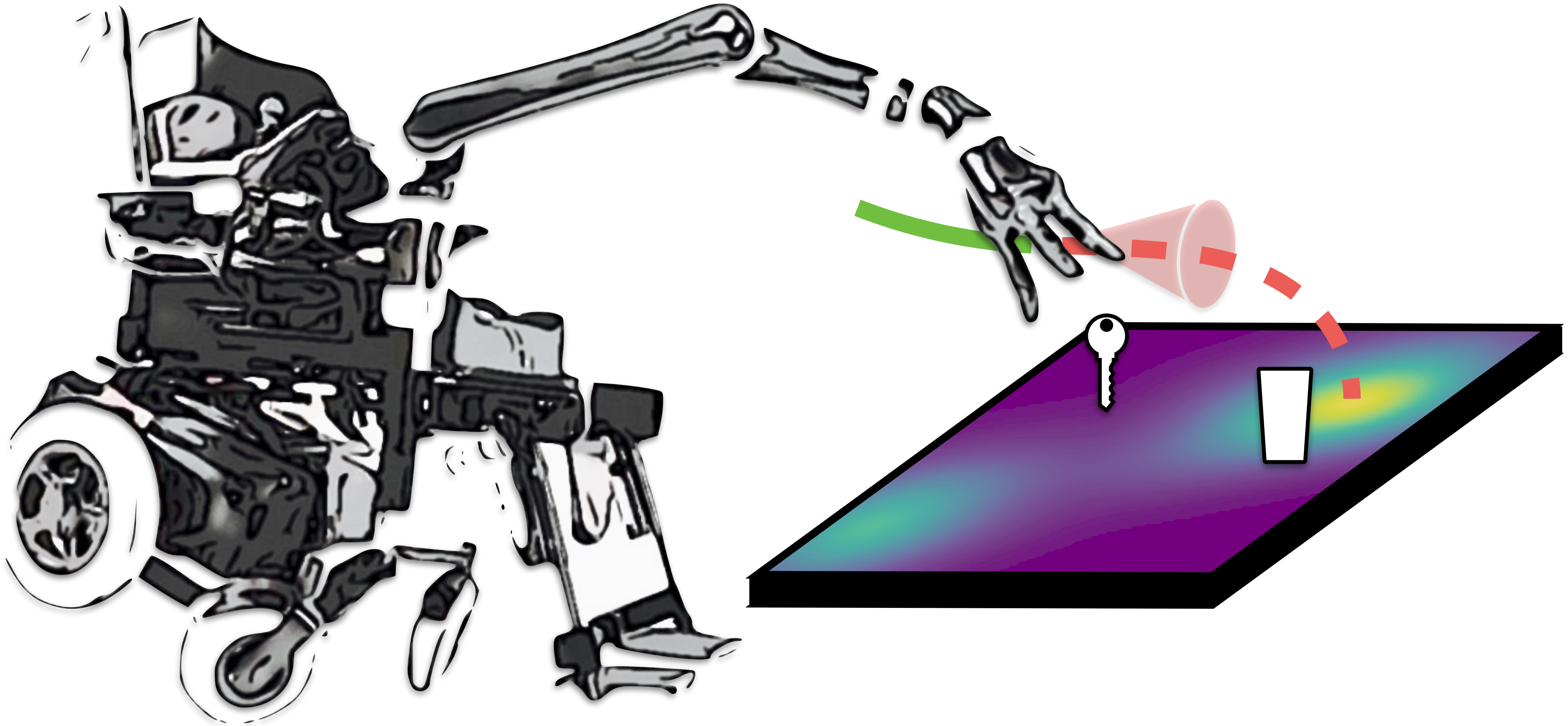
Assistive Manipulation: Reducing the Control Effort with Multi-Modal Intent Prediction



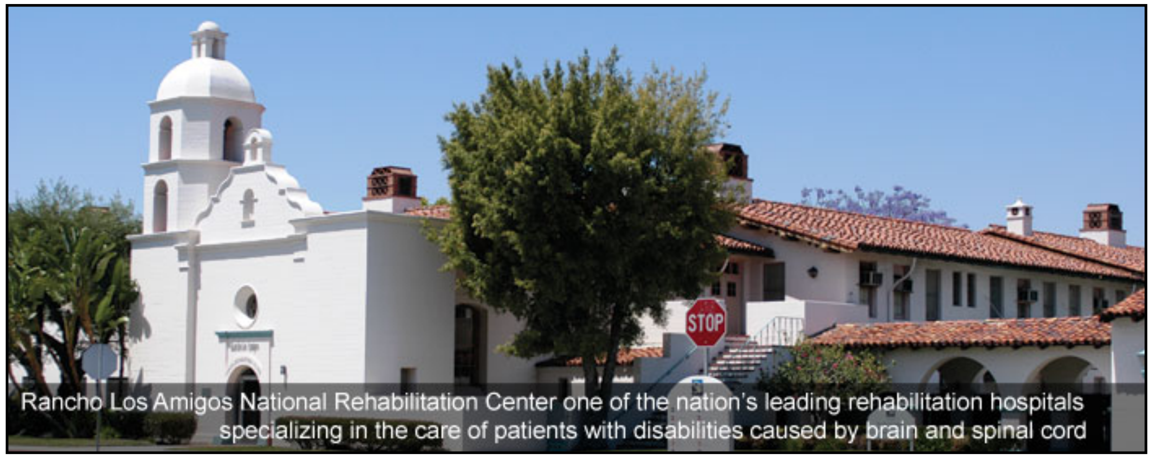
Renaud Detry – KU Leuven

https://www.nature.com/articles/nature11076

Lesson learned from a discussion with SCI Group: Lower the burden of robot operation

Challenges: charging a chair's battery, eating, picking up keys, picking up groceries, opening doors.

Major limitation of current solutions: even the simplest tasks are monopolizing the user's attention.



Our project aims to facilitate the guidance of a robot with a noisy, low-bandwidth input device

Neurobotics: blended brain-machine control for human assistance using hybrid smart systems

Pinhao Song, Ophelie Saussus, Santiago Rondón, Sofie De Schrijver, Erwin Aertbelien, Tom Theys, Renaud Detry, Peter Janssen

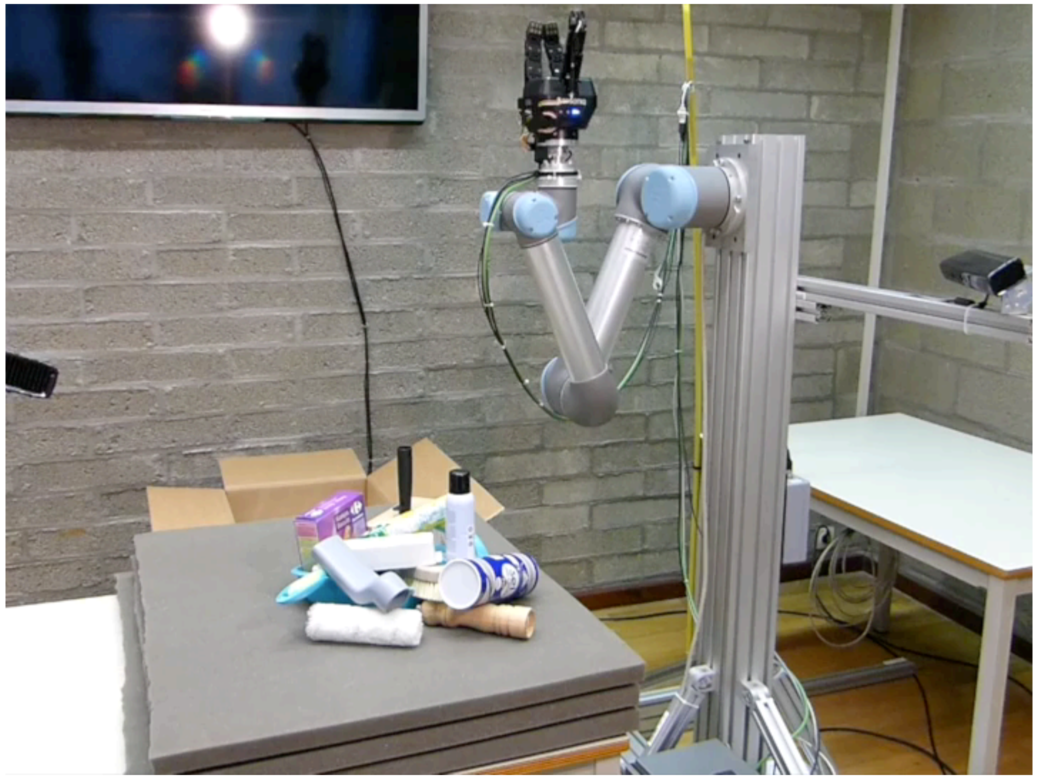
Funded by KU Leuven

KU LEUVEN

Objective: lower the burden of robot operation (with today's very capable data-driven robot models)

1: Parse user intent

2: Move robot in a direction aligned with intent



We focus on predicting user intent from the user-guided robot motion

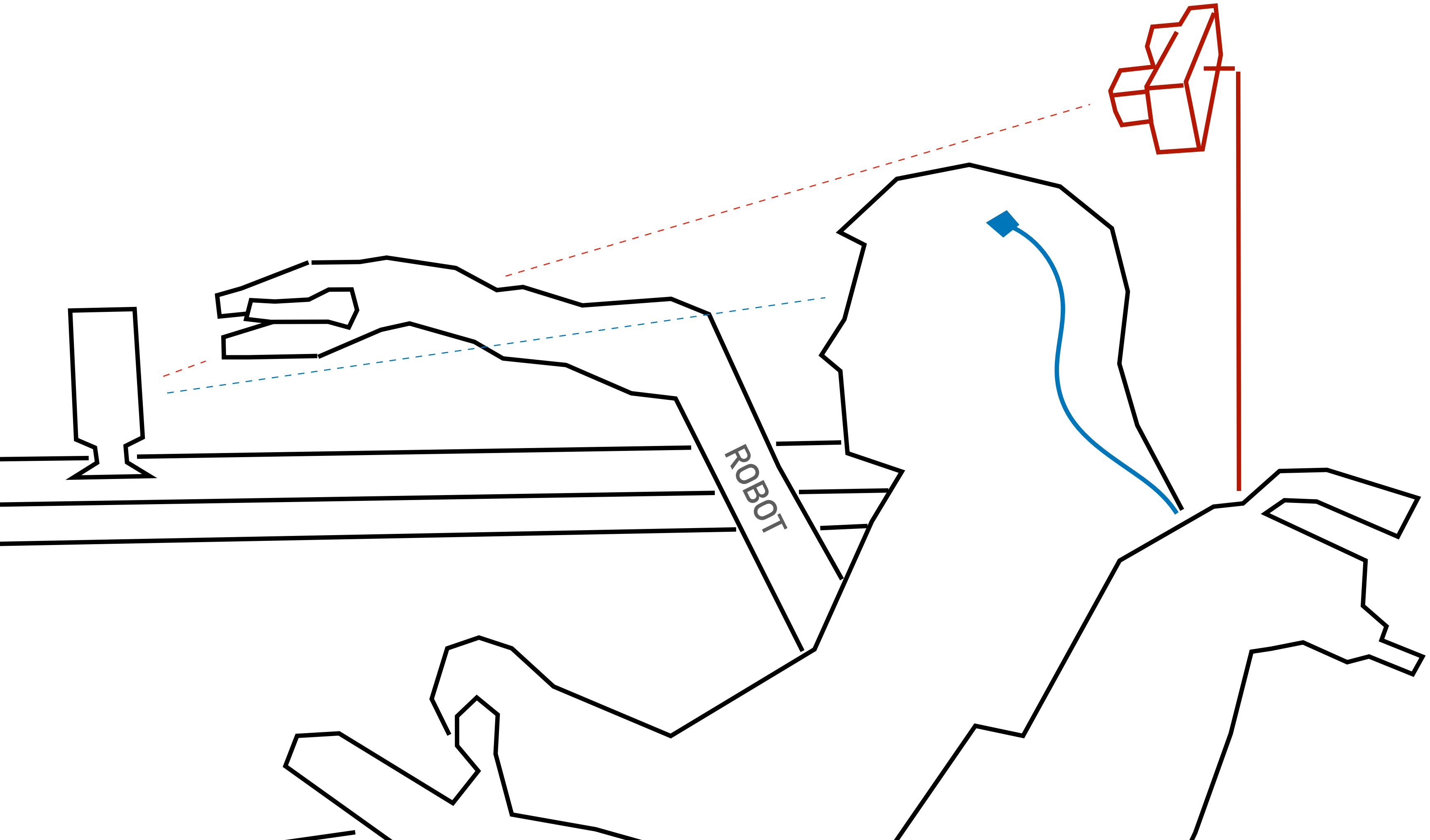
1: Parse user intent

2: Move robot in a direction aligned with intent

To understand the user's intent, we must reconcile user input with a contextual (visual) scene understanding

1. The subject initiates a reaching motion towards the desired object,

2. The robot infers the subject's intention - i.e., which object the subject intends to grasp and from where to approach.



