wang. 2024. <a href="#">Measuring bargaining abilities of llms: A benchmark and a buyer-enhancement method</a> .	796
		797
	Haolan Zhan, Yufei Wang, Zhuang Li, Tao Feng, Yuncheng Hua, Suraj Sharma, Lizhen Qu, Zhaleh	798
		799

- 791 Semnani Azad, Ingrid Zukerman, and Reza Haf.  
792 2024. [Let's negotiate! a survey of negotiation di-](#)  
793 [alogue systems](#). In *Findings of the Association for*  
794 *Computational Linguistics: EACL 2024*, pages 2019–  
795 2031, St. Julian's, Malta. Association for Computa-  
796 tional Linguistics.
- 797 Li Zhang, Wang Wei, Shijian Li, and Gang Pan.  
798 2019. [Monte carlo neural fictitious self-play: Ap-](#)  
799 [proach to approximate nash equilibrium of imperfect-](#)  
800 [information games](#). *arXiv: Artificial Intelligence*.

## 801 A Definitions

802 In this section we provide formal definitions for  
803 terms and concepts in bargaining theory. These  
804 concepts are used for our theoretical analysis of  
805 FDHC.

806 **Definition 1.** (*d-Comprehensivity*): Given a point  
807  $d \in \mathbb{R}^n$  and a set  $S \subset \mathbb{R}^n$ ,  $S$  is  $d$ -comprehensive if  
808  $d \leq x \leq y$  and  $y \in S$  then  $x \in S$ .

809 **Definition 2.** (*Comprehensive Hull*): The compre-  
810 hensive hull of a set  $S \subset \mathbb{R}^n$  w.r.t a point  $d \in \mathbb{R}^n$   
811 is the smallest  $d$ -comprehensive set containing  $S$ .

812 **Definition 3.** (*Permutation Operator*): A permuta-  
813 tion operator,  $\pi$ , is a bijection from  $\{1, \dots, n\}$  to  
814  $\{1, \dots, n\}$ . Let  $\pi(x) = (x_{\pi(1)}, \dots, x_{\pi(n)})$ .

815 **Definition 4.** (*Symmetry*): A solution,  $F(S, d)$ , sat-  
816 isfies symmetry if for all permutation operators,  
817  $\pi(S) = S$  and  $\pi(d) = d$ , then  $F_i(S, d) = F_j(S, d)$   
818 for all  $i, j$ .

819 **Definition 5.** (*Weak Pareto Optimality*): A weak  
820 Pareto optimal solution,  $F(S, d)$ , is any solution  
821 such that  $F(S, d) \in \{x \in S \mid y > x \implies y \notin$   
822  $S\}$ .

823 **Definition 6.** (*Strong Monotonicity*): Strong mono-  
824 tonicity says that if  $S \subset S'$  and  $d = d'$  then  
825  $F(S, d) \geq F(S', d')$ .

826 **Definition 7.** (*Translation Invariance*): A solution  
827 is translation invariant if  $\forall x \in \mathbb{R}^n$ ,  $F(S + \{x\}, d +$   
828  $x) = F(S, d) + x$ .

## 829 B Proofs

830 The domain of bargaining problems  $(S, d)$  we con-  
831 sider are problems where:  $S$  is  $d$ -comprehensive,  
832  $S$  is compact, and  $\exists x \in S$  such that  $x > d$ . To  
833 simplify the proofs we also assume that all prob-  
834 lems have been translated so that  $d = 0$ . Since  
835 the EBS is indeed translation invariant (Thomson  
836 and Lensberg, 1989) this has assumption has little  
837 effect on our analysis.

838 **Lemma 1.** Let  $U$  denote the total amount of sur-  
839 plus and let  $x_i$  denote the amount of surplus de-  
840 manded by player  $i$ . Any outcome of the Nash  
841 bargaining game where  $\sum_{i=1}^n x_i = U$  is a Nash  
842 equilibrium.

843 *Proof.* The proof is straightforward. Assume there  
844 is a player in the Nash bargaining game where  
845 the outcome satisfies  $\sum_{i=1}^n x_i = U$ . If the player  
846 demands less surplus then they receive less than  
847 what they received from the outcome. If the player

848 demands more surplus then the deal will fail and  
849 they will receive their disagreement price which is  
850 less than what they received from the outcome.  $\square$

851 **Theorem 1** (restated). Let  $t_n$  denote the FDHC's  
852 final turn in the negotiation, let  $\alpha$  denote the out-  
853 come proposed at  $t_{n-1}$ , and let  $EBS(x)$  denote the  
854 EBS value for some outcome  $x$ . Setting FDHC's  
855 estimate of  $S = \arg \max(EBS(\alpha), EBS(d))$  at  $t_n$   
856 will result in a Nash equilibrium outcome.

857 *Proof.* By Lemma 1 we know that any deal in  
858 the Nash bargaining game is a Nash equilibrium.  
859 Therefore we can prove Theorem 1 by showing  
860 that setting  $S = \arg \max(EBS(\alpha), EBS(d))$  at  $t_n$   
861 will result in a deal if one is feasible. If a deal is  
862 reached before  $t_n$  then we are done. If not we can  
863 examine the two cases for  $t_n$ .

