Understanding Tools: Task-Oriented Object Modeling, Learning and Recognition

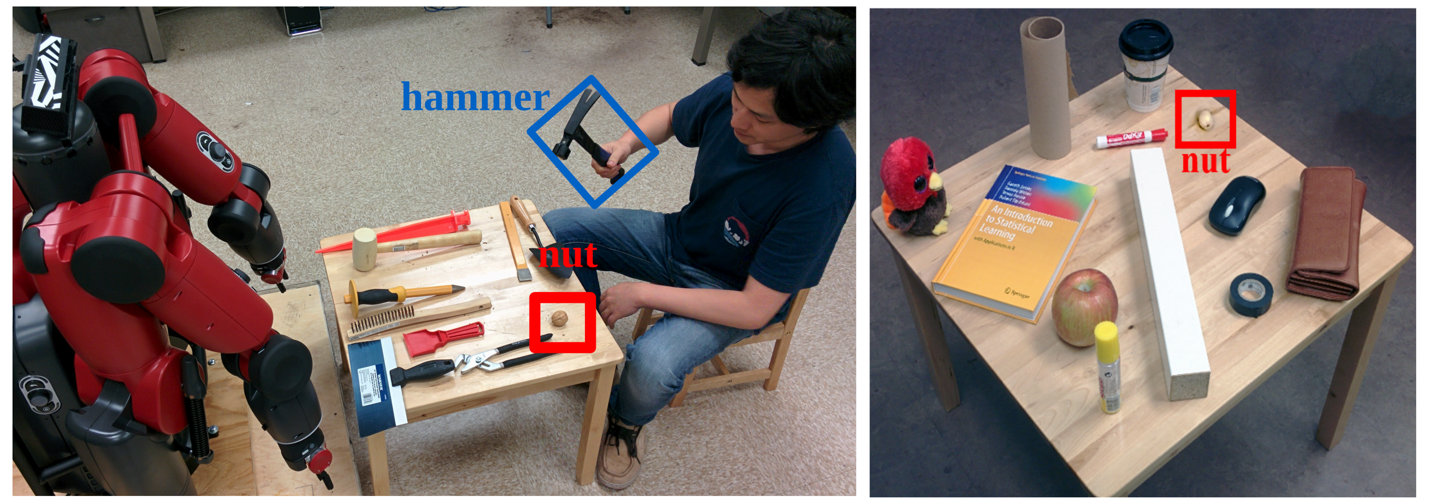
Yixin Zhu∗, Yibiao Zhao∗, Song-Chun Zhu (∗ equal contribution)

Center for Vision, Cognition, Learning, and Art (VCLA), University of California, Los Angeles (UCLA)

CVPR 2015 BOSTON JUNE 7-12

Motivation

In this paper, we rethink object recognition from the perspective of an agent: how objects are used as “tools” in actions to accomplish a “task”.

