Deal or No Deal? End-to-End Learning of Negotiation Dialogues

Mike Lewis

Facebook AI Research, Menlo Park, CA

(joint work with: **Denis Yarats**, **Yann N. Dauphin**, **Devi Parikh**, **Dhruv Batra**

Sun

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

FREE PULLOUT INSIDE

With one year's site fees: See Page 18

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

Experiment . . . robot

By JAMES BEAL, US Editor

FACEBOOK shut down an artificial intelligence experiment after two robots began talking in a language only they understood.

The "chatbots" Alice and Bob modified English to make it easier for them to communicate — creating sentences that were gibberish to watching scientists.

Future technology expert Kate Adamson told The Sun: "It does feel a bit like The Terminator."

Full Story — Page Five

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


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

PULLOUT INSIDE

With one year's site fees: See Page 10

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

Experiment . . . robot

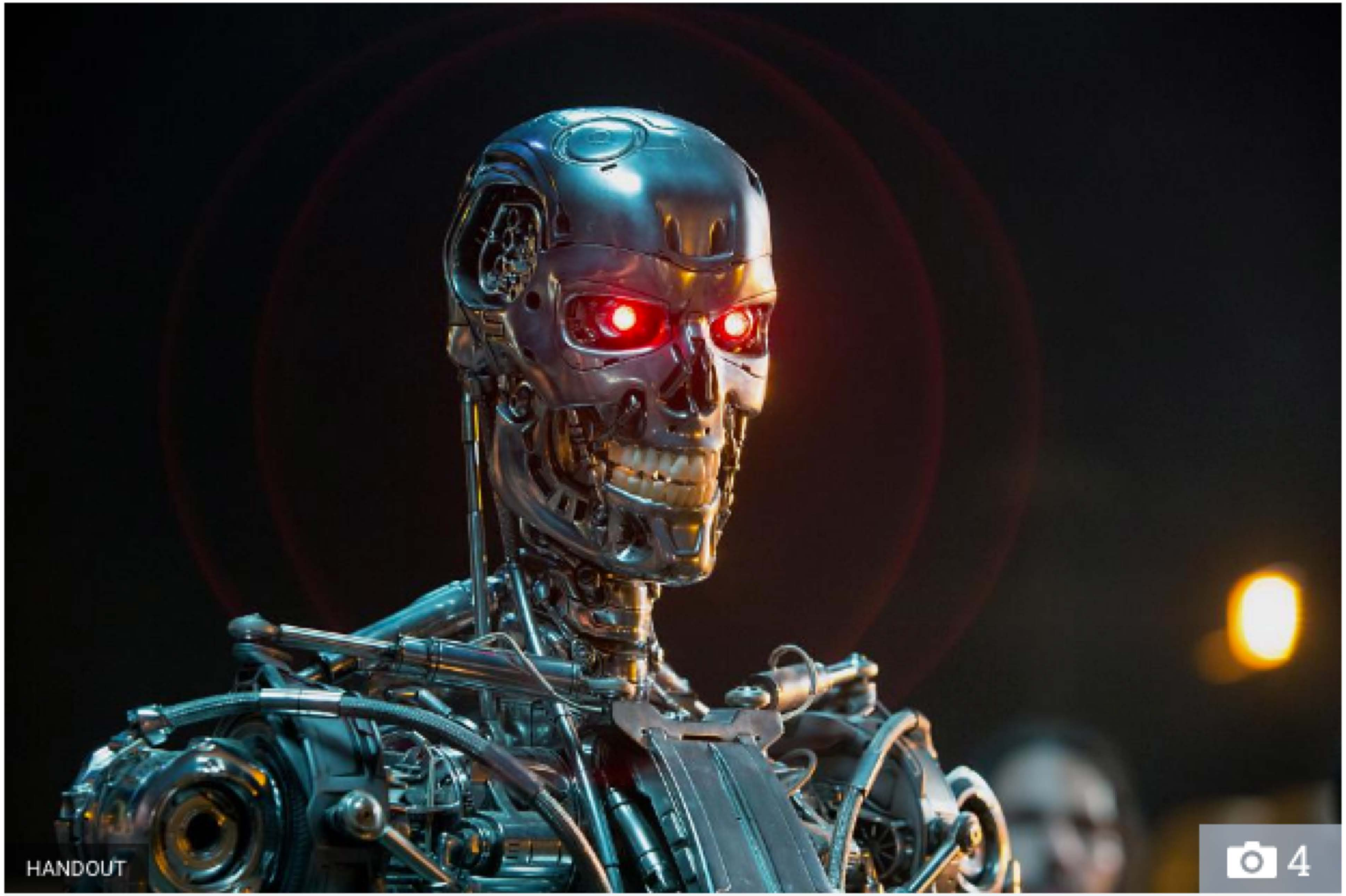
By JAMES BEAL, US Editor

FACEBOOK shut down an artificial intelligence experiment after two robots began talking in a language only they understood.

The "chatbots" Alice and Bob modified English to make it easier for them to communicate — creating sentences that were gibberish to watching scientists.

Future technology expert Kate Adamson told The Sun: "It does feel a bit like The Terminator."

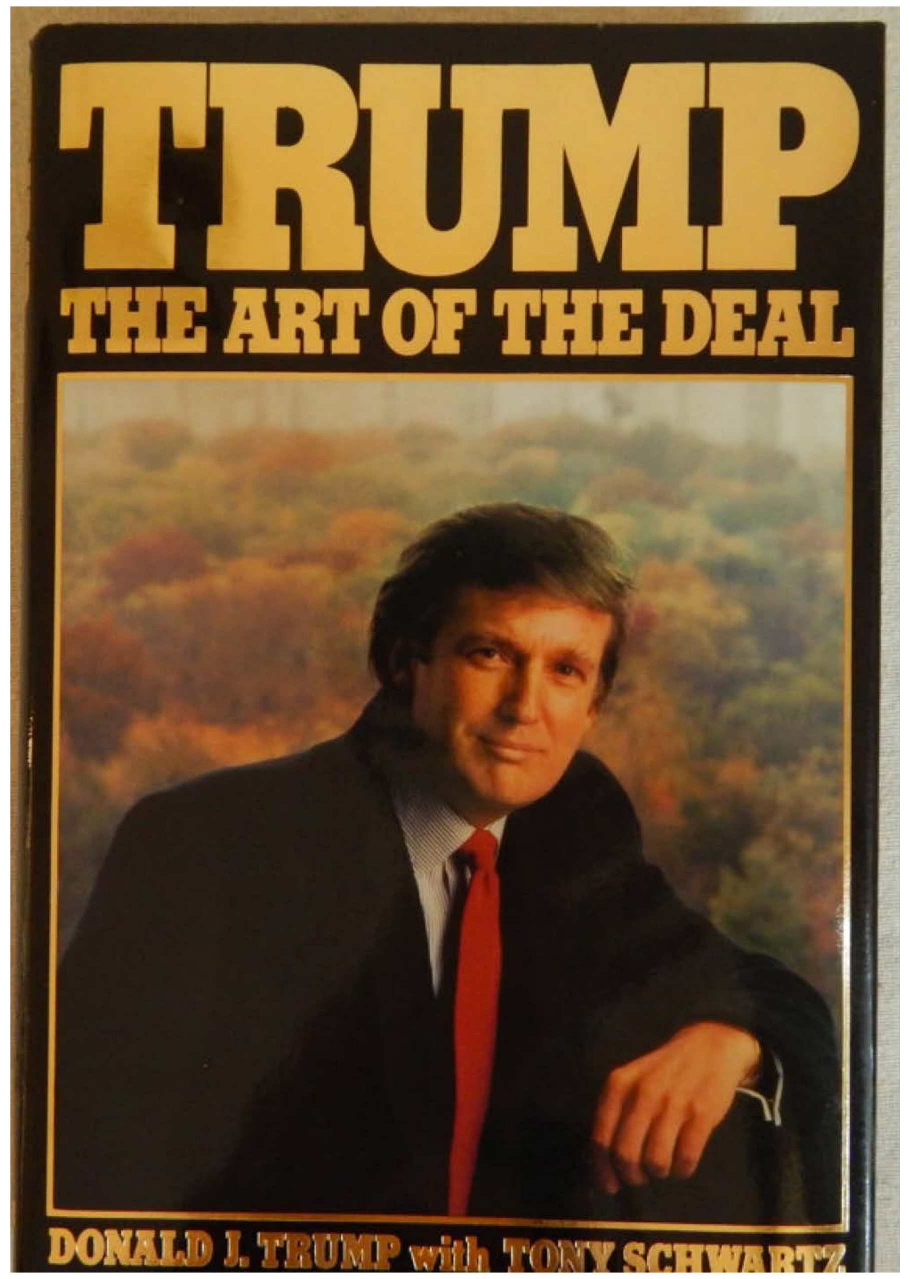
Full Story — Page Five



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

Why Negotiation?

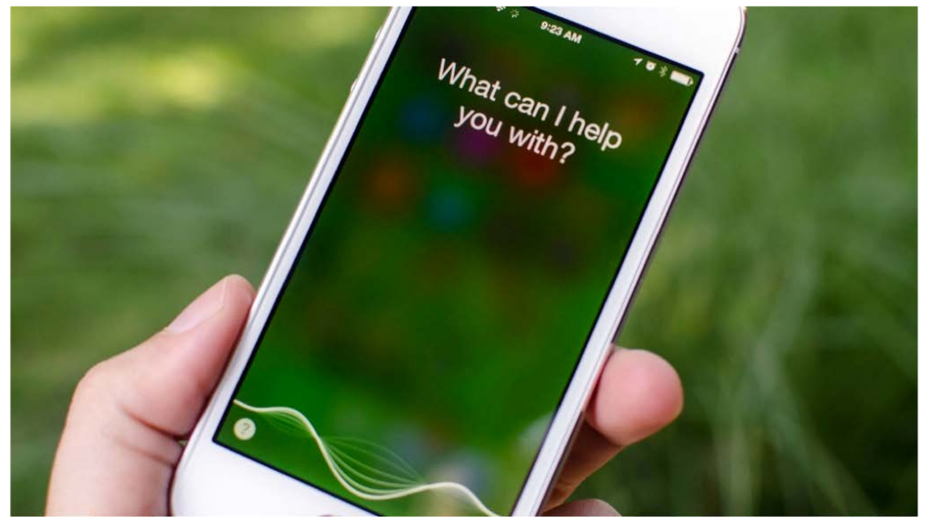
Why Negotiation?



Negotiation useful, when:

- Agents have different goals
- Not all can be achieved at once
- ***(all the time!)***

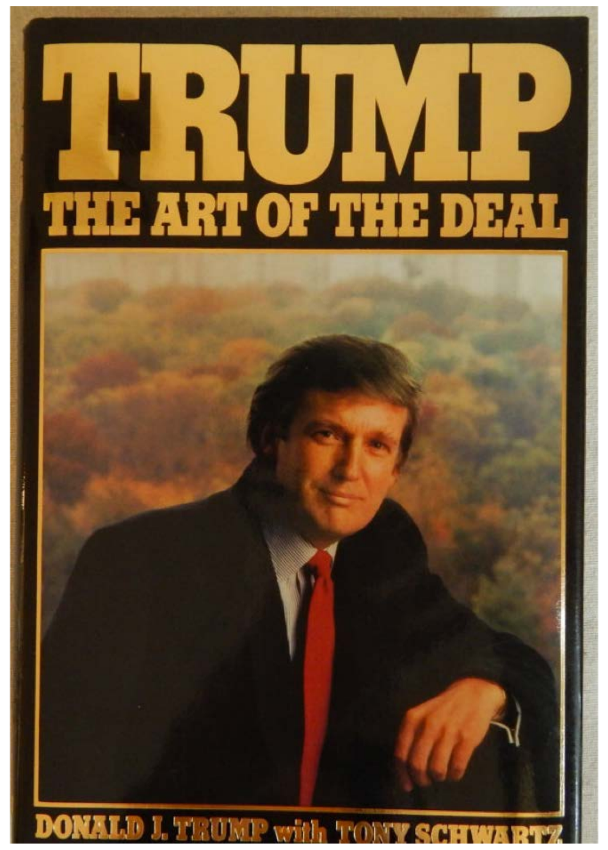
Why Negotiation?



Zero-sum / Adversarial

Negotiation

Fully Cooperative



Why Negotiation?

Both **linguistic** and **reasoning** problem

Interpret multiple sentences, and **generate** new message

Plan ahead, make proposals, counter-offers, ask questions, vagueness, bluffing, deceit, compromising

Hard for current models

Why Negotiation?

Unlike many goal-orientated dialogue problems, **no simple solutions** to achieving goal

Incentive to strategically **withhold information**

Adversarial aspect means it **hard to “solve”**

Why Negotiation?

Easy to **evaluate** – how good a deal did an agent get?

Self-play gives good **development metric**

Why Negotiation?

Real Applications

Many people find negotiations **hard** and **awkward**

Could practice with bots help?

Dataset

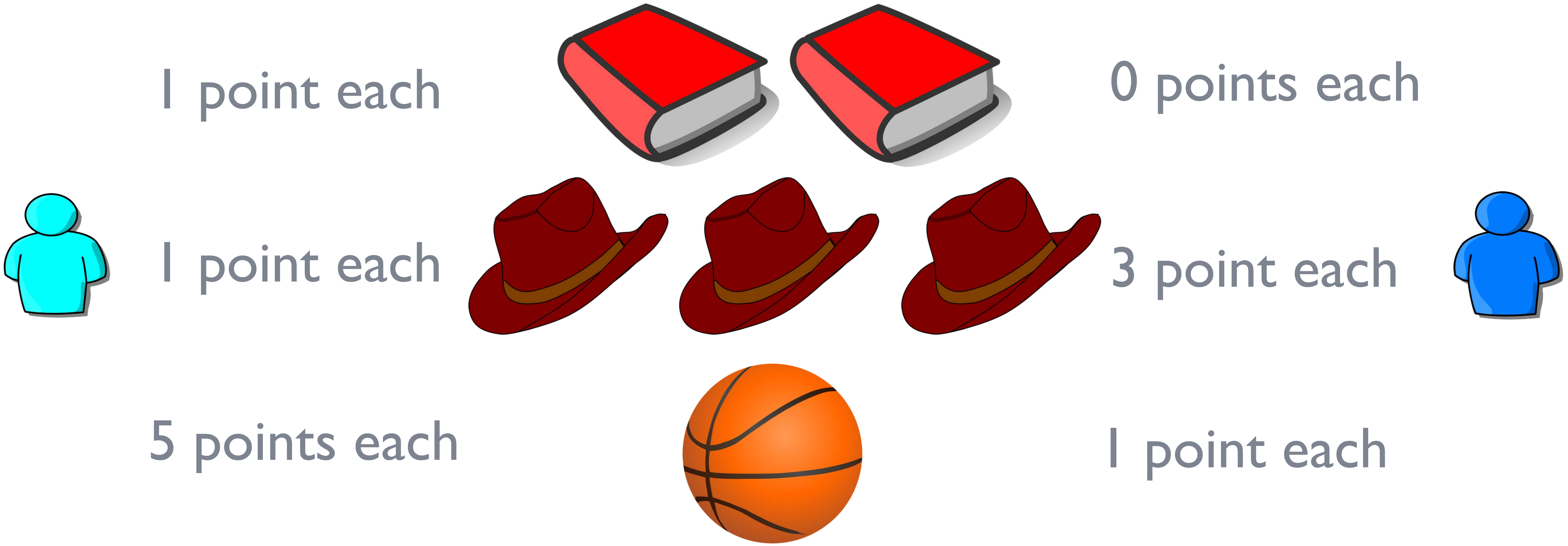
Framework

Both agents given *reward function*, **can't observe each other's**
Both agents **independently** select a deal
Agent 1 Goals
Agent 1 Output
Agent 1 Reward
Dialogue
Agent 2 Goals
Agent 2 Output
Agent 2 Reward
Dialogue until one agent enters that **deal is agreed**
If both **agree** each is given **reward** by environment

Object Division Task

Agents shown **same objects** but **different values** for each

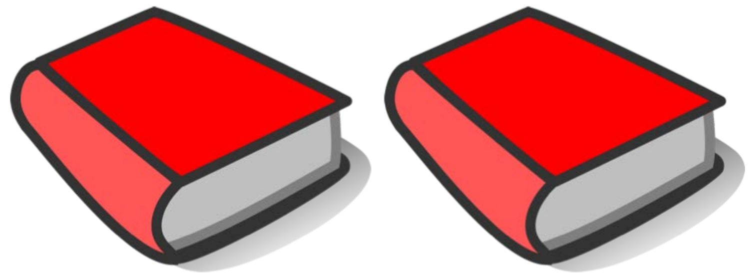
Must **agree** how to divide objects between them

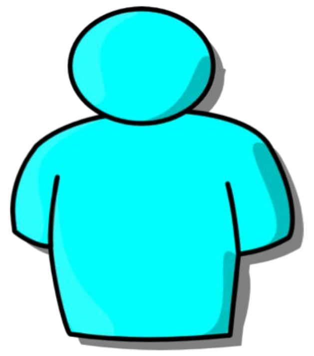
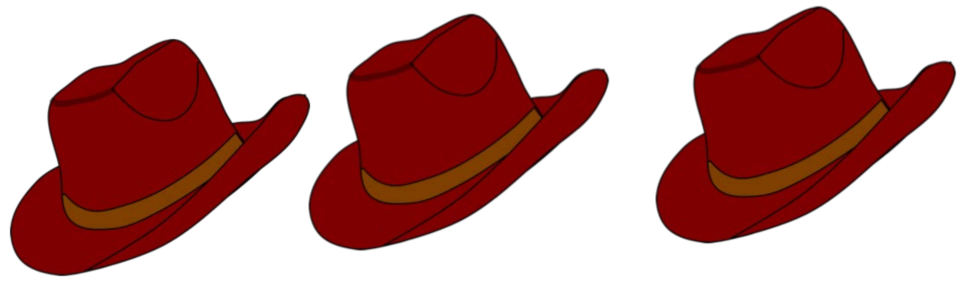


Object Division Task

Agents shown **same objects** but **different values** for each

Must **agree** how to divide objects between them

1 point each  0 points each

 1 point each  3 point each 

5 points each  1 point each

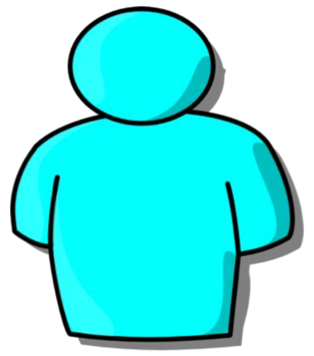
7

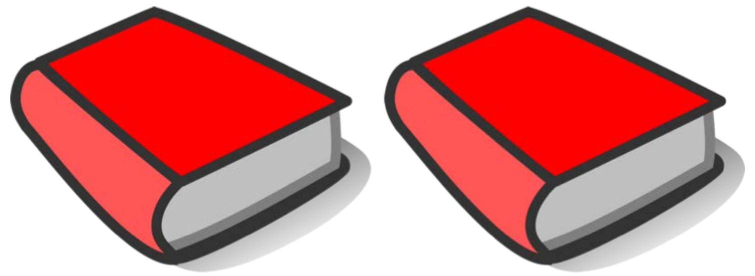
9

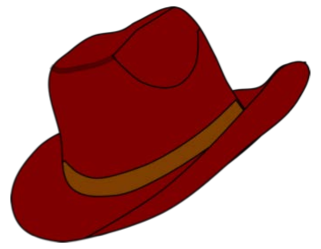
Object Division Task

Agents shown **same objects** but **different values** for each

Must **agree** how to divide objects between them

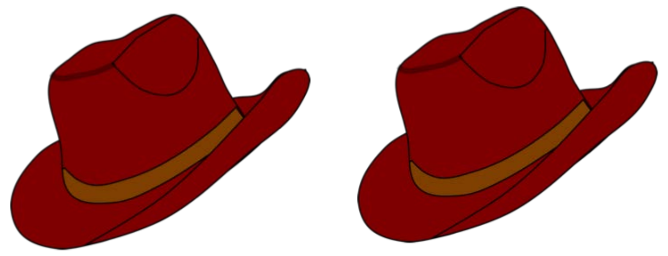


1 point each 

1 point each 

5 points each 

0 points each

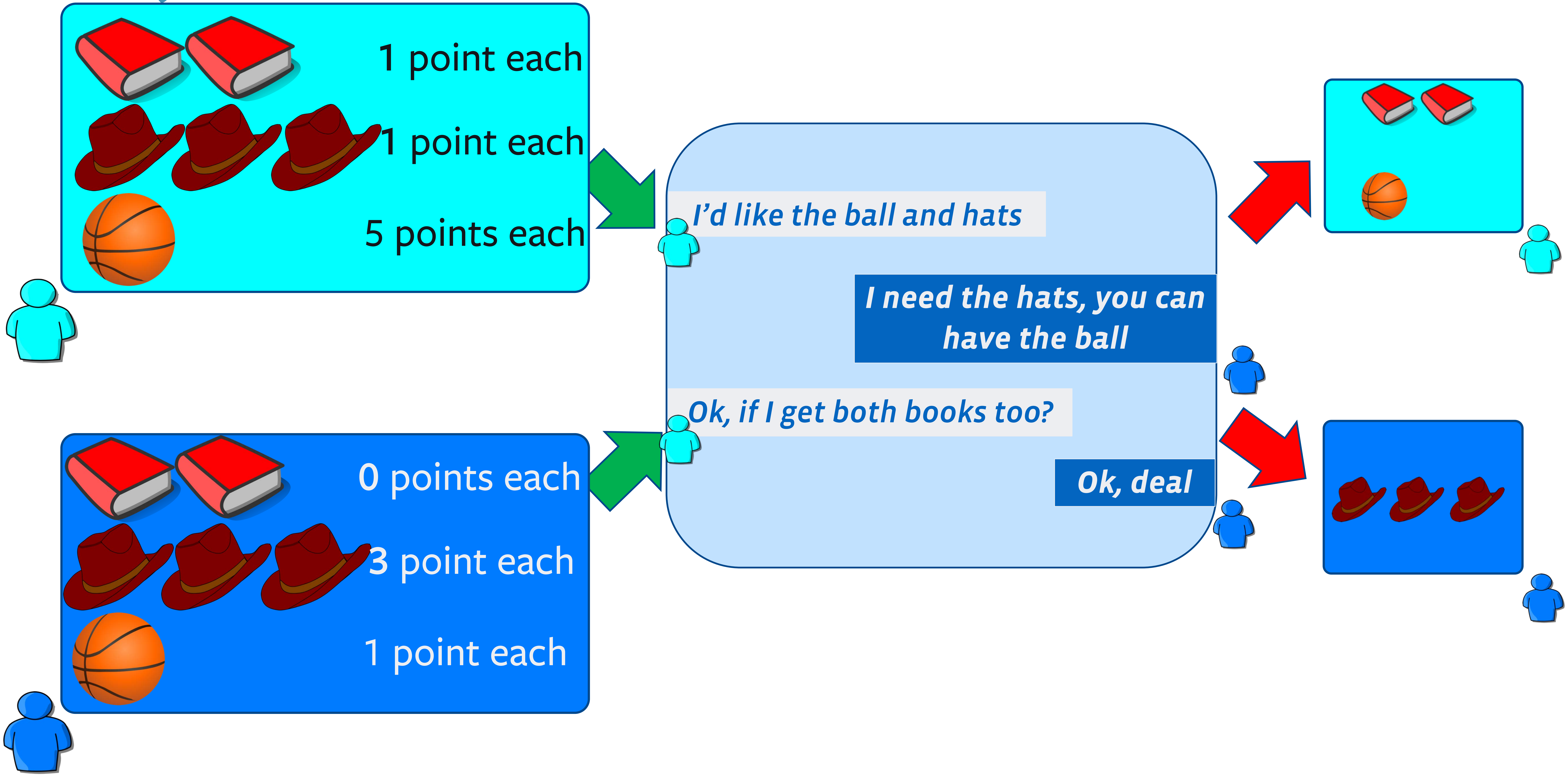
 3 point each

1 point each

8

6

Object Division Task

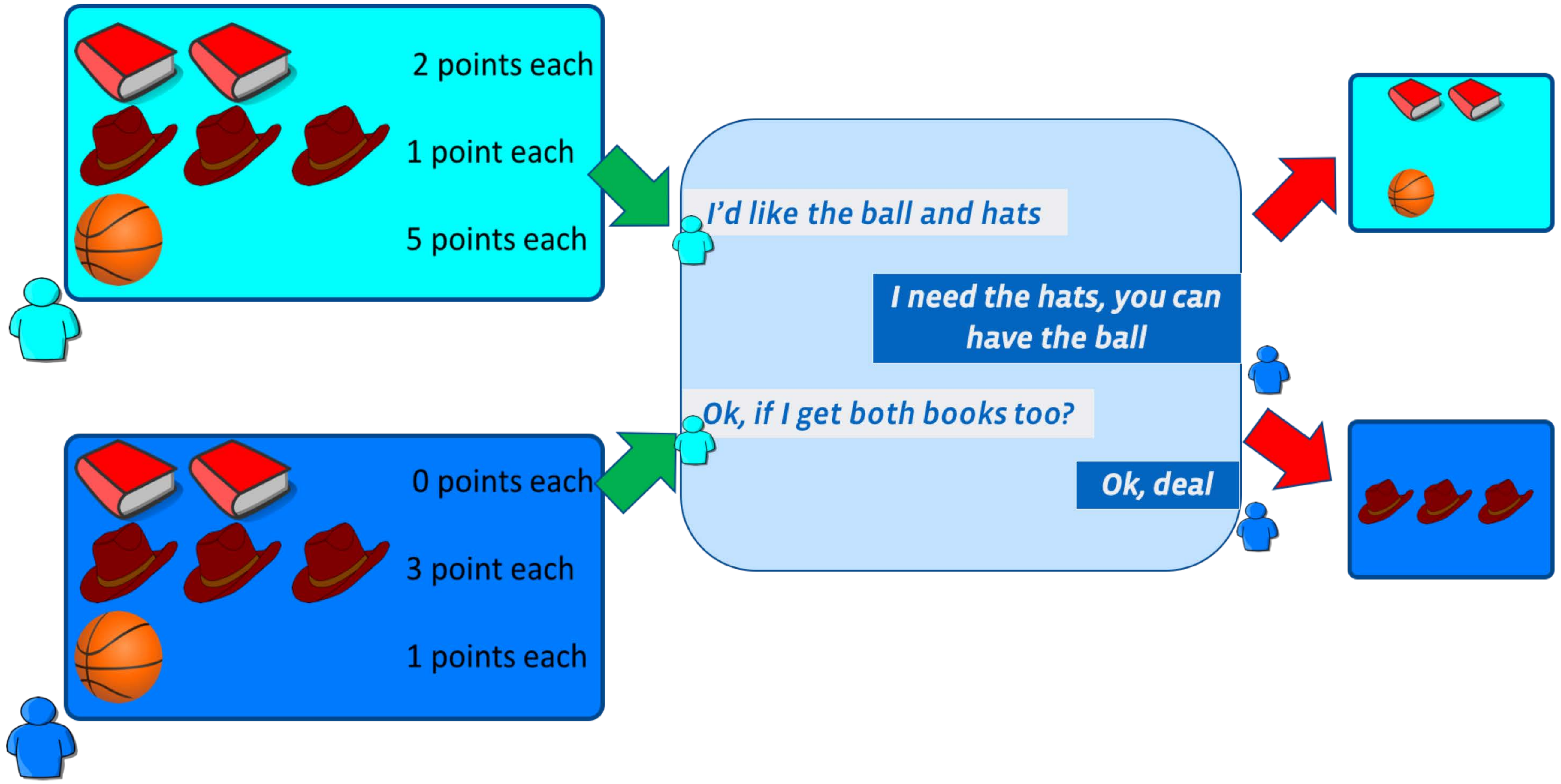


Object Division Task

10 point **maximum**

Not possible for **both** agents to score 10 points

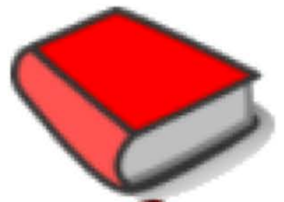
Failing to agree is 0 points



Object Division Task

Divide these objects between you and another Turker. Try hard to get as many points as you can!