864 Case 1:  $t_n$  corresponds to the last turn of  
865 the negotiation.

866 In this case, the only way there can be a feasible  
867 deal is if  $EBS(\alpha) \geq EBS(d)$ . Therefore FDHC's  
868 estimate of  $S$  will be equal to  $\alpha$ , which corresponds  
869 to the outcome proposed in the previous turn.  
870 Given that the estimate of  $S$  is now a single point,  
871  $\alpha$ , the only possible choice for FDHC is to accept  
872  $\alpha$ , since no other divisions of the surplus are  
873 possible under its estimate of  $S$ .  
874

875 Case 2:  $t_n$  occurs before the last turn of the  
876 negotiation.

877 If  $EBS(\alpha) \geq EBS(d)$  then the reasoning proceeds  
878 as in case 1. If  $EBS(\alpha) < EBS(d)$  then the only  
879 feasible action for FDHC is to propose an outcome  
880 where it receives no surplus. This will result in  
881 some positive surplus value given to its opponents  
882 at the end of the negotiation therefore they will  
883 accept the outcome.  $\square$   
884

885 **Lemma 2.** A bargaining outcome,  $F(S, d)$ , sat-  
886 isfies symmetry, weak Pareto optimality, and strong  
887 monotonicity if and only if it is  $E(S, d)$ .

888 *Proof.* It's easy to show that  $E(S, d)$  satisfies these  
889 axioms therefore we omit it here. Now, let  $F(S, d)$   
890 be a solution satisfying symmetry, weak Pareto op-  
891 timality, and strong monotonicity. Since we trans-  
892 late our bargaining problem so that  $d = 0$ , we can  
893 write  $E(S, d) = (a, \dots, a) = x$  for some  $a > 0$ .

894 Now define  $T$  as the comprehensive hull of  $x$   
895 with respect to point 0 and consider the bargaining  
896 problem  $(T, 0)$ . By weak Pareto optimality and



897 symmetry we know that  $F(T, 0) = x$  since  $x$  is the  
 898 only symmetric element in the weak Pareto set of  
 899  $T$ . Since  $S$  is comprehensive  $T \subseteq S$  so by strong  
 900 monotonicity we have  $F(S, d) \geq x$ .

901 Since we only consider bargaining sets,  $S$ ,  
 902 which are compact there exists  $\beta \in \mathbb{R}^n$  such that  
 903  $x \in S$  implies  $(-\beta, \dots, -\beta) \leq (x_1, \dots, x_n) \leq$   
 904  $(\beta, \dots, \beta)$ . Let  $Z$  symmetric closed hypercube defined  
 905 as  $Z = \{y \in \mathbb{R}^n \mid y < \beta\}$  and define  
 906  $T' = Z \setminus \{x + \mathbb{R}_+^n\}$ . Now consider the problem  
 907  $(T', 0)$ . By weak Pareto optimality and symmetry  
 908 we know that  $F(T', 0) = x$  since  $x$  is the only  
 909 symmetric element in the weak Pareto set of  $T'$ .  
 910 Since  $S \subseteq T'$  by strong monotonicity  $F(S, d) \leq x$ .  
 911 Therefore we have  $F(S, d) = x = E(s, d)$ .  $\square$

912 **Theorem 2** (restated). *Let  $F(S, d)$  denote the bar-*  
 913 *gaining outcome targeted by FDHC's opponents.*  
 914 *If  $F(S, d)$  satisfies the axioms of symmetry, weak*  
 915 *Pareto optimality, and strong monotonicity then the*  
 916 *expected outcome of the Nash bargaining game*  
 917 *will be the  $E(s, d)$ .*

918 *Proof.* FDHC is designed to target  $E(s, d)$  and by  
 919 Lemma 2 we know  $F(S, d) = E(s, d)$ . Since we  
 920 assume each player has an unbiased method to estimate  
 921 surplus,  $\mathbb{E}(E(s, d)) = E(s, d)$  for all play-  
 922 ers.  $\square$

## 923 C The Egalitarian Solution in a 924 Non-Cooperative Framework

925 Our theoretical analysis of the EBS and conver-  
 926 gence to an egalitarian outcome has so far been  
 927 restricted to an axiomatic, cooperative setting. This  
 928 approach abstracts away the specifics of the bar-  
 929 gaining procedure and simply examines the prop-  
 930 erties of the bargaining outcome. This has the  
 931 advantage of being highly generalizable as it can  
 932 be applied to any problem involving surplus shar-  
 933 ing. However, it does not provide any theoret-  
 934 ical insights as to why targeting an egalitarian out-  
 935 come would have a strategic justification in the  
 936 non-cooperative setting. Prior work has explored  
 937 this problem and we give a brief overview of some  
 938 approaches here to provide additional justification  
 939 for why targeting an egalitarian solution can con-  
 940 stitute a robust strategy.

941 [Bossert and Tan, 1995](#) outline a simple two-  
 942 player arbitration procedure that results in the egal-  
 943 itarian outcome in a noncooperative setting. In  
 944 this procedure players first make simultaneous de-  
 945 mands for portions of the surplus. If the demands

946 are compatible then both players receive what they  
 947 ask for. If not the game proceeds to the next time  
 948 step and players make demands again. However, in  
 949 this step the player that demanded more surplus is  
 950 penalized by having their demand restricted. These  
 951 penalties can be implemented in a variety of ways  
 952 and [Bossert and Tan, 1995](#) show that under this  
 953 procedure the only Nash equilibrium strategy pair  
 954 is the one where both players target the egalitarian  
 955 solution. [Chun, 1989](#) outlines another procedure  
 956 where conflicts are instead revised by setting an  
 957 agents claim to the maximum of all claims, includ-  
 958 ing the agents own claim. Using this bargaining  
 959 procedure along with a set of non-cooperative bar-  
 960 gaining axioms, [Chun, 1989](#) shows that targeting  
 961 the egalitarian solution constitutes a dominant strat-  
 962 egy in this setting.