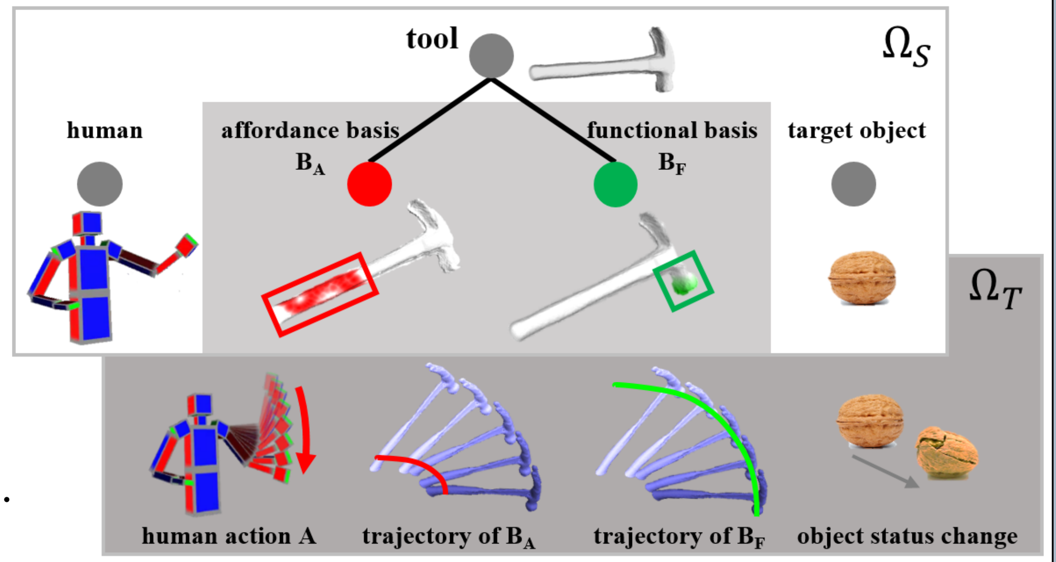


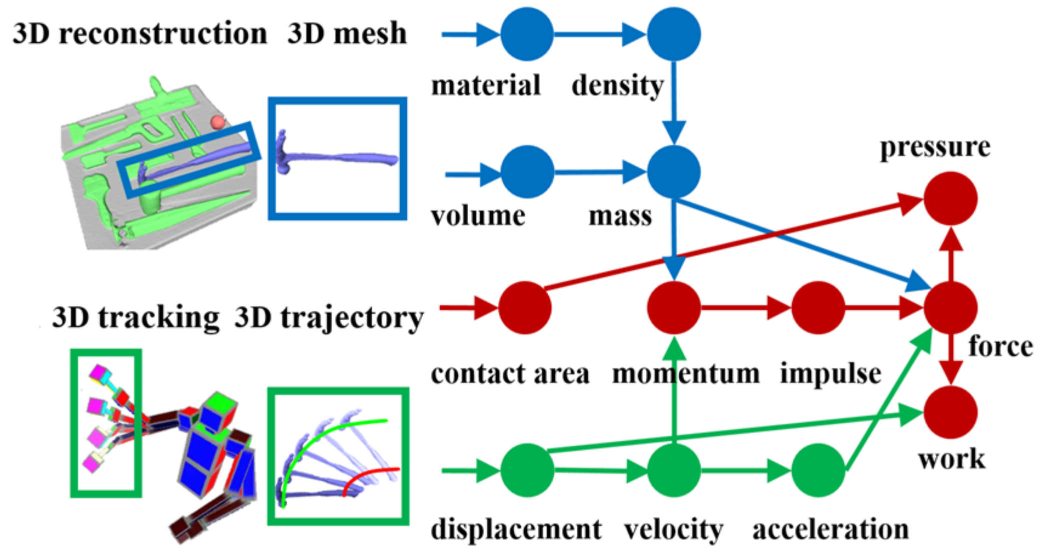
Learning from a human demonstration | Inference in a novel situation

The task-oriented representation of tool and tool-use

A spatial space represents object decomposition and 3D relations with human pose.

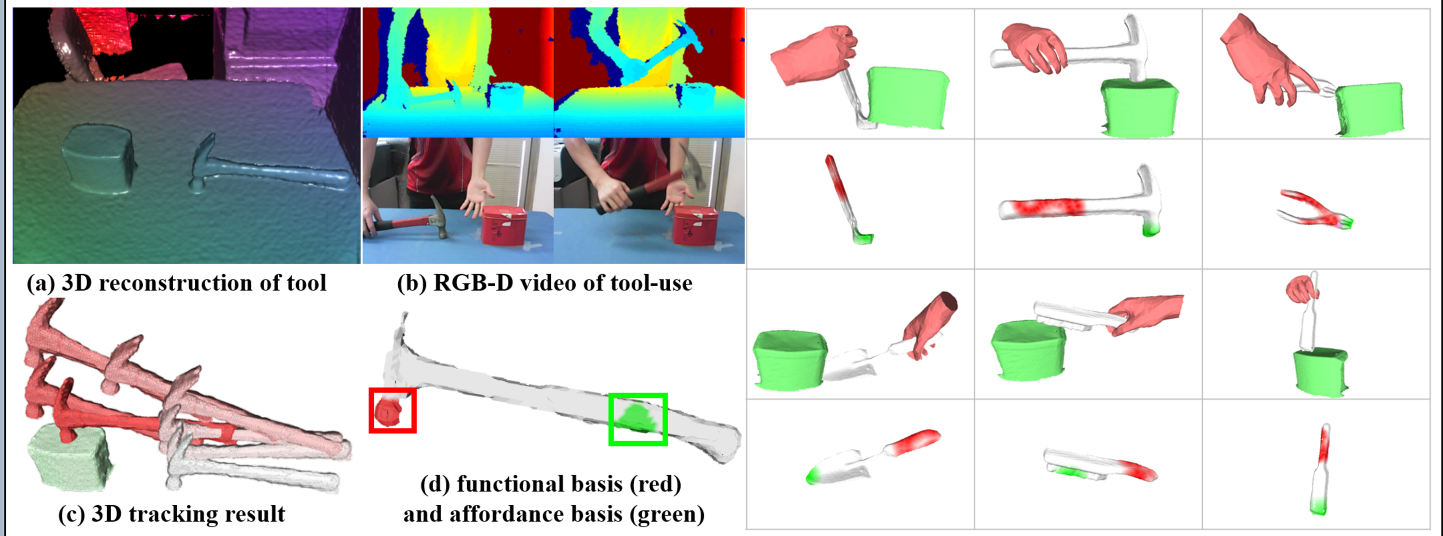
A temporal space represents the pose sequence in actions.



