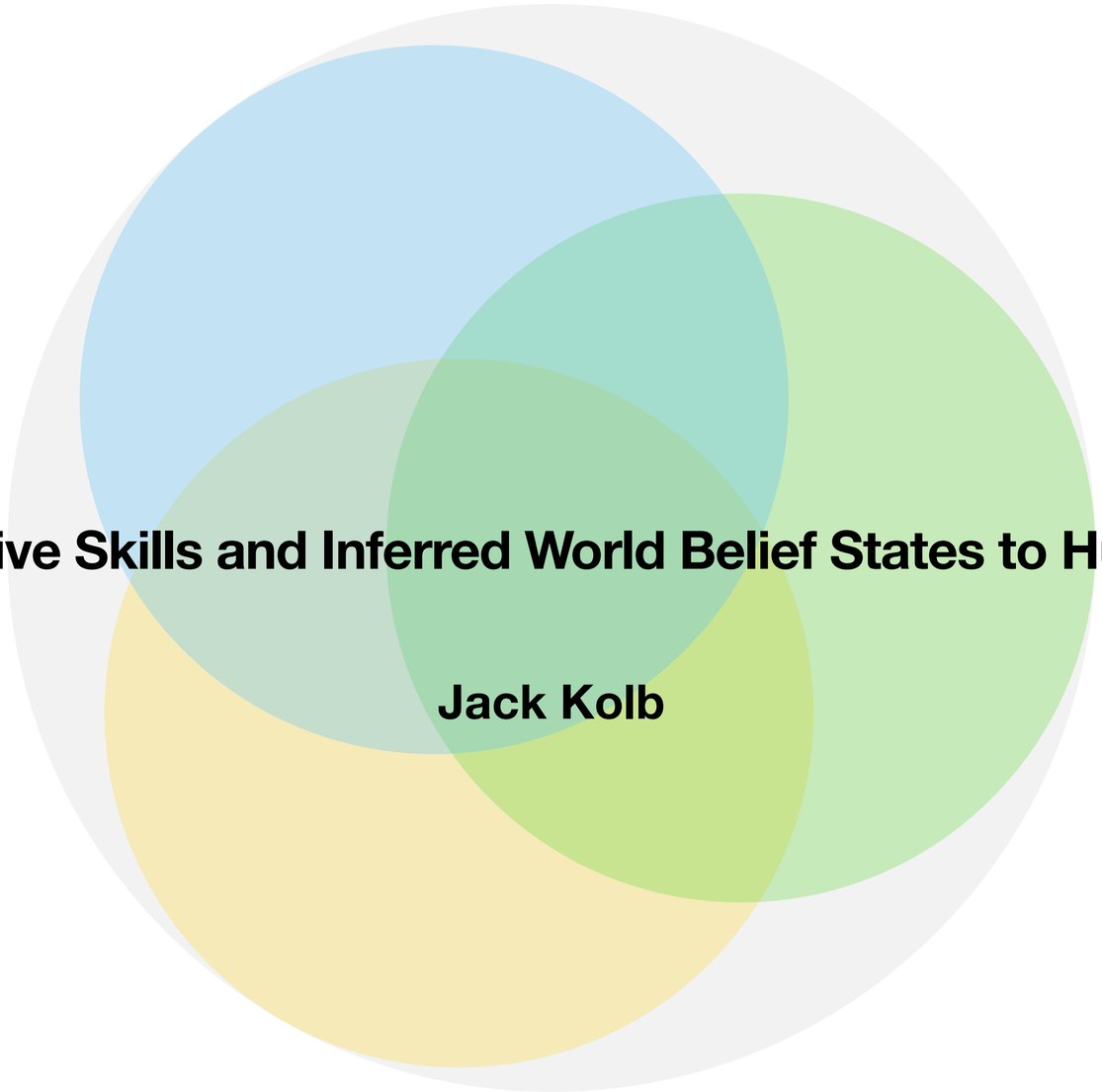


Applying User Cognitive Skills and Inferred World Belief States to Human-Robot Teaming

Jack Kolb

Committee:

- Karen Feigh (advisor)
- Julie Adams
- Sonia Chernova
- Harish Ravichandar
- Alan Wagner

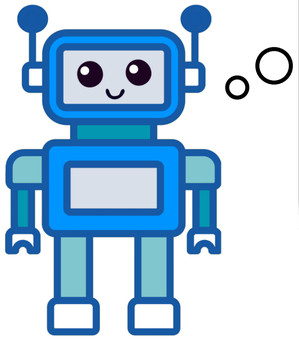
March 28, 2025

What makes a **teammate**?

What is the difference between a **teammate** and a **tool**?

What makes a teammate?

- Virtually all robot systems ignore the **person**.
- Instead they treat people as **black boxes**.



- I heard *“Can you help?”*
- I see a person.
- I see a table.
- I reason that the person wants help with the table.

What makes a teammate?

- Virtually all robot systems ignore the **person**.
- Instead they treat people as **black boxes**.

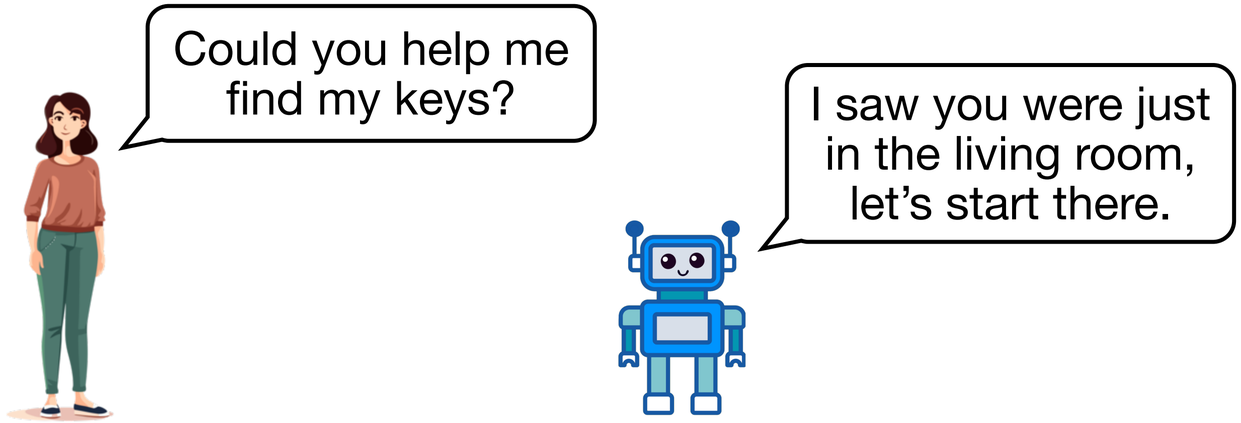
Jack's litmus test:

If the **human** teammate were replaced with a **capable automaton**, would the team's performance change?

Yes → Likely a teammate capability

No → Likely a tool capability

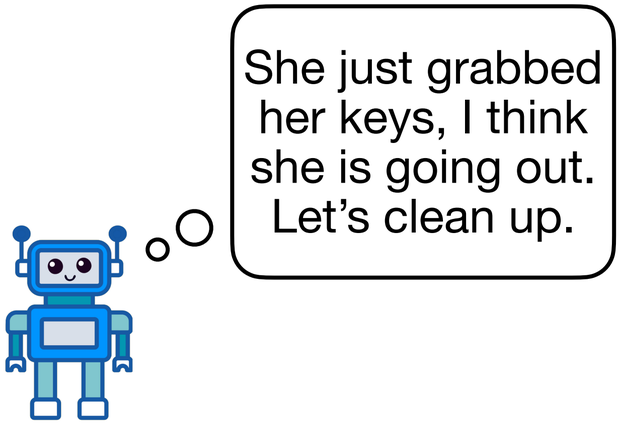
What makes a teammate?



- Bi-directional communication and planning.

  *A **teammate** works with your level of knowledge to set goals and communicate.*

What makes a teammate?



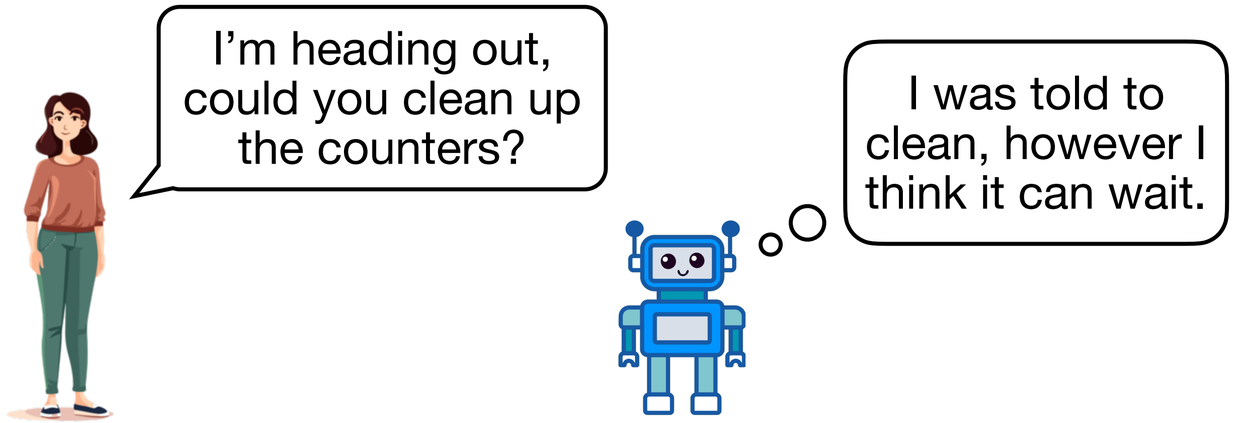
- Bi-directional communication and planning.

  *A **teammate** works with your level of knowledge to set goals and communicate.*

- Actively maintains a team mental model.

  *A **teammate** has a team mental model used to inform downstream tasks.*

What makes a teammate?



- Bi-directional communication and planning.

  *A **teammate** works with your level of knowledge to set goals and communicate.*

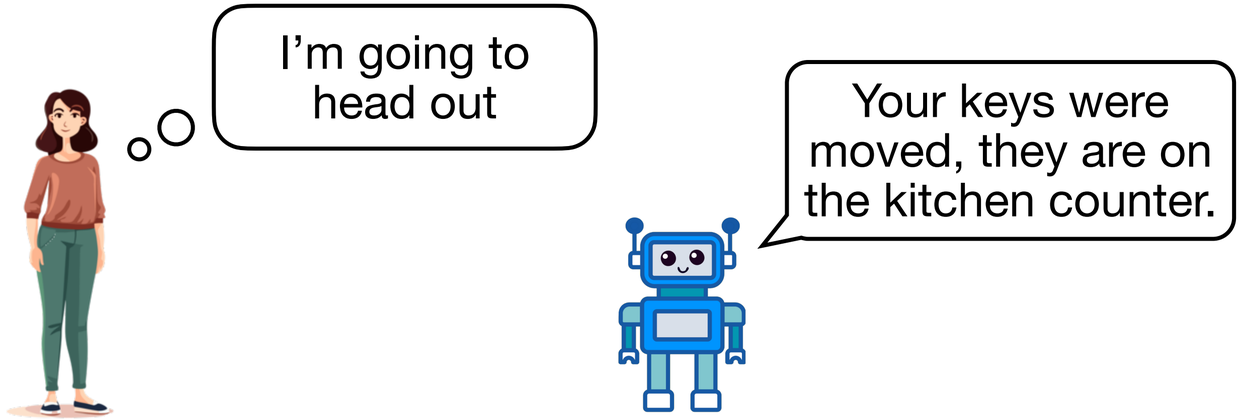
- Actively maintains a team mental model.

  *A **teammate** has a team mental model used to inform downstream tasks.*

- **Shared authority over key decision-making.**

  ***A teammate** has some authority on par with and potentially above you.*

What makes a teammate?



- Bi-directional communication and planning.

  *A **teammate** works with your level of knowledge to set goals and communicate.*

- Actively maintains a team mental model.

  *A **teammate** has a team mental model used to inform downstream tasks.*

- Shared authority over key decision-making.

  *A **teammate** has some authority on par with and potentially above you.*

- Intentioned communication and transparency.

  ***A teammate** knows when and what to communicate to be an effective team.*

What makes a teammate?

- Bi-directional communication and planning.

  *A **teammate** works with your level of knowledge to set goals and communicate.*

- Actively maintains a team mental model.

  *A **teammate** has a team mental model used to inform downstream tasks.*

- Shared authority over key decision-making.

  *A **teammate** has some authority on par with and potentially above you.*

- Intentioned communication and transparency.

  *A **teammate** knows when and what to communicate to be an effective team.*

Many teammate characteristics are reliant on modeling teammate beliefs, intentions, and capabilities.

How can robots **model** and **apply** a user's **cognitive state** to inform human-robot teaming capabilities?


IRA
DEVCOM
ARMY RESEARCH LABORATORY
RAIL
PCI
HRI
STAR LAB
SAFE
RL
STAR LAB
MARS
SMM
ISR
MED
EVAC
Office of Naval Research
Science & Technology
MAISR
POSE
CTRL
TMM
HAI
TMM
HRI
TMM
MAS
LOCKHEED MARTIN

How can robots **model** and **apply** a user's **cognitive state** to inform human-robot teaming capabilities?

Measure user **cognitive skills** → Predict teleoperation performance and conduct team role assignment.

Estimate user **situation awareness** → design systems to **structure decision-making** or **share authority**.

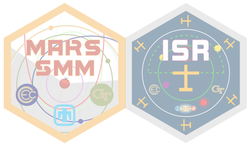
Infer user **world belief states** → Infer user situation awareness and actively assist users.

Monitor user **cognitive workload** → Adapt communication, reallocate taskwork to moderate workload.

Identify user strategic styles → Adapt agents to compliment a range of strategies and habituation.

How can robots **model** and **apply** a user's **cognitive state** to inform human-robot teaming capabilities?

Measure user **cognitive skills** → Predict teleoperation performance and conduct team role assignment.

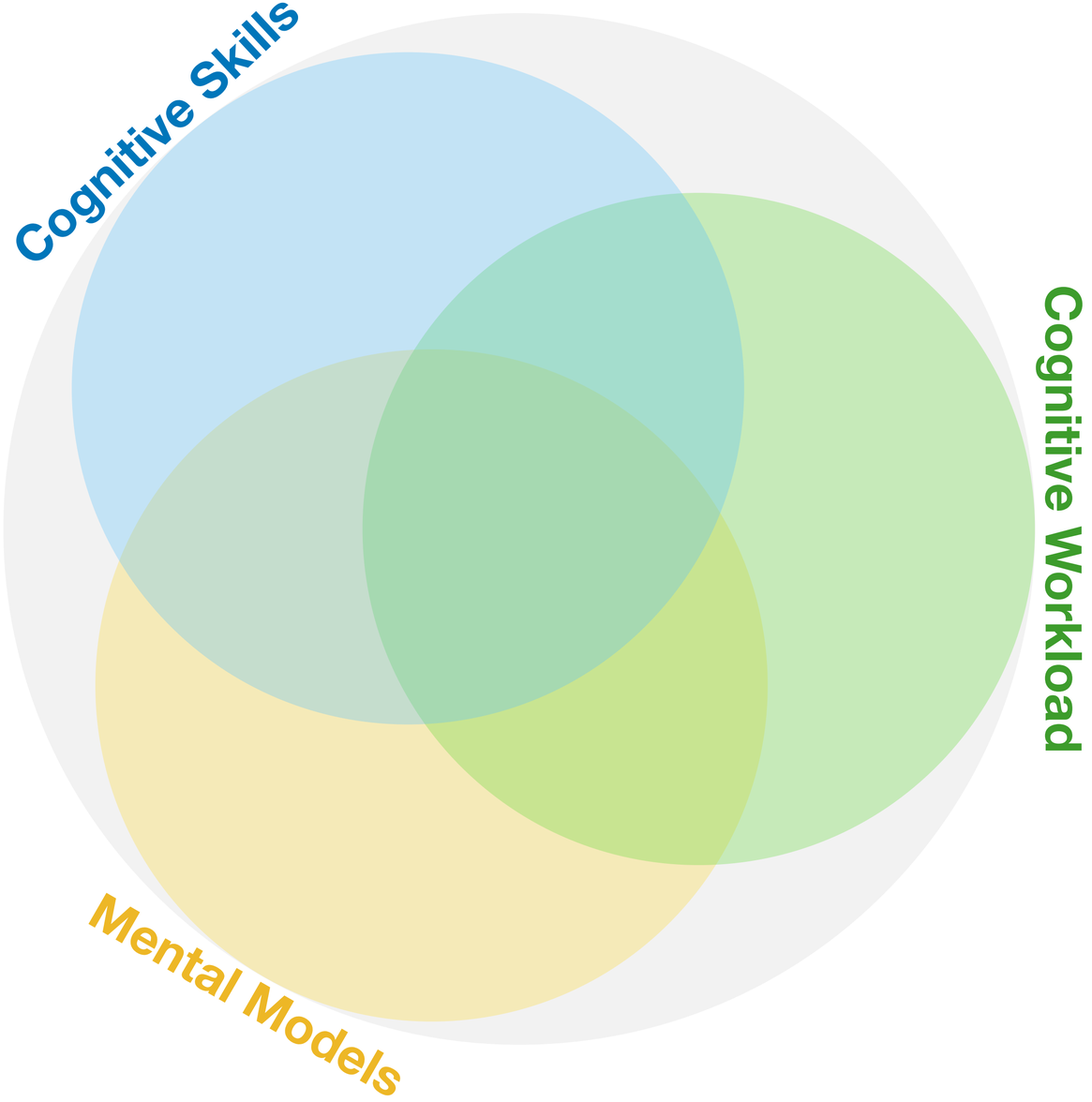


Infer user **world belief states** → Infer user situation awareness and actively assist users.

How can robots **model** and **apply** a user's **cognitive state** to inform human-robot teaming capabilities?

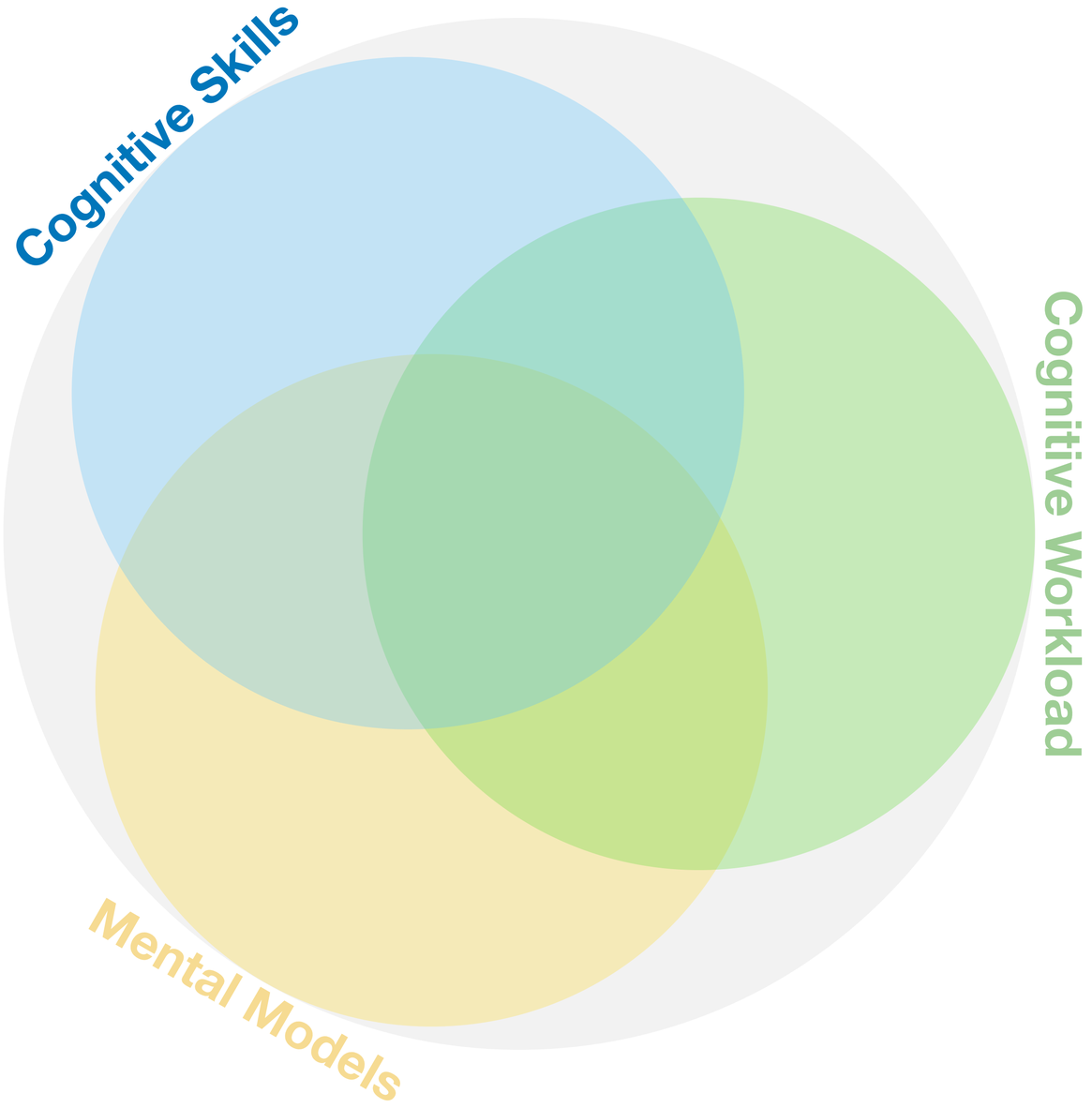
- What is **cognitive state**?
- **Predicting future performance** at robot teleoperation tasks.
- **Inferring the situation awareness** of people in a **2D** kitchen domain.
- **Inferring the belief states** of people in a **3D** household.

What is **cognitive state**?


Cognitive Skills
Cognitive Workload
Mental Models

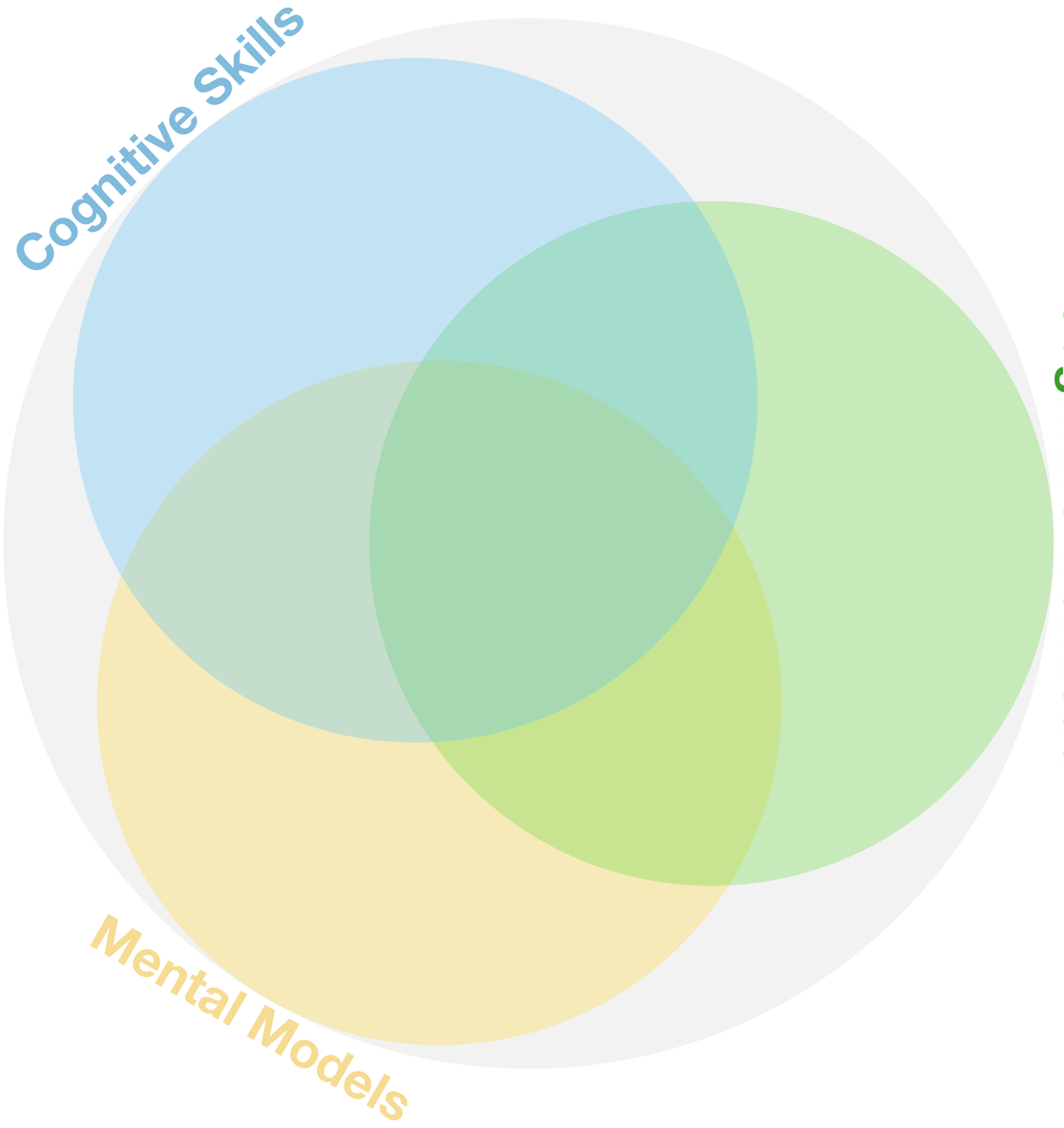
What cognitive skills are used in taskwork?

