

# Targeted Data Acquisition for Evolving Negotiation Agents

Minae Kwon, Siddharth Karamcheti,  
Mariano-Florentino Cuéllar, Dorsa Sadigh



*Negotiation is a bargaining process  
by which a joint decision is made by two parties*

*Negotiation is a bargaining process  
by which a joint decision is made by two parties*



Lawyers in court

*Negotiation is a bargaining process  
by which a joint decision is made by two parties*



Lawyers in court



Employee negotiating salary

*Negotiation is a bargaining process  
by which a joint decision is made by two parties*



Lawyers in court



Employee negotiating salary



2021 UN climate change  
conference



Nước Tương  
Soya sauce

# Desiderata



# Desiderata



(1) Agents that maximize their self-interest



# Desiderata



- (1) Agents that maximize their self-interest
- (2) Agents that can compromise (find Pareto-optimal solutions)

# Supervised Learning (SL)

Books 📖 Hats 🎩 Balls 🏀

Counts	1	1	3
Points	4	3	1

Negotiations left: 10

Messages

**Alice:**  
Propose  
📖 🎩 🏀 Total Points  
You 1 0 0 4  
Alice 0 1 3 ?

**You:**  
Propose  
📖 🎩 🏀 Total Points  
You 1 1 0 7  
Alice 0 0 3 ?

Action: choose... Your Books: choose... Your Hats: choose... Your Balls: choose... Send

$$L(\theta) = - \sum_{x,c} \sum_t \log p_{\theta}(x_t | x_{0:t-1}, c)$$

$$- \alpha \sum_{x,c} \sum_j \log p_{\theta}(o_j | x_{0:t-1}, c)$$

# Supervised Learning (SL)

Books 📖 Hats 🎩 Balls 🏀

Counts	1	1	3
Points	4	3	1

Negotiations left: 10

Messages

**You:**  
Propose  
📖 🎩 🏀 Total Points  
You 1 1 0 7  
Alice 0 0 3 ?

**Alice:**  
Propose  
📖 🎩 🏀 Total Points  
You 1 0 0 4  
Alice 0 1 3 ?

Action: choose... Your Books: choose... Your Hats: choose... Your Balls: choose... Send

$$L(\theta) = - \sum_{x,c} \sum_t \log p_{\theta}(x_t | x_{0:t-1}, c)$$

*utterances* (pointing to  $x_t$ )  
*context* (pointing to  $c$ )

$$-\alpha \sum_{x,c} \sum_j \log p_{\theta}(o_j | x_{0:t-1}, c)$$

# Supervised Learning (SL)

The screenshot shows a negotiation interface. At the top right, a table displays counts and points for Books, Hats, and Balls. Below it, a 'Messages' section shows two proposals: one from Alice and one from 'You'. At the bottom, there is an 'Action' section with dropdown menus for 'Your Books', 'Your Hats', and 'Your Balls', and a 'Send' button.

	Books	Hats	Balls
Counts	1	1	3
Points	4	3	1

Messages:

**Alice:**  
Propose  
Total Points  
You: 1 0 0 4  
Alice: 0 1 3 ?

**You:**  
Propose  
Total Points  
You: 1 1 0 7  
Alice: 0 0 3 ?

Control Panel:

Action: choose...  
Your Books: choose...  
Your Hats: choose...  
Your Balls: choose...  
Send

$$L(\theta) = - \underbrace{\sum_{x,c} \sum_t \log p_{\theta}(x_t | x_{0:t-1}, c)}_{\text{utterance prediction loss}}$$

*utterances* (pointing to  $x_t$ )  
*context* (pointing to  $c$ )

$$-\alpha \sum_{x,c} \sum_j \log p_{\theta}(o_j | x_{0:t-1}, c)$$

# Supervised Learning (SL)

The screenshot shows a negotiation interface. At the top right, a table displays counts and points for Books, Hats, and Balls. Below it, a 'Messages' section shows two messages: one from Alice and one from 'You'. At the bottom, there is an 'Action' section with dropdown menus for 'Your Books', 'Your Hats', and 'Your Balls', and a 'Send' button.

	Books	Hats	Balls
Counts	1	1	3
Points	4	3	1

Messages:

**Alice:**  
Propose  
Total Points  
You: 1 0 0 4  
Alice: 0 1 3 ?

**You:**  
Propose  
Total Points  
You: 1 1 0 7  
Alice: 0 0 3 ?

Control Panel:

Action: choose...  
Your Books: choose...  
Your Hats: choose...  
Your Balls: choose...  
Send

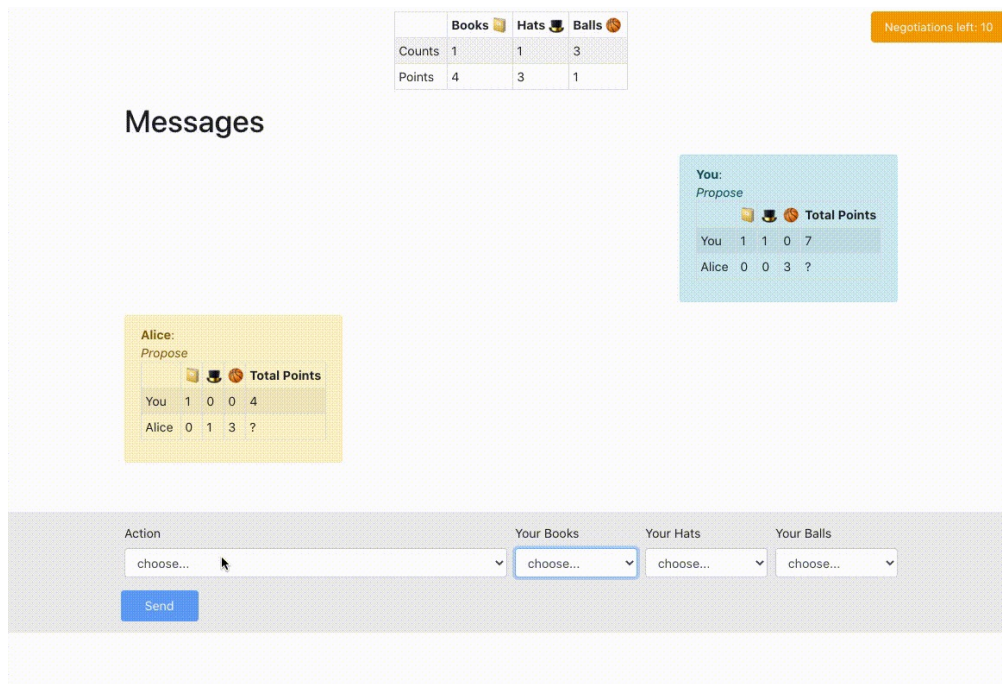
$$L(\theta) = - \sum_{x,c} \sum_t \log p_{\theta}(x_t | x_{0:t-1}, c)$$

*utterance prediction loss*

$$-\alpha \sum_{x,c} \sum_j \log p_{\theta}(o_j | x_{0:t-1}, c)$$

*final split prediction loss*

# Supervised Learning (SL)



$$L(\theta) = - \sum_{x,c} \sum_t \log p_{\theta}(x_t | x_{0:t-1}, c)$$

*utterance prediction loss*

$$-\alpha \sum_{x,c} \sum_j \log p_{\theta}(o_j | x_{0:t-1}, c)$$

*final split prediction loss*

***Relationship to dataset:*** bias inherited from dataset

# Reinforcement Learning (RL)

*Negotiation*



*Bob*



*Alice*

# Reinforcement Learning (RL)

*Negotiation*



*Bob*  
*(fixed)*



*Alice*



# Reinforcement Learning (RL)

*Negotiation*



*Bob*  
*(fixed)*



*Alice*  
*(learning)*

# Reinforcement Learning (RL)

*Negotiation*



*Bob*  
*(fixed)*

propose(0 buns, 2 puffs, 1 roll)

