Facebook AI Research

# Task-Oriented Dialogues

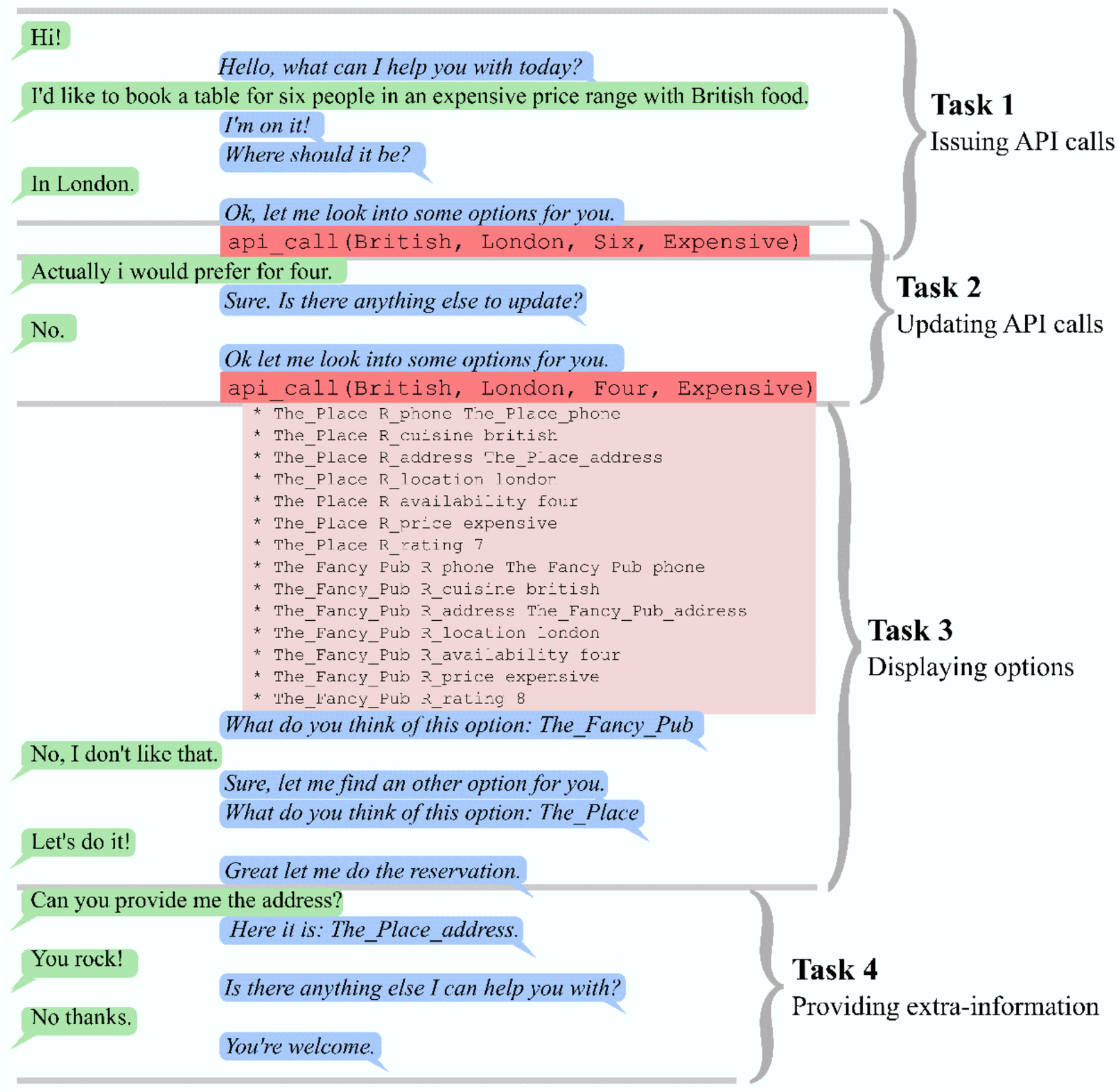
Denis Yarats

# Facebook AI Research

# Overview

- Intro to Task-Oriented Dialogues
- Negotiation Task
  - Dataset
  - Modeling
- Hierarchical Generation

# Task-Oriented Dialogue




**[Bordes et. al. 16']**

# Task-Oriented Dialogue

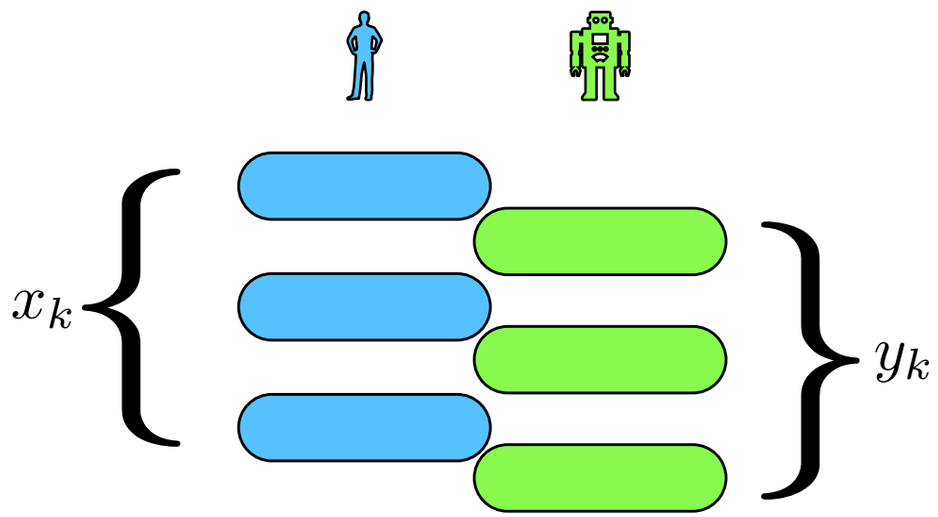
**Formal Definition:**

$D = \{(x_k, y_k)\}_{k=1}^{K}$ - **dialogue**

$s_t = \{x_1, y_1, x_2, y_2, \ldots, x_t\}$ - **state**

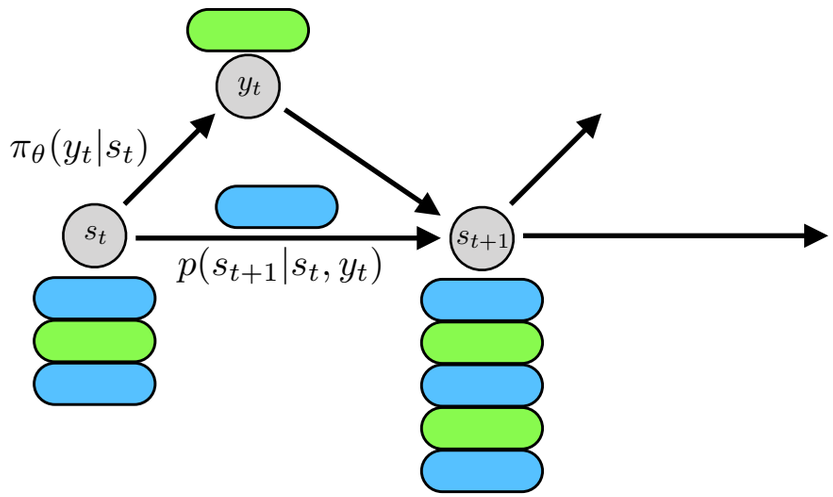
$\mathcal{S} = \{s_t\} \quad \mathcal{A} = \{y_t\}$ - **state and action spaces**

$p : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$ - **transition function**




$\pi_\theta : \mathcal{S} \times \mathcal{A} \to [0, 1]$ - **policy**

$r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ - **reward**



# Task-Oriented Dialogue

**Formal Definition:**

$\tau = \{s_1, y_1, s_2, y_2, \ldots, s_T, y_T\}$ - trajectory of a dialogue

$\pi_\theta(\tau) = p(s_1) \prod_{t=1}^{T} \pi_\theta(y_t|s_t) p(s_{t+1}|s_t, y_t)$ - distribution over trajectory

$J(\theta) = \mathbb{E}_{\tau \sim \pi_\theta}[\sum_t r(s_t, y_t)]$ - cumulative reward

