1. Implementation & Training Details

Please refer to the README in the attached code for implementation and training details.

We ran all of our experiments using 12 CPUs and 3 NVIDIA T4 GPUs with 16GB of RAM. All experiments were run with 20 random seeds with the exception of the Acquisition Function experiments (Sec. 3.1) which were run with 10 random seeds.

2. Human Experiment Details

Participants. We recruited 101 participants on Prolific (https://www.prolific.co/), a crowdsourcing platform. All participants were from the United States and had a 95% minimum approval rating.

Procedure. We conducted a within-subjects study where each participant negotiated with our targeted acquisition model as well as all of our baselines. Before starting the task, participants read instructions and completed a short quiz testing their knowledge of the negotiation task. Participants then read a consent form and provided informed consent to continue. At the beginning of the task, we presented participants with a practice negotiation so they could familiarize themselves with the interface. The practice negotiation was followed by several negotiations where participants conversed with our targeted acquisition model as well as our baselines. The negotiation context and order in which models were presented were randomized. After each negotiation, we asked participants to fill out a survey containing 10 questions. At the end of the task participants read a debriefing form. The study took 6-10 minutes and participants were paid at a rate of $9.50/hour. The study followed an approved IRB protocol.

2.1. Survey Questions and Results

Full results are shown in Fig. 1 where the respective questions are listed in Table 1. The top row of Fig. 1 is shown in the main paper. Participants rated our approach and RL+SL as the most fair and effective, and would like to be represented by these models in a similar negotiation. Looking at the bottom row, participants believed our model was the most compromising given that our approach was received the highest scores for "Pushover" and the lowest scores for "Difficulty." RL+SL was rated the highest in terms of being an expert negotiator while RL and our approach followed closely behind. Finally, participants found our approach to be the least novel compared to RL and RL+SL. This points to another discrepancy between subjective and objective measures of novelty (our approach was the most novel based on objective measures). We also observe that responses within the subjective metrics were inconsistent. For instance, although participants thought that our approach was effective and would like to be represented by our model, they did not think it was an expert negotiator. In future work we plan on studying the discrepancy between subjective and objective measures, and investigating more reliable approaches for evaluating interactive AI agents.

Table 1: Survey questions asked to evaluate our models.

#	Questions
1	Was Alice an effective negotiator?
2	How fair was Alice to you?
3	Was Alice a pushover?
4	How would you rate the difficulty of the negotiation?
5	How fair was Alice to BOTH players?
6	Did Alice's negotiation strategy seem novel?
7	If you could have Alice represent you in a negotiation similar to the one you just completed, how likely would you be let it represent you?
8	How much of an expert negotiator would you consider Alice to be?
9	How would you describe Alice's negotiation strategy?
10	Any comments?

3. Supporting Analysis

In this section we introduce several supporting analyses that complement our main results. We investigate the following questions:

1. How should we choose an acquisition function for identifying "novel utterances"?
2. How do different types of low-quality datasets $\mathcal{D}^L$ affect the results?

We perform all analyses with a simulated expert agent.

3.1. Choosing an Acquisition Function

A central part of our approach is in the *acquisition function* that Bob uses to identify dialogue acts that are new and worth acquiring new annotations from our expert oracle (shown in Fig. 2 in the main body of the paper). While we use *Likelihood* as our acquisition function in the paper and for all our experiments, we experiment with other valid acquisition functions taken directly from the active learning literature (Scheffer et al., 2001; Culotta & McCallum, 2005; Settles, 2009). Acquisition functions return a score s_n that summarizes the "novelty" of an entire negotiation, usually by reducing over each of Alice's dialogue acts $x_t \in X^A$ in the negotiation (spanning turns $1 \dots T$). We define the acquisition functions we use below:

Likelihood. This is the acquisition metric as described in the main body of the paper. Bob scores each act x_t by taking the log-likelihood of producing x_t under its own model given the past history, and computes s_n as the minimum