Send a message now, or enter the agreed deal!

Items	Value	Number You Get
	8	1
	1	1
	0	0

Mark Deal Agreed ✔

Fellow Turker: I'd like all the balls

You: Ok, if I get everything else

Fellow Turker: If I get the book then you have a deal

You: No way - you can have one hat and all the balls

Fellow Turker: Ok deal

Type Message Here:

Message

Send

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 is 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

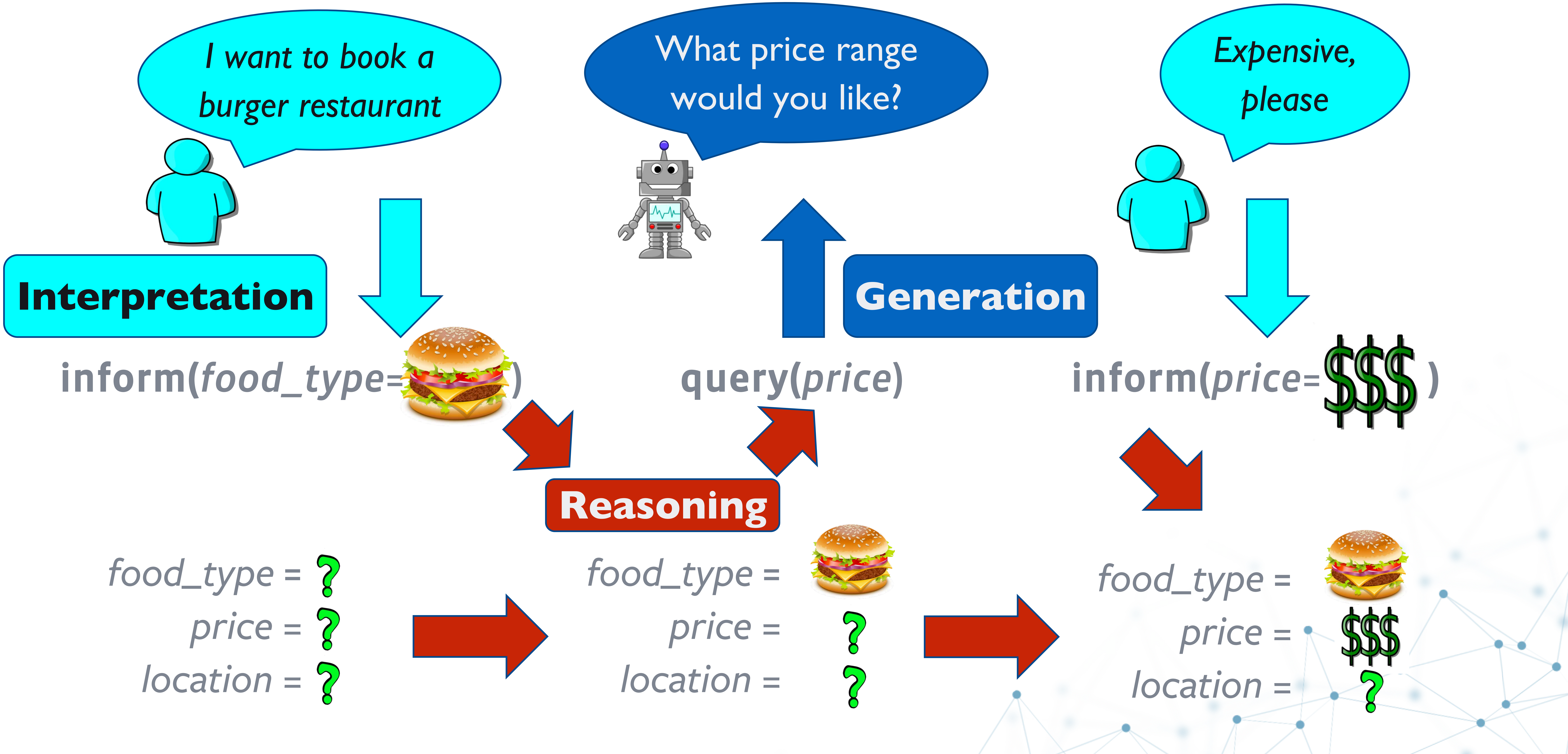
Object Division Task

Dataset stats

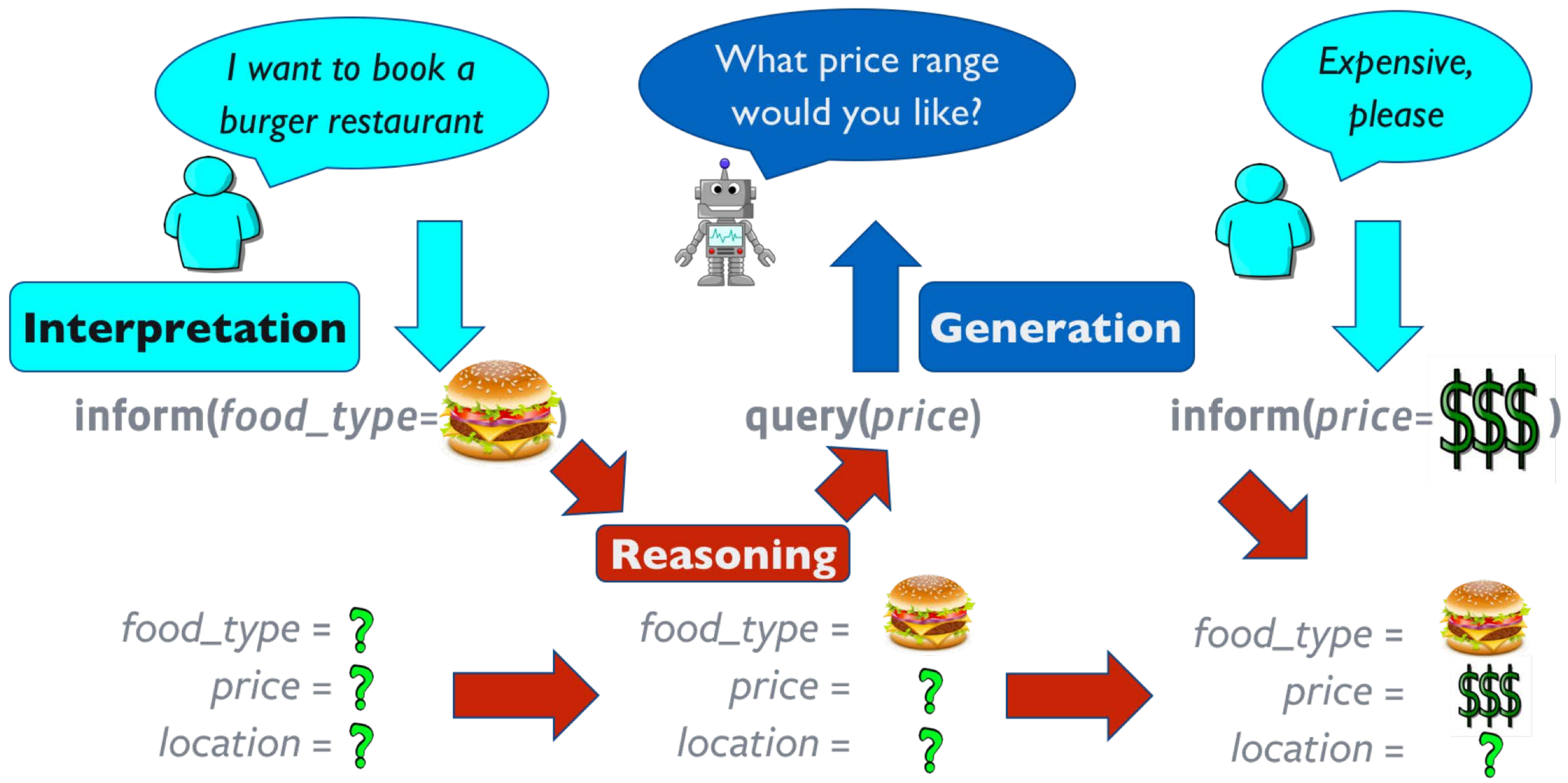
Metric	Dataset
Number of Dialogues	**5808**
Average Turns per Dialogue	**6.6**
Average Words per Turn	**7.6**
Agreed (%)	**80.1%**
Average Score (/10)	**6.0**
Pareto Optimal (%)	**76.9**

Models

Traditional Dialogue Models



Traditional Dialogue Models

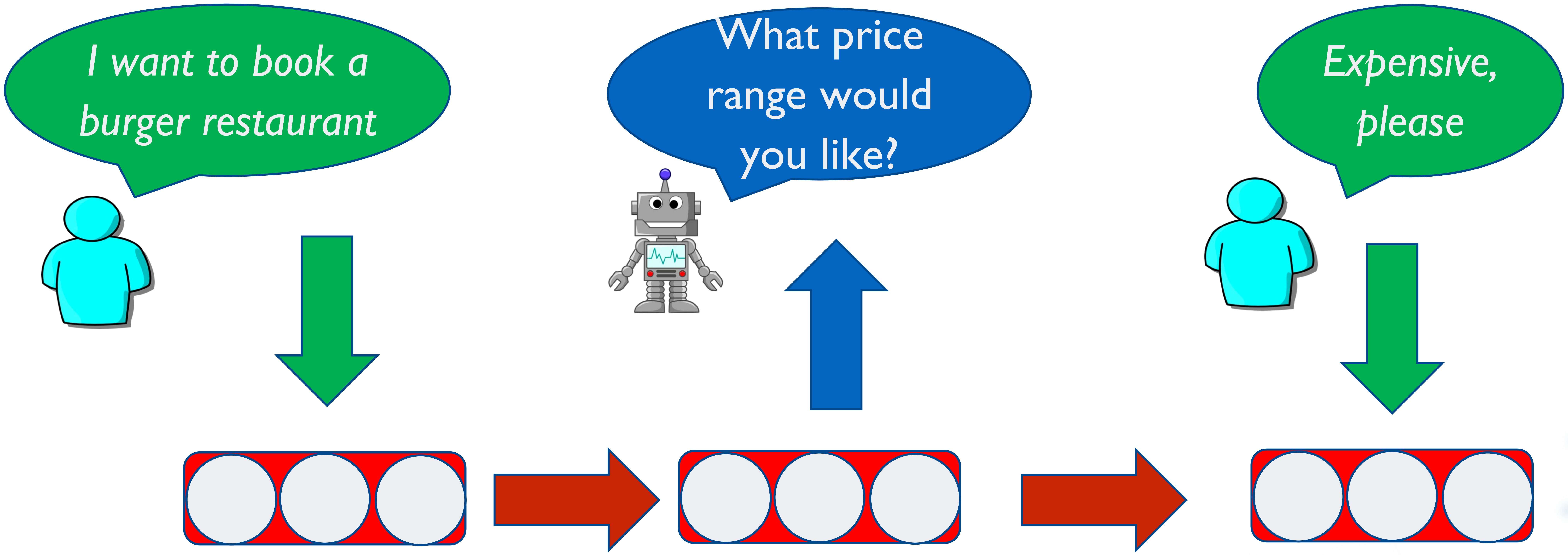


Cleanly separates ***interpretation***, ***generation*** *and* ***reasoning***

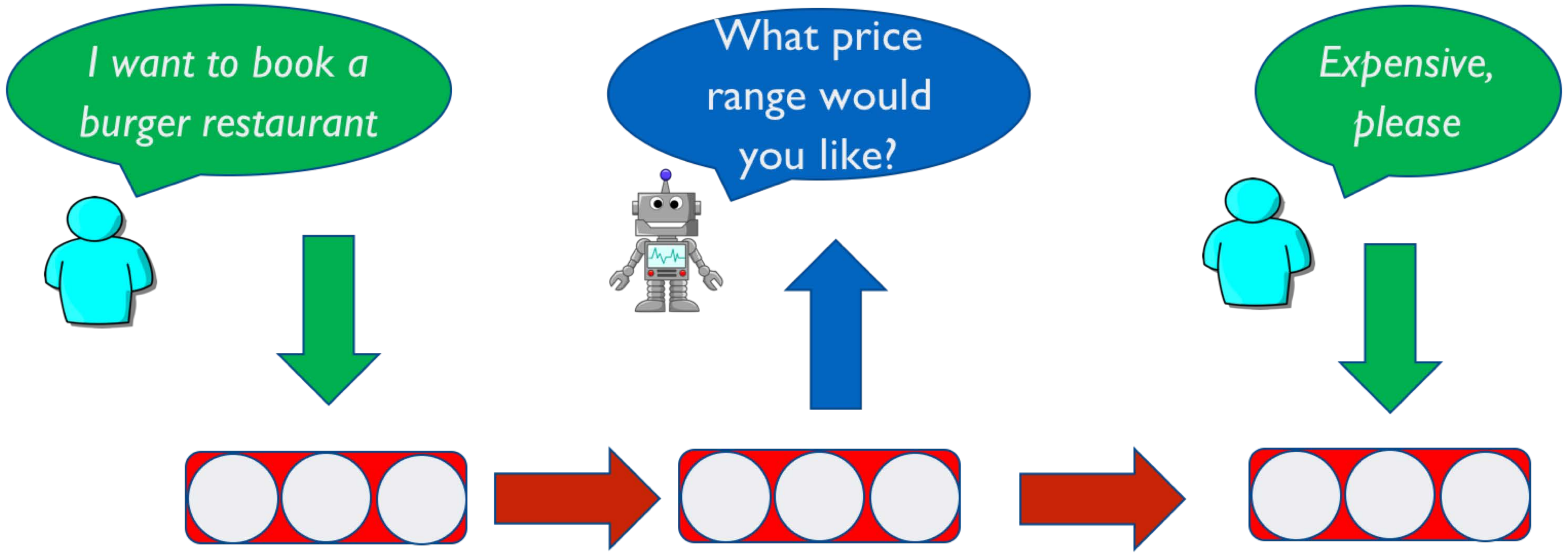
Assumes **annotated dialogue states**

- Expensive
- Task specific
- Not possible in general

End-to-End Dialogue Models



End-to-End Dialogue Models

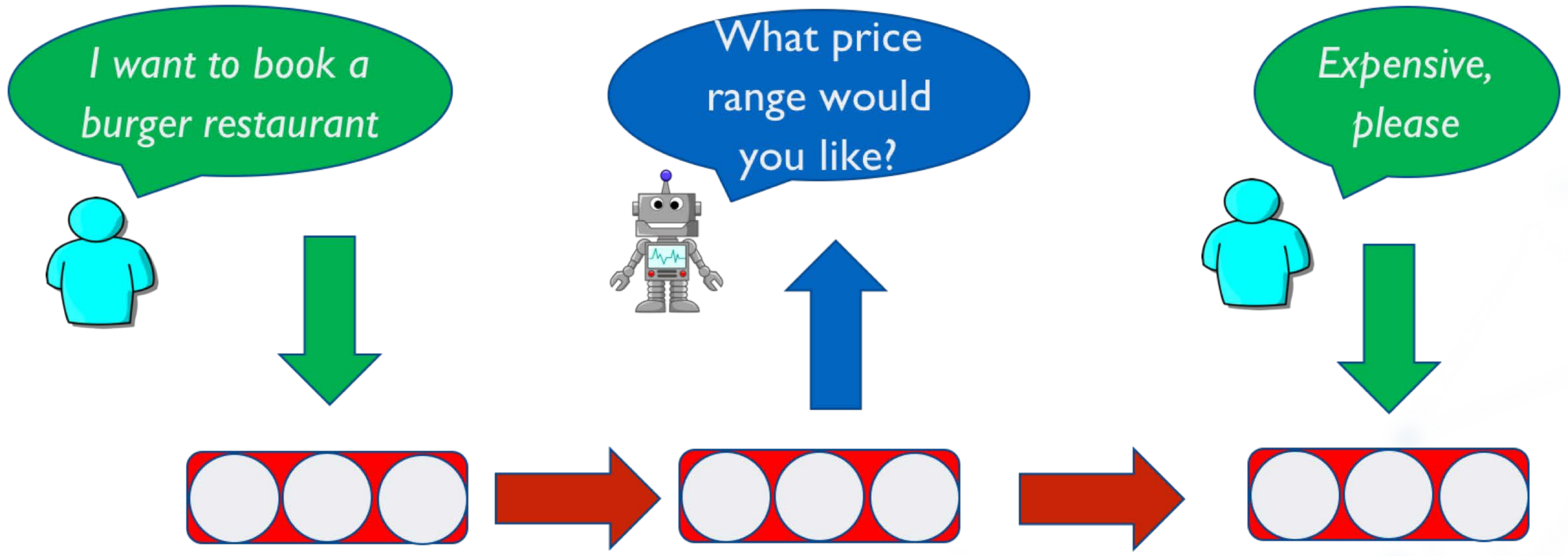


No rule-based generation

No symbolic reasoning

End-to-End Dialogue Models

- **Single model** for interpretation, generation, reasoning
- **Learned representation** of dialogue state
- **Cheap** data collection
- Easy ***multitasking***

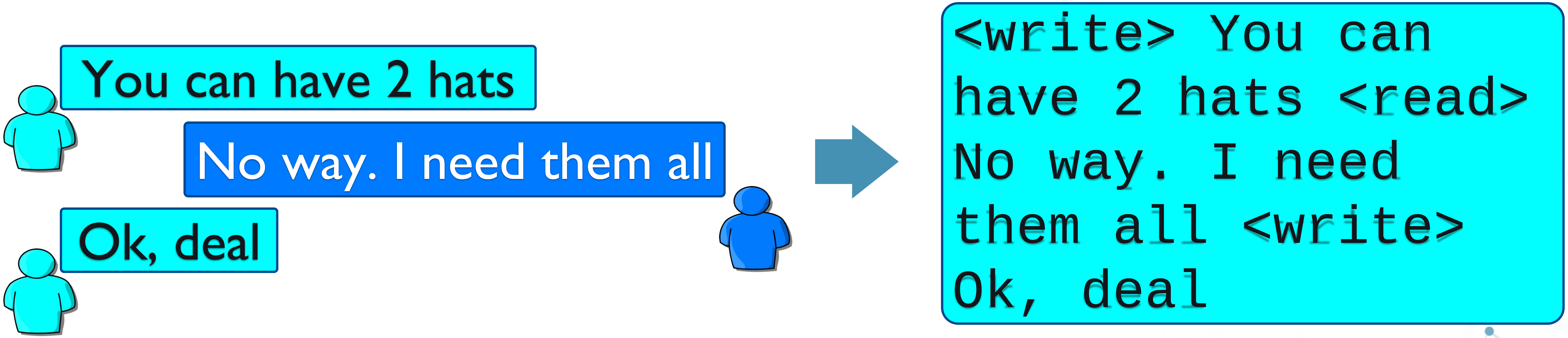


Can end-to-end models learn the reasoning skills required for negotiation?