1: "What are the possible actions given visible objects?"

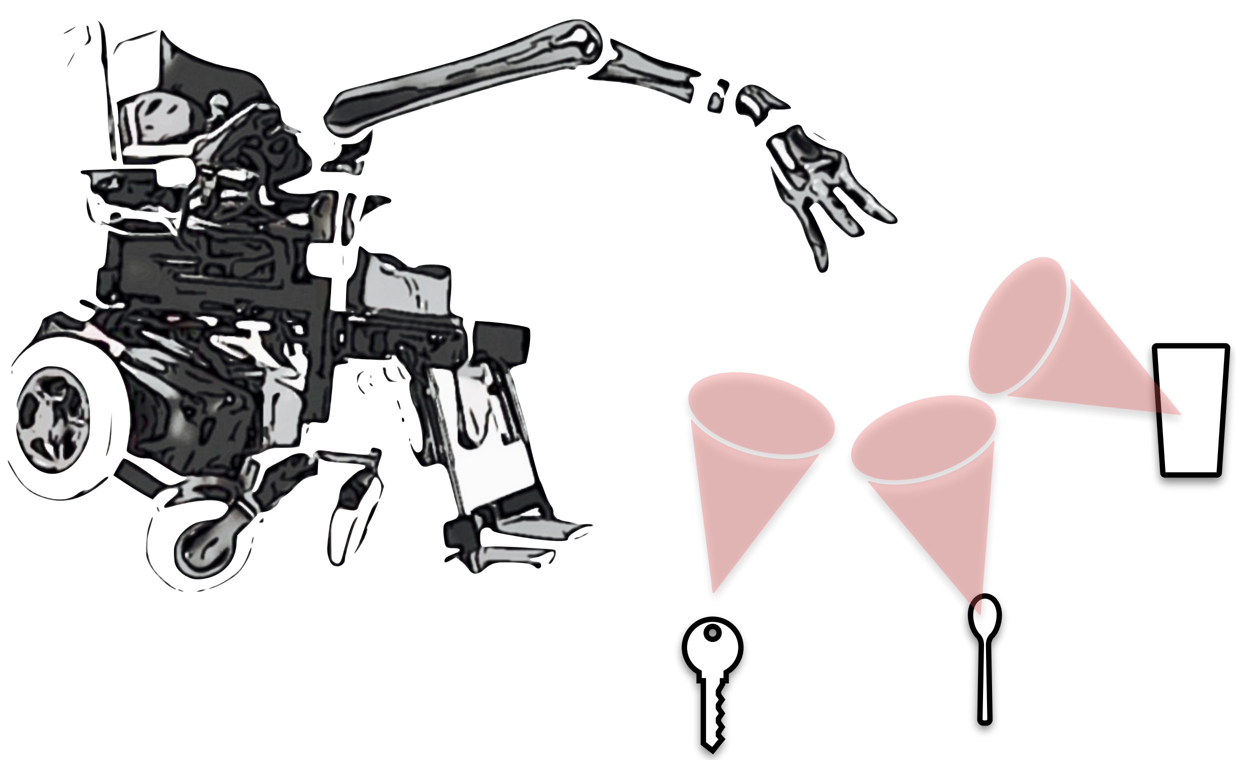
1. The subject initiates a reaching motion towards the desired object,

2. The robot infers the subject's intention - i.e., which object the subject intends to grasp and from where to approach.

1: "What are the possible actions given visible objects?"
→ Robot inventories all possible actions offered by what it sees

2: "Where is the user headed?"
→ Robot infers plausible targets given onset of motion (decoded from brain signals)

3: "What does the subject want?"
Merging 1 and 2 yields an unambiguous target



2: “What action is the subject attempting?”

1. The subject initiates a reaching motion towards the desired object,

2. The robot infers the subject’s intention - i.e., which object the subject intends to grasp and from where to approach.

1: “What are the possible actions given visible objects?”

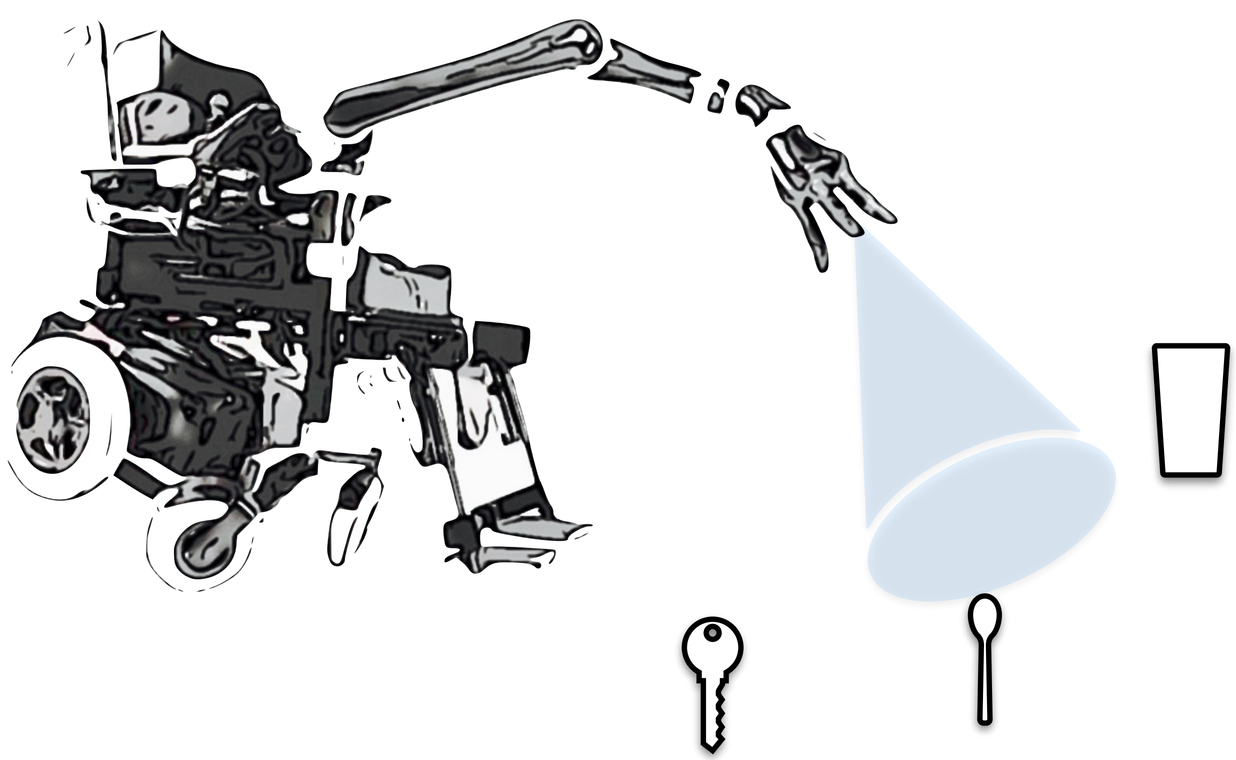
→ Robot inventories all possible actions offered by what it sees

2: “Where is the user headed?”

→ Robot infers plausible targets given onset of motion (decoded from brain signals)

3: *“What does the subject want?”*

Merging 1 and 2 yields an unambiguous target



3: Reconcile 1 and 2 to infer intent and derive a control command

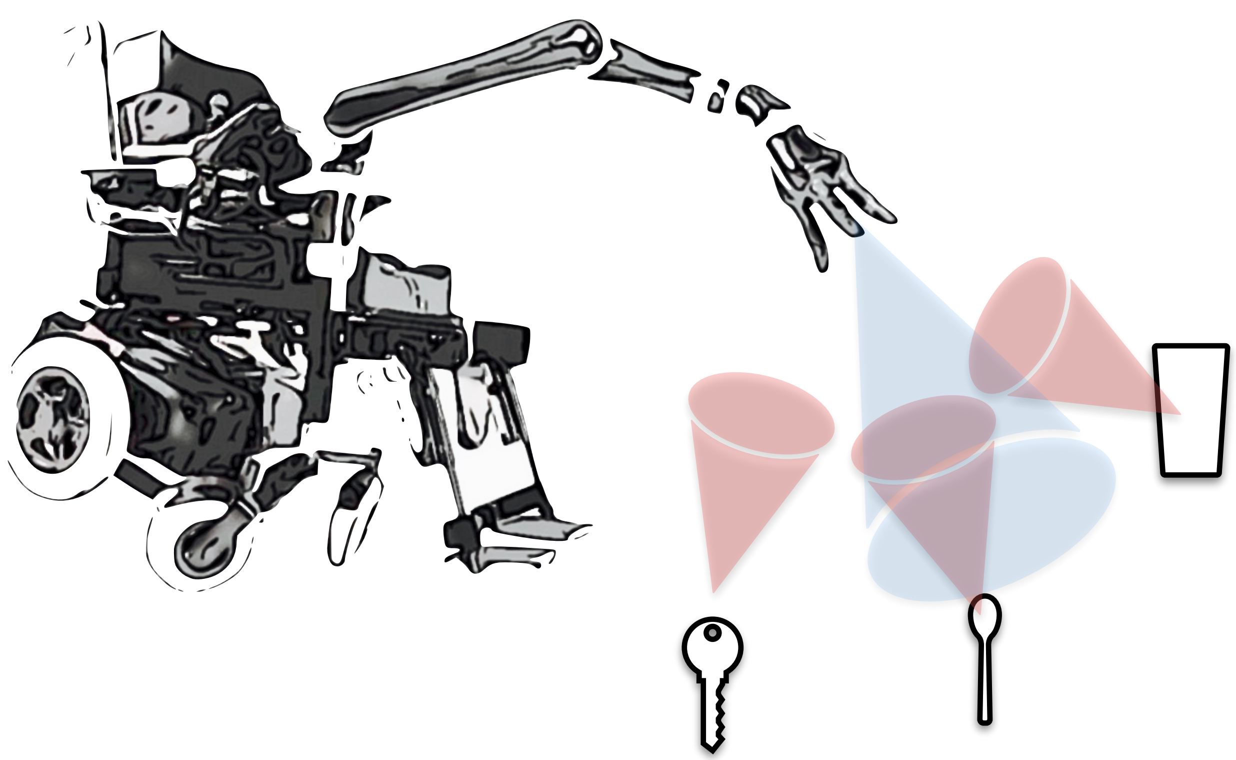
1. The subject initiates a reaching motion towards the desired object,
2. The robot infers the subject's intention - i.e., which object the subject intends to grasp and from where to approach.

1: "What are the possible actions given visible objects?"

→ Robot inventories all possible actions offered by what it sees

2: "Where is the user headed?"