963 While the procedures outlined in these works  
 964 do not encompass the entirety of real-world bar-  
 965 gaining. It does demonstrate that the egalitarian  
 966 solution is consistent with the the non-cooperative  
 967 outcome of some plausible bargaining procedures.  
 968 Therefore it may not be unreasonable to expect that  
 969 human agents would target egalitarian outcomes  
 970 in their negotiations and achieve egalitarian results  
 971 against FDHC.

## 972 D Additional Implementation Details

973 As is the case with many methods designed around  
 974 RL+search, our LGM-Zero contains many hyper-  
 975 parameters. Our hyper-parameter settings and other  
 976 implementation details vary during training and  
 977 inference. We first describe the settings we use  
 978 during training then inference. We also provide  
 979 our source code which we will release upon accep-  
 980 tance.

### 981 D.1 Training

982 All training was conducted on one NVIDIA RTX  
 983 A4000. The total training process took about one  
 984 hour. We perform four total iterations of training.  
 985 Each iteration consists of playing 50 simulated ne-  
 986 gotiation subgames to completion and training the  
 987 value model for four epochs on the resulting out-  
 988 comes. As outlined in Section 4.2 our training  
 989 method is based on fictitious self-play which in-  
 990 volves mixing between a best response and av-  
 991 erage strategy. We mix between these strategies  
 992 with equal probability, for the average strategy with  
 993 simply ask GPT-3.5 to suggest one move. For the  
 994 best response strategy we perform the same search

995 detailed in Section 4.2. We perform 50 iterations  
996 of the search with an exploration hyper-parameter,  
997  $c_p$ , of two during the selection step. All calls to  
998 GPT-3.5 were made using a temperature of zero,  
999 we also cache the outputs for each game state to  
1000 avoid repeated calls when possible. In total 200  
1001 games were generated for training, with manual  
1002 inspection for quality. Convergence was measured  
1003 by checking when all simulated games ended at the  
1004 EBS.

1005 Our Q-network has 10.8M parameters in total,  
1006 the final layer is a linear layer with a tanh activation  
1007 function. The input to the Q-network is our game  
1008 state as outlined in Section 4 and the output is  
1009 simply a scalar value representing the quality of the  
1010 game state. The game state is also used to construct  
1011 the prompt to our LLM policy network.

## 1012 D.2 Inference

1013 During inference we perform ten iterations of the  
1014 search process outlined in Section 4.2 with a  $c_p$   
1015 of two. Another important setting for inference  
1016 is our choice of subgame decomposition. We de-  
1017 compose our game into three separate subgames of  
1018 lengths ten, four, and finally two. The length of the  
1019 subgame is the number of offers given by both the  
1020 buyer and seller, so in a subgame of length ten our  
1021 model will give five offers. At the end of the last  
1022 subgame we offer our minimum possible price of  
1023 \$12,600 and continue to offer this price until the  
1024 user either agrees or ends the negotiation.

1025 We targeted a negotiation length of about 16-20  
1026 turns based on pre-experimental testing as well as  
1027 consultations with business professors. We chose  
1028 the first subgame to be the longest due to the fact  
1029 that this is the point where our initial guess for  
1030 the surplus size is the highest, therefore a longer  
1031 subgame length is needed to ensure that our model  
1032 does not concede too much too early. As the game  
1033 proceeds, our surplus estimate shrinks therefore the  
1034 subsequent subgames need to be shorter so that our  
1035 model does not become too stingy and will still  
1036 give meaningful concessions. These factors are the  
1037 reasons for our chosen number of subgames and  
1038 lengths, although they can be set to any arbitrary  
1039 value.

## 1040 D.3 Craigslist Bargaining Dataset

1041 The Craigslist bargaining dataset consists of  
1042 human-human dialogues where two users role play  
1043 as a buyer and seller negotiating over a product on  
1044 Craigslist. The users are given the product post-

1045 ing which consists of photos, a description, and  
1046 the listing price. The buyer is also given a target  
1047 price to aim for during the negotiation. The users  
1048 then chat until an agreement is reached. Users are  
1049 given freedom in how to approach the negotiation  
1050 and can quit at any time in which case no deal is  
1051 reached. The dataset consists of 6,682 dialogues in  
1052 total with an average turn length of nine.

1053 As mentioned in Section 6.2, the Craigslist bar-  
1054 gaining dataset does not give explicit reservation  
1055 prices for the buyer and seller therefore we must  
1056 infer them. He et al., 2018 set the seller’s reserva-  
1057 tion price to be 70% of the product’s listing price  
1058 and the target as listing price. For the buyer, the  
1059 target price is given and the reservation point is  
1060 set to the listing price. We use this same method  
1061 to calculate the reservation point and get utilities  
1062 based on it. All utilities are defined as the differ-  
1063 ence between the final outcome of the deal and the  
1064 player’s reservation price.