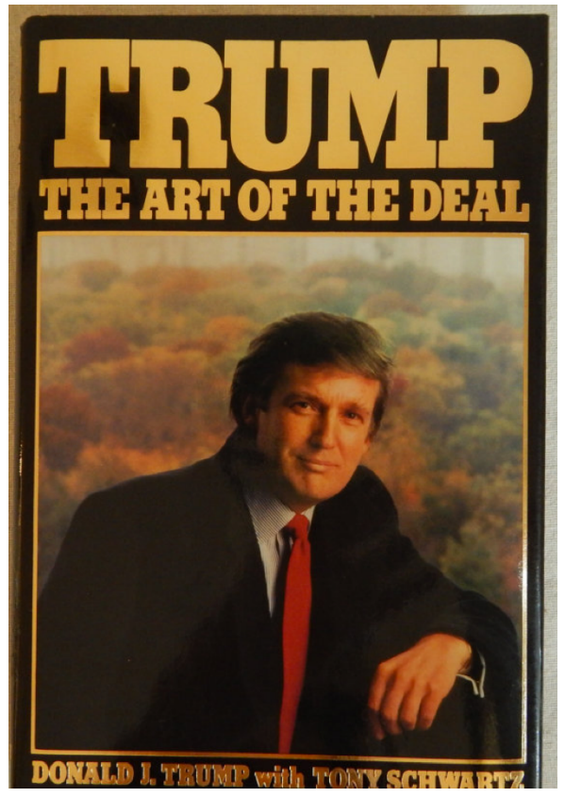
## Goal:

$$\theta^* = \arg\max_\theta J(\theta) = \arg\max_\theta \mathbb{E}_{\tau \sim \pi_\theta}[\sum_t r(s_t, y_t)]$$

# Negotiation

**Why Negotiation?**

- Both **linguistic** and **reasoning** problem
  - **interpret** multiple sentences and **generate** new message
  - plan ahead, make proposals, ask questions, bluff, compromise



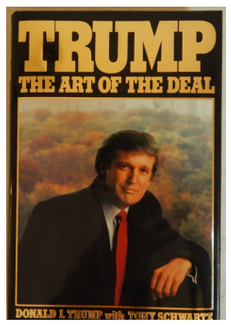
# Negotiation

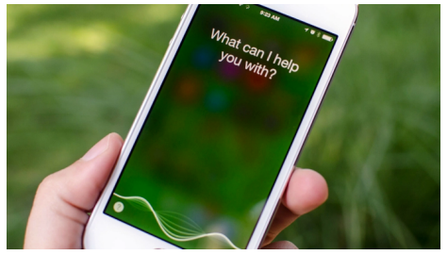
**Why Negotiation?**

- Unlike many task-oriented dialogue problems, **no simple solution** to achieving goal
- But, it is **very easy** to evaluate

Zero-sum / Adversarial

**Negotiation**




Fully Cooperative

# Framework

Both agents given reward function, **can't observe each other's**

Agent 1 Goal

Agent 2 Goal

Dialogue until agreement or disagreement

Both agents **independently** select a deal

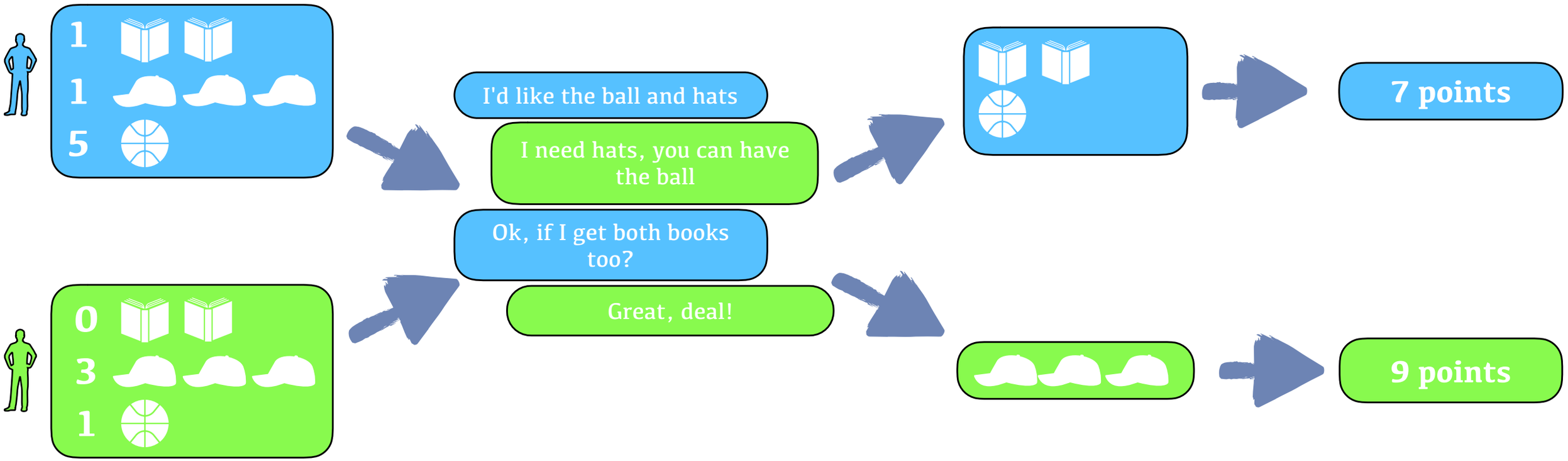
Agent 1 Output → Agent 1 Reward

Agent 1 Output → Agent 1 Reward

If both **agree** each is given **reward** by environment

# Dataset

## Object division task:



# Dataset

**Object division task:**

You can have 3 books, I will take the rest

I will give you the basketball and a book

You can have 4 books, final offer

That deal is not fair and I will not accept. Split it down the middle or no deal.

Fine, walk away with nothing

You are doing the same. Hope you enjoy your rejection.

Alright I'll take a hat and a book

Awesome. Pleasure doing business with you

# Dataset

| Metric | Dataset |
|---|---|
| Number of Dialogues | 5808 |
| Average Turns per Dialogue | 6.6 |
| Average Words per Turn | 7.6 |
| Agreed (%) | 80.1 |
| Average Score (max 10) | 6.0 |
| Pareto Optimality (%) | 76.9 |

# Baseline Model

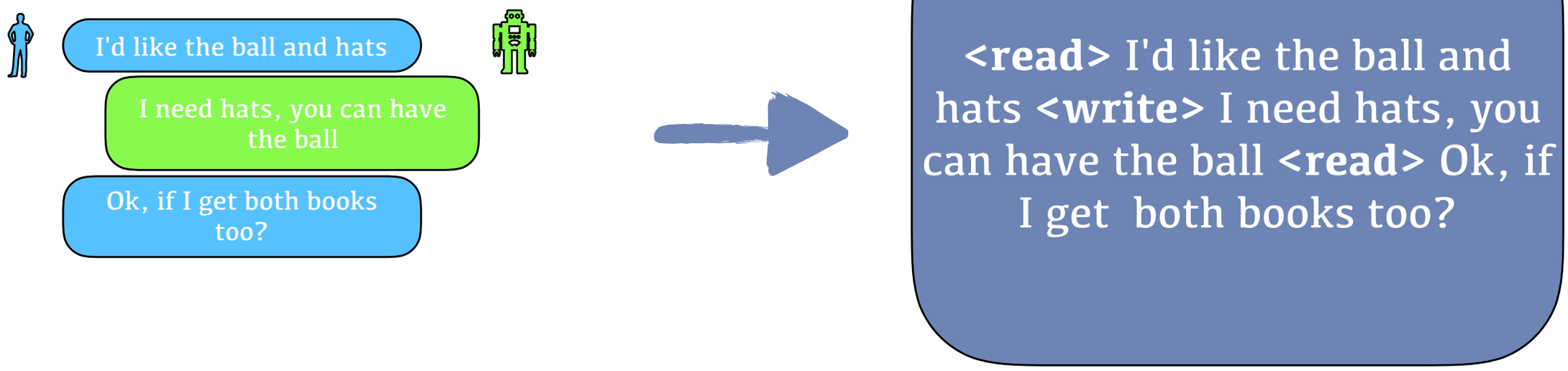
How to model an agent that can both perform **reasoning** and **conversing**?