→ Robot infers plausible targets given onset of motion (decoded from brain signals)



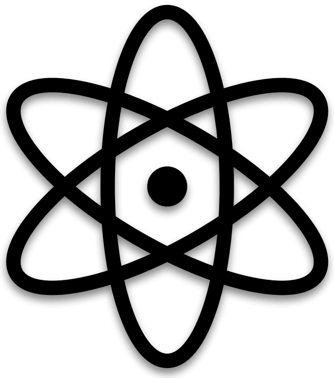
3: "*What does the subject want?*"

Merging 1 and 2 yields an unambiguous target

Outline

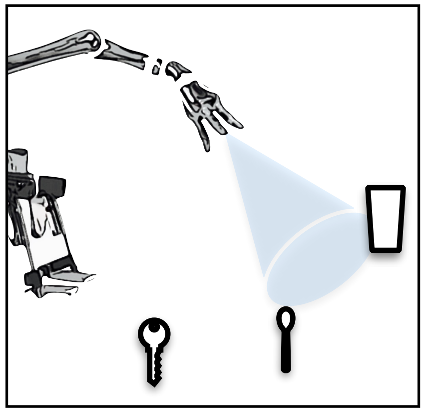
State of Art

Innovation: dynamics and multimodality



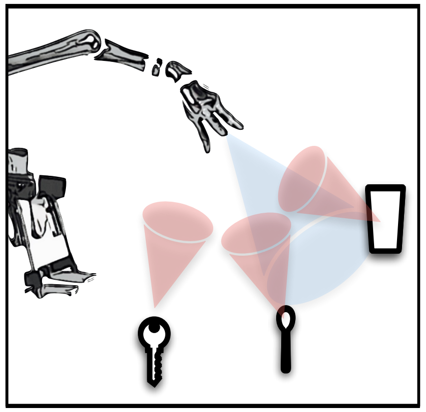
Motion Prediction

From the user's motion onset



Intention Prediction

Motion prediction + goal assessment



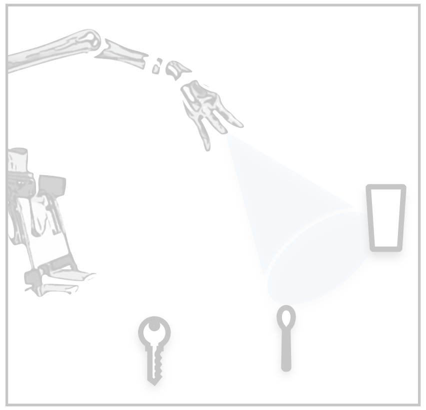
Outline

State of Art

Innovation: dynamics and multimodality

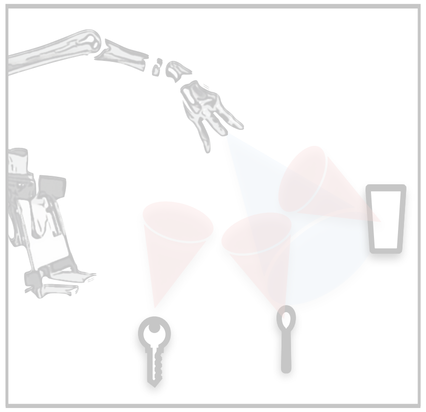
Motion Prediction

From the user's motion onset

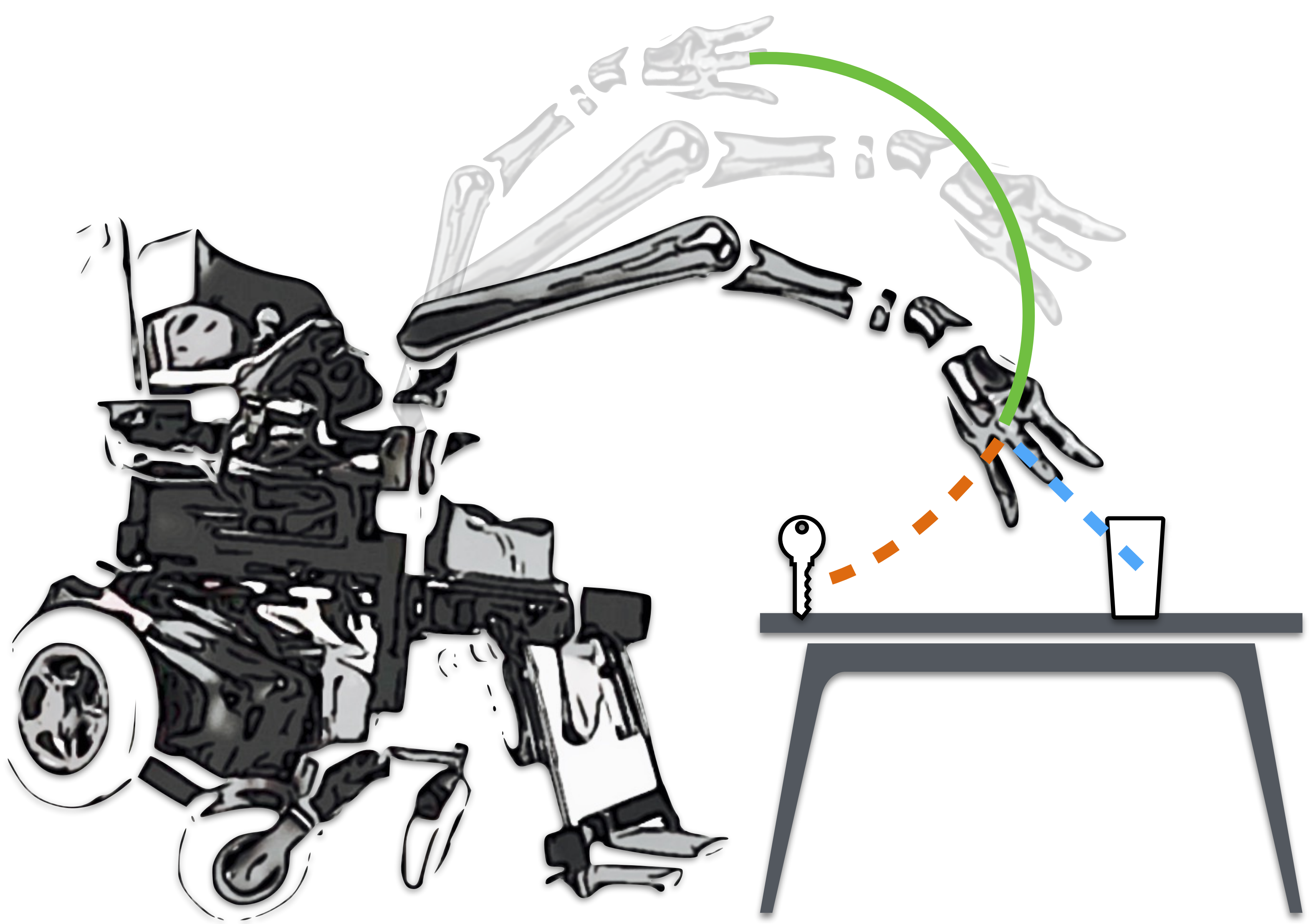


Intention Prediction

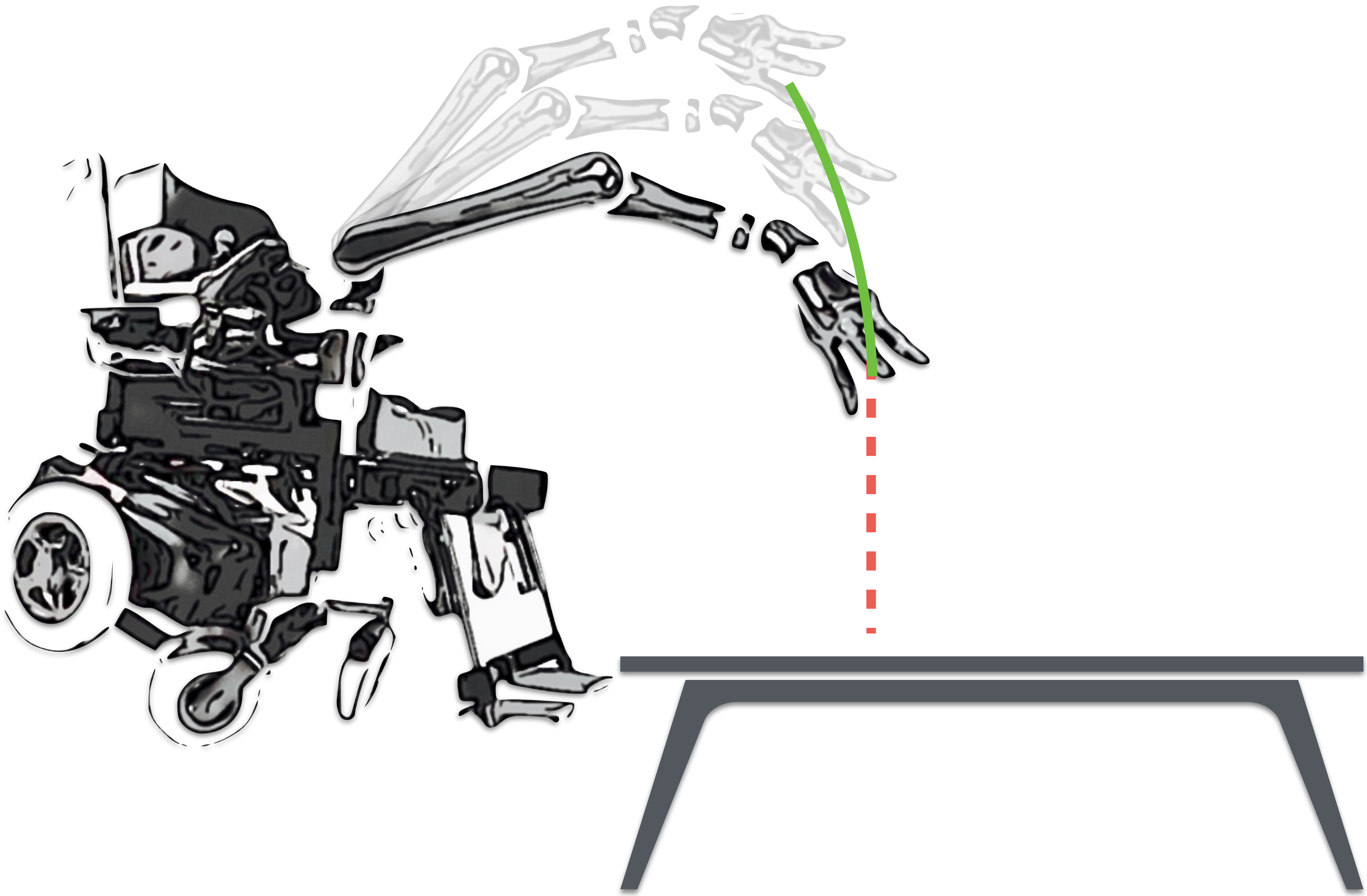
Motion prediction + goal assessment



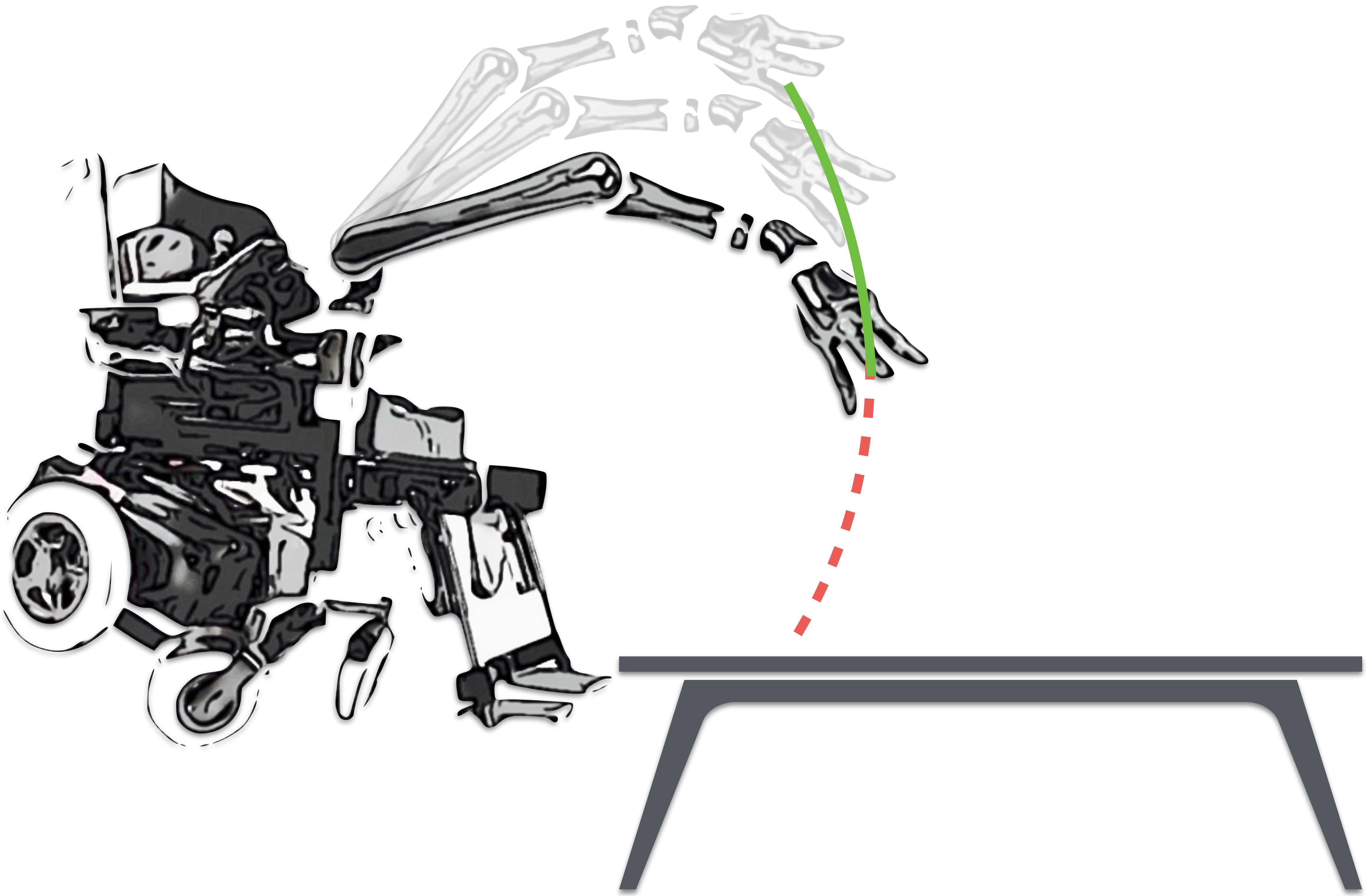
We improve the state of art by leveraging motion dynamics



