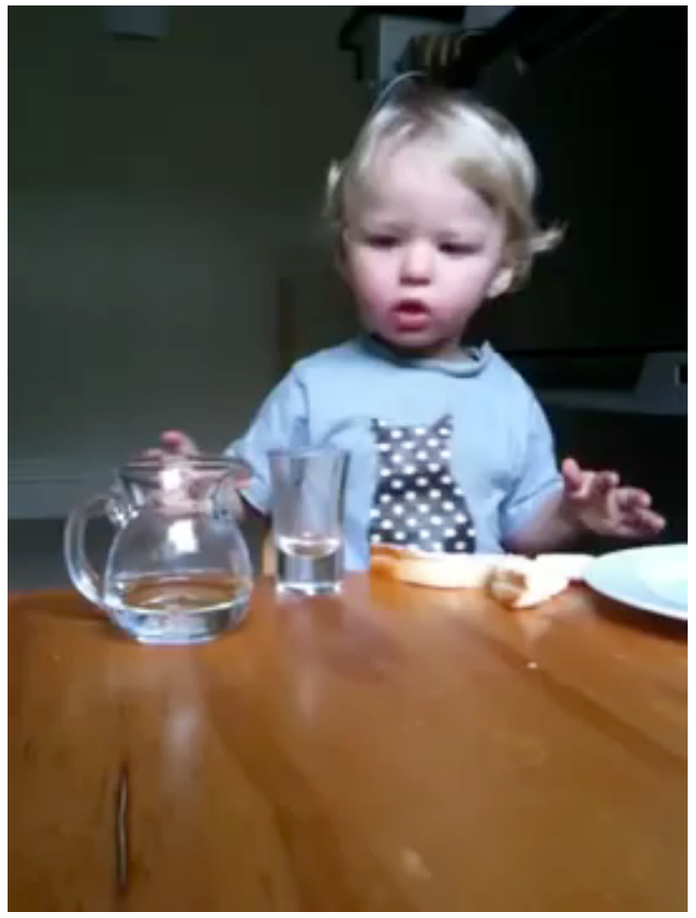
Beyond Learning from Reward

reward

Mnih et al. ’15

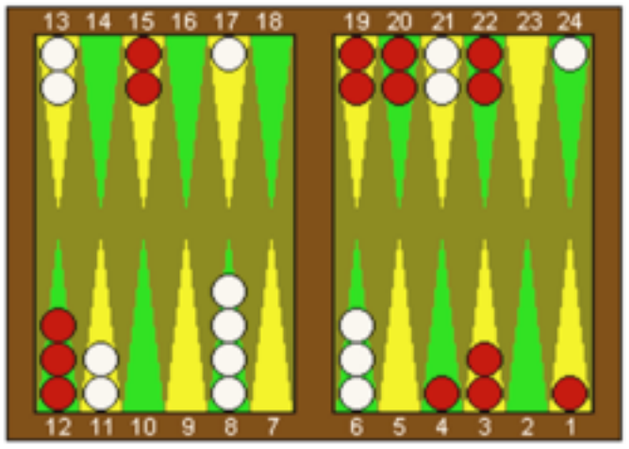


reinforcement learning agent

what is the reward?

In the real world, humans don’t get a score.

video from Montessori New Zealand



Tesauro ’95

Kohl & Stone, ’04

Mnih et al. ’15

AlphaGo

Silver et al. ‘16

reward function is essential for RL

real-world domains: reward/cost often difficult to specify

- robotic manipulation
- autonomous driving
- dialog systems
- virtual assistants
- and more…

What other forms of supervision?

1. demonstrated behavior -> imitation, inferring intention
2. self-supervision, prediction -> model-based control
3. auxiliary objectives and additional sensing modalities

Learning from Demonstrated Behavior

Human Learning via imitation

8 months: imitate simple actions & expressions

18 months: imitate after a delay and multi-step actions

36 months: imitate multi-step actions after a delay

Developmental Psychology
1988, Vol. 24, No. 4, 470–476

0012-1649/88/$00.75

Infant Imitation After a 1-Week Delay: Long-Term Memory for Novel Acts and Multiple Stimuli

Andrew N. Meltzoff
University of Washington

Deferred imitation after a 1-week delay was examined in 14-month-old infants. Six actions, each using a different object, were demonstrated to each infant. One of the six actions was a novel behavior that had a zero probability of occurrence in spontaneous play. In the imitation condition, infants observed the demonstration but were not allowed to touch the objects, thus preventing any immedi-

Learning from imitation

For autonomous driving:

