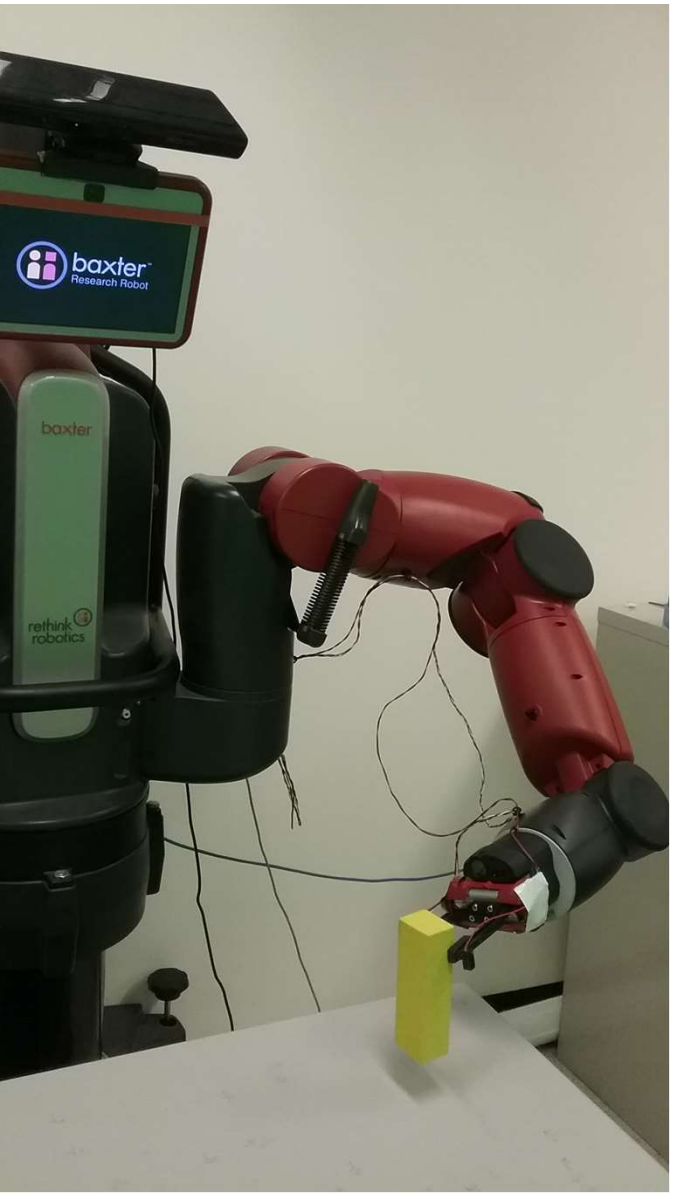


Towards Learning the Foundations of Manipulation Actions from Unguided Exploration

Jon Juett

Thesis Defense

Computer Science and Engineering
University of Michigan

July 19, 2021

What Problem Are We Solving?

- Human infants learn to reach, grasp, and place objects.
  - Infants are not born with these capabilities.
  - In well under a year, infants learn to perform these actions reliably.
    - (Berthier, 2011)

- We have developed a computational model of how an embodied learning agent can learn to reach, grasp, and place objects.
  - We have implemented and evaluated it on a physical robot.

Why Is This Problem Important?

- Learning actions for purposefully reaching, grasping, and placing objects is foundational for an agent interacting with its world.

- Solving this problem may provide insights into:
  - how to learn other actions from unguided experience;
  - human infant developmental learning;
  - other kinds of spatial knowledge.

What Does the Literature Tell Us?

- There is a huge literature on robot motion planning based on precise models of environmental geometry and robot geometry, kinematics, and dynamics.
  - Our focus is on learning reliable actions from experience without precise prior models.

- Developmental Psychology
  - (von Hofsten, 1991), (Thelen et al., 1993, 1996), (Clifton et al., 1993, 1994, 1999), (Adolph and Berger, 2005), (Bremner et al., 2008), (Berthier, 2011), (Corbetta et al., 2014)
  - Well-documented timeline and characteristics of typical infant action learning
- Neural Modeling
  - (Oztop et al., 2004), (Chinellato et al., 2011), (Savastano and Nolfi, 2013), (Roncone et al., 2016), (Serino, 2019)
  - Carefully designed networks can demonstrate human-like learning stages and behaviors.
- Sensorimotor Modeling
  - (Hulse et al., 2010), (Jamone et al., 2012, 2014), (Law et al., 2014a, 2014b), (Ugur et al., 2015), (Luo et al., 2018), (Kumar et al., 2021)
  - Learning a mapping between proprioceptive and visual sensors allows learning actions from experience.

Learning Efficiency

- Robotics traditionally relies on reinforcement learning.
  - Very high sample complexity for RL, especially for deep RL
    - (Pinto and Gupta, 2016), (Levine et al., 2016, 2018)

- Infants learn very quickly (a few months)
  - from autonomous practice;
  - without formal instruction;
  - without extrinsic rewards.

- Infants learn reaching and grasping in two phases.
  - Early Phase: *"When infants first learn to reach at about 4 months, their hand paths are jerky and tortuous"* (Thelen et al., 1996)
  - Late Phase: *"With age, infants' reaches become straighter and more directly aimed toward the target and show fewer movement units."* (Thelen et al., 1996)

Early and Late Action Learning

- The two phases suggest two learning processes.
  - The first process efficiently learns jerky actions that reliably achieve goals.
  - The second process uses RL to make those actions smooth and efficient.

- The result of early learning concentrates behavior in the neighborhood of trajectories that achieve the goal.
  - This will make RL much more efficient.

- This thesis models the first learning process.

Our Learning Model

- The agent learns the dynamics of its body and its environment.
- The agent observes the usual and unusual results of its actions.
- The agent defines actions with goals of repeating a type of unusual event.
- The agent identifies features to improve the reliability of actions.
- Features are derived from an agent-constructed spatial representation.
- The learned action will resemble an early action in infant development.

Learning Sequence

- Learning to move the arm with a self-built spatial representation
- Learning to reach and bump
- Learning to grasp
- Learning to ungrasp
- Learning to place into specific locations
- Learning to pick-and-place (grasping then placing)

- This sequence is determined by the prerequisites of each action.
- Intrinsic motivation can also produce the order of this sequence.
  - The agent receives a reward based on how quickly it improves its performance. (Schmidhuber, 1991, 2011), (Baldassarre and Mirolli, 2013)
  - Prioritizes actions with enough prerequisites learned that they can be improved efficiently

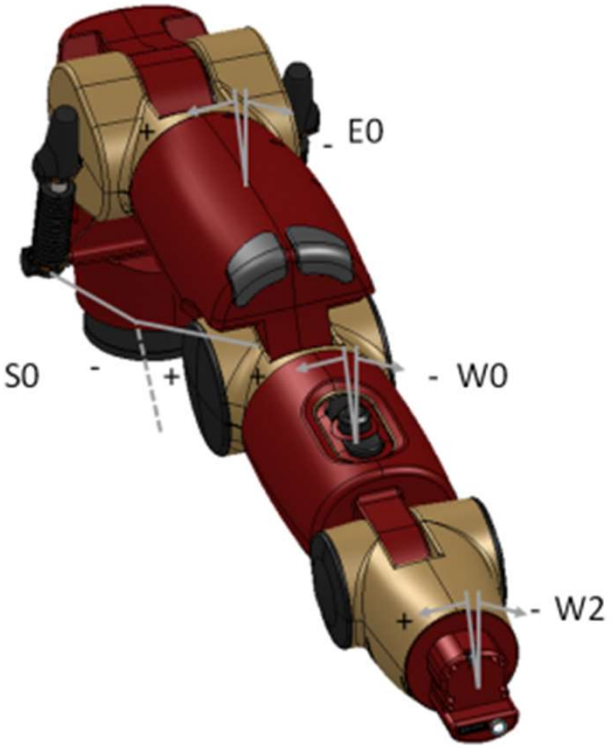
Learning to Move the Arm

Why is a representation for peripersonal space necessary?

- Peripersonal space:
  - The space surrounding an agent and in reach of its manipulators
- The agent has the ability to move the arm from one configuration to another, but…
  - the move will not necessarily be safe for all pairs of configurations.
  - the effect on the environment cannot be predicted.
- The Peripersonal Space (PPS) Graph resolves these issues.
  - Trajectories planned along graph edges will be feasible and safe.
  - A mapping between configuration space and image space allows predictions.