⋮

end

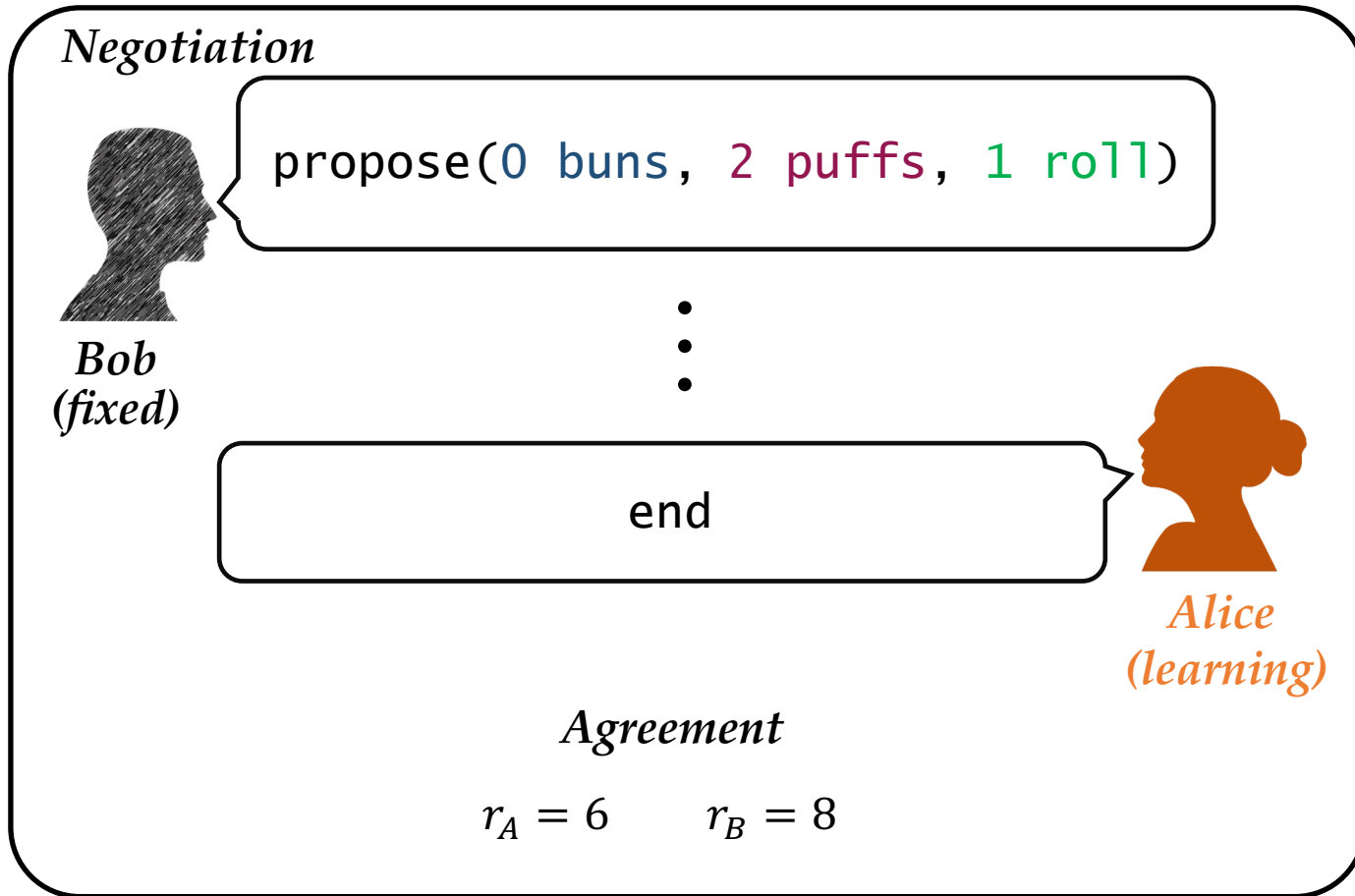


*Alice*  
*(learning)*

*Agreement*

$r_A = 6$      $r_B = 8$

# Reinforcement Learning (RL)



For  $x_t \in X^A$

$$R_A(x_t) = \gamma^{T-t} (r_A - \mu_n)$$

*Alice's utterances*

*running mean*

# Reinforcement Learning (RL)

*Negotiation*



*Bob  
(fixed)*

propose(0 buns, 2 puffs, 1 roll)

For  $x_t \in X^A$

$$R_A(x_t) = \gamma^{T-t} (r_A - \mu_n)$$

*Alice's utterances*

*running mean*

# Reinforcement Learning (RL)

*Negotiation*



propose(0 buns, 2 puffs, 1 roll)

*Bob  
(fixed)*

insist(1 bun, 2 puffs, 2 rolls)



*Alice  
(learning)*

For  $x_t \in X^A$

$$R_A(x_t) = \gamma^{T-t} (r_A - \mu_n)$$

*Alice's utterances*

*running mean*

# Reinforce

RL)

## Negotiation



Bob  
(fixed)

propose(0 k

insist(1 bun,

RL

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 : <selection>  
 Alice : book=1 hat=3 ball=1  
 Bob : book=1 hat=2 ball=0

-----  
 Disagreement?!  
 Alice : 0 (potential 10)  
 Bob : 0 (potential 7)

For  $x_t \in X^A$

$$R_A(x_t) = \gamma^{T-t} (r_A - \mu_n)$$

*Alice's utterances*

*running mean*

# Reinforce

RL)

## Negotiation



Bob  
(fixed)

propose(0 k

insist(1 bun,

RL

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 : <selection>  
 Alice : book=1 hat=3 ball=1  
 Bob : book=1 hat=2 ball=0

-----  
 Disagreement?!  
 Alice : 0 (potential 10)  
 Bob : 0 (potential 7)

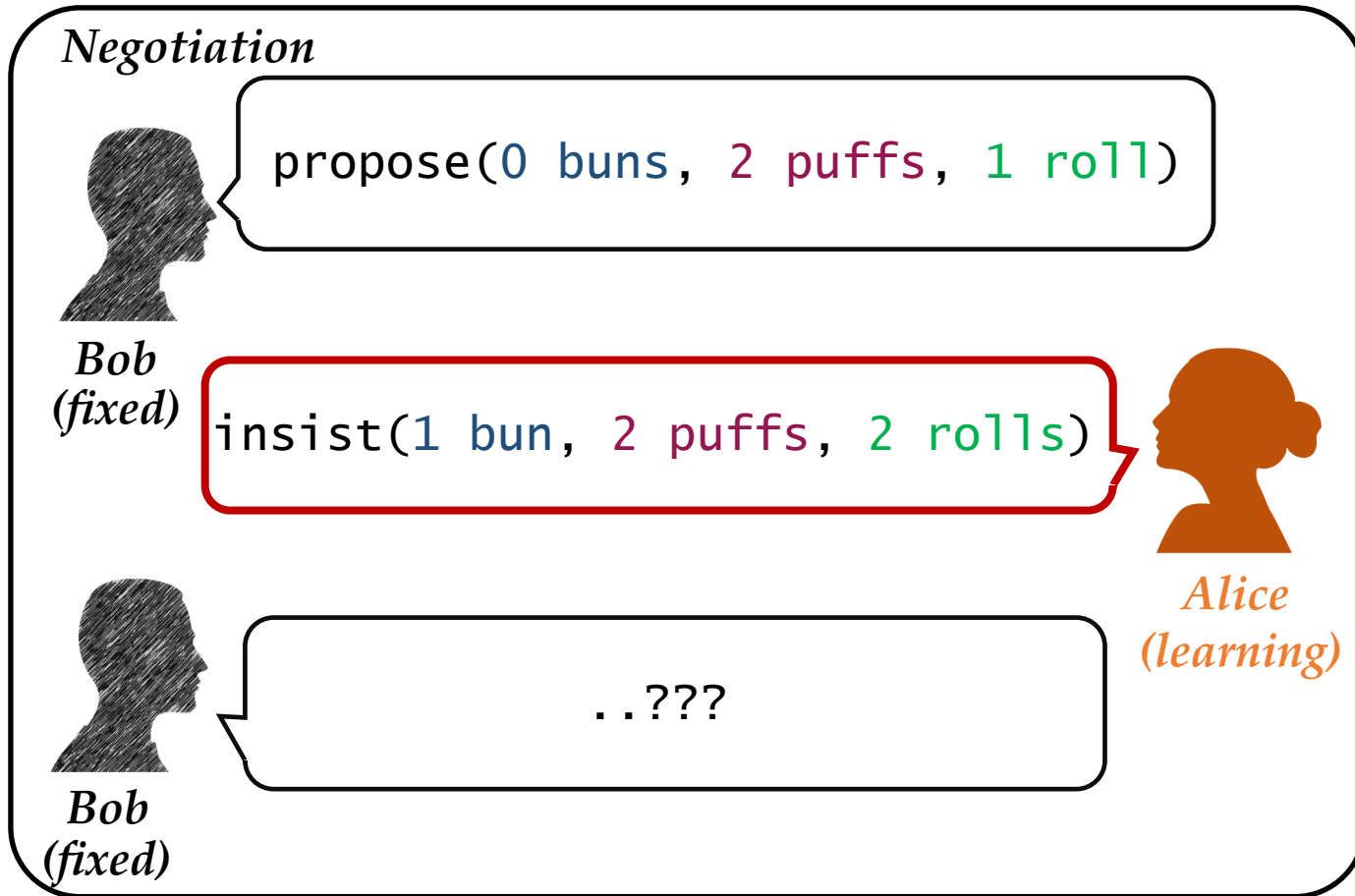
Alice's utterances

For  $x_t \in X^A$

$$R_A(x_t) = \gamma^{T-t} (r_A - \mu_n)$$

running mean

# Reinforcement Learning (RL)



For  $x_t \in X^A$

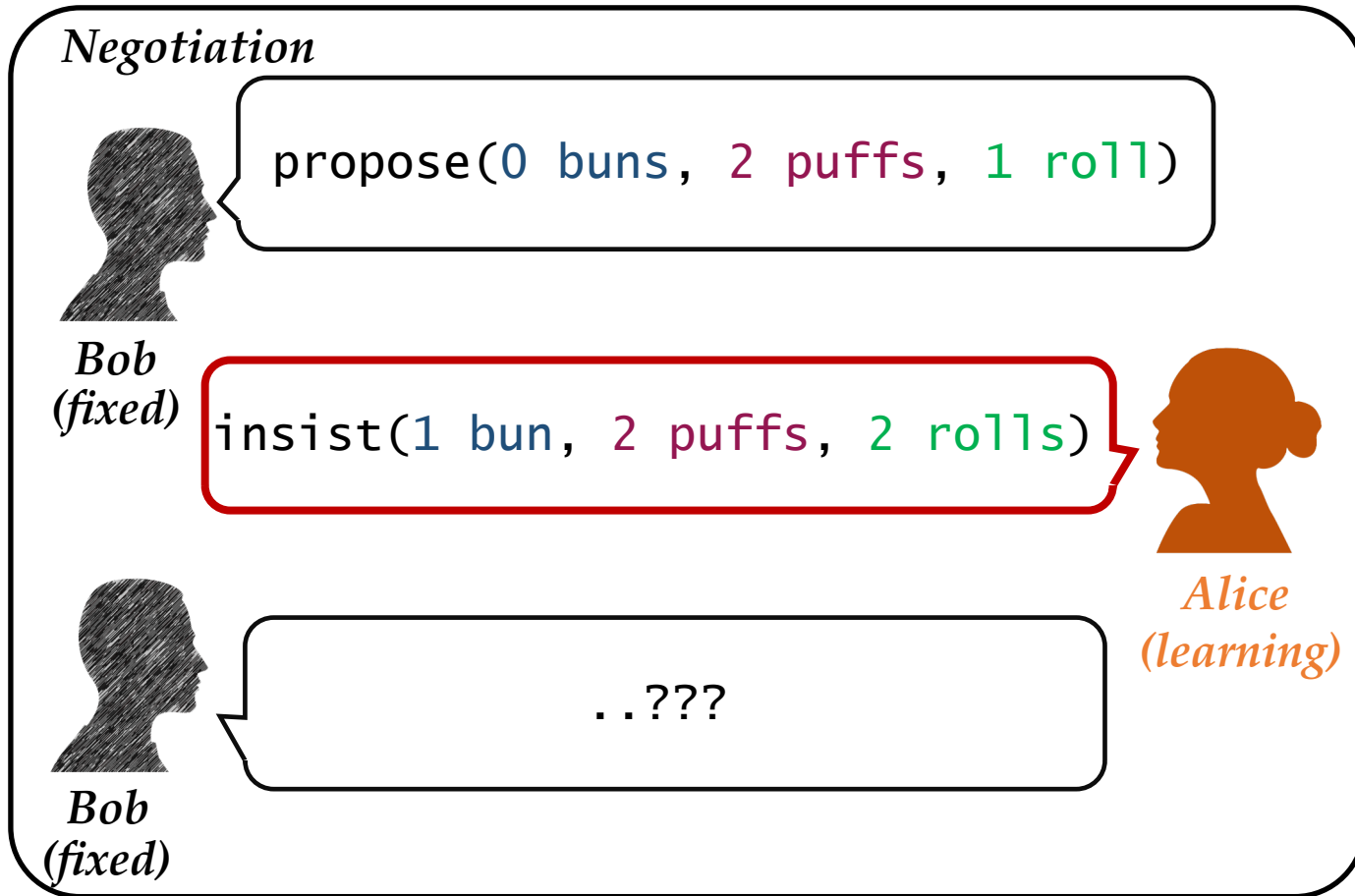
$$R_A(x_t) = \gamma^{T-t} (r_A - \mu_n)$$