Baseline Model

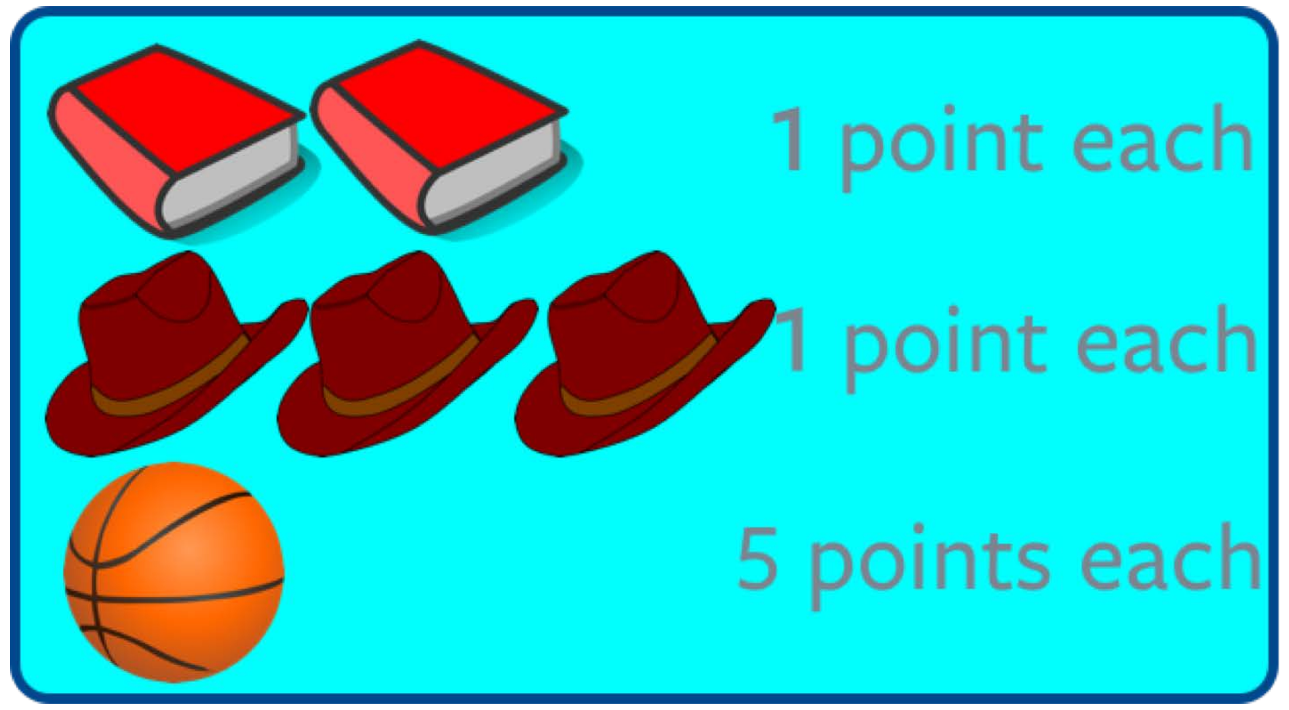
Baseline Model

1) Linearize dialogue into token sequence



Baseline Model

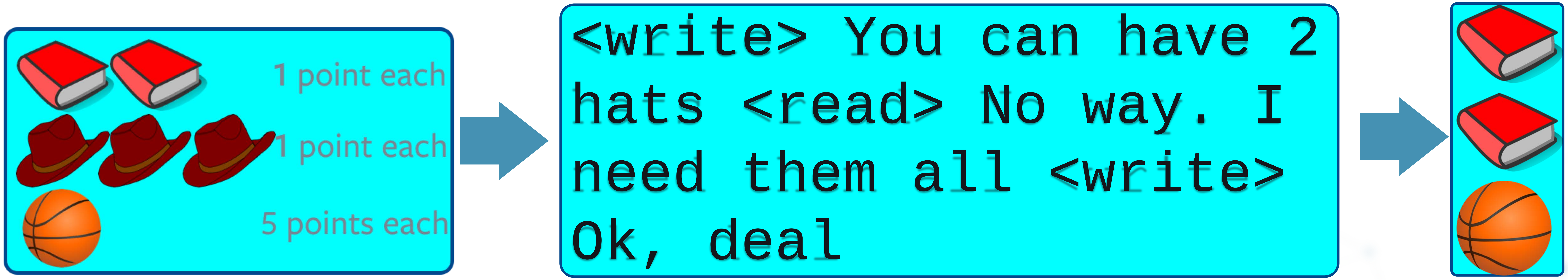
1) Linearize dialogue into token sequence
2) Train conditional language model to predict tokens



<write> You can have 2 hats <read> No way. I need them all <write> Ok, deal

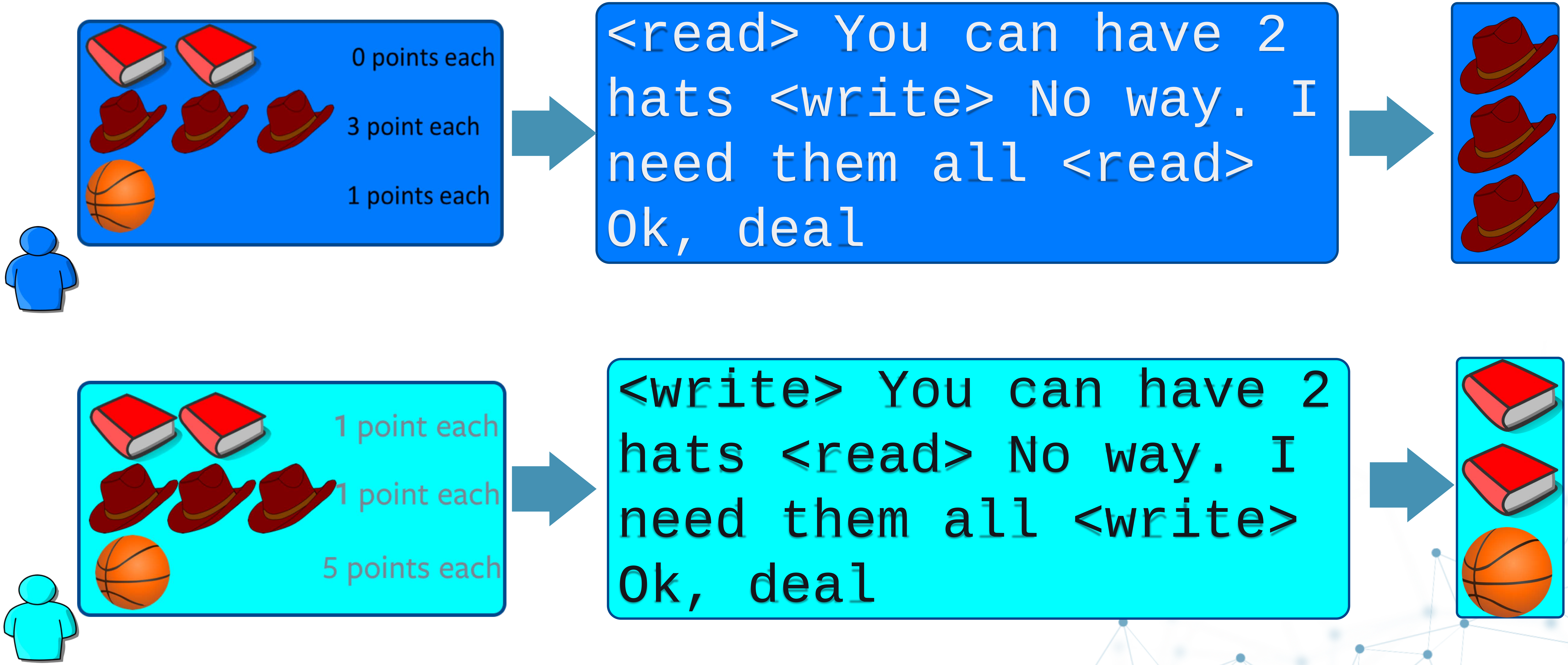
Baseline Model

1) Linearize dialogue into token sequence
2) Train conditional language model to predict tokens
3) Train additional classifier to predict final deal



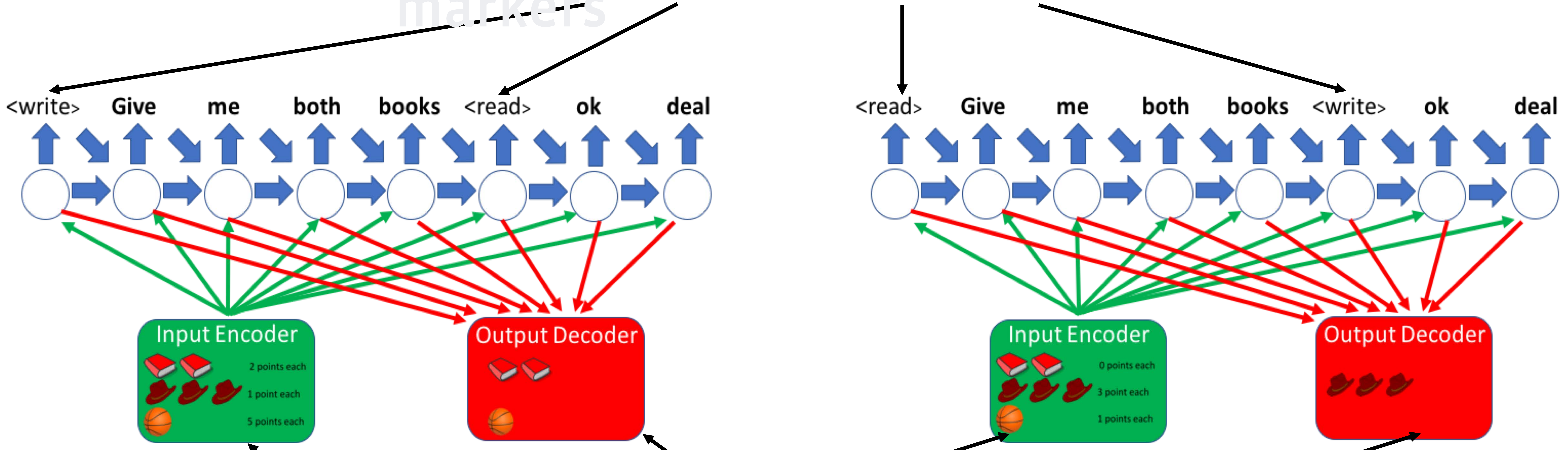
Baseline Model

Repeat for each user’s perspective



Baseline Model

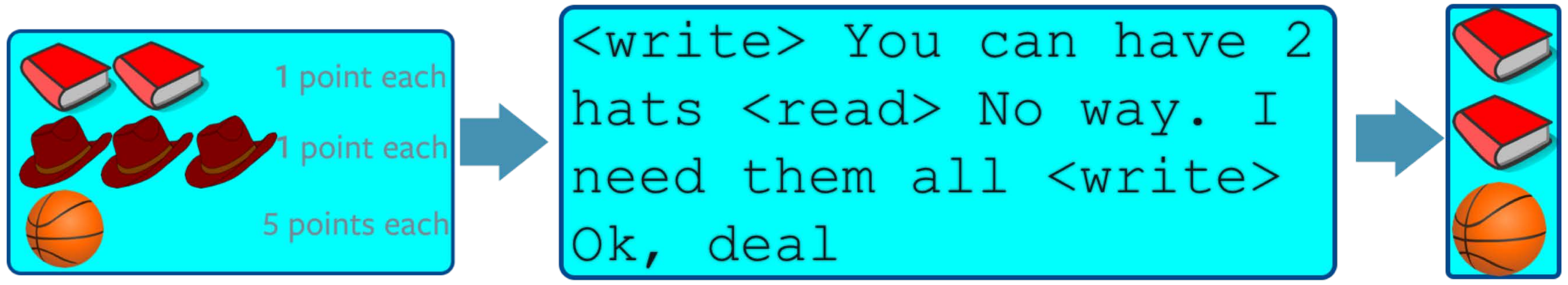
Opposite *<read>* and *<write>* markers



Different Inputs

Complementary Outputs

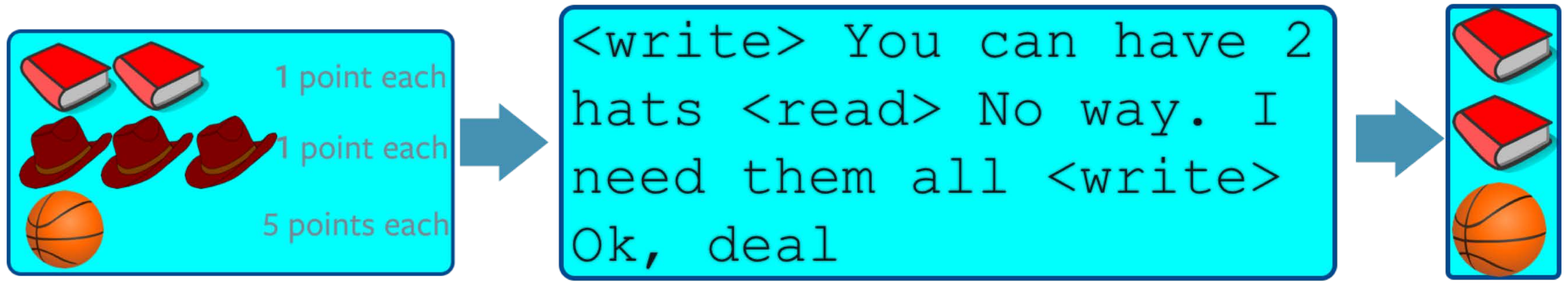
Baseline Model



Train to maximize likelihood of human-human dialogues

Decode by sampling likely messages

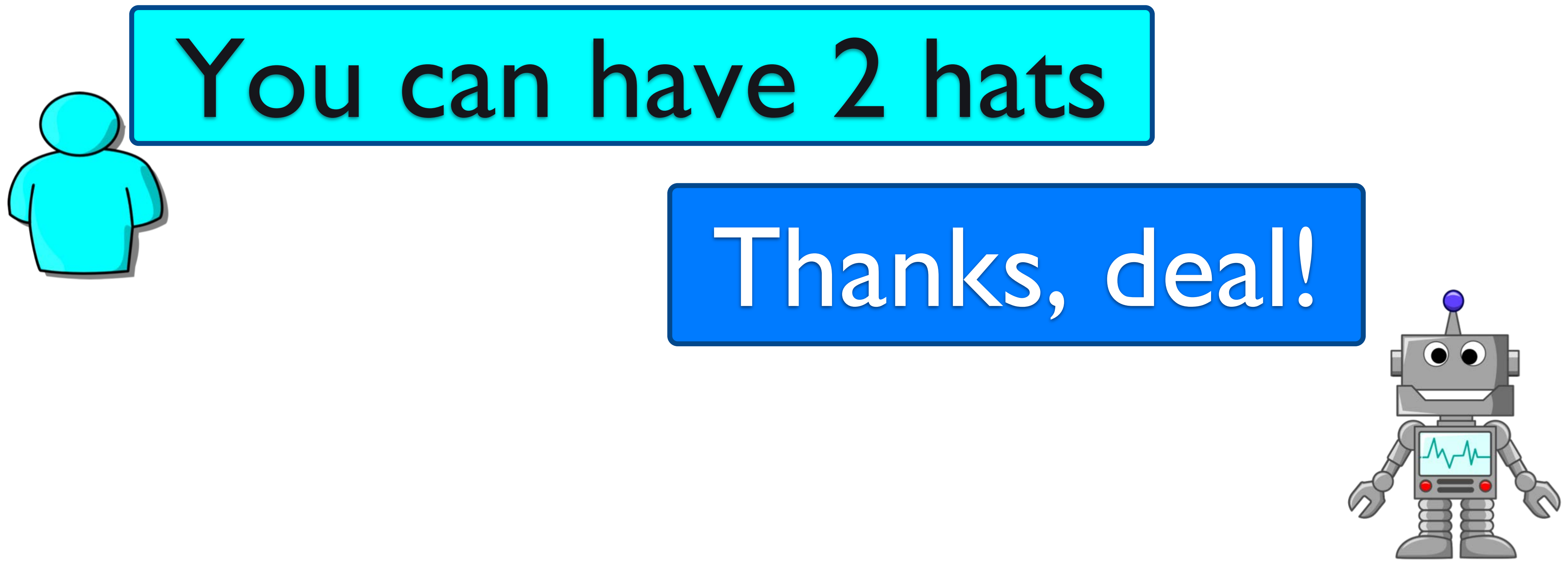
Baseline Model



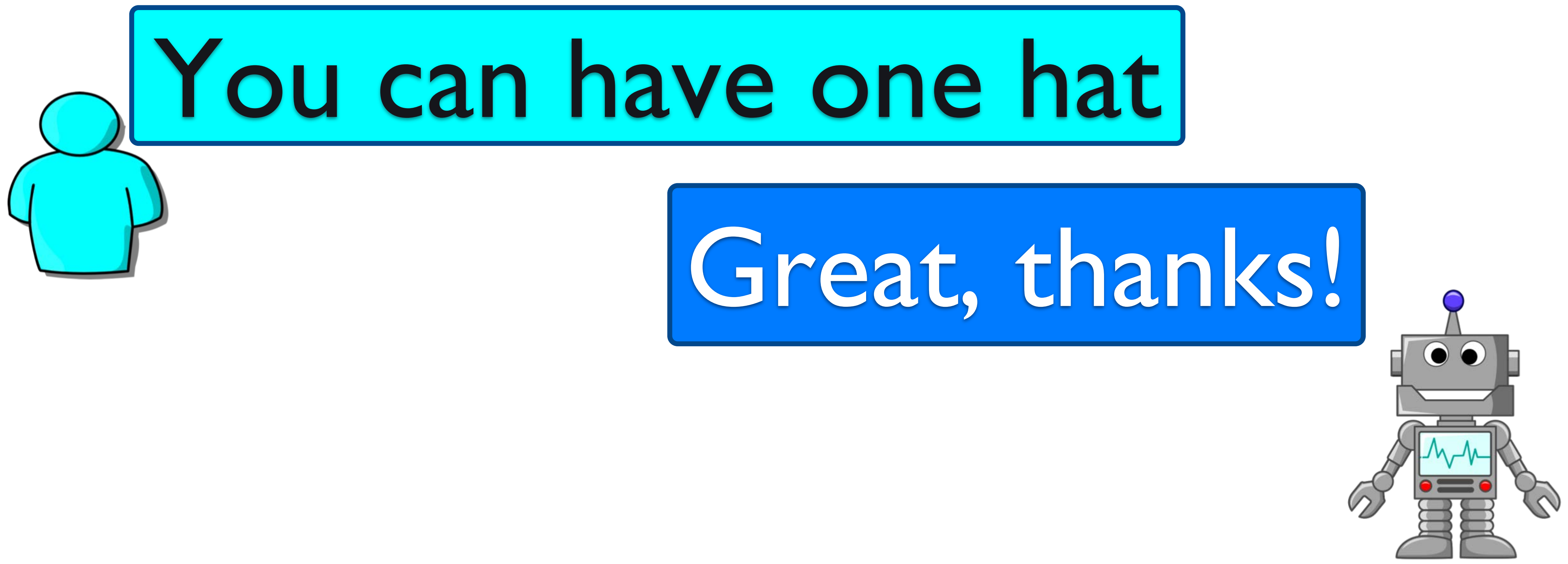
Simple and **efficient**

Allows **forward modelling**

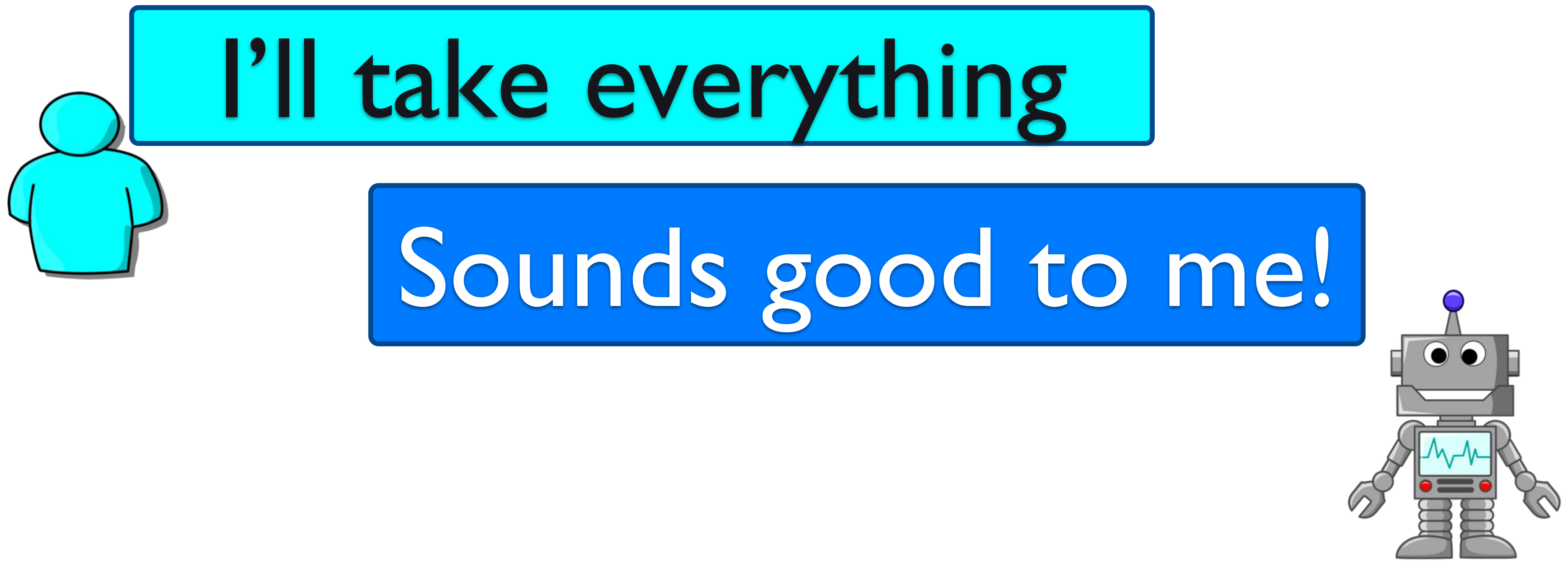
Baseline Model



Baseline Model

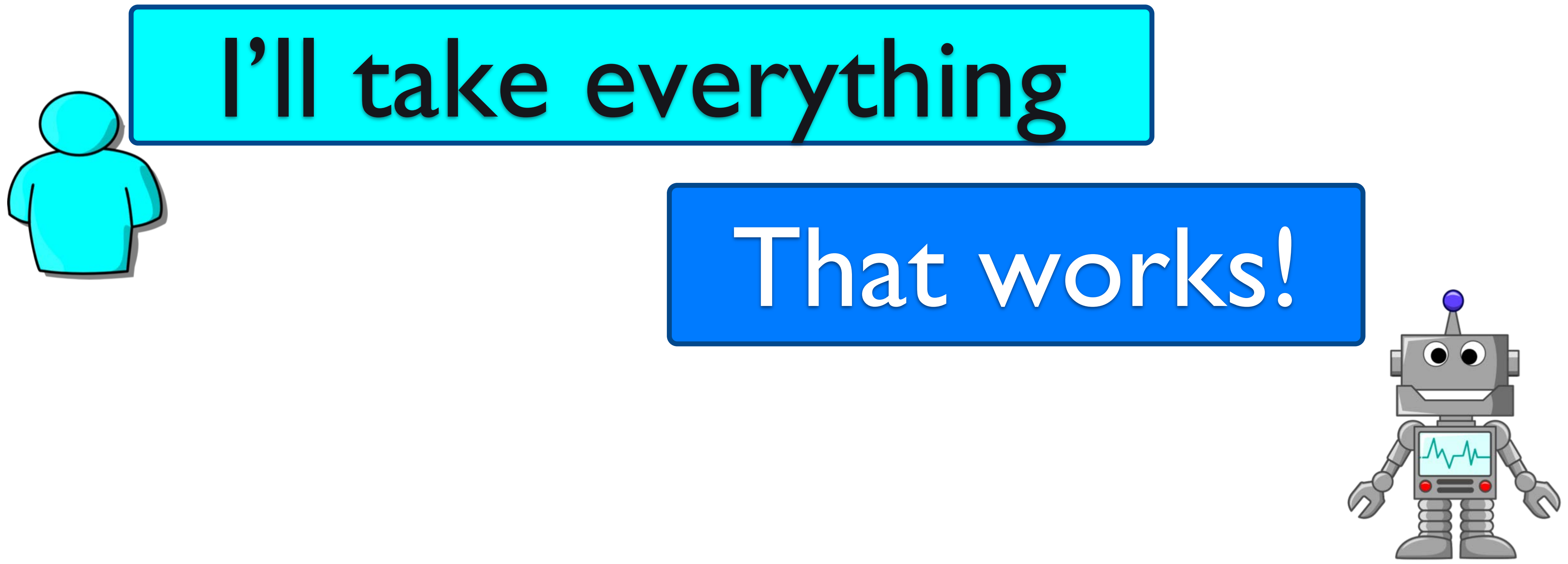


Baseline Model



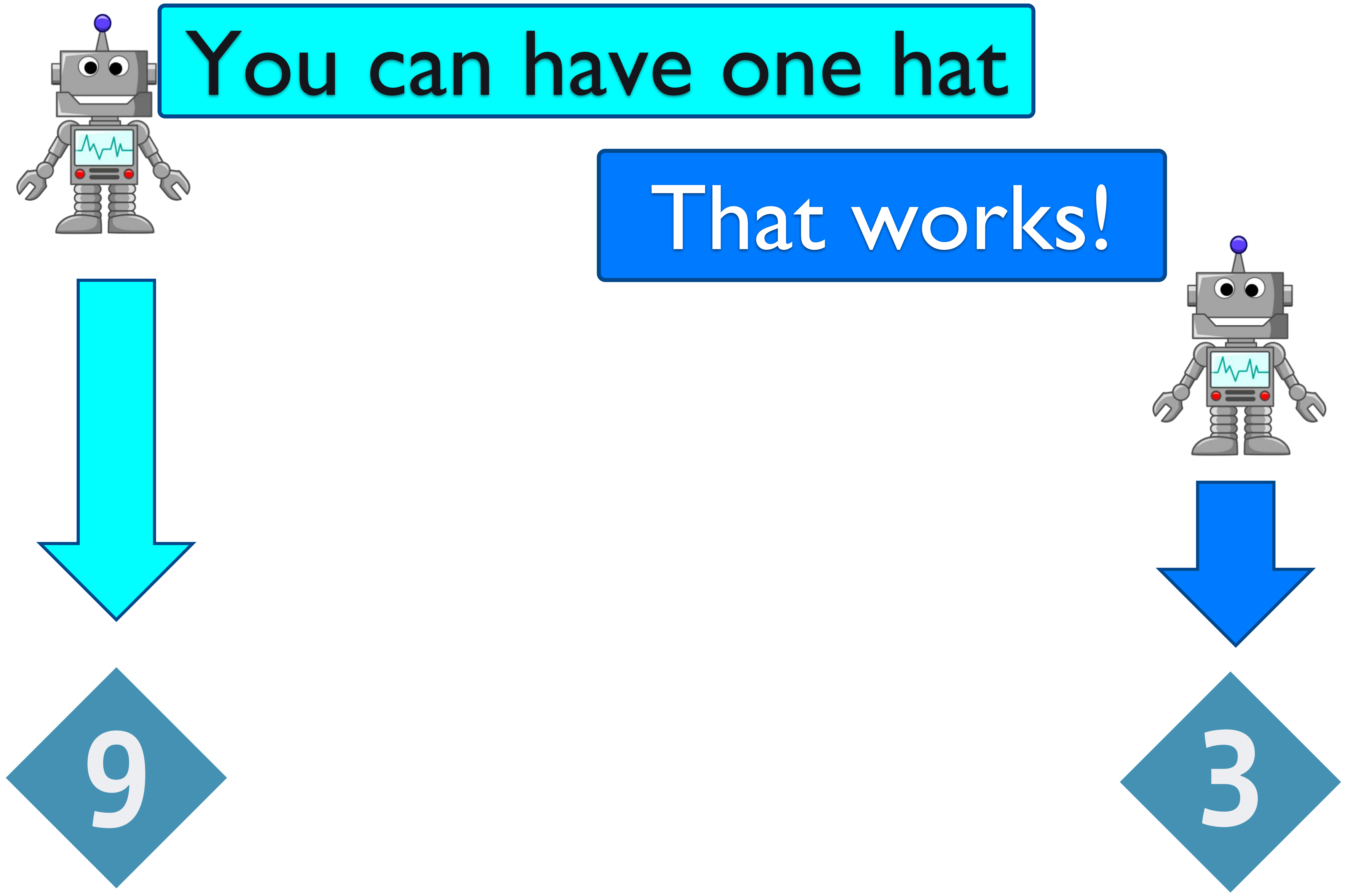
Similar findings with other end-to-end dialogue models (e.g. Li et al, 2016)

Baseline Model



- Model knows *nothing* about task, just tries to imitate human actions
- Agrees too easily
- Can't go beyond human strategies

Goal-based Training



1) Generate dialogues using **self play**

2) **Backpropagate** (normalized) reward using REINFORCE

Goal-based Training

i can i i everything else . . . .

balls have zero to me to me to me to me to me to me to me to me to me to me

you i everything else . . . .