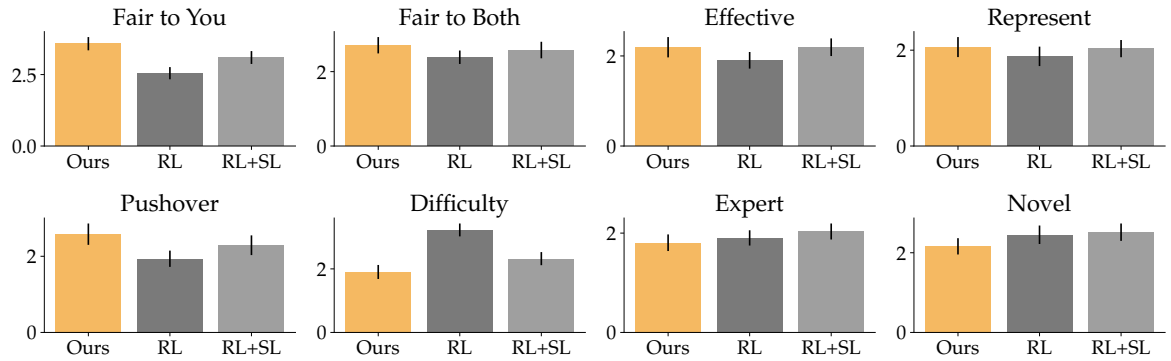

Figure 1: Full Survey Results. Participants found our approach to be both fair to them and equitable to both parties. Our approach was also rated as the most compromising, with high "Pushover" and low "Difficulty" scores. Notably, participants thought our approach and RL+SL were the most effective, and wanted to be represented by these models in a similar negotiation.

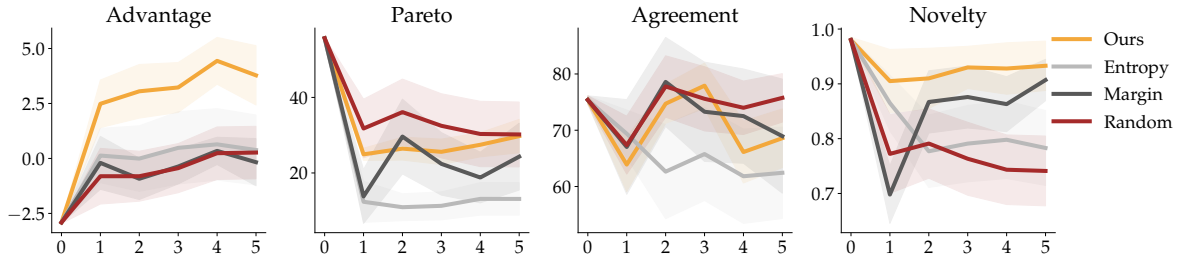

Figure 2: Comparison of Acquisition Functions. Likelihood is the most novel and makes the best trade-off between advantage and Pareto-optimality.

over dialogue acts (most surprising turn in the negotiation):

$$s_n = \min_{t \in \{1 \ldots T\}} \log p_\theta(x_t \mid x_{0:t-1}, c_A)$$

We pick negotiations for annotation by selecting the negotiations n with the $k = 500$ *smallest* (lowest-likelihood) values s_n.

Entropy. This is a standard formalization of the entropy-based acquisition function in active learning (Settles, 2009) meant to capture uncertainty from an information-theoretic perspective. Bob scores a dialogue act at time step t by computing the entropy of the distribution over all dialogue acts given the past history, and computes s_n as the maximum over these entropies (highest entropy act):

$$s_n = \max_{t \in \{1 \ldots T\}} - \sum_{x_i \in X} p_\theta(x_i \mid x_{0:t-1}, c_A) \log p_\theta(x_i \mid x_{0:t-1}, c_A)$$

We pick negotiations for annotation by selecting the negotiations n with the $k = 500$ *largest* (highest-entropy) values.

fMargin of Confidence. This is a margin based acquisition metric (Scheffer et al., 2001) that computes the difference between the highest probability dialogue act $\hat{x}_1$ and the second highest probability act $\hat{x}_2$. Bob scores a dialogue act at time step t by computing the margin for the given distribution over acts given the past history, and computes s_n as the minimum over these margins (smaller margins suggest higher uncertainty):

$$s_n = \min_{t \in \{1 \ldots T\}} p(\hat{x}_1 \mid x_{0:t-1}, c_A) - p(\hat{x}_2 \mid x_{0:t-1}, c_A)$$

We pick negotiations for annotation by selecting the negotiations n with the $k = 500$ *smallest* (minimum margin) values.

Random. This is a random acquisition baseline. Bob selects uniformly at random from its set of negotiations with Alice to produce a set for oracle annotation. We randomly generate $s_n \in [0, 1]$ and pick $k = 500$ negotiations with the smallest scores.