Bojarski et al. ‘16

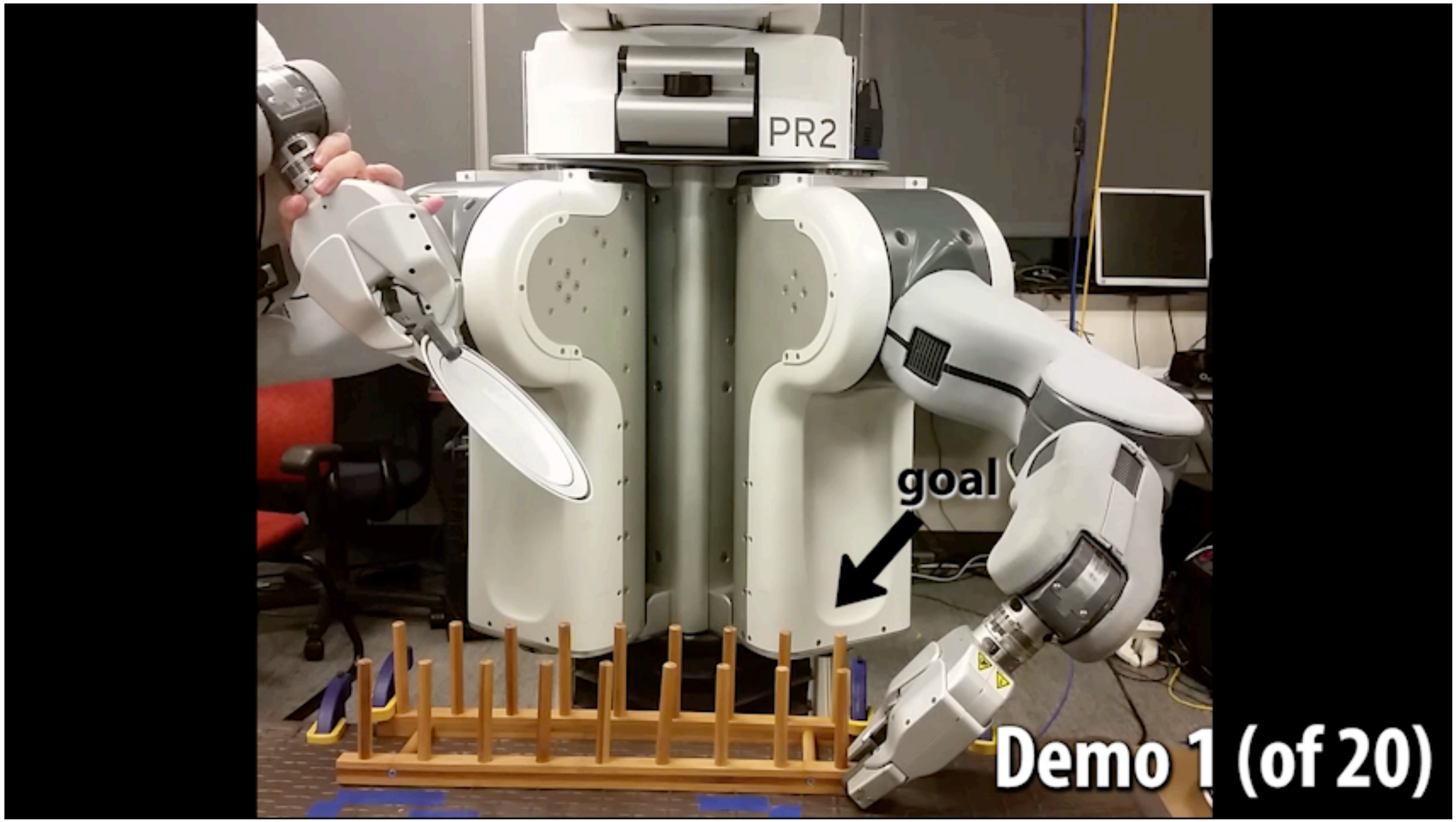
More than imitation: inferring intentions