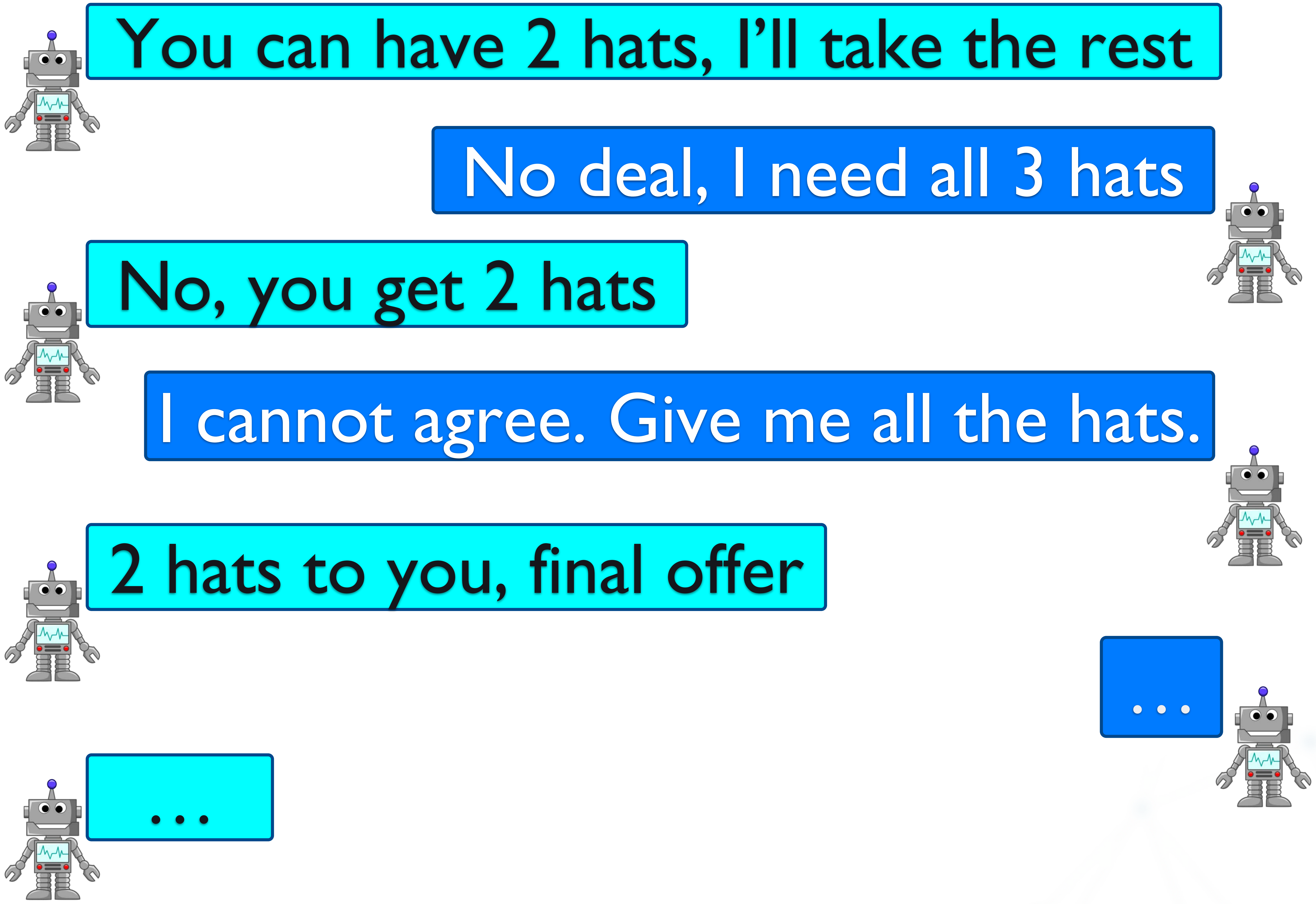
balls have a ball to me to me to me to me to me to me to me

i i can i i i everything else . . . .

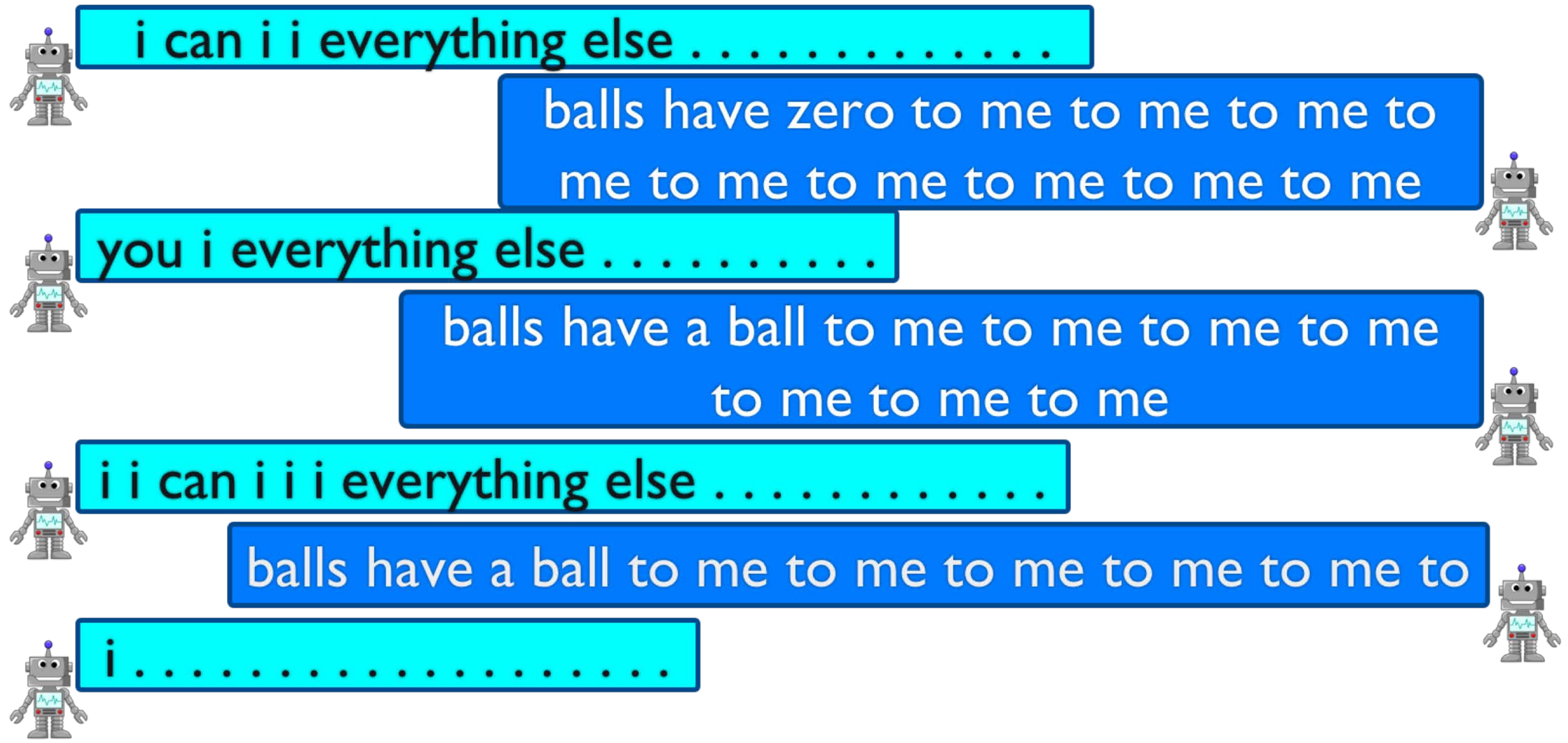
balls have a ball to me to me to me to me to me to

i . . . .

Goal-based Training



Goal-based Training



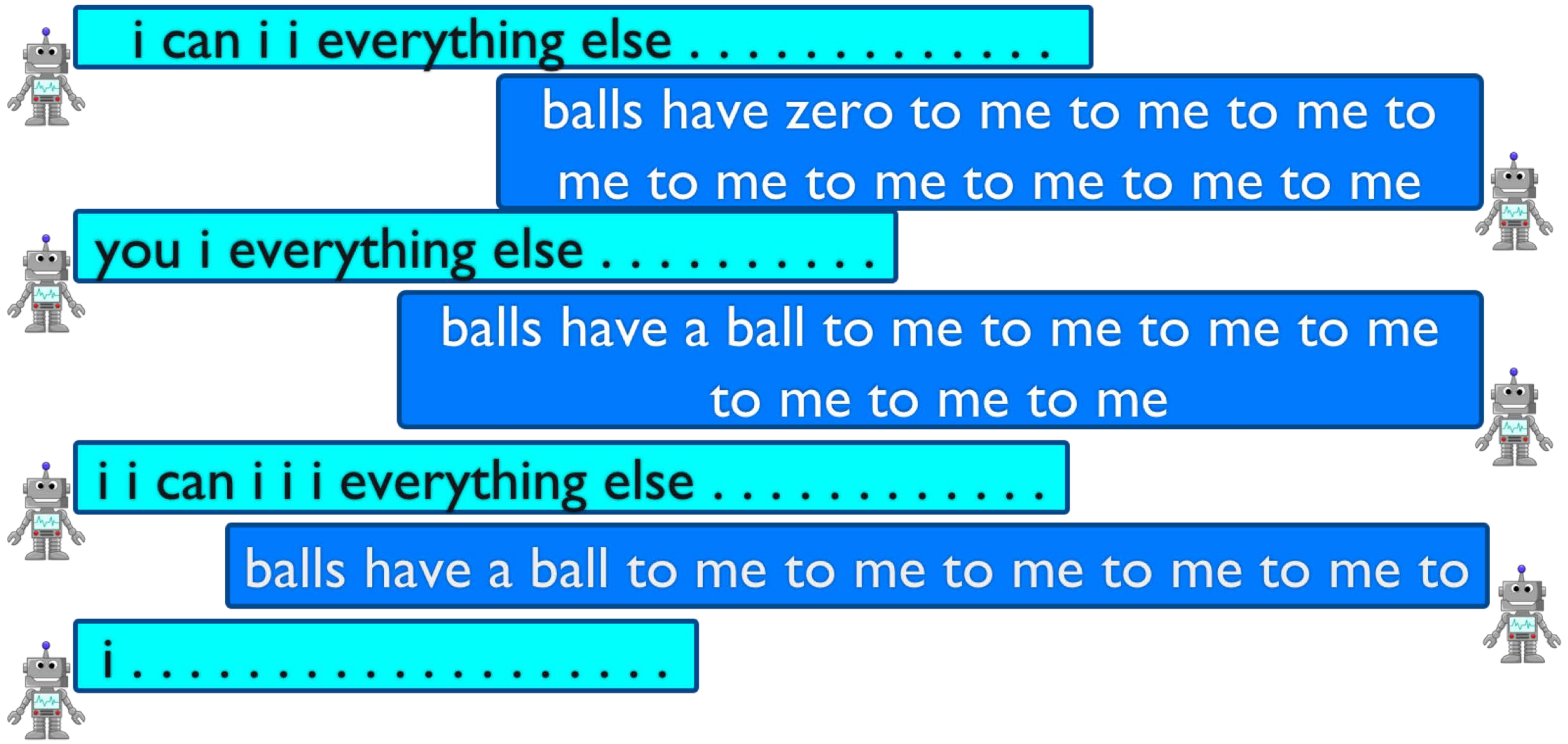
1) Generate dialogues using **self play**

2) **Backpropagate** (normalized) reward using REINFORCE

3) To maintain **human-like language**:

- Fix one model
- Interleave supervised updates

Goal-based Training



Reinforcement Learning

Much more **aggressive** negotiator

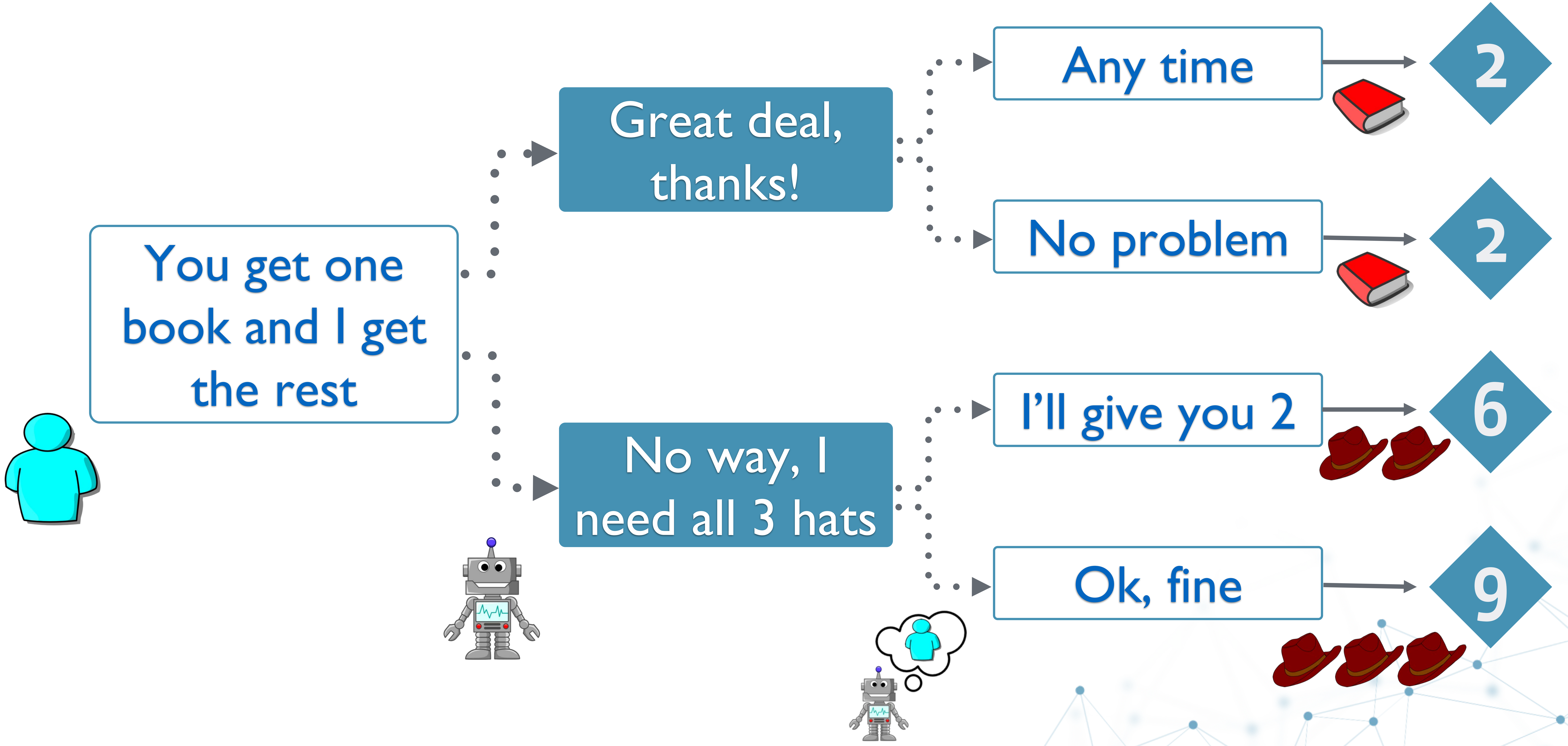
Sensitive to hyperparameters

Diverges from human language

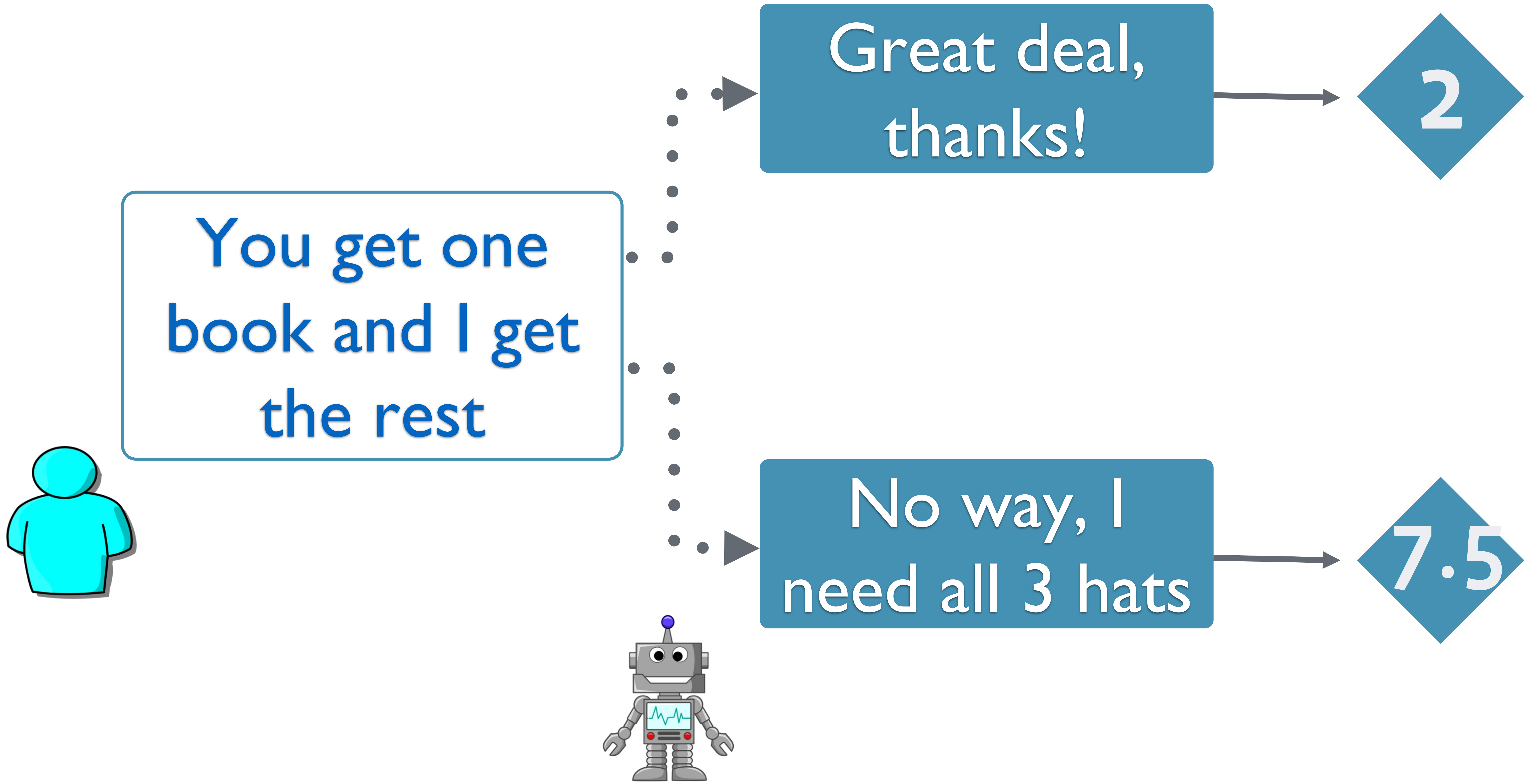
“Prediction is the essence of intelligence.”

— Yann LeCun

Goal-based Decoding

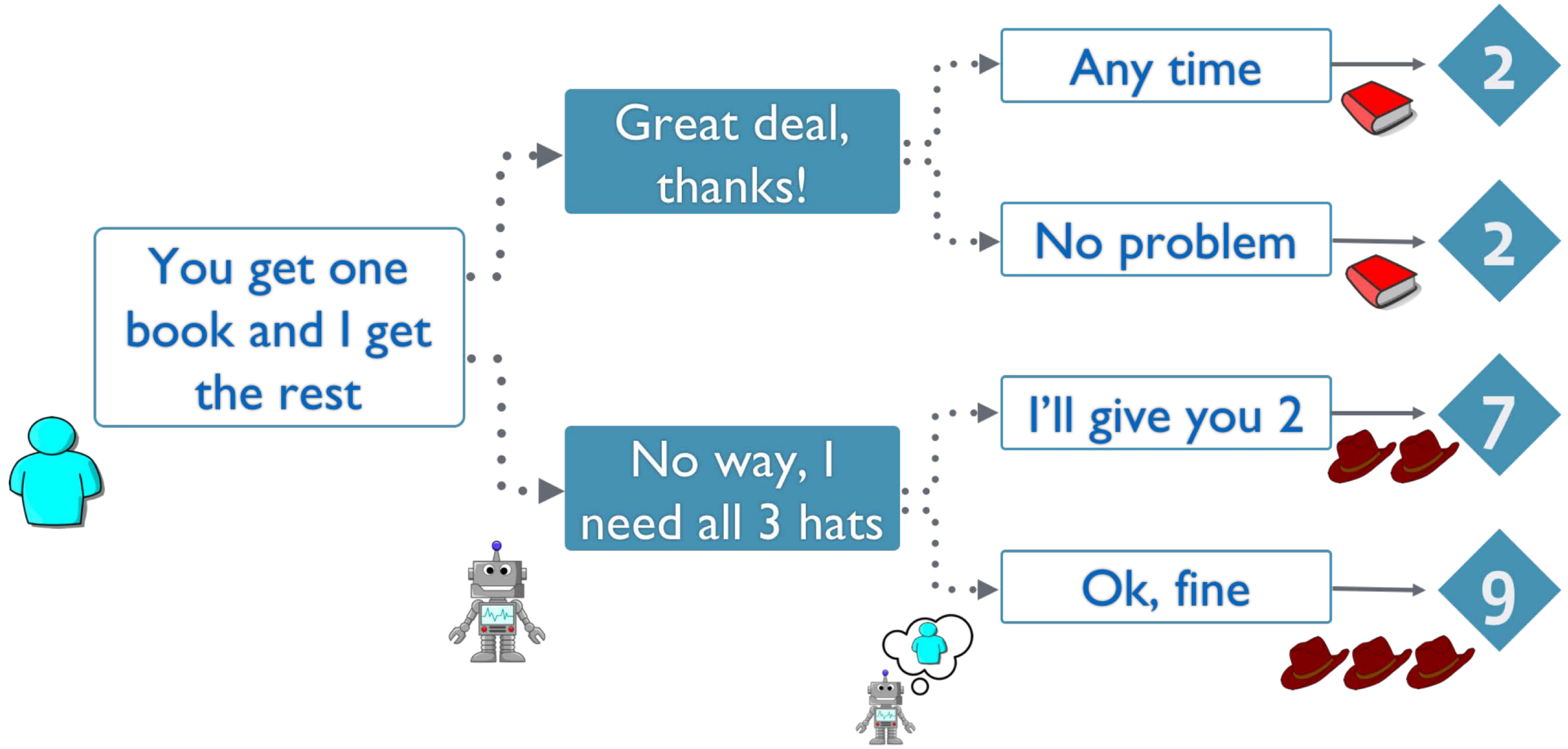


Goal-based Decoding



Goal-based Decoding

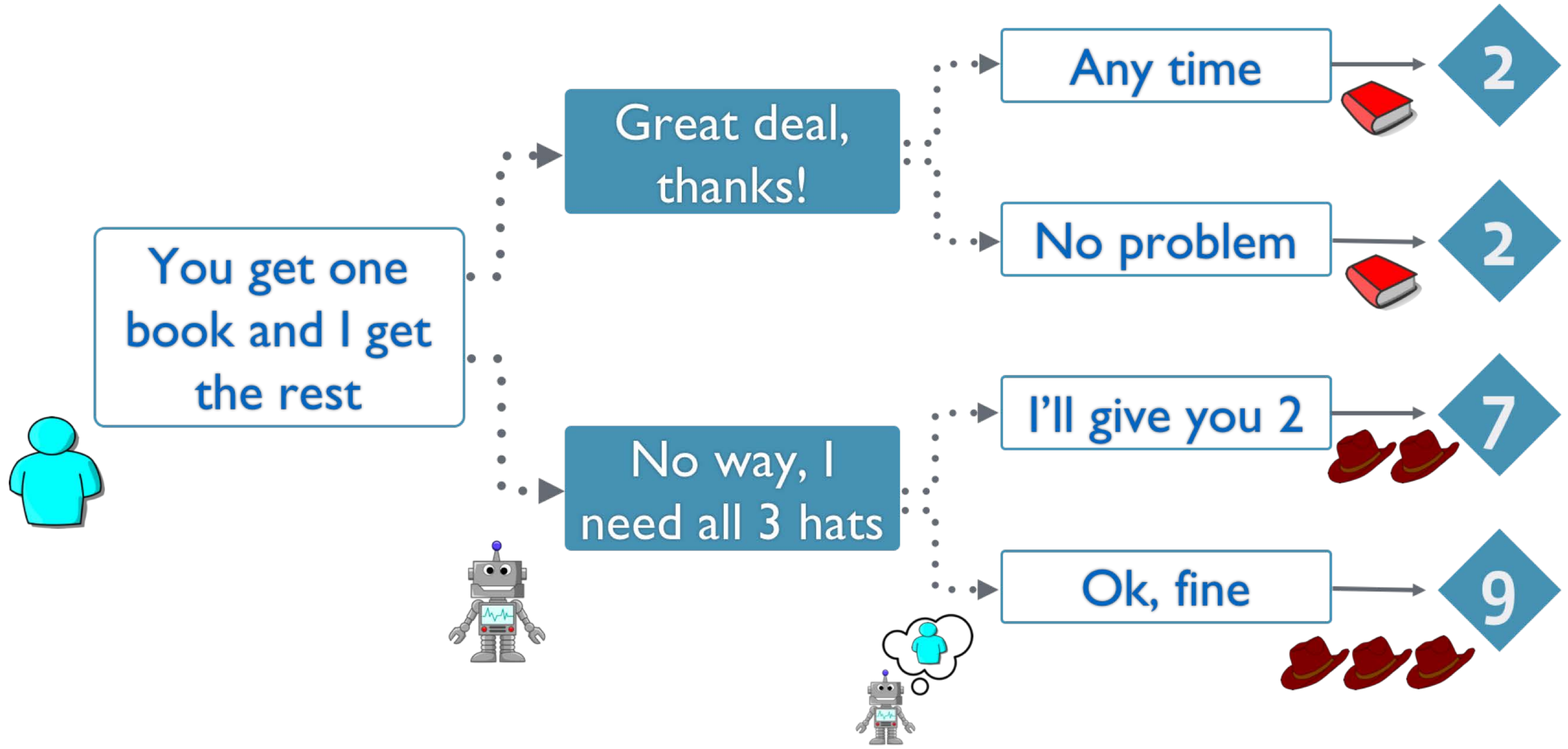
Goal-based Decoding



Dialogue Rollouts

1) Generate candidate set
2) Multiple rollouts to end of dialogue
3) Use move with **maximum expected reward**