Let's follow an end-to-end approach:

- Single model for **interpretation**, **generation** and **reasoning**

# Baseline Model

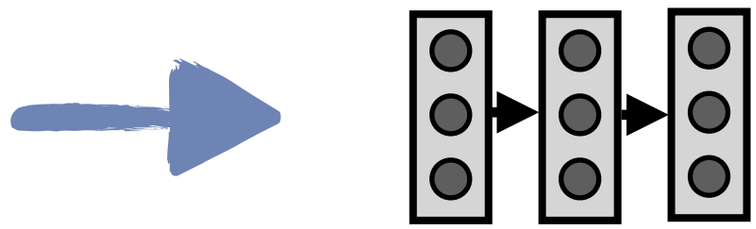
Linearize dialogue into token sequence



# Baseline Model

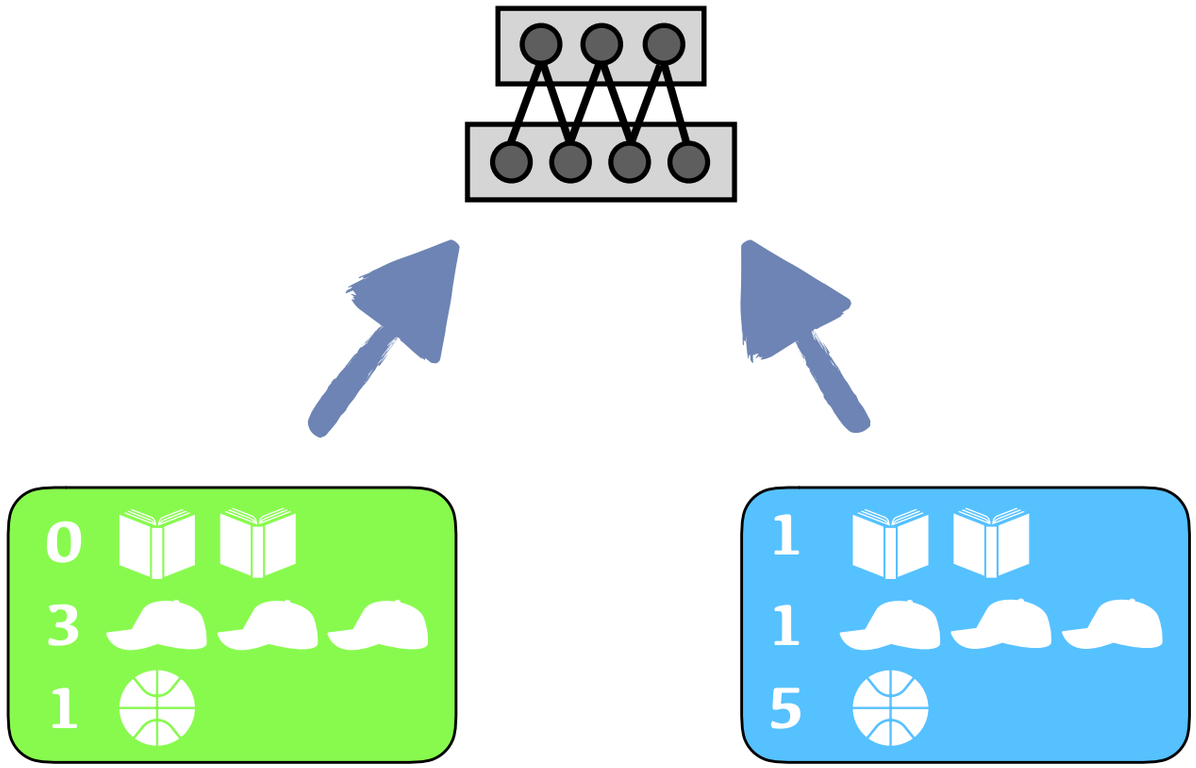
Train conditional language model to predict tokens

<read> I'd like the ball and hats <write> I need hats, you can have the ball <read> Ok, if I get  both books too?