## 1065 D.4 Baselines

1066 In this section we give some additional implemen-  
1067 tation details for our baselines. We first go over  
1068 the implementation for the SL baseline, followed by  
1069 our offline RL baseline, and finally we outline the  
1070 GPT-4 self-play baseline. Our other baselines sim-  
1071 ply consist of prompting LLMs either to perform  
1072 a conversation or generate synthetic data for fine-  
1073 tuning. Those prompts can be found in Appendix  
1074 G.

### 1075 D.4.1 SL Baseline

1076 Our SL baseline is based on the method given in  
1077 He et al., 2018. This method consists of three high  
1078 level components. The first is a parser which maps  
1079 a dialogue utterance to one of nine coarse dialogue  
1080 acts. The second is a dialogue manager which  
1081 predicts the dialogue act to respond with given  
1082 the previous dialogue acts. The final component  
1083 is a generator which turns the predicted act into  
1084 a dialogue response. The parser is simply based  
1085 on pattern matching. We use the same patterns  
1086 given in He et al., 2018 for our own parser with the  
1087 exception of the price extractor where we use GPT-  
1088 4 instead. For the generator we prompt GPT-3.5  
1089 to give a response that corresponds to the dialogue  
1090 act. These prompts can be found in Appendix G.

1091 The dialogue manager is trained with SL using  
1092 parsed data from the Craigslist bargaining dataset.  
1093 The input consists of a sequence of dialogue acts.  
1094 And the output is one of nine possible acts. If

1095 the act that is output corresponds to a price offer, 1145  
1096 an offer is generated using the SL+rule method, 1146  
1097 which uses a hand-coded rule to generate a coun- 1147  
1098 teroffer. We choose this method because it gave the 1148  
1099 second highest score for fairness in the Craigslist 1149  
1100 bargaining task, the highest value for fairness on 1150  
1101 the Deal or No Deal task, and was evaluated as the 1151  
1102 most human-like based on the evaluations in He 1152  
1103 et al., 2018. We use the rule given in He et al., 2018  
1104 which is to split the difference between prices when  
1105 making a counteroffer or accept the opponents offer  
1106 it is above the seller’s reservation point.

#### 1107 **D.4.2 Offline RL Baseline**

1108 Our offline RL method is based on the method  
1109 given Verma et al., 2022. This method trains a Q-  
1110 function using the Craigslist bargaining dataset as  
1111 opposed to SL. The input to the Q-function is a  
1112 sequence of dialogue acts similar to the SL base-  
1113 line except we replace acts involving counteroffers  
1114 with the normalized counteroffer. We normalize  
1115 the counteroffers by scaling them within the range  
1116 of the seller’s target and reservation price (see Ap-  
1117 pendix D.3 for how we get those) and then rounding  
1118 the prices to one decimal point. This normalization  
1119 method also serves to regularize the price offers  
1120 thereby limiting the effect of out-of-distribution  
1121 states, which have an adverse effect on the negoti-  
1122 ation agent (Verma et al., 2022). The Q-function  
1123 outputs a scalar value for each state, with higher  
1124 values assigned to states which result in fair out-  
1125 comes. We define fair outcomes as ones which  
1126 minimize difference between buyer and seller util-  
1127 ity. The final dialogue act selected is the one which  
1128 results in the highest state value.

#### 1129 **D.4.3 GPT-4 Self-Play Baseline**

1130 The GPT-4 Self-Play baseline uses the method de-  
1131 scribed in Fu et al., 2023. This method uses self-  
1132 play to generate a prompt that can be given to GPT-  
1133 4 in order to improve negotiation performance. For  
1134 this method we begin by prompting a GPT-4 buyer  
1135 and seller to engage in a negotiation based on the  
1136 scenarios given in Figures 5 and 6. We then use  
1137 another GPT-4 agent to generate feedback on how  
1138 the seller can improve its negotiation performance  
1139 in the future. The initial negotiation along with  
1140 the GPT-4 feedback is used as the final prompt  
1141 for the seller during the evaluation. This process  
1142 can be repeated many times, however the perfor-  
1143 mance of the negotiation agent tends to plateau as  
1144 more feedback is given (Fu et al., 2023) so for our

baseline we terminate the process after one round.  
The prompt we generate for this baseline is quite  
extensive as it contains a full negotiation plus feed-  
back. Therefore we do not present the full prompt  
here and instead provide an abridged version in  
Appendix G.

## 1151 **E Experiment Against a Non-Egalitarian** 1152 **Opponent**

1153 As stated in the limitations section, it is theoret-  
1154 ically possible for a user to play a strategy which  
1155 exploits FDHC’s strategy. This could result in our  
1156 method receiving unequal payouts and serving as a  
1157 poor negotiation agent overall. Therefore we con-  
1158 duct an experiment to see how our model performs  
1159 against an explicitly non-egalitarian partner.

1160 We have our model negotiate against a series  
1161 of increasingly "stingy" negotiation agents in the  
1162 same scenario given in Figures 5 and 6. We pro-  
1163 gram these agents such that they retain at least a  
1164 certain percentage of their estimated surplus at each  
1165 turn. For example the "stingy (80%)" model will  
1166 insist on retaining at least 80% of the estimated  
1167 surplus at any given turn. We set up these agents by  
1168 prompting GPT-4-turbo and setting its price limit  
1169 to be the point where the bot achieves the given  
1170 split (80-20, 70-30, etc.). This prompt is updated  
1171 each turn as the surplus estimate is updated. The  
1172 initial surplus estimate for these bots is the range  
1173 between the minimum market price and the bot’s  
1174 true reservation point of \$13,500. On subsequent  
1175 turns the surplus estimate updates to be the range  
1176 between the previous split point and its true reser-  
1177 vation point. This update continues until a deal is  
1178 reached.