The Baxter Research Robot's Arm

4 "twist" joints



3 "bend" joints

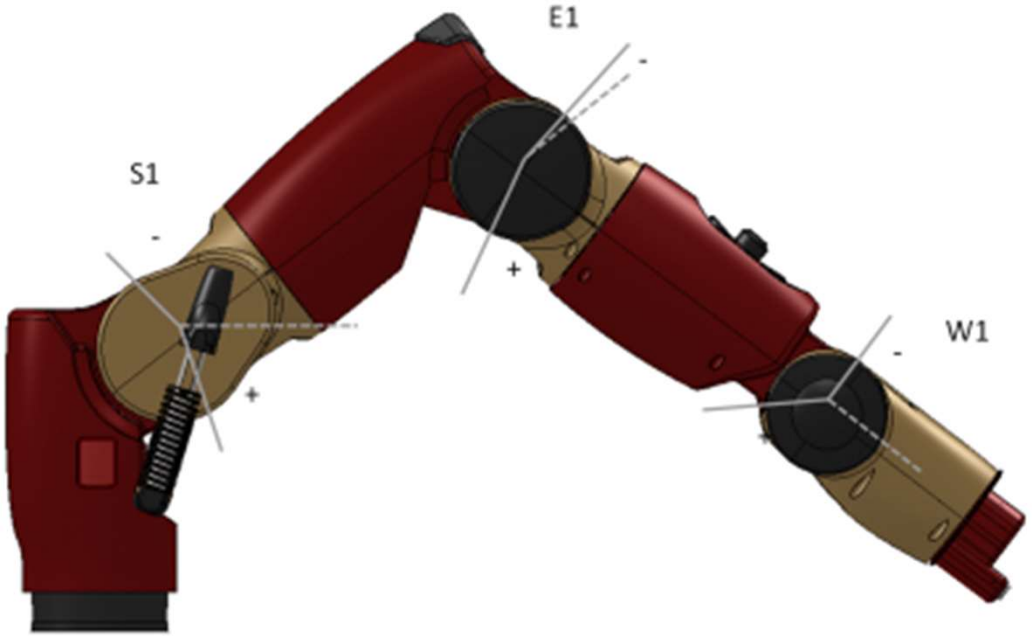
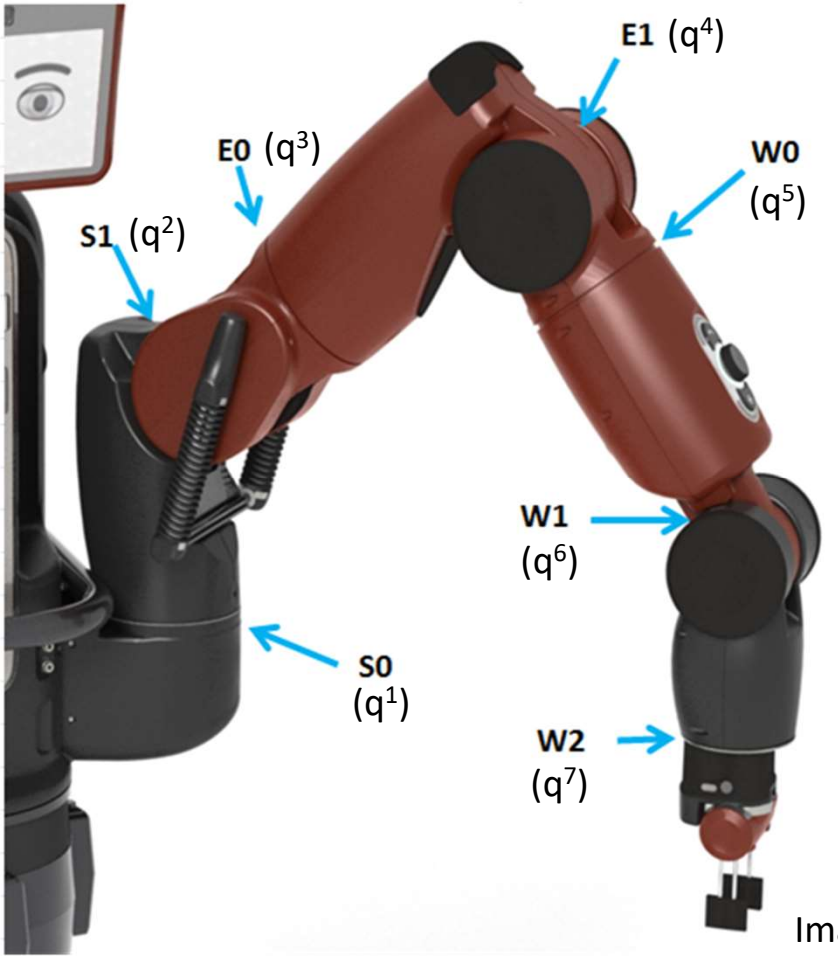


Image source: https://sdk.rethinkrobotics.com/wiki/Hardware_Specifications

Representing the Arm's Configuration

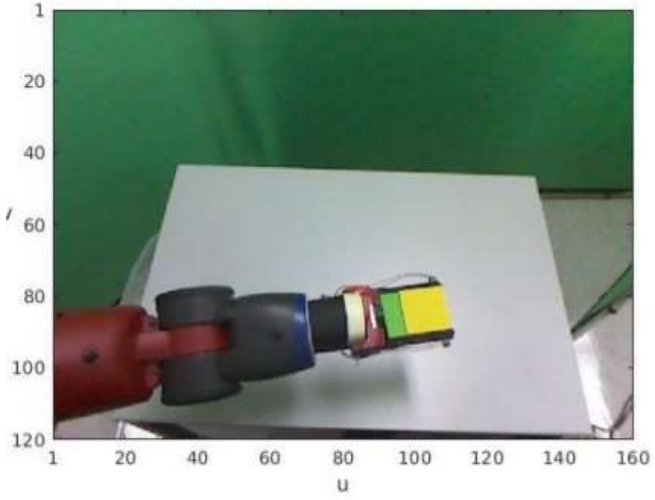


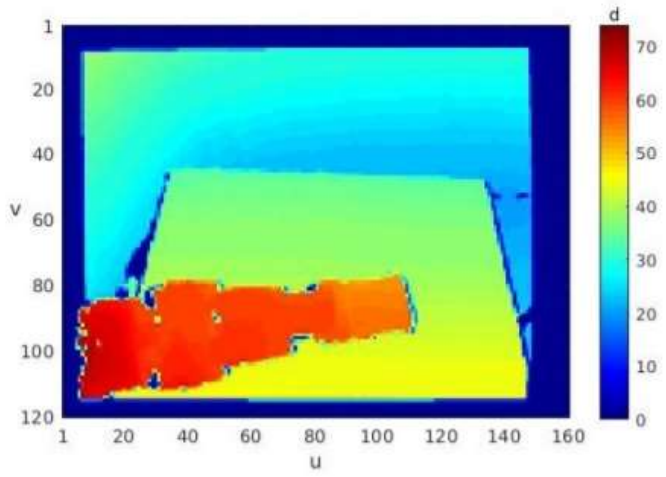
- The agent can sense each joint angle with proprioception.
- The configuration is given by an ordered vector of joint angles $q = \langle q^1, q^2, \cdots, q^7 \rangle$.
- The configuration stored in a node n_i will be denoted $q_i = \langle q_i^1, q_i^2, \cdots, q_i^7 \rangle$.

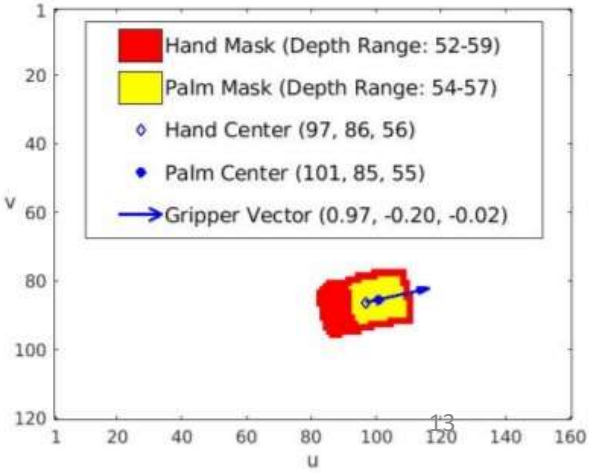
Image source: https://sdk.rethinkrobotics.com/wiki/Hardware_Specifications

The Agent's Vision

- The agent can observe the environment at any time.
- Images are captured with a single RGB-D camera with a fixed perspective.
  - Incorporate components of other models to consider changeable gaze
    - (Hulse et al., 2010), (Jamone et al., 2012, 2014)
- Simple representations for the hand and other objects can be identified.
  - Binary image masks, centers, depth ranges, and orientations







PPS Graph Nodes and Edges

- Nodes represent poses visited during motor babbling.
  - These poses are safe and in view.
  - Stored configurations allow precise return to the same pose.
  - Stored visual data is used to decide which pose to return to.

- Edges connect pairs of nodes with a feasible motion between them.
  - Motor babbling moves are feasible.
  - Moves shorter than the median length of motor babbling edges are assumed feasible.

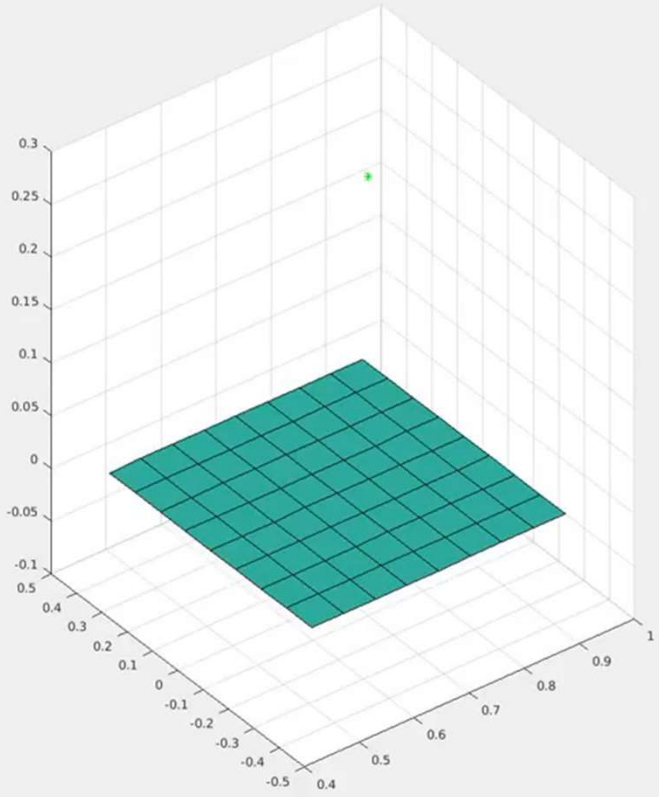
- Random exploration creates a PPS Graph of 3,000 nodes
  - 117,717 edges

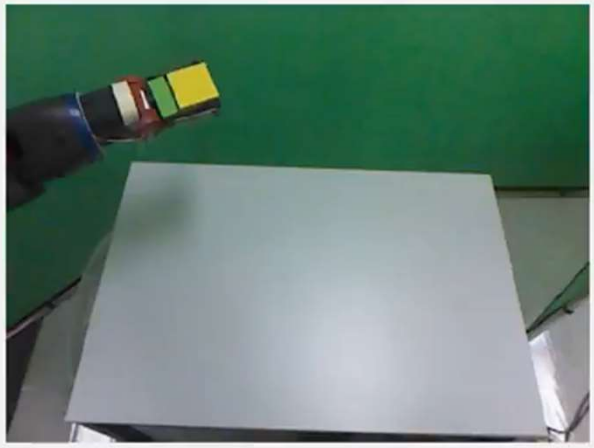
Probabilistic Roadmaps (PRMs)

- Our Peripersonal Space (PPS) Graph has a similar structure to a PRM.
  - A PRM represents the configuration space (Kavraki et al., 1996).
- The PPS Graph is inspired by the Visual Roadmap (VRM).
  - Nodes store the configuration and associated perceptual information.
  - Intersections between arm percepts and object percepts suggest collisions.
  - (Ramaiah et al., 2015)