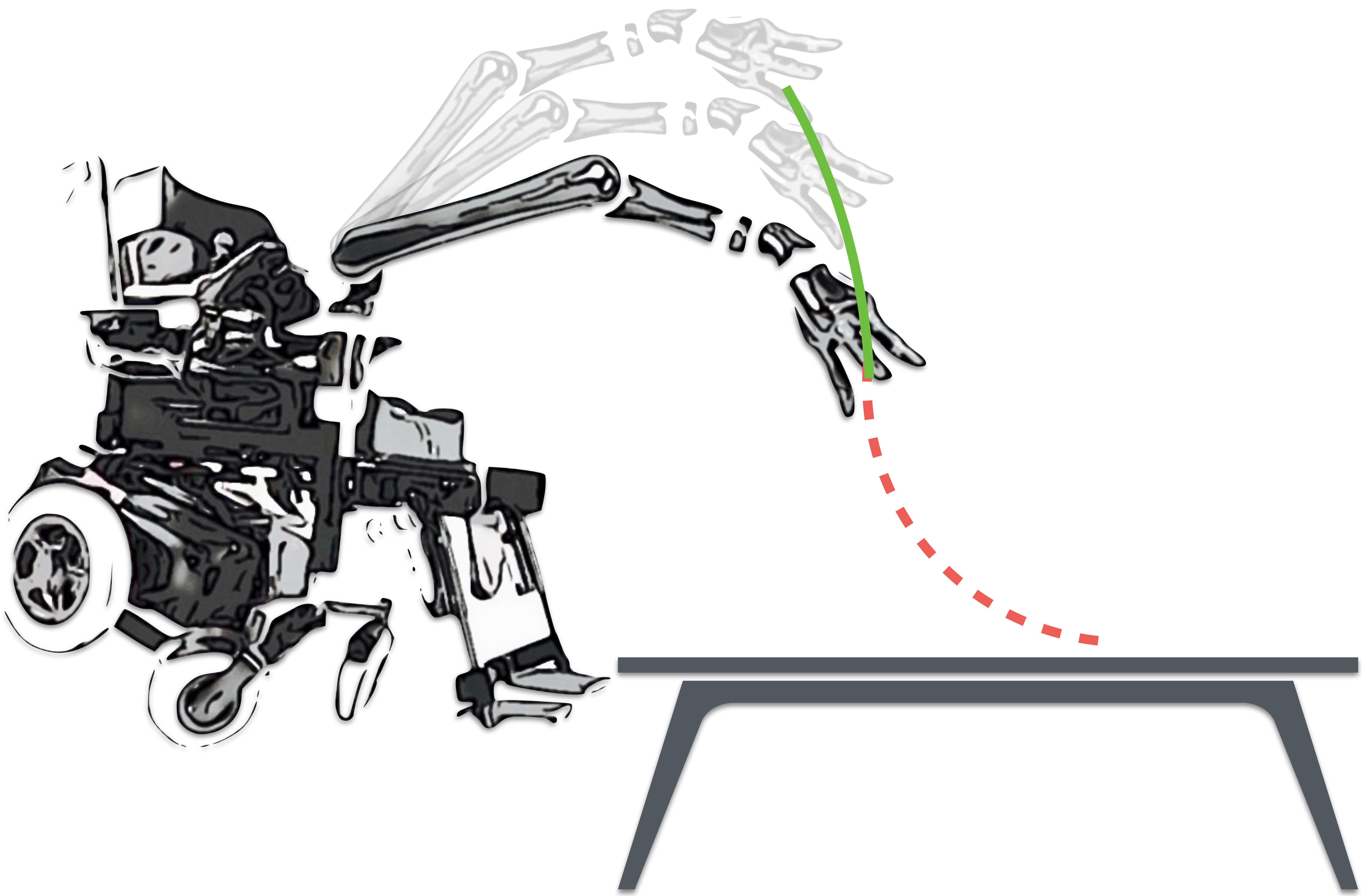
We improve the state of art by modeling a trajectory distribution



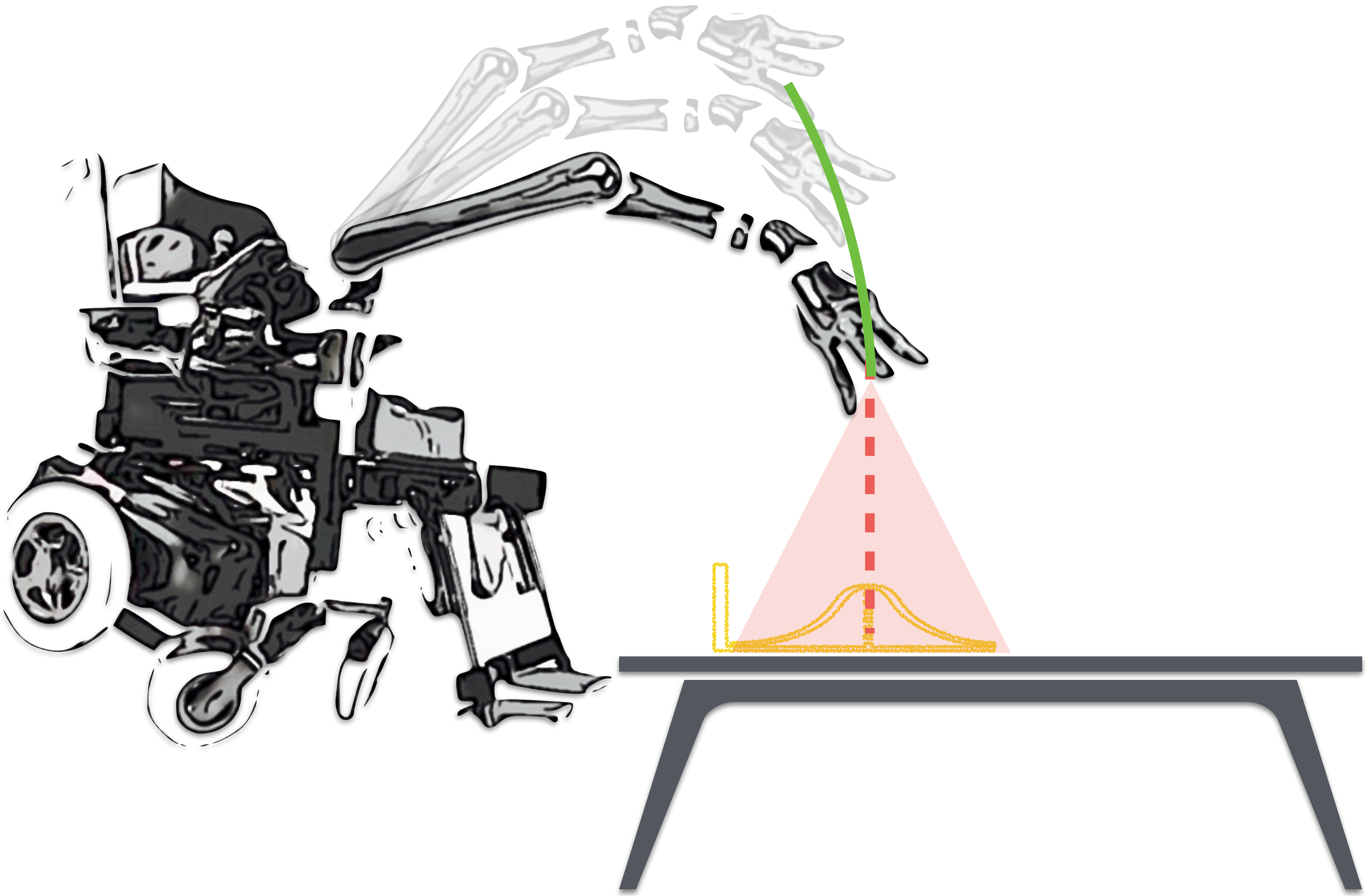
We improve the state of art by modeling a trajectory distribution



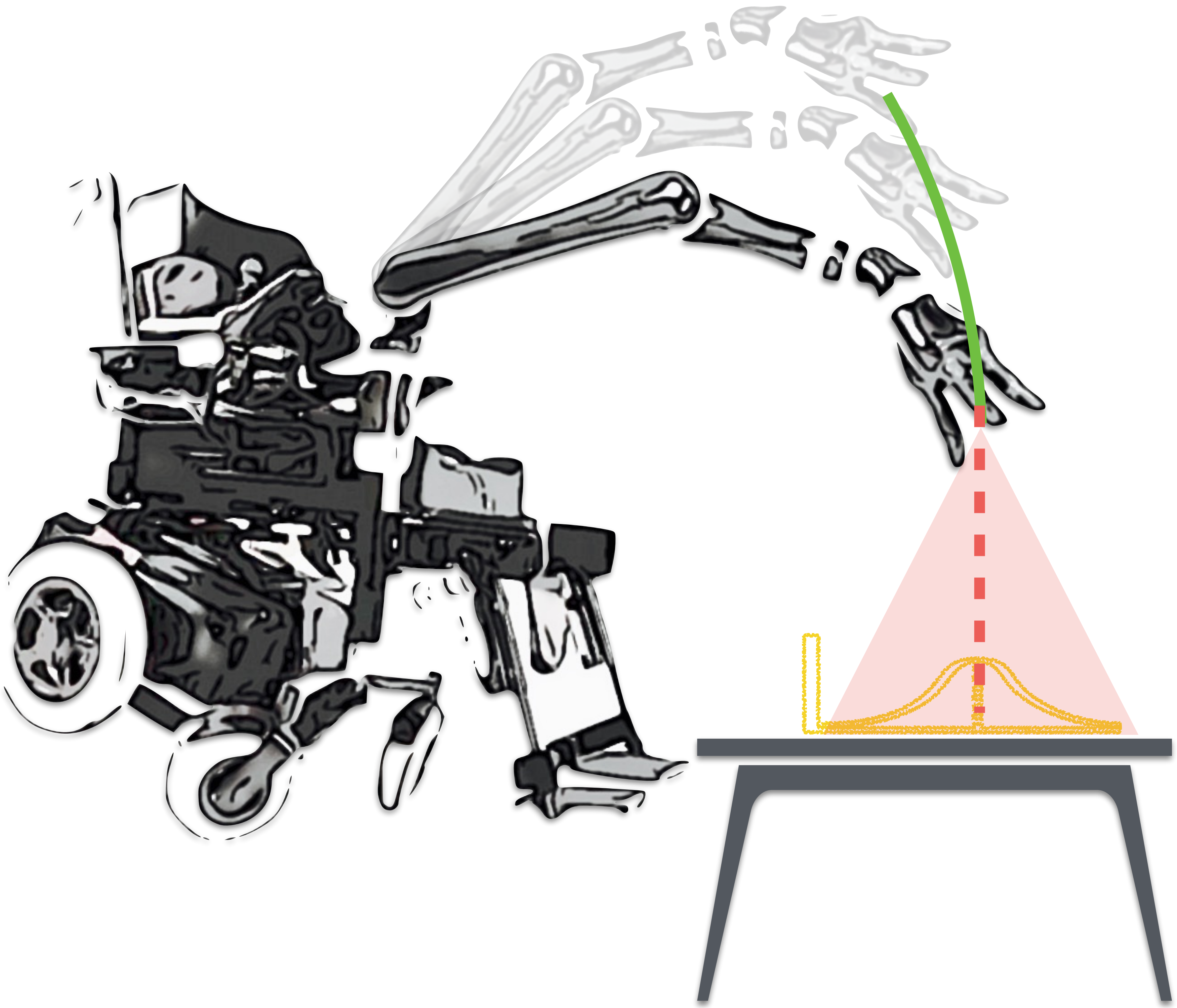
We improve the state of art by modeling a trajectory distribution