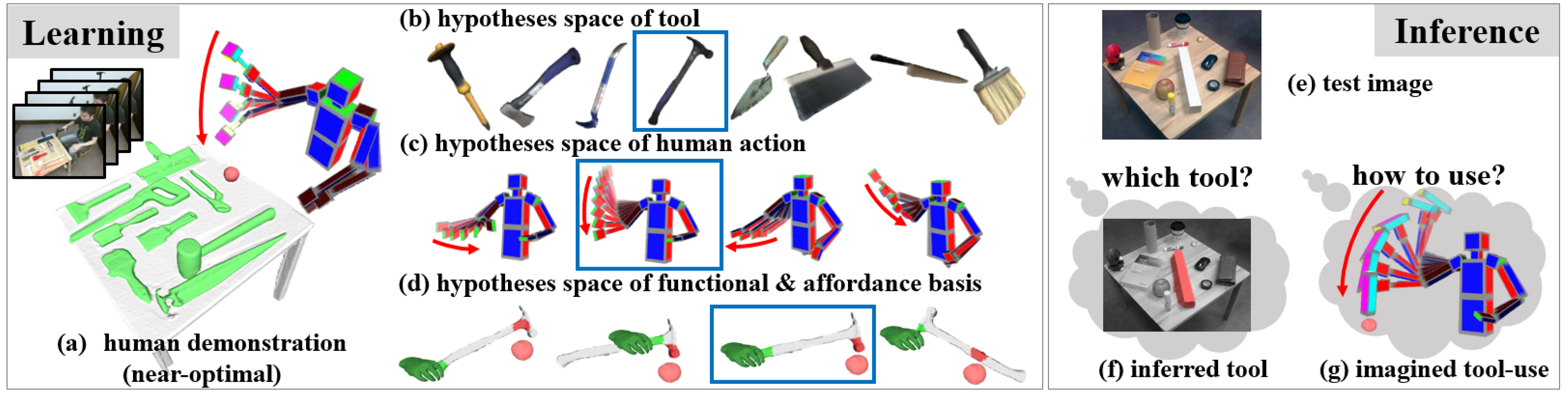


A causal space represents physical quantities estimated from 3D tool mesh, tool-use trajectory or jointly.

Parsing human demonstration



An illustration of learning and inference framework



Problem definition

Rational human choice assumption: we assume human chooses the optimal tool and tool-use based on the essential physical concept.