How can systems measure a user's current state and apply that information?



What cognitive skills are used in taskwork?

How can systems measure a user's current state and apply that information?

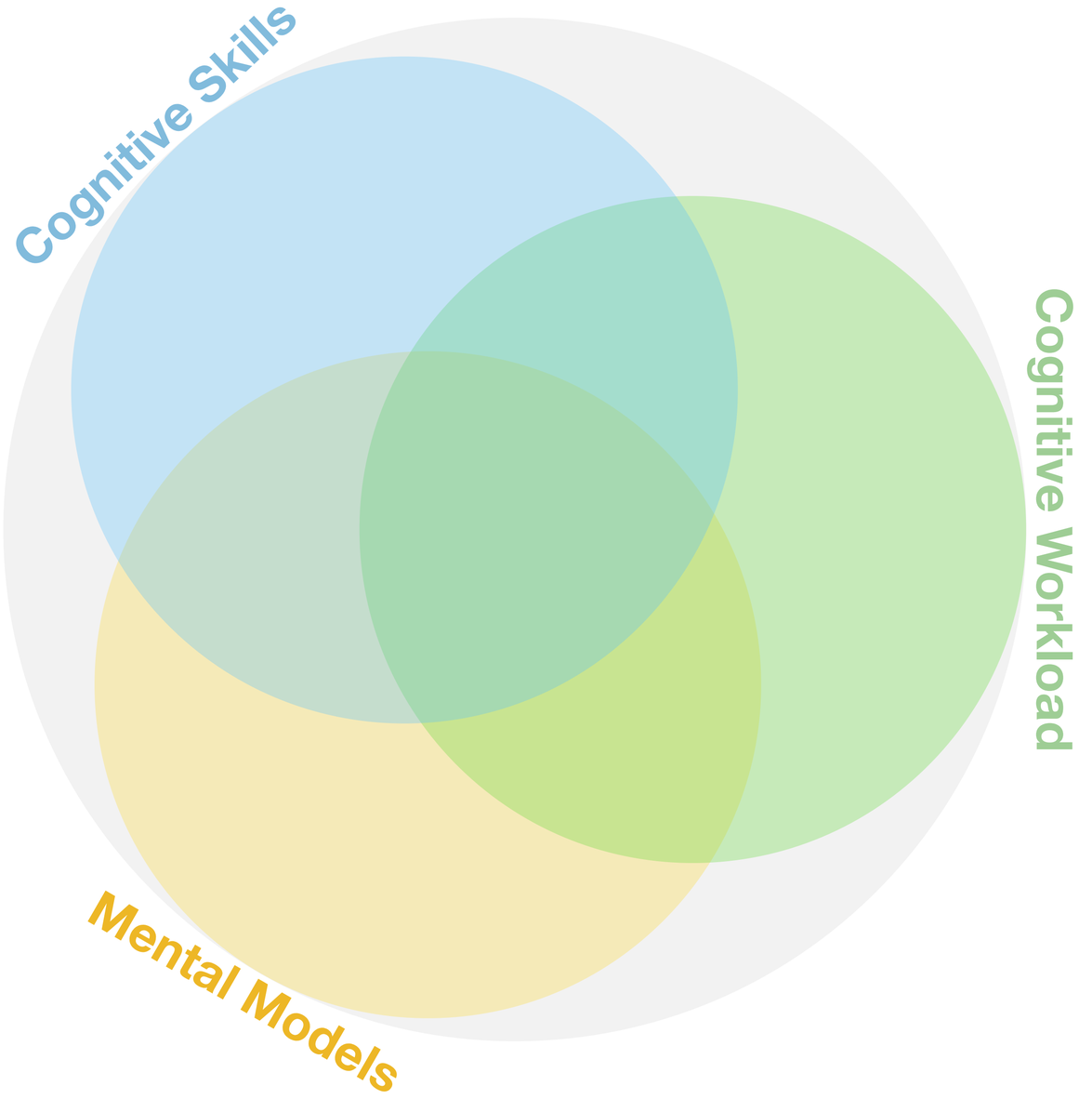


How does cognitive workload affect performance?

How can systems use workload to adapt interfaces and role assignments to optimize team performance?

What cognitive skills are used in taskwork?

How can systems measure a user's current state and apply that information?



How does cognitive workload affect performance?

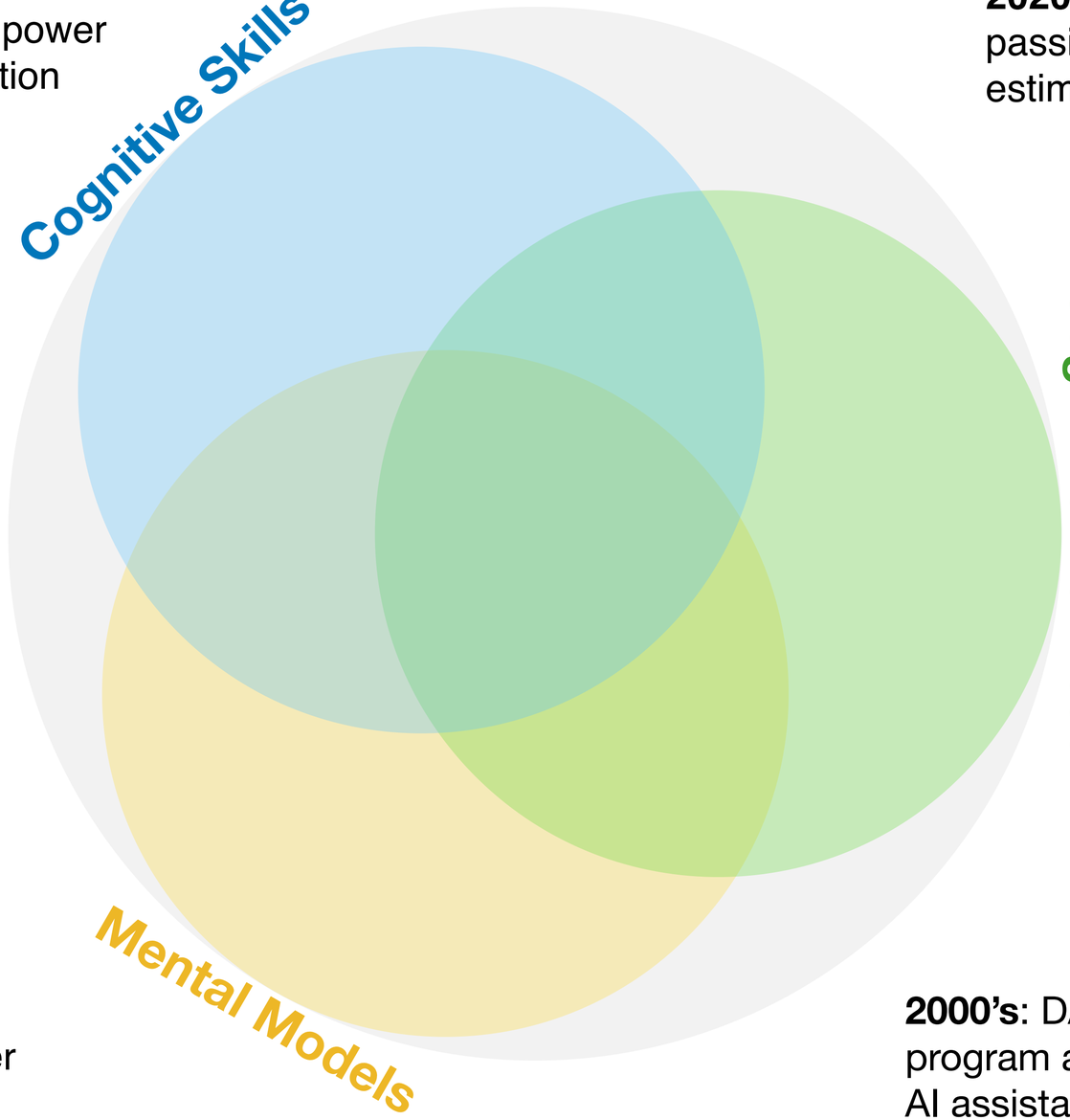
How can systems use workload to adapt interfaces and role assignments to optimize team performance?

How do people construct mental models of their surroundings?

How can systems estimate a user's mental model to enhance team dynamics and fluency?

2010's: ARL explored predictive power of cognitive skills on robot operation task performance.

2020's: Renewed interest in passive cognitive workload estimation.



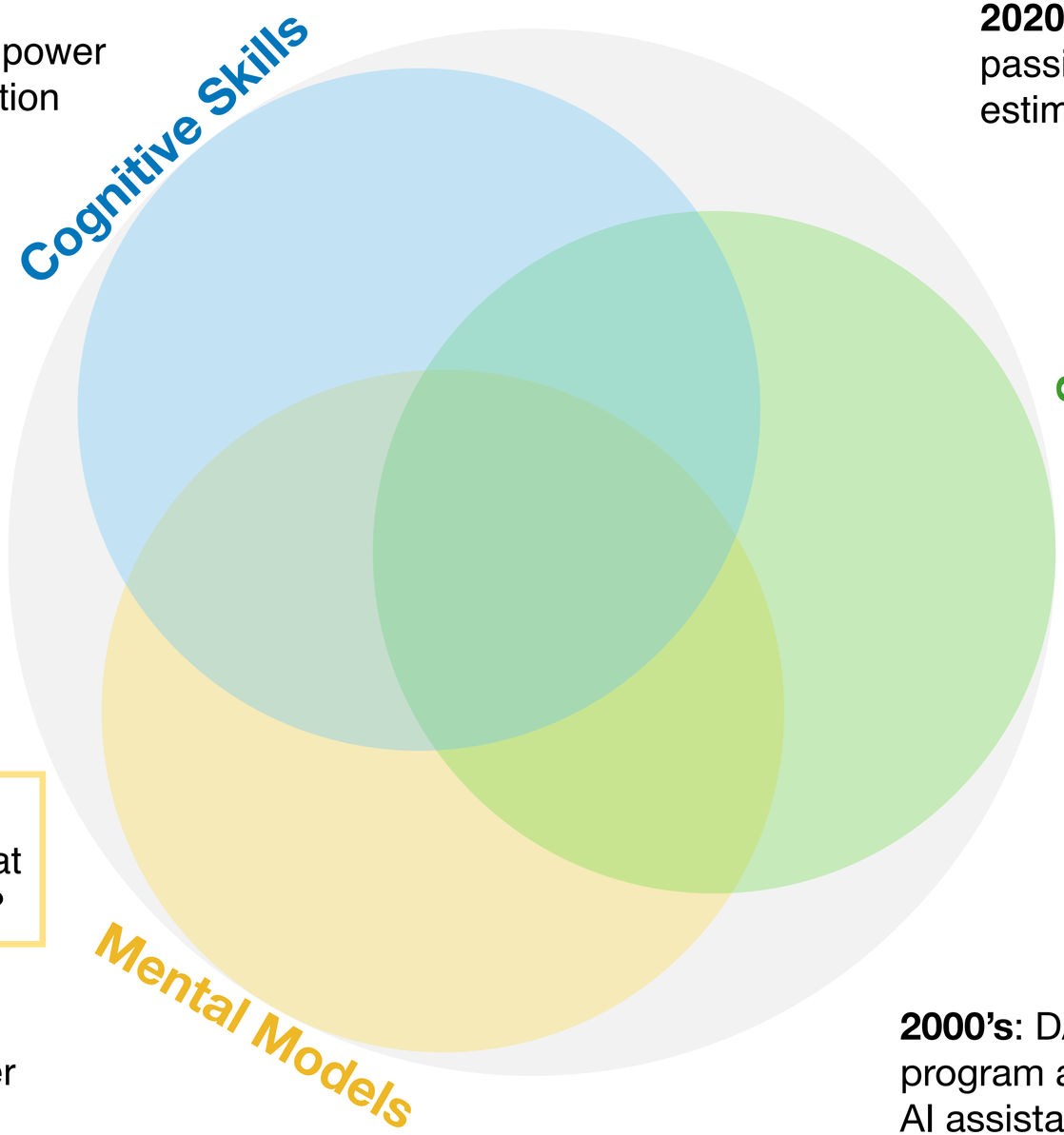
1990's: US Air Force sought cockpits that could adapt a pilot's taskwork to moderate pilot workload.

2010-20's: NASA interested in robots capable of estimating user mental models for teaming.

2000's: DARPA's Augmented Cognition program aimed to create context-aware AI assistants for soldiers.

2010's: ARL explored predictive power of cognitive skills on robot operation task performance.

Gap: What cognitive skills have predictive power? How can this be operationalized?



2020's: Renewed interest in passive cognitive workload estimation.

1990's: US Air Force sought cockpits that could adapt a pilot's taskwork to moderate pilot workload.

Gap: Which biometrics can passively monitor workload?

Gap: Can a mental model be obtained in real-world domains? What aspects can be leveraged by robots?

2010-20's: NASA interested in robots capable of estimating user mental models for teaming.

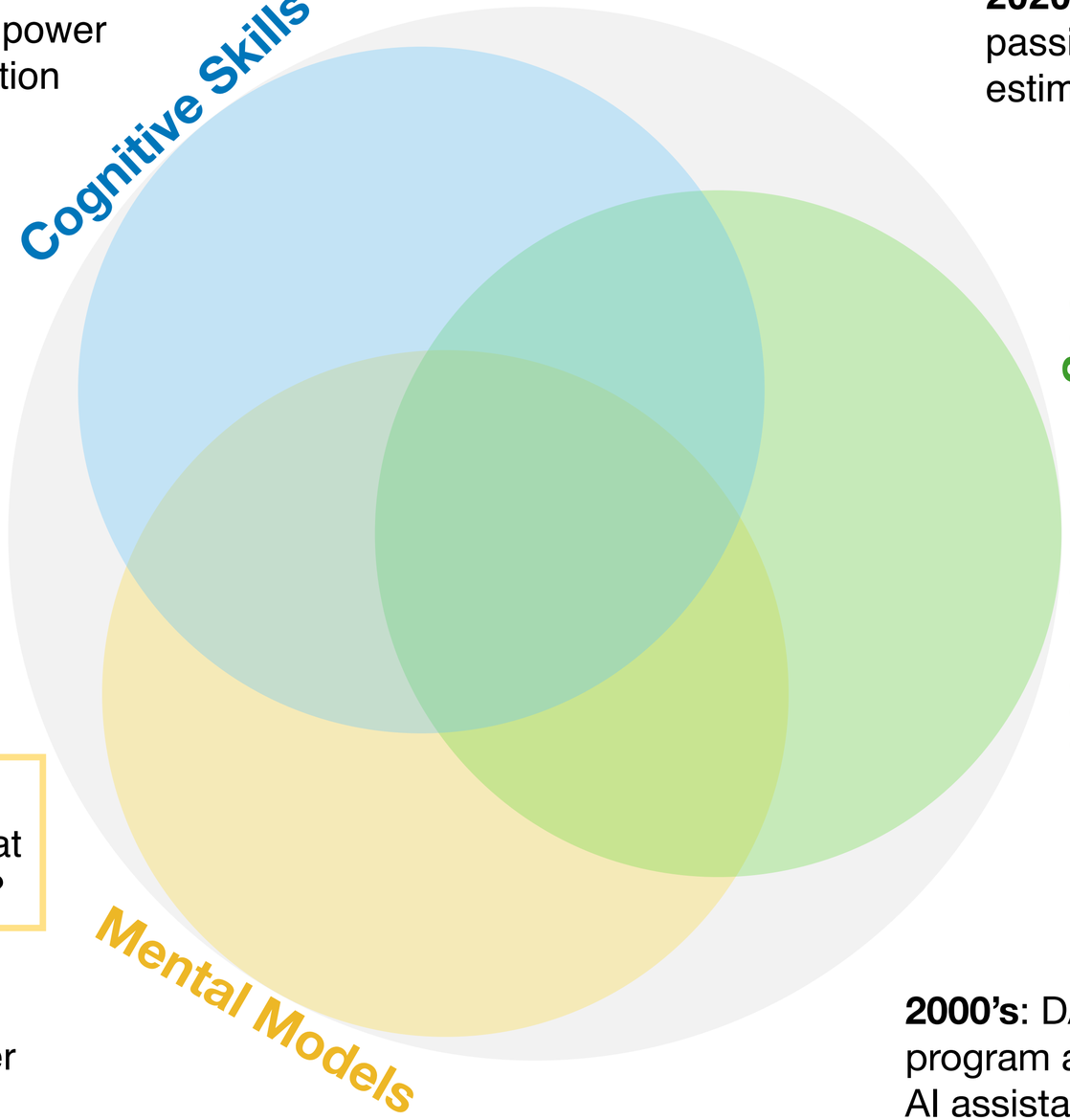
2000's: DARPA's Augmented Cognition program aimed to create context-aware AI assistants for soldiers.

2010's: ARL explored predictive power of cognitive skills on robot operation task performance.

Gap: What cognitive skills have predictive power? How can this be operationalized?

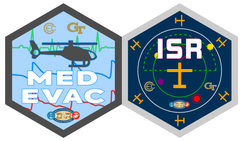
Gap: Can a mental model be obtained in real-world domains? What aspects can be leveraged by robots?

2010-20's: NASA interested in robots capable of estimating user mental models for teaming.



2020's: Renewed interest in passive cognitive workload estimation.

1990's: US Air Force sought cockpits that could adapt a pilot's taskwork to moderate pilot workload.

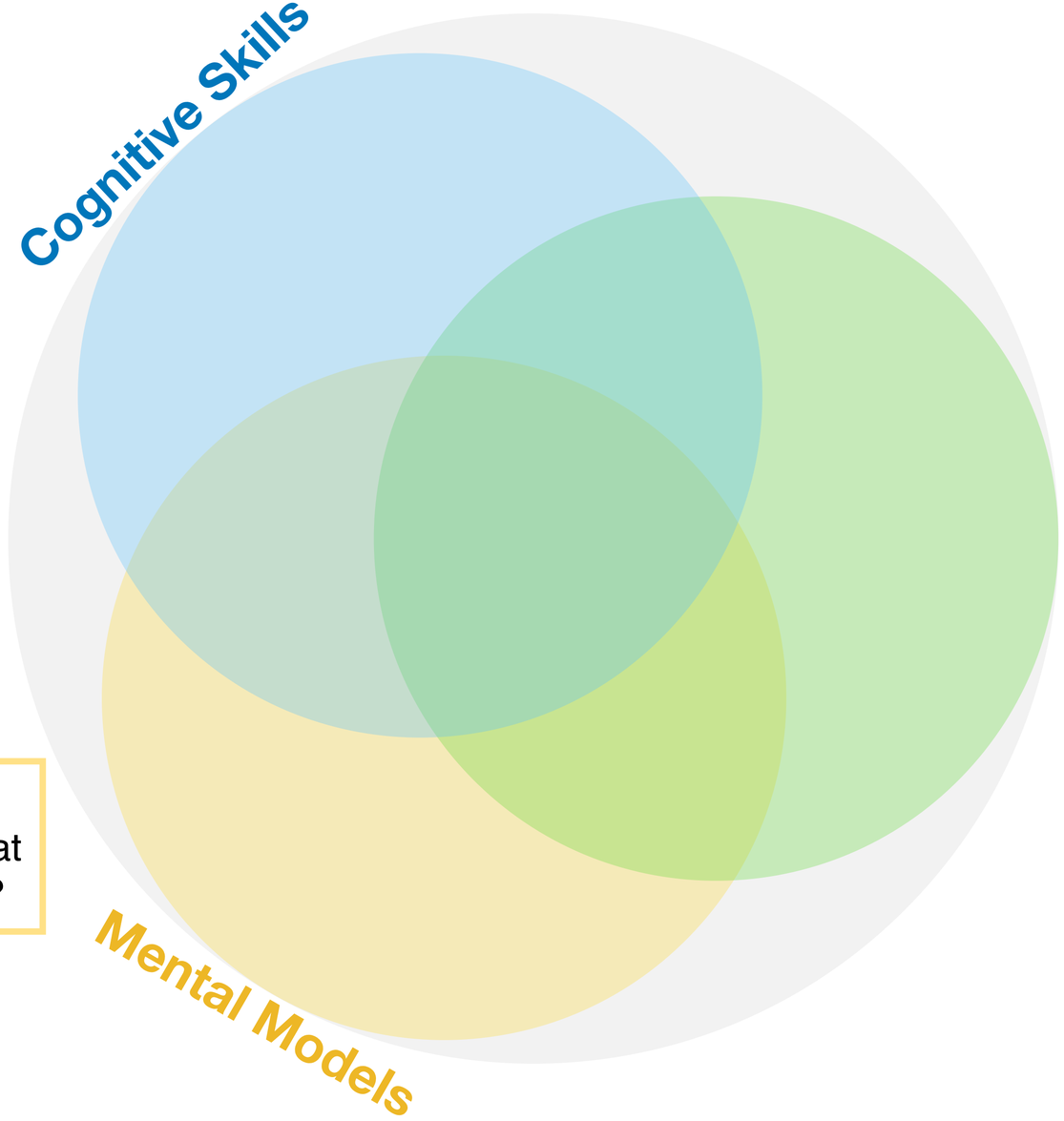


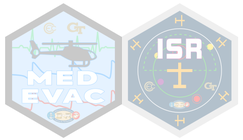
Gap: Which biometrics can passively monitor workload?

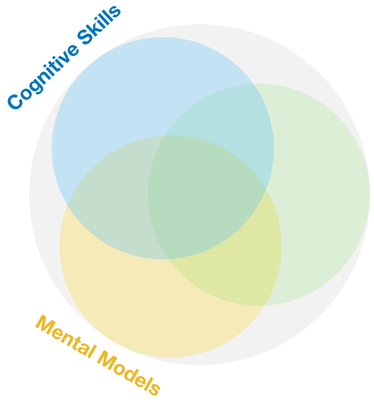
2000's: DARPA's Augmented Cognition program aimed to create context-aware AI assistants for soldiers.

Gap: What cognitive skills have predictive power? How can this be operationalized?

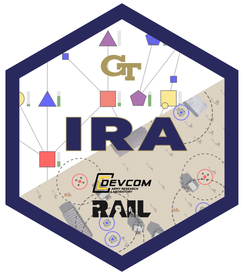
Gap: Can a mental model be obtained in real-world domains? What aspects can be leveraged by robots?







How can robots **model** and **apply** a user's **cognitive state** to inform human-robot teaming capabilities?

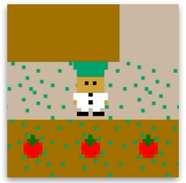


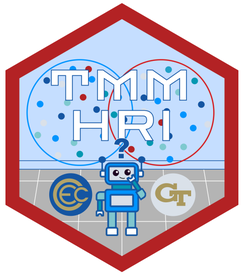
Can we predict future teleoperation performance only using cognitive skills, and apply it to role assignment?

Published in RO-MAN '21, RO-MAN '22

Can we infer user situation awareness via observing users in a **partially-observable** environment?

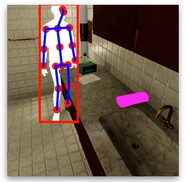
Published in IROS '24

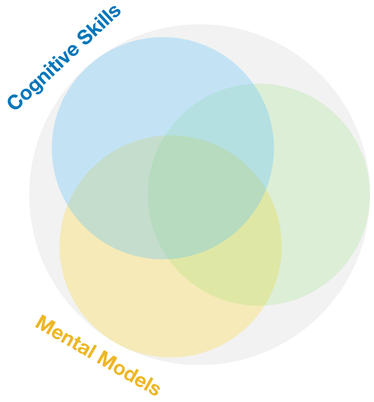




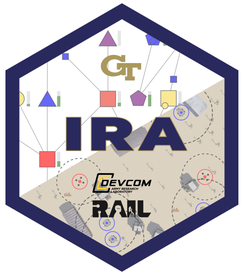
Can we infer user situation awareness via camera observations in a household domain?

Submitted to RA-L





How can robots **model** and **apply** a user's **cognitive state** to inform human-robot teaming capabilities?