We improve the state of art by modeling a trajectory distribution



We improve the state of art by modeling a multi-modal trajectory distribution

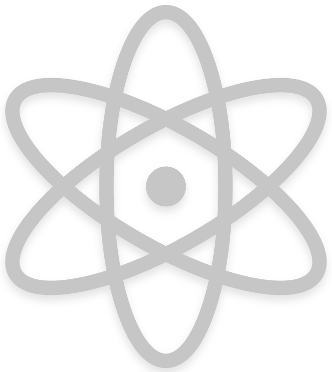


Top-down view:

Outline

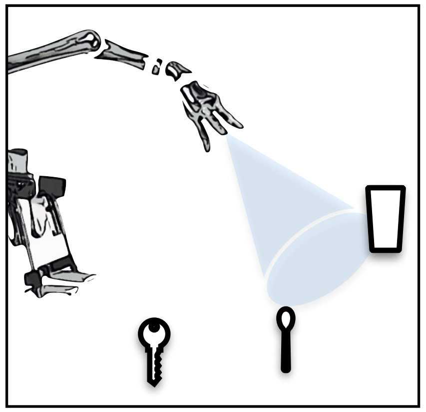
State of Art

Innovation: dynamics and multimodality



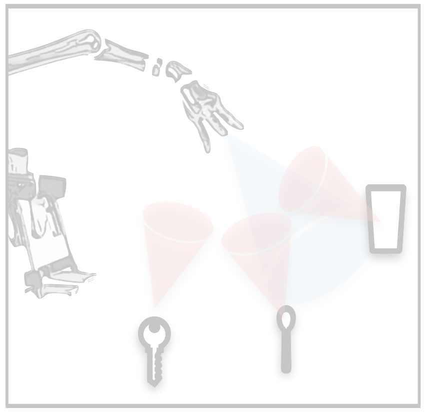
Motion Prediction

From the user's motion onset

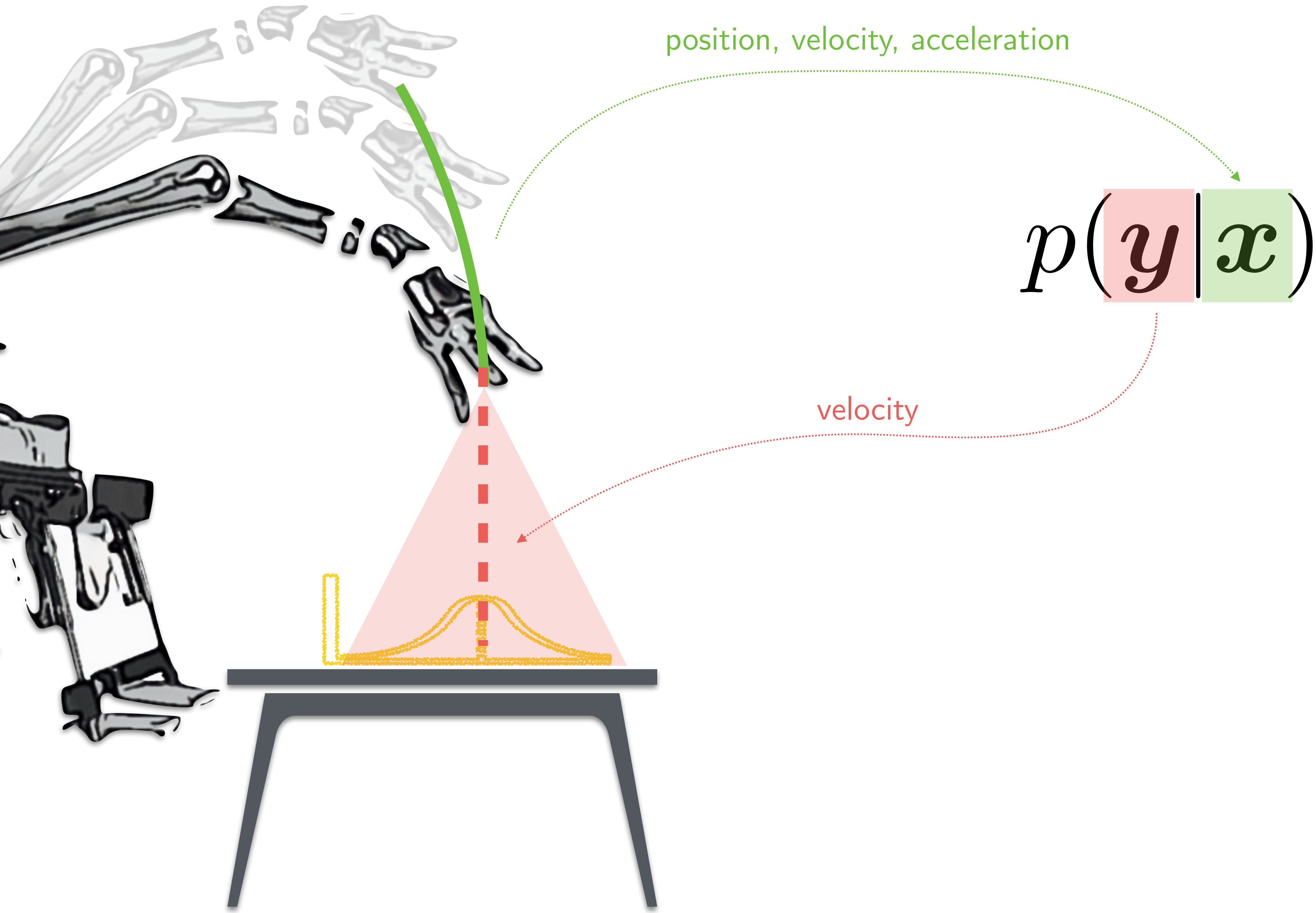


Intention Prediction

Motion prediction + goal assessment

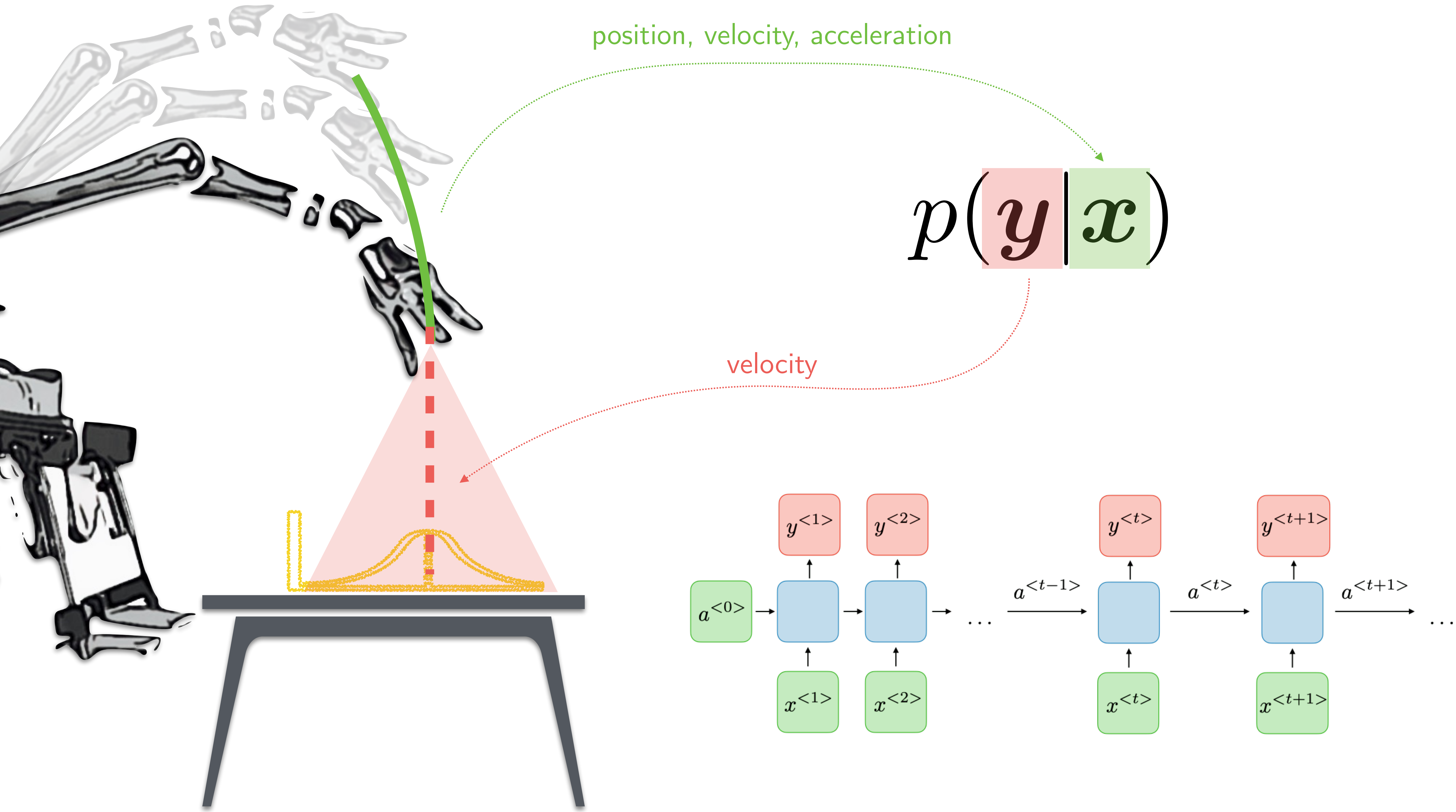


Our trajectory model encodes motion dynamics



$$p(\boldsymbol{y} \mid \boldsymbol{x})$$

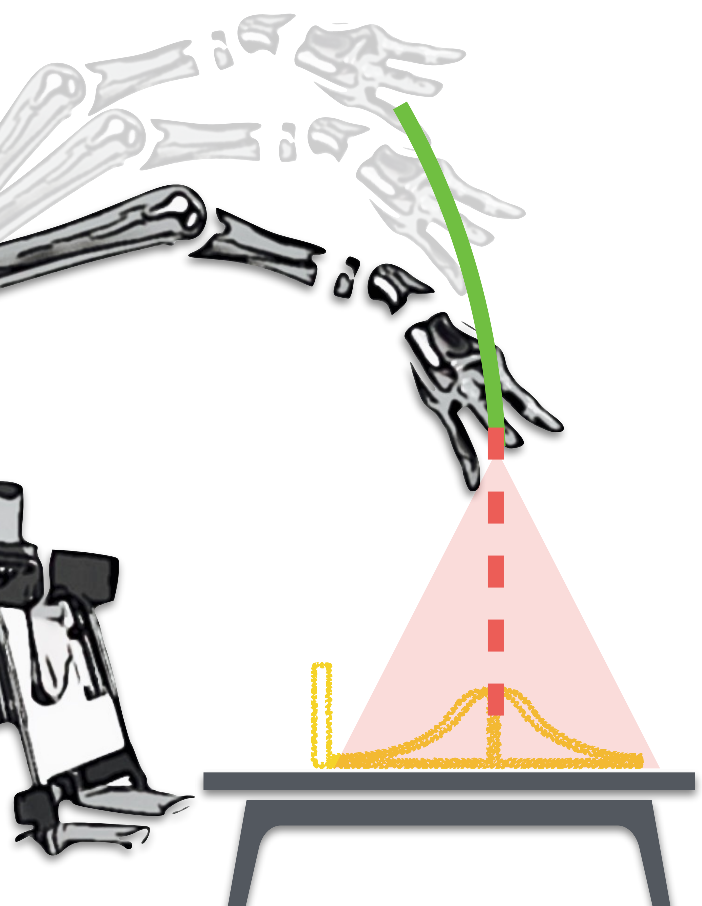
Our trajectory model encodes motion dynamics using recurrent neural networks