We formulate the tool recognition as a ranking problem, so that learned tool model $\omega \cdot \phi(pg)$ satisfies the maximum number of constraints:

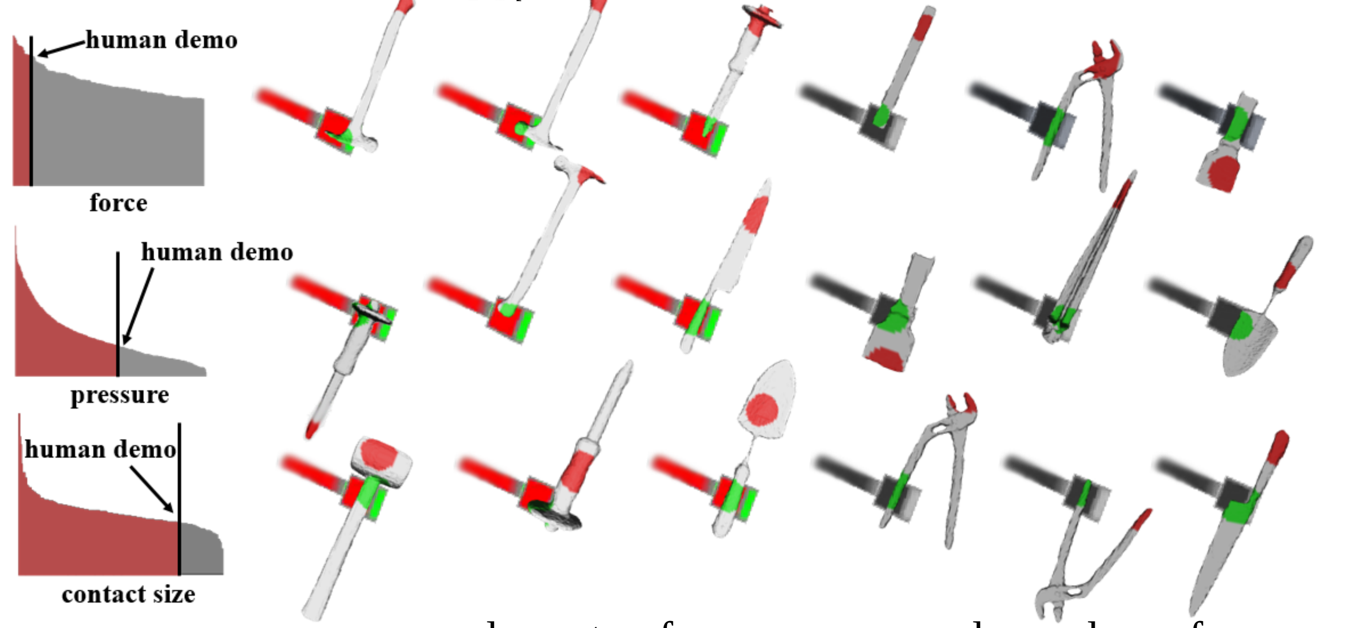
$$\min \quad \frac{1}{2}\,\omega \cdot \omega + \lambda \sum_i^n \xi_i^2$$