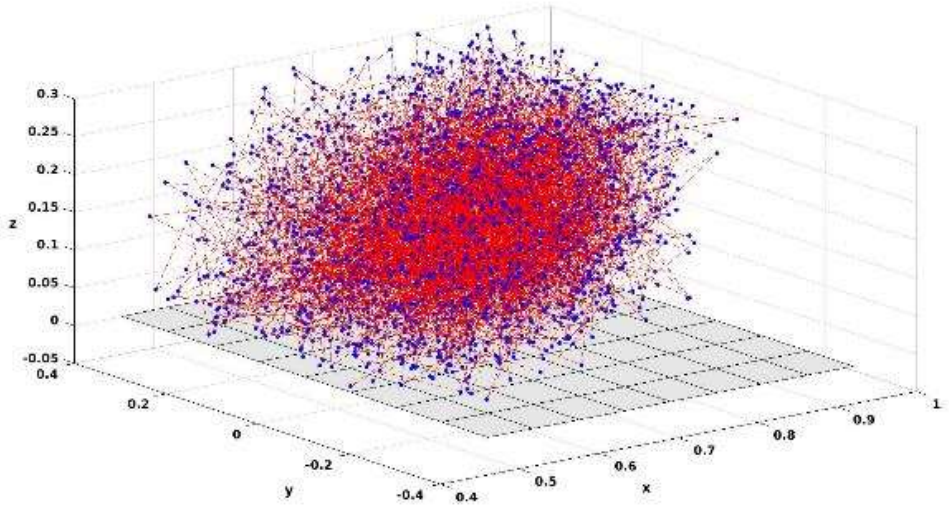
- The PPS Graph has unique features:
  - Constructed during naturally plausible motor babbling.
  - Construction requires only limited sensory information without interpretation.
  - The agent must learn to use the stored information.

Building the PPS Graph with Motor Babbling

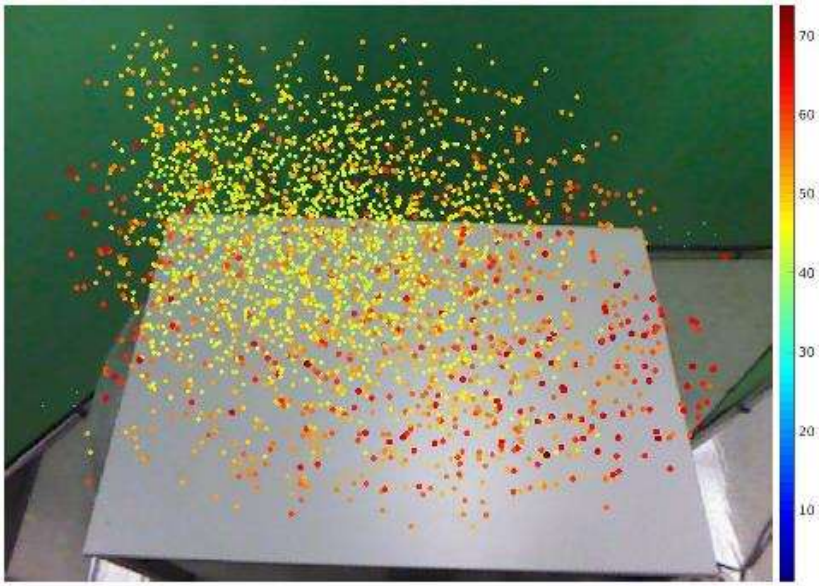




Visualizations of the PPS Graph after Motor Babbling is Completed



Nodes plotted at their (x,y,z) coordinate's in Baxter's default coordinate frame. For visualization only – not accessible to the agent.



Nodes plotted by their centers of mass in (u,v,d) image-space, as the agent sees them.

Learning to Reach

Observing the Unusual Bump Event

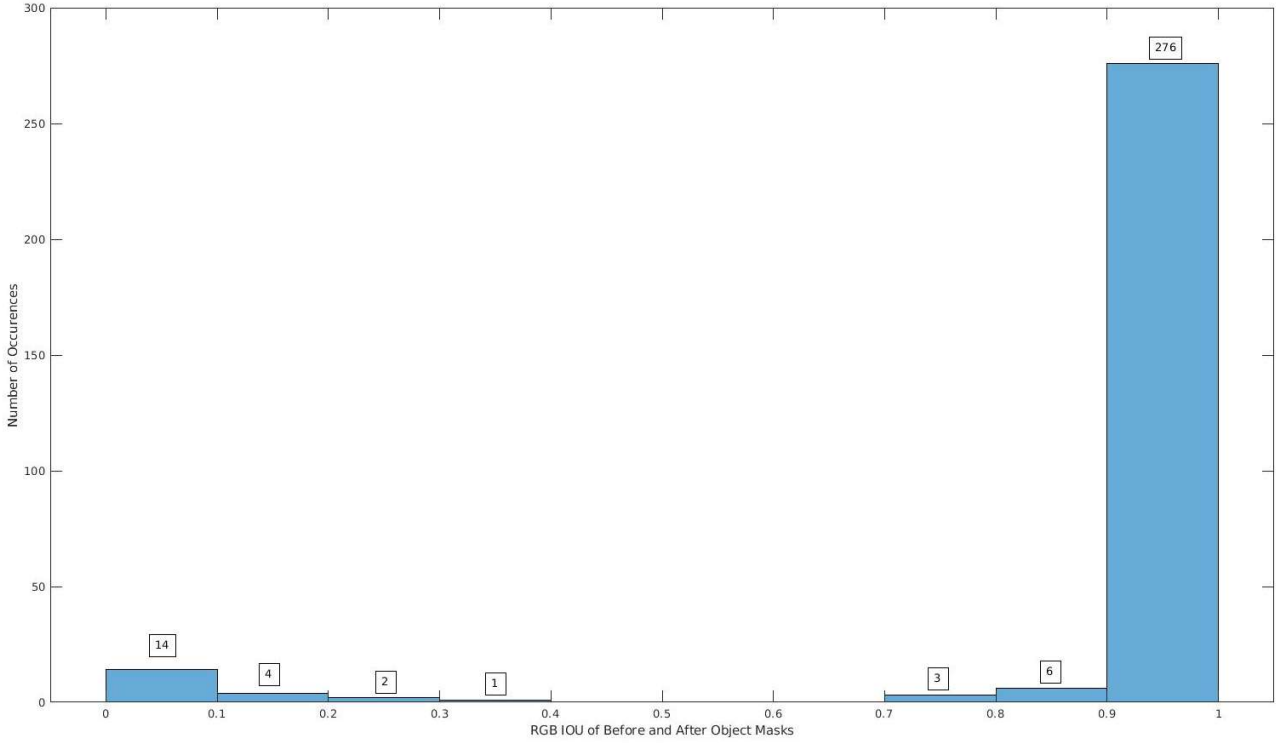
- The agent performs move trajectories to random nodes.
  - The agent compares the “before” and “after” appearances of the object.
  - The typical result is no change to the object’s appearance.
  - In rare cases, the motion “bumps” the object and causes a significant change.

Move trajectory:

Return trajectory:

Differentiating Bumps from the Typical Result

- $\mathrm{IOU}(\mathrm{A}, \mathrm{B}) = \frac{|\mathrm{A} \cap \mathrm{B}|}{|\mathrm{A} \cup \mathrm{B}|}$
- Clear typical and atypical clusters emerge
- IOU ≈ 1 → No change
- IOU << 1 → Bump



Examples of No Change and Bump Results

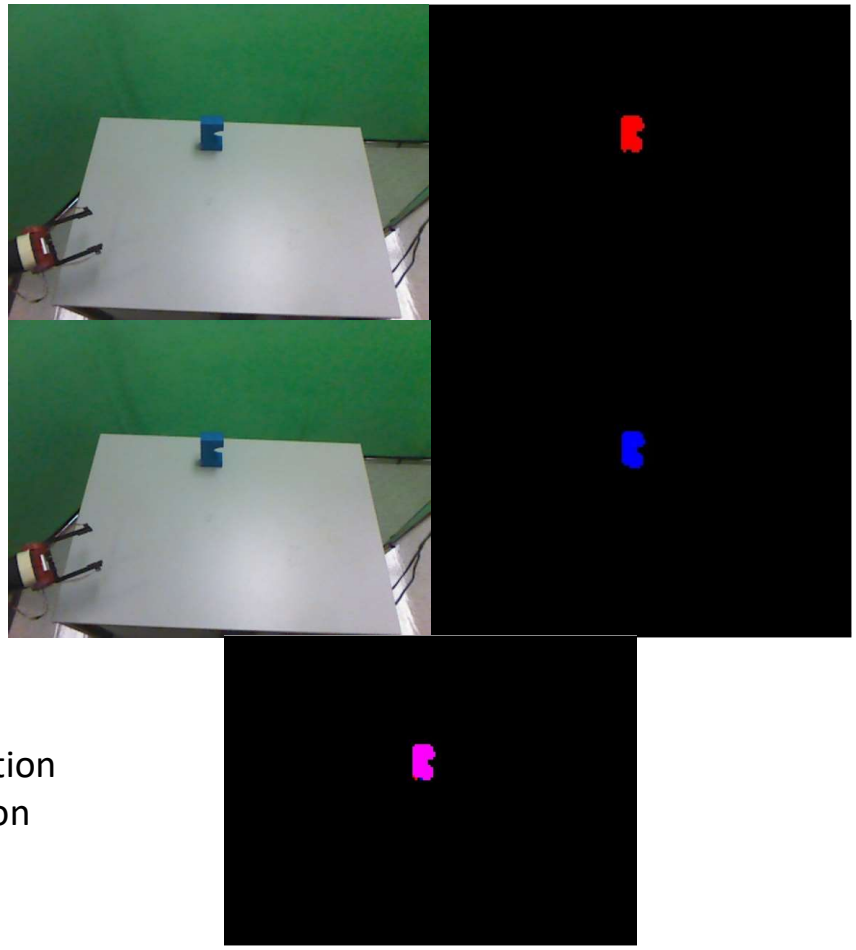
No Change (Typical, IOU ≈ 1)