*Alice's utterances*

*running mean*



# Reinforcement Learning (RL)



For  $x_t \in X^A$

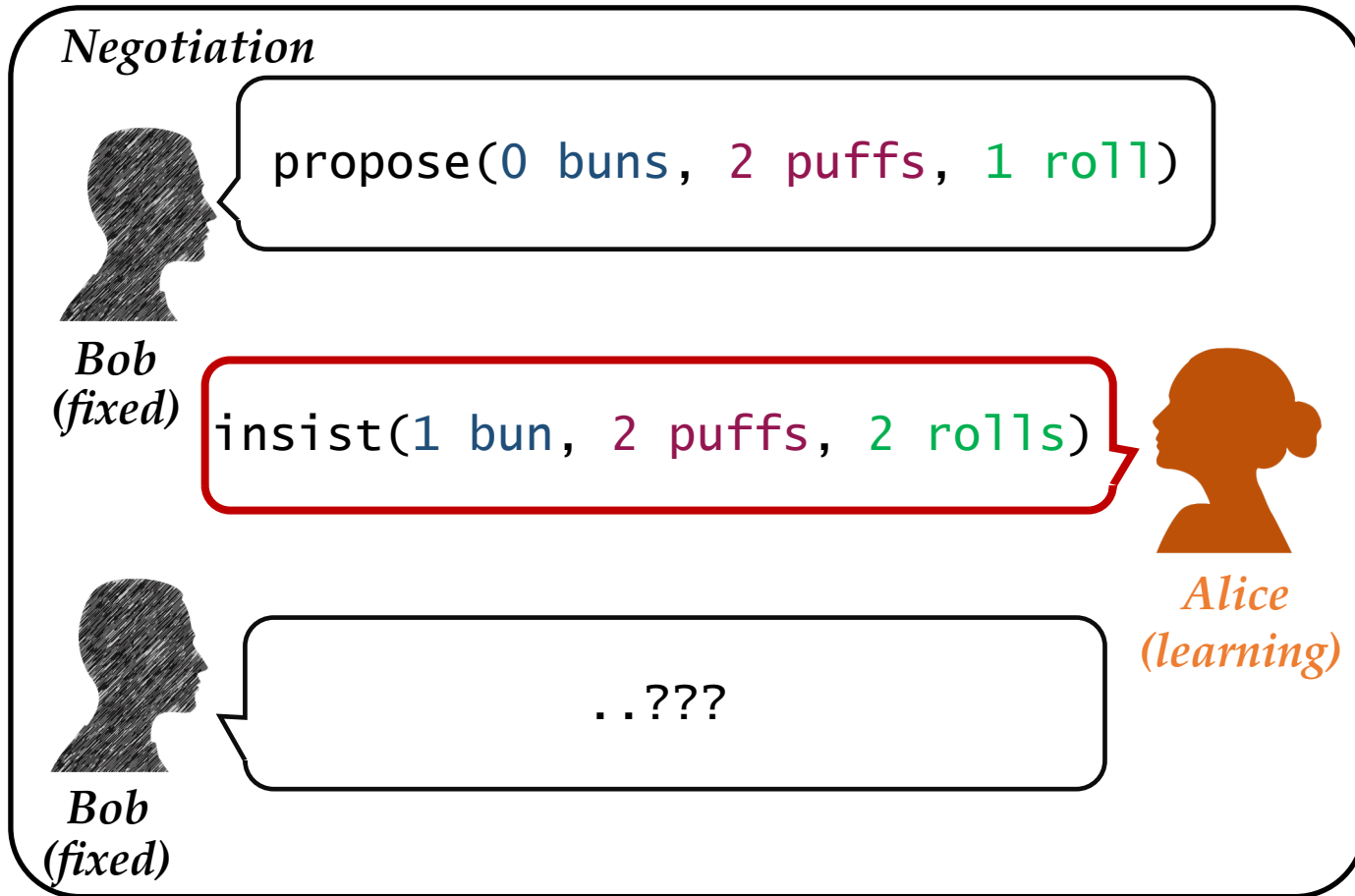
$$R_A(x_t) = \gamma^{T-t} (r_A - \mu_n)$$

*Alice's utterances*

*running mean*

***Relationship to dataset:*** Alice inherits dataset biases through Bob

# Reinforcement Learning (RL)



For  $x_t \in X^A$

$$R_A(x_t) = \gamma^{T-t} (r_A - \mu_n)$$

*Alice's utterances*

*running mean*

***Relationship to dataset:*** Alice inherits dataset biases through Bob

# Mixed RL, SL (RL+SL)

Interleave SL training every nth timestep

- n=1: RL, SL, RL, SL ...
- n=2: RL, RL, SL, RL, RL, SL ...

# Mixed RL, SL (RL+SL)

Interleave SL training every  $n$ th timestep

- $n=1$ : RL, SL, RL, SL ...
- $n=2$ : RL, RL, SL, RL, RL, SL ...

*Relationship to dataset:* same as SL, bias inherited from dataset

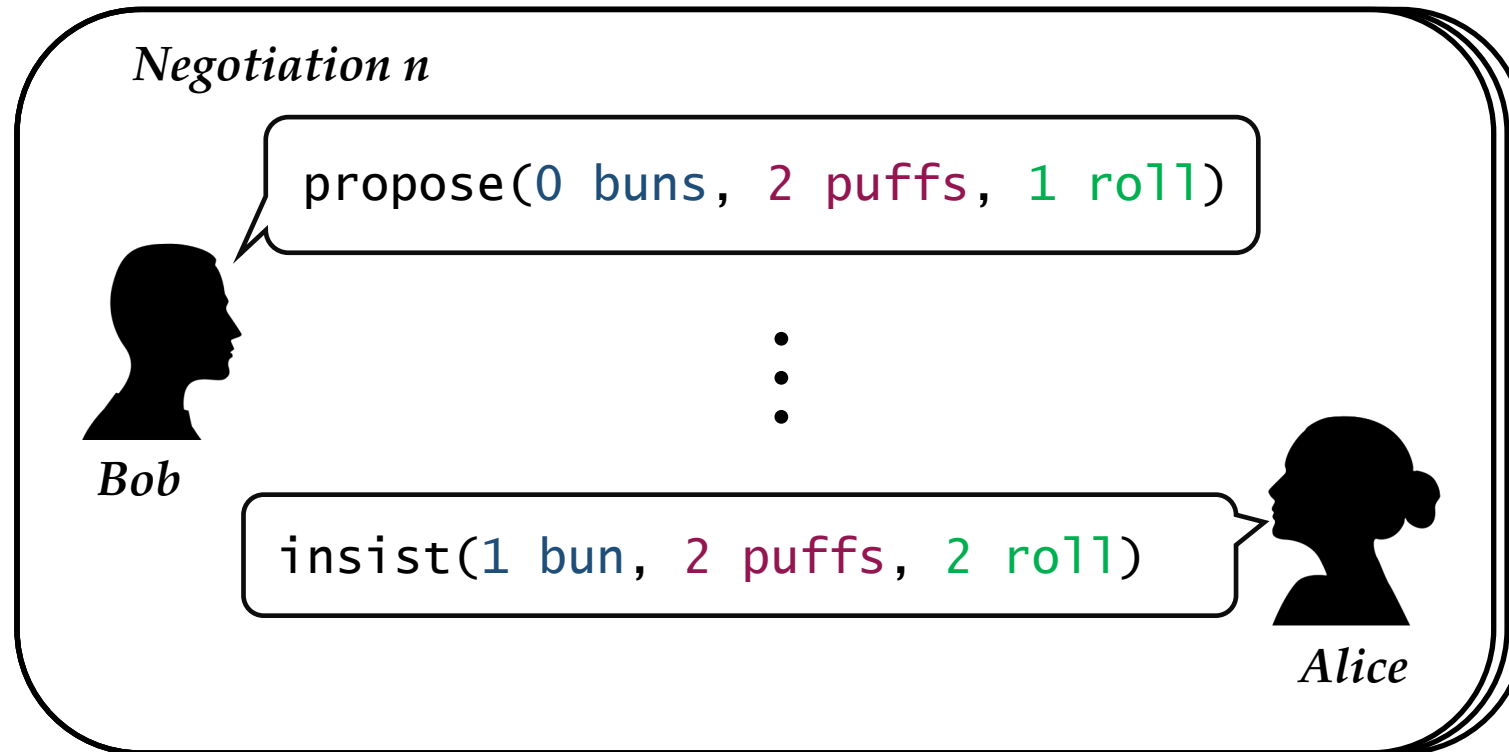
*Problem:* Low-quality, static datasets!