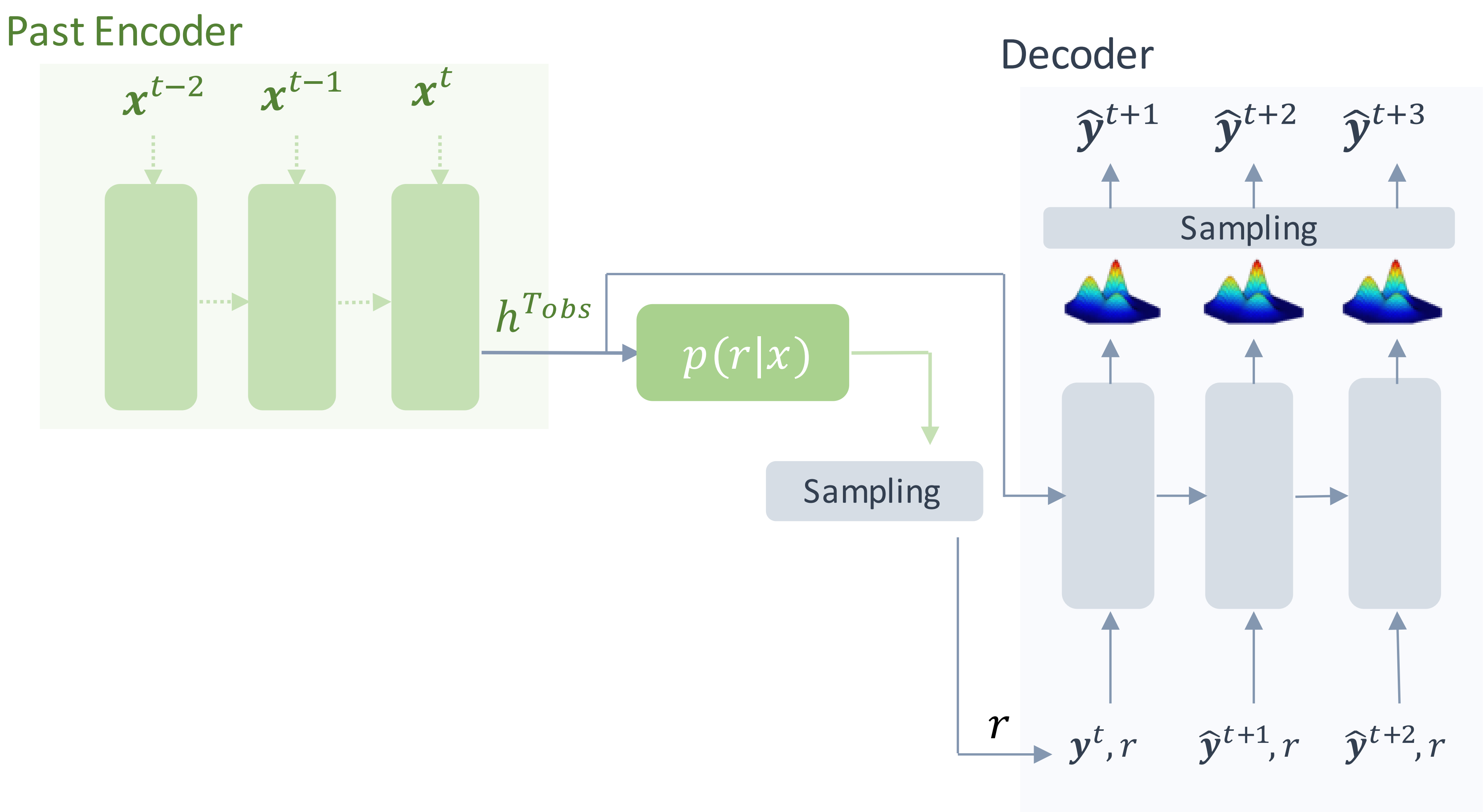


https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks

We introduce a latent variable r, to facilitate the encoding of a low-dimensional, multi-modal representation of trajectory data

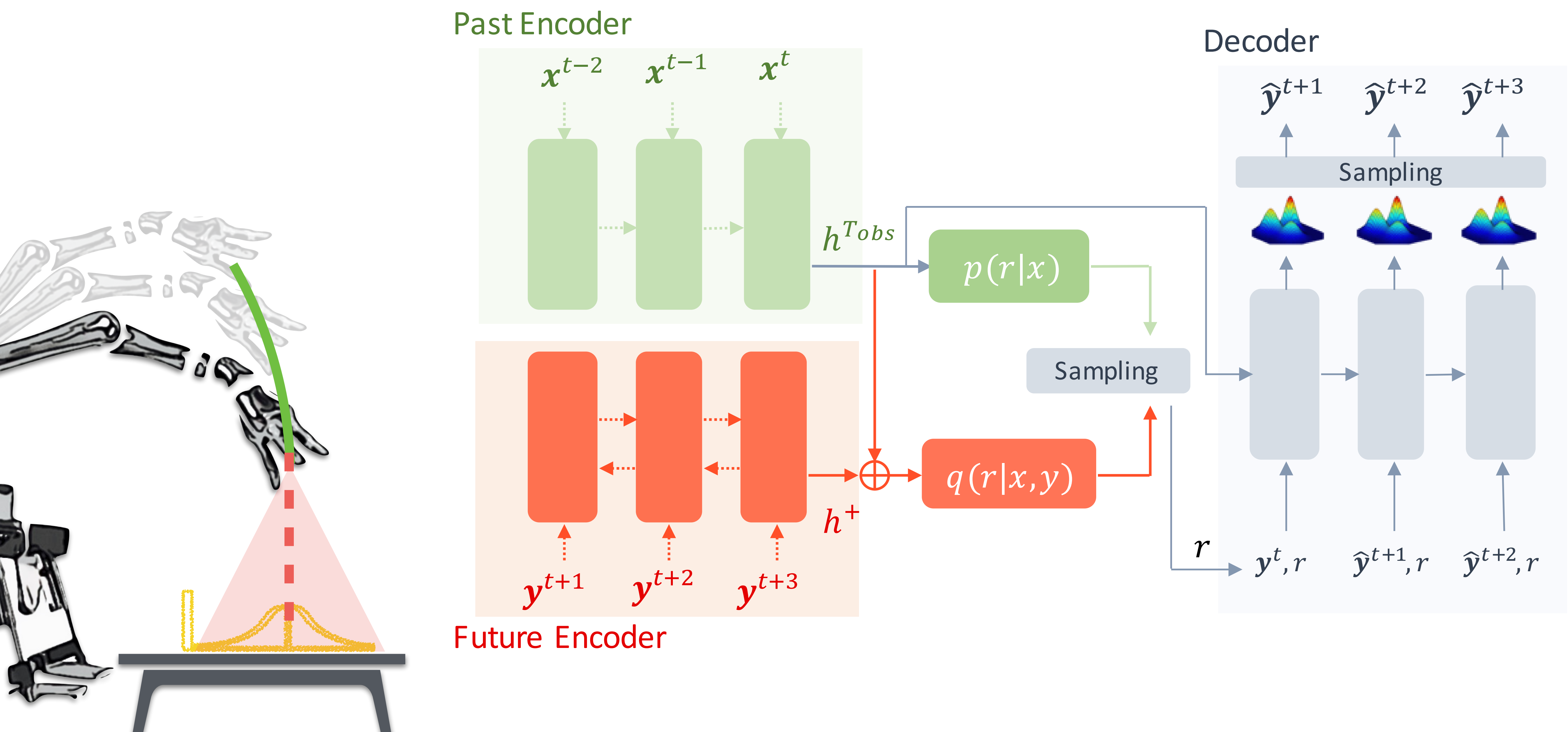
$$p(\boldsymbol{y}|\boldsymbol{x}) = \sum_{\boldsymbol{r}} p_{\psi}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{r}) p_{\theta}(\boldsymbol{r}|\boldsymbol{x})$$





This model identifies to a CVAE.
We approximate $p_\theta(r|x)$ with $q(r|x,y)$.

$$p(\boldsymbol{y}|\boldsymbol{x}) = \sum_{\boldsymbol{r}} p_\psi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{r})\,p_\theta(\boldsymbol{r}|\boldsymbol{x})$$



We call our model *Robot Trajectron* ("RT")

Robot Trajectron: Trajectory Prediction-based Shared Control for Robot Manipulation

Pinhao Song[1], Pengteng Li[4], Erwin Aertbeliën[1,2], Renaud Detry[1,3]

ICRA2024 YOKOHAMA | JAPAN

***Abstract*— We address the problem of (a) predicting the trajectory of an arm reaching motion, based on a few seconds of the motion's onset, and (b) leveraging this predictor to facilitate shared-control manipulation tasks, by reducing the**

that goal [1], [2], [3], which does not always hold true. Furthermore, most intent estimators rely on position-based methods, which consider only the distance between gripper *position* (past or predicted) and each goal to infer the user's

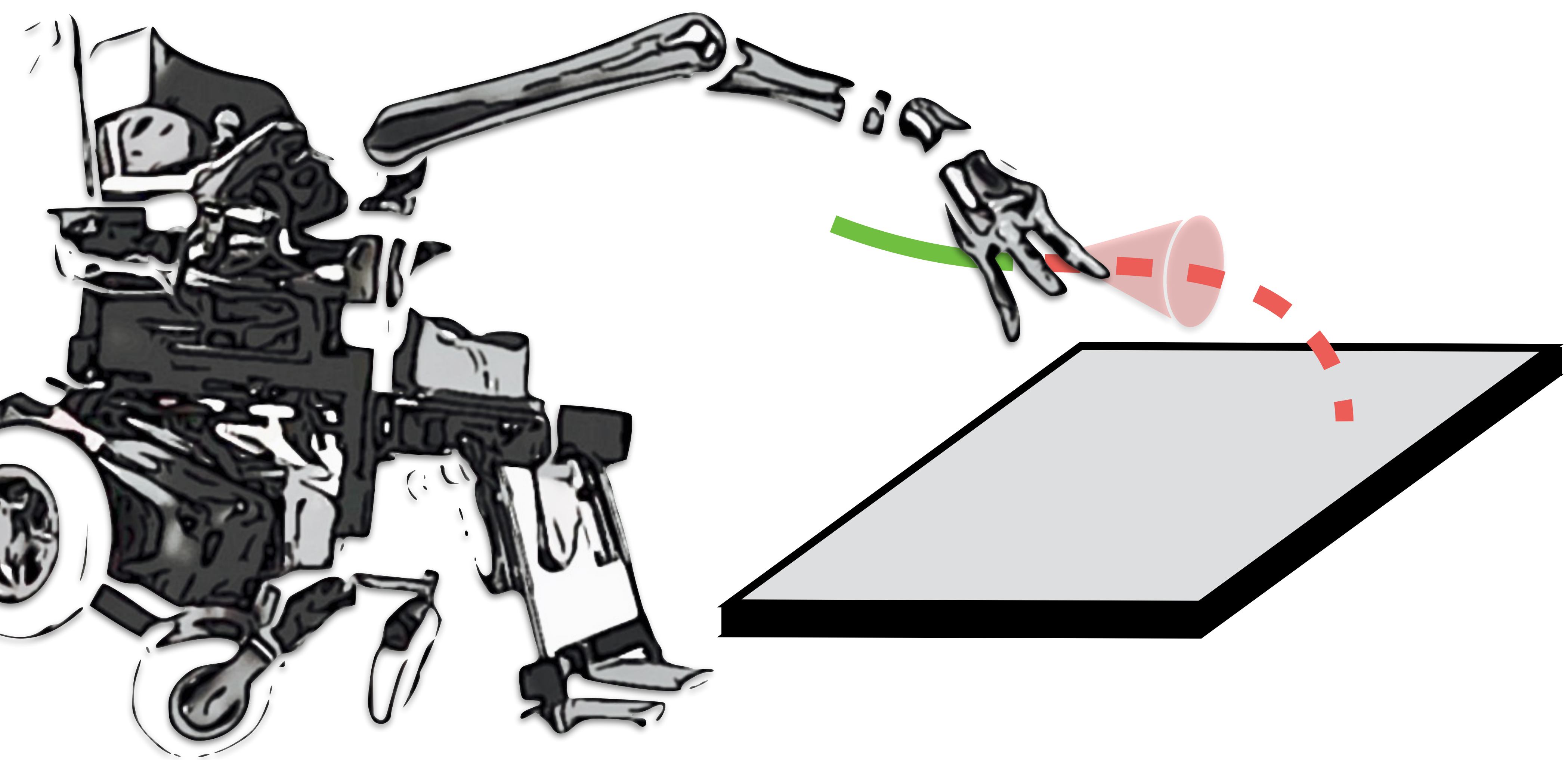
Inspired by "Boris Ivanovic, Marco Pavone
The Trajectron: Probabilistic Multi-Agent Trajectory Modeling With Dynamic Spatiotemporal Graphs"

Our Experiments showed that RT outperforms a direct LSTM on a trajectory prediction task.