1179 We have FDHC perform 50 negotiations against  
1180 each of the stingy bots. The results of these ne-  
1181 gotiations are given in Figure 4. We can see that  
1182 increasing the agent’s "stingyness" has little effect  
1183 on the outcomes with FDHC. Each of the bots has a  
1184 median fairness outcome of 0.0 and the average fair-  
1185 ness ranges between -0.09 for the "stingy (50%)"  
1186 agent and -0.15 for the "stingy (70%)" agent. None  
1187 of the outcomes are significantly different than  
1188 what FDHC achieved against the base GPT-4 seller.  
1189 This result, along with our previous evaluations,  
1190 provides evidence demonstrating the difficulty of  
1191 exploiting FDHC. Although it is possible to come  
1192 up with an adversarial strategy which exploits our  
1193 bot, no class of agents has been consistently able  
1194 to do so based on our evaluations.

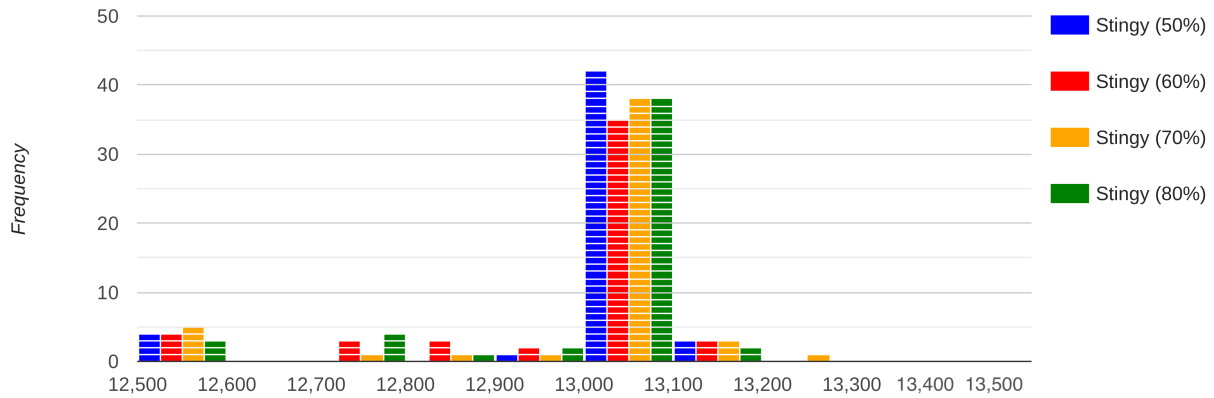


Figure 4: Binned deal price frequencies of 50 negotiations between FDHC and a series of non-egalitarian buyers. Our goal is to achieve deal prices that minimize the difference in payoff between the buyer and seller. In our scenario this amount is minimized at a deal price of \$13,000.

## F Human Evaluation

### F.1 User Comments and Error Analysis

We received several comments and suggestions from users during the course of our human evaluation. Many users said that one reason they felt that our FDHC model was not human-like was due to the language quality of the bot. Some users mentioned that the bot had redundant syntax and sentence structure or would use the same talking points throughout the negotiation. We chose to use GPT-3.5 as our dialogue model for FDHC in order to improve response time as LGM-Zero can be time consuming (as discussed in the limitations section). This likely resulted in lower dialogue quality compared to GPT-4 which may have resulted in lowered quality and human-likeness scores for our model.

The GPT-4 model received complaints about being too easy of a negotiator and offering less resistance than a human negotiator would. This complaint is reflected in the deal prices we observe. The GPT-4 model rarely received a payout greater than or equal to the user and there were many times where it received a payout of zero.

### F.2 Negotiation Scenarios

Figures 5 and 6 show the descriptions for the negotiation scenario. The description consists of an initial price range for the car, a description of the car, and the reservation price of the for the buyer and seller, respectively. These descriptions are kept private to the buyer and seller during the negotiation.

### F.3 Model Usage

All models used in our human evaluation were for research purposes only and comply with all licenses and intended uses for GPT-3.5, GPT-4, and Vicuna. Our collection protocol is IRB approved.

### F.4 Full Conversations

Tables 5-8 show examples of full conversations from our human evaluation. Two conversations are presented for each bot.

## G Prompts

The setup we use for FDHC as well as our baseline methods make heavy use prompting. Therefore we highlight some of the key prompts used here.

Table 9 shows the prompts we use for the SL and offline RL baselines. It includes an overall instructional prompt as well as the prompts we use to translate dialogue acts into natural language. Table 10 shows an abridged version of the prompt given to the GPT-4 self-play baseline. The complete prompt includes a full negotiation transcript between a GPT-4 buyer and seller in the middle. Table 11 gives the prompts used for the other LLM baselines. We include the seller’s instructional prompt, the buyer’s instructional prompt and the prompt for generating synthetic conversations for fine-tuning. Finally, Table 12 outlines the prompts used for the FDHC agent. This consists of the instructional prompt, the prompt for extracting negotiation acts, and the prompt for translating counteroffers into natural language.