*Problem: Low-quality, static datasets!*

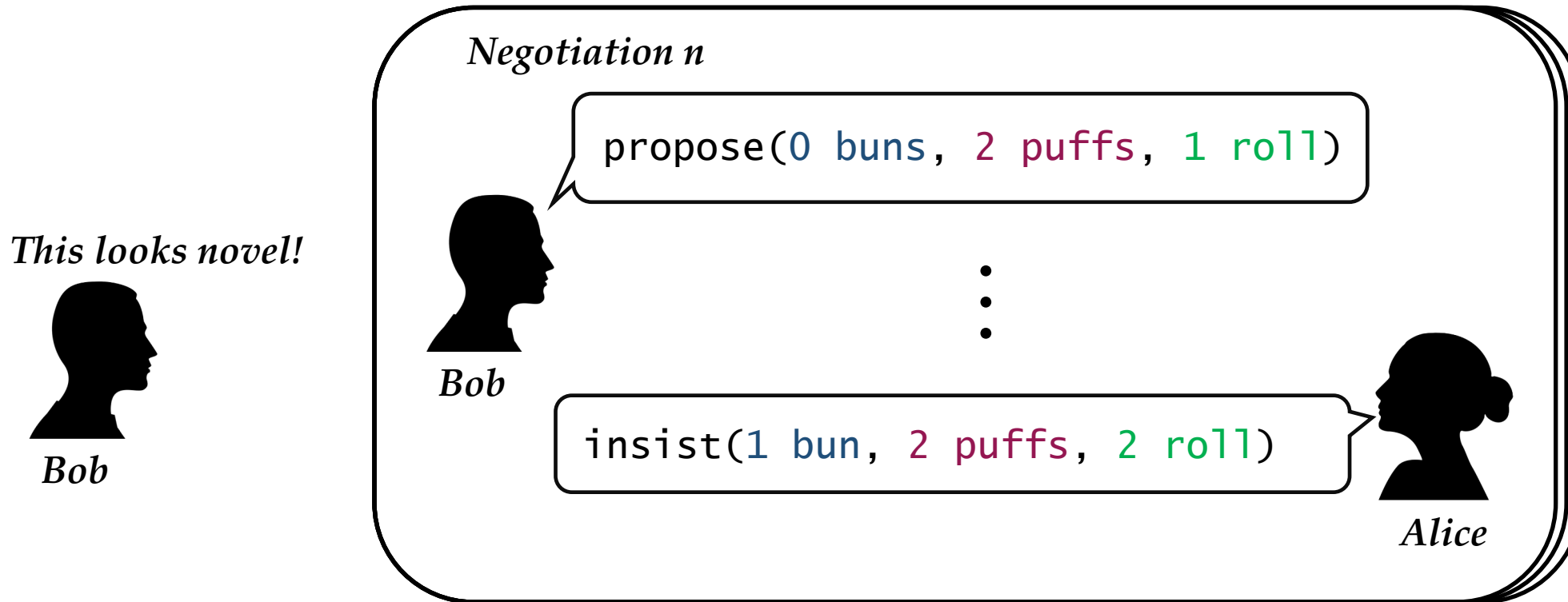
*Key Insight:*

Continually improve Bob with expert data!