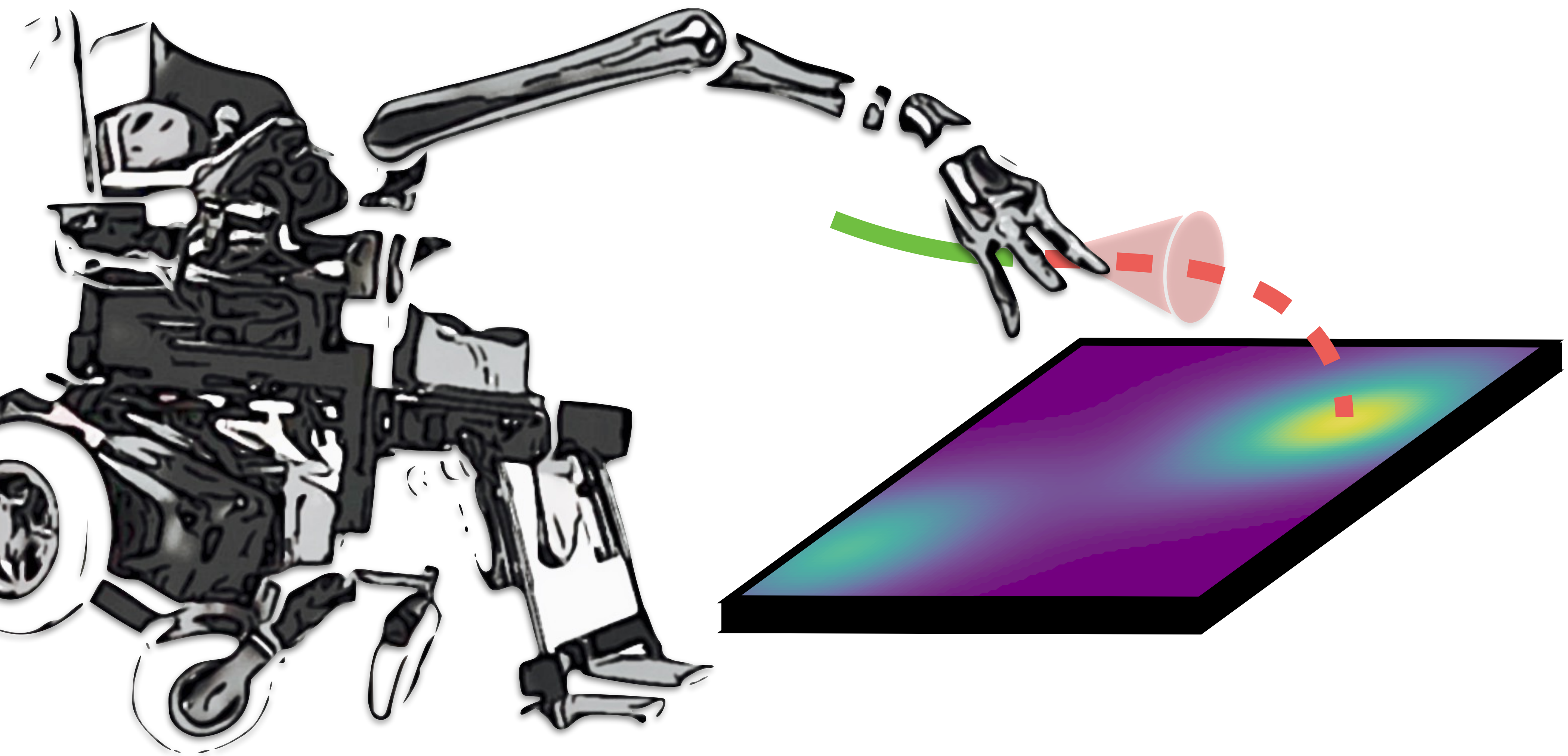
RT trained on 100k trajectories collected in sim.

Method	Most likely (mm)	
	ADE	FDE
Vanilla LSTM [29]	136.95	115.47
Robot Trajectron	30.58	49.94

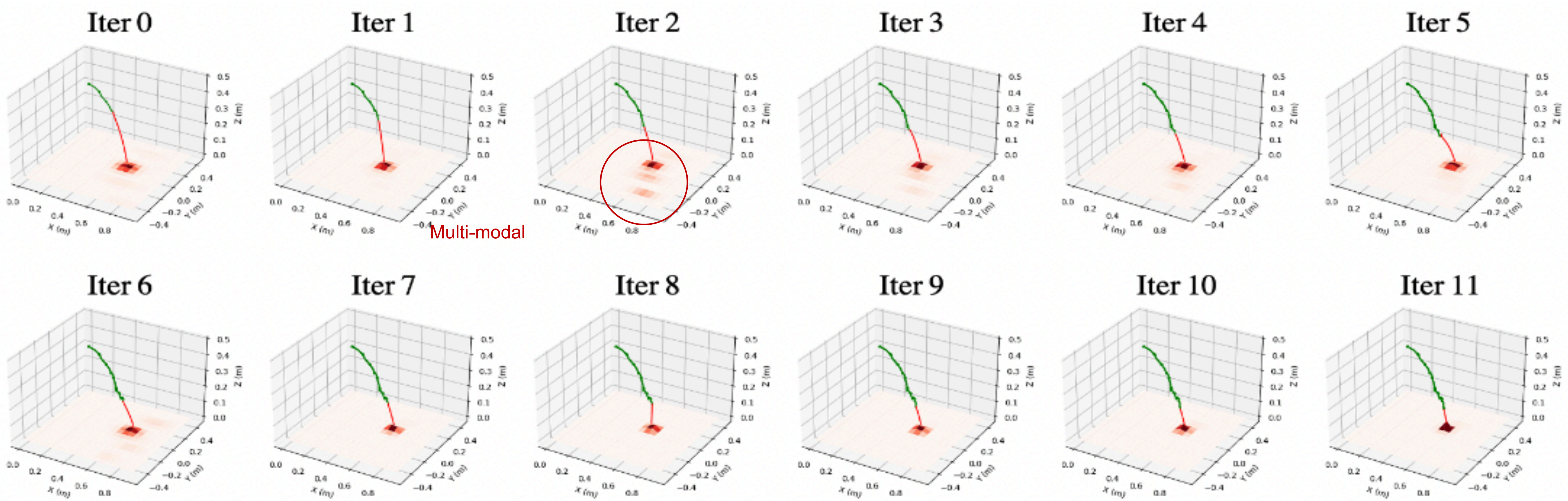
In a tabletop scenario, trajectory forecasts can be intersected with the table plane



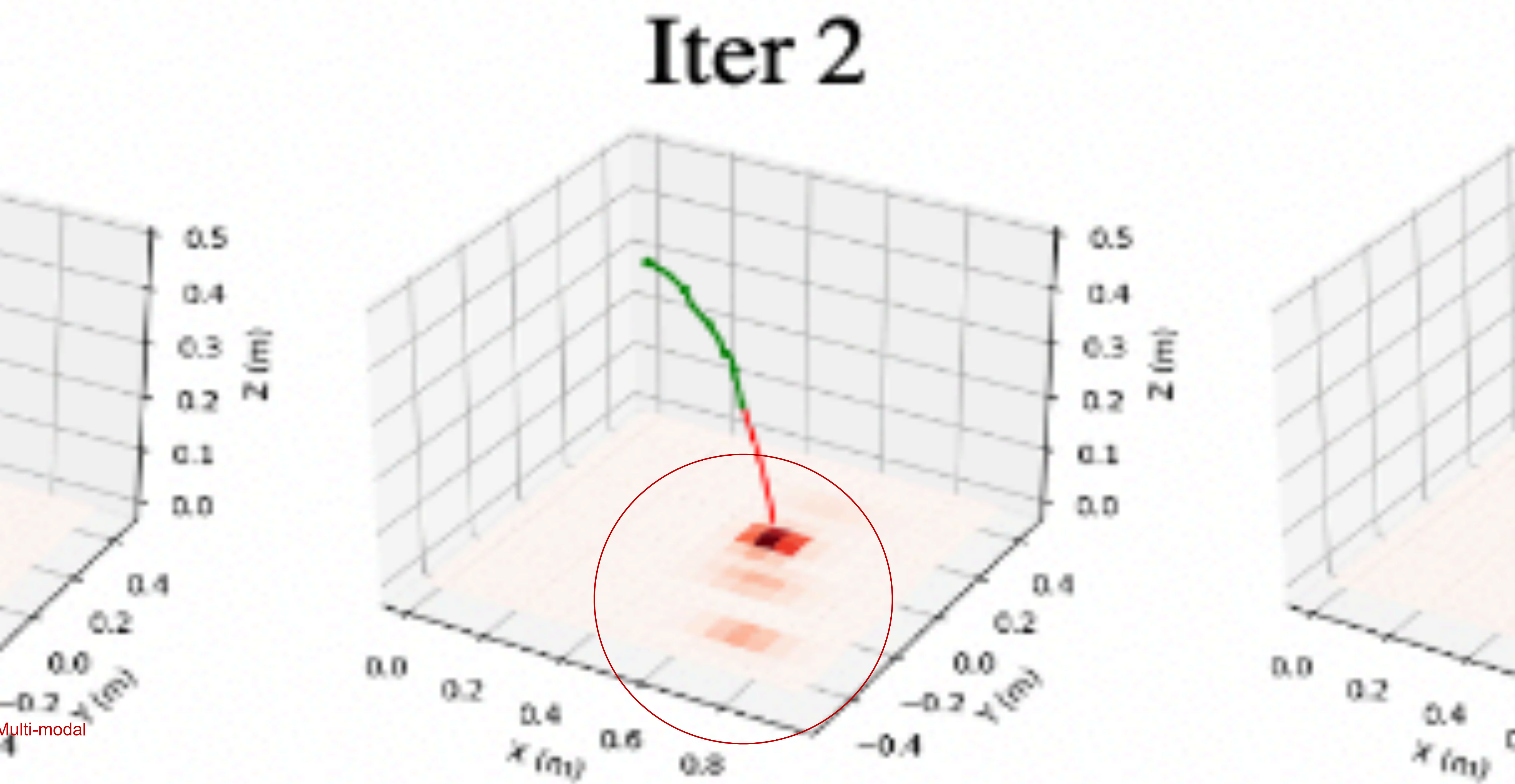
Similarly, RT's probabilistic motion predictions can be projected onto the table plane



Our sim experiments demonstrate RT's multi-modal prediction capacity



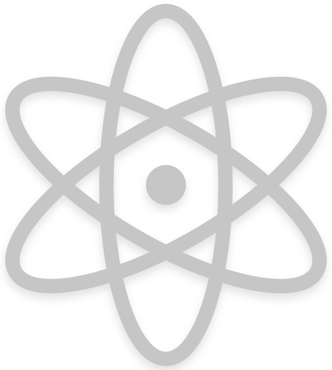
Our sim experiments demonstrate RT's multi-modal prediction capacity