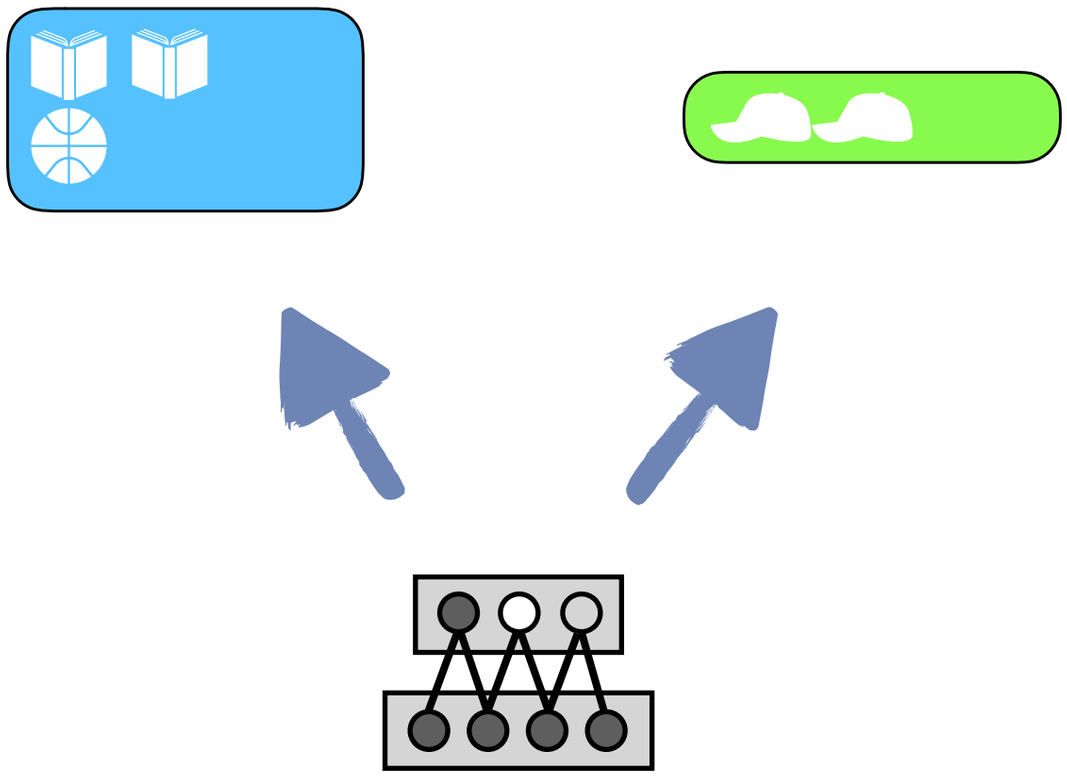
# Baseline Model

Train **Encoder** to read in the value function



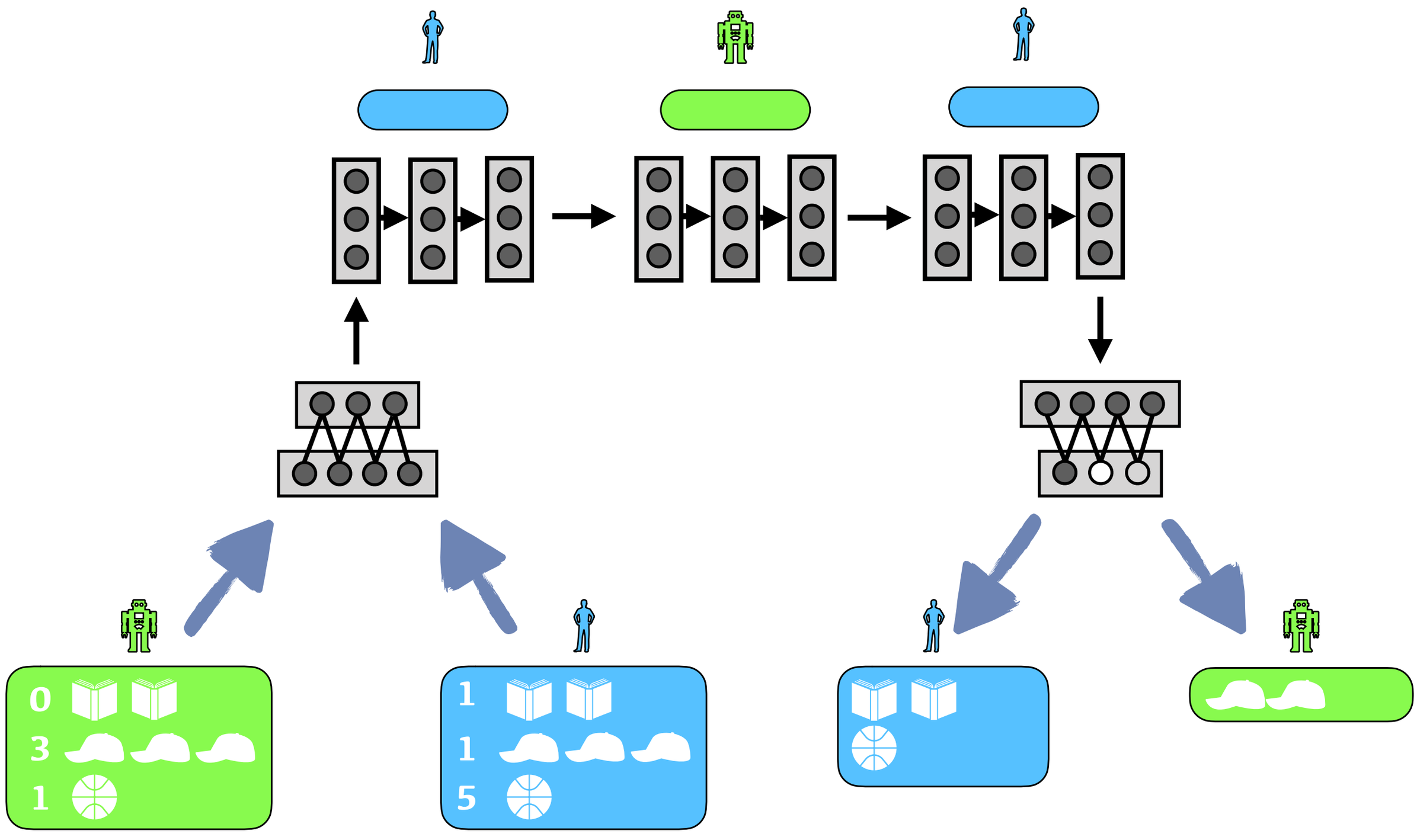
# Baseline Model

Train **Classifier** to predict the final deal



# Baseline Model

Combine everything **together**



# Training

Train language model using **supervised** learning to minimize NLL:

- model knows nothing about the task, just imitates human actions
- agrees to easily
- can't go beyond human strategies

$$J(\theta) = J_{\text{lm}}(\theta) + J_{\text{classifier}}(\theta)$$

$$J_{\text{lm}}(\theta) = -\sum_t \sum_i x_{t,i} \log \hat{x_{t,i}}$$

$$J_{\text{classifier}}(\theta) = -\sum_t \sum_i s_{t,i} \log \hat{s_{t,i}}$$

# Training

Use **reinforcement** learning to teach model to negotiate:

- generate dialogues using **self-play**
- backpropagate reward using **REINFORCE**

RL **adversely affects** model's language coherence:

- **interleave** supervised and reinforcement training

# Training

**Monte-Carlo estimate of objective function:**

$$J(\theta) = \mathbb{E}_{\tau \sim \pi_\theta}[\sum_t r(s_t, y_t)] \approx \frac{1}{N} \sum_i \sum_t r(s_{i,t}, y_{i,t})$$

**Policy Gradient:**

$$J(\theta) = \mathbb{E}_{\tau \sim \pi_\theta}[r(\tau)] = \int \pi_\theta(\tau) r(\tau) d\tau$$

$$\nabla_\theta J(\theta) = \int \nabla_\theta \pi_\theta(\tau) r(\tau) d\tau = \int \pi_\theta(\tau) \nabla_\theta \log \pi_\theta(\tau) r(\tau) d\tau =$$

$$= \mathbb{E}_{\tau \sim \pi_\theta}[\nabla_\theta \log \pi_\theta(\tau) r(\tau)]$$

**Update:**

$$\theta \leftarrow \theta + \alpha \nabla_\theta J(\theta)$$

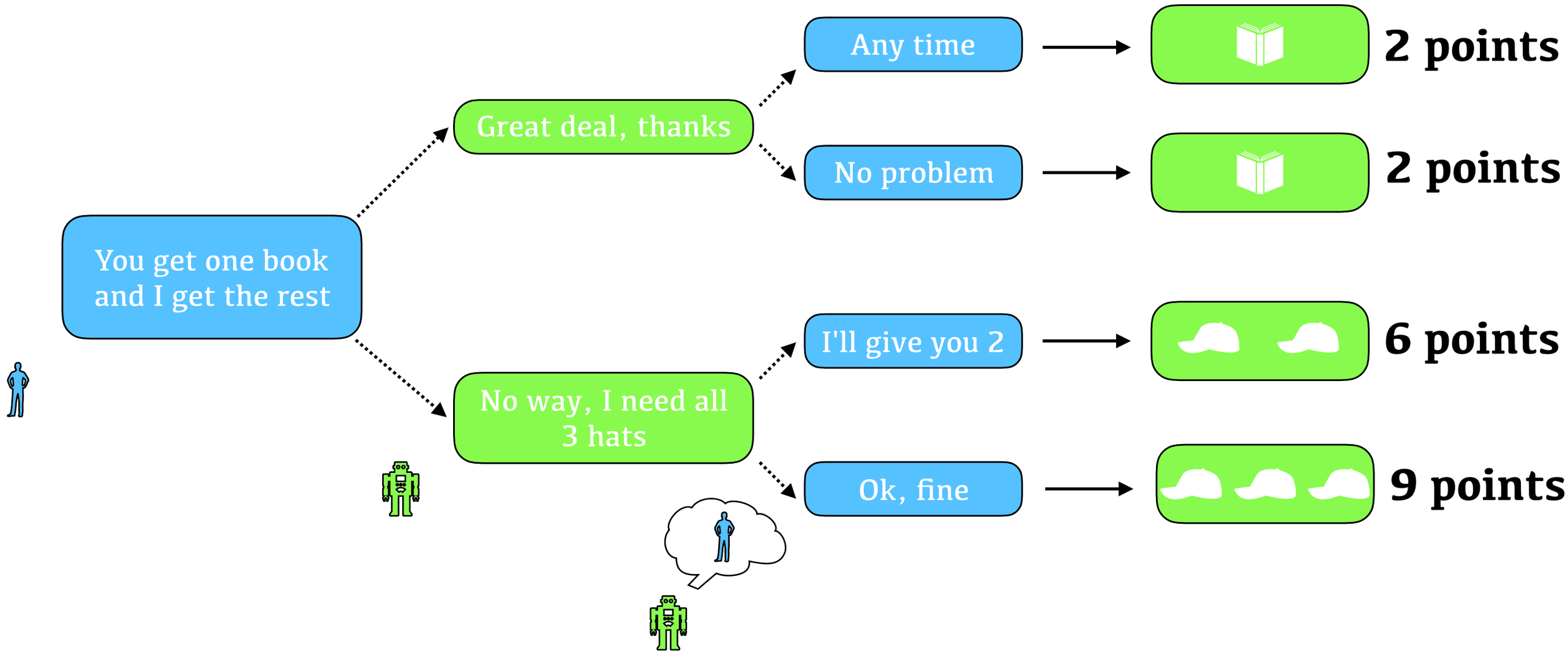
**[Figures from http://rll.berkeley.edu/deeprlcourse/f17docs/lecture_4_policy_gradient.pdf]**