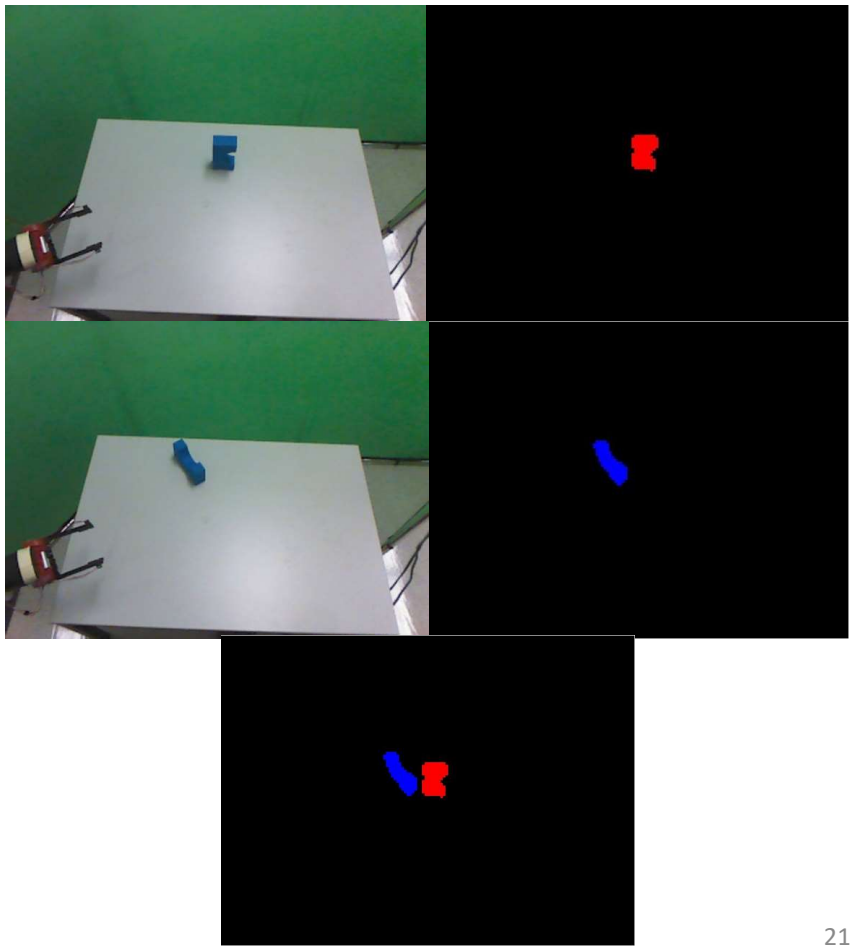
Bump (Unusual, IOU << 1)

Before

After

Intersection and Union





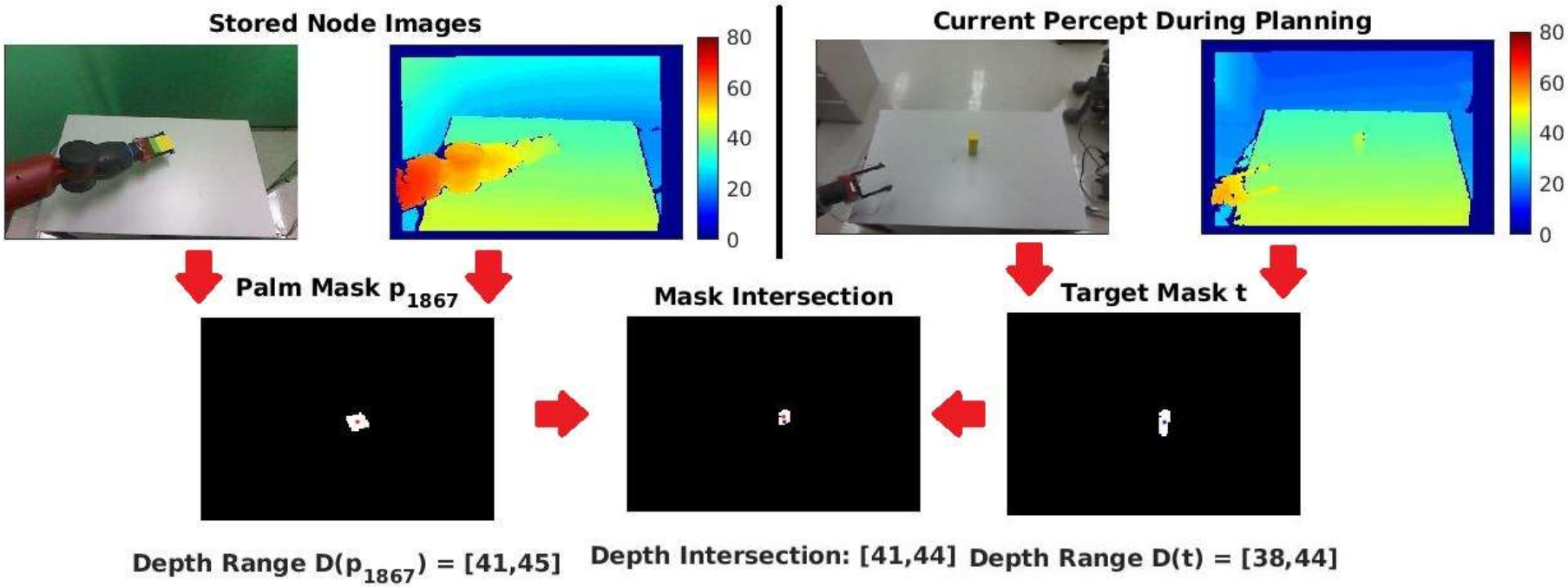
Planning a Reach Step 1:
Observe the Target with the Hand at the Home Node

The target object's mask and depth range are extracted from the current visual percepts.

Planning a Reach Step 2: Determine the Set of Final Node Candidates

Only 3 of 3000 nodes are reliable candidates when the target is placed here.

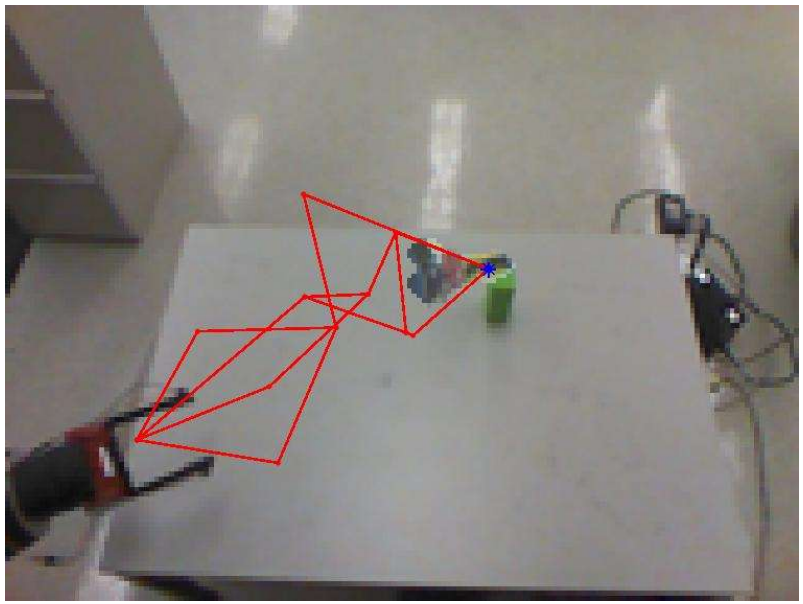
Determining the Candidacy of a Node using Intersection Features

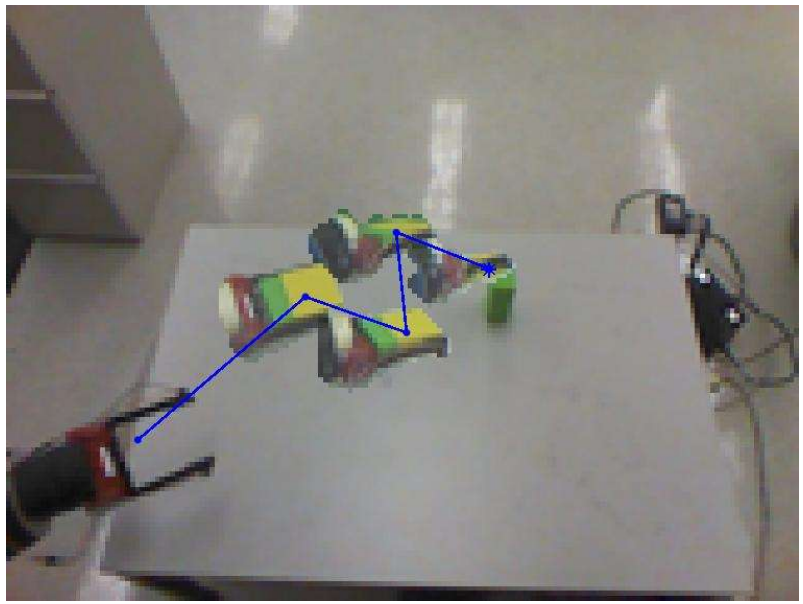


This node has the most reliable type of intersection features. Both its mask and depth range intersect with the target's. It will be included in the set of final node candidates.

Planning a Reach Step 3: Find the Shortest Graph Path Trajectory to a Final Node Randomly Chosen from the Candidates

Numerous paths exist between the home node and the selected final node. The shortest in configuration space distance will be used to perform the reach.





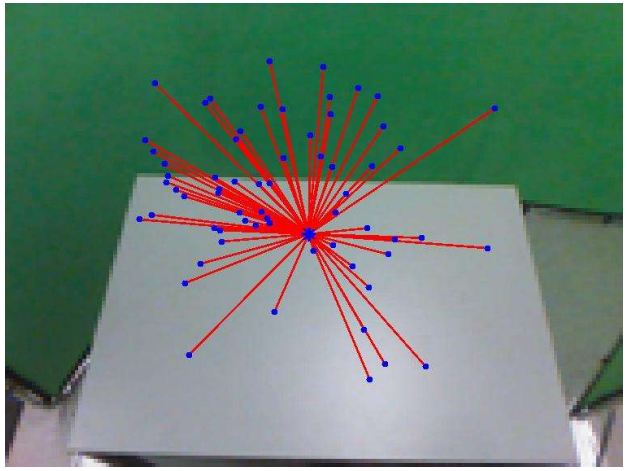
52.5% of reaches planned in this way successfully bumped the object.

Why Aren't All Reaches to these Final Node Candidates Successful?