18 months: understand others intentions and help

Warneken & Tomasello ‘06

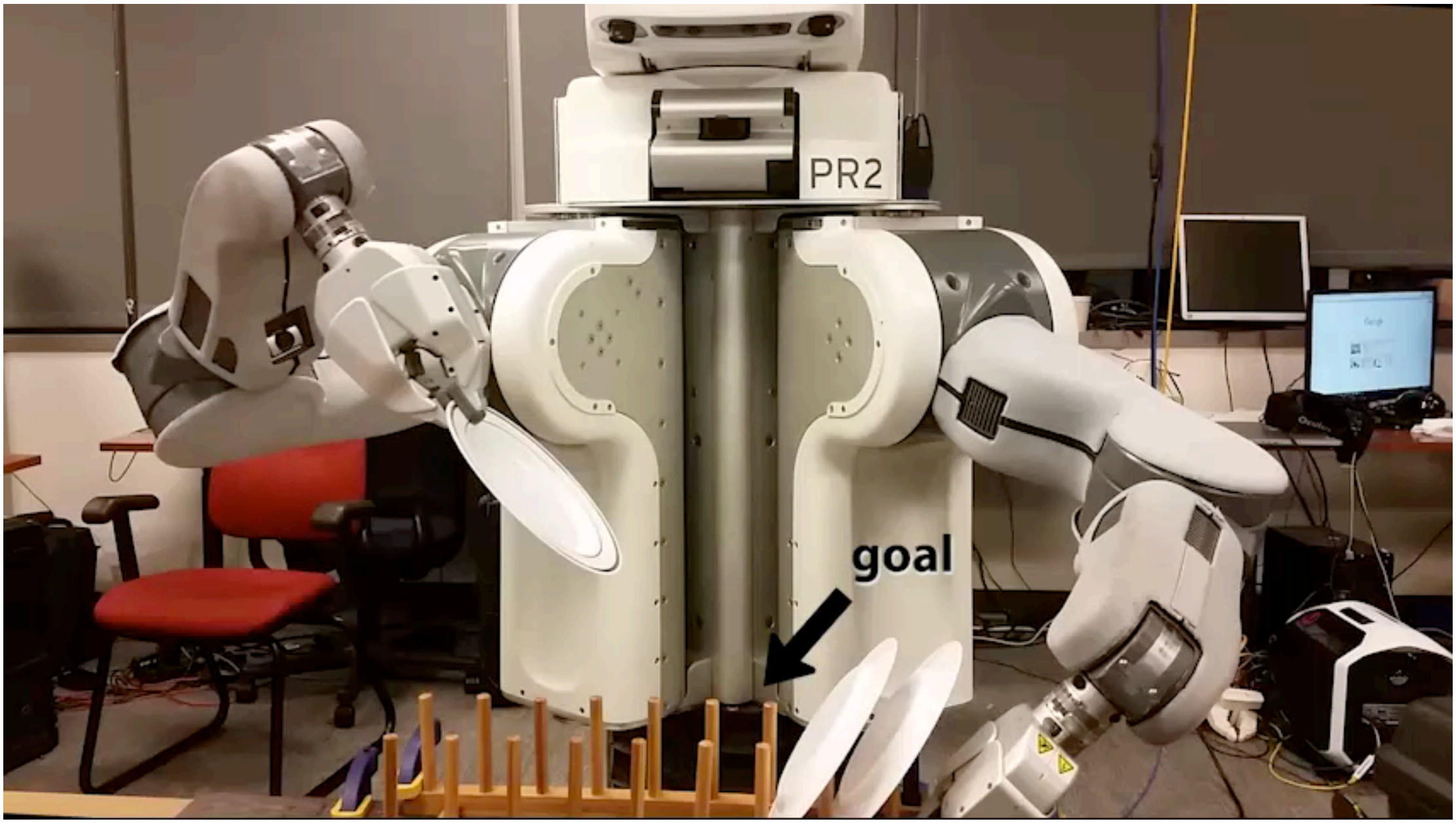
-> known as **inverse reinforcement learning**