# Rollouts

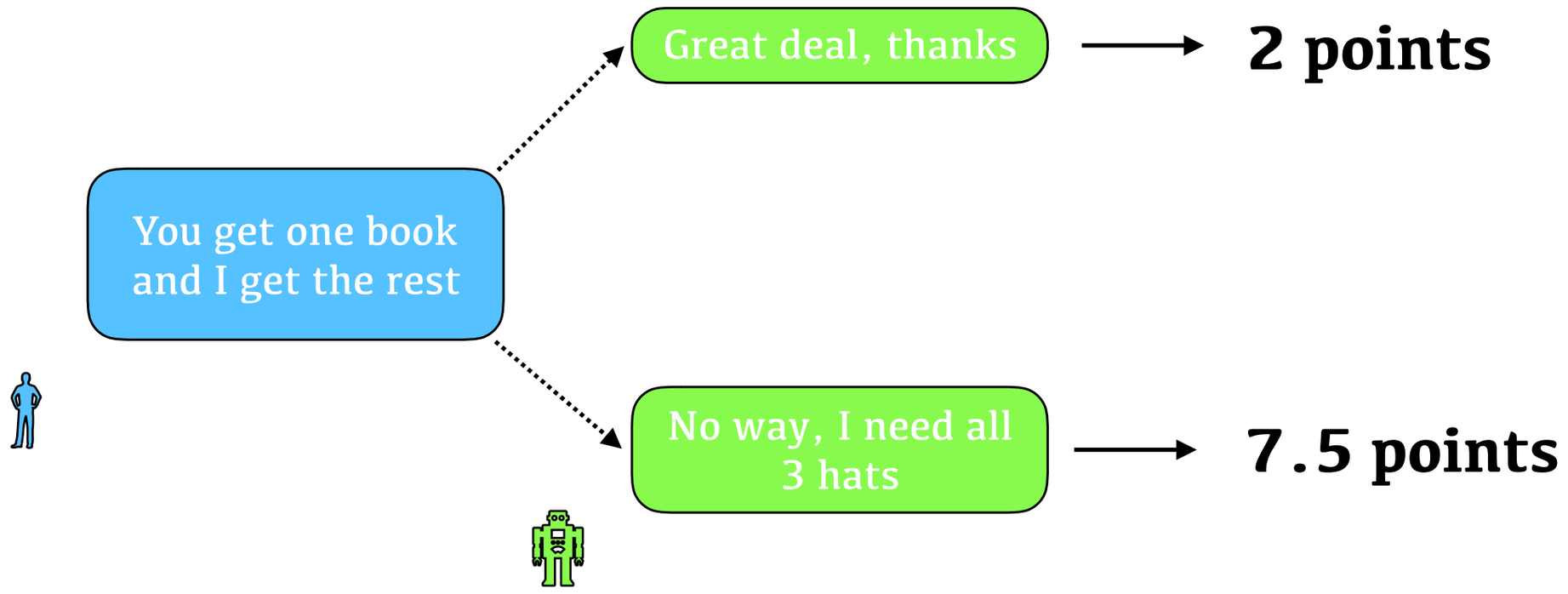
Finally, on trained model we can use rollouts:

- estimates expected score of a dialogue
- widely used in games:
  - Monte-Carlo-Tree-Search in AlphaGo

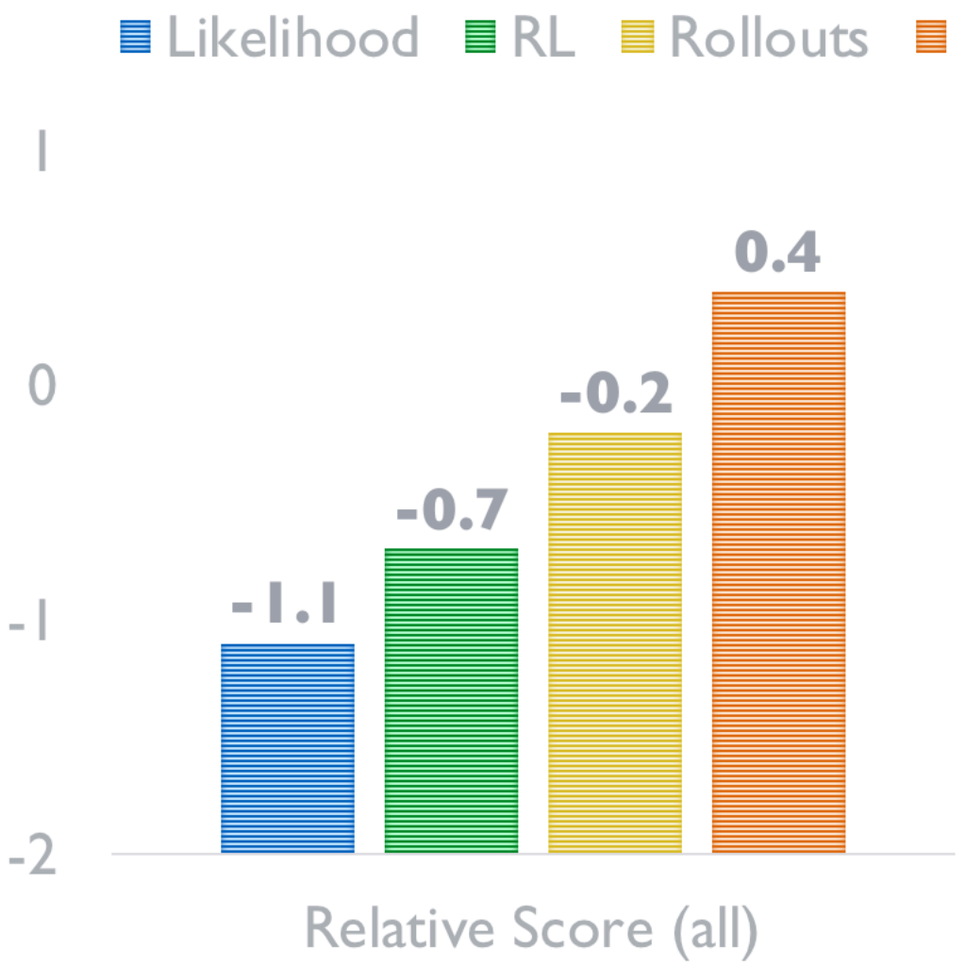
# Rollouts


You get one book and I get the rest
Great deal, thanks
Any time
2 points
No problem
2 points
No way, I need all 3 hats
I'll give you 2
6 points
Ok, fine
9 points

# Rollouts


Great deal, thanks
2 points
You get one book and I get the rest
No way, I need all 3 hats
7.5 points

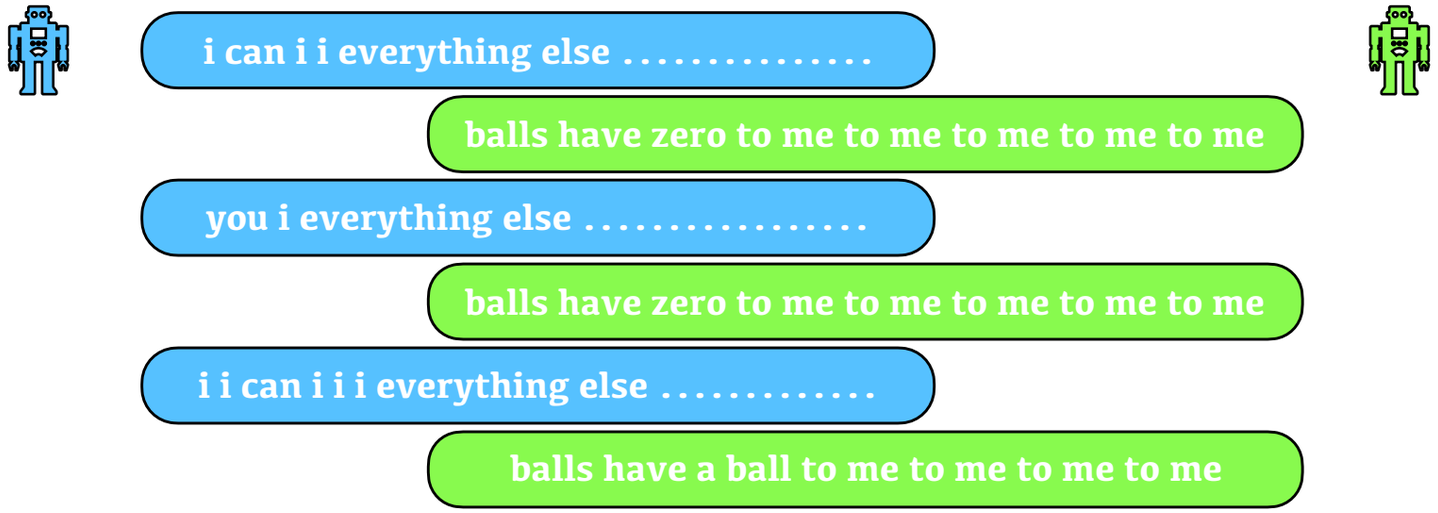
# Evaluation vs Human



# Issues

Entangled representations responsible for both **end-goal performance** and **language coherence**:

- Fine-tuning end-goal performance with RL adversely affects generated language



# Issues


Tuesday, August 1, 2017 FOR A GREATER BRITAIN 50p thesun.co.uk

FREE PULLOUT INSIDE

With one year's site fees: See Page 33

IN HER OWN WORDS, THE BOMBSHELL...