- Object representations overestimate the space they occupy.
- This leads to some intersections that imply false positive collisions.
- Changes caused by small collisions may not be observable.
- Reaches should end with the candidate node closest to the target.
  - 77.5% reliable
- Can the agent reach closer than the nearest node?

Estimating a Local Jacobian for the Neighborhood of a PPS Graph Node

- Each neighbor n_j of a node n_i provides an example of:
  - A change in configuration space $\Delta q = q_j - q_i$
  - And the resulting change of the hand's center position in image space $\Delta c = c_j - c_i$
- The changes for all neighbors can be combined into matrices ΔQ and ΔC.



- The local Jacobian estimate $\hat{J}(n_i)$ is the least squares solution of

$$\Delta Q \, \hat{J}(n_i) = \Delta C$$

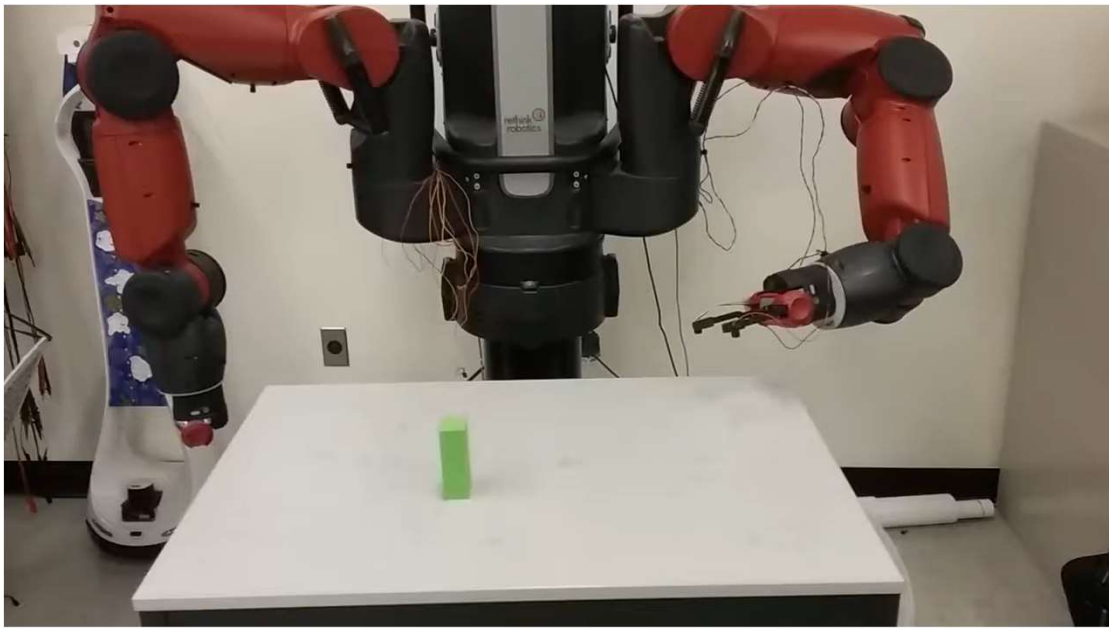
Using Local Jacobian Estimates to Extend the PPS Graph to a Locally Continuous Representation

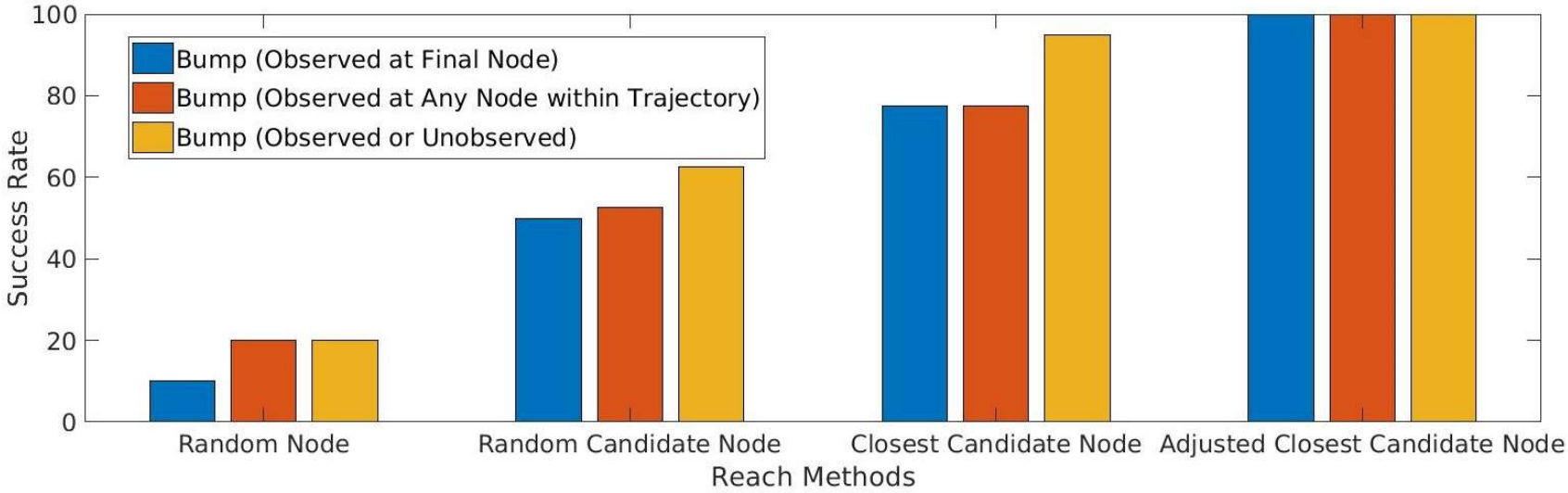
- $\Delta q\, \hat{J}(n_i) = \Delta c$ is a linear estimate and most accurate near the node n_i.
  - $\hat{J}(n_i)$ can be used to predict the change Δc caused by a change Δq from q_i.
  - $\hat{J}^{+}(n_i)$ can be used to estimate the Δq required to produce a desired Δc.

- When reaching, the final node can be adjusted to match the target's center.
  - The stored node image and current image of the target provide $\Delta c = c_{target} - c_i$.
  - The node stores its original arm configuration q_i.
  - The agent estimates the necessary change to the configuration $\Delta q = \Delta c\, \hat{J}^{+}(n_i)$.
  - The agent moves to $q_i + \Delta q$ instead of q_i during the trajectory.
    - The hand's center is expected to be zero distance from the target's at this configuration.

100% Reliable Reaching

The agent is now capable of causing observable (large) bumps with each reach attempt. These bump always occurred at the intended final node.





Our Model of the Learned Reach Action Matches Infant Behavior

- Reach trajectories are sequences of jerky submotions.
  - (von Hofsten, 1991), (Thelen et al., 1993), (Berthier, 2011)

- Reaches do not require current vision of the hand.
  - Before Clifton et al. (1993) the consensus was that infant reaches use visual servoing (Piaget, 1952).
  - Clifton et al. studied infant reaches without current vision of the hand.
    - Reaches for lit objects in the dark are equally successful (Clifton et al., 1993).
    - Reaches for lit objects in the dark use the same kinematics (Clifton et al., 1994).
  - In our model, reaches are planned using current vision of the object but only stored images of the hand.

Learning to Grasp

The Challenge of Observing Accidental Grasps

- The usual result of a reach is a bump (100% reliable).
  - The unusual result is an accidental grasp.

- An accidental grasp requires two low-probability events:
  - The grippers must surround the object during the trajectory.
  - The grippers must be closed at the correct time to firmly grip the object.

- The agent must observe multiple grasp examples to begin learning.
  - The co-occurrence of these events by chance is prohibitively rare.

The Palmar Reflex

- In human infants, the Palmar reflex causes the hand to close around an object pressed into the palm (Futagi et al., 2012).
- We simulate the Palmar reflex with a break-beam sensor.
- Closing the grippers is caused reflexively instead of by chance.
- Observing accidental grasps is now tractable.