Inverse RL: demonstrations



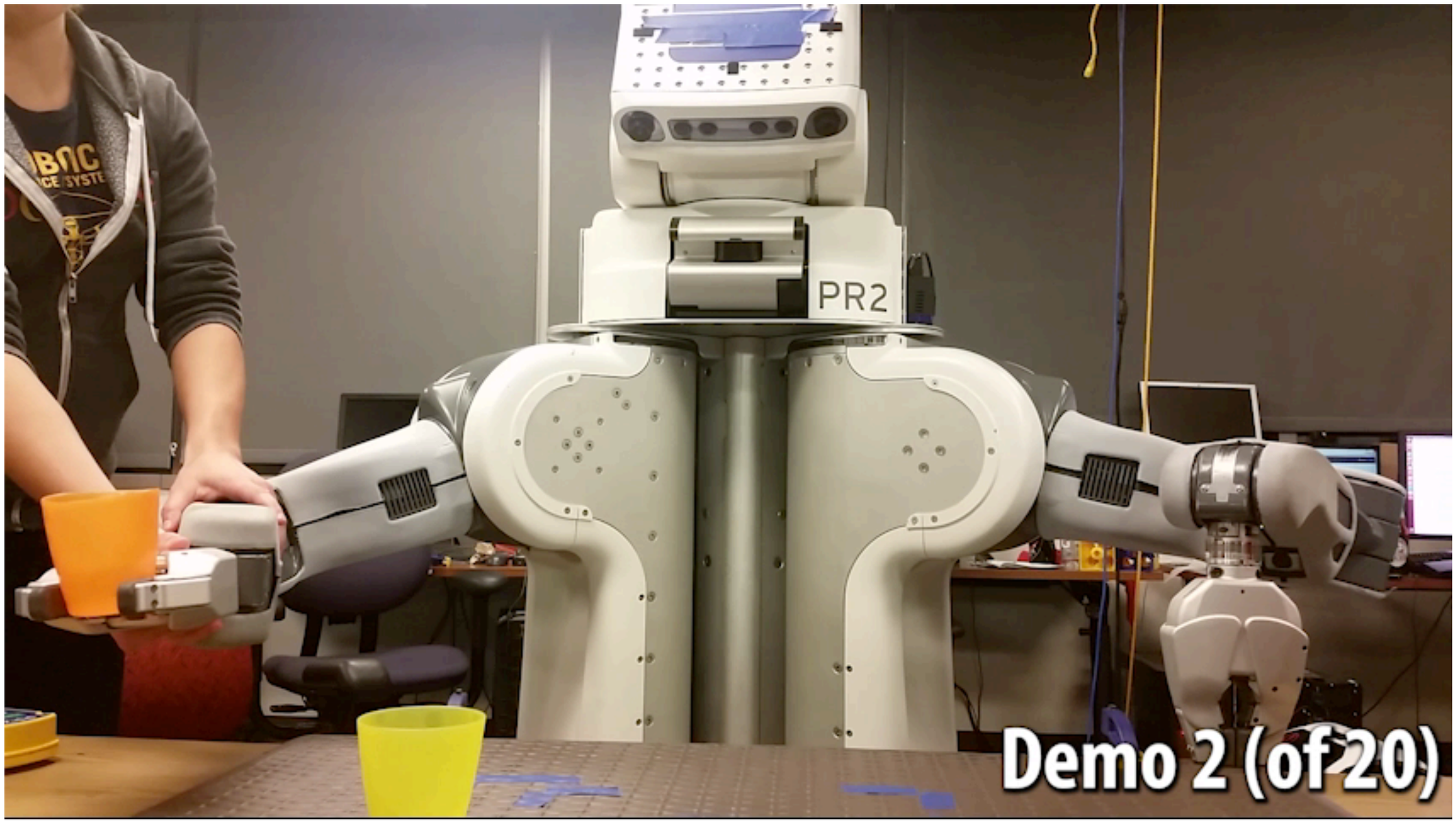
Finn et al. ‘16

Inverse RL: learned behavior



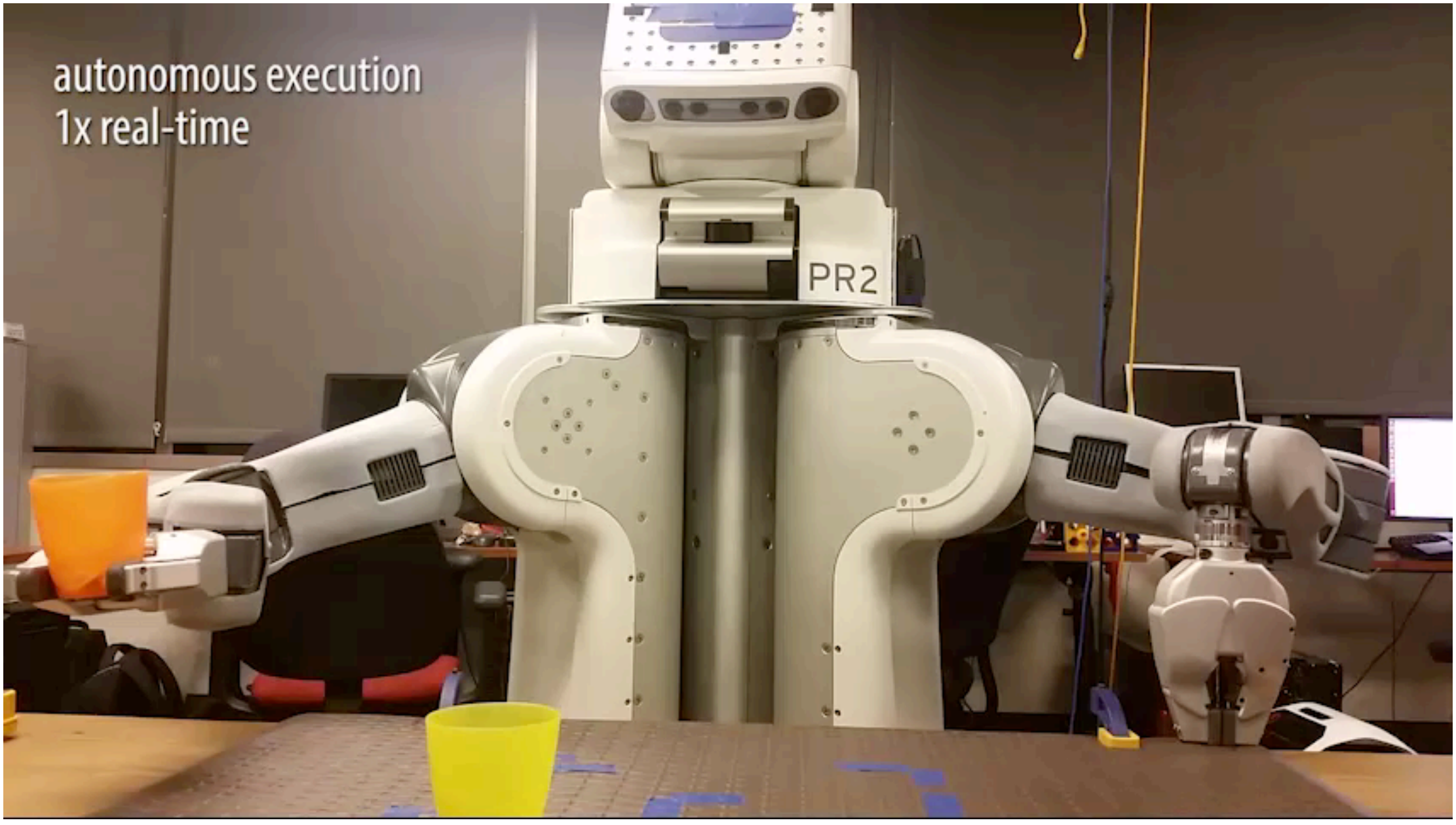
Finn et al. ‘16

Inverse RL: demonstrations



Finn et al. ‘16

Inverse RL: learned behavior



Finn et al. ‘16

Behavior via Prediction

Prediction

"the idea that we **predict the consequences of our motor commands** has emerged as an important theoretical concept in all aspects of sensorimotor control"

Prediction Precedes Control in Motor Learning