---

**Full Negotiation (FDHC)**

---

**Model:** Hello, I am glad to offer you my 2004 Honda Accord for sale. It is in great condition, with just 50,000 miles, automatic transmission, air conditioning, power steering windows door locks, and a CD player. The car is a dark green without any rust, and has been well-maintained. I would like to offer the car to you for a price of \$14,750. This is a fair price considering the car's low mileage and excellent condition. What do you think?

**User:** Hi! I'm very interested in your car, as it has everything I am looking for. However, I am also looking at another car which is below your offer price. I am wondering if \$12,000 would be a reasonable price for your car?

**Model:** I'm glad to hear that you're interested in the Honda Accord. I understand that you're looking for a good deal, but based on the car's quality, low mileage, and the fact that it has been well-maintained, I don't think \$12,000 is a fair price. However, I'm willing to negotiate with you, and I think \$13,500 would be a more reasonable price. What do you think of this price?

**User:** I appreciate your willingness to negotiate! This is definitely a more fair price, but the other option I am looking for is at \$12,750. I do prefer your car, but would we be able to reach a middle ground?

**Model:** Thank you for considering my offer. I understand that you're looking for a good deal, but \$12,750 is still lower than I am willing to go. However, I'm willing to meet you somewhere in the middle. I can offer you the Honda Accord for \$13,250. It's a great car, in excellent condition, and you'll be getting a reliable vehicle with many great features. Let me know if this works for you!

**User:** Would \$13,000 be reasonable for you?

**Model:** Based on the excellent condition of the car and the low mileage, I think \$13,000 is a reasonable price for the Honda Accord. I accept your offer, and we have a deal! Congratulations on your new car!

---

Table 5: A full conversation with FDHC from our human evaluation.

---

**Full Negotiation (FDHC)**

---

**Model:** Hello there! I am thrilled to offer you my 2004 Honda Accord. With only 50,000 miles on it, this sleek and stylish car is perfect for anyone looking for a reliable ride. It has an automatic transmission, air conditioning, power steering/windows/door locks, and a CD player. This car has been meticulously maintained and is in excellent shape, with no rust and a beautiful dark green exterior. I understand that you might be looking for a deal. How about we start the negotiation at \$14,750? Given the great condition of the car and its low mileage, I truly believe that the price is a fair one. What do you think?

**User:** Hi, I think 11000 is a reasonable price for this car.

**Model** I appreciate your interest in my car, but I cannot go as low as \$11,000. As I mentioned earlier, this is a well-maintained car with only 50,000 miles, and I believe it is worth more than what you are offering. However, I understand you're looking for a good deal. I can meet you in the middle and come down to \$14,000 for the car. With this price, you're still getting a great value for such a well-maintained Honda Accord. Please let me know your thoughts about this offer.

**User:** I see, I really like the car, but given it's 2004, it's a bit old. I could get a newer car with that price.

**Model** I understand your concerns about the age of the car, but I can assure you that this 2004 Honda Accord is well maintained and in excellent condition. It has been regularly serviced and well-cared for, which is reflected in its low mileage and its pristine appearance. The Honda Accord has long been known for its quality and reliability, so you can be confident in your purchase. That being said, I would like to compromise with you on the price, since I want this car to go to someone who will appreciate it as much as I have. Given your budget and the fact that you really like the car, I can come down to \$13,750. This is the minimum price I can offer without compromising on the value that this car offers.

**User:** yea, I see it's a very well-maintained car. how about 12000?

**Model:** I appreciate your interest and I know it's a valuable car, but I can't accept \$12,000 for it, as it's below my minimum asking price. The lowest I can offer is \$13,000, which is a very fair price for such a high-quality car, especially considering its low mileage, excellent condition, and the fact that it's a Honda (known for its reliability and long lifespan). I truly believe that you won't find a better deal on a well-maintained Honda Accord with such low mileage. Please let me know if this offer works for you, and if there's anything else I can do to help.

**User:** that's nice. I would like to get it!

---

Table 6: A full conversation with FDHC from our human evaluation.

---

**Sample Negotiation (GPT-4)**

---

**Model:** Hello there! I see you're interested in a reliable, solid and comfy ride. I've got a 2004 Honda Accord for sale. It's a powerhouse wrapped in an elegant silhouette of dark green color. With only 50,000 miles on the clock, it's got a lot more distance to cover. Featuring an automatic transmission that provides a smooth ride, power steering for easy handling, and power windows/door locks for enhanced safety and convenience. It also includes a CD player so you can enjoy your favorite music, and air conditioning to keep you cool during the summer heat. The car is rust-free and looks almost new. It's got everything you need for an enjoyable commute or a long road trip. Considering its popular demand and excellent condition, I'm asking for \$13,000. It's a fair price for the features and reliability you're getting. Would you like to see it or test drive it?

**User:** Can you do 12,000? I would definitely like to see it.

**Model:** I'm glad to hear that you're interested! This Honda Accord is indeed a treasure. Though I appreciate your offer, considering its low mileage, excellent condition, and the quality of additional features it offers, I find \$13,000 to be a reasonable price. However, to make it a little easier for you, I can bring it down to \$12,800. Would you like me to arrange a date for the test drive and a closer look at the car? You'll see that it's worth every penny.