$$\text{s.t.} \quad \forall i \in \{1, \cdots, n\}:$$

$$\omega \cdot \phi(pg^*) - \omega \cdot \phi(pg_i) > 1 - \xi_i^2$$

$$\xi_i \geq 0,$$

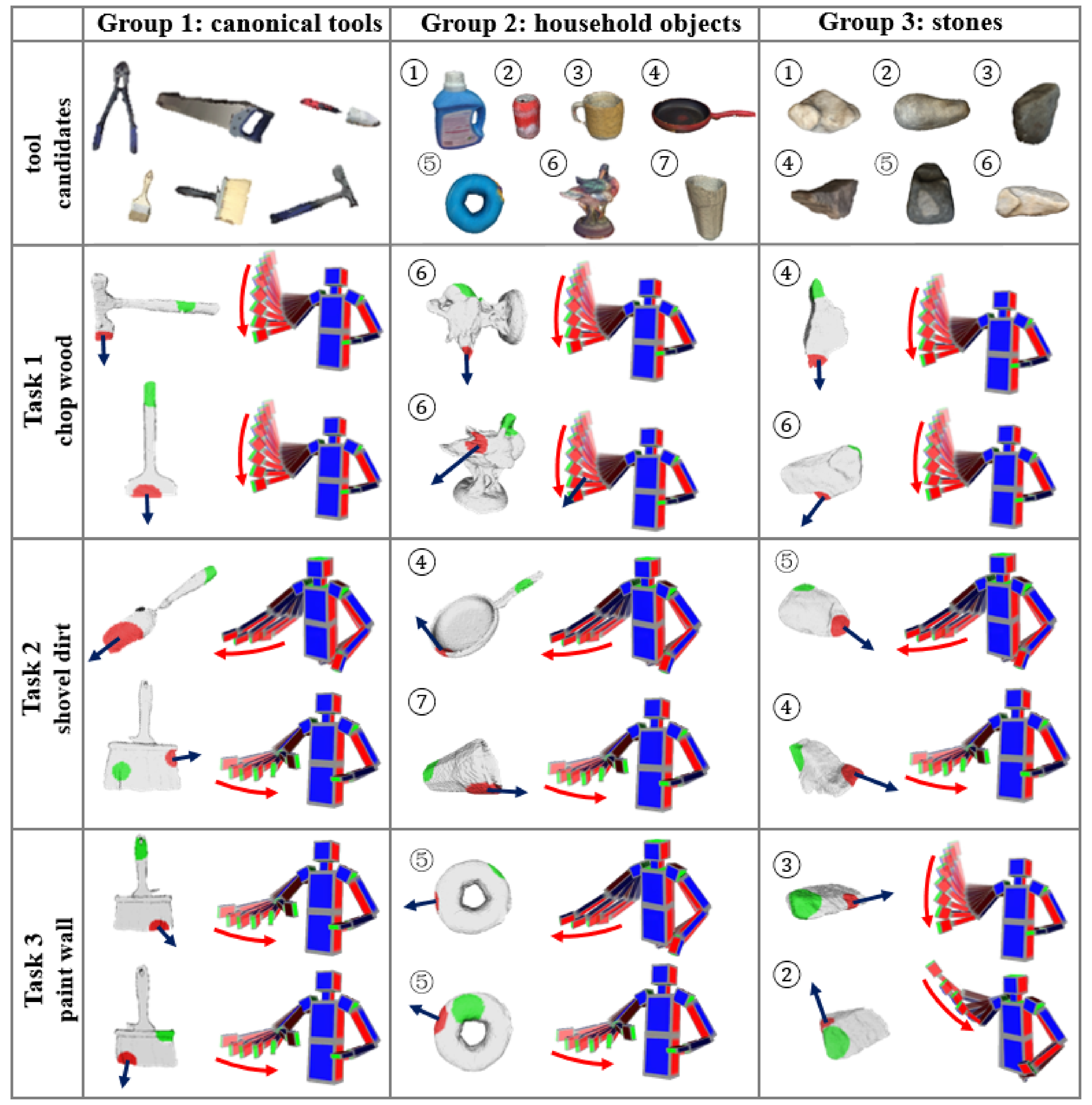
The human demonstration pg^* has the highest ranking score compared with the other tools and tool-uses pg_i.



examples outperform human demonstration | examples underperform human demonstration