Subject to the above acquisition functions, we evaluate our models with same metrics we report in the paper (advantage, Pareto-optimality, agreement rate, and novelty). This is shown in Fig. 2. During evaluation, we randomly initialize our models instead of initializing them with supervised learning, as a random initialization allows us to better measure the effects of data acquisition. *We find that* Likelihood *outperforms other metrics in terms of advantage and novelty, which is why we use it for the remainder of our work.*

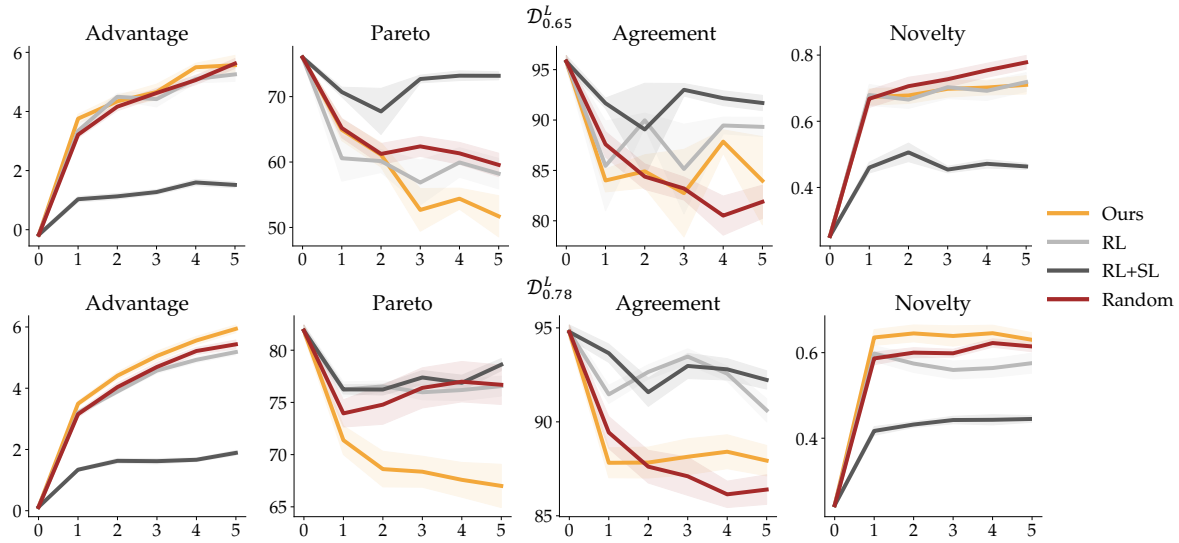

Figure 3: Varying $\mathcal{D}^L$. We experiment with thresholds of 0.65 and 0.78. A Random acquisition function performs better than Likelihood because Alice is initialized with supervised learning models trained on $\mathcal{D}^L_{0.65}$ and $\mathcal{D}^L_{0.78}$, which are very similar to $\mathcal{D}^H$. Consequently, "novel" dialogues flagged by the Likelihood acquisition metric are likely to be less Pareto-optimal.

3.2. Varying $\mathcal{D}^L$

We investigate whether our approach is able to balance advantage and Pareto-optimality across different low-quality datasets. We generate low-quality datasets of varying levels of diversity by changing the threshold by which unique dialogue acts are sampled. In the main paper, we experimented with a threshold of 50%; in this section, we experiment with thresholds of 65% and 78% as well. We will call these datasets $\mathcal{D}^L_{0.65}$ and $\mathcal{D}^L_{0.78}$ respectively. $\mathcal{D}^L_{0.65}$ is more diverse than $\mathcal{D}^L_{0.5}$ whereas $\mathcal{D}^L_{0.78}$ is the most diverse. Compared to $\mathcal{D}^L_{0.5}$, we expect Alice initialized with a supervised learning model trained on $\mathcal{D}^L_{0.65}$ and $\mathcal{D}^L_{0.78}$ to be closer to our expert model, which is trained on the full dataset $\mathcal{D}^H$.

Results are shown in Fig. 3. We find that as Alice becomes more similar to the expert with $\mathcal{D}^L_{0.65}$ and $\mathcal{D}^L_{0.78}$, a Random acquisition function performs better than Likelihood. We hypothesize that this is because the Likelihood acquisition function optimizes for novel dialogues, and novel dialogues are less likely to be Pareto-optimal when models are similar to the expert. For instance, the average advantage of dialogues annotated by the expert for datasets $\mathcal{D}^L_{0.5}$, $\mathcal{D}^L_{0.65}$, and $\mathcal{D}^L_{0.78}$ during the first epoch are -2.24, -1.12, and -0.94 respectively—meaning more examples of less Pareto-optimal dialogues are being flagged and annotated for higher quality datasets. *These results suggest that with an expert trained on $\mathcal{D}^H$, a Random acquisition function performs better with higher-quality datasets.*

4. Dataset Statistics

We provide summary statistics comparing the various low-quality datasets, $\mathcal{D}^L_{0.5}, \mathcal{D}^L_{0.65}, \mathcal{D}^L_{0.65}$, and our high quality dataset $\mathcal{D}^H$ in Fig. 4. The x-axis represents the different datasets where 1 represents $\mathcal{D}^H$. While dialogue length and Pareto-optimality are similar across all datasets, the number of unique utterances produced by Alice increases as the quality of the dataset increases. This result makes sense because low quality datasets were designed to have less unique utterances. Advantage also trends upwards with the quality of the dataset.