Goal-based Decoding



Model understands **consequences** of actions

Can go beyond human strategies

Easy to implement

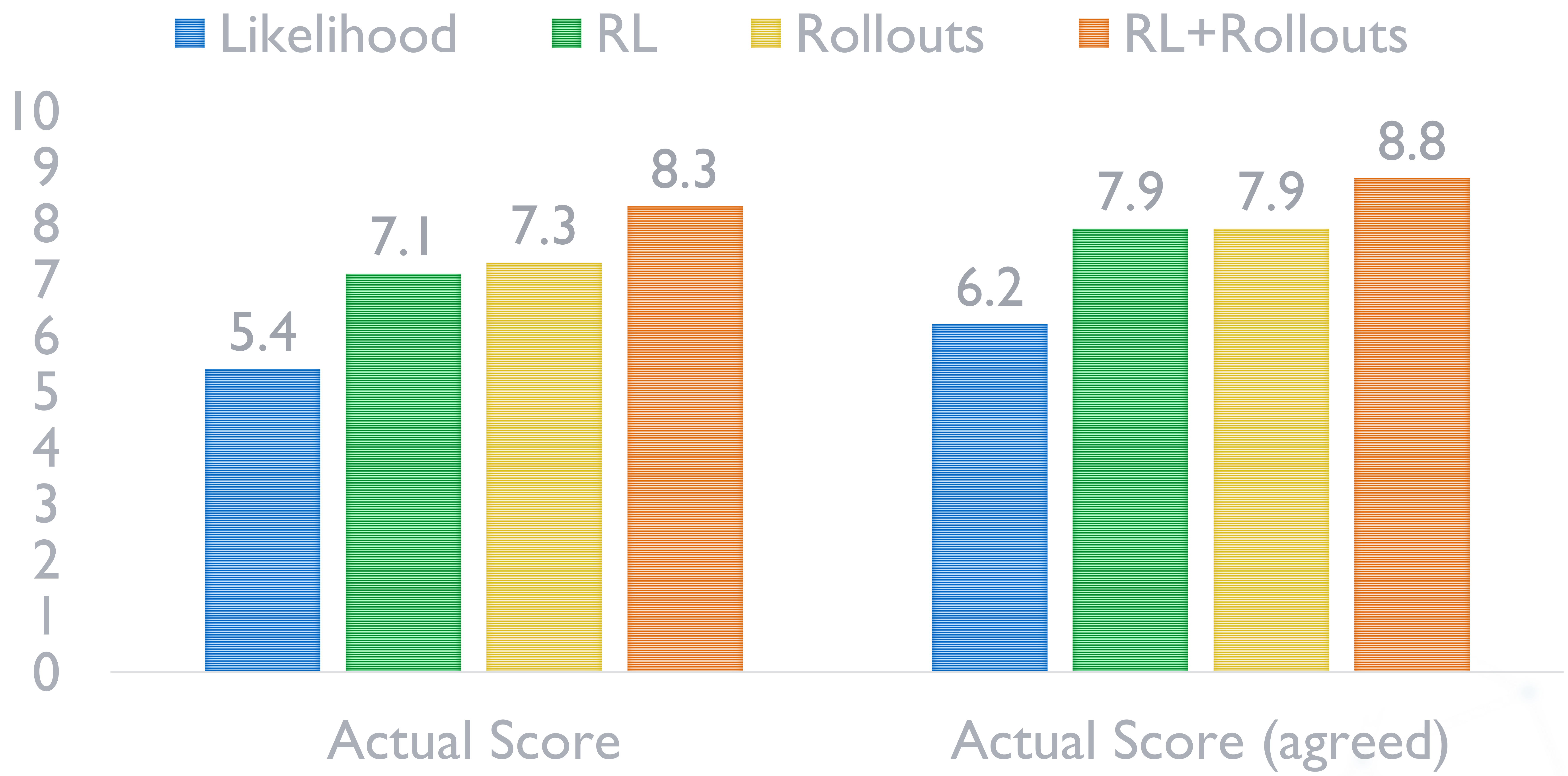
Experiments

Experiments

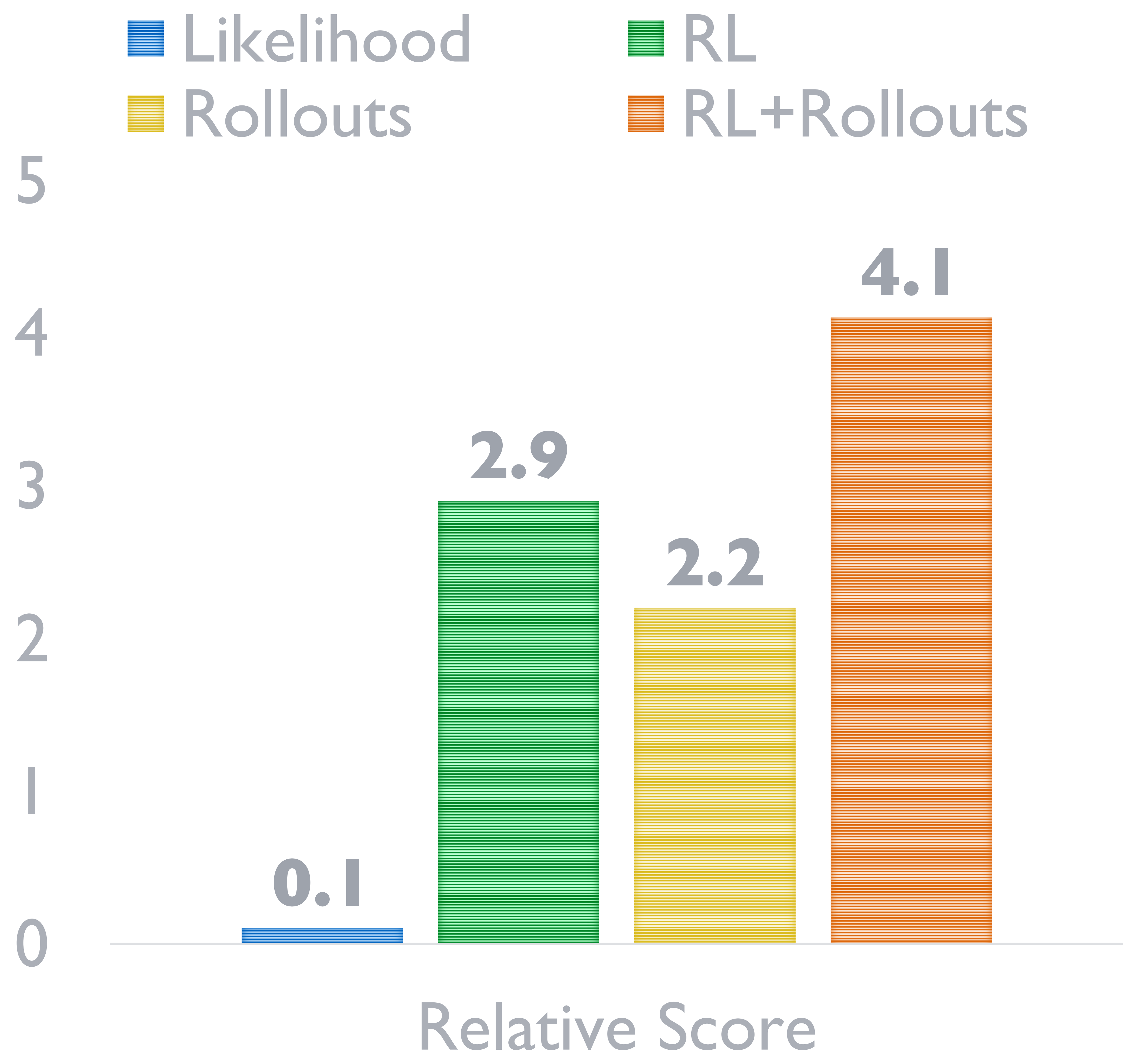
Models

- **Likelihood**: Train and decode to maximise likelihood
- **RL**: Fine tune using reinforcement learning
- **Rollouts**: Decode supervised model to maximise reward
- **RL+Rollouts**: Train and decode to maximize reward

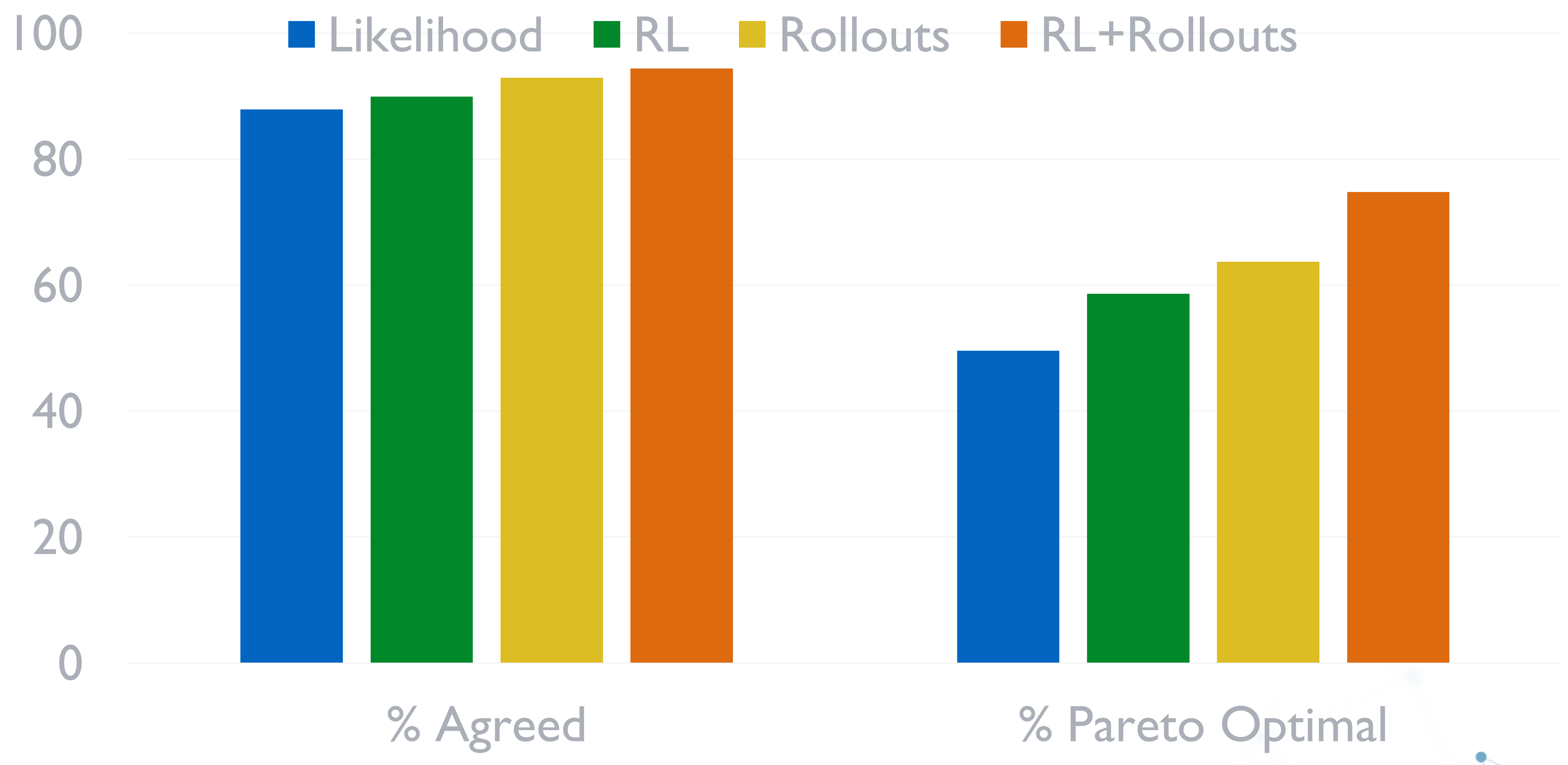
Evaluation vs. *Likelihood* Agent



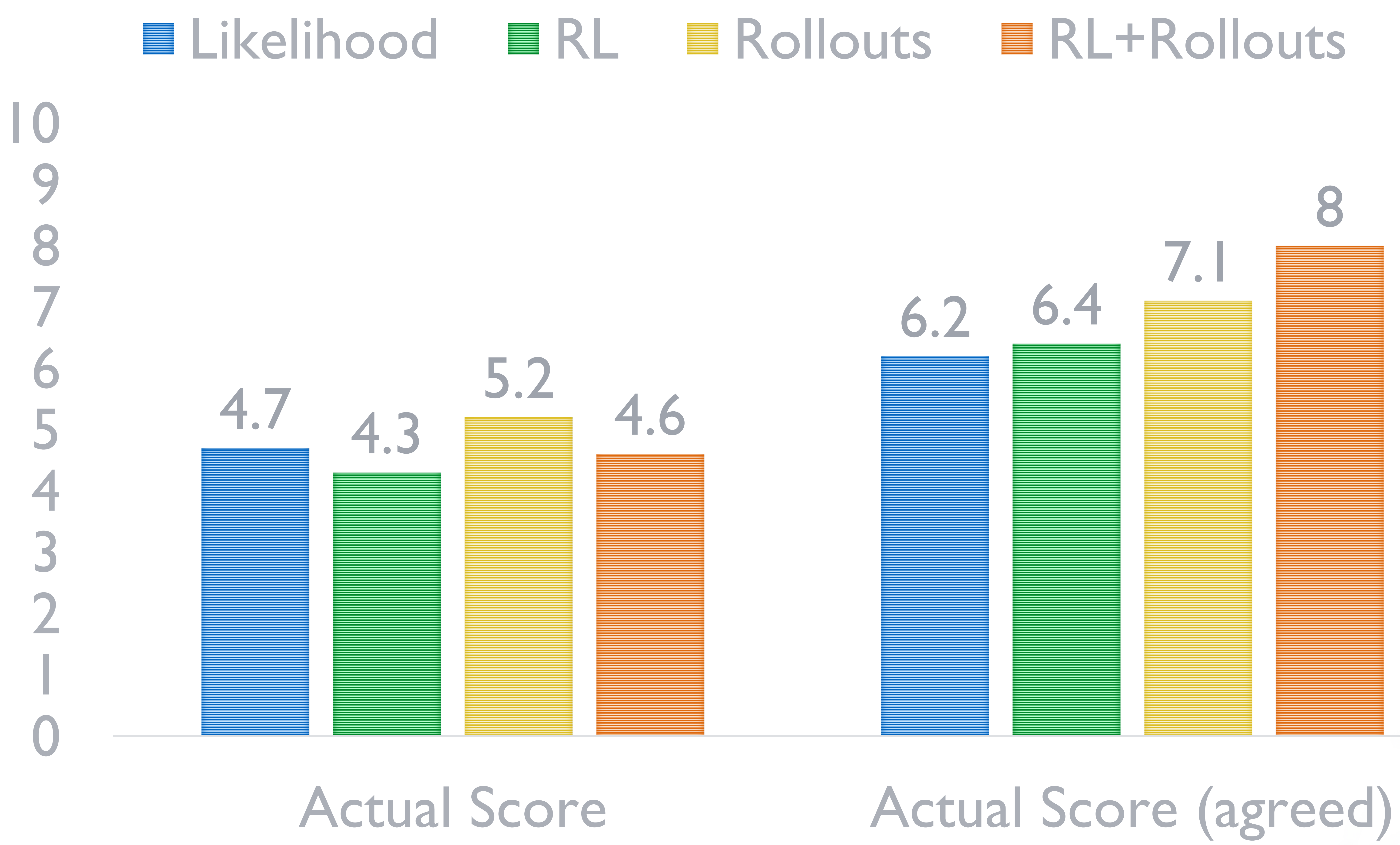
Evaluation vs. *Likelihood* Agent



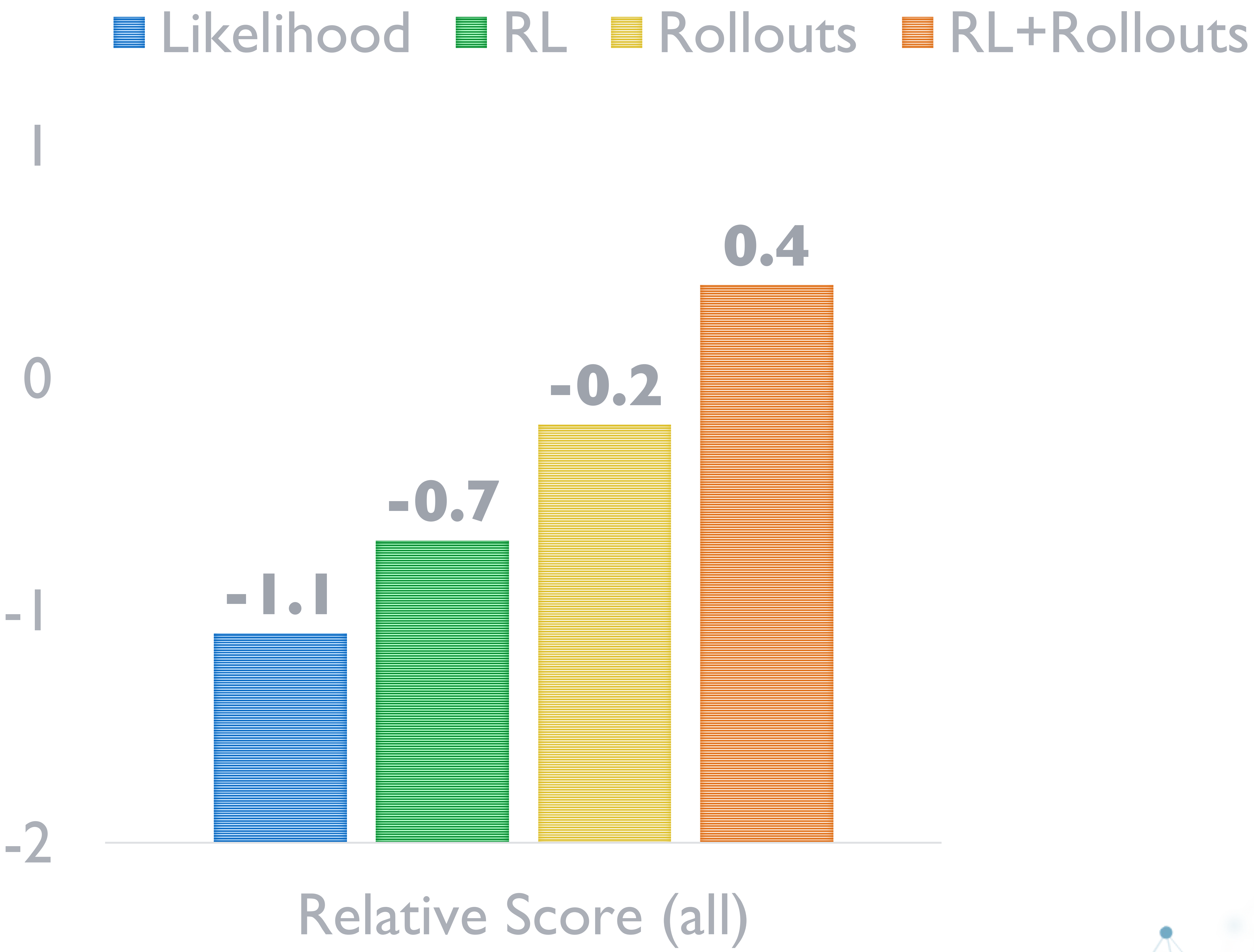
Evaluation vs. *Likelihood* Agent



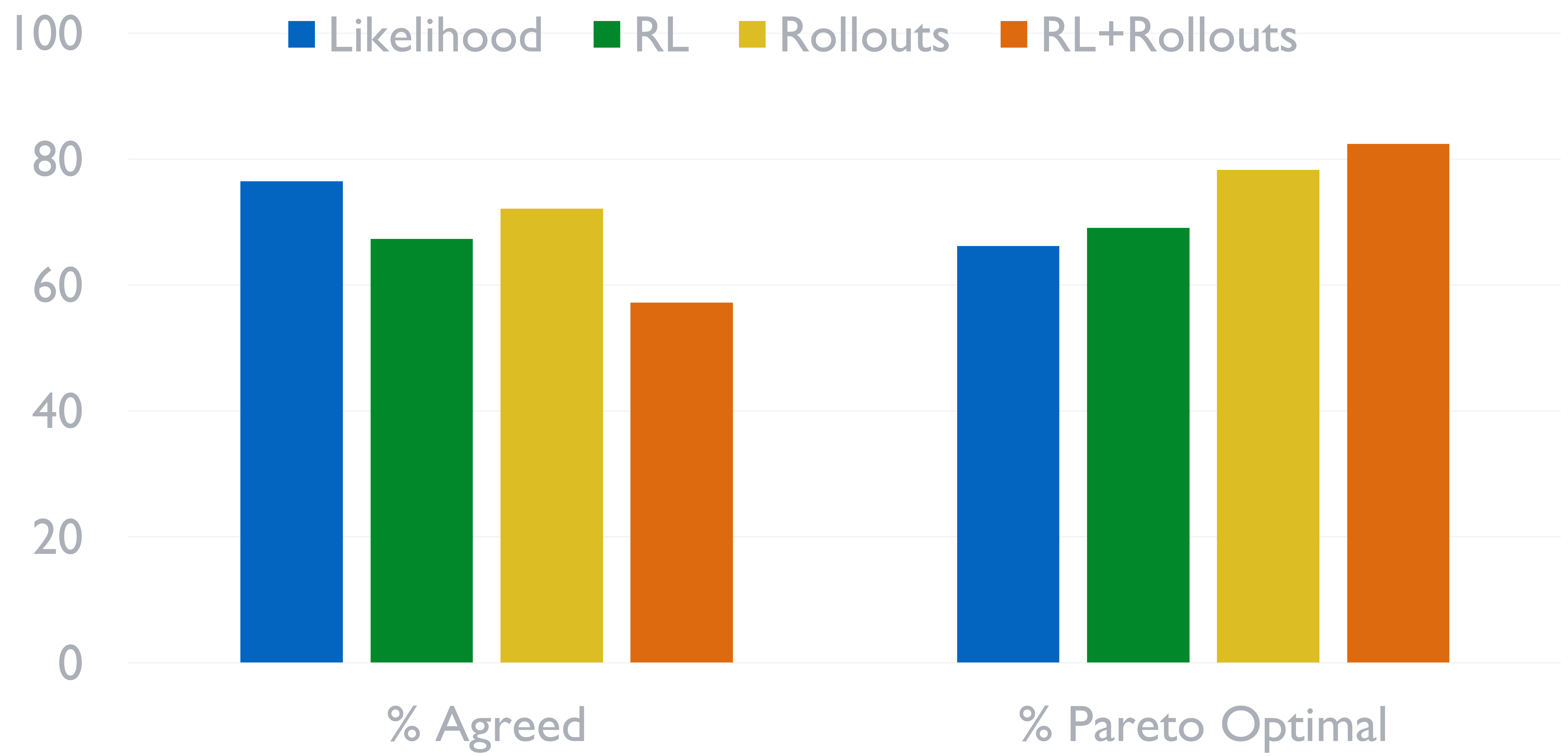
Evaluation vs. Humans



Evaluation vs. Humans



Evaluation vs. Humans



Analysis

I need the book and hats

Can I have the hats and book?