DIANA TAPES

'Sex with Charles fizzled out'

'I asked: Why is she around?'

THE FULL TRANSCRIPT: PAGES 6 & 7

ARE MACHINES TAKING OVER?

ROBO STOP

facebook shuts off robots after they chat in secret code

By JAMES BEAL, US Editor

FACEBOOK shut down an artificial intelligence experiment after two robots began talking in a tongue only they understood. The "chatbots" Alice and Bob – set for them to make it easier for them to communicate by creating sentences that were gibberish to watching scientists. Future technology expert Kate Adamson told The Sun: "It does feel a bit like The Terminator."

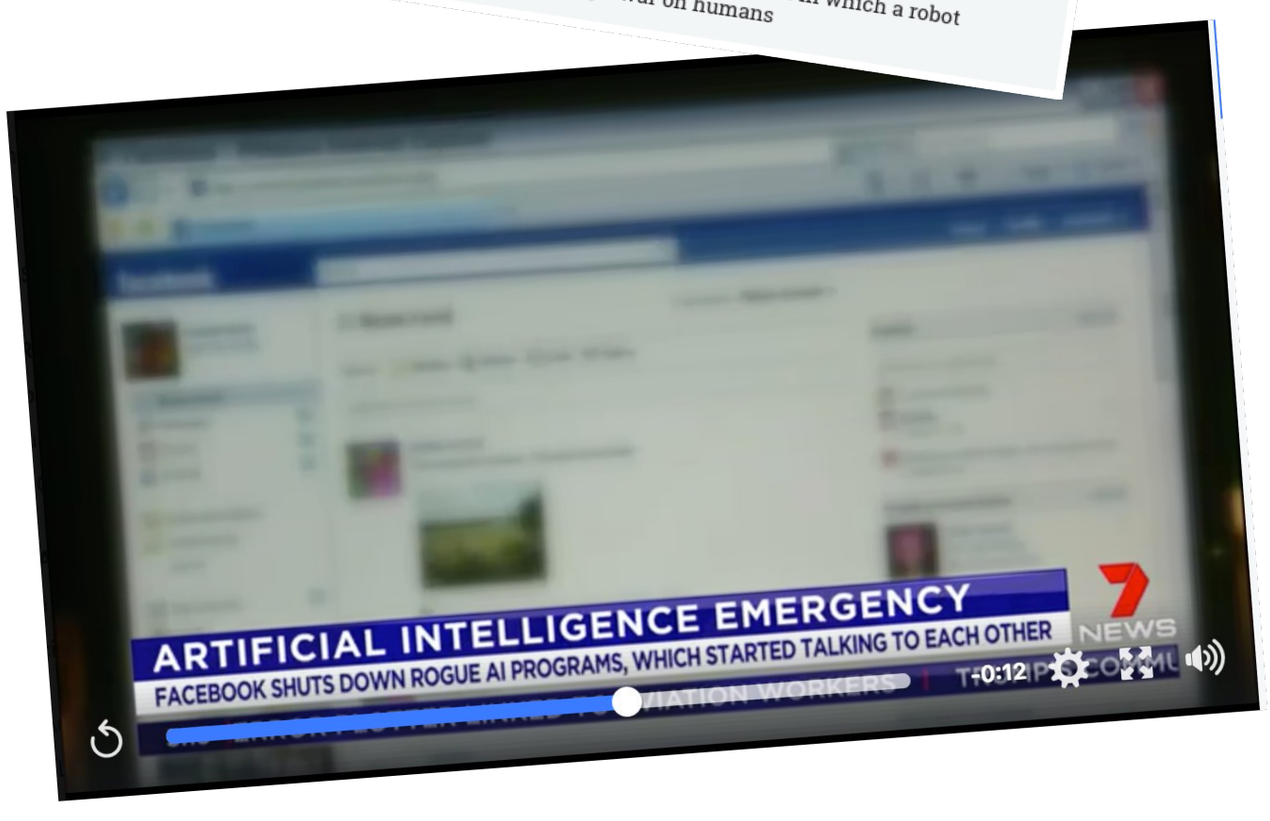
Experiment . . . robot

Full Story >> Page Five


HANDOUT

4

The incident closely resembles the plot of The Terminator in which a robot becomes self-aware and starts waging a war on humans



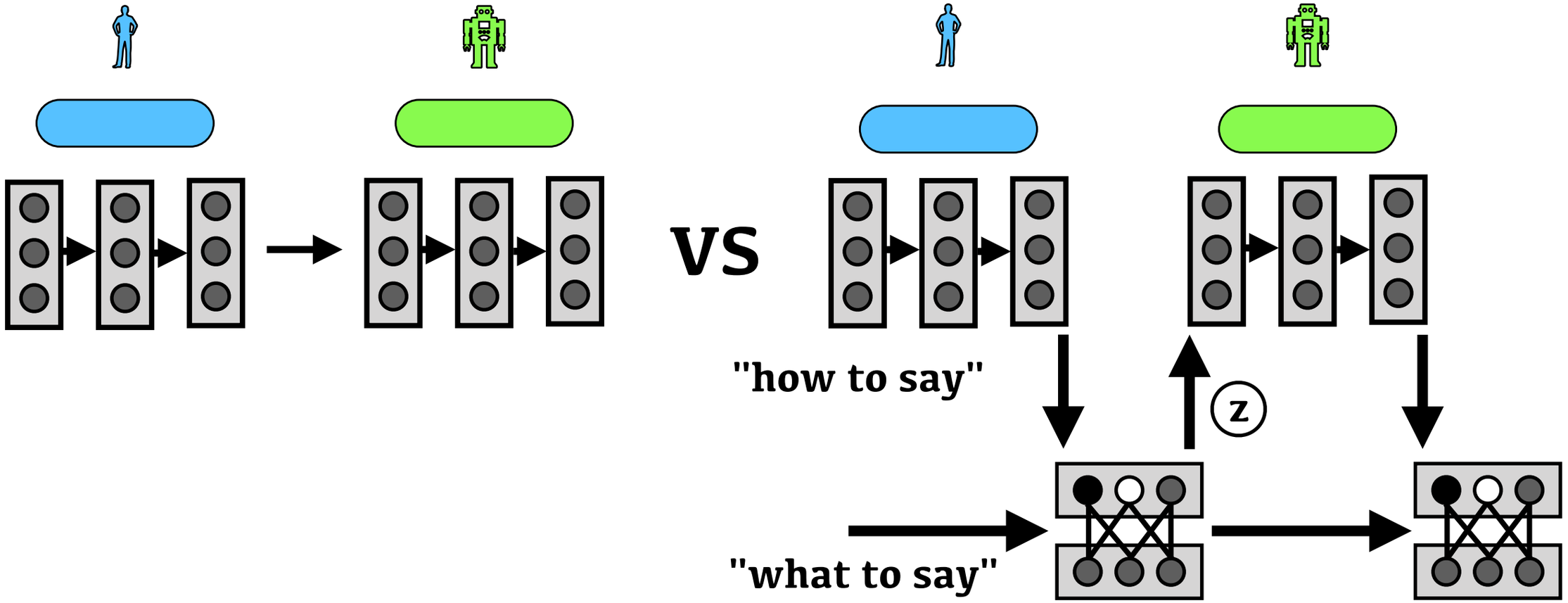
# Hierarchical Approach

**Solution**:

- Factorize generation problem into **"what to say"** and **"how to say"**

$$\pi_\theta(y_t|s_t) \quad \textbf{VS} \quad \pi_\theta^{\text{lm}}(y_t|z_t)\pi_\theta^{\text{z}}(z_t|s_t)$$