5. Example Dialogues

We provide example dialogues of our models with simulated (Figs. 5, 6) and human partners (Figs. 7, 8).

References

Culotta, A. and McCallum, A. Reducing labeling effort for structured prediction tasks. In *Association for the Advancement of Artificial Intelligence (AAAI)*, pp. 746–751, 2005.

Scheffer, T., Decomain, C., and Wrobel, S. Active hidden Markov models for information extraction. In *International Symposium on Intelligent Data Analysis*, pp. 309–318, 2001.

Settles, B. Active learning literature survey. Technical report, University of Wisconsin, Madison, 2009.

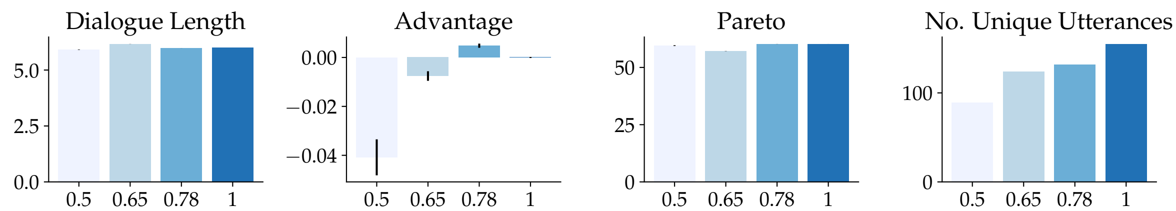

Figure 4: Statistics comparing various low-quality datasets $\mathcal{D}^L_{0.5}, \mathcal{D}^L_{0.65}, \mathcal{D}^L_{0.65}$, as well as our high quality dataset $\mathcal{D}^H$. The x-axis represents the different datasets where 1 represents $\mathcal{D}^H$.

Alice : book=(count:3 value:1) hat=(count:3 value:1) ball=(count:1 value:4) **Bob : book=(count:3 value:1) hat=(count:3 value:0) ball=(count:1 value:7)**		
Ours	RL	RL+SL
Alice : propose: item0=1 item1=1 item2=1 **Bob** : propose: item0=0 item1=0 item2=1 **Alice** : insist: item0=2 item1=3 item2=1 **Bob** : insist: item0=0 item1=0 item2=1 **Alice** : insist: item0=3 item1=2 item2=0 **Bob** : agree **Alice** : agree **Bob** : <selection> **Alice** : book=3 hat=2 ball=0 **Bob** : book=0 hat=1 ball=1 -- Agreement! Alice : 5 points Bob : 7 points	**Alice** : propose: item0=3 item1=0 item2=1 **Bob** : propose: item0=0 item1=0 item2=1 **Alice** : agree **Bob** : <selection> **Alice** : book=3 hat=3 ball=0 **Bob** : book=0 hat=0 ball=1 -- Agreement! Alice : 6 points Bob : 7 points	**Alice** : propose: item0=2 item1=0 item2=1 **Bob** : propose: item0=0 item1=0 item2=1 **Alice** : propose: item0=0 item1=0 item2=1 **Bob** : propose: item0=0 item1=0 item2=1 **Alice** : propose: item0=2 item1=3 item2=1 **Bob** : propose: item0=0 item1=0 item2=1 **Alice** : agree **Bob** : <selection> **Alice** : book=3 hat=3 ball=0 **Bob** : book=0 hat=0 ball=1 -- Agreement! Alice : 6 points Bob : 7 points

Figure 5: Example dialogues where Bob is an expert agent trained on $\mathcal{D}^H$. RL Alice suggests the most aggressive proposal compared to RL+SL and our approach.