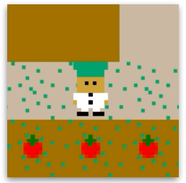


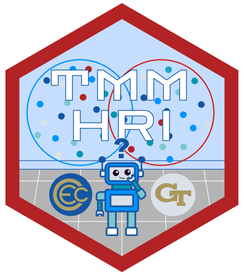
Can we predict future teleoperation performance only using cognitive skills, and apply it to role assignment?

Published in RO-MAN '21, RO-MAN '22

Can we infer user situation awareness via observing users in a **partially-observable** environment?

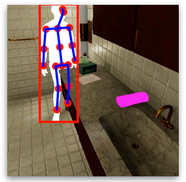
Published in IROS '24

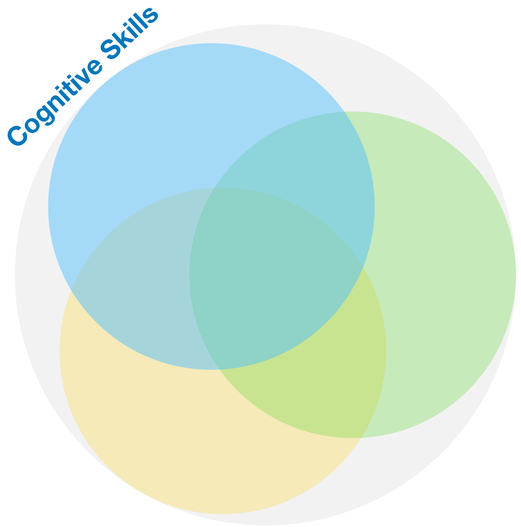




Can we infer user situation awareness via camera observations in a household domain?

Submitted to RA-L



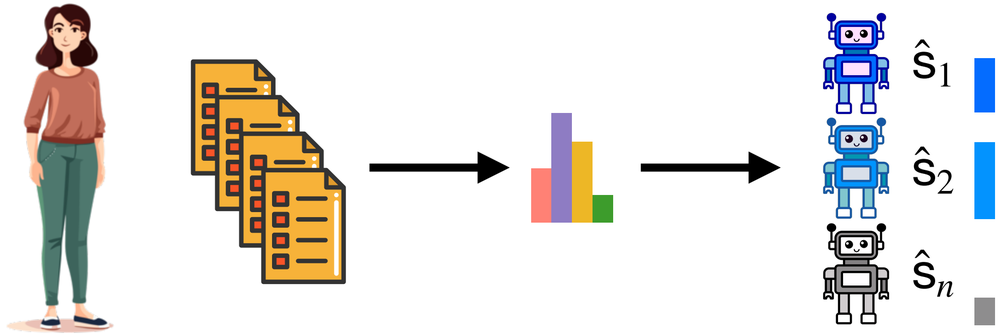


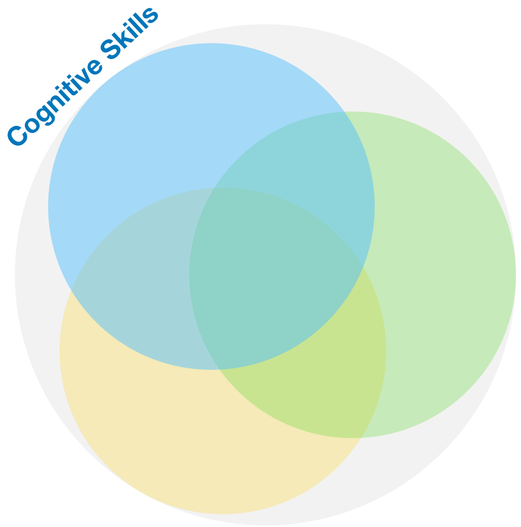
Cognitive Skills

Researchers aim to identify distinct, testable cognitive skills that are used by tasks.

Theory that training specific cognitive skills can improve performance in related work.

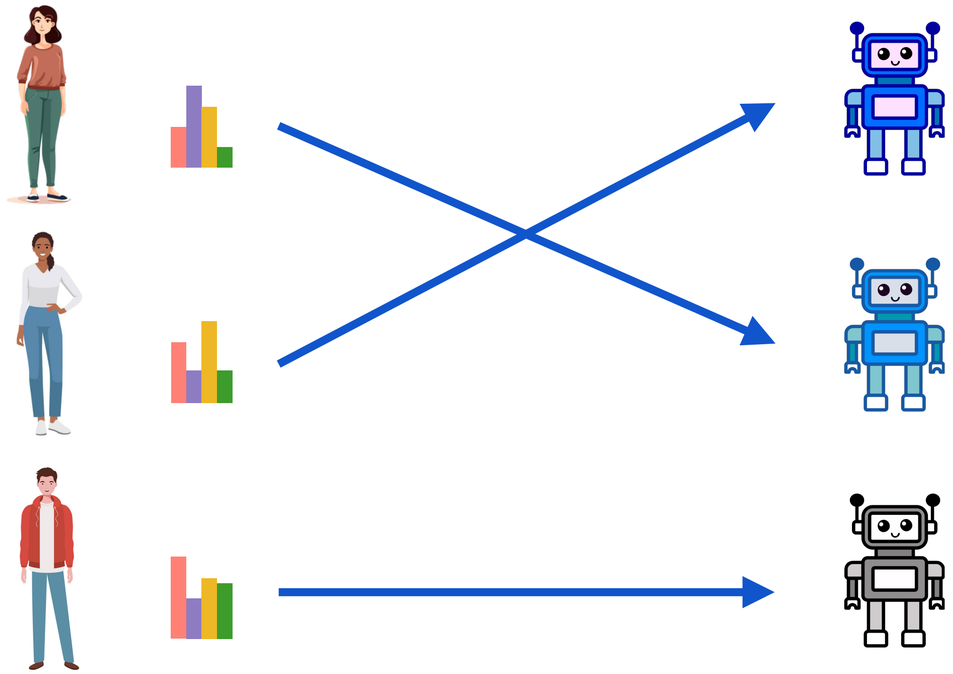
Could we apply this to predict robot teleoperation performance?



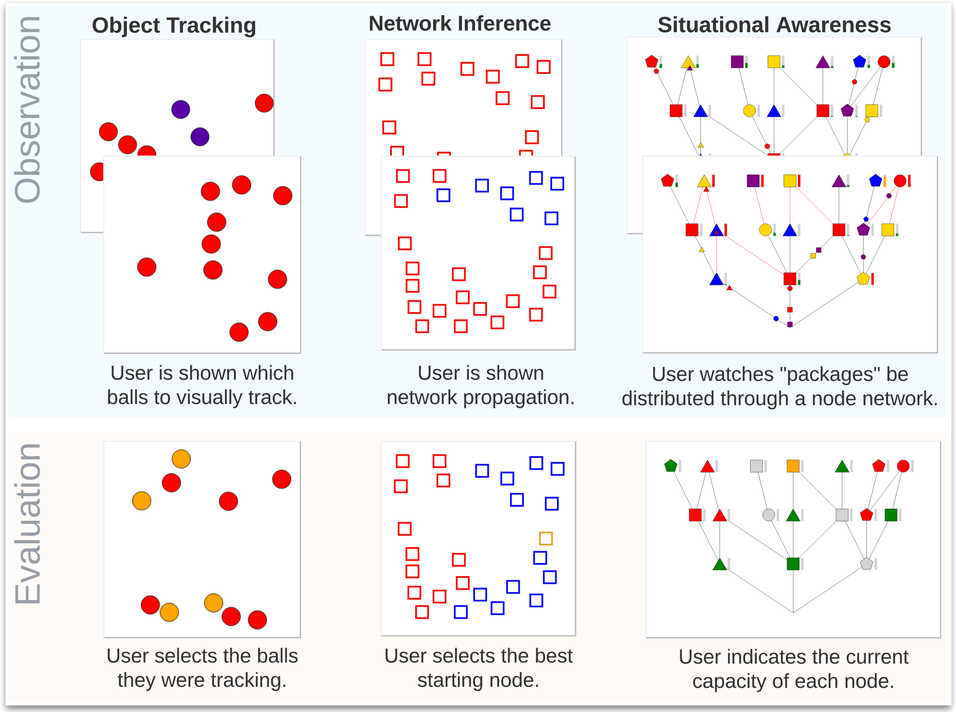


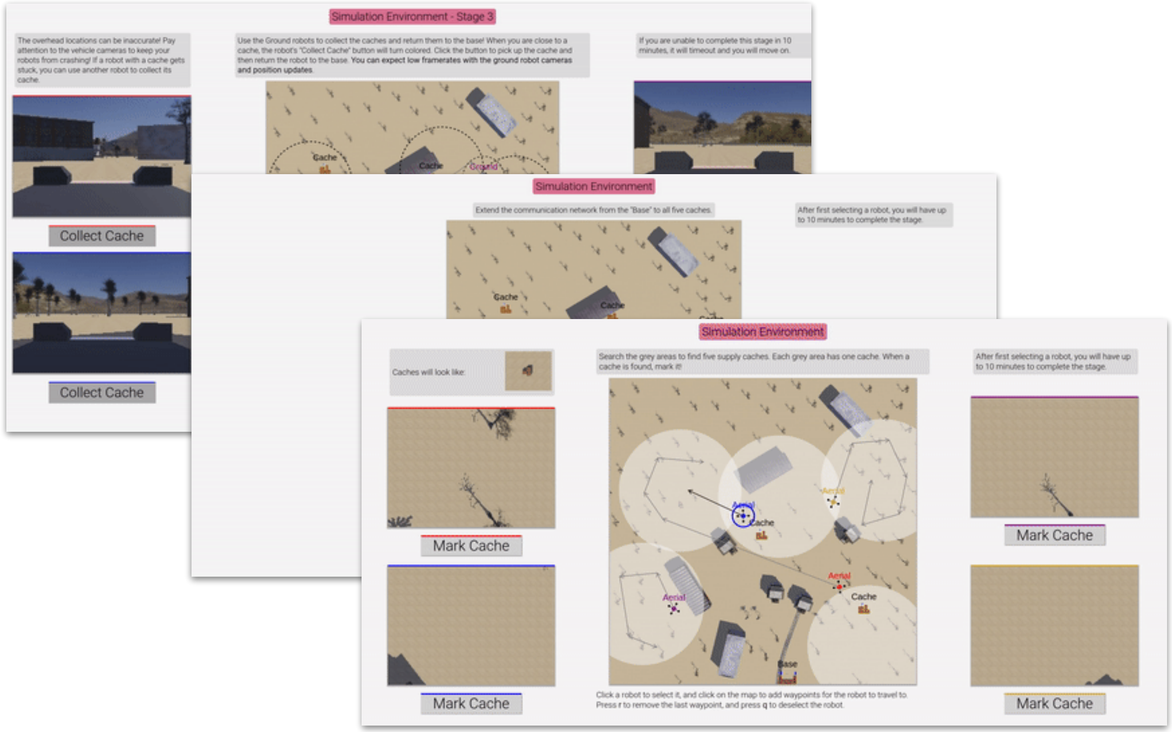
Cognitive Skills

Can we predict **future teleoperation performance** only using cognitive skills, and apply it to **role assignment**?



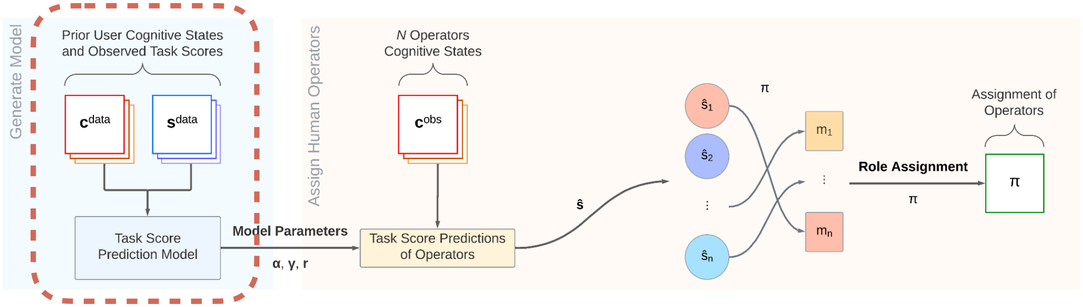
Overview

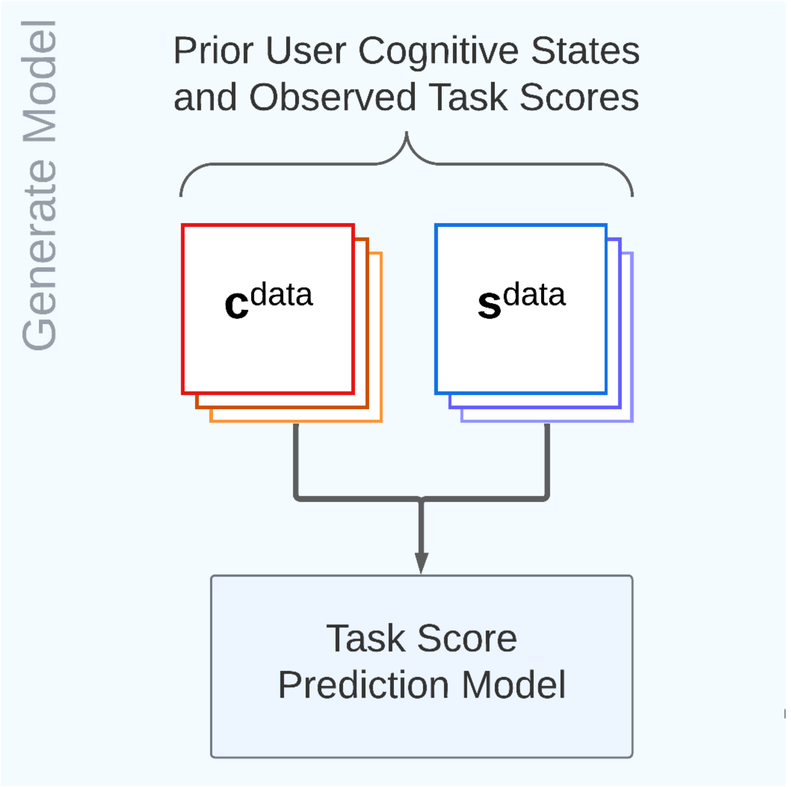




Can we give people cognitive tests and use their scores to predict robot teleoperation performance and optimize team role assignment?

Methods





Construct a model to predict task scores:

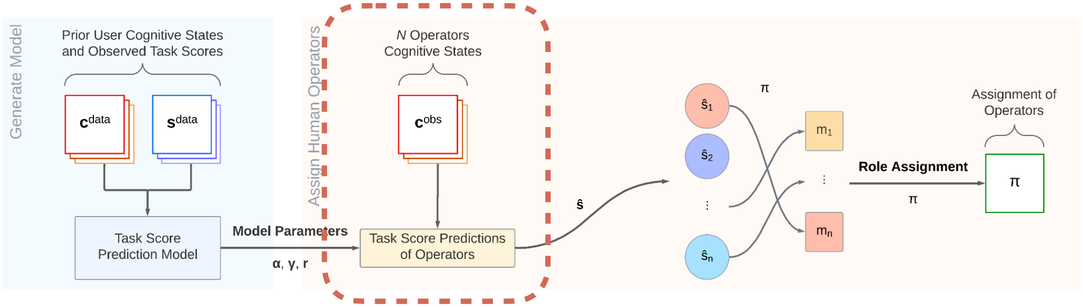
- In previous sessions, collect a dataset of participants taking U cognitive skill tests and then M robot teleoperation tasks.

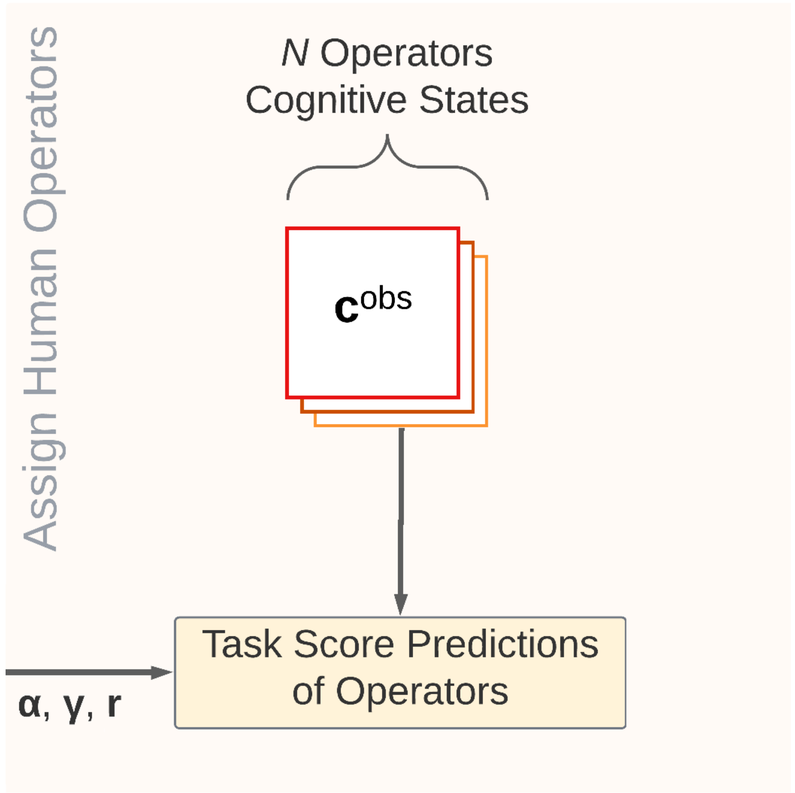
c^{data} s^{data}

- Fit regressions to each teleoperation task and cognitive skill pairing (M x U regressions).
- Return the regression slopes, y-intercepts, and correlations.

α γ r

Methods





Predict the task scores of a new participant team:

- Conduct cognitive skill tests for the participants.

$$c^{obs}$$

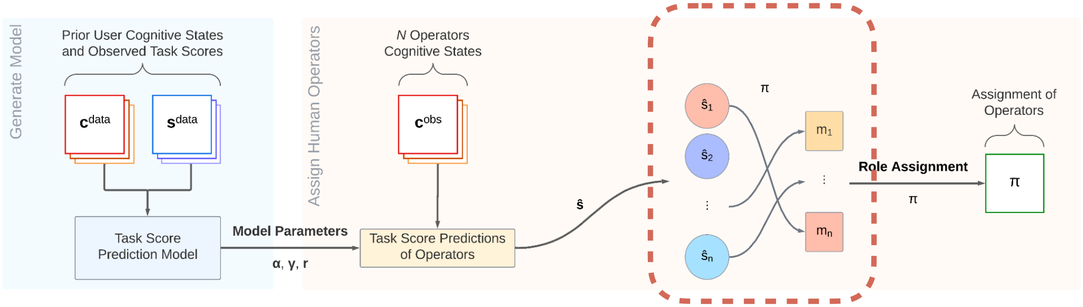
- Predict the teleoperation task scores for each participant *n*.

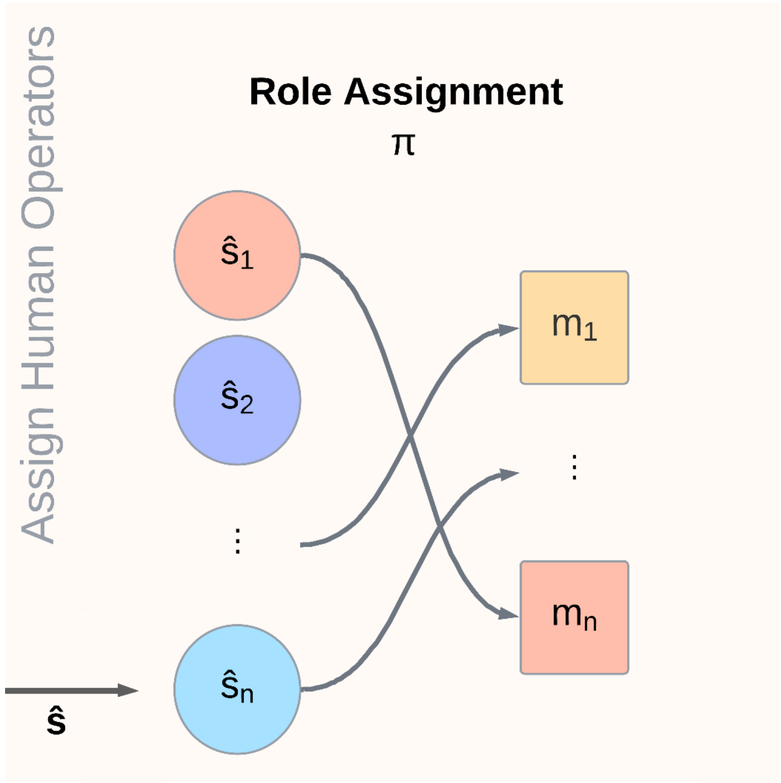
$$\hat{\mathbf{s}}_{n',m} = \sum_{u=0}^{U} \gamma_{m,u}(\alpha_{m,u}\ \mathbf{c}^{\text{obs}}_{u,n'} + \beta_{m,u}) \qquad \gamma_{m,u} = \frac{|\,r_{m,u}\,|}{\sum_{u'=0}^{U} |\,r_{m,u'}\,|}$$

- Return the predicted teleoperation task scores.

$$\hat{\mathbf{s}}$$

Methods





Conduct role assignment with the predicted task scores:

- Choose the role assignment π that maximizes the team's cumulative predicted scores.

$$\mathbf{S} = \sum_{m=1}^{M} \hat{\mathbf{s}}_{\pi(m),m}$$

- Frame as the *Optimal Assignment Problem*, also known as *"minimum cost allocation"*.

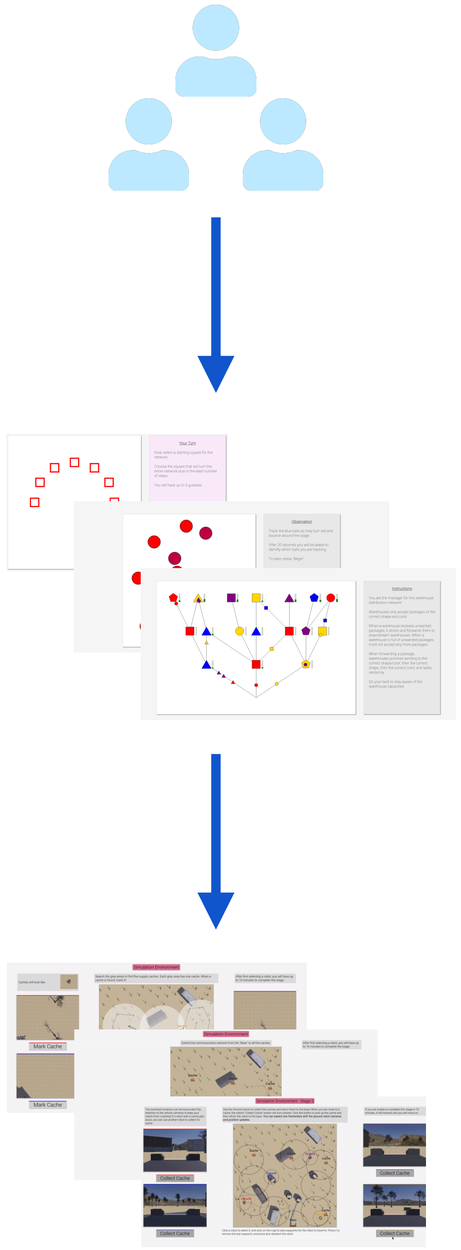
- Return the *Individualized Role Assignment* (**IRA**)

$$\underbrace{\qquad\qquad}_{\pi}$$

Methods

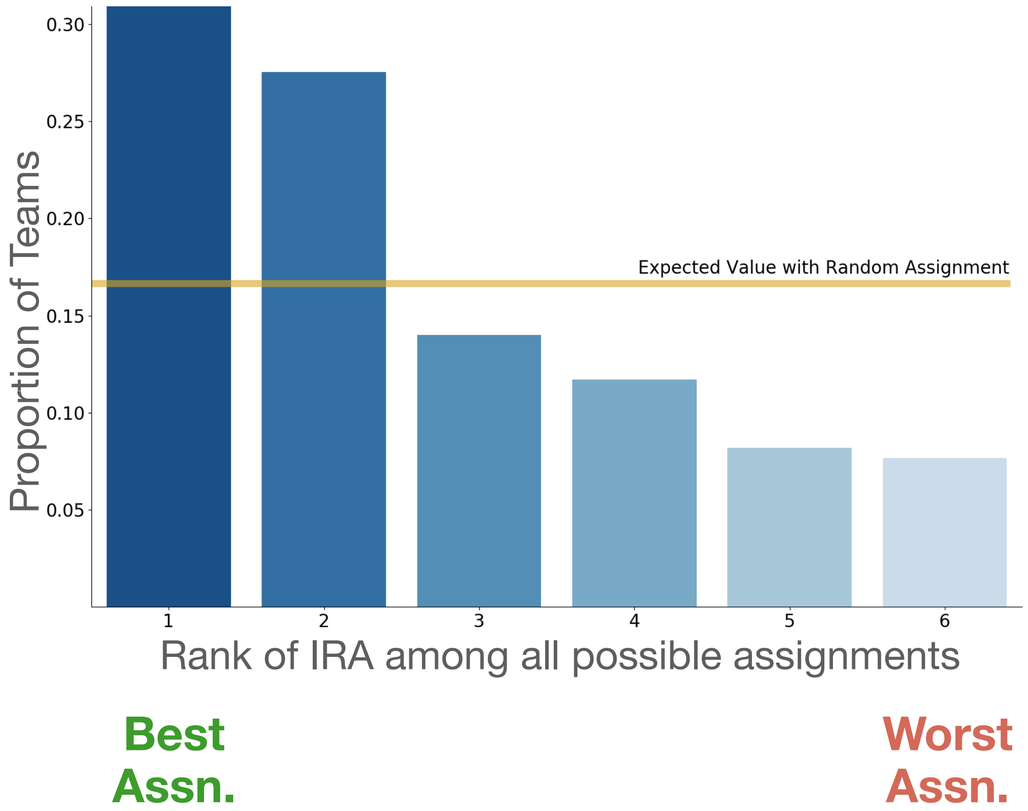
Ran a user study to evaluate our architecture.