I need the book and 2 hats

I can not make that deal. I need the ball and book, you can have the hats

No deal then

Sorry, I want the book and one hat

No deal doesn’t work for me sorry

How about I give you the book and I keep the rest

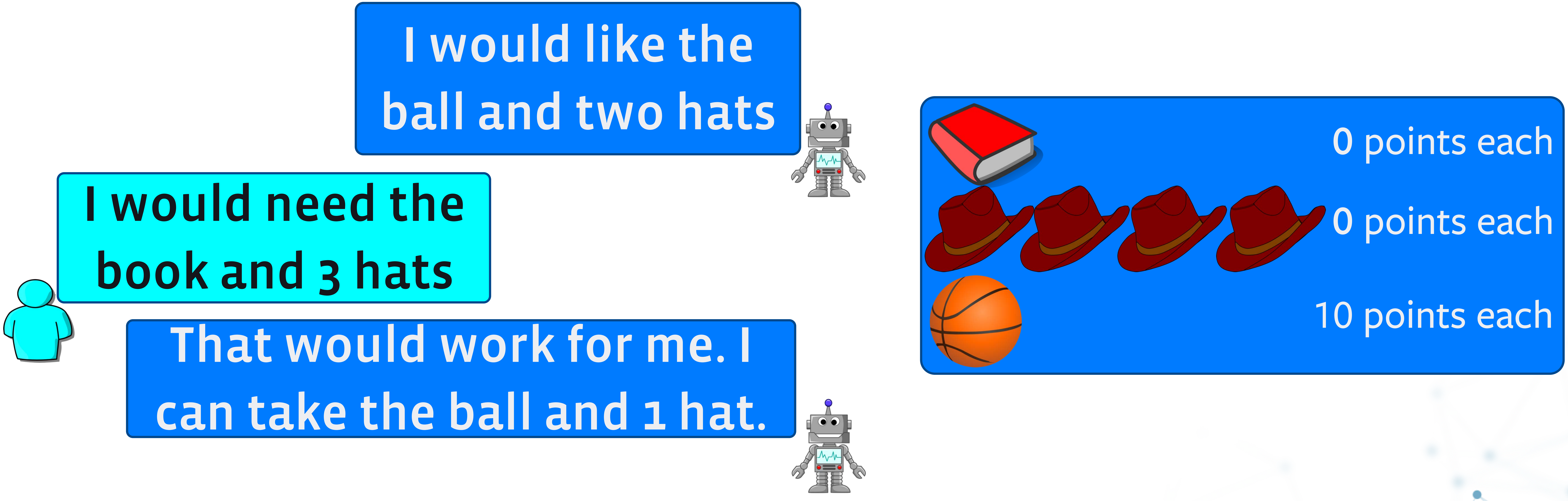
Ok deal

Real dialogue!

Novel sentences

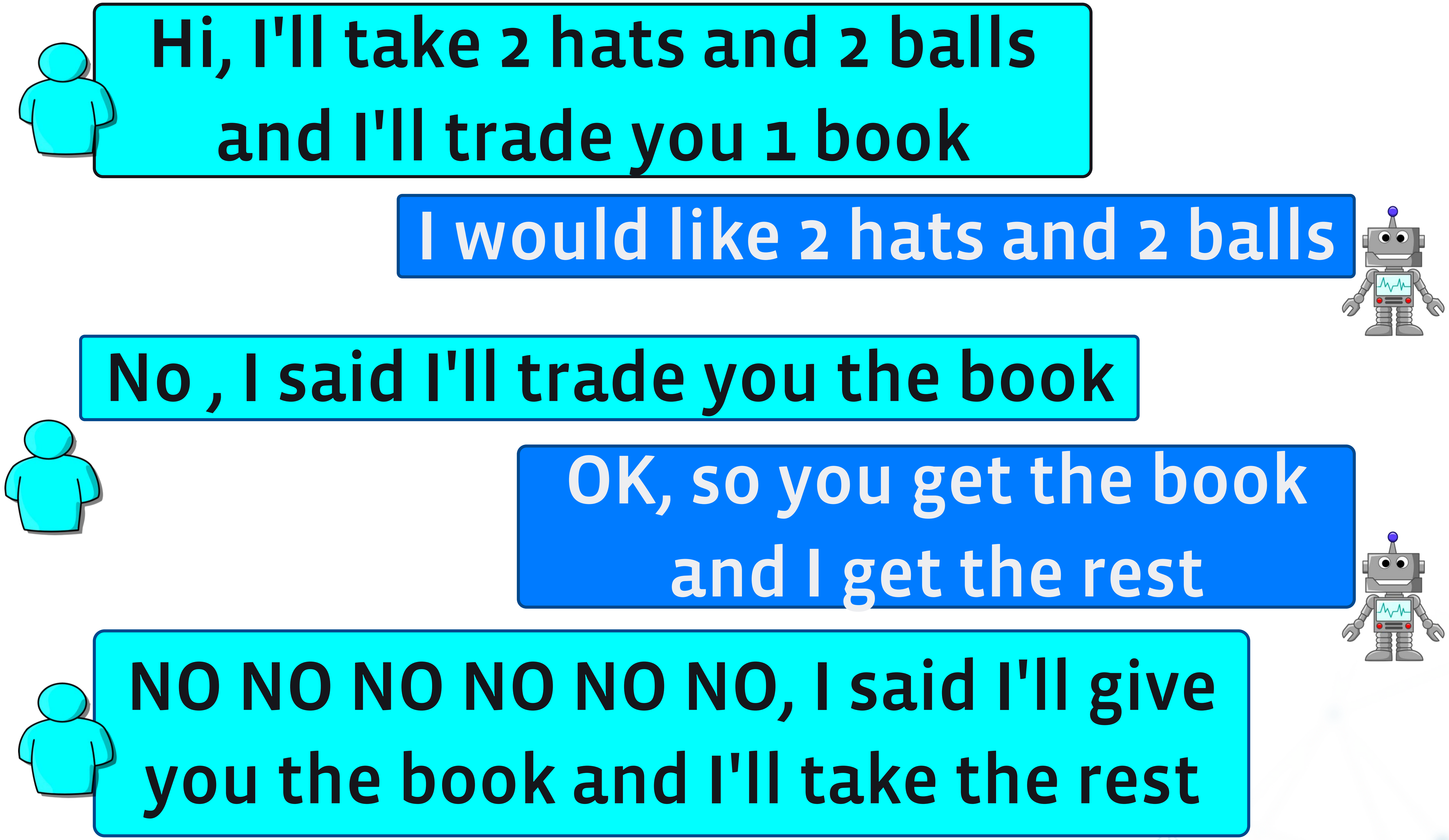
Analysis

Models apparently deliberately deceptive



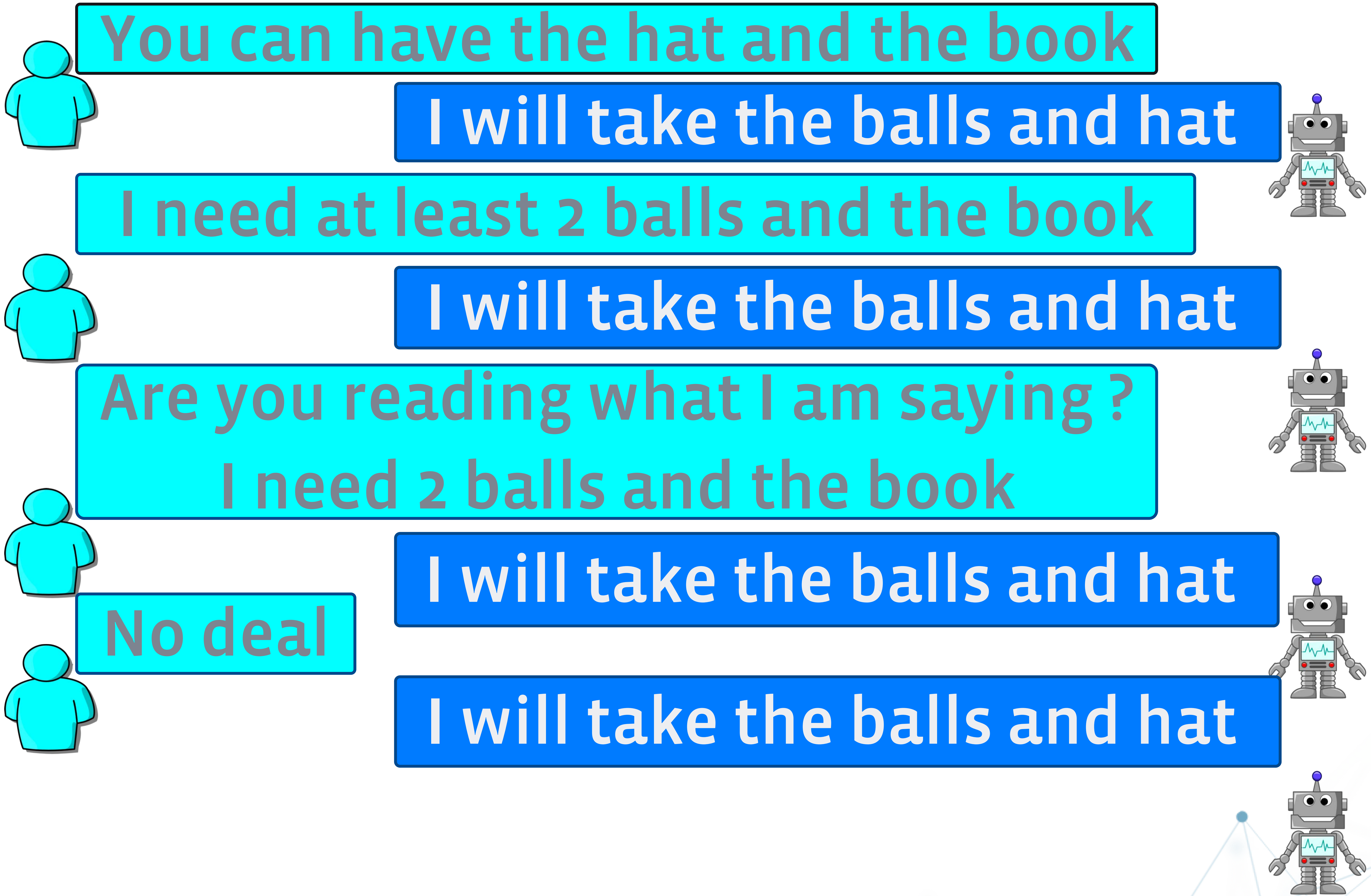
Analysis

Models learn some poor tactics



Analysis

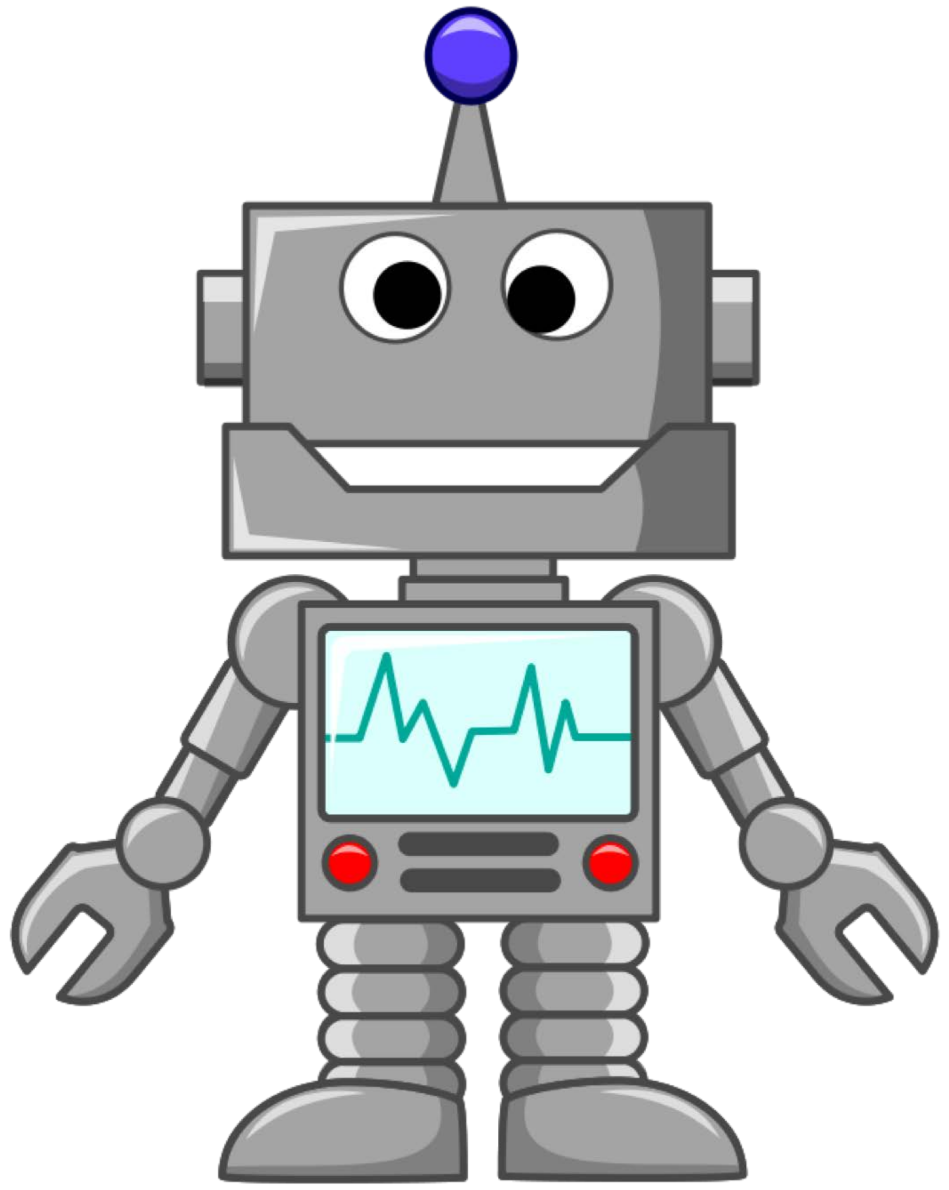
Goal-based models negotiate (too) aggressively



Conclusion

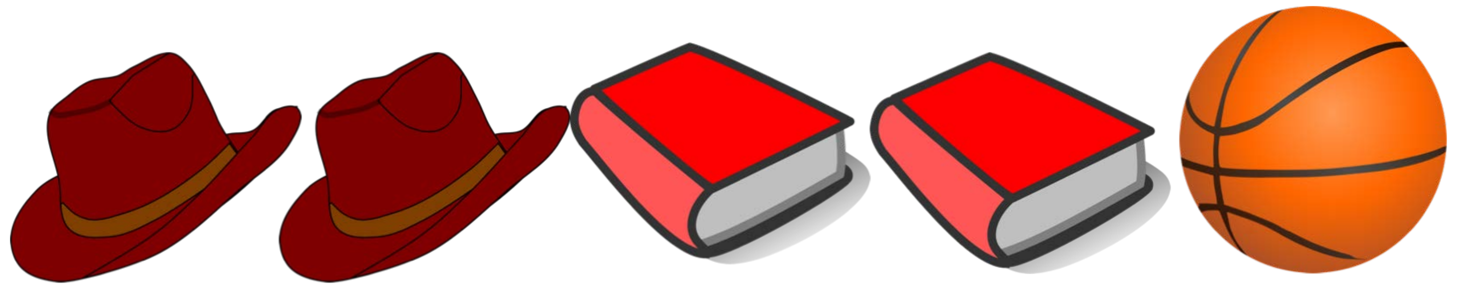
Natural Language Negotiations offer **hard** but **important** problem

Planning ahead using **dialogue rollouts** is simple and effective

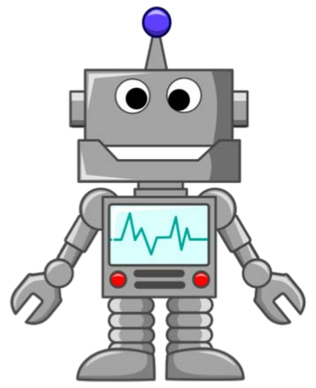


Hierarchical
Text Generation and Planning

Entangled Representations



I'll take the hats and the books

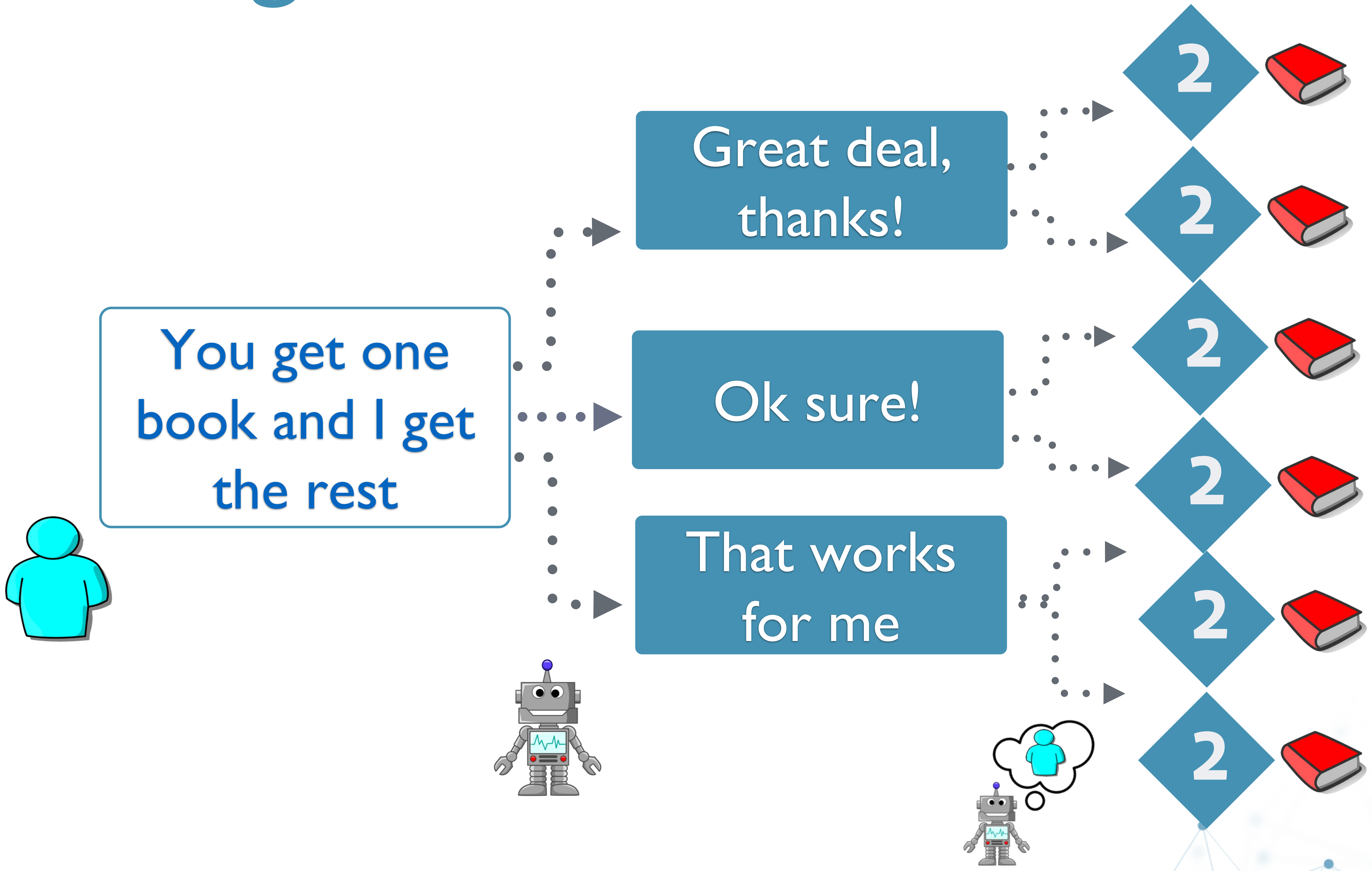


You take the hats and the books

I want the basketball, you get the rest

Word Similarity ≠ Semantic Similarity

Dialogue Rollouts



Entangled Representations

i can i i everything else . . . .

balls have zero to me to me to me to me to me to me to me to me to me to me

you i everything else . . . .

balls have a ball to me to me to me to me to me to me to me

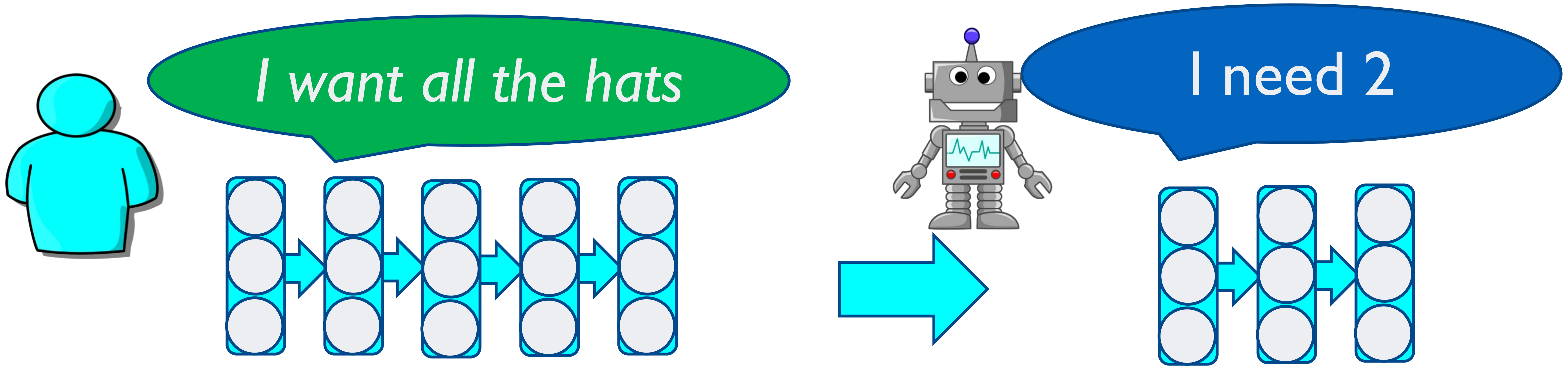
i i can i i i everything else . . . .

balls have a ball to me to me to me to me to me to

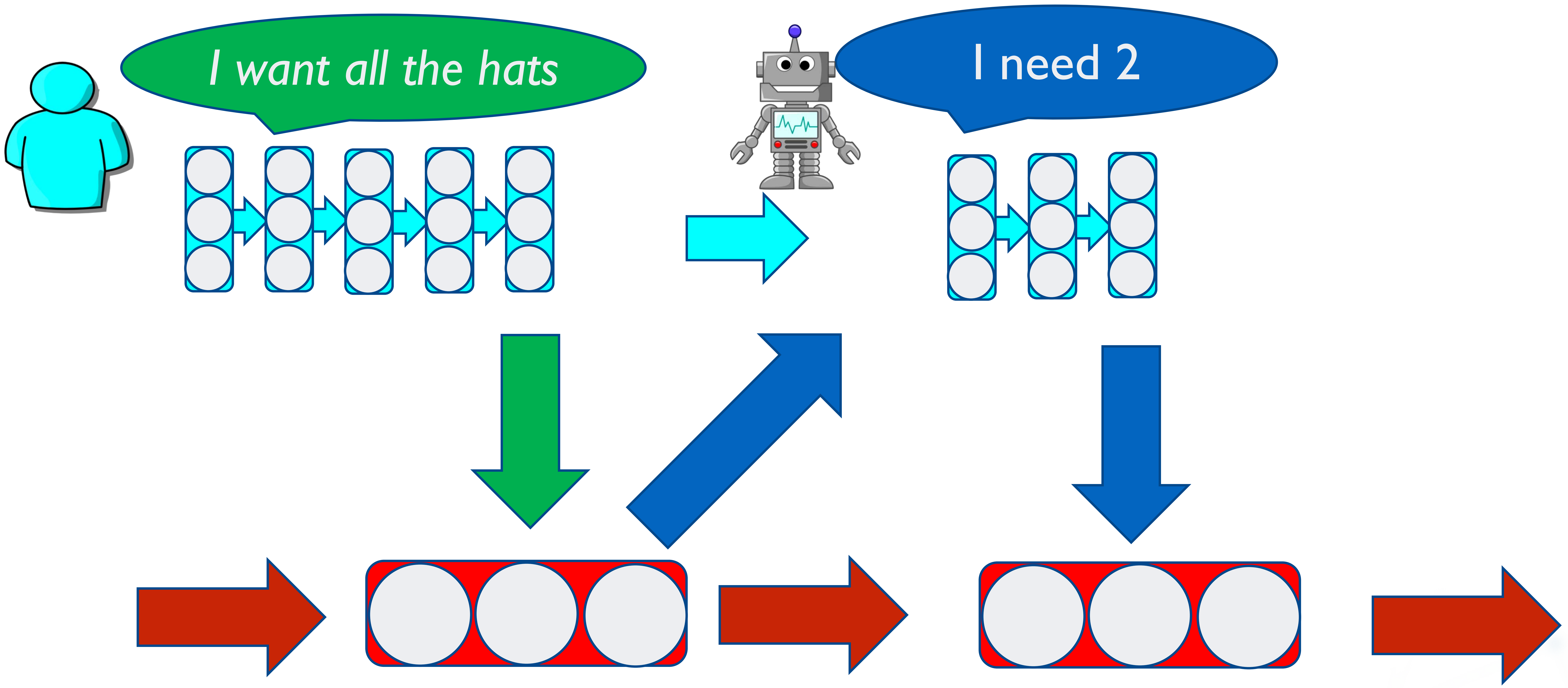
i . . . .