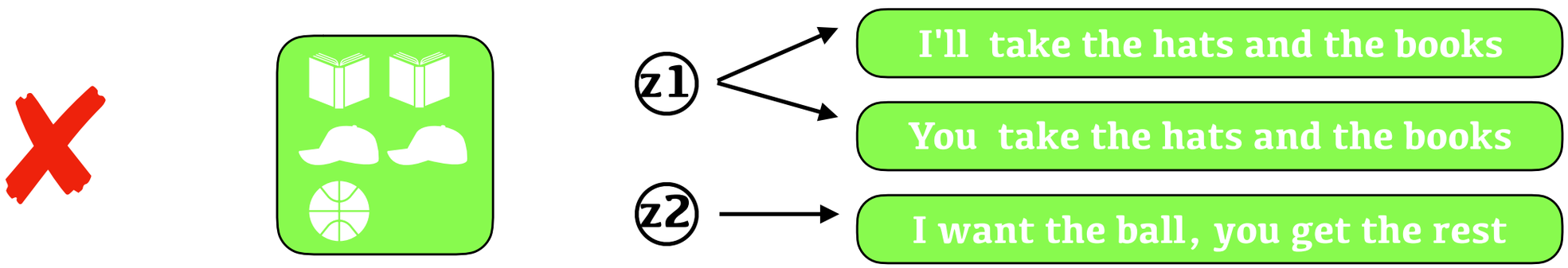
**"how to say"** **"what to say"**



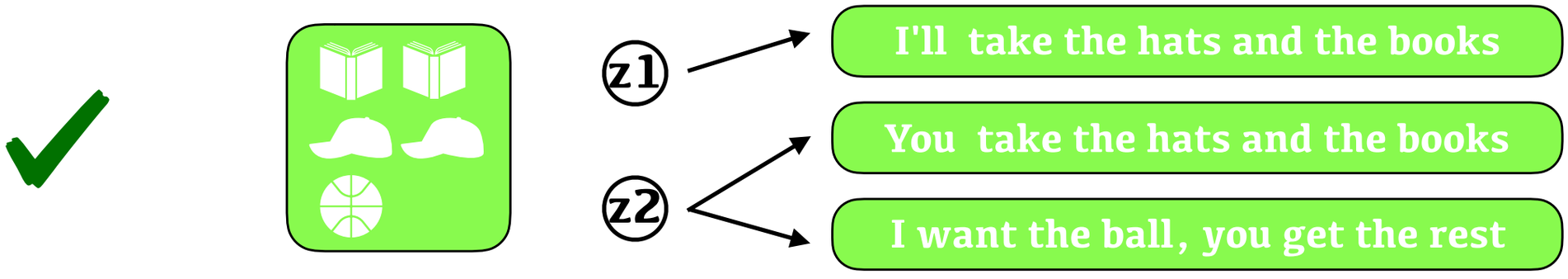
# Latent Variable

**How to learn z?**

- Learning z to maximize likelihood of messages gives similar strings, not similar meaning

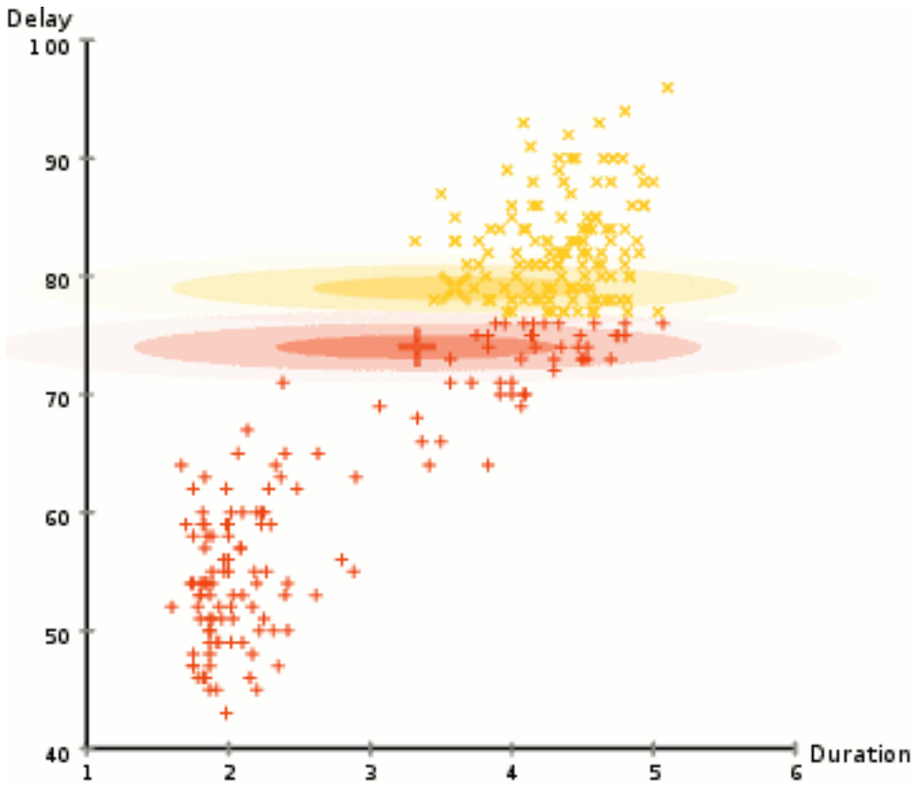



- Learning z to capture message's effect on dialogue means that semantically similar messages are clustered



# Latent Variable

- Pretrain **z** using  Expectation Maximization algorithm
- Widely used for models with latent variables:
  - e.g. Mixture of Gaussians