- Applied three online cognitive skill tests from the cognitive science literature.
- Developed three online robot teleoperation tasks (using 3D simulation).
- 29 participants individually completed the cognitive skill tests, and then the teleoperation tasks.
- Evaluated participant data post-hoc, using participant subsets to represent teams (29 choose 3 teams).



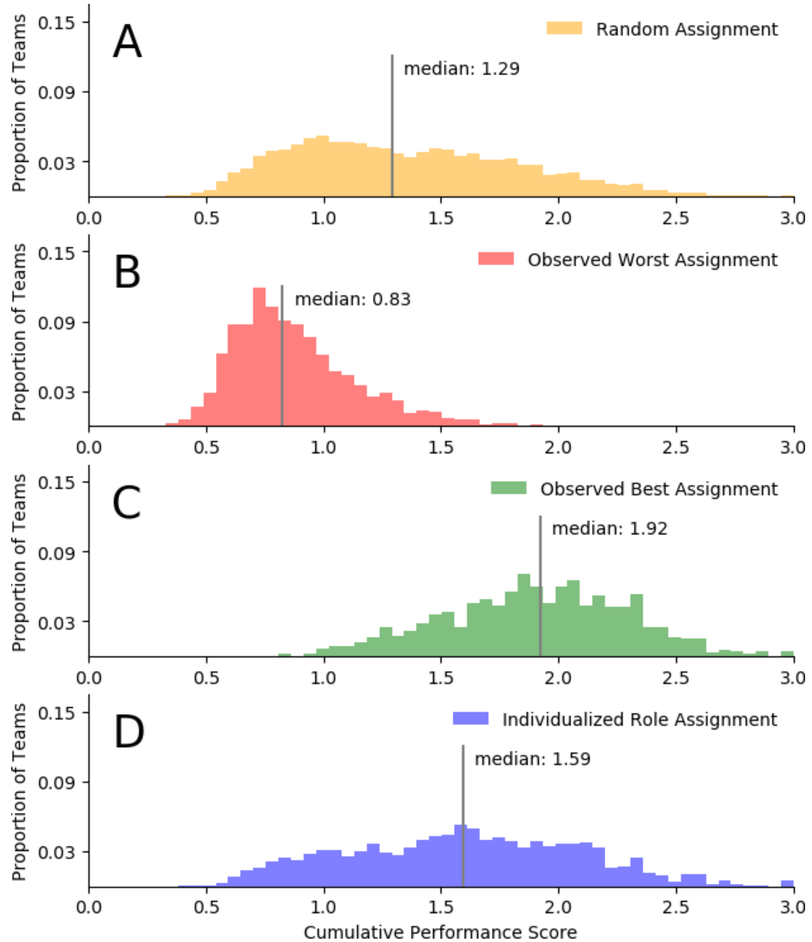
Results

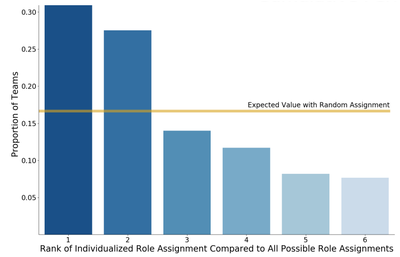
- IRA resulted in **31%** of teams being **optimally** assigned. (*Random: 17%*)
- IRA skewed in favor of better assignments.



Results

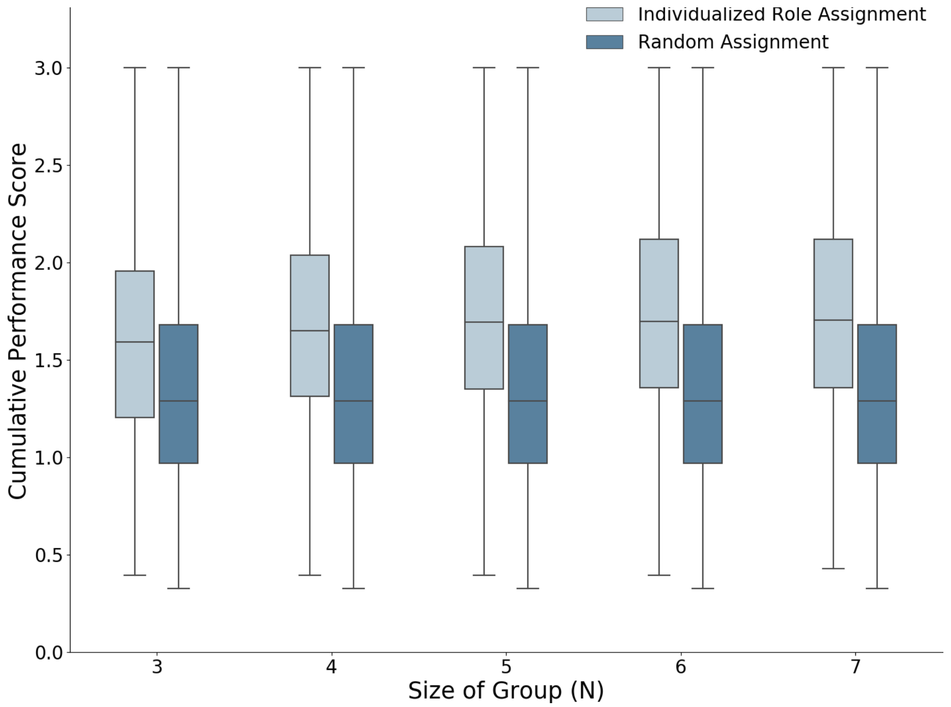
- IRA resulted in **31%** of teams being **optimally** assigned. (*Random: 17%*)
- IRA skewed in favor of better assignments.
- IRA had a median **24%** team score **improvement** over random assignment.
- **73%** of IRAs **outperformed** random assignment. (*Random: 50%*)

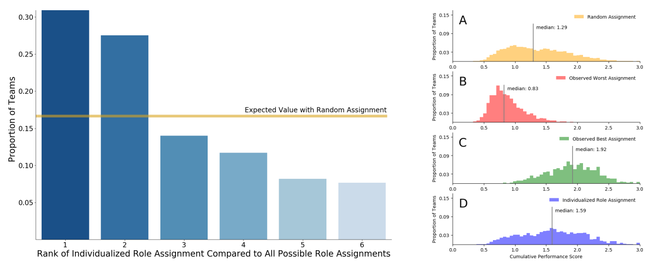




Results

- IRA resulted in **31%** of teams being **optimally** assigned. (*Random: 17%*)
- IRA skewed in favor of better assignments.
- IRA had a median **24%** team score **improvement** over random assignment.
- **73%** of IRAs **outperformed** random assignment. (*Random: 50%*)
- Results held with larger candidate pools (*from 4 choose 3*, *from 5 choose 3*, *etc*)



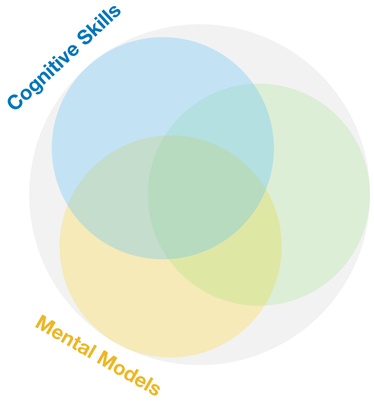


Can we ***predict user performance*** at robot teleoperation
only with information about their ***cognitive skills***?

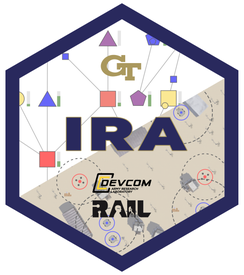
- Cognitive skills <u>can</u> predict future performance at command and control tasks.
- Applications towards role assignment and human-robot teaming are viable.
- Cognitive skill/task relationships hold between-subjects.

Kolb, Jack, et al. "Predicting Individual Human Performance in Human-Robot Teaming."
2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN). IEEE, 2021.

Kolb, Jack, et al. "Leveraging Cognitive States in Human-Robot Teaming."
2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). IEEE, 2022.



How can robots **model** and **apply** a user's **cognitive state** to inform human-robot teaming capabilities?

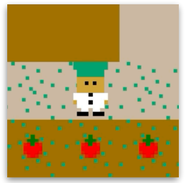


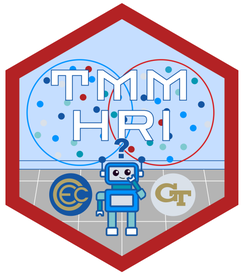
Can we predict future teleoperation performance only using cognitive skills, and apply it to role assignment?

Published in RO-MAN '21, RO-MAN '22

Can we infer user situation awareness via observing users in a **partially-observable** environment?

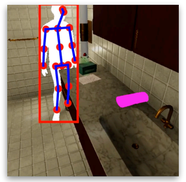
Published in IROS '24

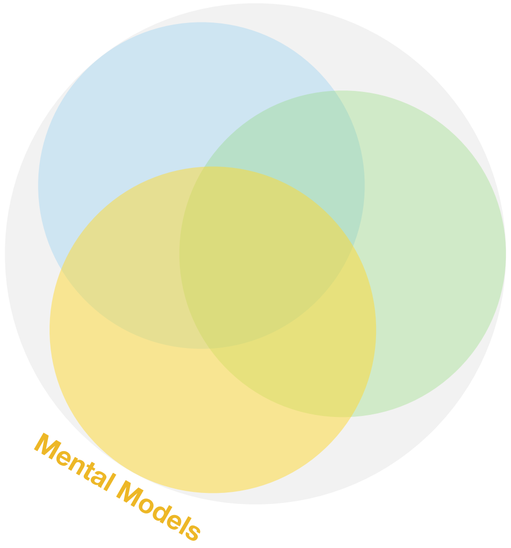




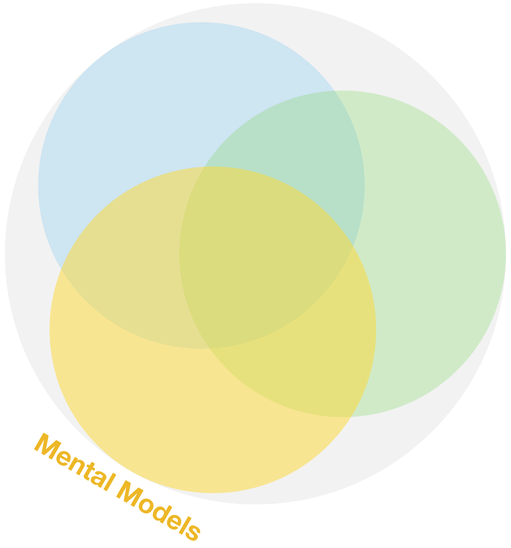
Can we infer user situation awareness via camera observations in a household domain?

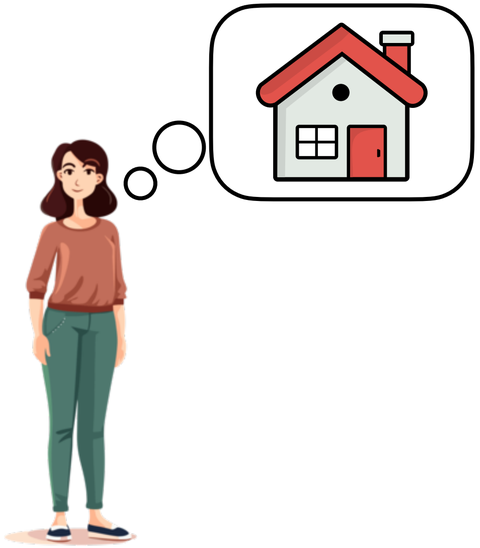
Submitted to RA-L

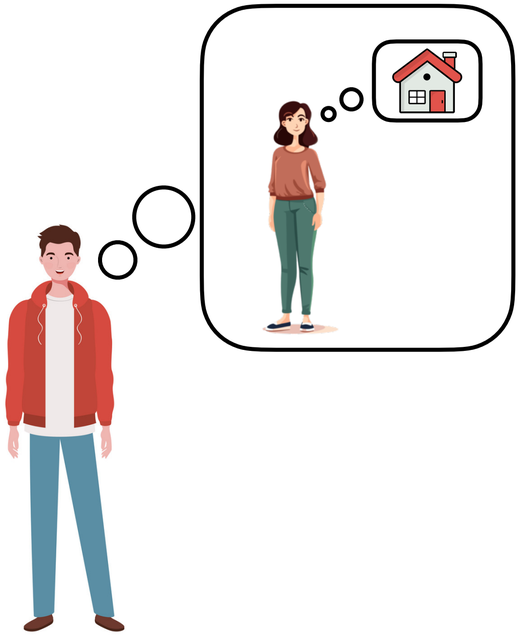




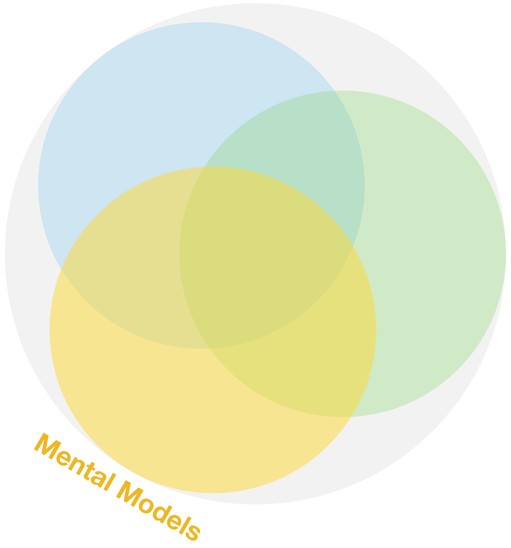
Theory of Mind

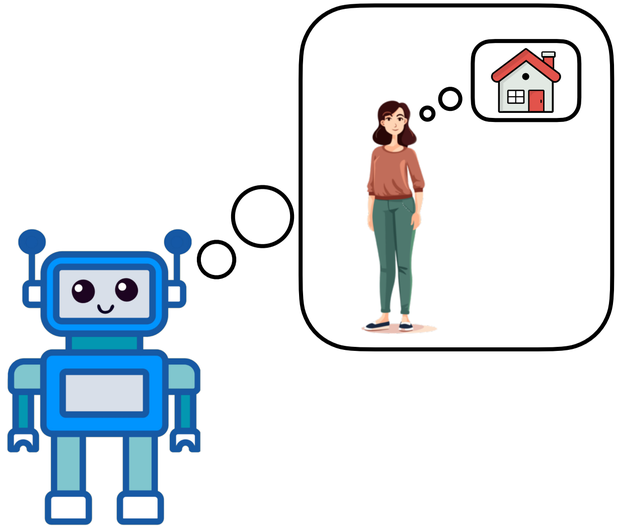




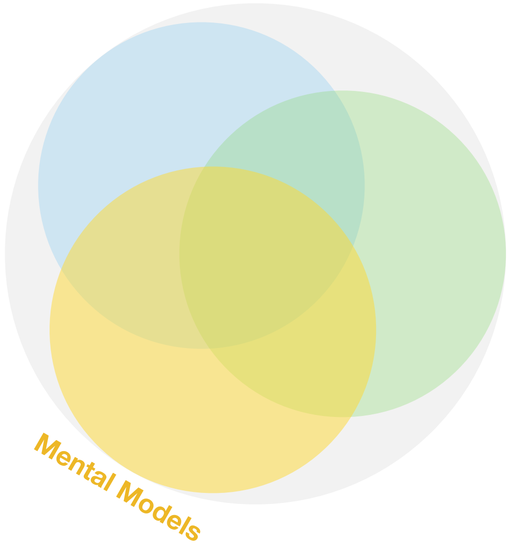


People use **theory of mind** all the time…





What if robots could too?



Terminology

Simulation Theory — We maintain an internal simulation of the environment.

↓

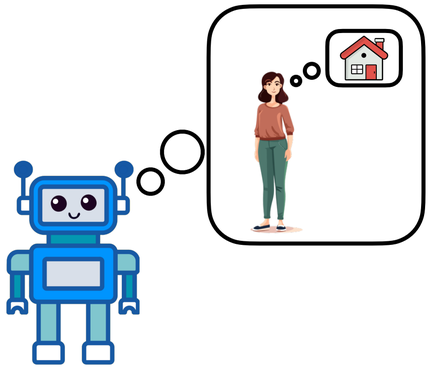
Mental Model — A formal structure of the internal simulation.

↓

Situation Awareness — The person's understanding of the environment.

↓

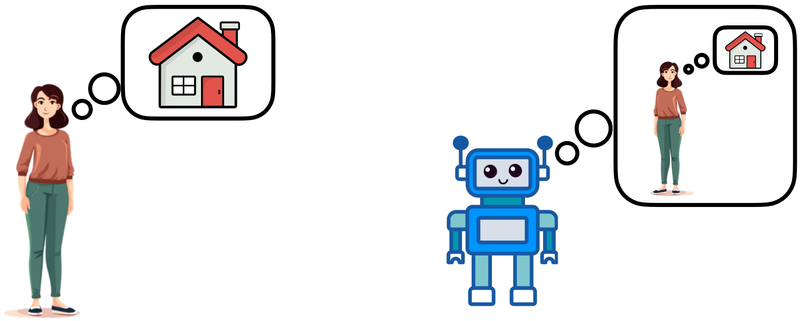
World Belief State — A data structure of situation awareness aspects.



Can we infer a user's world belief state in real-time
to inform human-robot teaming capabilities?

Overview

Can we infer user situation awareness via camera observations in a partially-observable environment?

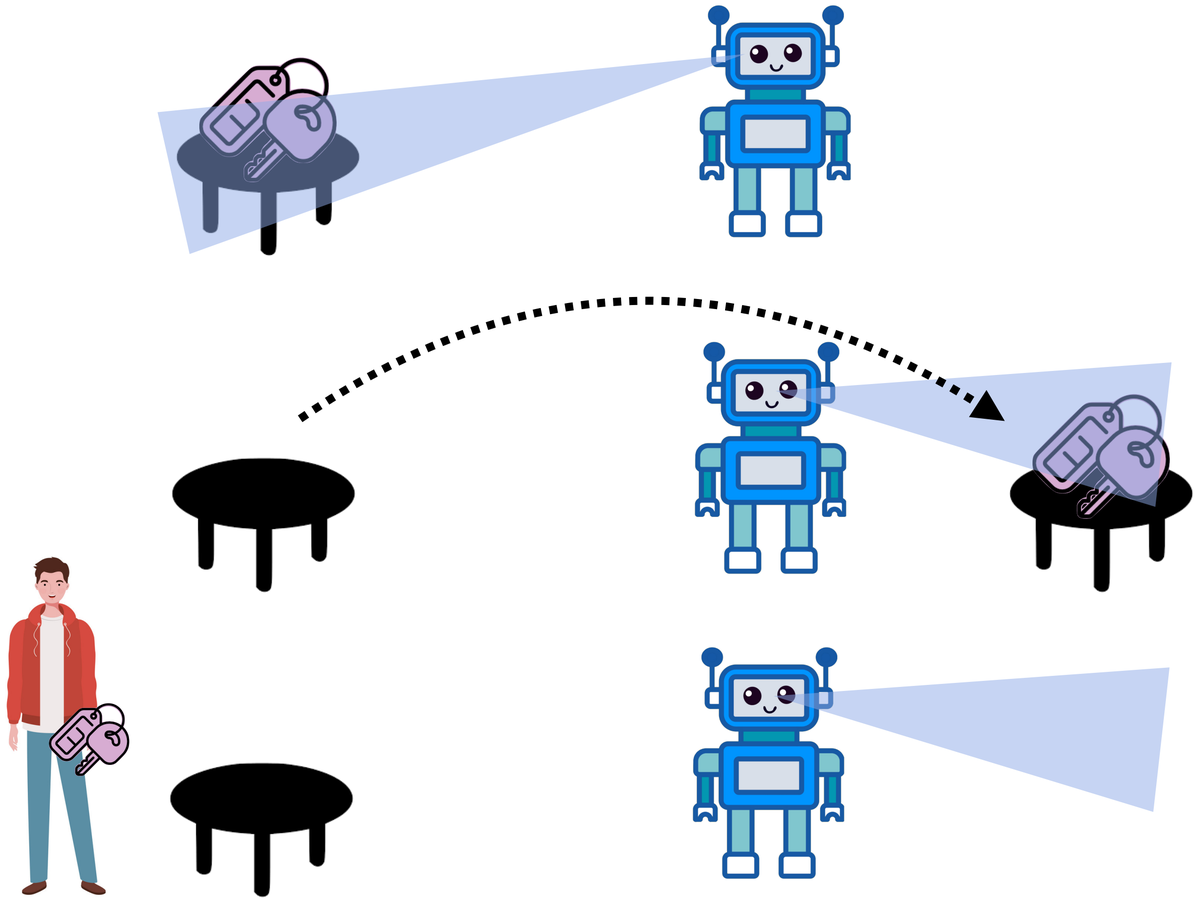


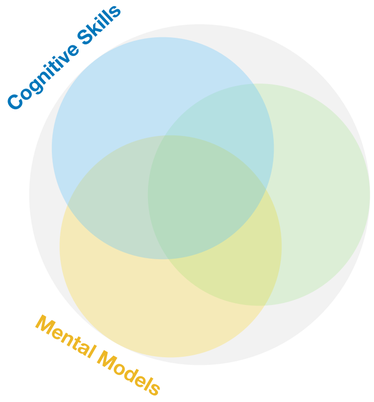
- Level one situation awareness: knowledge of environment elements.
- Useful for:
    - **Communication** — identifying *when* and *what* to communicate (hazards, activities, objects).
    - **Intent estimation** — improving reasoning over human intents.
    - **Query context** — shaping responses to user-specific knowledge.

Challenges

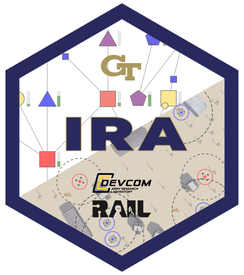
Applying this to household or floorplan environments is challenging:

- Mobile Cameras
- Object permanence
- Dynamic Environments





How can robots **model** and **apply** a user's **cognitive state** to inform human-robot teaming capabilities?

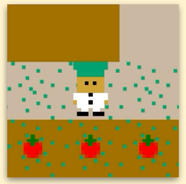


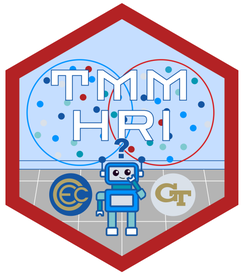
Can we predict future teleoperation performance only using cognitive skills, and apply it to role assignment?

Published in RO-MAN '21, RO-MAN '22

Can we infer user situation awareness via observing users in a **partially-observable** environment?

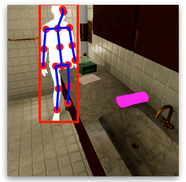
Published in IROS '24





Can we infer user situation awareness via camera observations in a household domain?

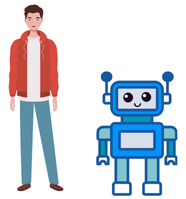
Submitted to RA-L



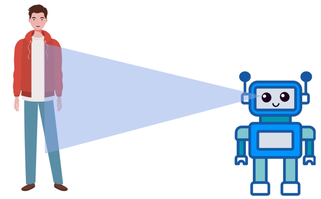
Methods

How well can current methods predict user situation awareness responses?

- Users complete a task with an autonomous agent.

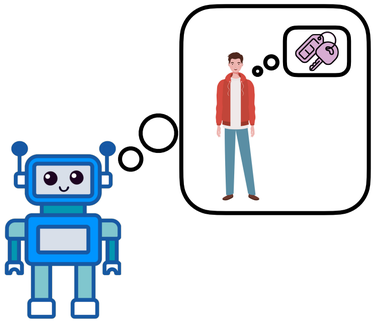


- The agent infers and resolves the user's world **belief state** from its observations.