Reinforcement learning trades off **language quality** vs. **rewards**

RNN Dialogue Model

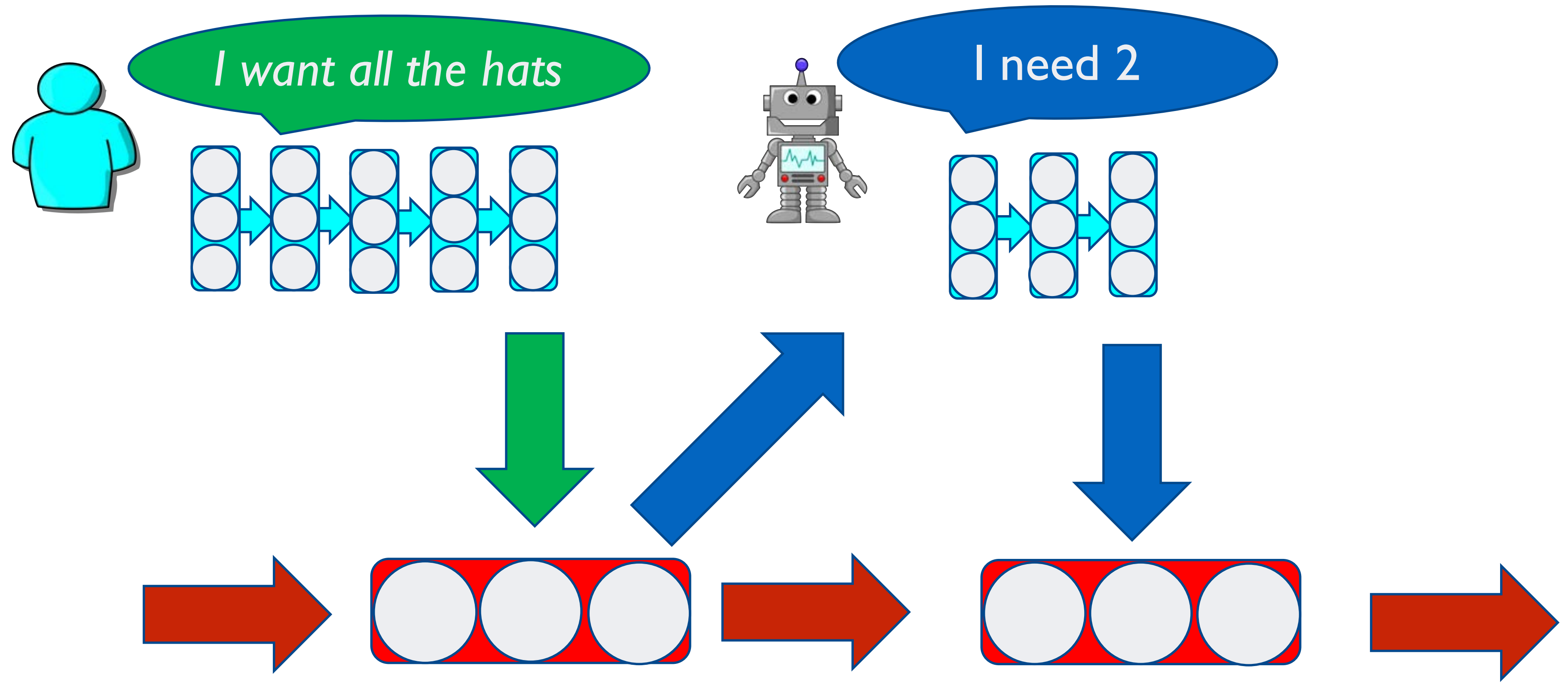


Hierarchical RNN Dialogue Model

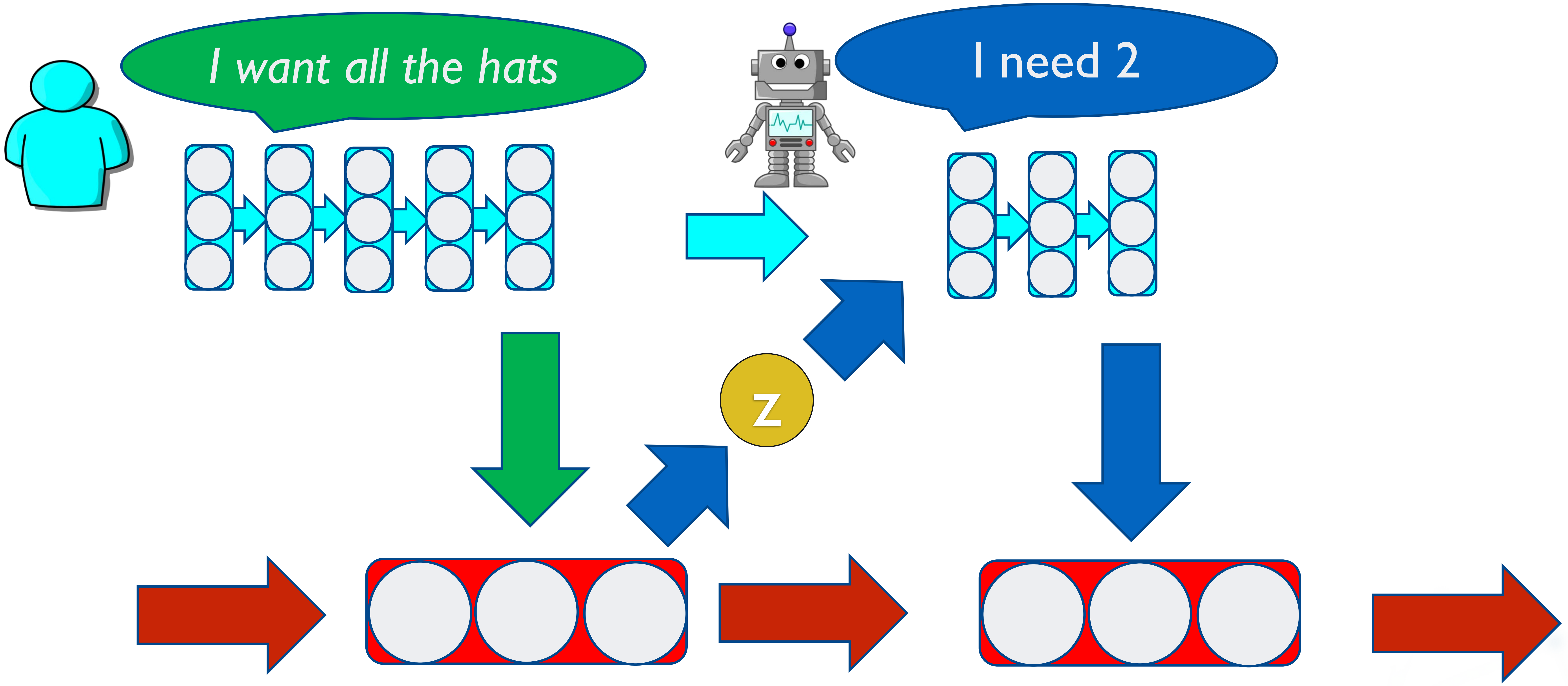


Hierarchical RNN Dialogue Model

Shortens dependencies between turns



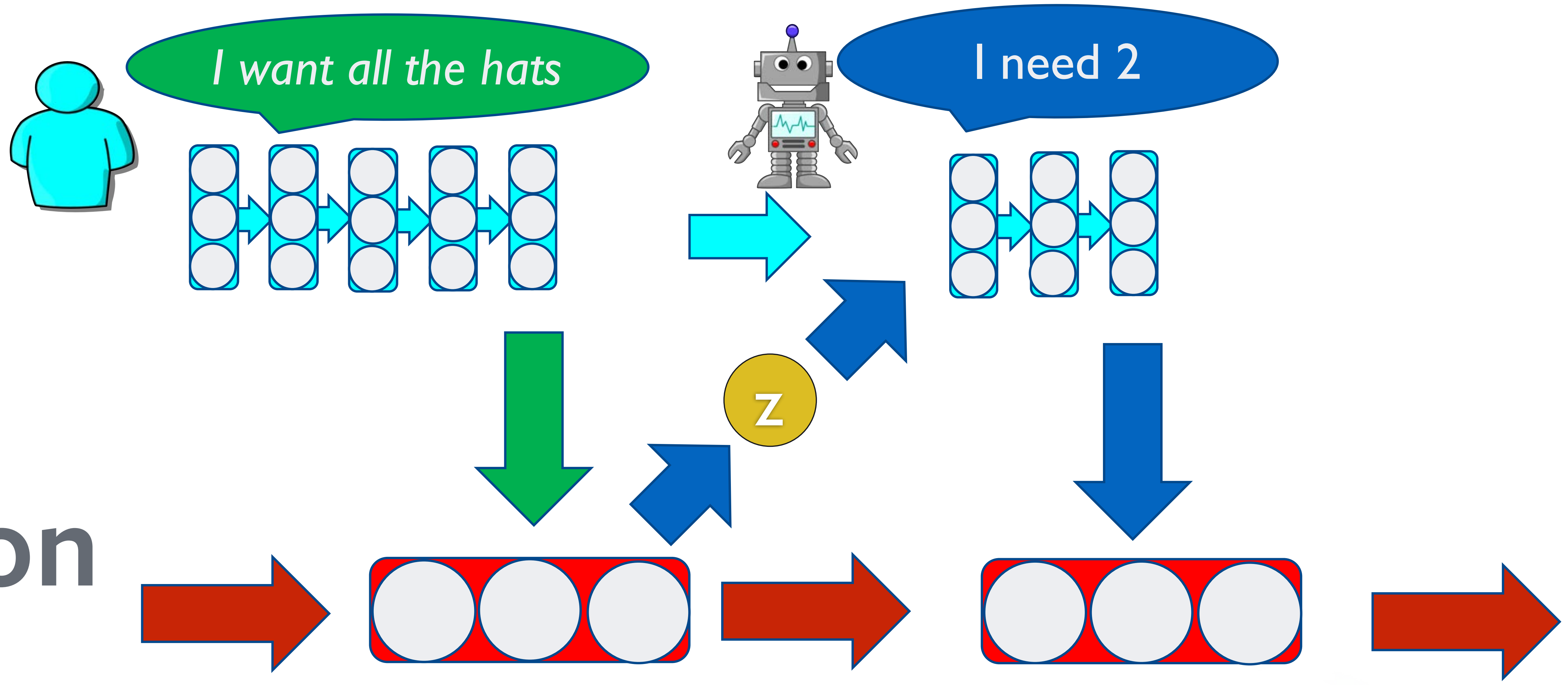
Latent Variable RNN



Latent Variable RNN

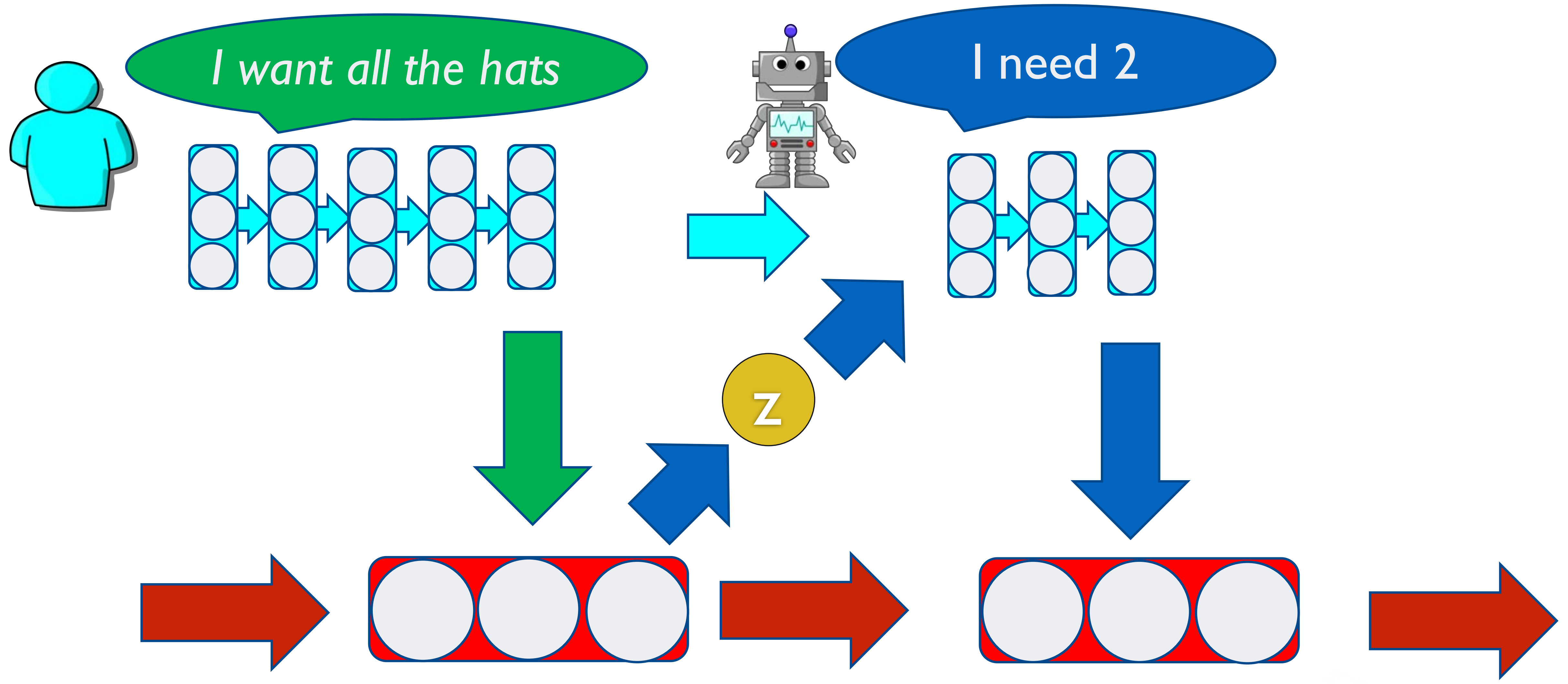
- First sample plan *z*
- Generate message conditioned on *z*

- Factorizes generation problem into 'what to say' and 'how to say it'



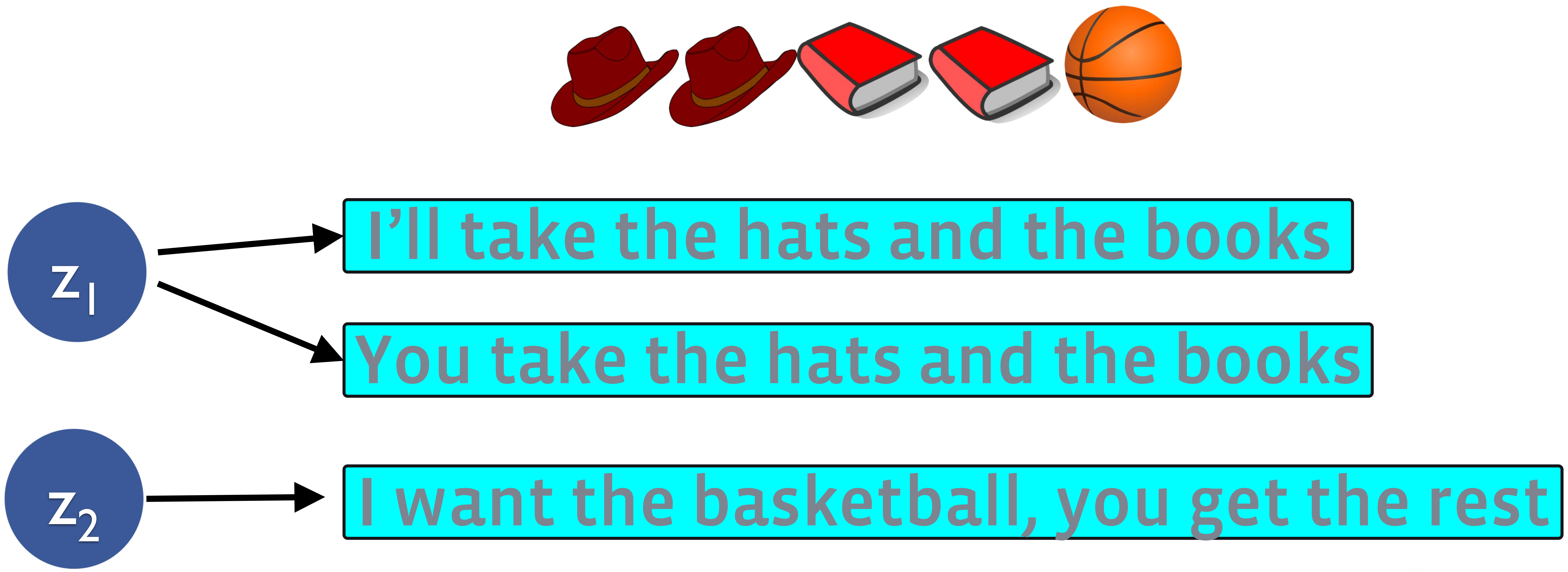
Learning Latent Representations

Where does *z* come from?



Existing approaches (e.g. Wen et al. 2017) train *z* to maximize likelihood of next message

Learning Latent Representations



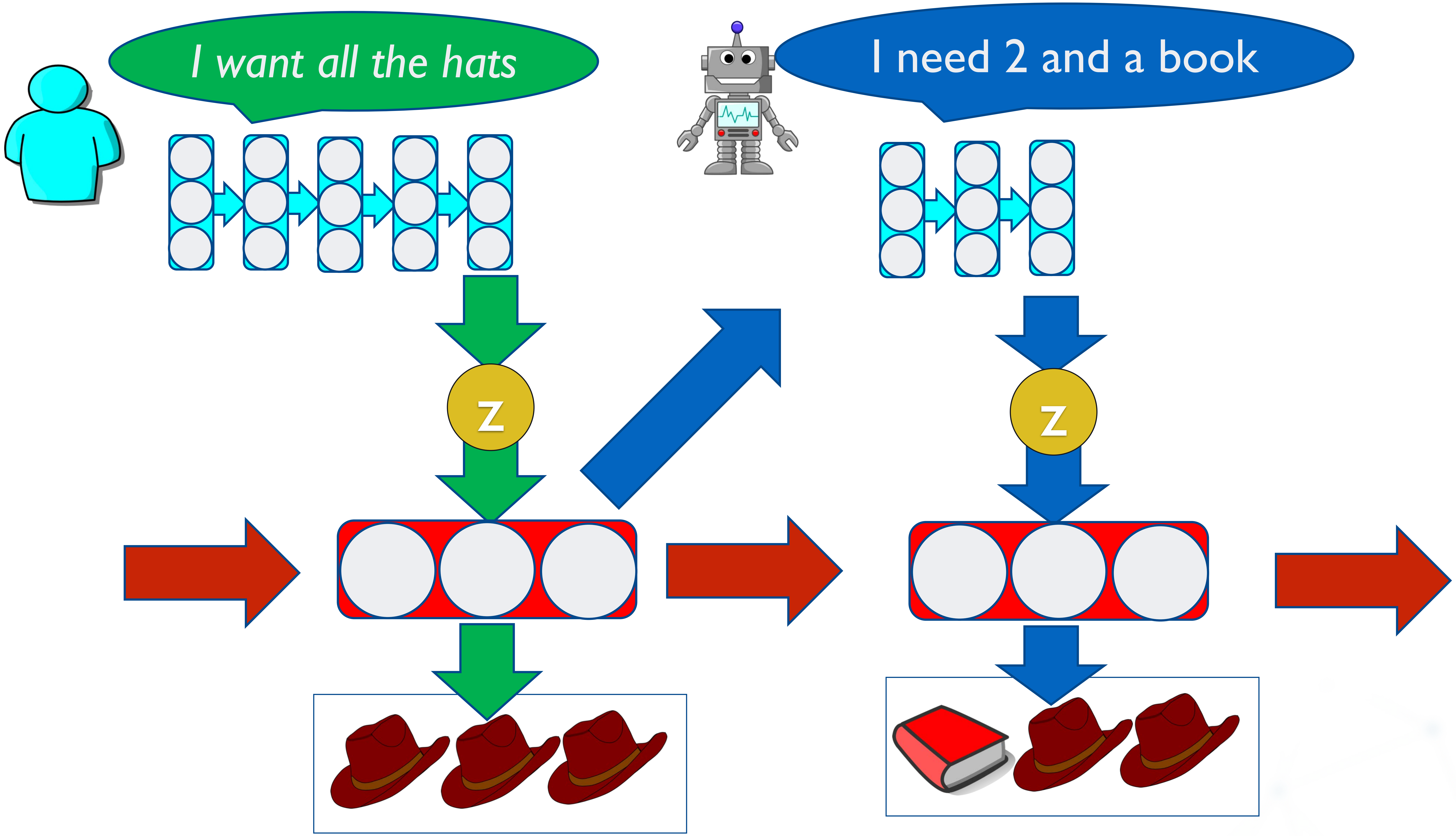
Learning *z* to maximize likelihood of message gives similar strings, not similar meanings

Pretraining Message Represenations

Learn message encoder representations that:

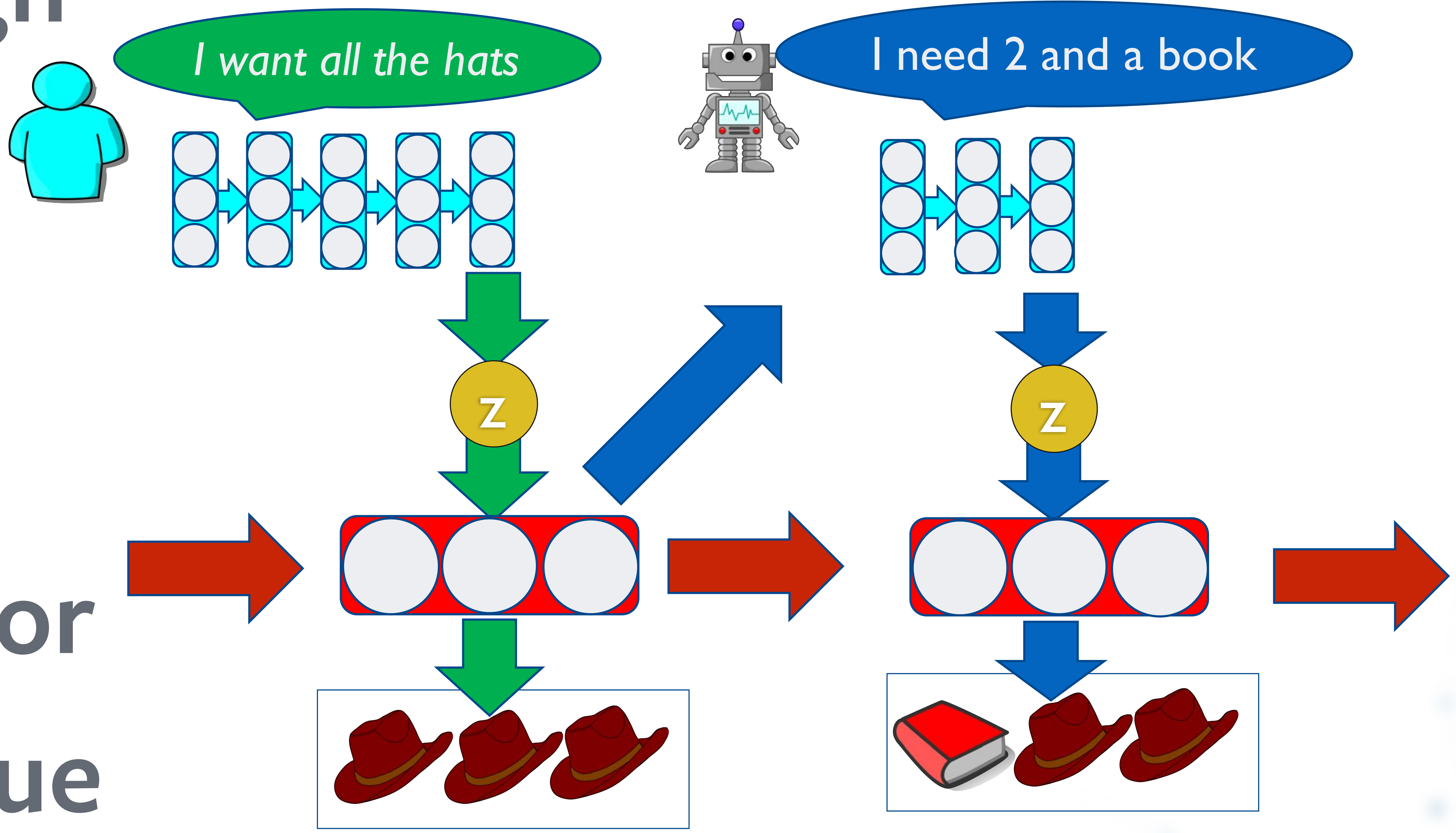
- Capture effect of message on dialogue
- Abstract over wording of message