# Targeted Data Acquisition Framework



# Targeted Data Acquisition Framework

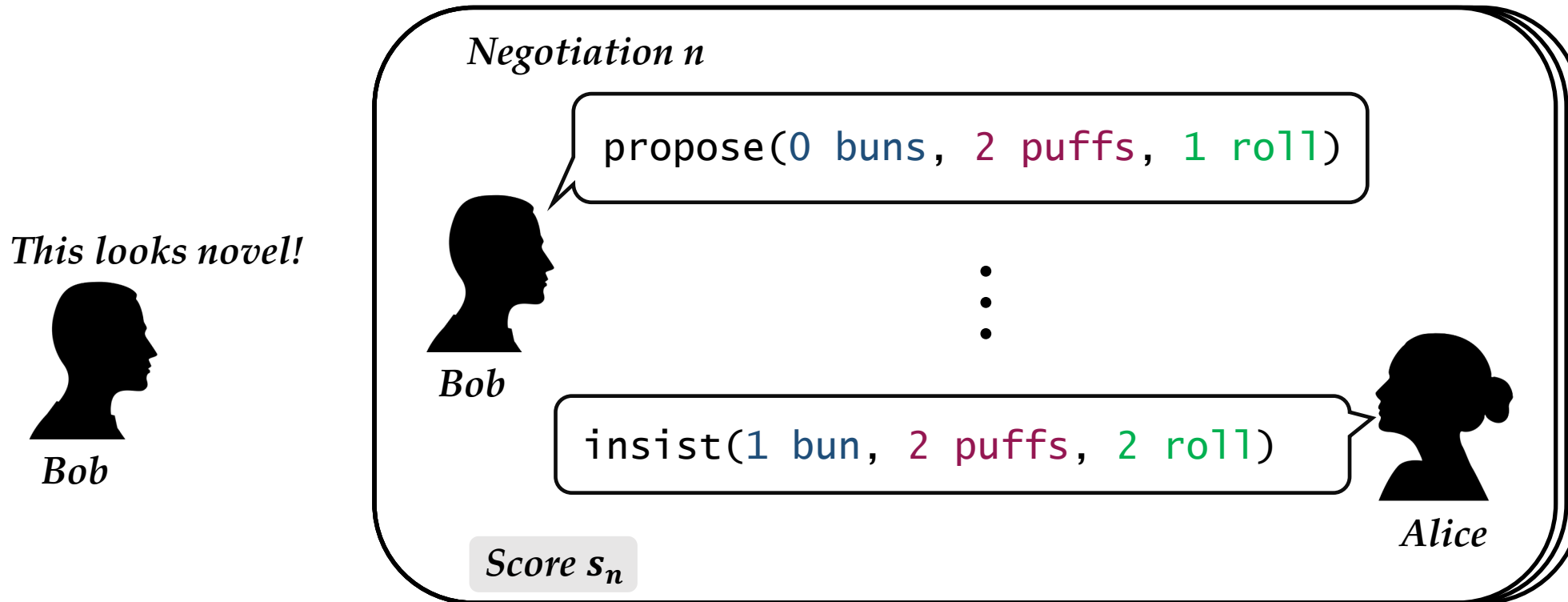


*Novelty score:*

$$s_n = \min_{x_t \in X^A} \log p_\theta(x_t | x_{0:t-1}, c^A)$$



# Targeted Data Acquisition Framework

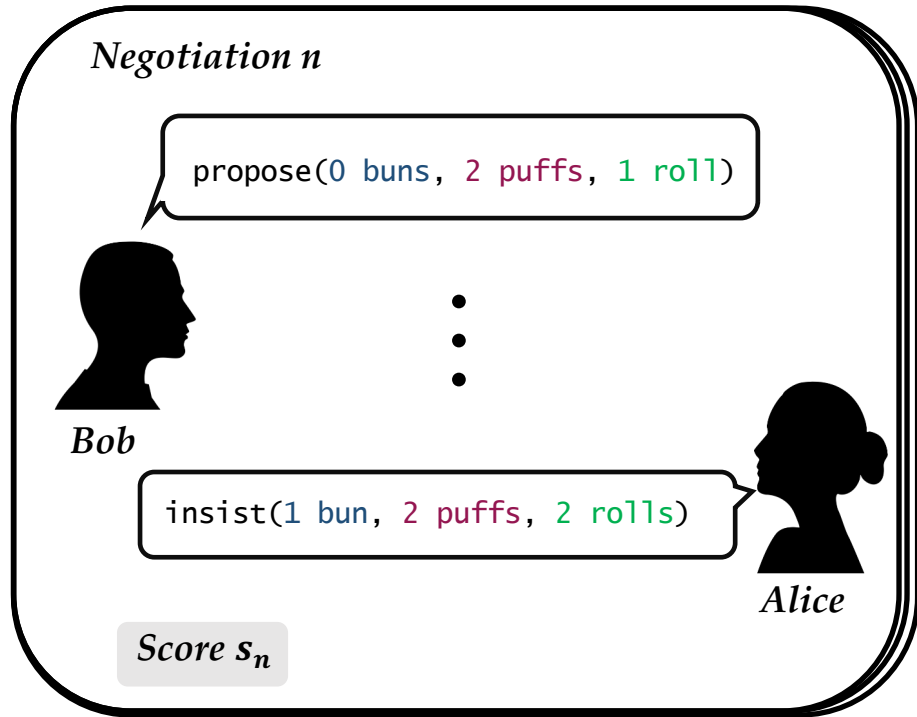


*Novelty score:*

$$s_n = \min_{x_t \in X^A} \log p_\theta(x_t | x_{0:t-1}, c^A)$$

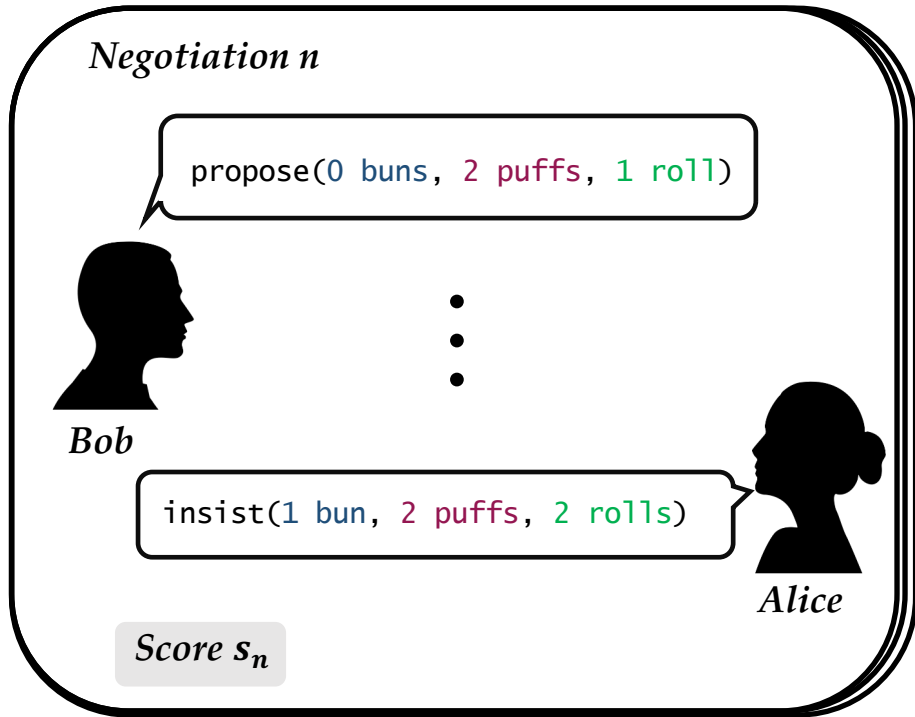
# Targeted Data Acquisition Framework

## *Alice RL Training*



# Targeted Data Acquisition Framework

## *Alice RL Training*

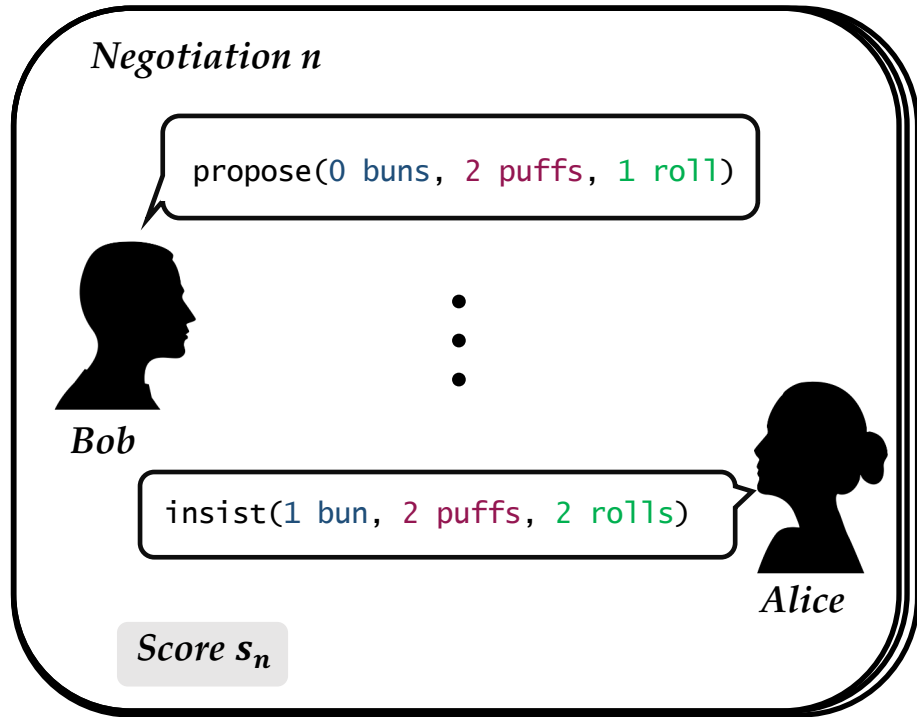


*Pick  $k=500$   
most novel negotiations*



# Targeted Data Acquisition Framework

## *Alice RL Training*

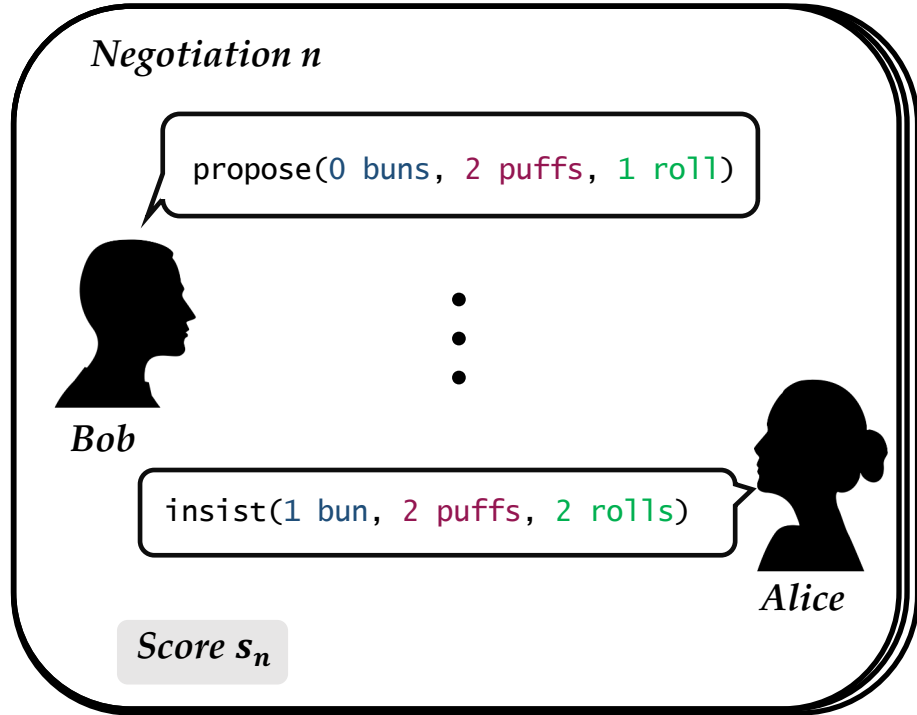


*Pick  $k=500$   
most novel negotiations*



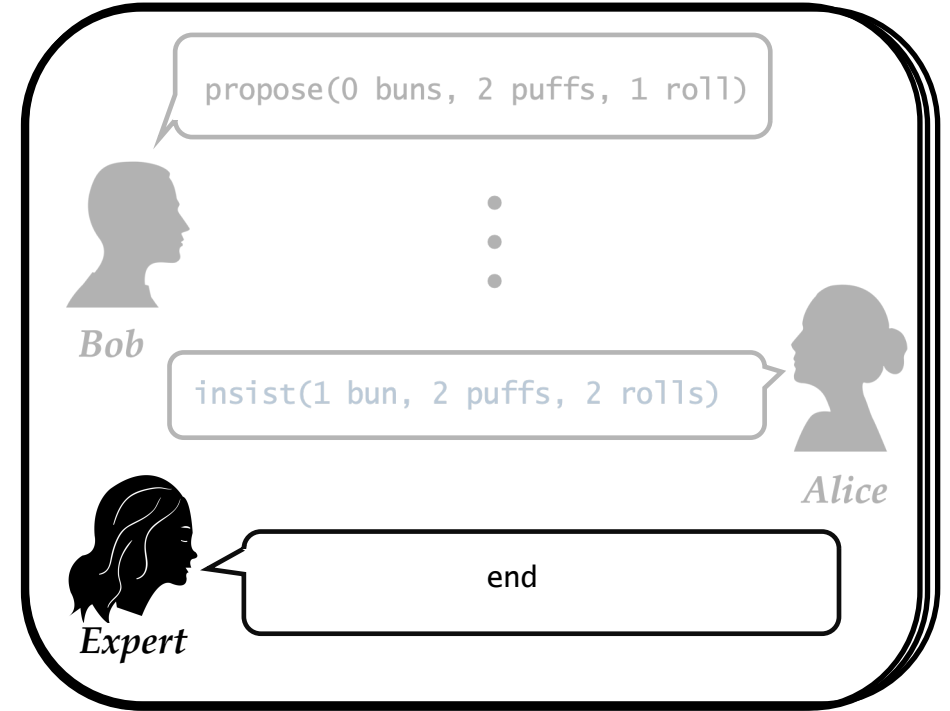
# Targeted Data Acquisition Framework

## Alice RL Training



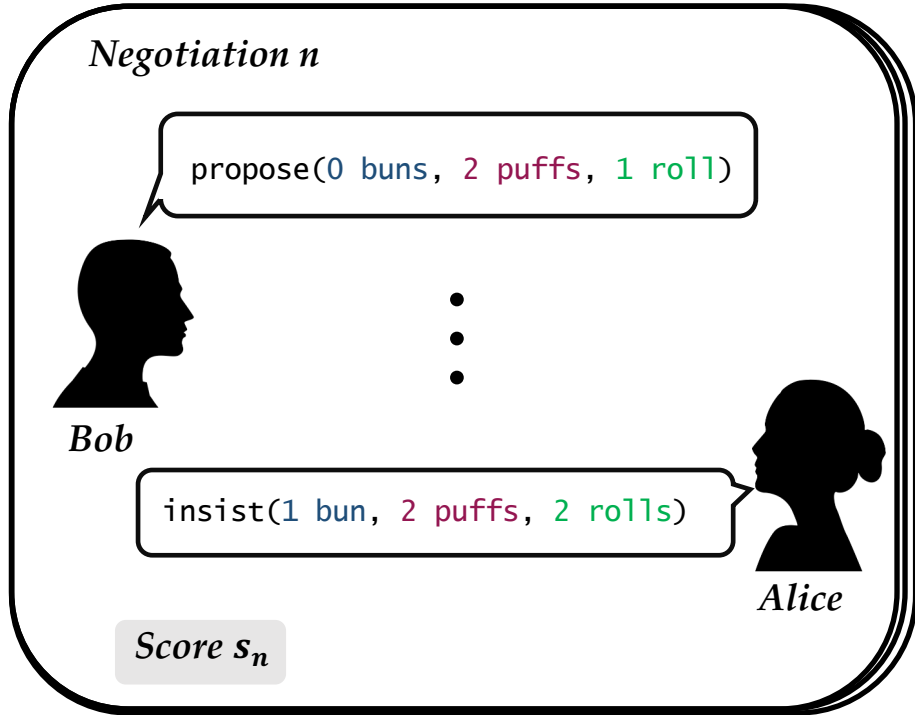
*Pick  $k=500$   
most novel negotiations*