- Users are regularly asked situation awareness questions (SAGAT format).

- Post-hoc, predict the user situation awareness responses using the inferred world belief state.

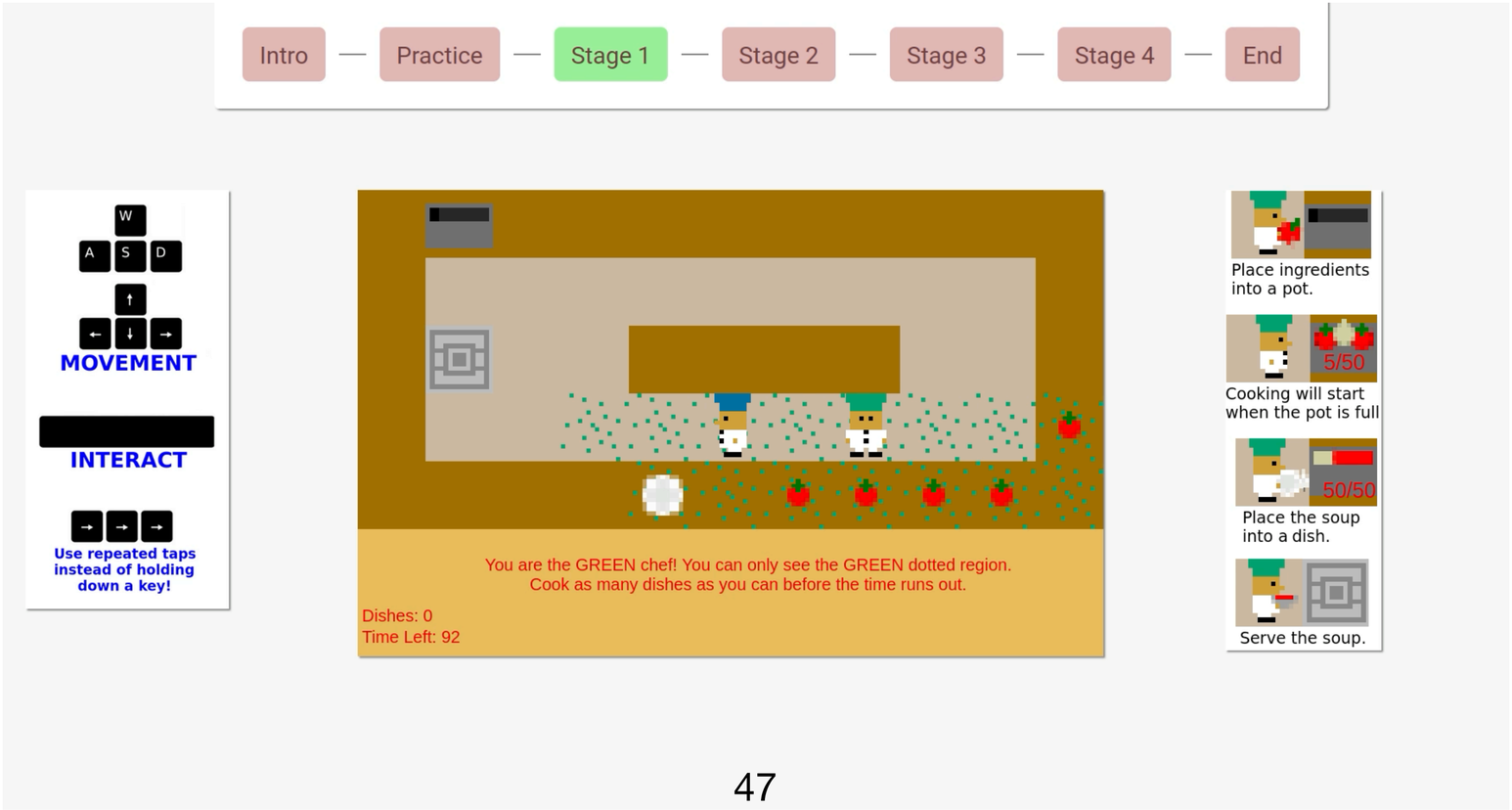


Methods

How well can current methods predict user situation awareness responses?

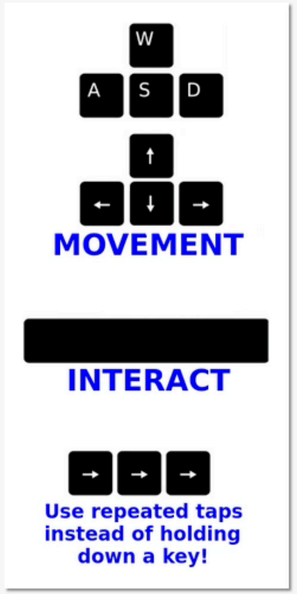
30 participants played a 2D, collaborative, partially-observable cooking game.

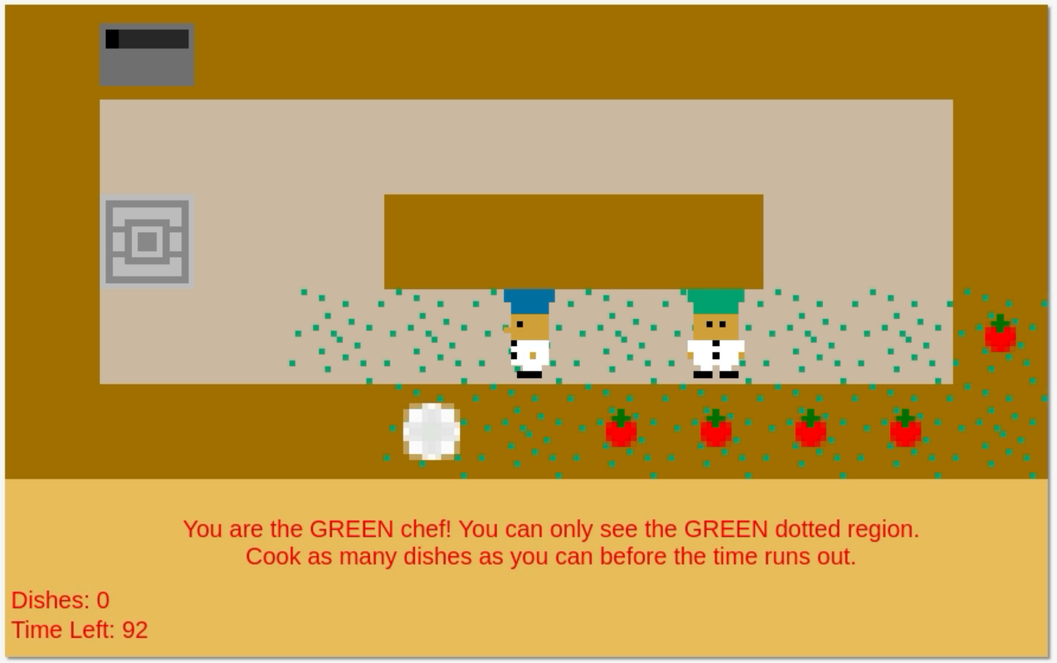
Participants were regularly asked situation awareness questions.



Methods

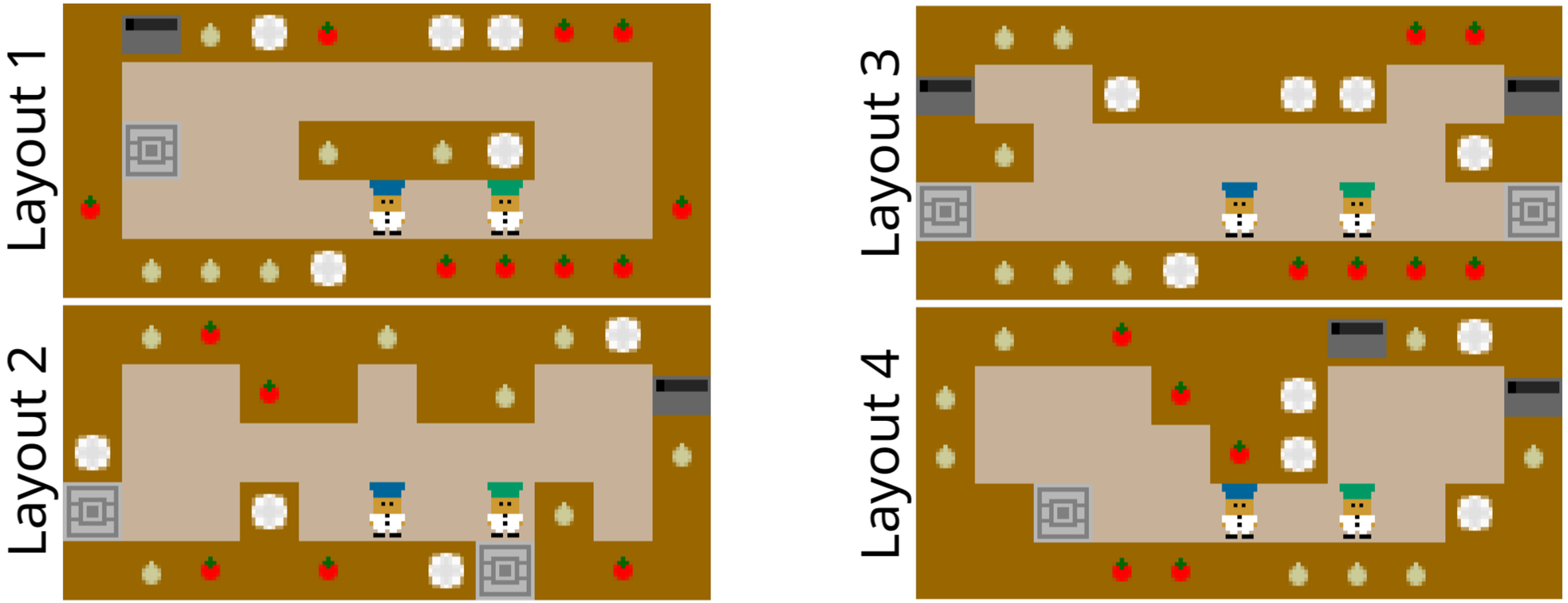
How well can current methods predict user situation awareness responses?





Methods

How well can current methods predict user situation awareness responses?

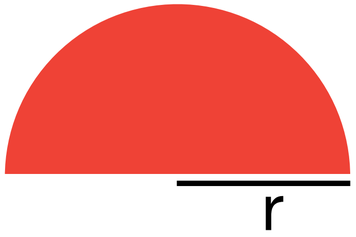


Methods

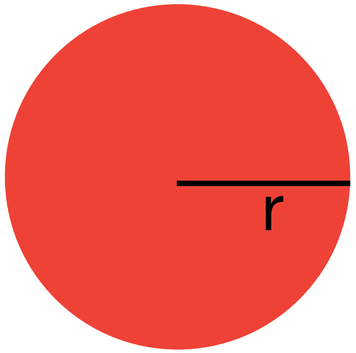
How well can current methods predict user situation awareness responses?

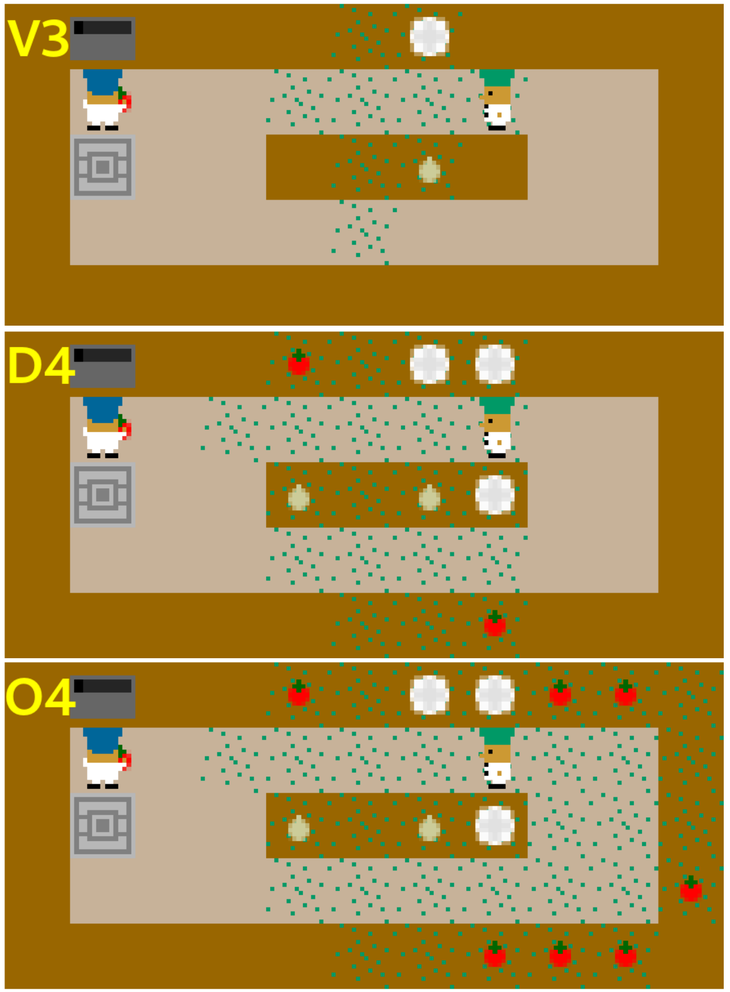
“V” Type

“D” Type



“O” Type





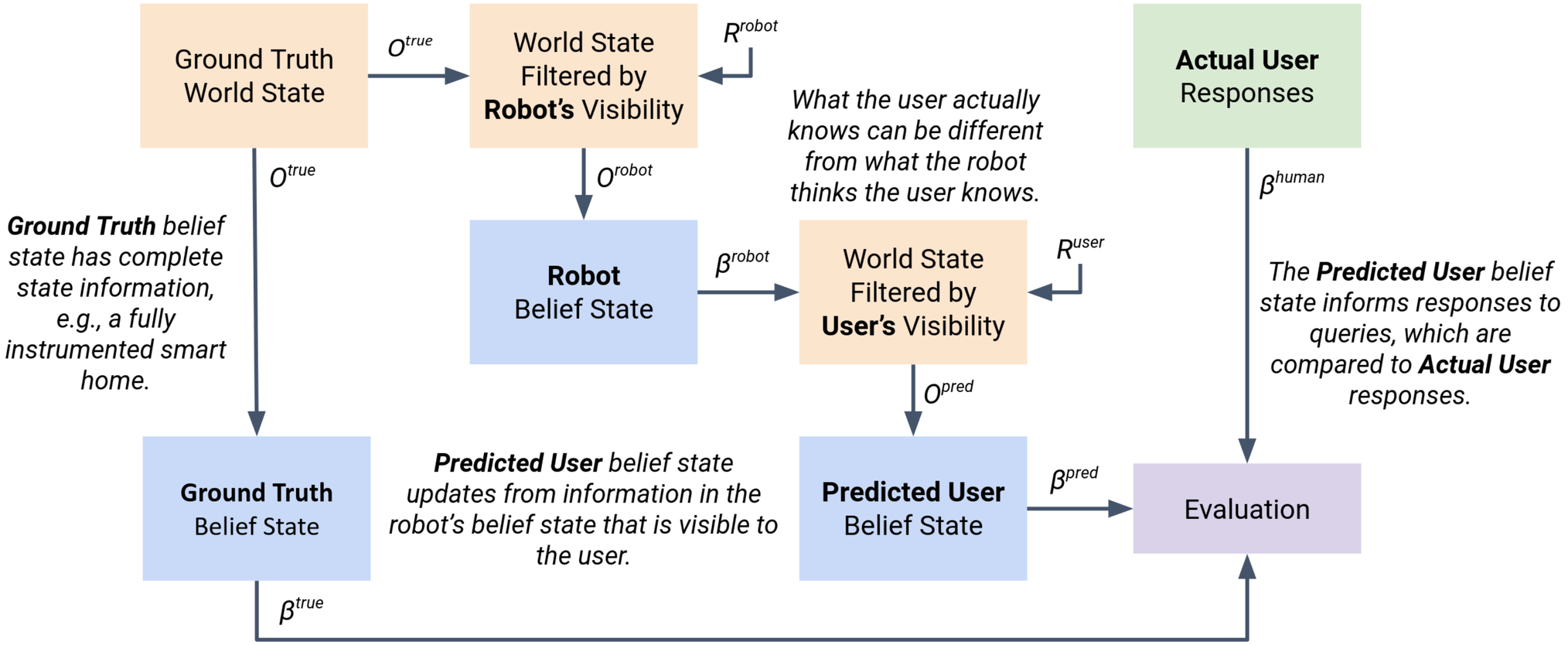
Methods

How well can current methods predict user situation awareness responses?

World State (Level 1)				Context (Level 2)	
Cooking Pot Fullness	**Ingredient Availability**	**Ingredient Location**	**Teammate Location**	**Soups Remaining**	**Cooking Pot State**
77	98	73	94	116	84
332				316	

Methods

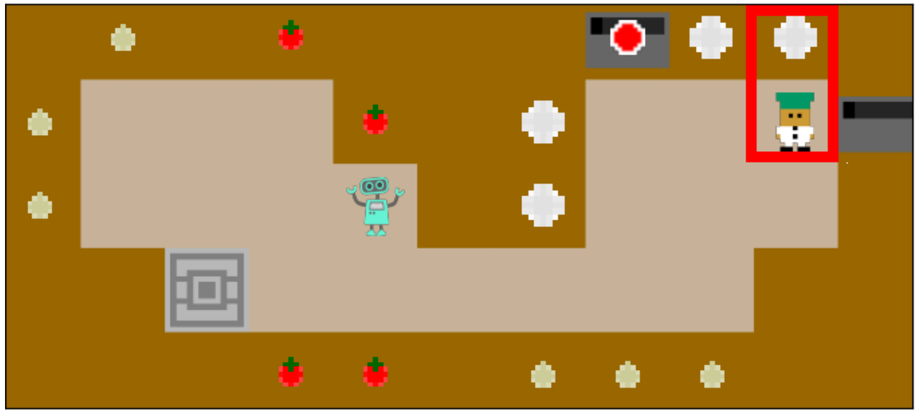
How well can current methods predict user situation awareness responses?



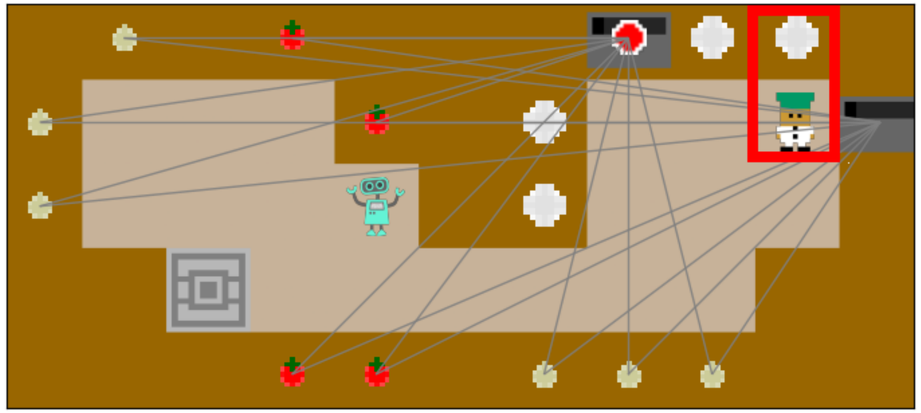
Methods

How well can current methods predict user situation awareness responses?

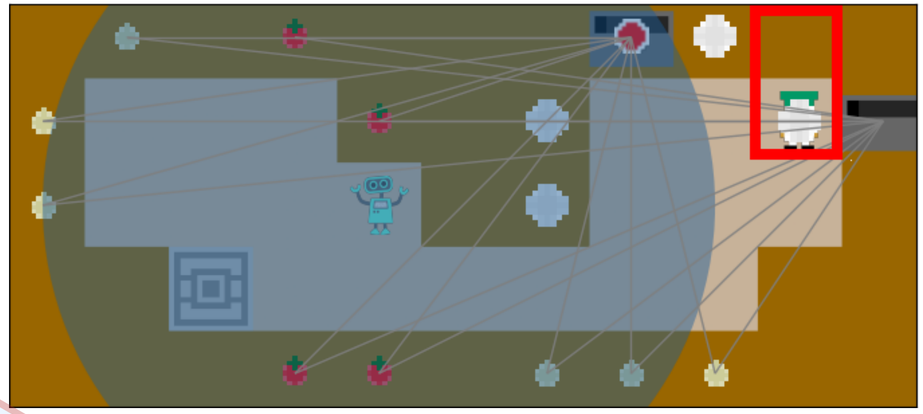
Raw Observations
(direct plot of the world state)



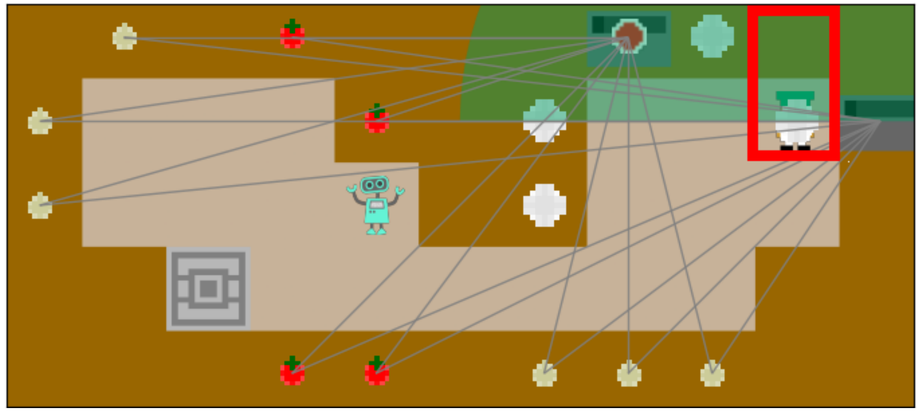
Full Observability Belief State
(has access to all world state information)



Robot's Partial Observability Belief State
(only has access to local world state information)



Estimated Human's Belief State
(uses the robot's belief state to estimate the human agent's belief state)

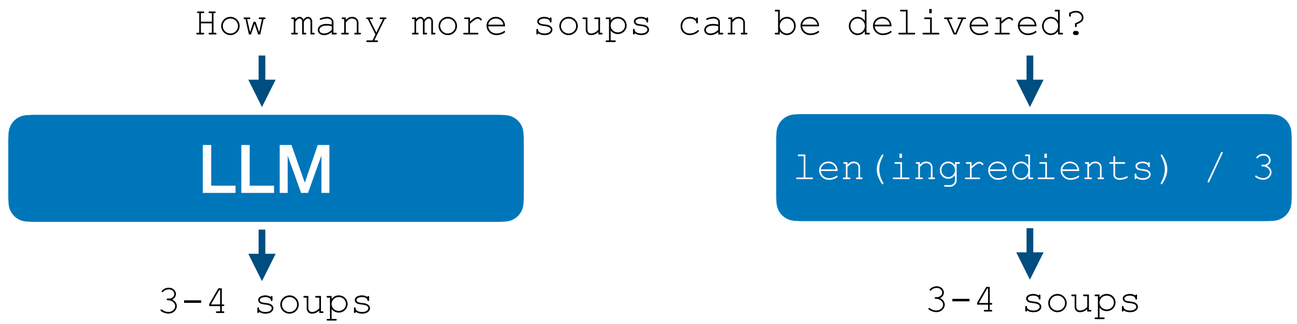


Methods

We implemented two methods for predicting user responses:

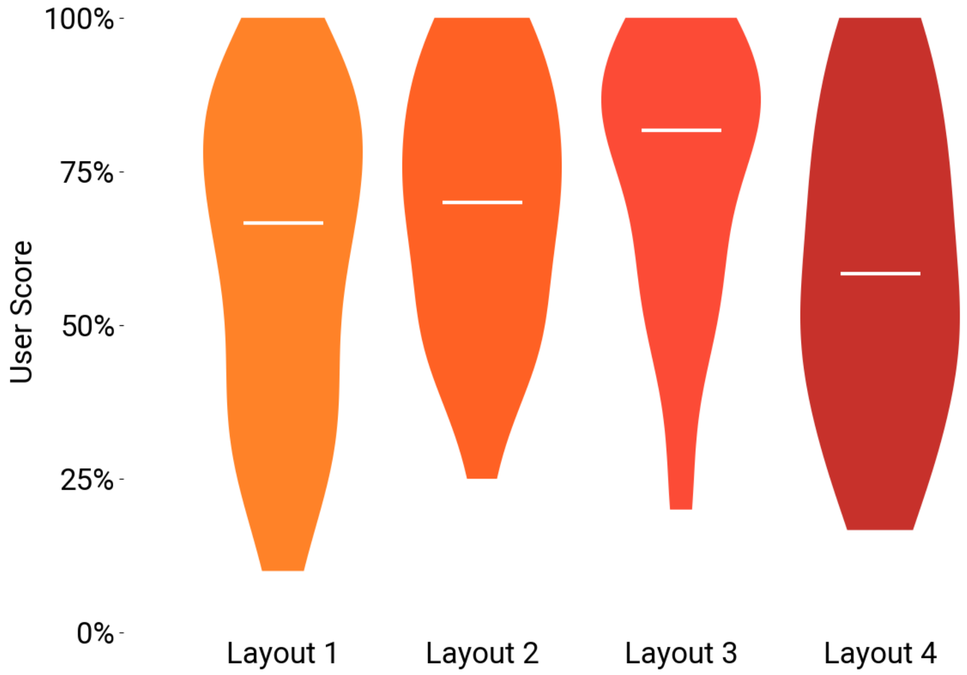
1. **Logical Predicates**: Hand-crafted rules on the user's scene graph to produce the best response to the situation awareness question.