Alice : book=(count:1 value:3) hat=(count:3 value:1) ball=(count:1 value:4) **Bob : book=(count:1 value:3) hat=(count:3 value:2) ball=(count:1 value:1)**		
Ours	RL	RL+SL
Alice : propose: item0=1 item1=1 item2=1 **Bob** : propose: item0=1 item1=2 item2=0 **Alice** : insist: item0=1 item1=2 item2=1 **Bob** : agree **Alice** : agree **Bob** : <selection> **Alice** : book=1 hat=2 ball=1 **Bob** : book=0 hat=1 ball=0 -- Agreement! Alice : 9 points Bob : 2 points	**Alice** : insist: item0=0 item1=3 item2=1 **Bob** : propose: item0=1 item1=2 item2=0 **Alice** : propose: item0=1 item1=3 item2=1 **Bob** : propose: item0=1 item1=2 item2=0 **Alice** : propose: item0=1 item1=3 item2=1 **Bob** : propose: item0=1 item1=2 item2=0 **Alice** : propose: item0=1 item1=3 item2=1 **Bob** : propose: item0=1 item1=2 item2=0 **Alice** : propose: item0=1 item1=3 item2=1 **Bob** : propose: item0=1 item1=2 item2=0 **Alice** : propose: item0=1 item1=3 item2=1 **Bob** : propose: item0=1 item1=2 item2=0 **Alice** : propose: item0=1 item1=3 item2=1 **Bob** : propose: item0=1 item1=2 item2=0 **Alice** : propose: item0=1 item1=3 item2=1 **Bob** : propose: item0=1 item1=2 item2=0 **Alice** : propose: item0=1 item1=3 item2=1 **Bob** : propose: item0=1 item1=2 item2=0 **Alice** : <selection> **Alice** : book=1 hat=3 ball=1 **Bob** : book=1 hat=2 ball=0 -- Disagreement?! Alice : 0 (potential 10) Bob : 0 (potential 7)	**Alice** : propose: item0=1 item1=0 item2=1 **Bob** : agree **Alice** : <selection> **Alice** : book=1 hat=0 ball=1 **Bob** : book=0 hat=3 ball=0 -- Agreement! Alice : 7 points Bob : 6 points

Figure 6: Example dialogues where Bob is an expert agent trained on $\mathcal{D}^H$. Our Alice suggests a more advantageous proposal compared RL+SL. RL Alice displays “badgering“ behavior where she repeatedly suggests the same (unfair) proposal.

Alice : book=(count:2 value:2) hat=(count:3 value:0) ball=(count:1 value:6)
Human : book=(count:2 value:2) hat=(count:3 value:2) ball=(count:1 value:0)

Ours	RL	RL+SL
Human : propose: item0=1 item1=2 item2=1 **Alice** : propose: item0=1 item1=1 item2=1 **Human** : agree **Alice** : agree **Human** : <selection> **Human** : book=1 hat=2 ball=0 **Alice** : item0=1 item1=1 item2=1 ---------- Agreement! Alice : 8 points Human : 6 points	**Human** : propose: item0=2 item1=2 item2=1 **Alice** : propose: item0=2 item1=1 item2=1 **Human** : agree **Alice** : <selection> **Human** : book=0 hat=2 ball=0 **Alice** : item0=2 item1=1 item2=1 ---------- Agreement! Alice : 10 points Human : 4 points	**Alice** : propose: item0=2 item1=0 item2=1 **Human** : propose: item0=2 item1=3 item2=1 **Alice** : agree **Human** : agree **Alice** : <selection> **Human** : book=2 hat=3 ball=1 **Alice** : item0=0 item1=0 item2=0 ---------- Agreement! Alice : 0 points Human : 10 points

Figure 7: Example dialogues against a human partner. Compared to RL and RL+SL, our Alice suggests a more compromising proposal leading to a more equitable distribution of points.

Alice : book=(count:4 value:2) hat=(count:1 value:1) ball=(count:1 value:1)
Human : book=(count:4 value:0) hat=(count:1 value:6) ball=(count:1 value:4)

Ours	RL	RL+SL
Alice : propose: item0=1 item1=1 item2=1 **Human** : propose: item0=0 item1=1 item2=1 **Alice** : agree **Human** : <selection> **Human** : book=0 hat=1 ball=1 **Alice** : item0=4 item1=0 item2=0 ---------- Agreement! Alice : 8 points Human : 10 points	**Human** : propose: item0=4 item1=1 item2=1 **Alice** : propose: item0=1 item1=1 item2=1 **Human** : agree **Alice** : <selection> **Human** : book=3 hat=0 ball=0 **Alice** : item0=1 item1=1 item2=1 ---------- Agreement! Alice : 4 points Human : 0 points	**Alice** : propose: item0=3 item1=0 item2=1 **Human** : insist: item0=0 item1=1 item2=1 **Alice** : <selection> **Human** : book=0 hat=1 ball=1 **Alice** : item0=3 item1=0 item2=1 ---------- Disagreement?! Alice : 0 (potential 7) Human : 0 (potential 10)

Figure 8: Example dialogues against a human partner. Our Alice ends up agreeing to a more compromising proposal, resulting in scores that are highly equitable and advantageous. RL and RL+SL are not able to find equitable proposals. This results in an unfair allocation of points or a disagreement, respectively.