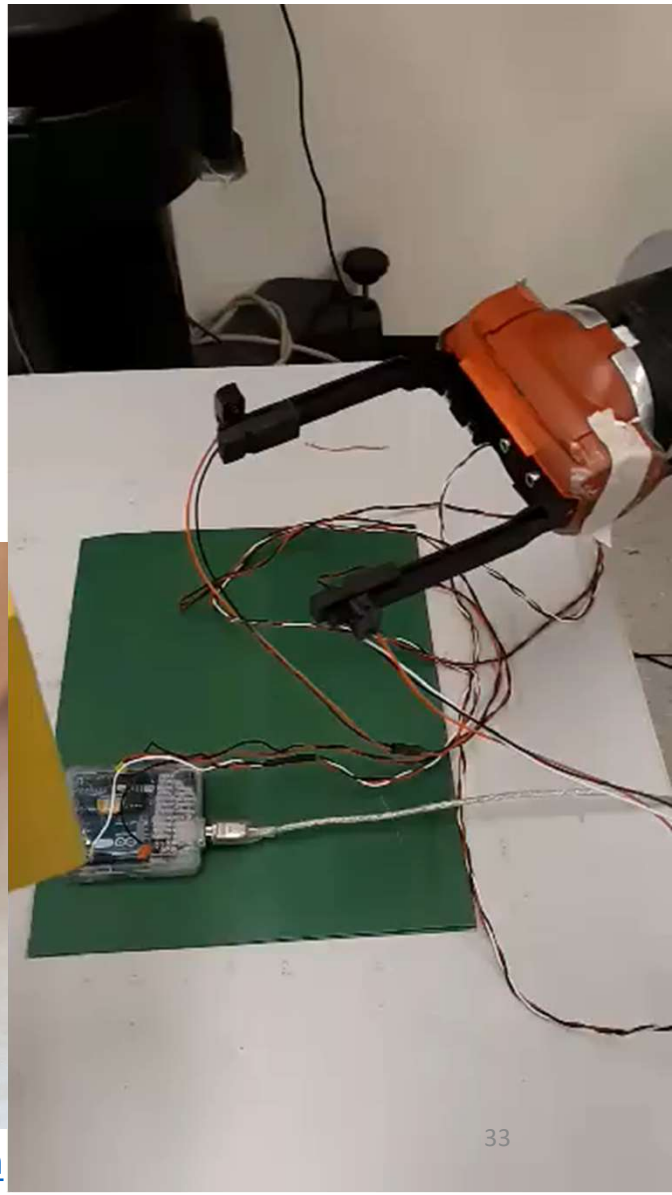
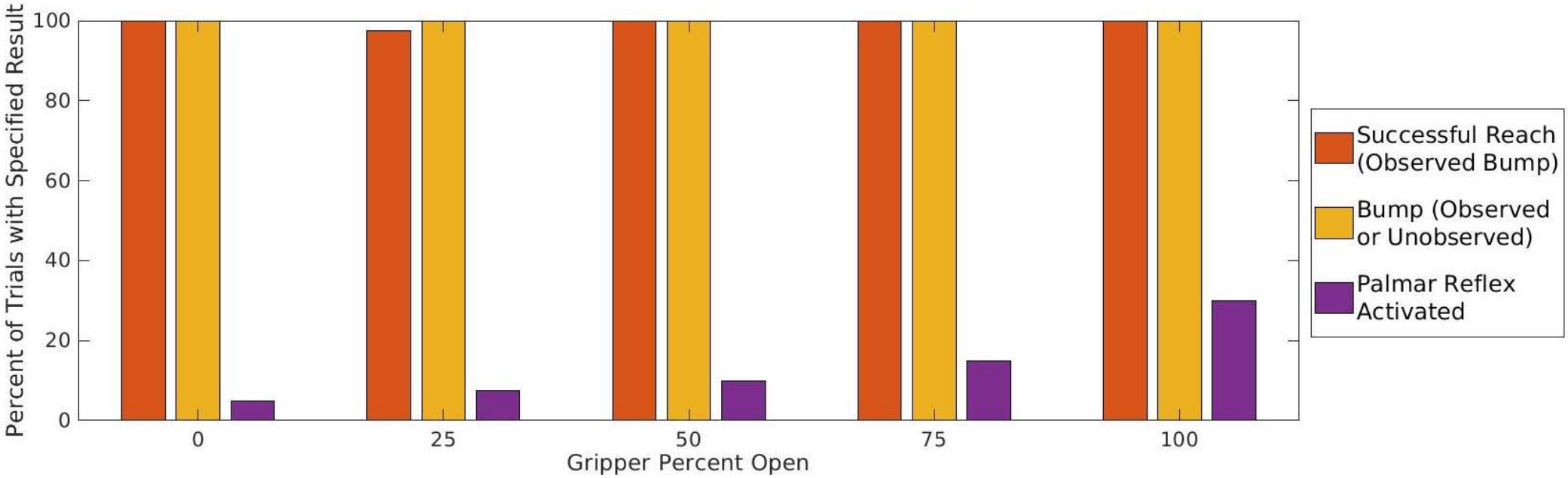


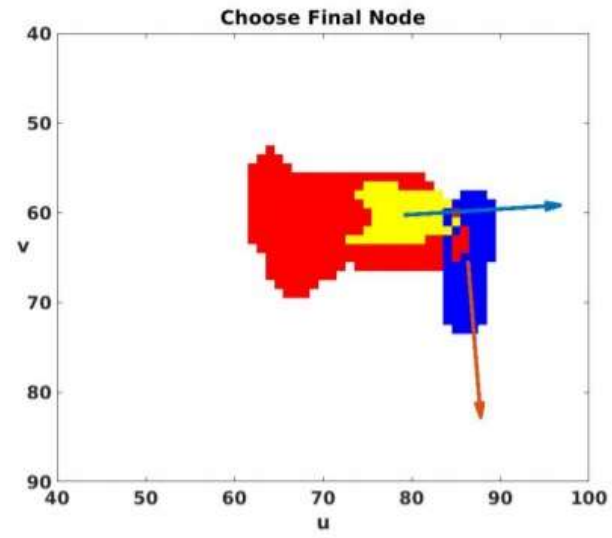
Image source: https://nashvilleparent.com/8-infant-reflexes-baby-s-born-with

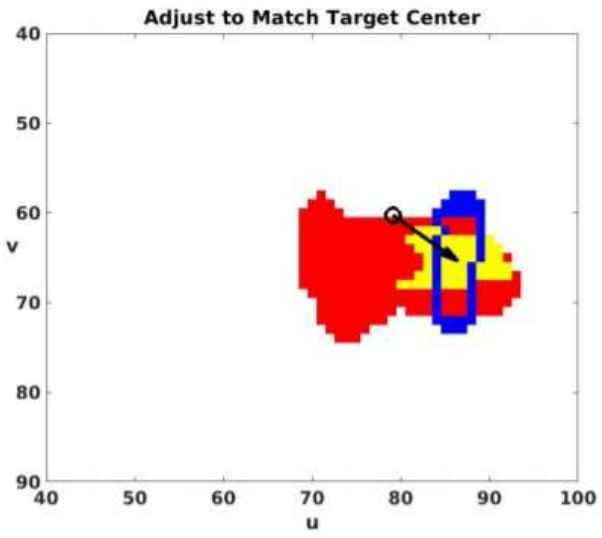
The Gripper should be Fully Open during the Grasp Approach

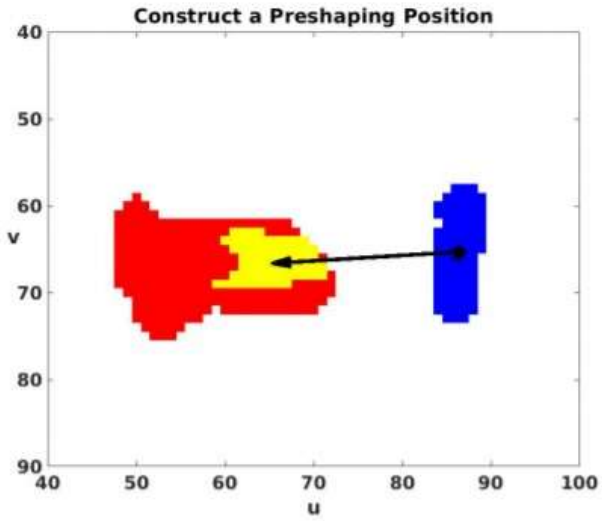
- Open grippers allow the Palmar reflex to be activated more often.



The Approach and Hand Orientation should be Perpendicular to the Target Major Axis







The agent uses two local Jacobian adjustments to create a final pose and a preshaping pose based on the most perpendicular final node candidate.

Additional Improvements to Grasp Reliability

- The w2 joint should be set to orient the hand to surround the object.
- The displacement of the final node from the target should be well-aligned with the orientation of the hand before adjustment.

Grasp Method (All with open grippers and Palmar reflex)	Reliability
Move to a random node	2.5%
Reach to a random final node candidate	2.5%
Reach to the closest final node candidate	5.0%
Reach to adjusted closest candidate	12.5%
Approach vector perpendicular to target major axis	35.0%
… and orient the hand with the wrist	50.0%
… and select a final node with aligned displacement and orientation vectors	57.5%

Why Isn’t Grasping Fully Reliable?

- The structure of the hand is not conducive to grasping.
  - The parallel grippers require very specific trajectories.
  - The inflexible hand cannot be reconfigured to improve a failing grasp.
- The agent may need improved visual perception and representations.
  - The agent relies on distorted image space coordinates and object hulls.
  - Workspace representations or more detailed object models may be required.
- Smooth and precise motions of late action learning may be necessary.

Learning to Ungrasp

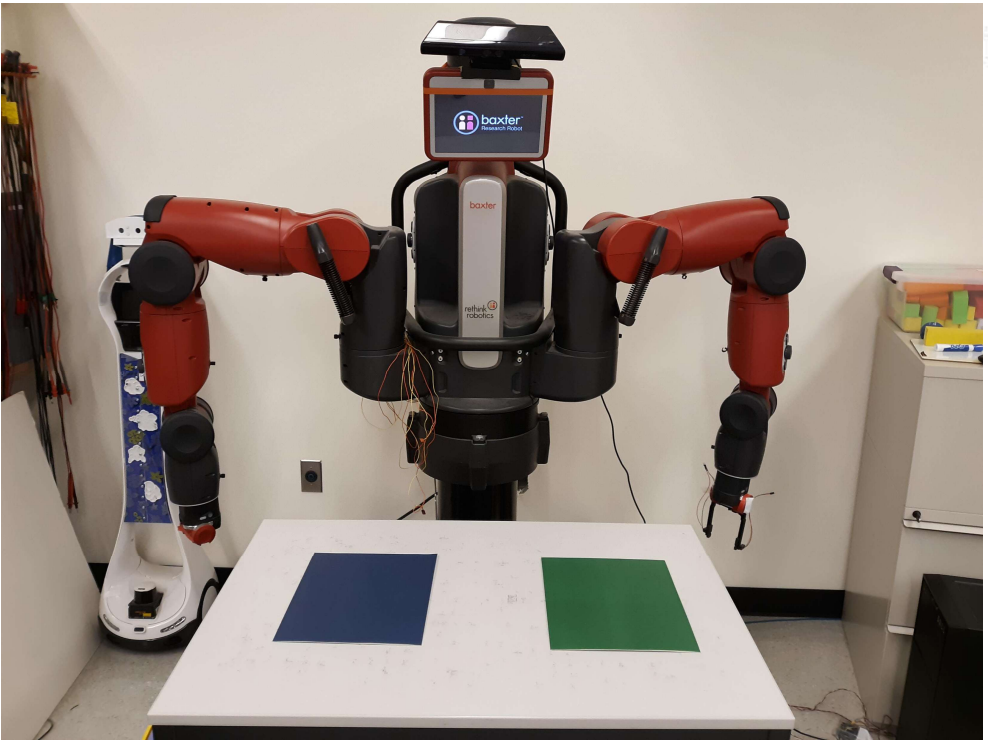
- With an object grasped, an attempt to…
  - Rotate any arm joint maintains the grasp
  - Decrease or slightly increase the gripper aperture maintains the grasp
  - Significantly increase the gripper aperture ends the grasp