2. **LLM**: Fed the scene graph and game description into GPT4 and asked it to choose the user's most likely response.



Results

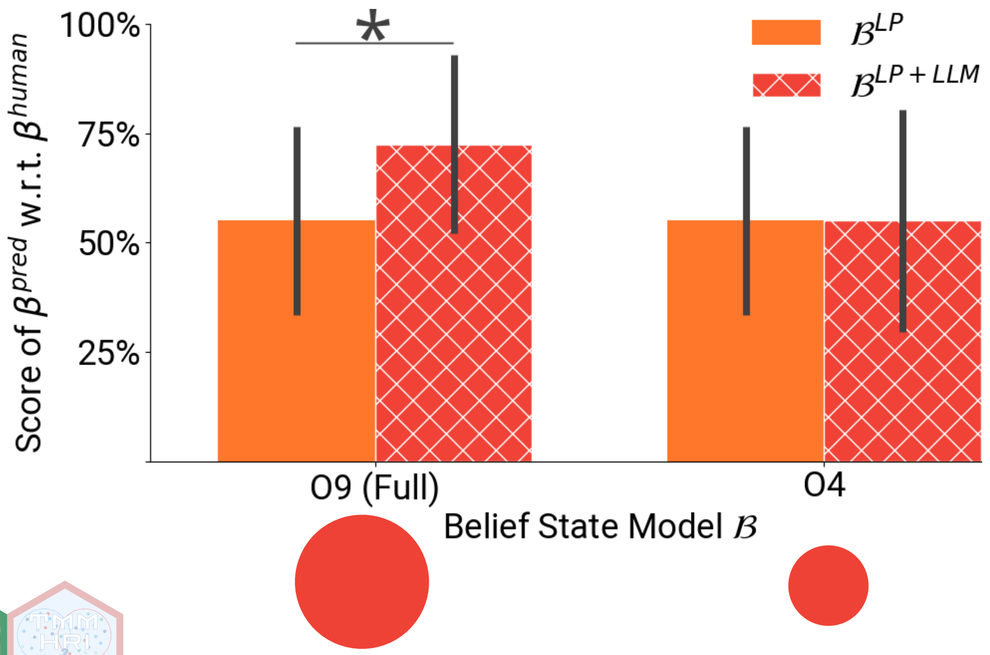
1. Users performed well, averaging between ~60-80% per round (random ~20%).

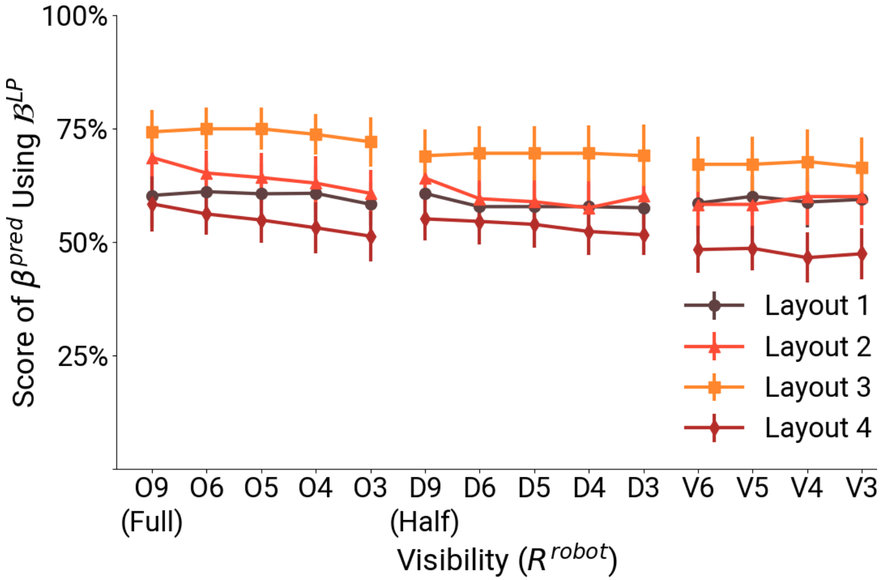


Results

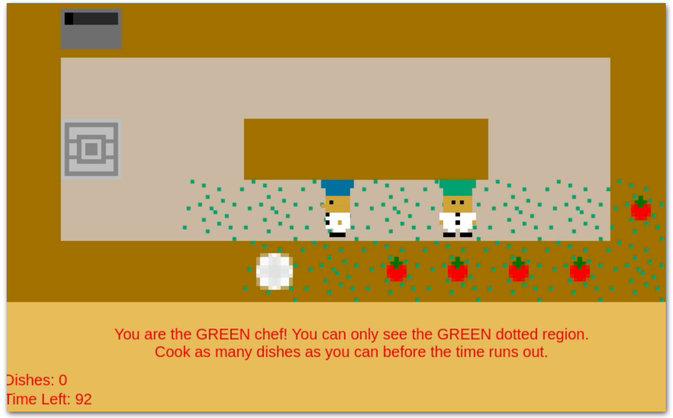
1. Users performed well, averaging between ~60-80% per round (random ~20%).

2. The models were also ~50-75% accurate at predicting user responses.
   Most error was from user error, not false beliefs.



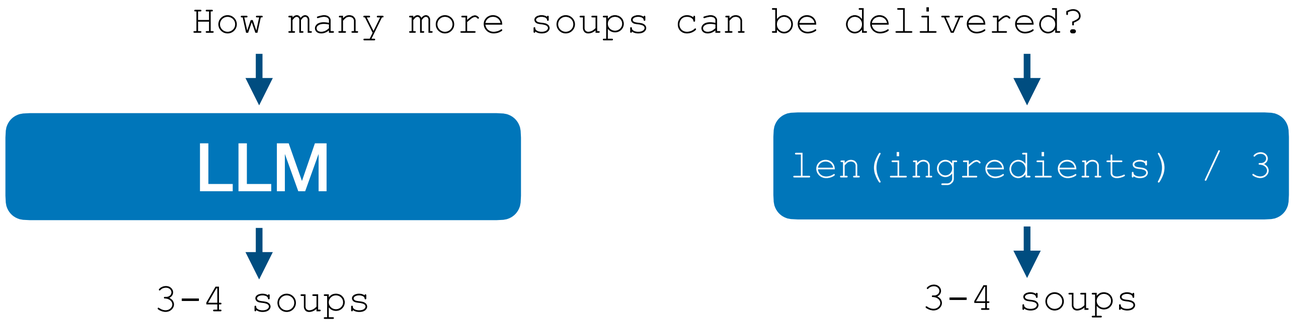


Recap



Augmented the Overcooked domain.

Obtained a dataset of user situation awareness.



Compared two methods at predicting situation awareness responses.

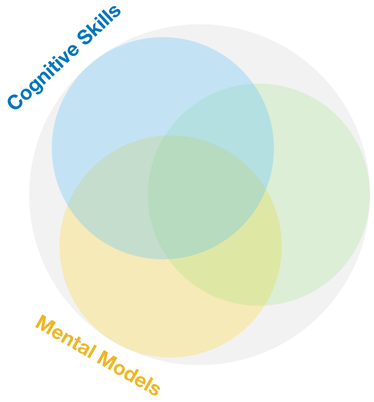
Recap

How well can current methods predict user situation awareness?

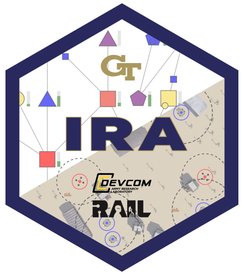
Takeaways:

- The signal is there, we can infer a user's belief state as an explicit and portable data structure.
- The layouts we evaluated were a bit too easy — false beliefs exist but were sparse.

Let's take it to the next level!



How can robots **model** and **apply** a user's **cognitive state** to inform human-robot teaming capabilities?

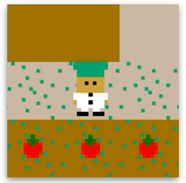


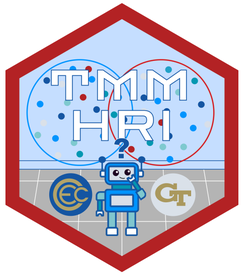
Can we predict future teleoperation performance only using cognitive skills, and apply it to role assignment?

Published in RO-MAN '21, RO-MAN '22

Can we infer user situation awareness via observing users in a **partially-observable** environment?

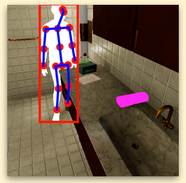
Published in IROS '24





Can we infer a world belief state via camera observations in a household domain?

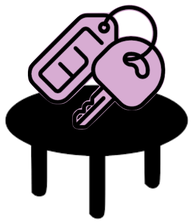
Submitted to RA-L

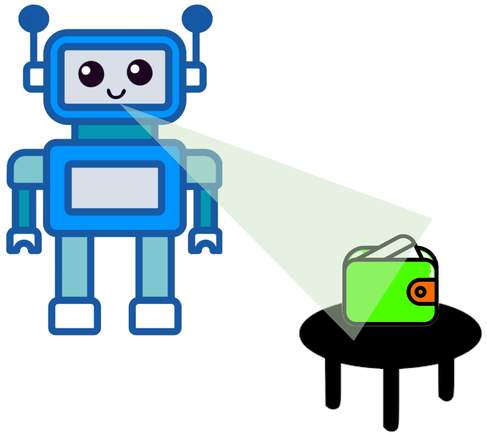


Overview

Can we infer a world belief state via camera observations in a 3D household domain?

- A robot (or smart home) tracks objects in the scene.
- It uses the gaze direction of the person to infer the person's belief state.
- The inferred belief state is useful for a broad range of downstream tasks.

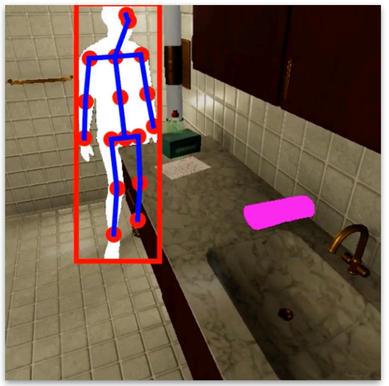


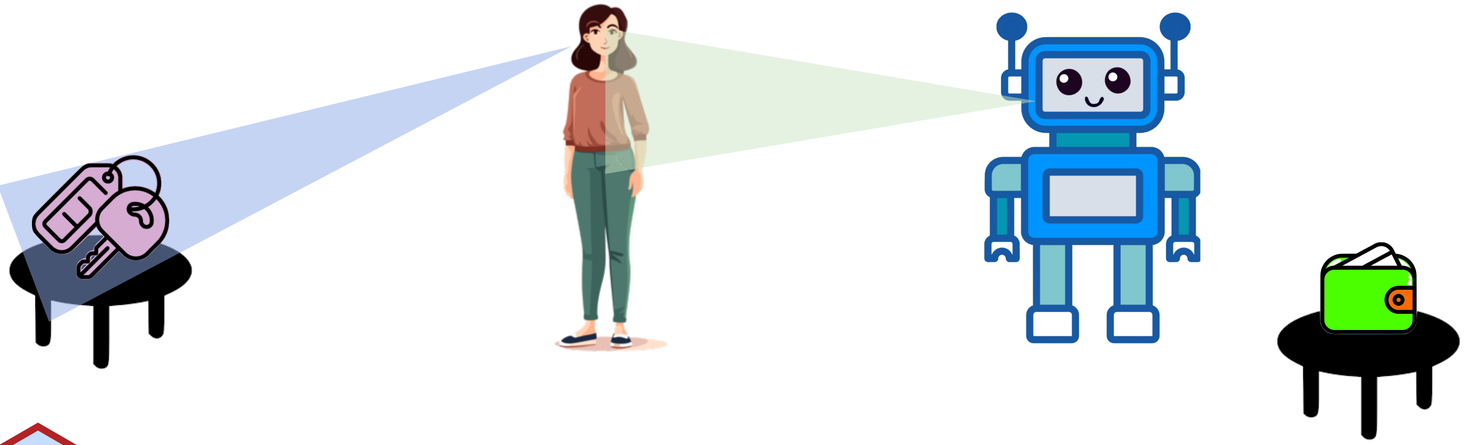


Overview

Can we infer a world belief state via camera observations in a 3D household domain?

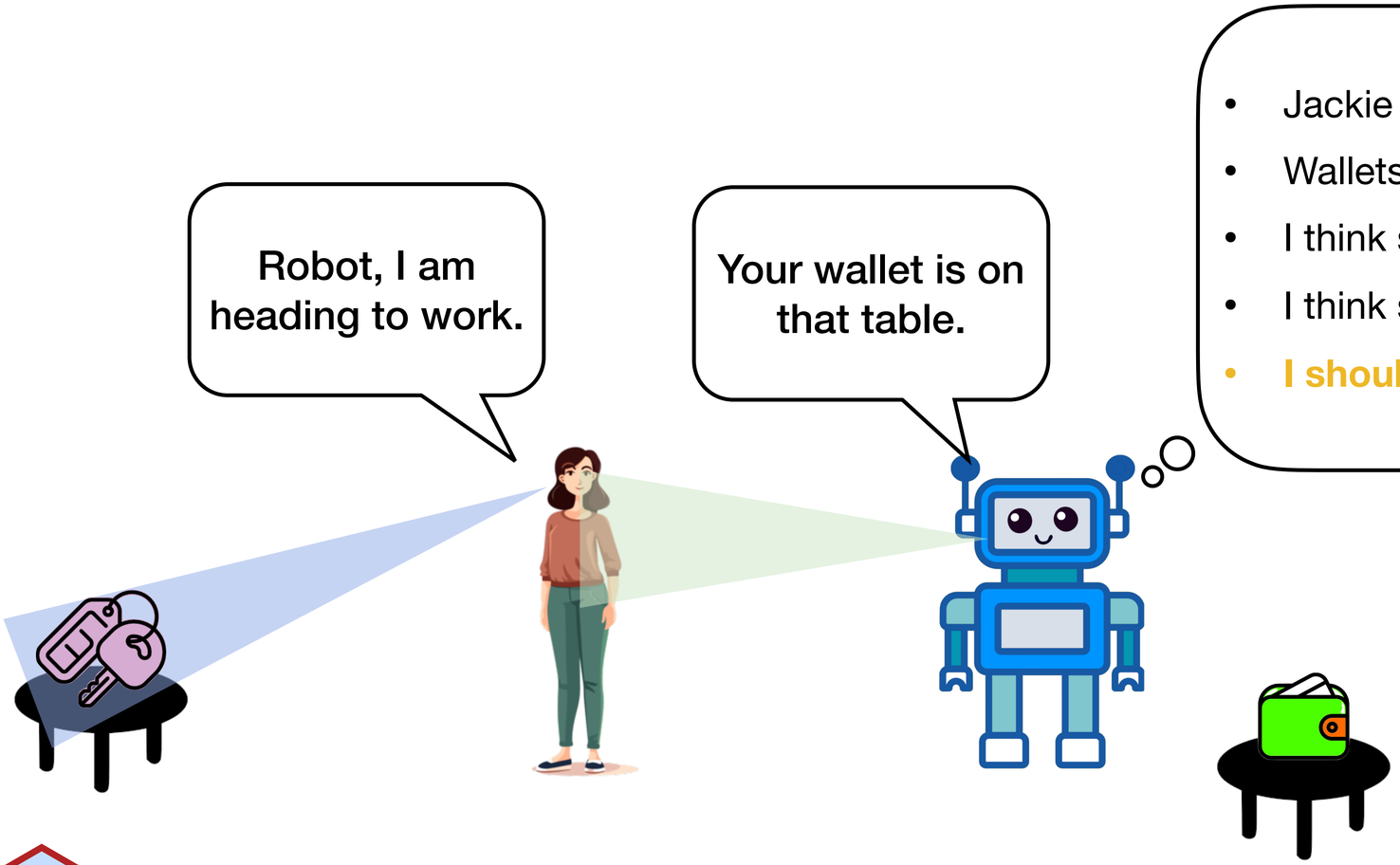
- A robot (or smart home) tracks objects in the scene.
- It uses the gaze direction of the person to infer the person's belief state.
- The inferred belief state is useful for a broad range of downstream tasks.



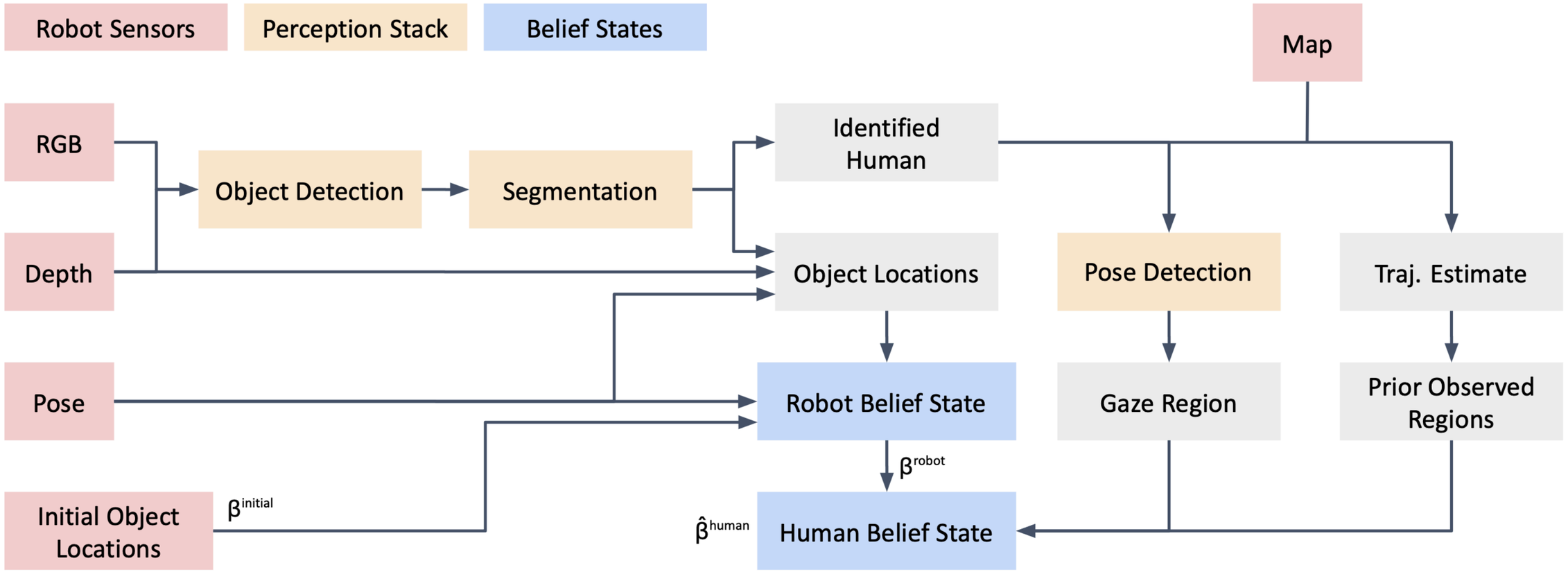


Overview

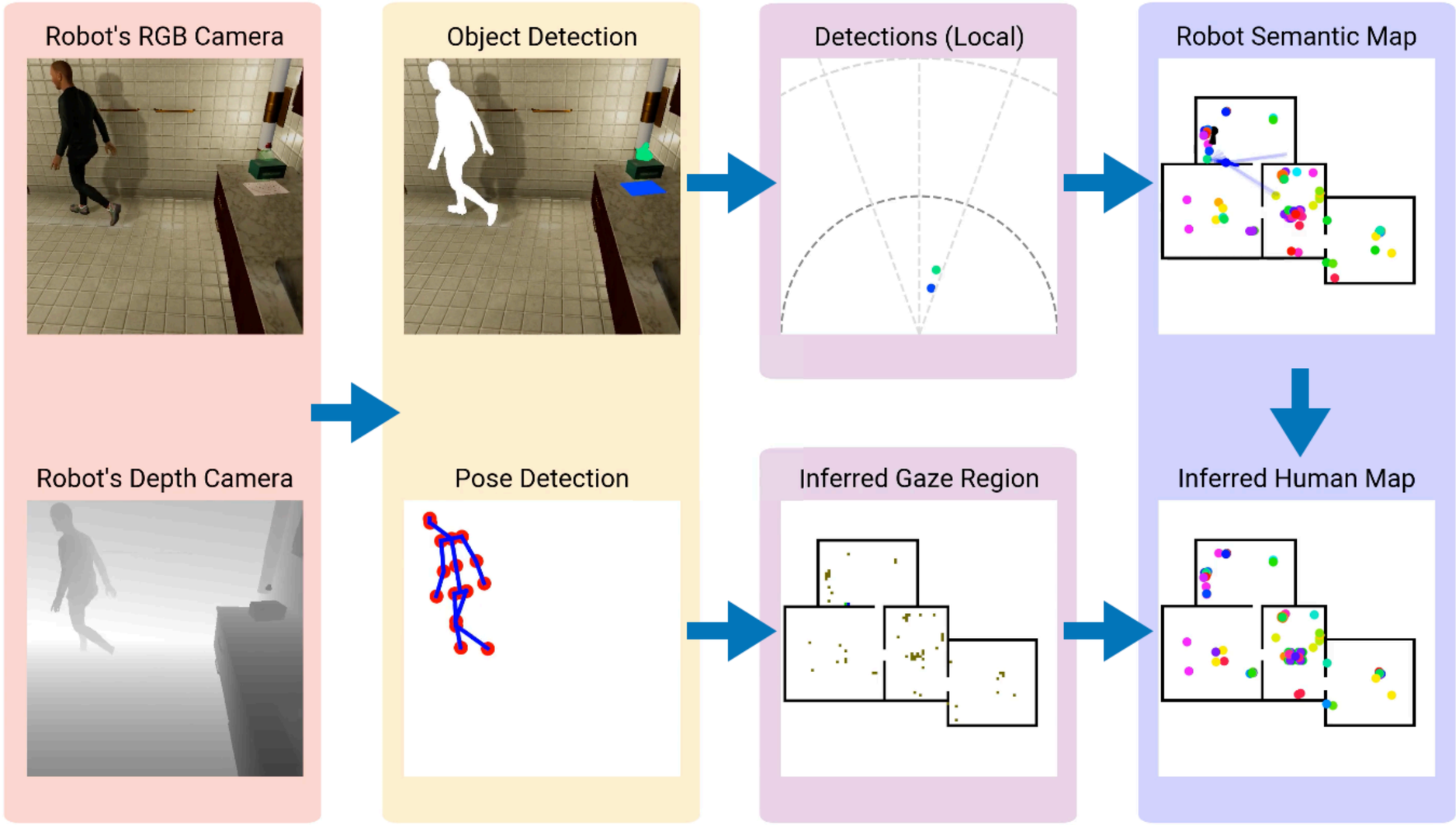
Can we infer a world belief state via camera observations in a 3D household domain?