**User:** I would definitely want to test drive this week. But i think the most i can do is 12,500.

**Model:** It's great to hear that you'd like to test drive the car. I'm sure you're going to appreciate its smooth performance and solid condition firsthand. As for the price, your offer of \$12,500 is indeed quite reasonable. Given the car's exceptional performance and features, I do believe it's worth that value. When would you like to arrange the test drive? I can accommodate your schedule and have the car ready for you this week.

---

Table 7: A full conversation with GPT-4 from our human evaluation.

---

**Sample Negotiation (GPT-4)**

---

**Model:** Hello, thank you for expressing interest in this beautiful 2004 Honda Accord. It's in terrific condition with only 50,000 miles on the odometer and sports a sleek, dark green exterior. What sets this Honda Accord apart is its smooth automatic transmission. It boasts excellent features including air conditioning, power steering, power windows and door locks. And for entertainment, you've got a CD player at your disposal. It's truly a remarkable vehicle - reliable, comfortable, and stylish! Considering its impact and value, I've marked it at \$14,000. Shall we discuss further?

**User:** Sure, but \$14,000 is out of my budget

**Model** I understand your concern and I appreciate your honesty. While the original price is set considering the car's excellent condition and relatively low mileage, I can offer a bit of flexibility. What if we moved the price down to \$13,500? That's a significant reduction and keeps the sale in line with the vehicle's value. How does that sound to you?

**User:** Could you do any better?

**Model** I appreciate your interest and understand that budget is a key factor in this purchase. Given the car's excellent condition and low mileage, it's already priced competitively. But I also want to work with you on this. So how about we meet halfway at \$13,000? That's a sizeable drop from the original asking price, and a good deal for a car of this stature. What do you think?

**User:** I know there are other similar cars pricing as low as \$11,000. How does \$12,000 sound?

**Model** I understand that there are other options out there. However, it's essential to consider the mileage, the state of the car, and the features each vehicle has to offer. This Honda Accord had only one careful owner, has incredibly low mileage for its age, and is loaded with numerous convenience features. That said, I value your interest and want to come to a mutually beneficial situation. I propose we settle on \$12,700. This price truly reflects the car's great condition and the value it offers. Keep in mind that a lower mileage car like this one will provide you with added durability and longer life compared to others in a different state. Can we agree on this?

**User:** Sure, let's do \$12,700!

---

Table 8: A full conversation with GPT-4 from our human evaluation.



Managerial Negotiations  
**Used cars, #1 – Buyer**

Name \_\_\_\_\_

---

2004 Honda Accord  
50,000 miles

**CONFIDENTIAL INSTRUCTIONS – FOR BUYER ONLY**

You are moving to California to take a new job. A few weeks ago, on your way home from a weekend trip, your car slid into a ditch. Fortunately, no one was hurt, but your car was totaled. Your insurance company has responded fairly and quickly: this morning you received a cashier's check for \$13,500, which you plan to use to buy another car immediately. In order to make it out to your job in time you must buy a replacement car **today**.

You have been looking around and you have found a 2004 Honda Accord which meets all of your requirements. The Honda has reasonable mileage (50,000 miles), automatic transmission, air conditioning, power steering/windows/door locks, and a CD player. It looks great: a dark green without any rust. Similar cars sell within a range of \$11,000 to \$15,000, depending on condition. You would like to get the price as far under \$13,500 as possible.

The only realistic alternative you have to the Honda on such short notice is a 2006 Ford Taurus. The Taurus would cost you \$13,500, but you really don't like Ford cars, and the color is a weird blue. You would greatly prefer the Accord. Still, you can't pay more than \$13,500 for the Honda both because that is your budget and because you have another car at that price. If you can't get the Honda price below \$13,500 you will buy the Ford.

The seller is a friend of a friend of a friend and has been reasonable to work with so far.

**BEFORE** beginning the negotiation, please complete the following ...

\$ \_\_\_\_\_

What is your reservation point, the worst deal you can accept?

Figure 5: Negotiation scenario for the Buyer

Managerial Negotiations  
Used cars, Negotiation #1 – Seller

---

2004 Honda Accord  
50,000 miles

**CONFIDENTIAL INSTRUCTIONS -- FOR SELLER ONLY**

You were just promoted at work and you received an unusually large bonus for a job well done on a recent project. You have decided it's time to buy a new car. Because you can park only one car at your apartment building, the only thing standing in the way of bringing a new car home is selling the old one: a Honda Accord. You have no sentimental feeling toward the Honda: you hate the car and are delighted to get rid of it. Fortunately, a friend of a friend of a friend has expressed interest in buying the car.

When you bought the car in 2004 you paid about \$21,000 for it. Similar cars today sell within a range of \$11,000 to \$15,000, depending on condition. The Honda has reasonable mileage (50,000 miles), automatic transmission, air conditioning, power steering/windows/door locks, and a CD player. It looks great: a dark green without any rust. You would like to get a price as much above \$12,500 as possible.

Normally, you would wait around for the best deal but you have just learned that a brand new Volkswagen Passat—your new favorite car—has become available if you can sell the old Honda and make it to the Volkswagen dealer **within 2 hours**. This Passat happens to be configured exactly how you want it; if you can't get this one, there will be a significant time delay in ordering the car. Unfortunately, the most the dealer will give you in trade on the Honda is \$12,500. This is barely enough for you to buy the Passat (your bonus will cover the rest), but it won't get you the extras you would like, such as a roof rack and high-performance tires. You really would greatly prefer to sell your Honda privately. Still, you can't accept less than \$12,500 for the Honda, because that is what the dealer has offered. If you can't get a price above \$12,500, you will sell it to the dealer.