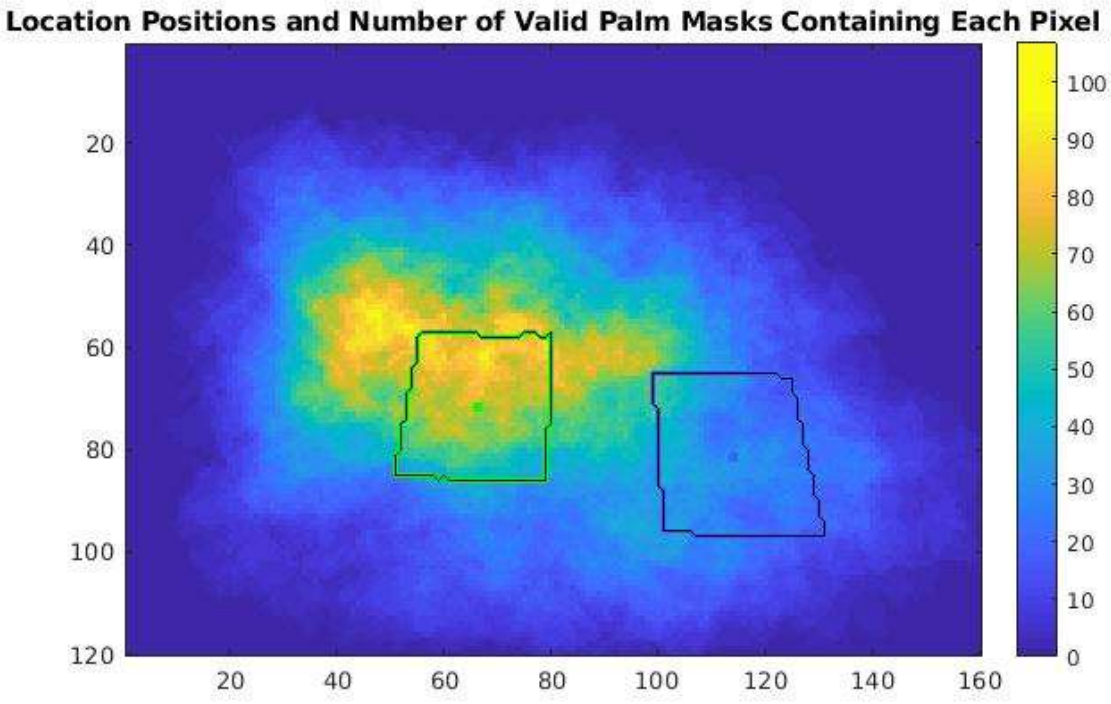
- An ungrasp is an unusual event

- The agent learns that fully opening the grippers is 100% reliable

Learning to Place

Representing Qualitative Locations

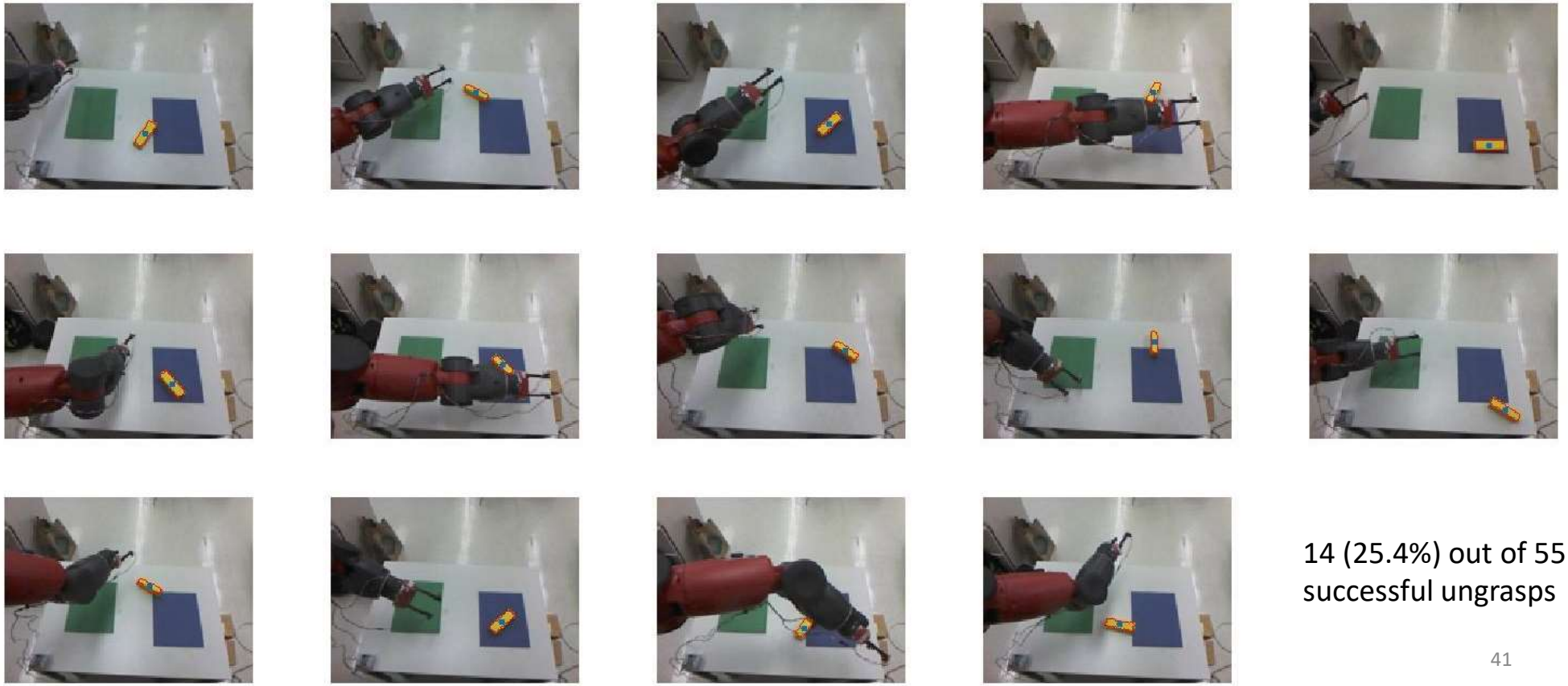




The agent defines two locations using visually distinct colored patches

Ungrasps that Placed into the Blue Location by Chance

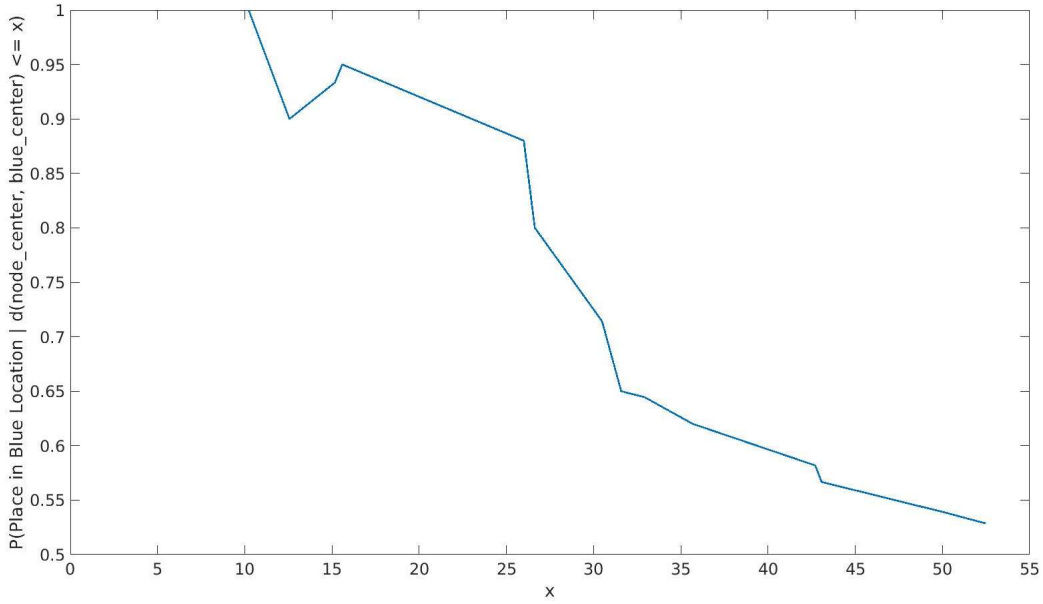
Ungrasps with a Blue Location Result



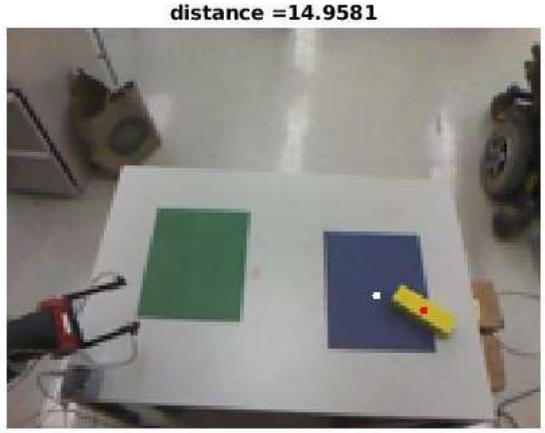
14 (25.4%) out of 55 successful ungrasps

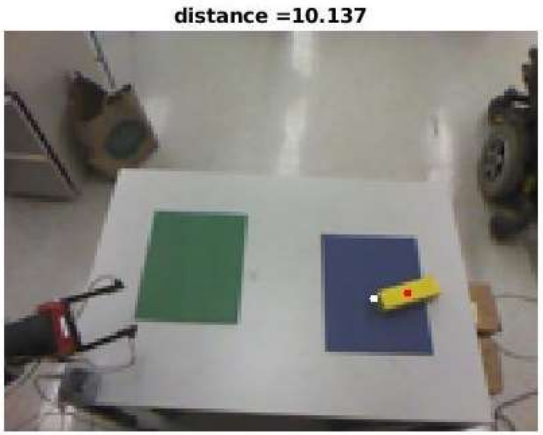
Learning that Shorter Distances are More Reliable for Placing

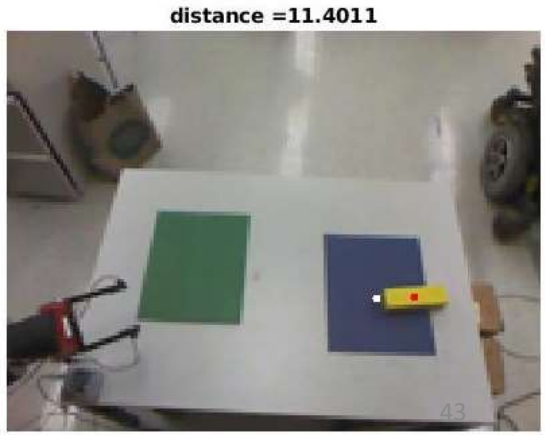
- The agent repeats the ungrasps that placed into the blue location.
- Ungrasps performed at more distant nodes are less likely to be successful placements.
- The agent learns to move to the node closest to the location before ungrasping.