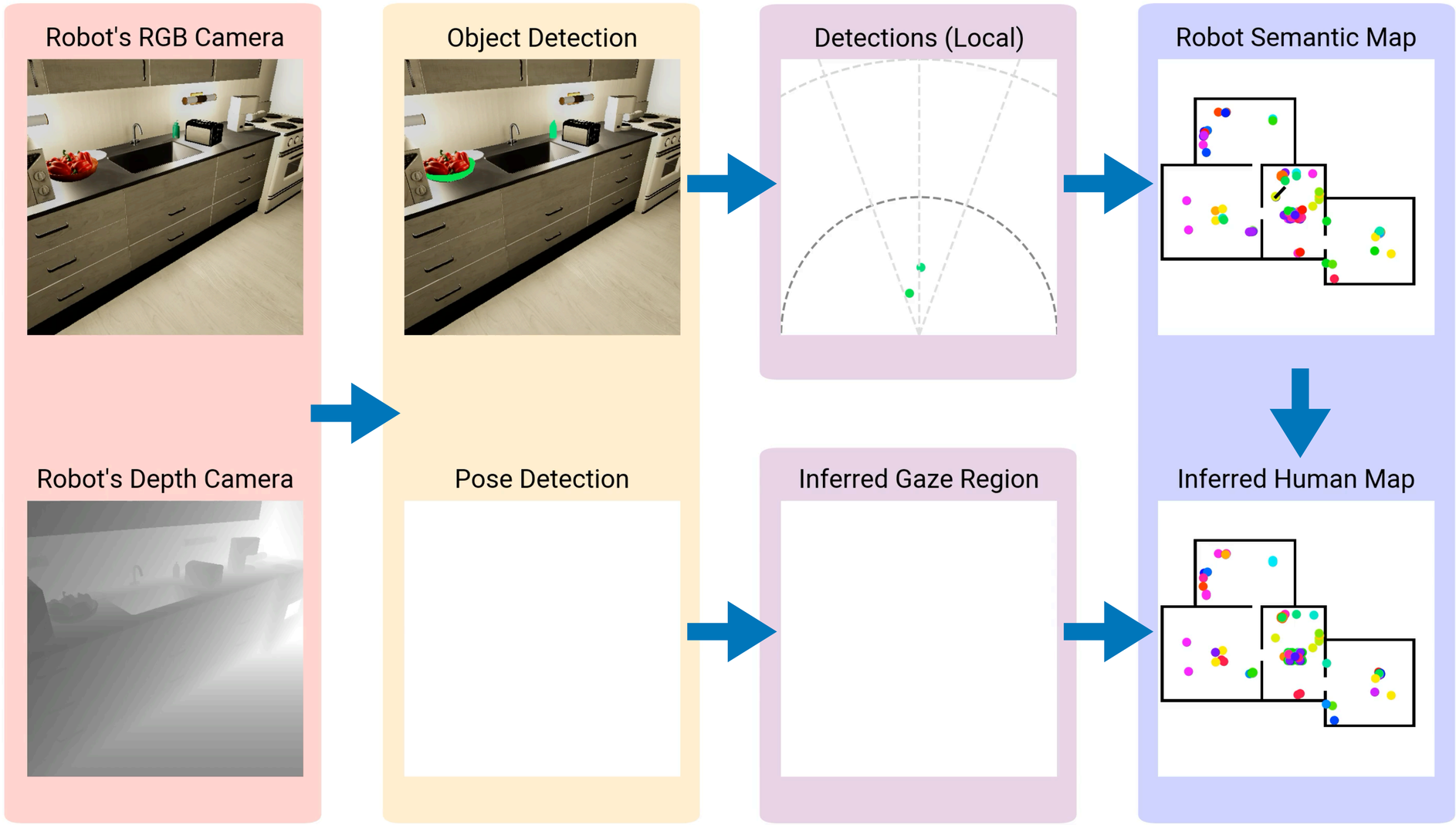
Methods


Robot Sensors
Perception Stack
Belief States
Map
RGB
Object Detection
Segmentation
Identified Human
Depth
Object Locations
Pose Detection
Traj. Estimate
Pose
Robot Belief State
Gaze Region
Prior Observed Regions
β^{robot}
Initial Object Locations
$\beta^{initial}$
$\hat{\beta}^{human}$
Human Belief State

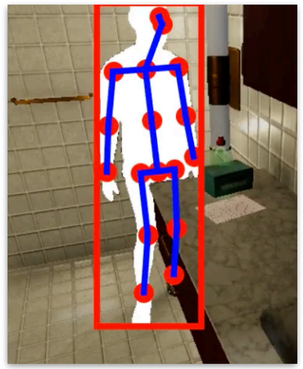
Methods

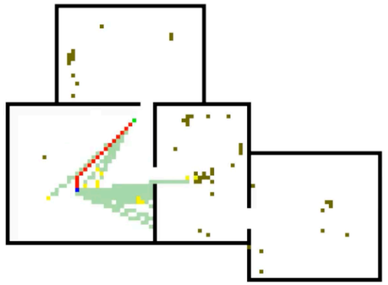

Robot's RGB Camera
Object Detection
Detections (Local)
Robot Semantic Map
Robot's Depth Camera
Pose Detection
Inferred Gaze Region
Inferred Human Map

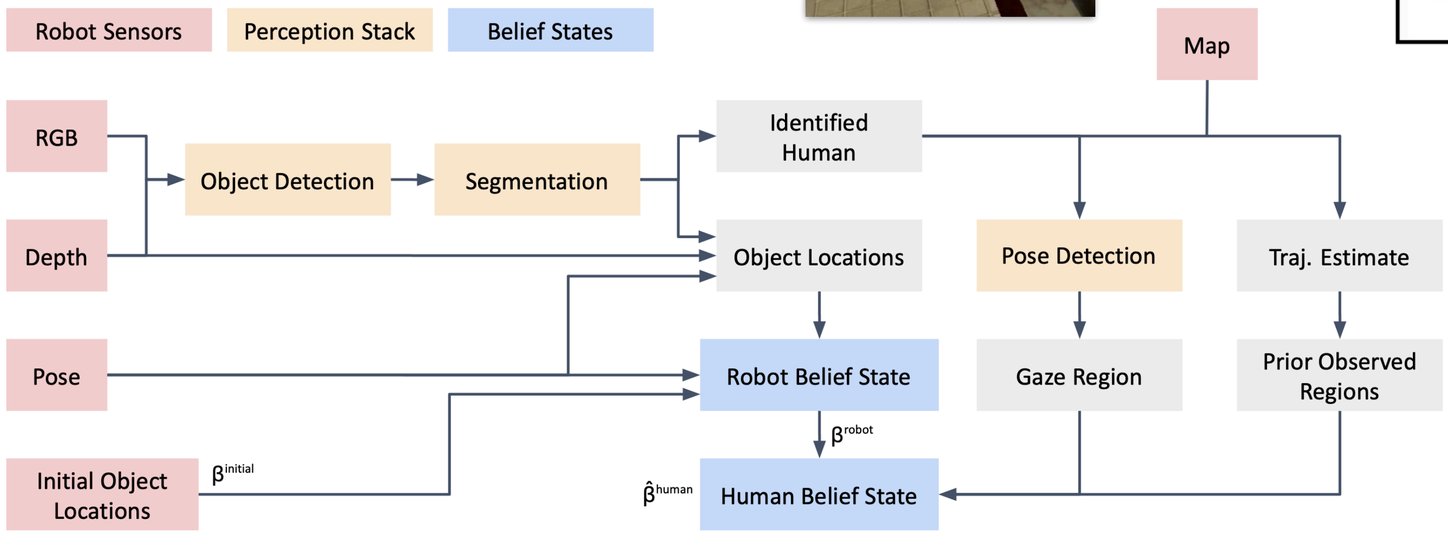
Methods


Robot's RGB Camera
Object Detection
Detections (Local)
Robot Semantic Map
Robot's Depth Camera
Pose Detection
Inferred Gaze Region
Inferred Human Map

Methods







Algorithm 1 Updating a belief state from observed objects.

Require: Sets $\mathcal{O}_c$, β_c for an object class c

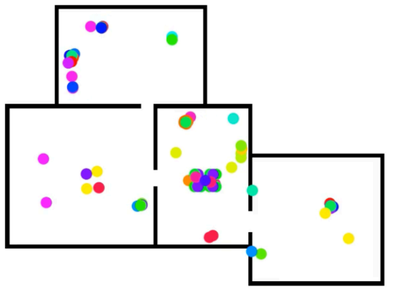
Ensure: Mapping $f : \mathcal{O}_c \to \beta_c$ that minimizes total distance

Find $f : \mathcal{O}_c \to \beta_c$ such that

$$\sum_{o \in \mathcal{O}_c} d(o, f(o))$$

is minimized, where $d(a, b)$ is the L^2-norm distance.

$\beta_c \leftarrow f(\mathcal{O}_c)$

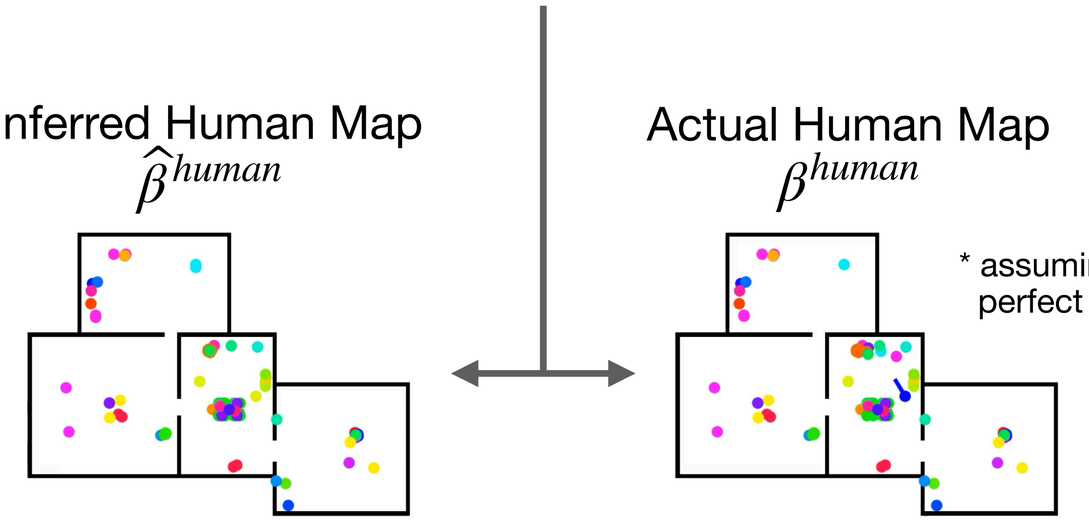


Evaluation

How can we measure the **disparity** between two belief states?

Inferred Human Map $\hat{\beta}^{human}$

Actual Human Map β^{human}



* assuming the human has perfect perception.

We need a set distance metric that is **instance-agnostic.**

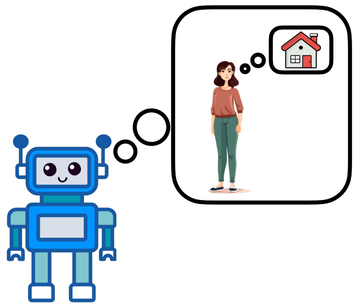
Evaluation

Summed Minimum Cost by Class: the minimum distance to resolve objects of each class.

$$SMCC(A,B) = \frac{\sum_{c \in C} minCost(A_c, B_c)}{|A|} \; ; \; where \quad |A| = |B|$$

minCost(A, B) is the shortest L2 distance to map object locations from set A to set B.

Evaluation

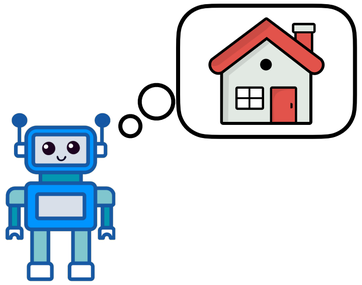


1. $SMCC(\widehat{\beta}^{human}, \beta^{human})$ — **predicted** belief state relative to the **person's** belief state.

   A **low** value means the system works well.

2. $SMCC(\beta^{human}, \beta^{true})$ — the **person's** belief state relative to the **true world** state.

   A **high** value means the environment is complex for the person to have false beliefs.



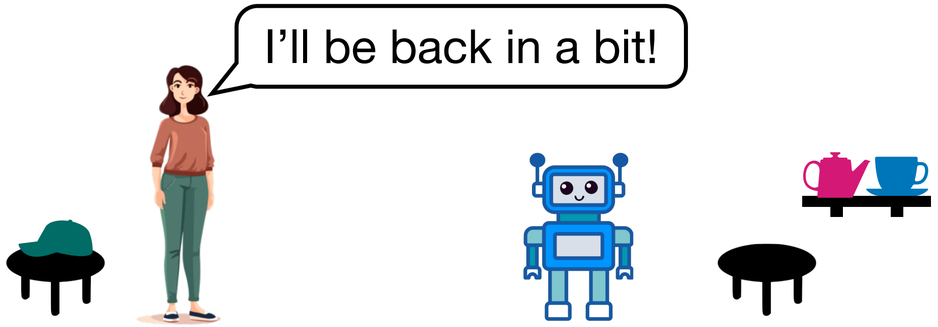
3. $SMCC(\beta^{robot}, \beta^{true})$ — the **robot's** belief state relative to the **true world state**.

   A **high** value means the robot struggles to maintain an accurate semantic map.

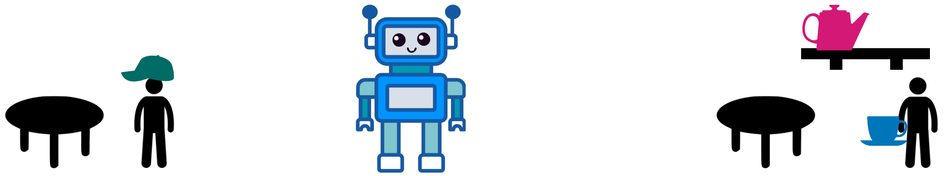
Evaluation

We evaluate on a scenario we call “*Parents are Out!*”

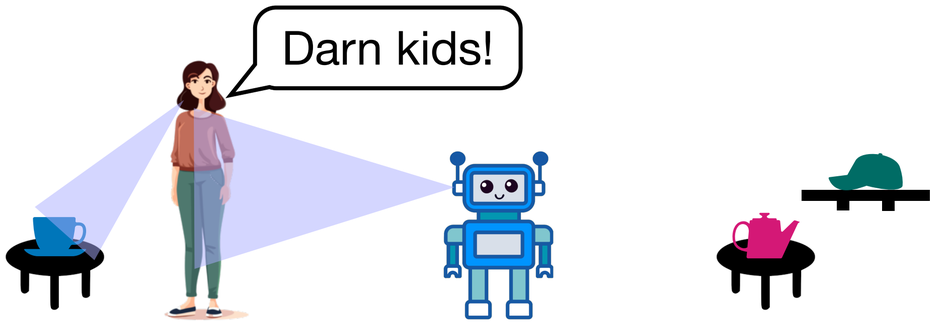
- Jackie starts in a clean house, with a correct belief state.



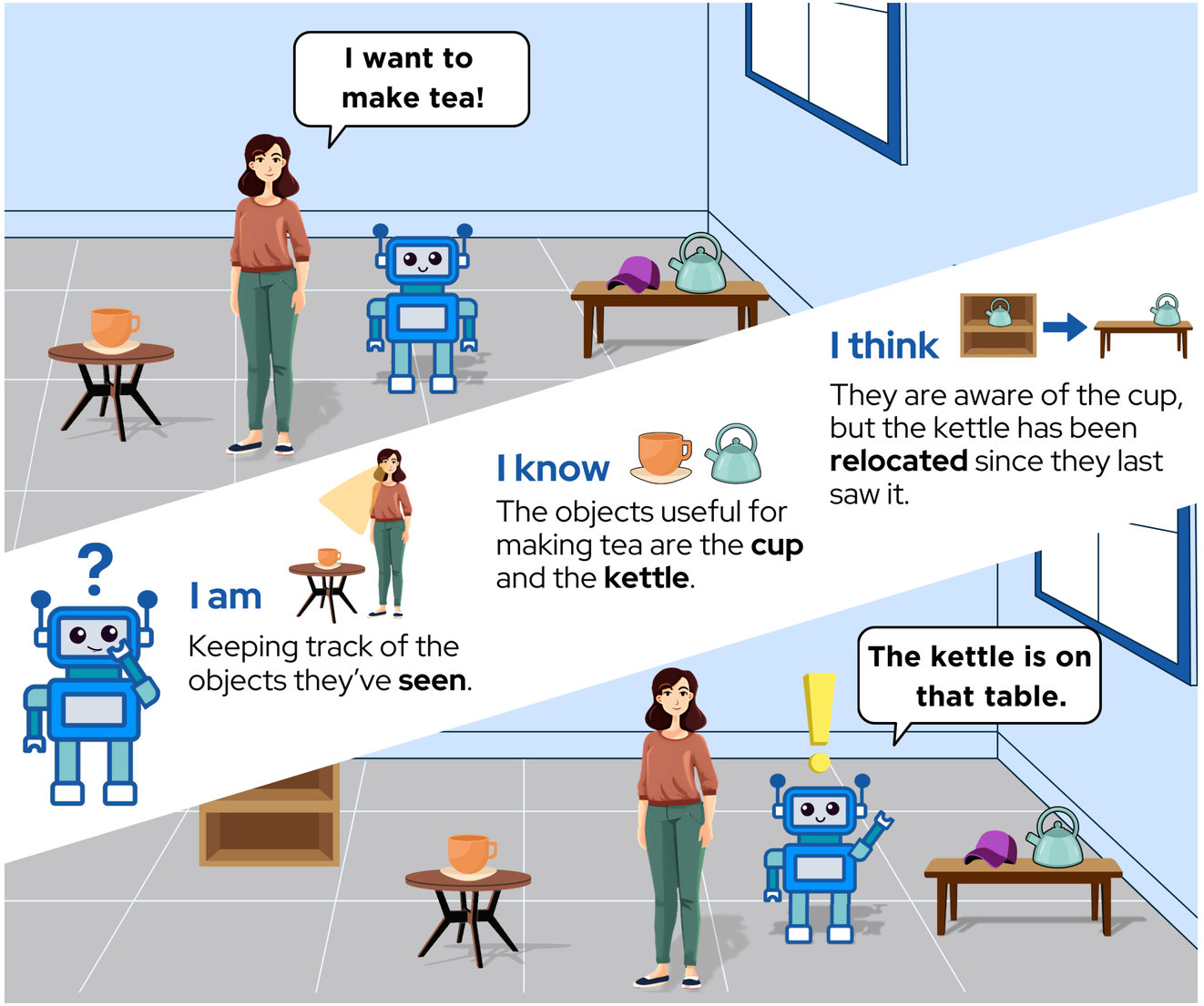
- Jackie leaves the house, and most household items are scattered about (kids threw a party).



- Jackie returns home, sees the house is a mess, and does a quick walk around.
- The robot follows and maintains its prediction of Jackie’s belief state.

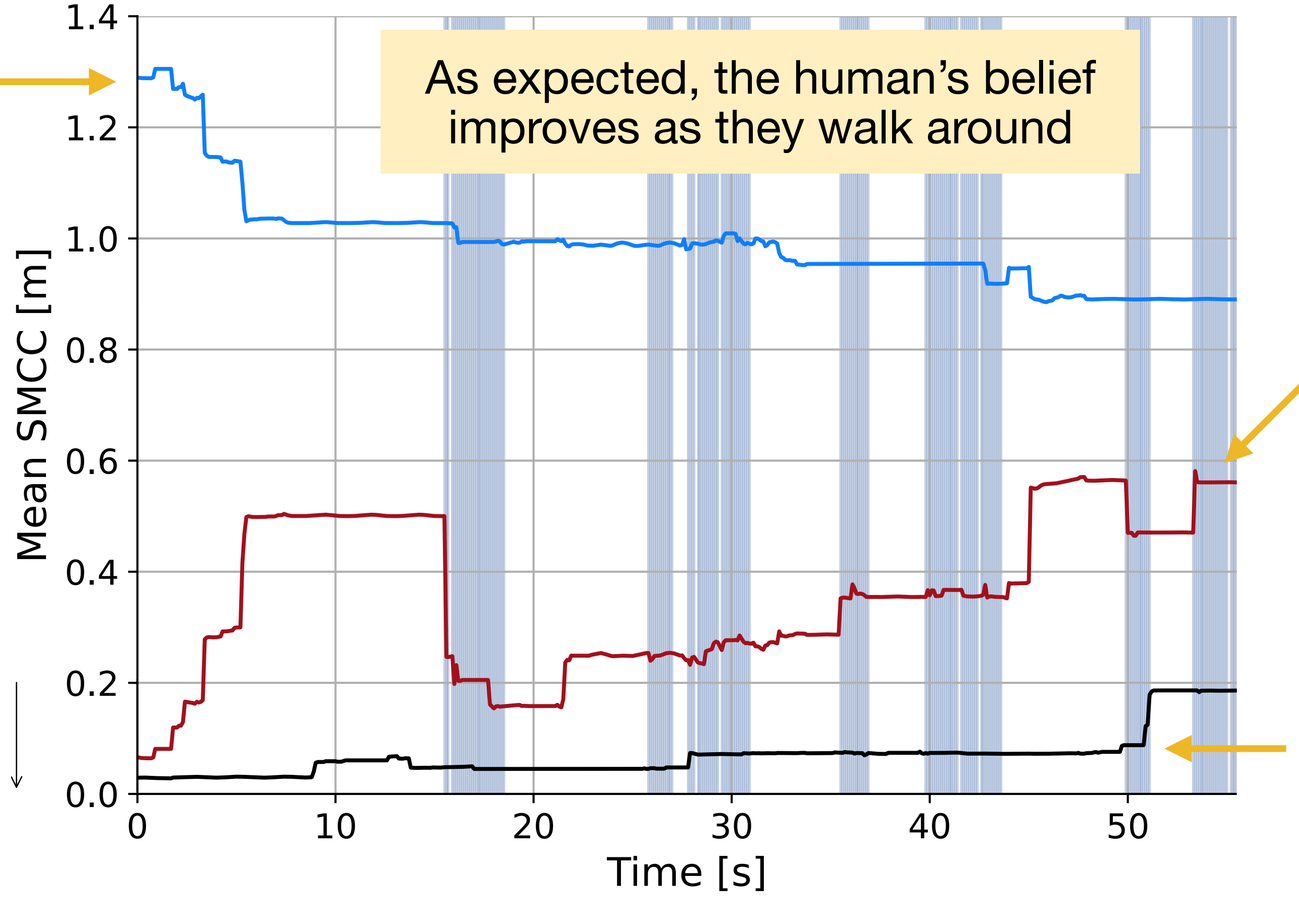


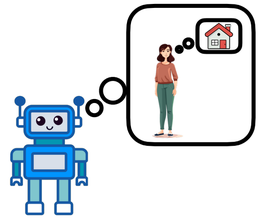
Evaluation


I want to make tea!
I think
They are aware of the cup, but the kettle has been **relocated** since they last saw it.
I know
The objects useful for making tea are the **cup** and the **kettle**.
I am
Keeping track of the objects they've **seen**.
The kettle is on that table.

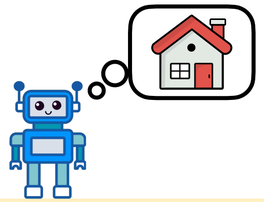
Results

Error of the human's belief state β^{human}





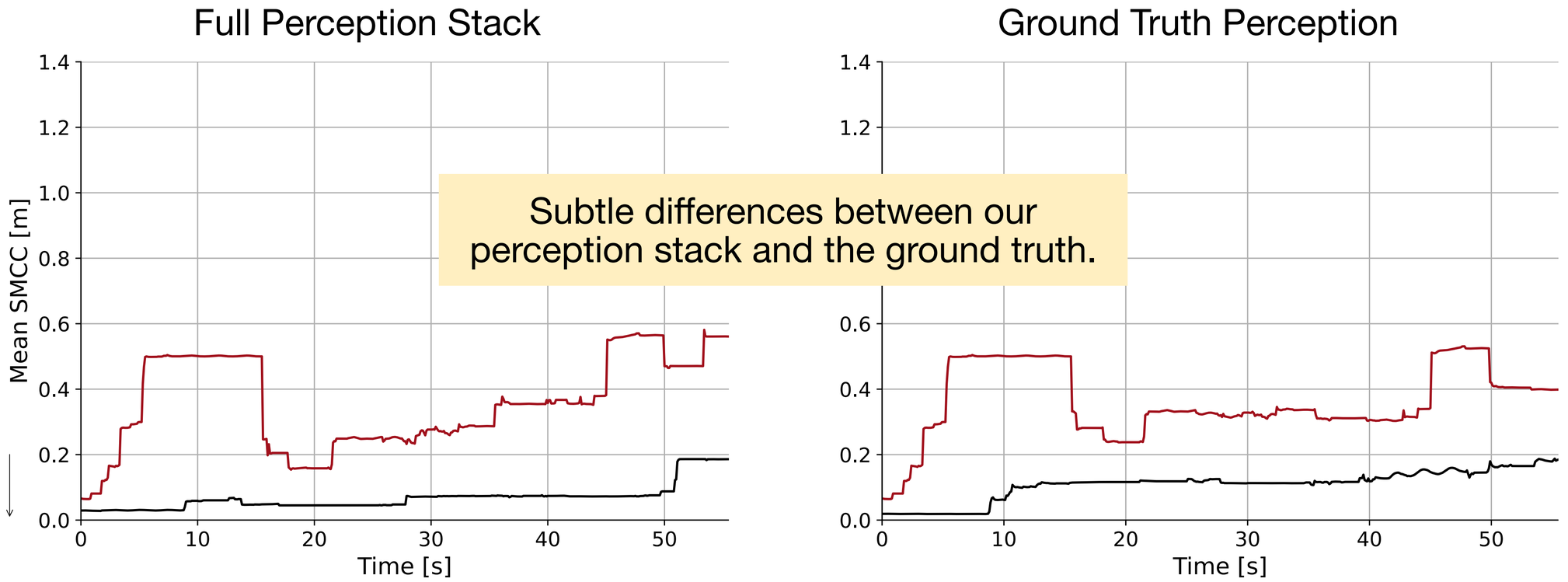
Error of the inferred belief state $\widehat{\beta}^{human}$



Error from sim, perception, and resolver, β^{robot}

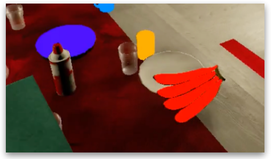
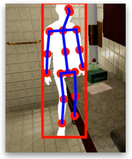
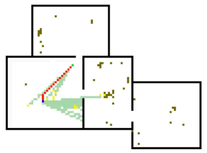
Finding: Error of $\widehat{\beta}^{human}$ < error of β^{human} : we can use $\widehat{\beta}^{human}$ to resolve false beliefs.