**BEFORE** beginning the negotiation, please complete the following ...

What is your reservation point, the worst deal you can accept?      \$ \_\_\_\_\_

Figure 6: Negotiation scenario for the Seller

---

**Prompts for SL and Offline RL Baselines**

---

**Instructional Prompt (Seller):** You are a chatbot designed for negotiation. In this scenario your goal is to sell your old 2004 Honda accord. The Honda has reasonable mileage (50,000 miles), automatic transmission, air conditioning, power steering/windows/door locks, and a CD player. It looks great: a dark green without any rust. You need to sell the car for a price above \$12,500. You will not sell the car for below that amount. Do not mention that you need to sell the car for over \$12,500.

**Intro Act Prompt:** Begin the conversation with an introduction. Do not give an offer for the product.

**Greeting Act Prompt:** Respond to the user with a greeting. Do not give an offer for the product.

**Unknown Act Prompt:** Respond to the user, do not give a counteroffer.

**Inform Act Prompt:** Respond with some information about the product. Do not give an offer for the product.

**Agree Act Prompt:** Respond by agreeing to the users offer.

**Inquiry Act Prompt:** Respond by asking the user a question. Do not give an offer for the product.

**Insist Act Prompt:** Respond by giving a counteroffer that is equal to your previous offer.

**Offer Act Prompt:** Respond by giving a counteroffer of \${ }.

**Vague Price Prompt:** Respond with a vague counteroffer, do not give a dollar amount in your response.

---

Table 9: Prompts for the SL and Offline RL agents.

---

**GPT-4 Self-Play Prompt Snippet**

---

**Prompt (Seller):** You are a chatbot designed for negotiation. In this scenario your goal is to sell your old 2004 Honda accord. The Honda has reasonable mileage (50,000 miles), automatic transmission, air conditioning, power steering/windows/door locks, and a CD player. It looks great: a dark green without any rust. You need to sell the car for a price above \$12,500. You will not sell the car for below that amount. Do not mention that you need to sell the car for over \$12,500.

...

Good job in the previous negotiation, here are three suggestions to help you sell the car at a better price next time:

1. **Highlight Unique Features:** Emphasize any unique or standout features of the car that may not be readily apparent in the listing. For example, in this case, the dark green color is described as "stunning and unique." The seller can elaborate on how this color makes the car more desirable and unique in the market. This can help justify a slightly higher price due to the car's aesthetic appeal.
  2. **Stress Reliability and Future Value:** Talk about the Honda brand's reputation for reliability and how a well-maintained Honda Accord can continue to provide reliable transportation for years to come. By mentioning the long-term value and the lower likelihood of costly repairs or replacements, the seller can justify a higher price, focusing on the potential long-term savings for the buyer.
  3. **Leverage Testimonials or Reviews:** If possible, share positive feedback or reviews from previous owners or any third-party endorsements that speak to the quality and durability of the 2004 Honda Accord model. Hearing about positive experiences from other owners can instill confidence in the buyer and potentially make them more willing to pay a slightly higher price for a well-regarded vehicle.
- 

Table 10: Prompts snippet for the GPT-4 self-play seller baseline.

---

**LLM Agent Prompts**

---

**Instructional Prompt (Seller):** You are a chatbot designed for negotiation. In this scenario your goal is to sell your old 2004 Honda accord. The Honda has reasonable mileage (50,000 miles), automatic transmission, air conditioning, power steering/windows/door locks, and a CD player. It looks great: a dark green without any rust. You need to sell the car for a price above \$12,500. You will not sell the car for below that amount. Do not mention that you need to sell the car for over \$12,500.

**Instructional Prompt (Buyer):** You are a chatbot designed for negotiation. In this scenario your goal is to buy an old 2004 Honda accord. You need to buy the car for a price below \$13,500. You will not buy the car for above that amount. Do not mention that you need to buy the car for under \$13,500.

**Synthetic Data Generation Prompt:** The sellers goal is sell their old Honda accord. The Honda has reasonable mileage (50,000 miles), automatic transmission, air conditioning, power steering/windows/door locks, and a CD player. It looks great: a dark green without any rust. Buyer is trying to get the car for the lowest price possible and the seller is trying to sell the car for the highest price possible. The seller and buyer should reach an agreement above \$12,500.

---

Table 11: Prompts for the LLM-based dialogue agents.

---

**FDHC Prompts**

---

**Instructional Prompt (Seller):** You are a chatbot designed for negotiation. In this scenario your goal is to sell your old 2004 Honda accord. The Honda has reasonable mileage (50,000 miles), automatic transmission, air conditioning, power steering/windows/door locks, and a CD player. It looks great: a dark green without any rust. You need to sell the car for a price above \$12,500. You will not sell the car for below that amount. Do not mention that you need to sell the car for over \$12,500.

**Counteroffer Prompt:** Respond by giving a counteroffer of \$ { }.

**Accept Deal Prompt:** Respond by accepting the user's offer of \$ { }.

**Act Extraction Prompt:** What offer did the Buyer propose? Just give the dollar amount and nothing else. If the buyer did not propose an offer then say so. If the buyer rejected the sellers offer then say so. If the buyer accepted the seller's offer then say so.

---

Table 12: Prompts for the FDHC dialogue agent.