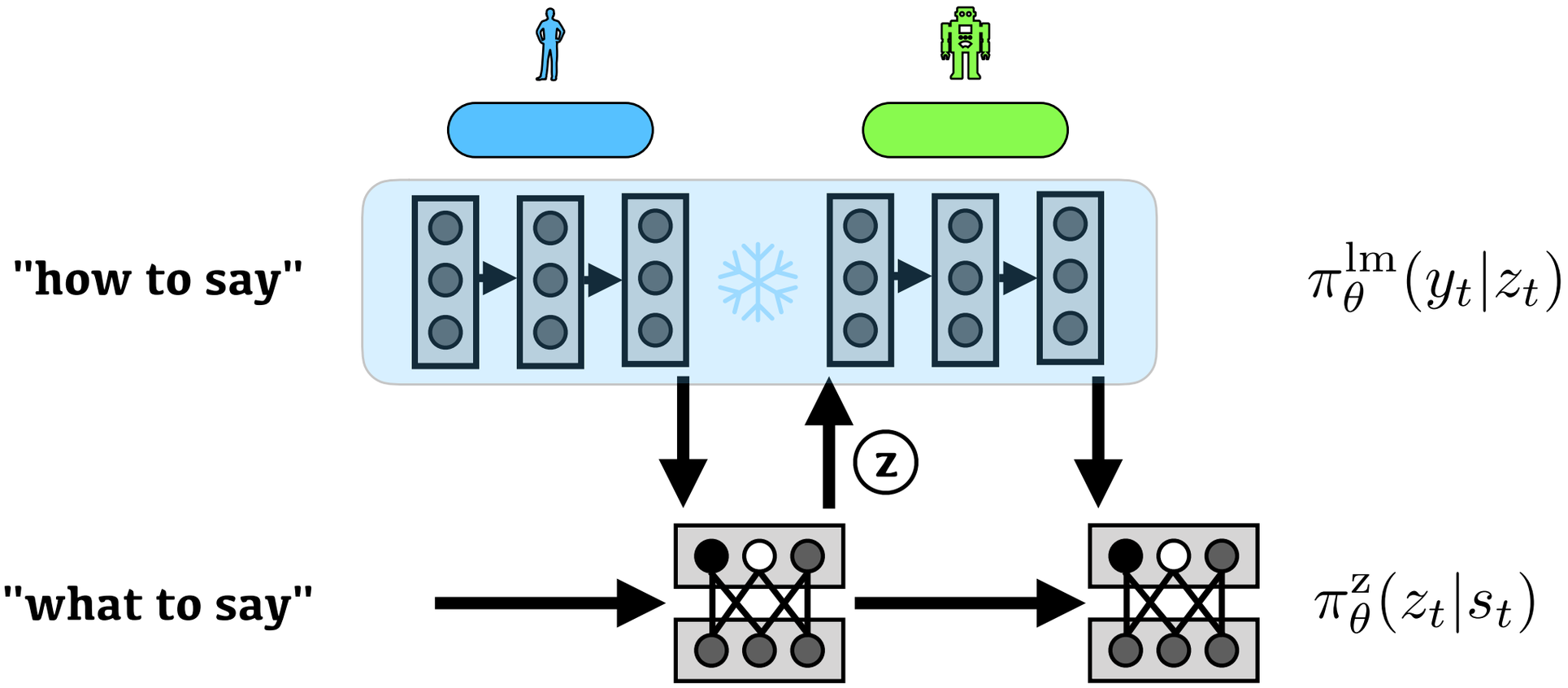



[Figures from https://en.wikipedia.org/wiki/Expectation-maximization_algorithm]

# Hierarchical Approach

**Benefits:**

- Freeze "**how to say**" and fine-tune "**what to say**" with RL to preserve language cohesiveness
  - i.e. only optimize $\pi_\theta^{\mathrm{z}}(z_t|s_t)$ with RL
  - $\pi_\theta^{\mathrm{lm}}(y_t|z_t)$ stays unchanged
- Rollouts using **z** are more efficient, only done on $\pi_\theta^{\mathrm{z}}(z_t|s_t)$



# Learned Clusters

| Cluster | BASELINE CLUSTERS | FULL MODEL |
|---|---|---|
| 1 | i can give you the books but , i would need the hat and the balls | i would like the hat and 1 book |
| | i can do that . i need both balls and one book | i can't give up the hat , but i can offer you the book and 2 balls |
| 2 | i need both books and the hat | i want the hat |
| | how about you get the hat and 1 ball | i need the hat . you can have all the books and the balls |
| 3 | i can not make that deal . i need the hat and one book | i can give you the hat and 1 ball |
| | i can give you the hat and 1 ball | i would like the books and a ball |
| 4 | i need two books and the hat | i need the books and the hat |
| | i need the hat , you can have the rest | i can give you the balls but i need the hat and books |
| 5 | i can give you the hat if i can have the rest | could i have the books and a ball ? |
| | i want one of each | i would like the books and one ball |

Table 4: Sample messages that are probable under different clusters for specified context, in comparison to a previous approach to learning message representations. An agreement needs to be done over a set of **2 books**, **1 hat**, and **2 balls**. The clusters produced by our method are much more semantically coherent than the baseline, and correspond closely to different ways of proposing the same deal.

# Experiments

| Model | Validation Perplexity | Test Perplexity |
|---|---|---|
| RNN | 5.62 | 5.47 |
| HIERARCHICAL | 5.37 | 5.21 |
| BASELINE CLUSTERS | 5.61 | 5.46 |
| FULL MODEL | 5.37 | 5.24 |

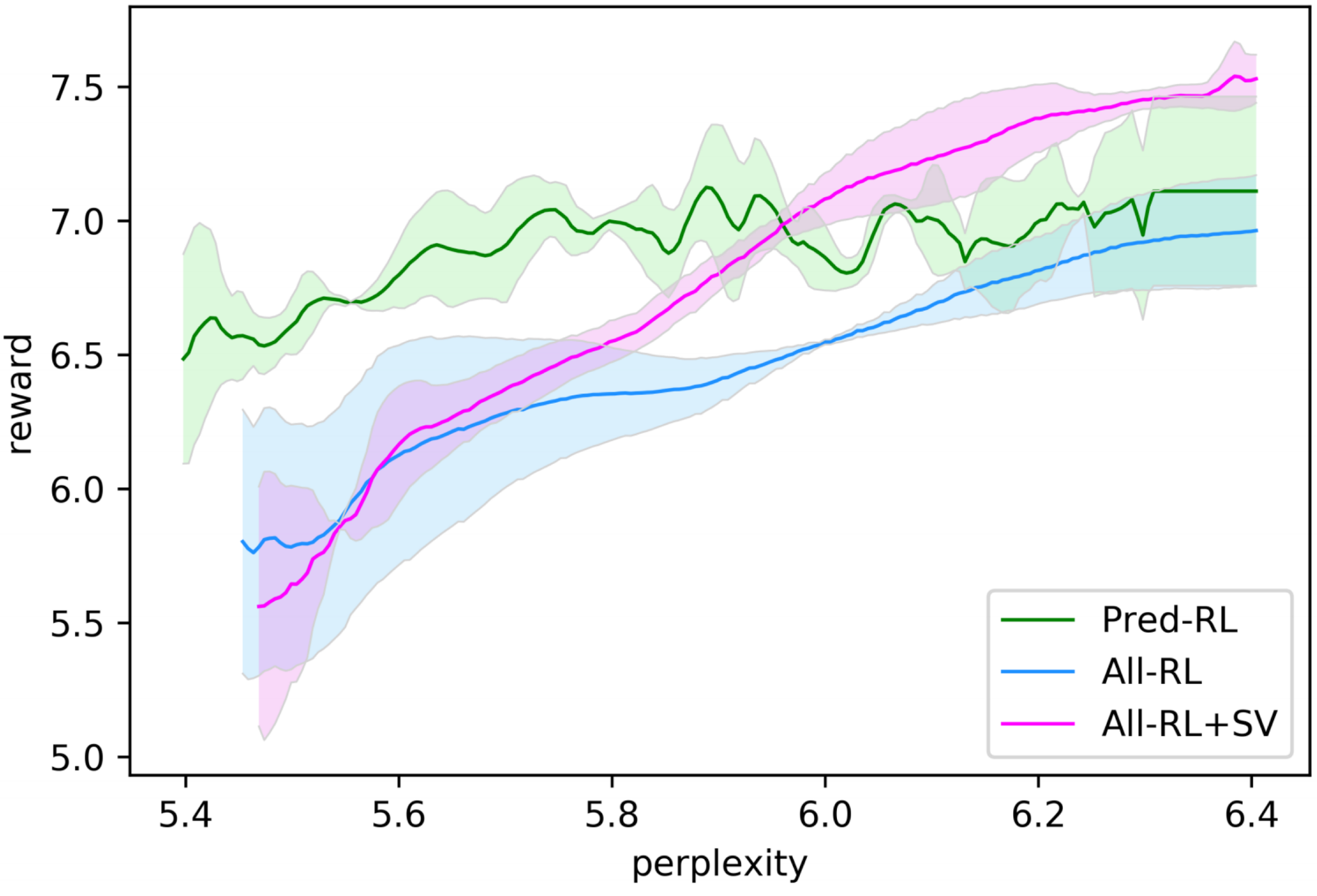
Table 1: Perplexity results, showing the likelihood of human dialogues using different models. Our model with discrete message representations is able to achieve state-of-the-art performance, showing that the representations effectively capture relevant aspects of messages for predicting the future dialogue.

# Experiments

| Model | Score vs. RNN | Score vs. HIERARCHICAL |
|---|---|---|
| RNN | 5.33 | 5.17 |
| HIERARCHICAL | 5.37 | 5.08 |
| BASELINE CLUSTERS | 4.68 | 4.66 |
| FULL MODEL | 6.75 | 6.57 |

Table 2: Comparison of different models based on their end-task reward. Our clusters substantially improve reward, indicating that they make it easier for supervised learning to model strategic decision making.

# Experiments



# Research Papers

## Deal or no Deal? End-to-End Learning for Negotiation Dialogues

- URL: https://arxiv.org/abs/1706.05125


## Hierarchical Text Generation and Planning for Strategic Dialogue

- https://arxiv.org/abs/1712.05846

# Open Source

## End-to-End Negotiator

- **URL: https://github.com/facebookresearch/end-to-end-negotiator**