J. Randall Flanagan,[1,*] Philipp Vetter,[2] Roland S. Johansson,[3] and Daniel M. Wolpert[2]

Procedures for details). Figure 1 shows, for a single subject, the hand path (top trace) and the grip (middle)

Predicting the Consequences of Our Own Actions: The Role of Sensorimotor Context Estimation

Sarah J. Blakemore, Susan J. Goodbody, and Daniel M. Wolpert

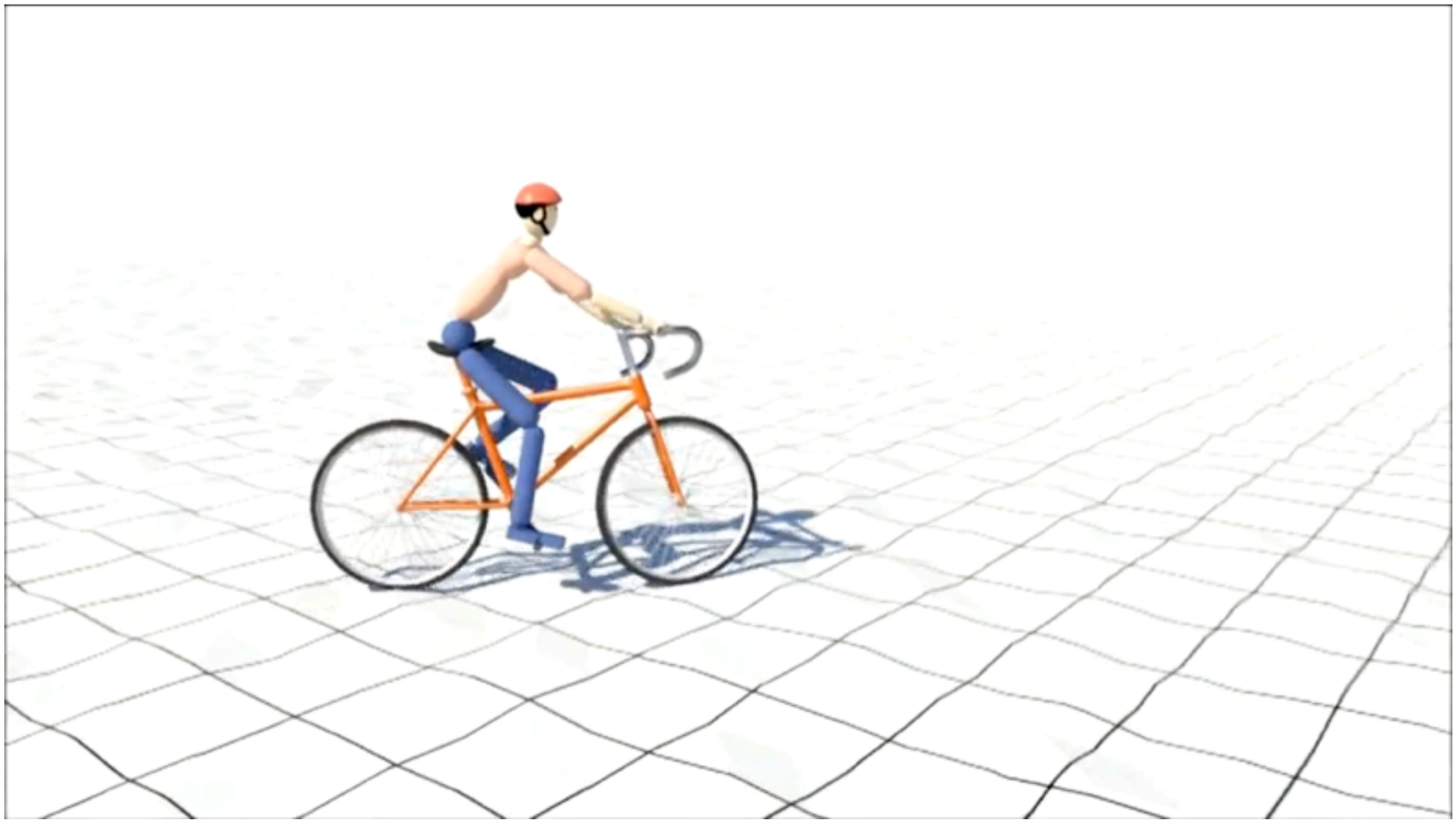
Sobell Department of Neurophysiology, Institute of Neurology, University College London, London WC1N 3BG,

Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects

Rajesh P. N. Rao[1] and Dana H. Ballard[2]

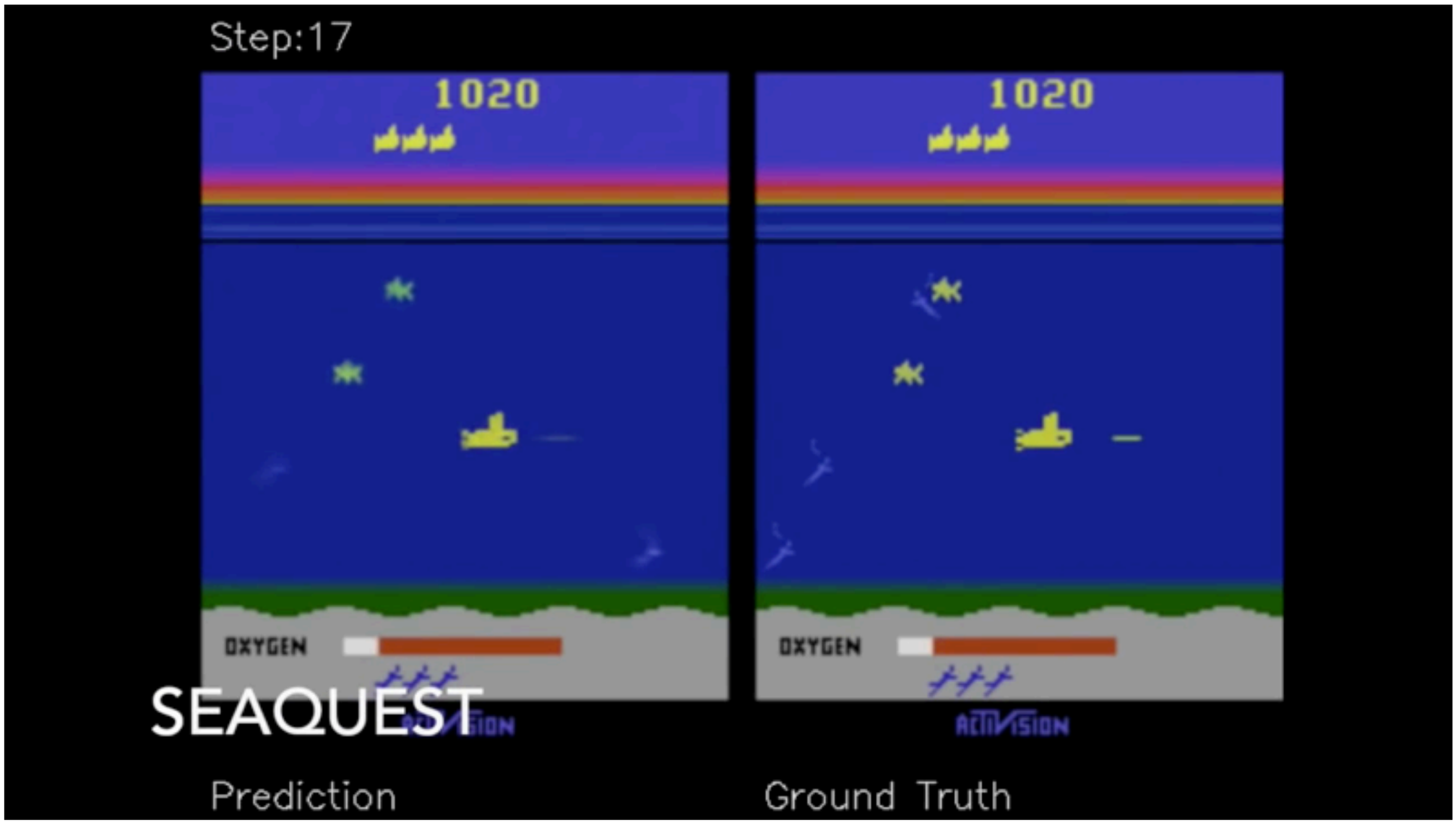
Prediction

With a perfect model & optimization:



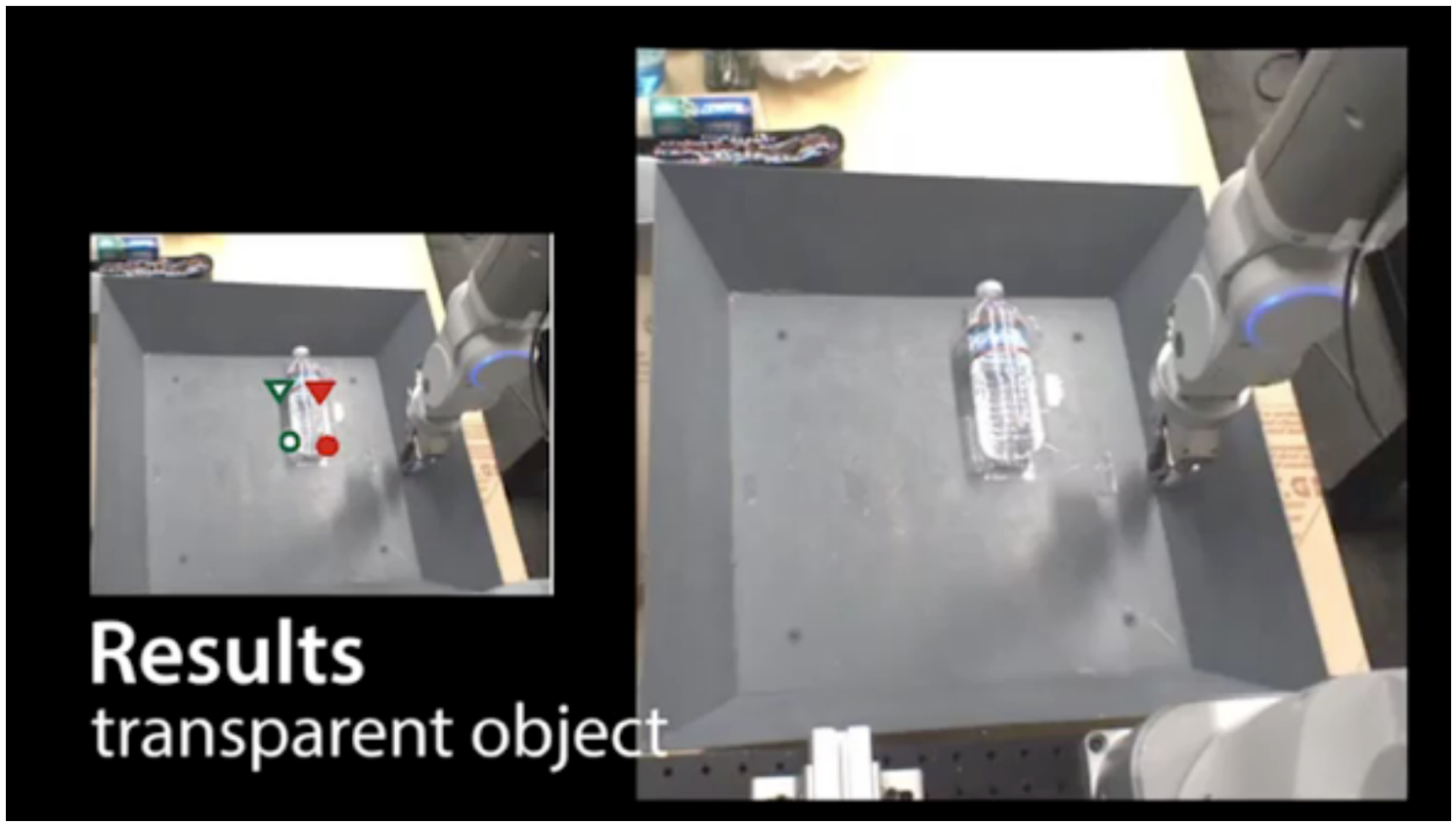
Tan et al. ‘14

Learning to Predict



Oh et al. ‘15

Learning to Predict

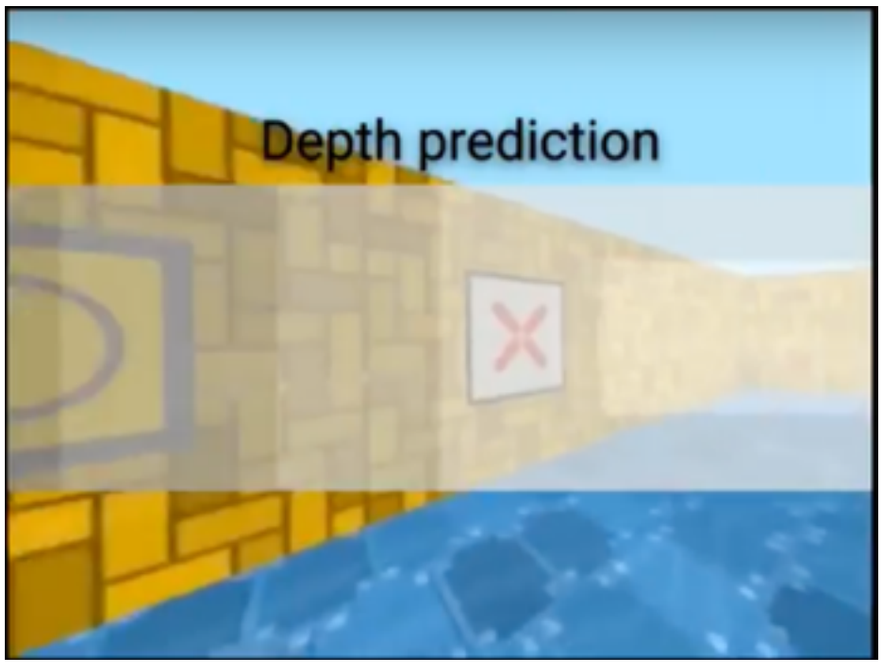


Finn et al. ’16, ‘17

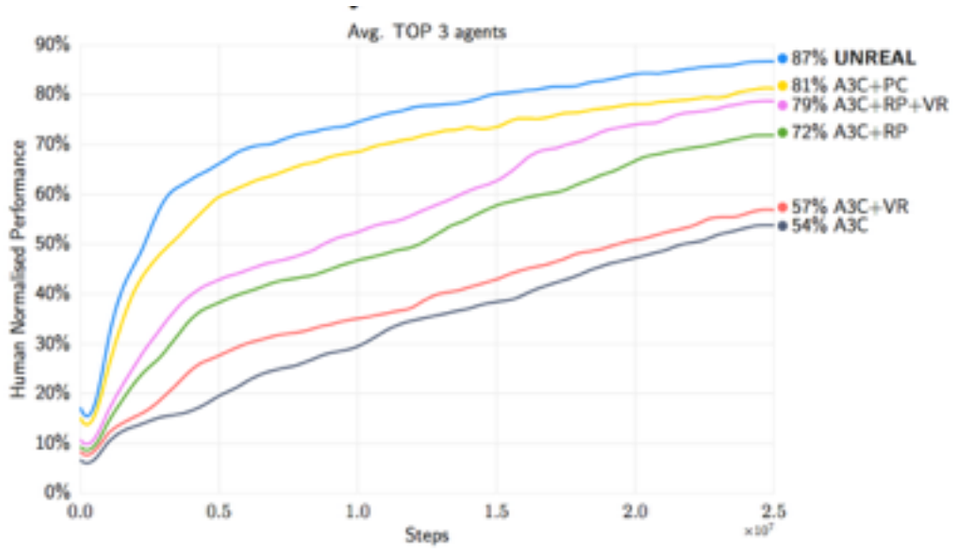
Auxiliary Objectives and Sensing Modalities

Sources of Auxiliary Supervision

- other sensing modalities (touch, audio, depth)
- learning multiple, related tasks
- task-relevant properties of the world



Mirowski et al. ‘17



Jaderberg et al. ‘17

In this class…

1. demonstrated behavior -> imitation, inferring intention
2. self-supervision, prediction -> model-based control
3. auxiliary objectives and additional sensing modalities

Note: RL can be combined with these other forms of information!