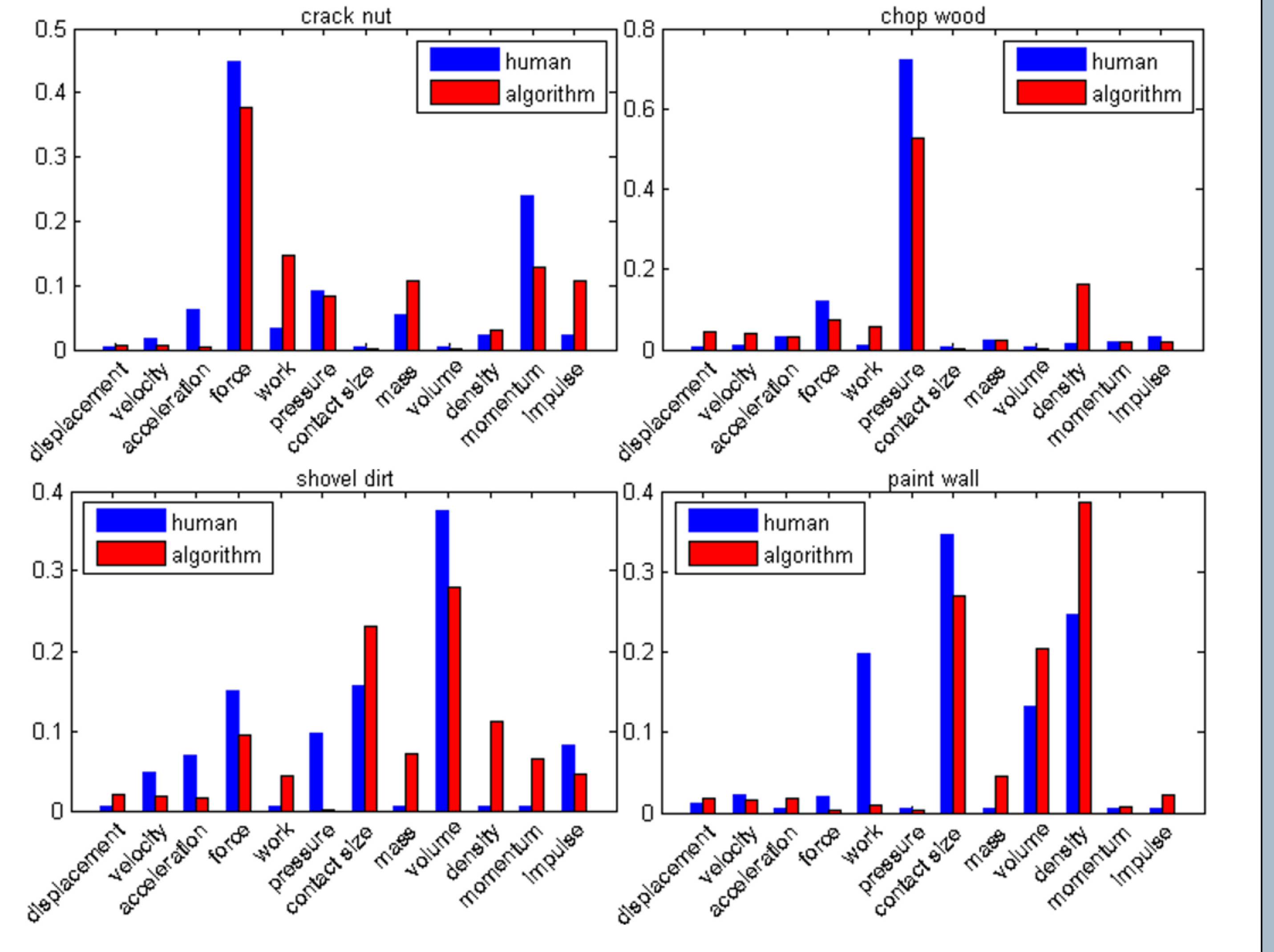
Qualitative results

Given three tasks: chop wood, shovel dirt, and paint wall, our algorithm picks and ranks objects for each task among objects in three groups: 1) conventional tools, 2) household objects, and 3) stones, and output the imagined tool-use: affordance basis, functional basis, and the imagined action pose sequence.