Placing into the Blue Location from the Closest Node is 100% Reliable

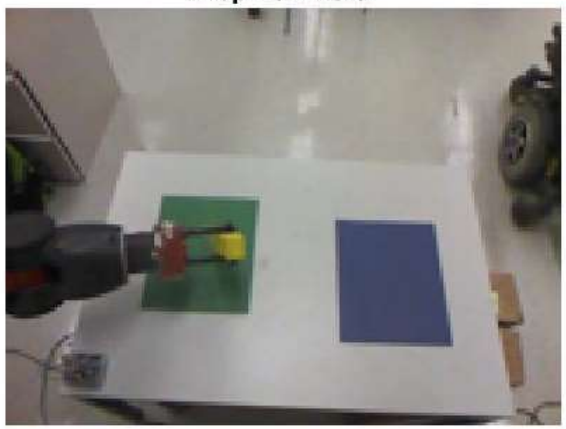




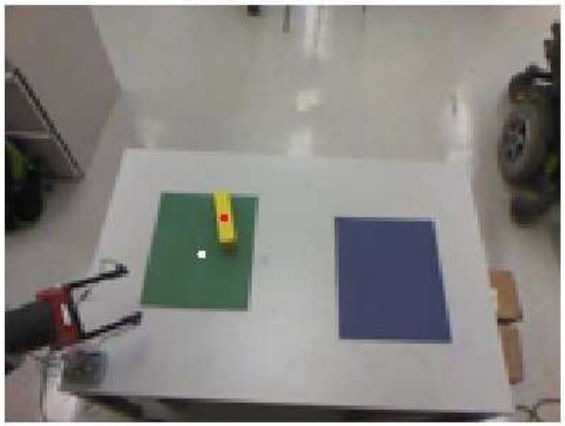


This Policy is also 100% Reliable for the Green Location

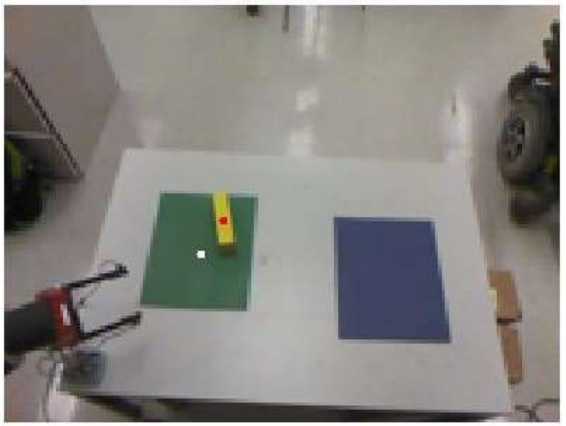
Drop from here:



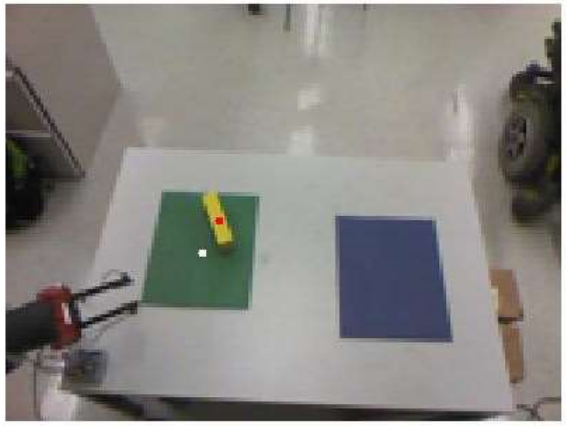
distance =12.248



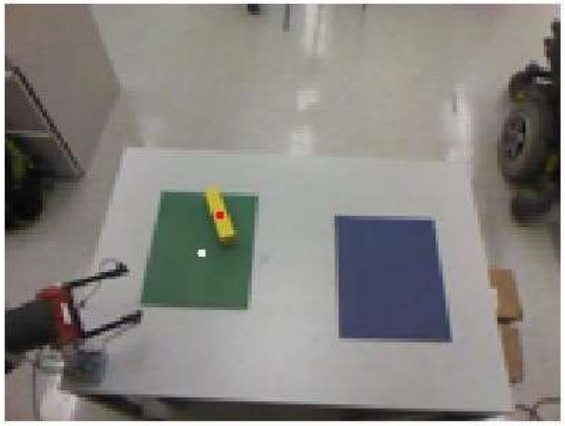
distance =11.7744



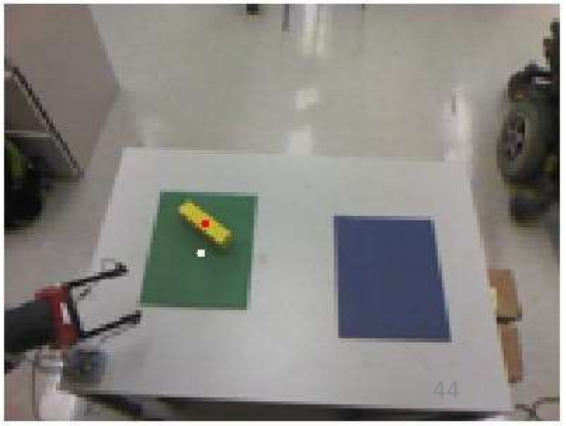
distance =10.4022



distance =12.0147



distance =8.6836

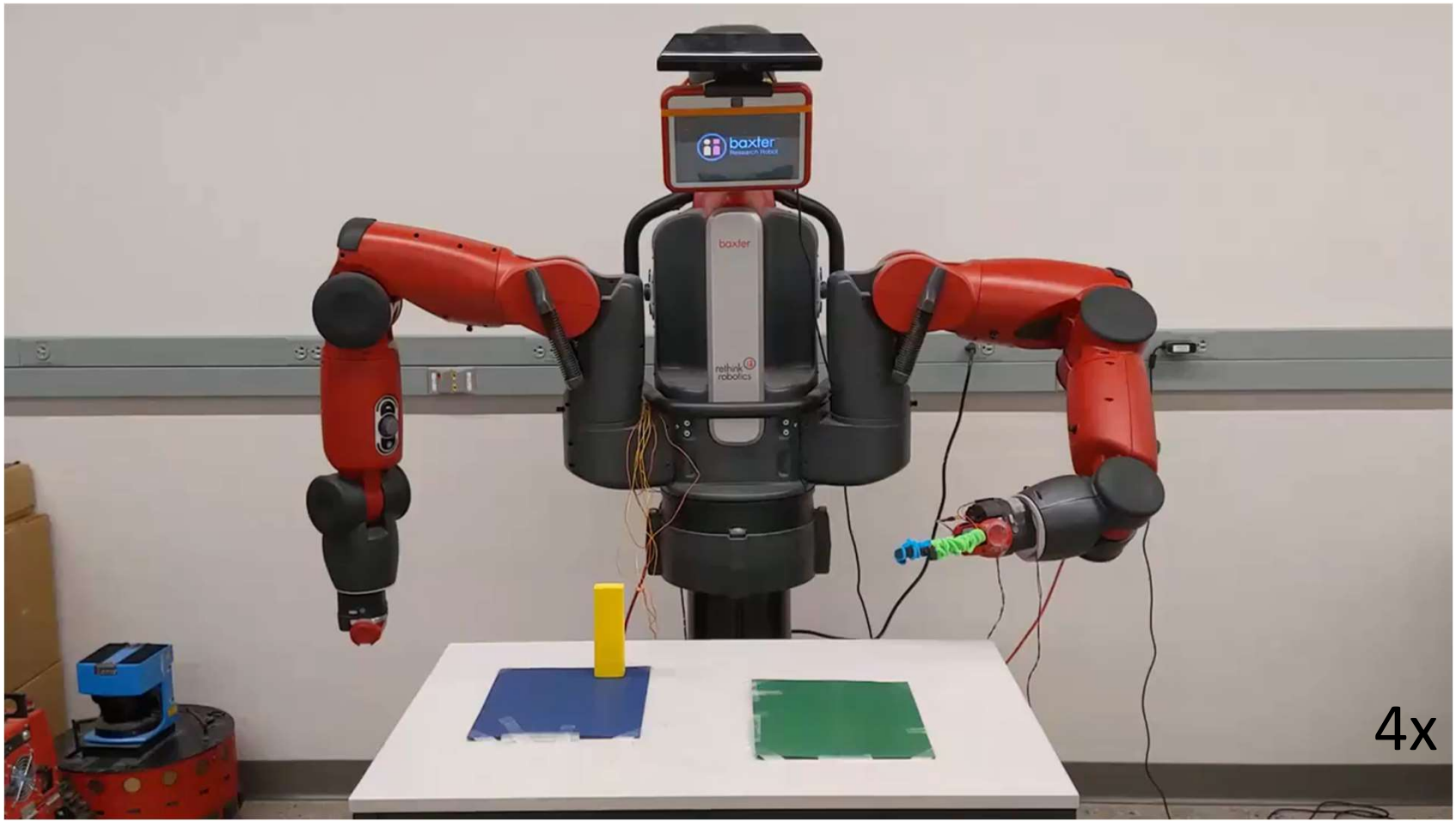


Learning to Pick-and-Place

Performing the Pick-and-Place Sequence

- Grasping a block satisfies the preconditions of the place action.
- The agent attempts to move the block from one location to the other.
- The pick-and-place action is 52.5% reliable.
- Reliability is determined by the reliability of grasping.

Potential for Learning through Repetitive Play (Piaget, 1977), (Schmidhuber, 2011)



Conclusion

- We present the Peripersonal Space (PPS) Graph model.
  - Minimal required sensory capabilities and prior knowledge
  - Allows safe motion throughout PPS that provides experience

- We demonstrate successful manipulation action learning.
  - All actions are defined by the agent based on observed unusual events.
  - Moving, reaching, ungrasping, and placing become fully reliable.
  - Grasping and pick-and-place become semi-reliable.
  - Our model predicts behaviors matching surprising observations of infants.
  - These early phase actions provide enriched experience that could make late phase reinforcement learning more efficient.

Thank you!

- Questions?