## Expert Annotations



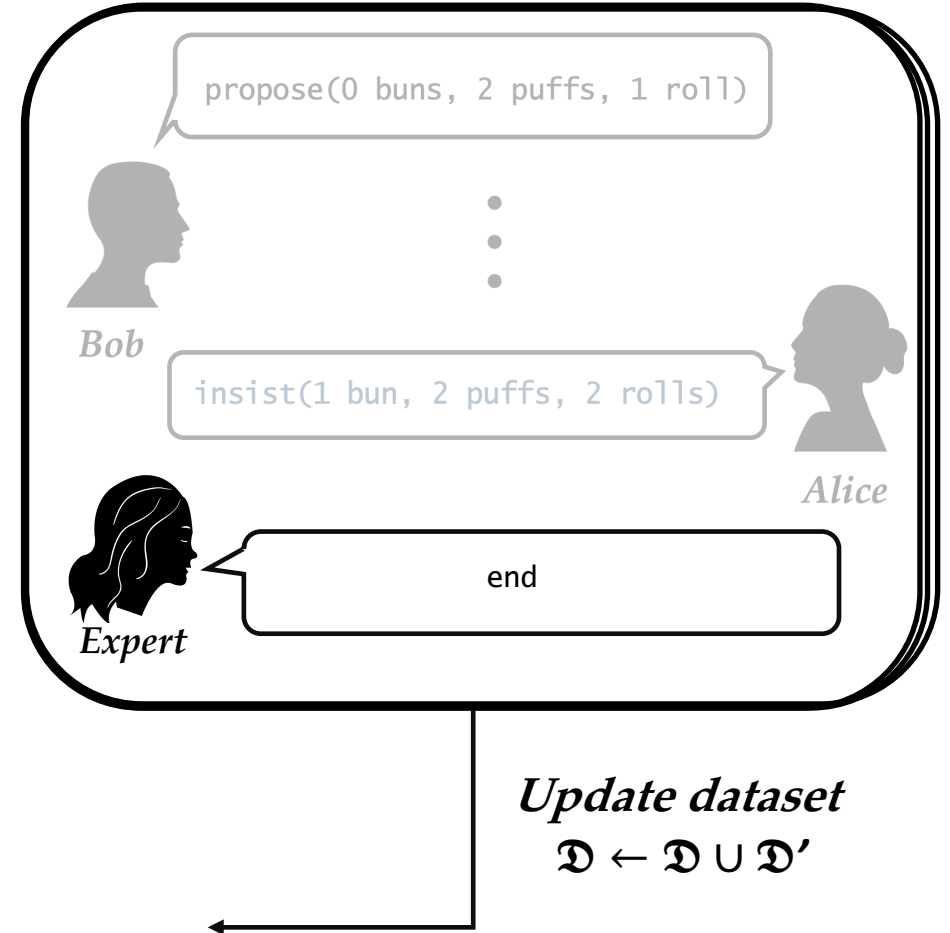
# Targeted Data Acquisition Framework

## Alice RL Training



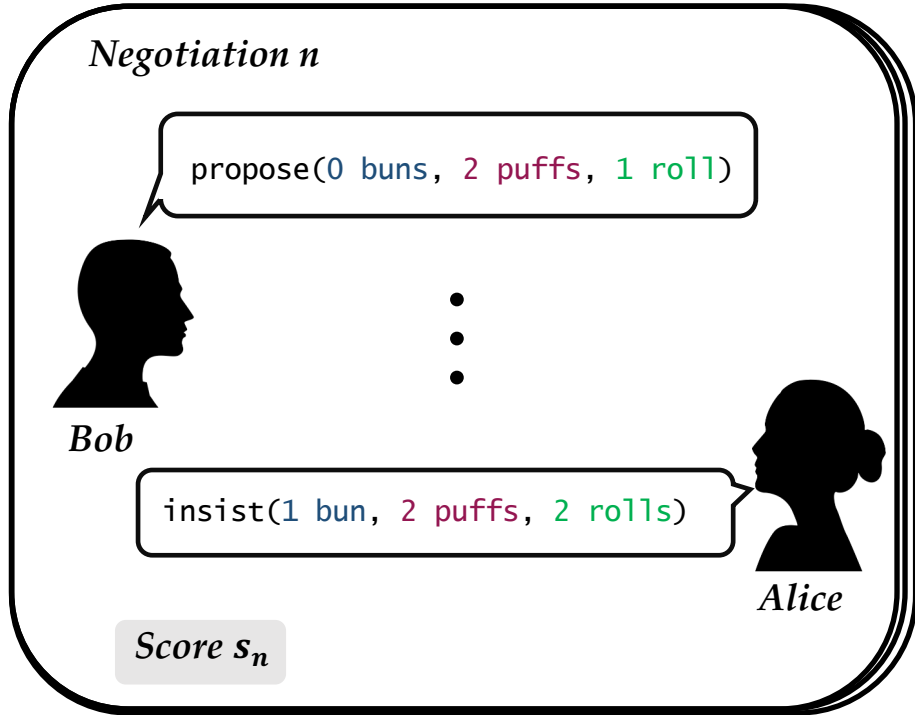
*Pick  $k=500$   
most novel negotiations*

## Expert Annotations



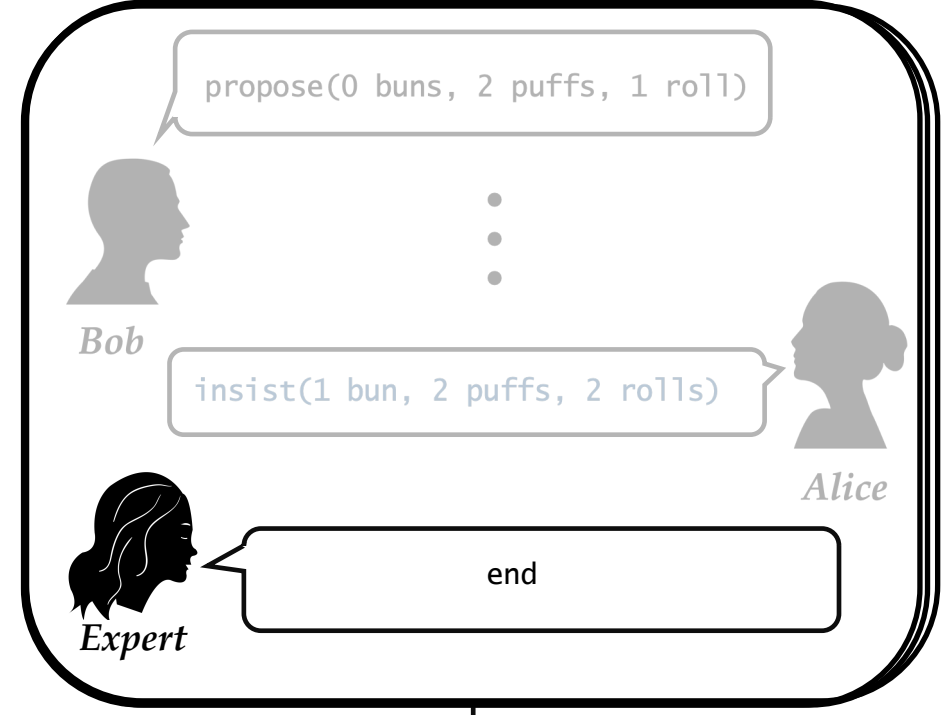
# Targeted Data Acquisition Framework

## Alice RL Training



*Pick  $k=500$   
most novel negotiations*

## Expert Annotations

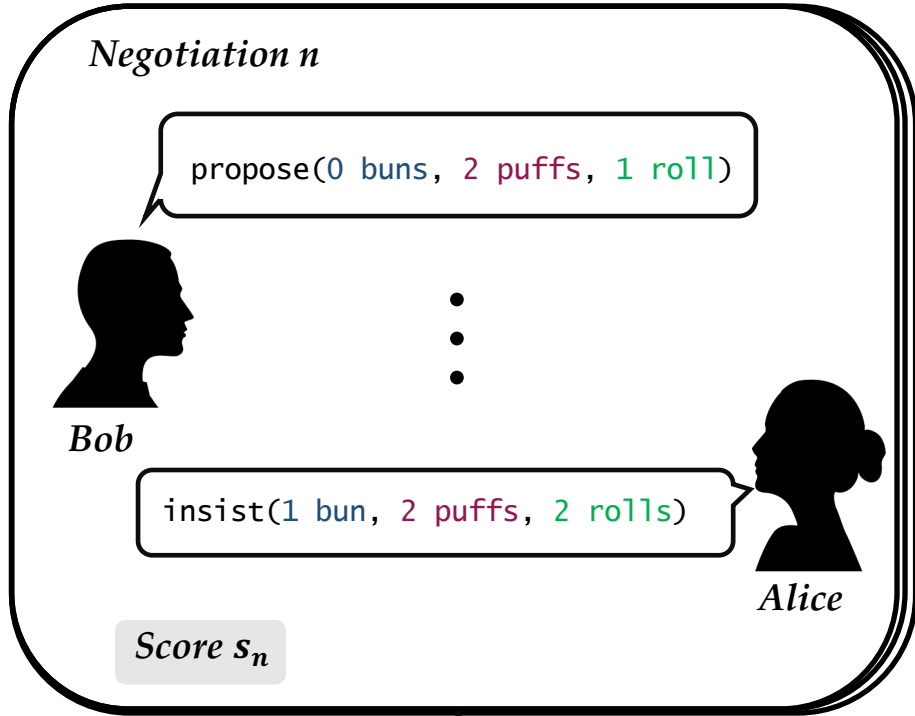


*Update dataset  
 $\mathcal{D} \leftarrow \mathcal{D} \cup \mathcal{D}'$*

*Bob SL Training*

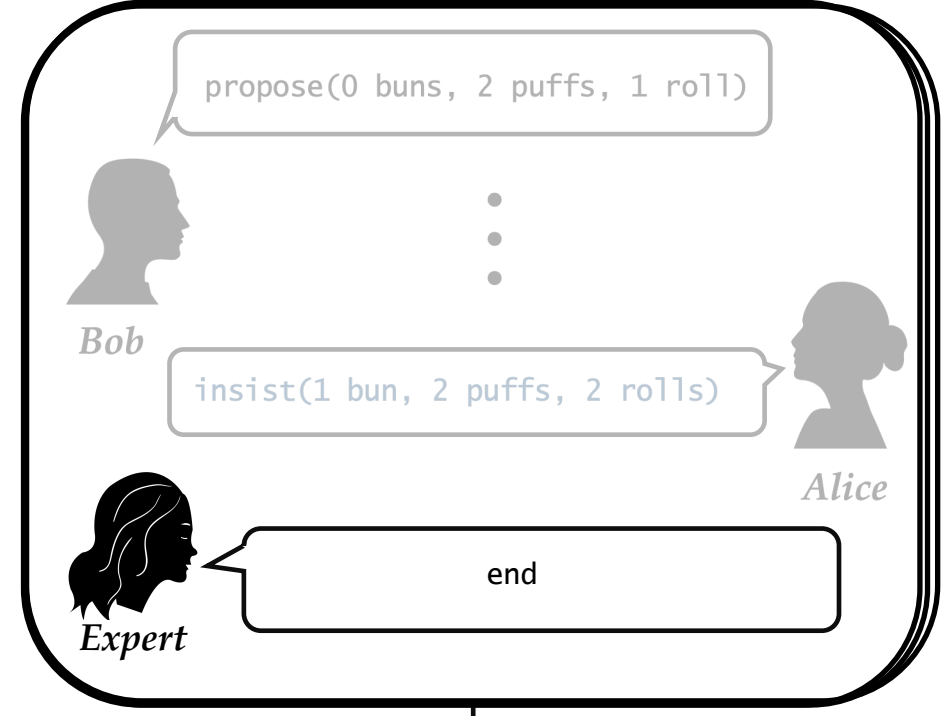
# Targeted Data Acquisition Framework

## Alice RL Training



*Pick  $k=500$   
most novel negotiations*

## Expert Annotations



*Continue  
training Alice*

*Bob SL Training*

*Update dataset  
 $\mathcal{D} \leftarrow \mathcal{D} \cup \mathcal{D}'$*



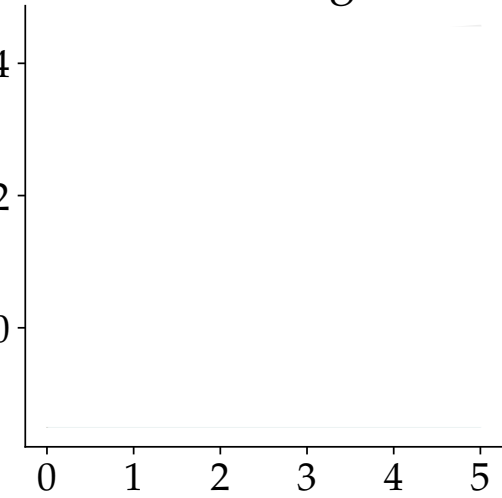
# Evaluation

Can we balance self-interest and Pareto-optimality?