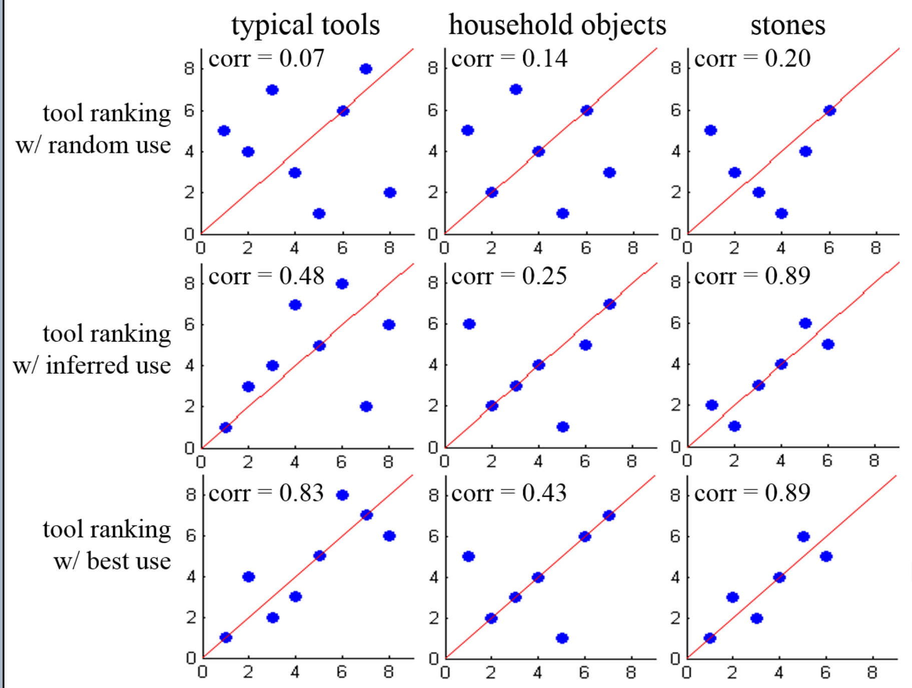


Quantitative results

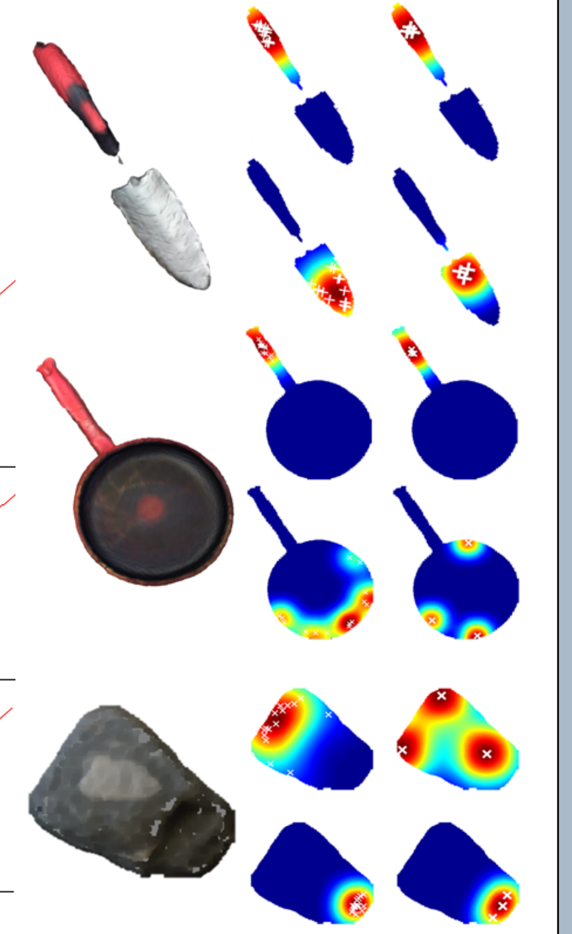
Exp 1. The distribution of human judgments about what the essential physical concepts are vs. learned coefficients of different physical concepts.



Exp 2. Recognizing tools for chopping wood. The scatters show tool candidates ranked by our algorithm (y-axis) with respect to the average ranking by human subjects (x-axis).



Exp 3. Imagined tool-uses.



correlation of ranking algorithm vs. human	chop wood			shovel dirt			paint wall		
	tool	object	stone	tool	object	stone	tool	object	stone
tool + random use	0.07	0.14	0.20	0.52	0.32	0.09	0.12	0.11	0.31
tool + inferred use	0.48	0.25	0.89	0.64	0.89	0.14	0.10	0.64	0.20
tool + best use	0.83	0.43	0.89	0.64	0.89	0.14	0.10	0.64	0.20

3D distance (cm) algorithm vs. human	chop wood			shovel dirt			paint wall		
	tool	object	stone	tool	object	stone	tool	object	stone
B_A - top 1	1.75	3.02	3.19	1.17	2.03	3.28	0.43	2.48	2.86
B_A - top 3	1.04	2.17	2.81	0.97	0.52	2.21	0.31	2.32	2.67
B_F - top 1	0.48	5.97	3.91	6.98	6.38	0.23	2.35	2.74	2.65
B_F - top 3	0.27	5.92	3.95	2.85	3.29	0.31	1.43	2.64	2.71