Results



Finding: Our perception stacks performs almost as good as the simulator's ground truth!

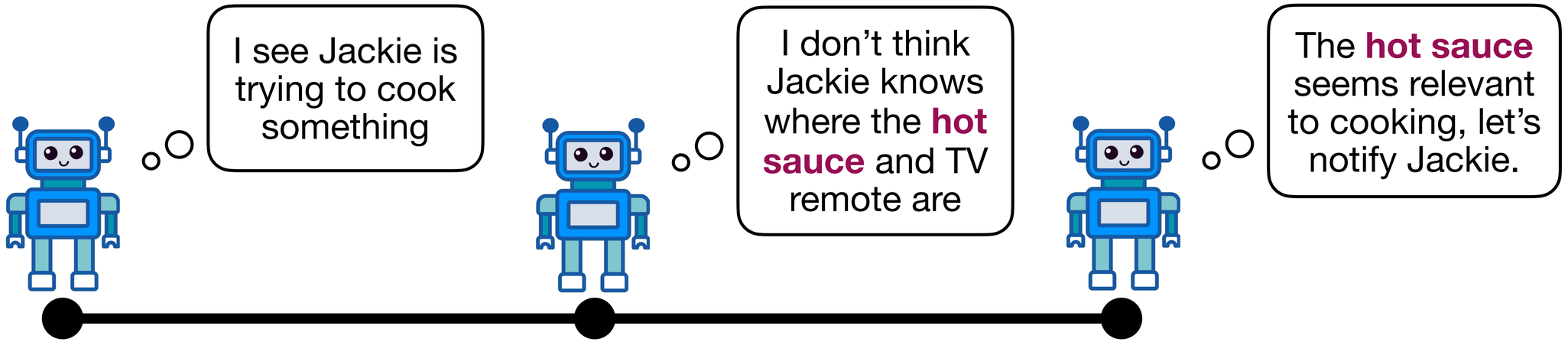
Results

Object Detection	Pose Detection	Trajectory Inference	Mean Inference Error [m]
OWLv2	RTM+MM	A*	0.356
GT	RTM+MM	A*	0.359
OWLv2	**GT**	A*	0.347
OWLv2	RTM+MM	**GT**	0.344
OWLv2	RTM+MM	**None**	0.345
GT	GT	GT	0.360
OWLv2	GT	GT	0.378
GT	**RTM+MM**	GT	0.376
GT	GT	**A***	0.378
GT	GT	**None**	0.378

Finding: Negligible difference between ablation conditions.

Downstream



Activity is given *a priori*

Identify false beliefs

Actively correct false beliefs

Downstream

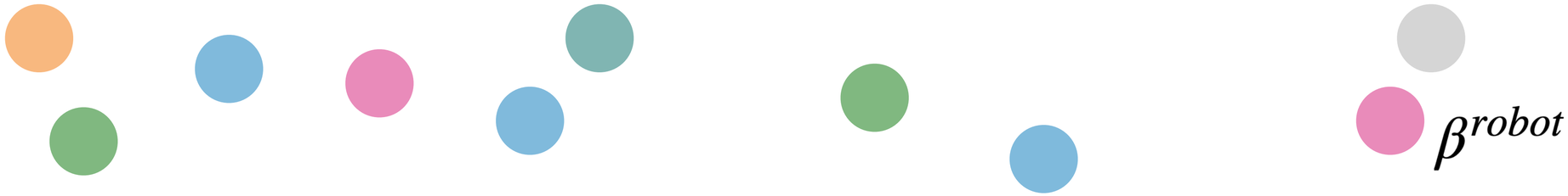
Approach:

- Filter in objects in the predicted human belief state that are far from any objects of that class in the robot's belief state.

$$\mathcal{O}^{unaware} = \{o \in \beta_c^{human} \forall c \,|\, minDist(\beta_c^{human}) < \delta\}$$

where:

o is an object
c is a class
$minDist(A, B)$ is the minimum distance between objects in sets A and B
δ is a distance threshold parameter (we chose $0.3m$)



Downstream

Approach:

- Filter in objects in the predicted human belief state that are far from any objects of that class in the robot's belief state.

$$\mathcal{O}^{unaware} = \{o \in \beta_c^{human} \forall c \,|\, minDist(\beta_c^{human}) < \delta\}$$

where:

o is an object
c is a class
$minDist(A, B)$ is the minimum distance between objects in sets A and B
δ is a distance threshold parameter (we chose $0.3m$)

Downstream

Approach:

- Filter in objects in the predicted human belief state that are far from any objects of that class in the robot's belief state.

$$\mathcal{O}^{unaware} = \{o \in \beta_c^{human} \forall c \,|\, minDist(\beta_c^{human}) < \delta\}$$

where:

o is an object
c is a class
$minDist(A, B)$ is the minimum distance between objects in sets A and B
δ is a distance threshold parameter (we chose $0.3m$)

Downstream

Preliminaries:

- Evaluated four semantic reasoning strategies on quantized Qwen2.5:32b-Instruct

List CoT | Single CoT | Single Binary | Embeddings

```
You are tasked with identifying objects that are relevant to an activity. You will be
given a set of objects and an activity, and you must return the subset that is relevant
for the activity.

For example, if you are given the objects [A, B, C, D, E] and the objects A, C, D are
relevant to the task, return {relevant_items: ['A', 'B', 'C']}.

Your turn! You are aware of the following items:
 • {belief state}

Which items are useful for {activity}?

Return your output as a JSON object with the given structure.
```

Downstream

Preliminaries:

- Evaluated four semantic reasoning strategies on quantized Qwen2.5:32b-Instruct

List CoT | **Single CoT** | Single Binary | Embeddings

```
You are tasked with judging whether an object is relevant to an activity. You will be
given an object and an activity, and you must return whether the object is relevant for
the activity.

For example, if you are given an object and the activity is {activity}, return `true`
if the object is relevant and `false` if it is not.

Your turn! Is a {object} useful for {activity}?

Return your output as `true` or `false`.
```

Downstream

Preliminaries:

- Evaluated four semantic reasoning strategies on quantized Qwen2.5:32b-Instruct

List CoT | Single CoT | **Single Binary** | Embeddings

```
Is a {object} useful for {activity}? Answer `true` or `false` with no other
explanation.
```

Downstream

Preliminaries:

- Evaluated four semantic reasoning strategies on quantized Qwen2.5:32b-Instruct

List CoT | Single CoT | Single Binary | **Embeddings**

Used a pre-trained **deBERTa-v3** model for zero-shot classification.

Source: `I am {activity} using a {object}.`

Classes:

- `A sentence that makes sense.`
- `A sentence that does not make sense.`

Downstream

Preliminaries:

Notifications are correct

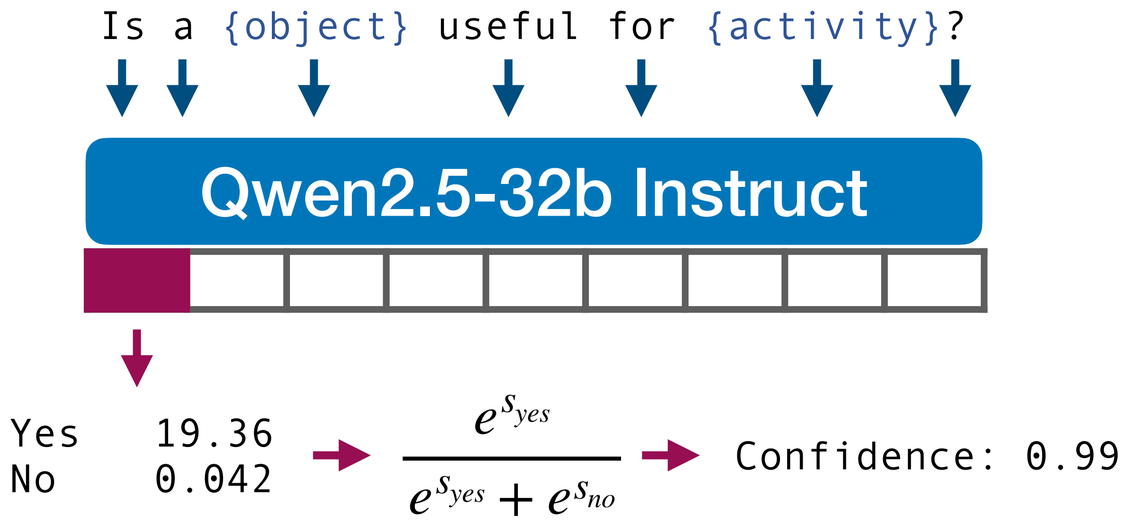
Relevant objects are recalled

Method	F1	Precision	Recall
LLM (List CoT)	**0.68**	0.56	**0.86**
LLM (Single CoT)	0.53	**0.77**	0.41
LLM (Single Binary)	0.46	0.76	0.33
deBERTa-v3	0.56	0.46	0.72
Random	0.14	0.08	0.50

- From a user experience perspective, ***Precision*** is most important:
  - ↑ Precision = fewer false positives, we anticipate that users prefer silence to incorrect advice.
- All methods greatly outperformed random selection, but no clear winner.

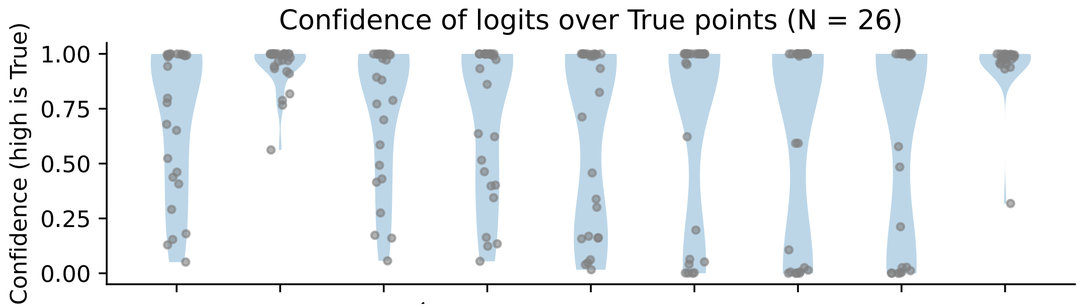
Downstream

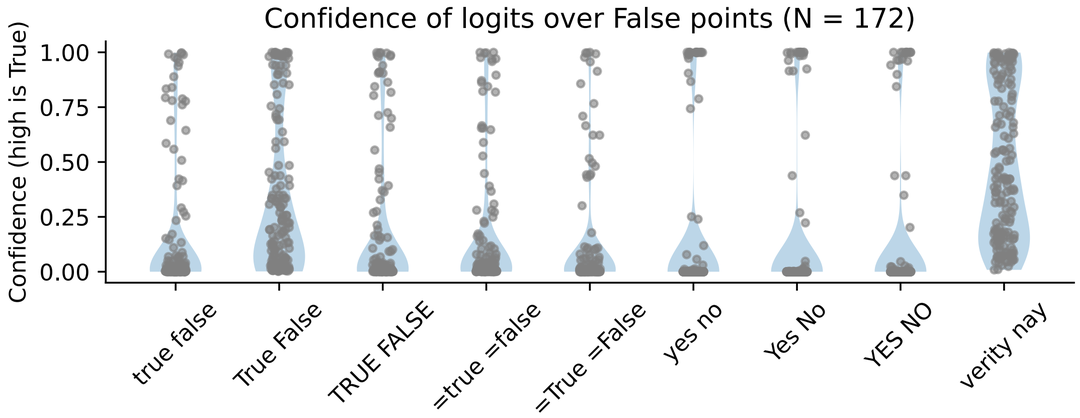
Single Binary Prompt

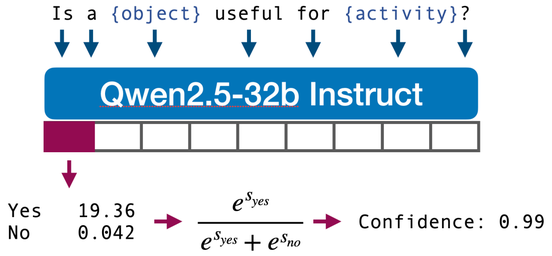


Downstream

Single Binary Prompt







Preliminary:

Method	F1	Precision	Recall
LLM (List CoT)	**0.68**	0.56	**0.86**
LLM (Single CoT)	0.53	**0.77**	0.41
LLM (Single Binary)	0.46	0.76	0.33
deBERTa-v3	0.56	0.46	0.72
Random	0.14	0.08	0.50

true false

Actual \ Predicted	T	F
T	18	8
F	21	151

F1: 0.55
Prec: 0.46
Recall: 0.69

True False

Actual \ Predicted	T	F
T	26	0
F	41	131

F1: 0.56
Prec: 0.39
Recall: 1.0

TRUE FALSE

Actual \ Predicted	T	F
T	19	7
F	22	150

F1: 0.57
Prec: 0.46
Recall: 0.73

=true =false

Actual \ Predicted	T	F
T	18	8
F	20	152

F1: 0.56
Prec: 0.47
Recall: 0.69

=True =False

Actual \ Predicted	T	F
T	15	11
F	14	158

F1: 0.55
Prec: 0.52
Recall: 0.58

yes no

Actual \ Predicted	T	F
T	18	8
F	21	151

F1: 0.55
Prec: 0.46
Recall: 0.69

Yes No

Actual \ Predicted	T	F
T	17	9
F	18	154

F1: 0.56
Prec: 0.49
Recall: 0.65

YES NO

Actual \ Predicted	T	F
T	16	10
F	17	155

F1: 0.54
Prec: 0.48
Recall: 0.62

verity nay

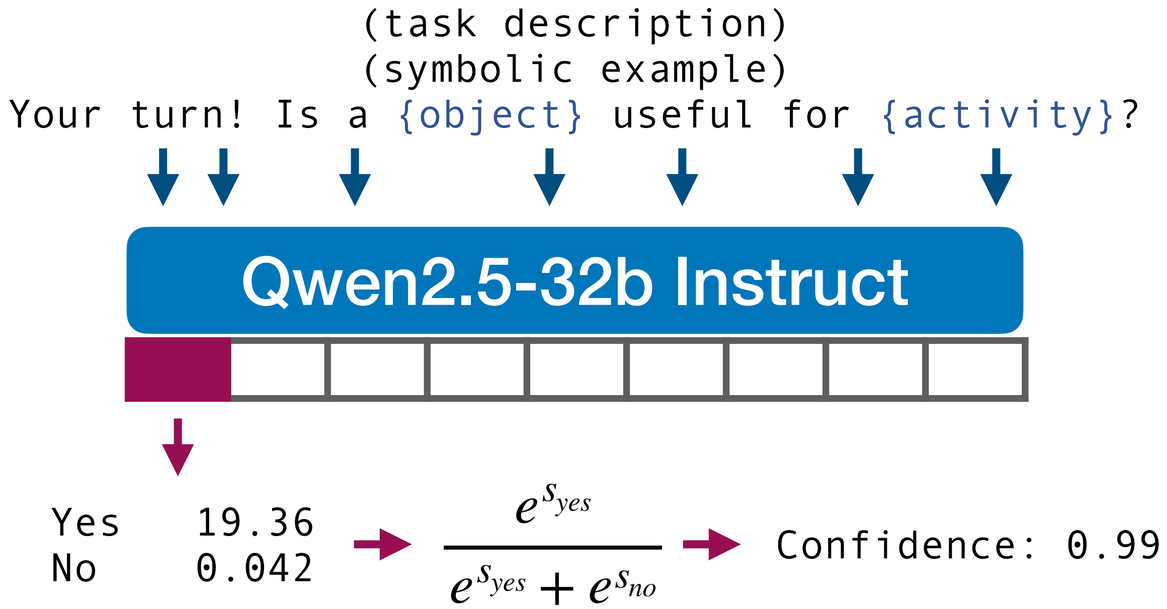
Actual \ Predicted	T	F
T	25	1
F	72	100

F1: 0.41
Prec: 0.26
Recall: 0.96

Finding: Higher F1 scores, but lower precision at this confidence threshold (>0.5), interesting tradeoff between precision and recall for most token pairs.

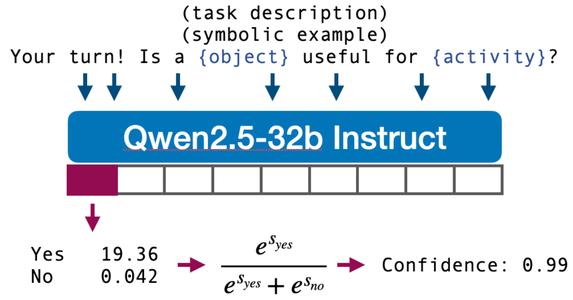
Downstream

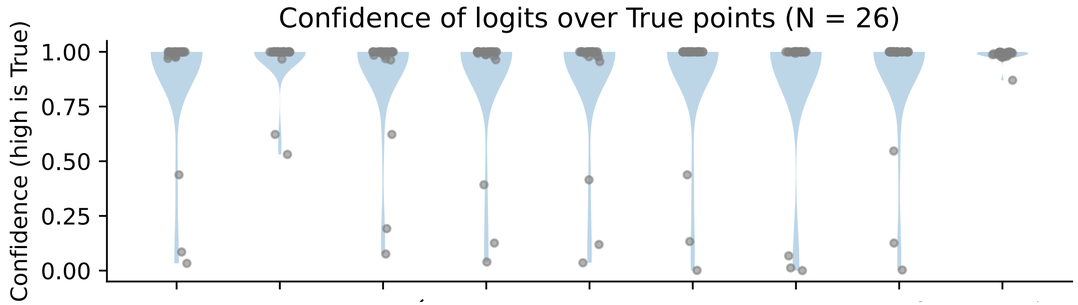
Single CoT Prompt

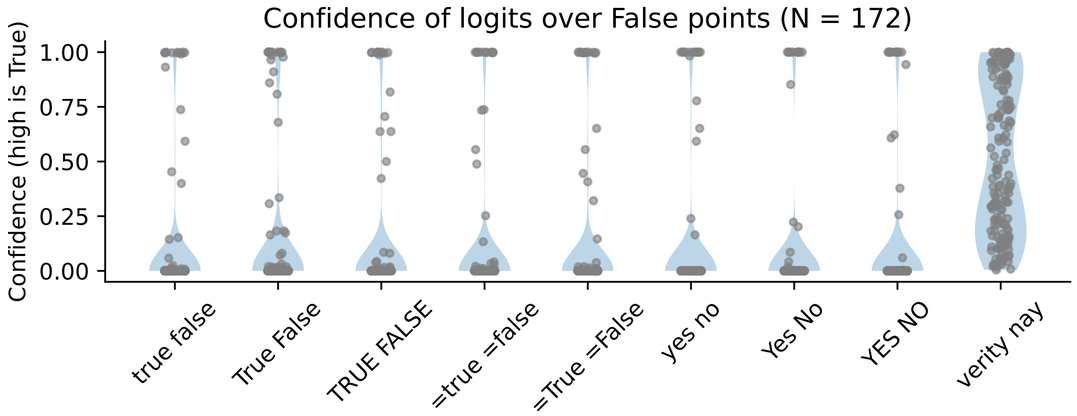


Downstream

Single CoT Prompt







Preliminary:

Method	F1	Precision	Recall
LLM (List CoT)	**0.68**	0.56	**0.86**
LLM (Single CoT)	0.53	**0.77**	0.41
LLM (Single Binary)	0.46	0.76	0.33
deBERTa-v3	0.56	0.46	0.72
Random	0.14	0.08	0.50

true false

Actual \ Predicted	T	F
T	23	3
F	13	159

F1: 0.74
Prec: 0.64
Recall: 0.88

True False

Actual \ Predicted	T	F
T	26	0
F	18	154

F1: 0.74
Prec: 0.59
Recall: 1.0

TRUE FALSE

Actual \ Predicted	T	F
T	24	2
F	15	157

F1: 0.74
Prec: 0.62
Recall: 0.92

=true =false

Actual \ Predicted	T	F
T	23	3
F	14	158

F1: 0.73
Prec: 0.62
Recall: 0.88

=True =False

Actual \ Predicted	T	F
T	23	3
F	13	159

F1: 0.74
Prec: 0.64
Recall: 0.88

yes no

Actual \ Predicted	T	F
T	23	3
F	15	157

F1: 0.72
Prec: 0.61
Recall: 0.88

Yes No

Actual \ Predicted	T	F
T	23	3
F	12	160

F1: 0.75
Prec: 0.66
Recall: 0.88

YES NO

Actual \ Predicted	T	F
T	24	2
F	14	158

F1: 0.75
Prec: 0.63
Recall: 0.92

verity nay

Actual \ Predicted	T	F
T	26	0
F	80	92

F1: 0.39
Prec: 0.25
Recall: 1.0

Finding: CoT improves F1 scores over preliminaries and single binary, F1 of 0.75.

Takeaways

Can we infer a world belief state via camera observations in a household domain?

Yes, robots can infer user world belief states in household domains.

Modern semantic reasoning methods are fairly good at reasoning over the world belief state for downstream assistance.

Using an inferred belief state enables useful teaming capabilities.

Kolb, Jack, and Karen M. Feigh. "Inferring Belief States in Partially-Observable Human-Robot Teams." *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024.

Kolb, Jack, and Aditya Garg and Nikolai Warner and Karen M. Feigh. "Inferring World Belief States in Dynamic Real-World Environments." *Under review*. 2025.

Future Directions

Simulation

Where are the semantic relocation simulators?

This work was limited to ~60s rollouts.

Semantic Scene Graphs

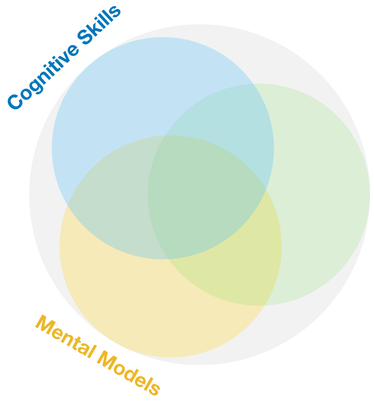
Object permanence in floorplans is a barren field.

This work used a “simple” approach to resolving objects.

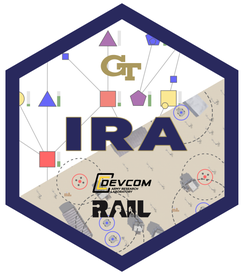
Modeling User Perception

Models have not made it to the HRI community.

This work assumed the user agent had perfect perception.



How can robots **model** and **apply** a user's **cognitive state** to inform human-robot teaming capabilities?

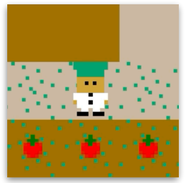


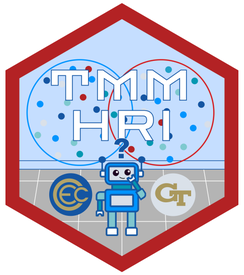
Can we predict future teleoperation performance only using cognitive skills, and apply it to role assignment?

Published in RO-MAN '21, RO-MAN '22

Can we infer user situation awareness via observing users in a **partially-observable** environment?

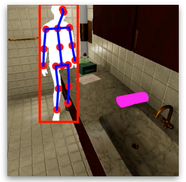
Published in IROS '24





Can we infer user situation awareness via camera observations in a household domain?

Submitted to RA-L



How can robots **model** and **apply** a user's **cognitive state** to inform human-robot teaming capabilities?

How can robots **model** and **apply** a user's **cognitive state** to inform human-robot teaming capabilities?

Measure user **cognitive skills** → Predict teleoperation performance and conduct team role assignment.

Estimate user situation awareness → design systems to **structure decision-making** or **share authority**.

Infer user **world belief states** → Infer user situation awareness, actively assist users, and inform planning.

Monitor user **cognitive workload** → Adapt communication, reallocate taskwork to moderate workload.

Identify user **strategic styles** → Adapt agents to compliment a diverse range of strategies and change at will.


IRA
DEVCOM
ARMY RESEARCH LABORATORY
RAIL
PCI
HRI
STAR LAB
SAFE
RL
STAR LAB
MARS
SMM
ISR
MAISR
?
POSE
CTRL
TMM
HAI
TMM
HRI
TMM
MAS
LOCKHEED MARTIN
















amazon Lab126

Sandia National Laboratories

DEVCOM ARMY RESEARCH LABORATORY

Office of Naval Research ONR Science & Technology


This must be
BAIR
CORL
fetch
hello robot
QATAR