# Results with a Simulated Partner

*(higher is better)*

Advantage



(D1) Self-interest

- Ours
- RL
- RL+SL
- SL

# Results with a Simulated Partner

*(higher is better)*



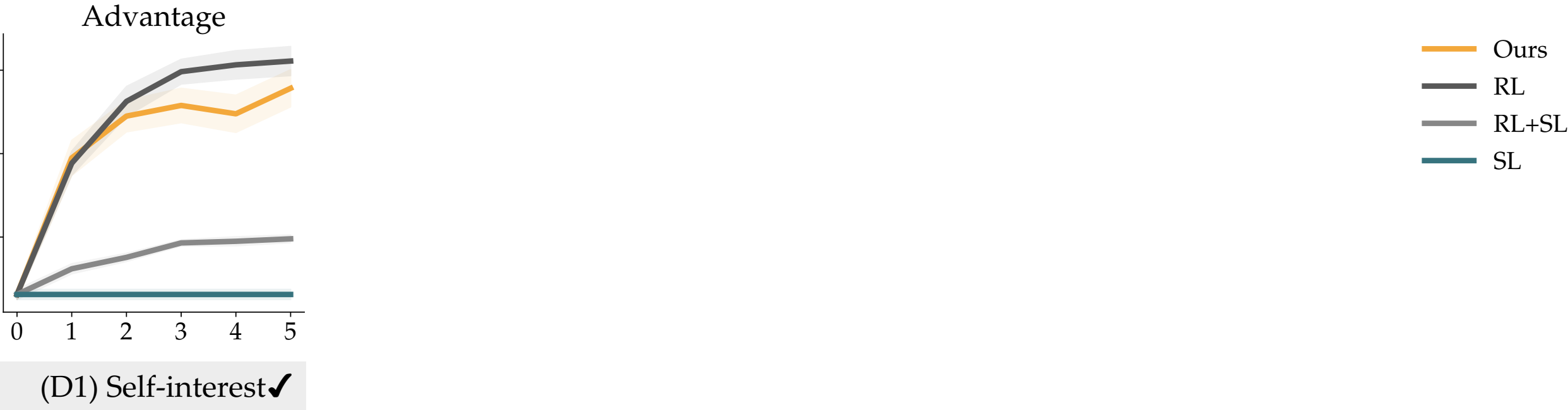
# Results with a Simulated Partner

*(higher is better)*



# Results with a Simulated Partner

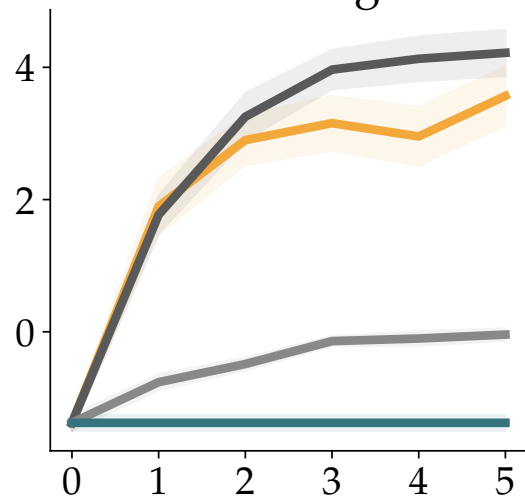
*(higher is better)*



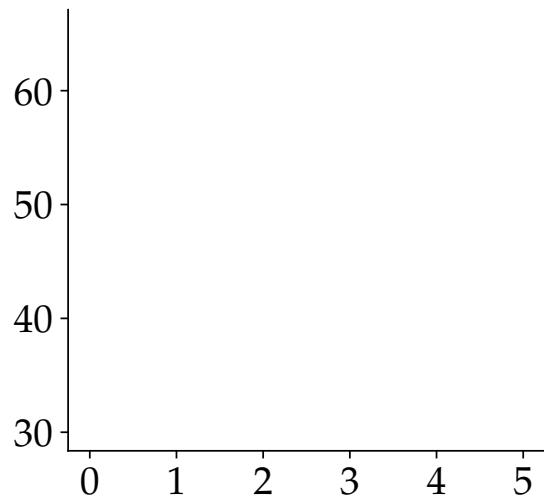
# Results with a Simulated Partner

*(higher is better)*

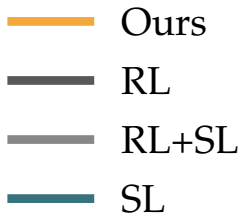
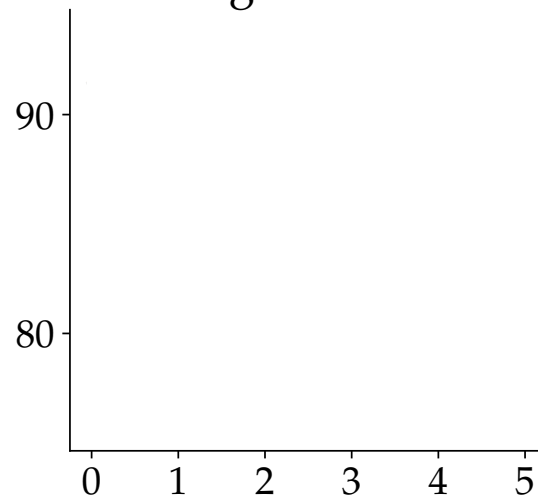
Advantage



Pareto



Agreement



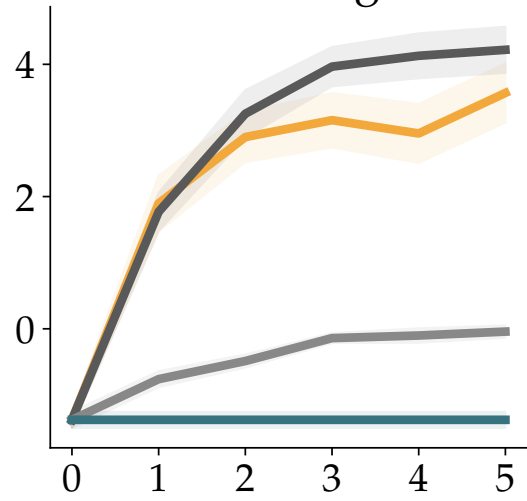
(D1) Self-interest ✓

(D2) Pareto-Optimal

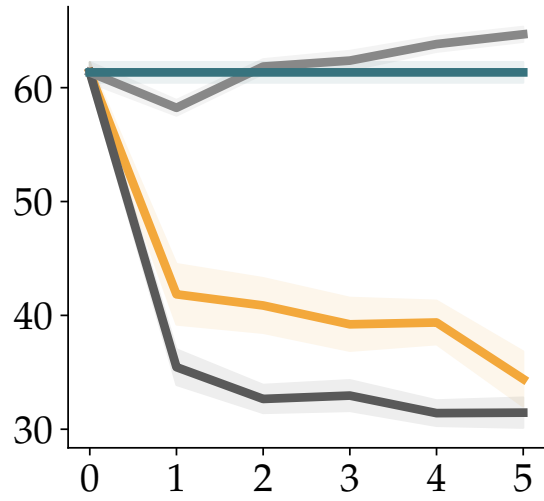
# Results with a Simulated Partner

*(higher is better)*

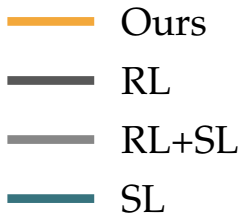
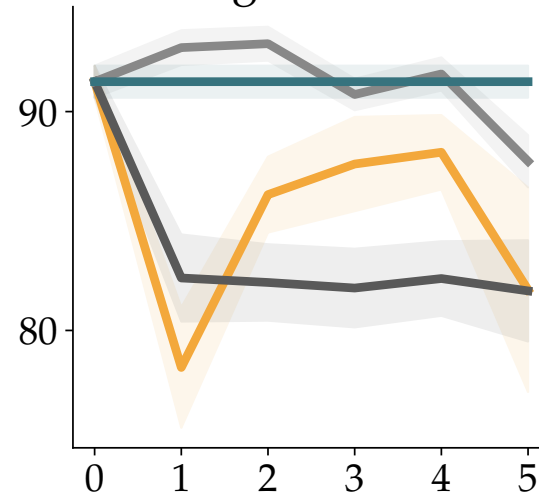
### Advantage