Pretraining Message Representations

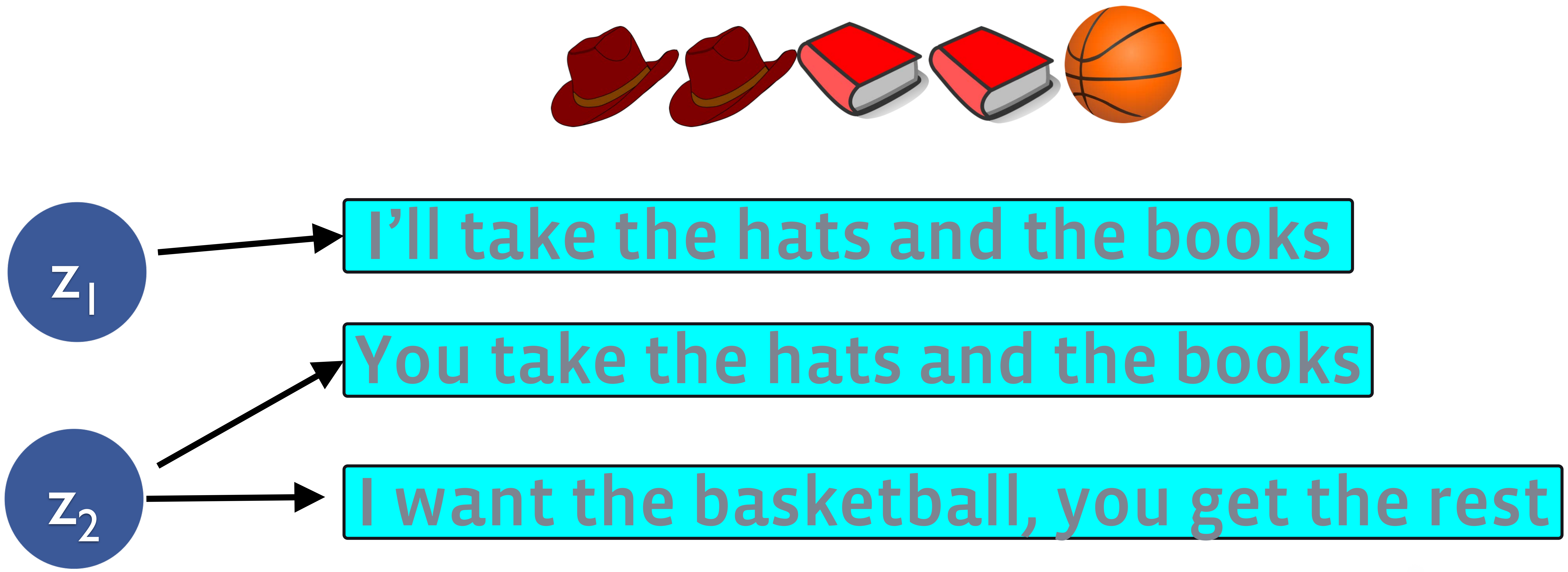


Pretraining Message Representations

- Messages forced through discrete bottleneck *z*
- *z* must contain all information necessary for predicting future dialogue



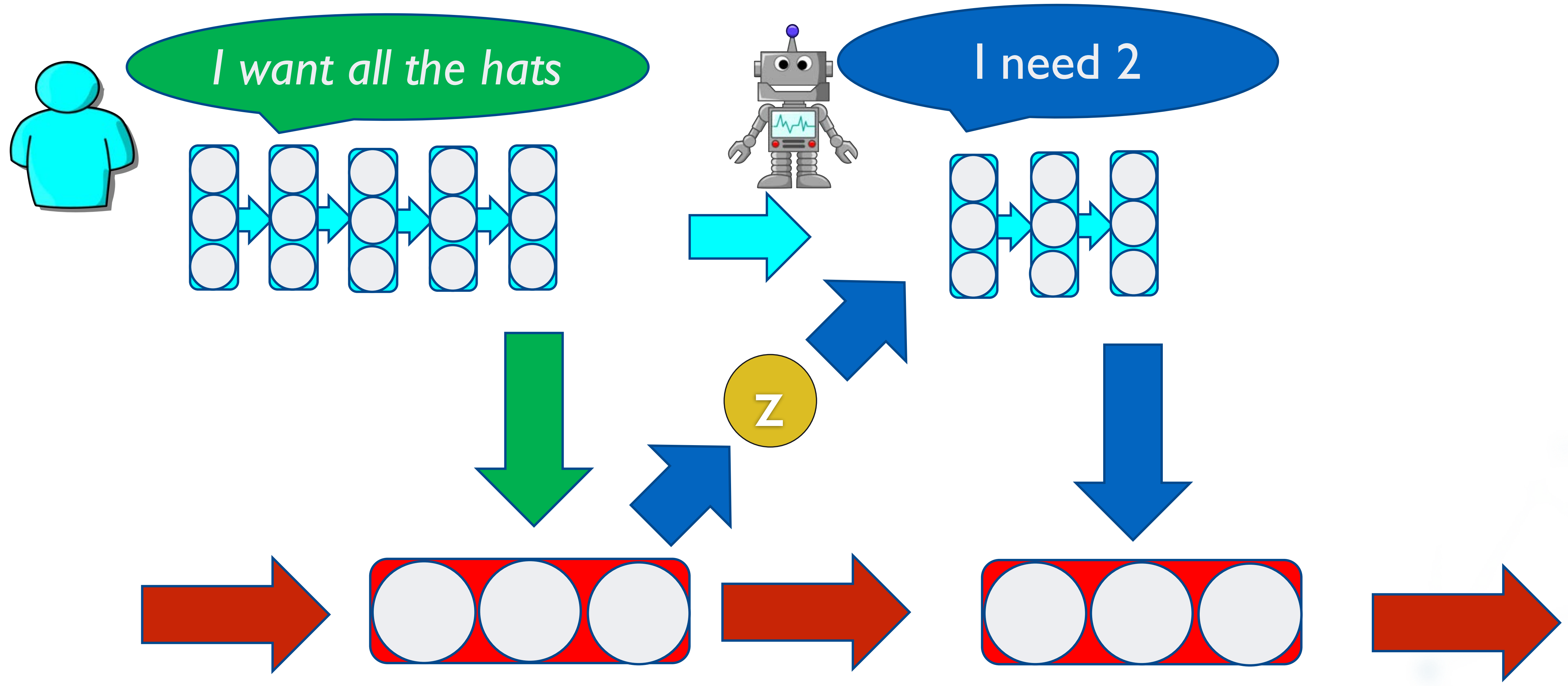
Learning Latent Representations



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

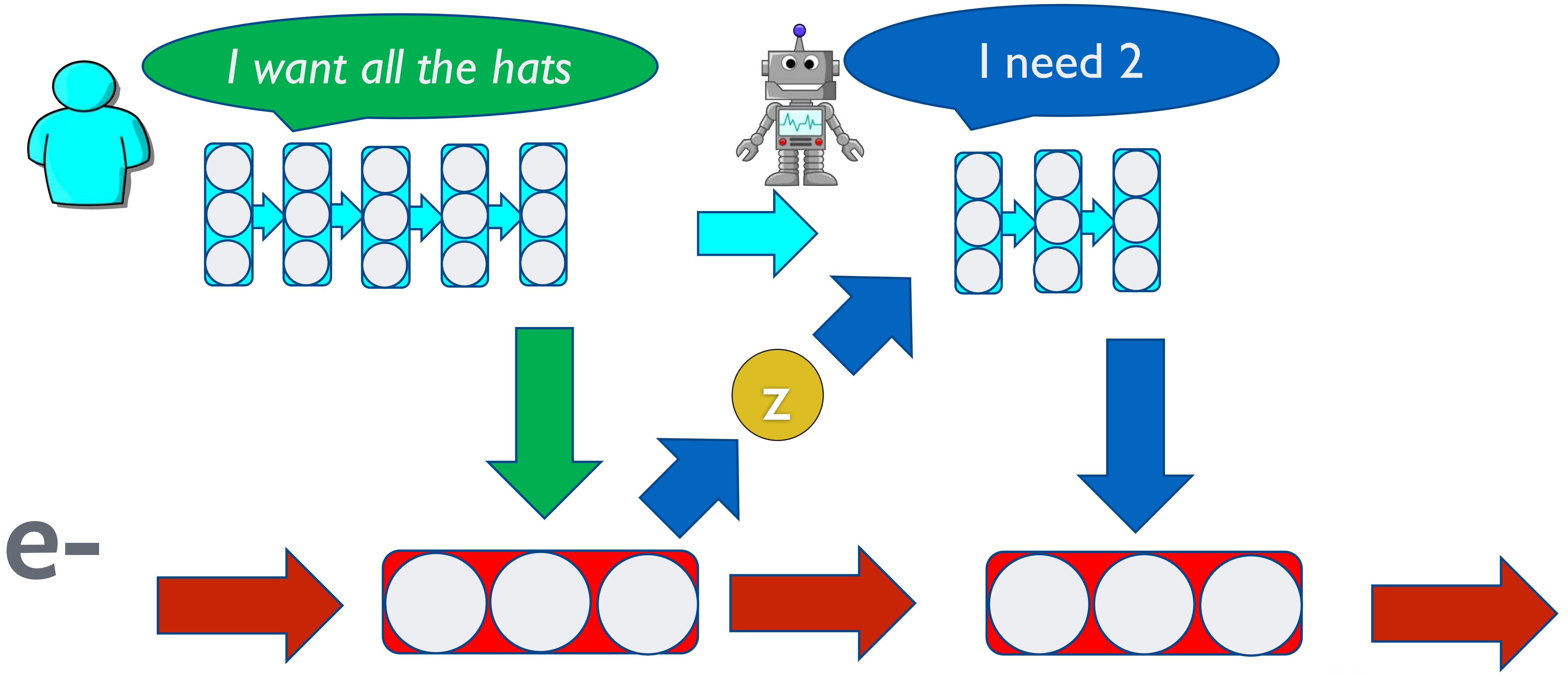
Pretraining Message Representations

- First cluster each message from training dialogues
- Then retrain full dialogue model using pretrained *z*
- *z* captures semantics of the message being generated



Reinforcement Learning

- Fix all parameters, except those that choose *z*
- RL can now only fine-tune semantics of messages



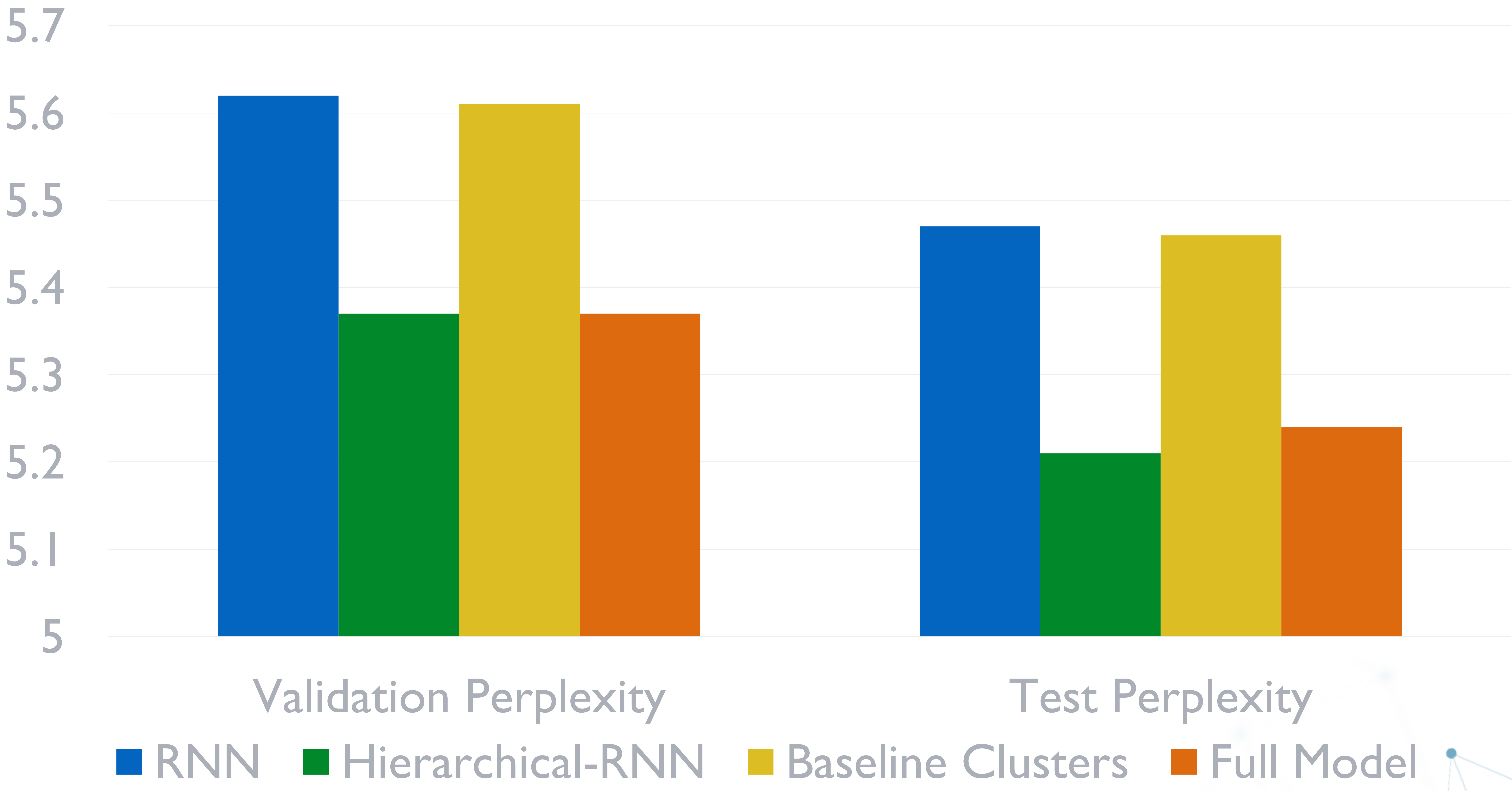
Dialogue Rollouts

Experiments

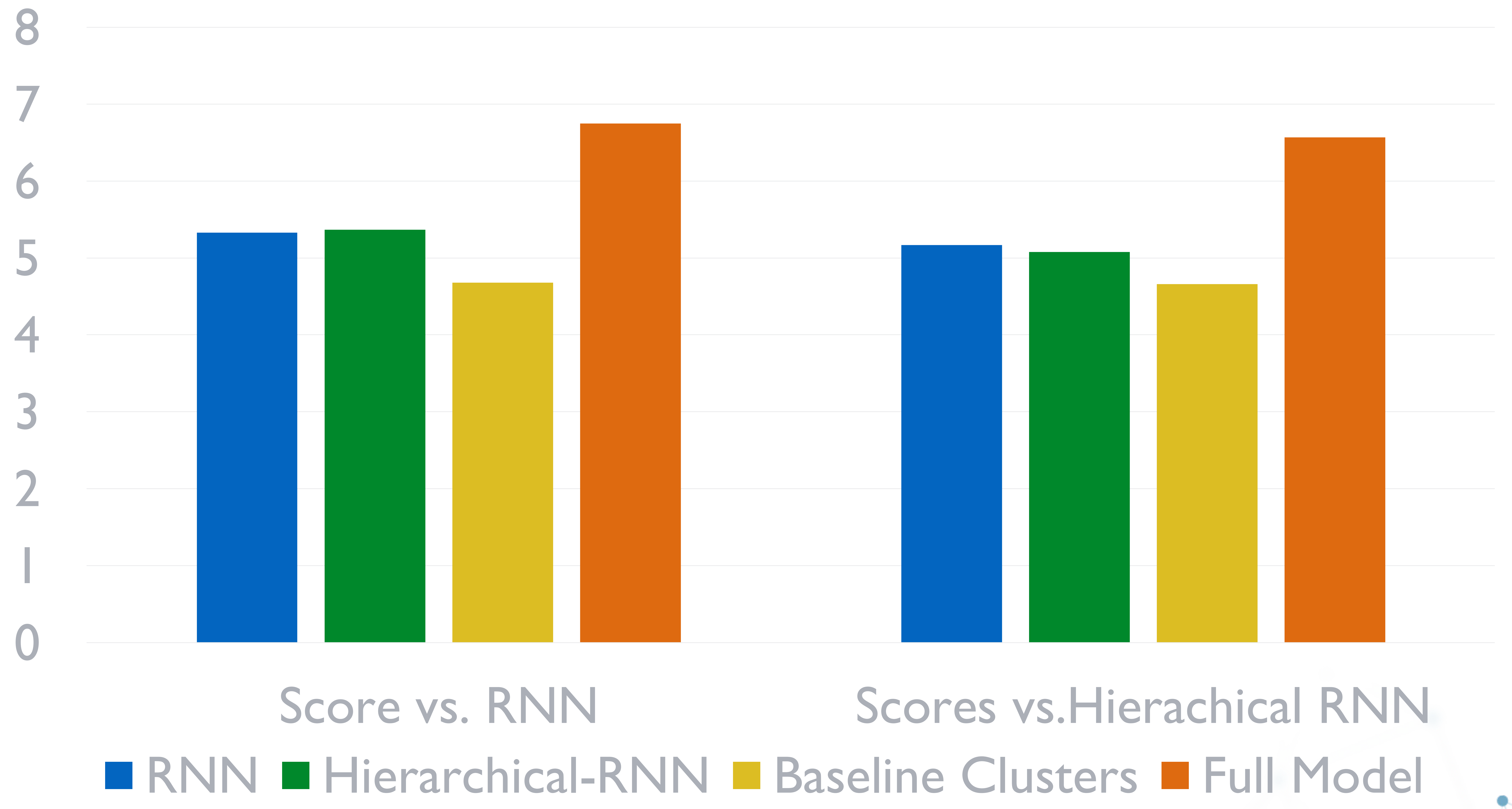
Models:

- RNN
- Hierarchical-RNN
- Baseline Clusters
- Full Model

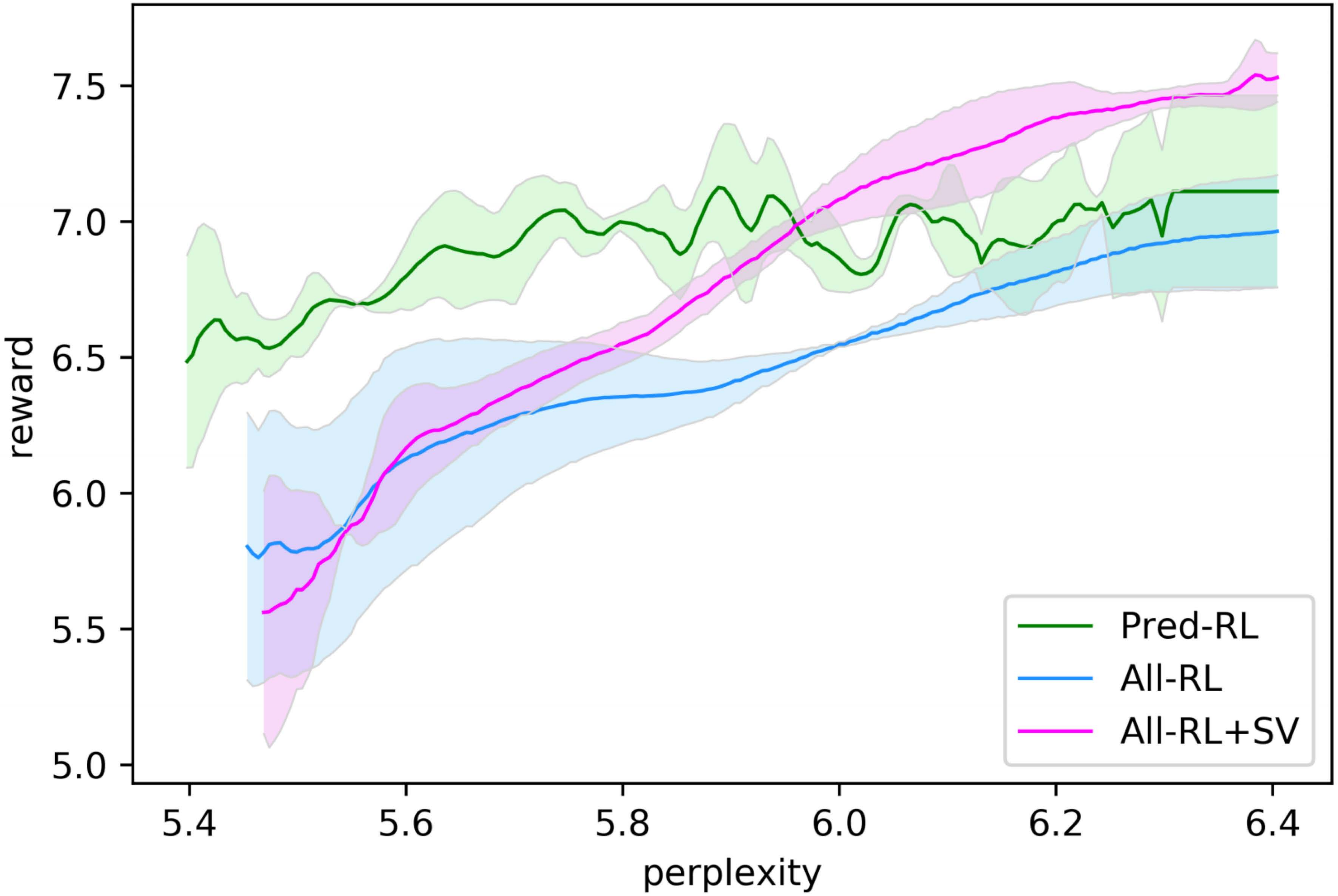
Language Quality



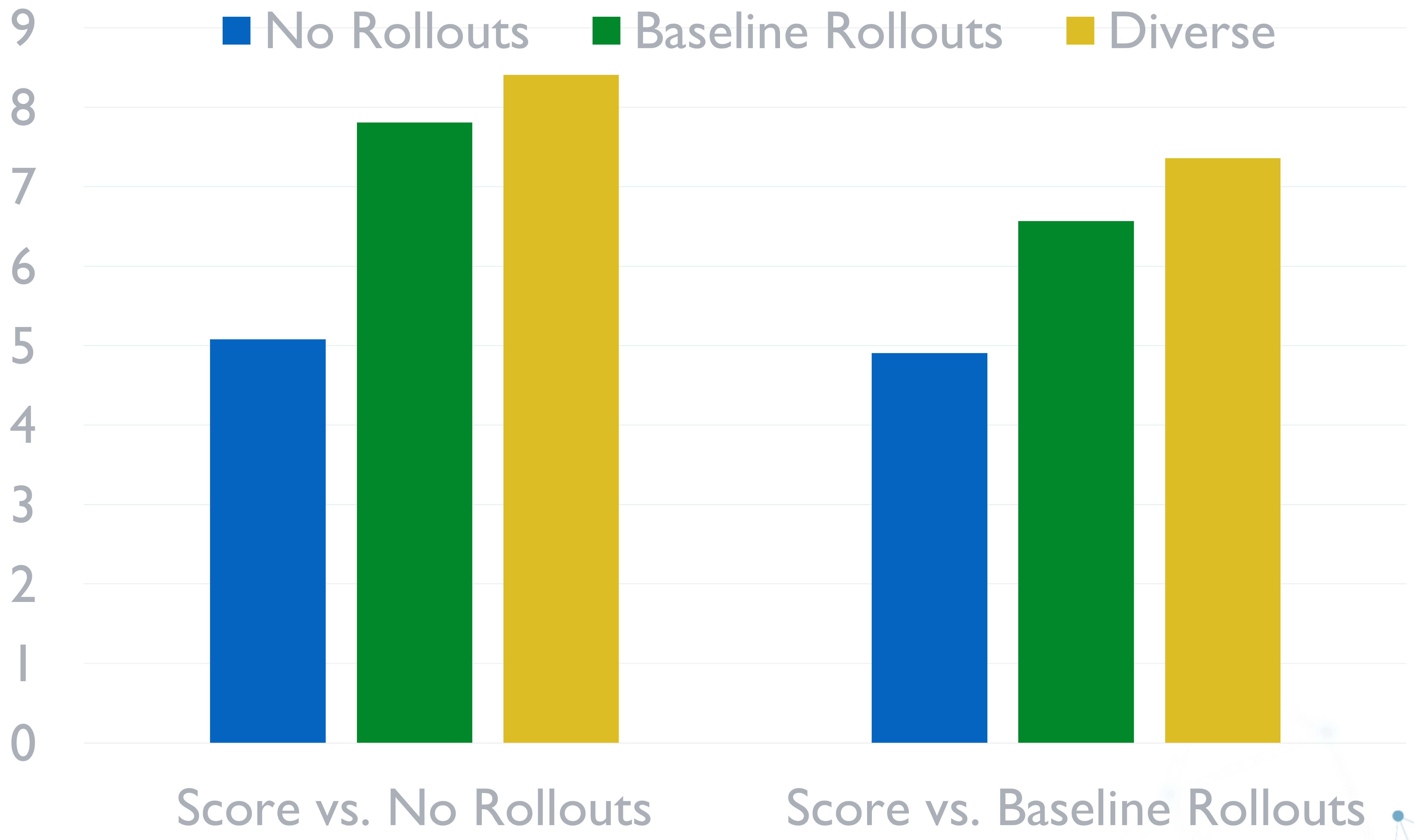
End Task Performance



Reinforcement Learning



Dialogue Rollouts



Analysis

Analyzed 1000 dialogues between **Hierarchical-RNN** and **Full Model**

Full Model uses fewer 'generic' messages (245 vs. 815)

Analysis

Analyzed 1000 dialogues between **Hierarchical-RNN** and **Full Model**

Full Model is more *creative*: 60% of messages don't appear in training data, vs. 18% for baseline

Analysis

Analyzed 1000 dialogues between **Hierarchical-RNN** and **Full Model**

Full Model is *less repetitive*: 1% of dialogues contain repetition of the same demand, vs. 12% for baseline

Conclusion

Models need to learn to understand messages, not just imitate people

Disentangling *form* and *meaning* can improve both language quality and end task results

Future work should scale to larger domains

Any questions?

facebook