Outline

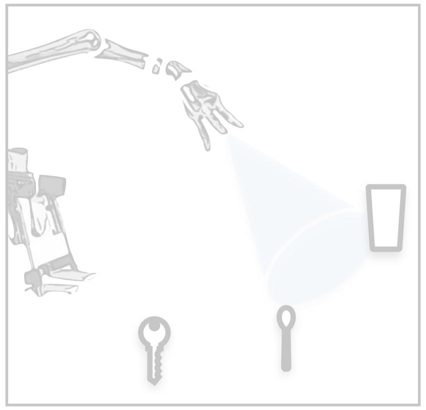
State of Art

Innovation: dynamics and multimodality



Motion Prediction

From the user's motion onset



Intention Prediction

Motion prediction + goal assessment

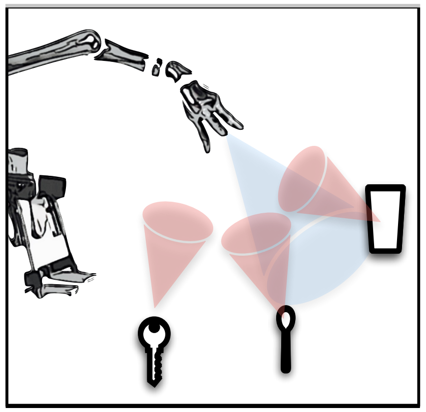
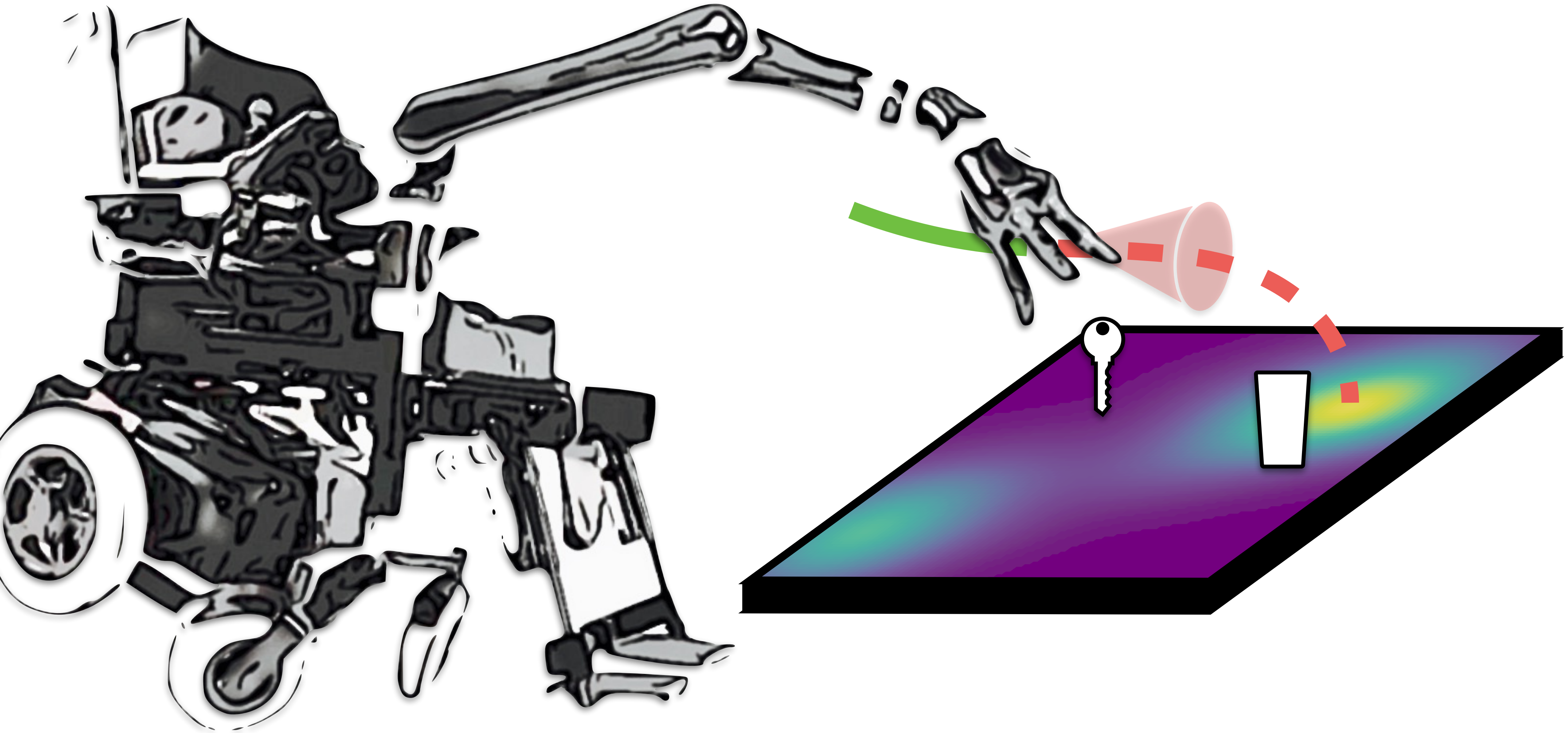
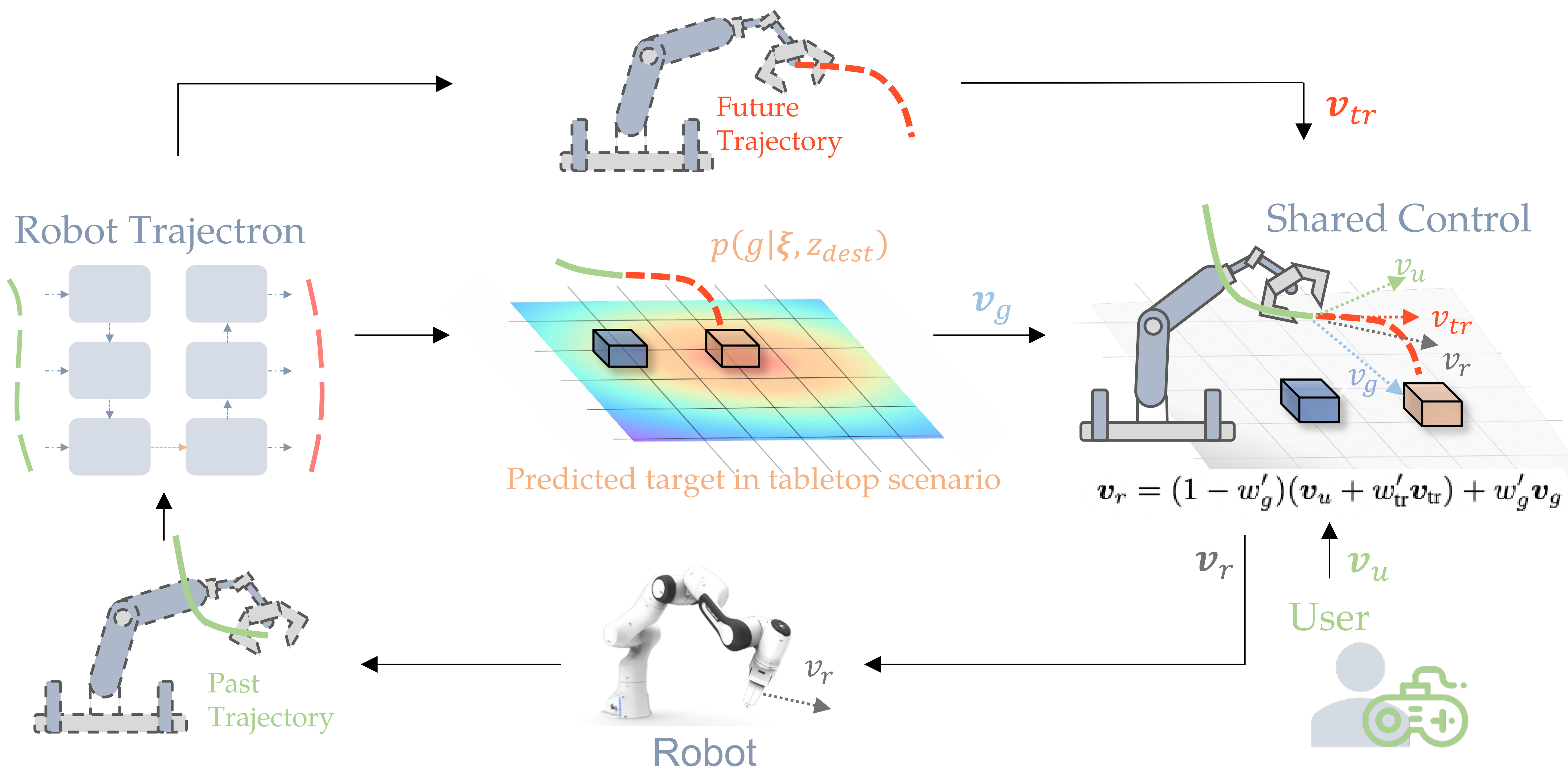


Table-plane projections allow us to intuitively reconcile goal locations and trajectory forecasts

Implicit Grasp Diffusion: Bridging the Gap between Dense Prediction and Sampling-based Grasping
Pinhao Song, Pengteng Li, Renaud Detry – CoRL 2024



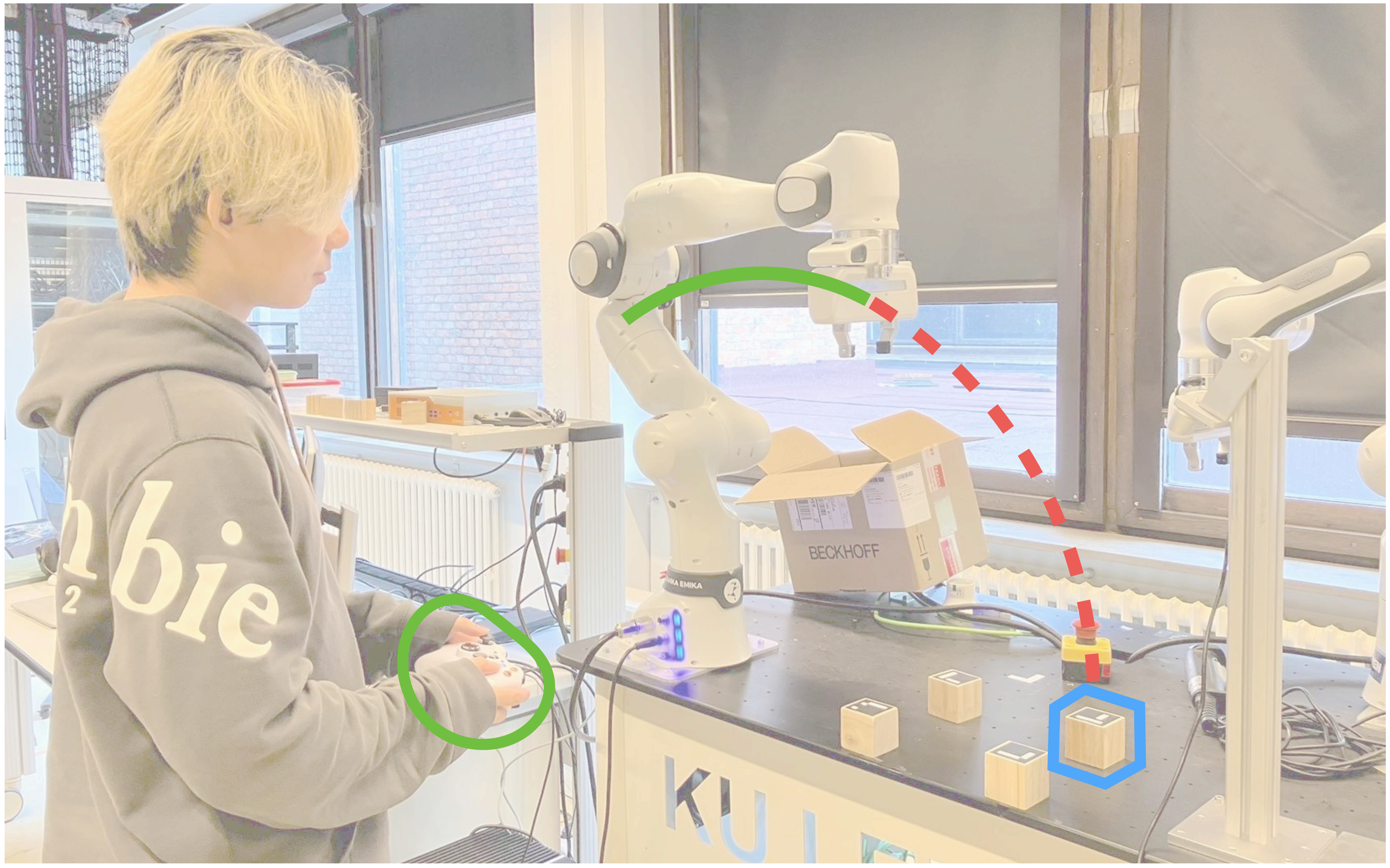
We leveraged RT to assemble a simple shared controller that combines user input, motion forecasts and goal locations



Our simple RT-based shared controller performs on-par with MaxEnt IOC, a state-of-art shared-control method

- **Subjects**: 10, from local community
- **Objets**: 4 small cubes equipped with ArUCo markers
- **Task**: the subject was required to gradually approach one object
- **Baseline 1**: no assistance (direct teleoperation)
- **Baseline 2**: MaxEnt IOC [1]

Method	Time (sec)	Input	Average l_{tr} (m)
Teleop. [3]	9.36±0.71	41.8±2.8	2.452±0.246
MaxEnt IOC [3]	7.24±0.33	33.8±1.2	2.007±0.060
Robot Trajectron	7.17±0.43	33.8±1.3	1.981±0.092



[1] Gottardi, A., Tortora, S., Tosello, E., & Menegatti, E. (2022). Shared control in robot teleoperation with improved potential fields. *IEEE Transactions on Human-Machine Systems*, *52*(3), 410-422.

Robot Trajectron ("RT") is a motion dynamics model that with multi-modal motion forecasting capabilities

Pinhao Song, Pengteng Li, Erwin Aertbelien, and Renaud Detry. *Robot Trajectron: Trajectory Prediction-based Shared Control for Robot Manipulation.* ICRA 2024.