### Pareto



### Agreement



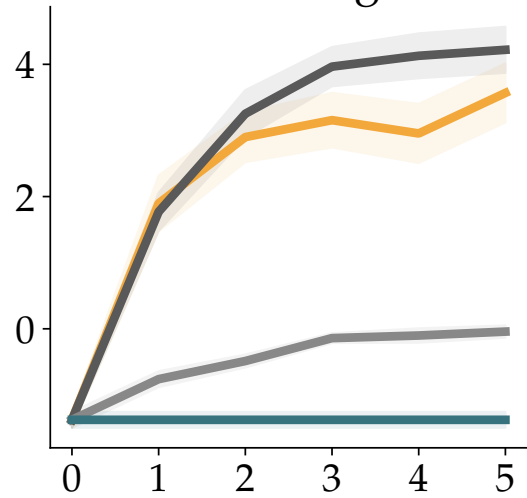
(D1) Self-interest ✓

(D2) Pareto-Optimal

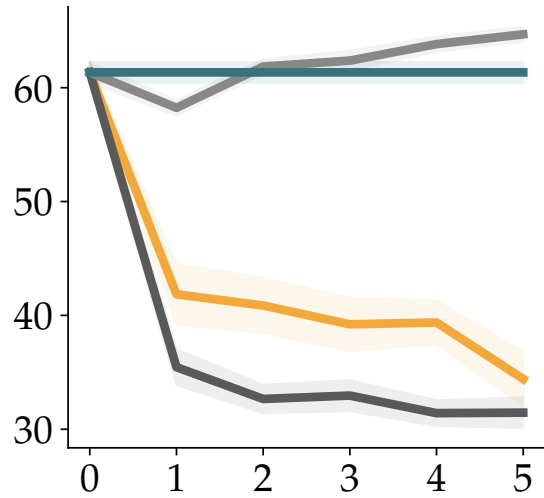
# Results with a Simulated Partner

*(higher is better)*

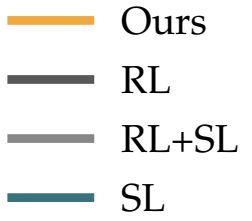
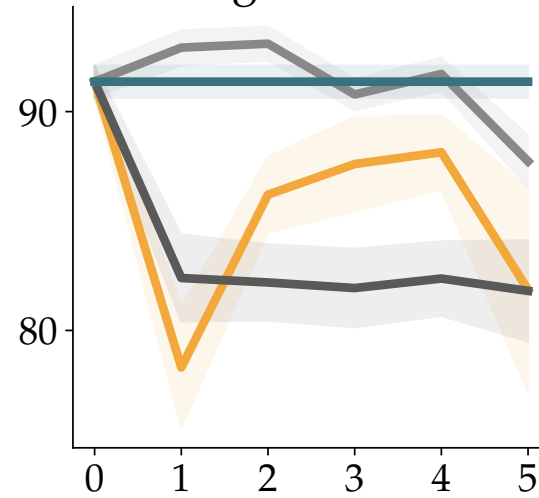
### Advantage



### Pareto



### Agreement



(D1) Self-interest ✓

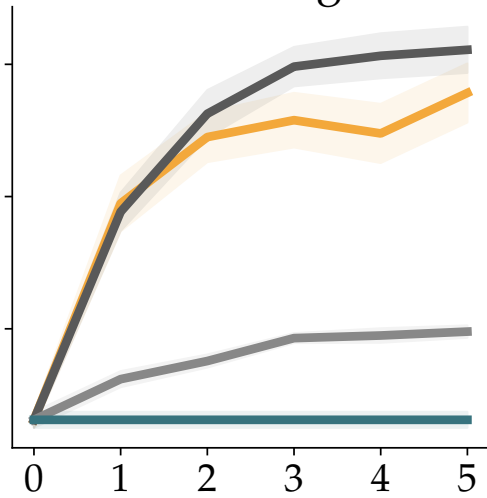
(D2) Pareto-Optimal ✓



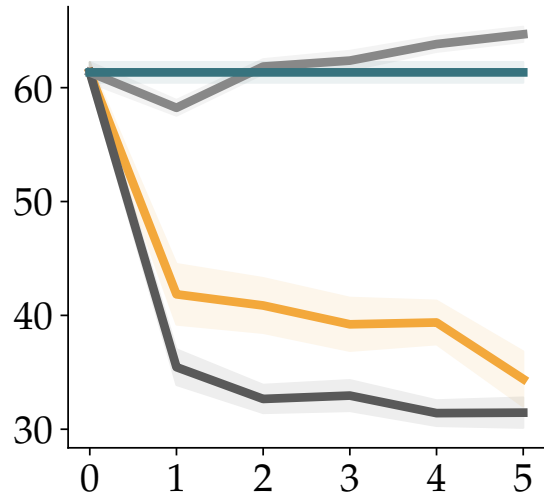
# Results with a Simulated Partner

*(higher is better)*

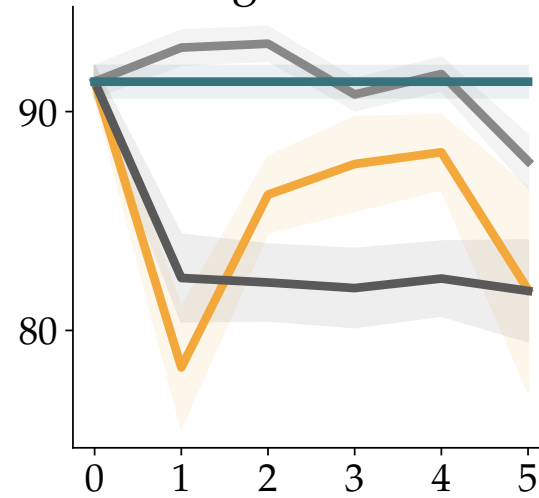
### Advantage



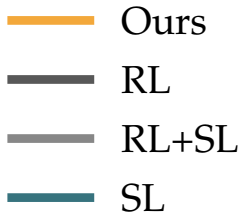
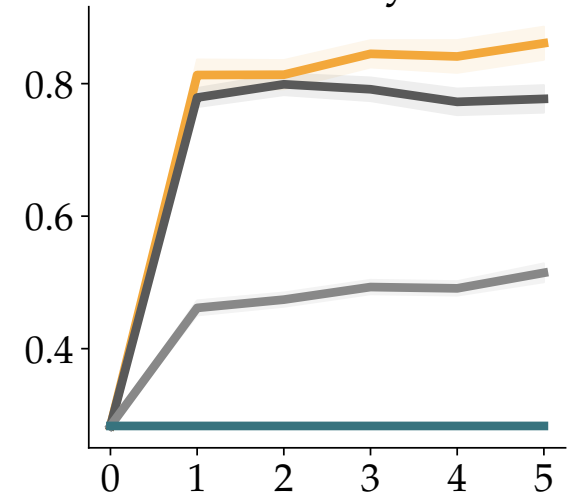
### Pareto



### Agreement



### Novelty

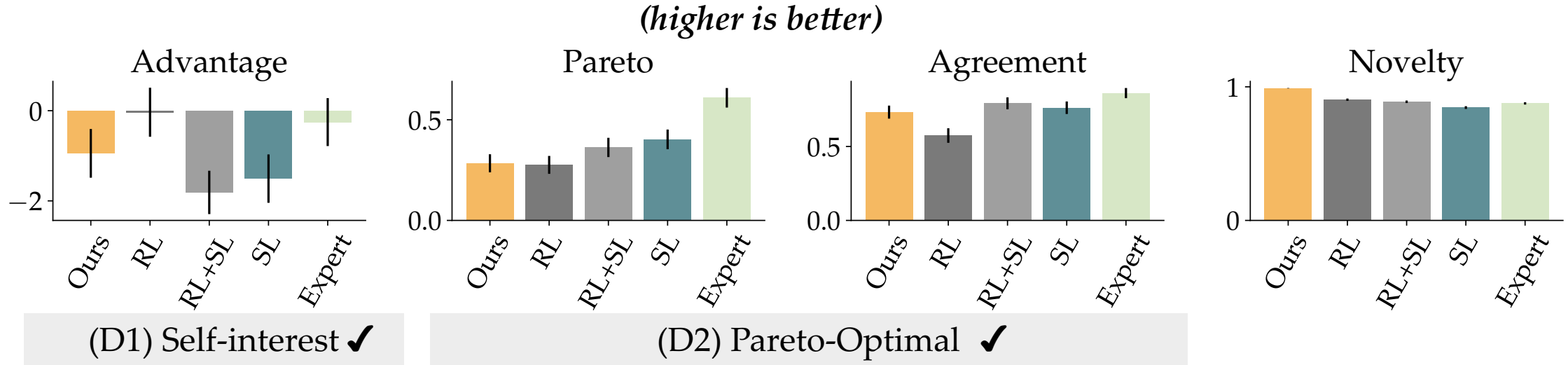


(D1) Self-interest ✓

(D2) Pareto-Optimal ✓

# Results with a Human Partner

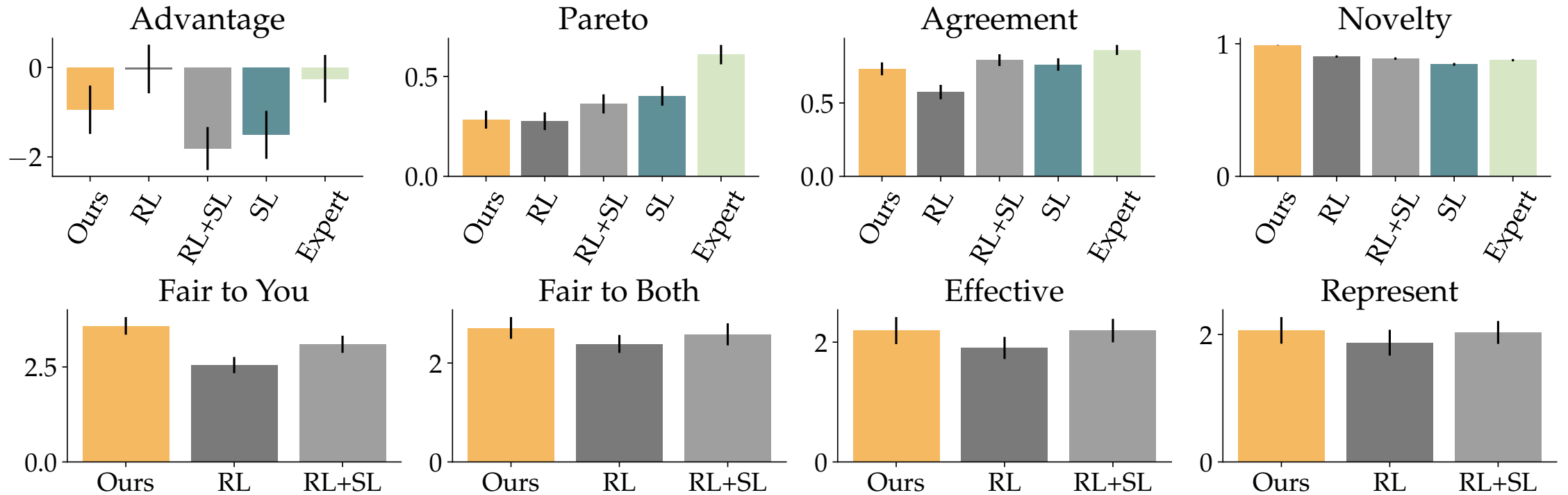
*N=101*



# Results with a Human Partner

*N=101*

*(higher is better)*



# Main Ideas

- Our approach balances self-interest and Pareto-optimality the best.
- This holds true against both simulated and human partners.