

Table 1: Real-world Evaluation

Language-conditioned						Goal-conditioned				
Task	TRA	GRIF	LCBC	Octo	AWR	TRA	GRIF	GCBC	Octo	AWR
(A) open the drawer	0.80 ^(±0.1) †	0.20 ^(±0.2)	0.60 ^(±0.2)	0.60 ^(±0.2)	0.40 ^(±0.2)	0.60 ^(±0.2) †	0.60 ^(±0.2)	0.40 ^(±0.2)	0.50 ^(±0.2)	0.80 ^(±0.2)
(A) mushroom in drawer	0.80 ^(±0.1)	0.80 ^(±0.2)	0.40 ^(±0.2)	0.00 ^(±0.0)	0.60 ^(±0.2)	0.90 ^(±0.1)	0.40 ^(±0.2)	0.80 ^(±0.2)	0.90 ^(±0.1)	0.60 ^(±0.2)
(A) close drawer	0.60 ^(±0.2)	0.60 ^(±0.2)	0.40 ^(±0.2)	0.60 ^(±0.2)	0.40 ^(±0.2)	1.00 ^(±0.0)	0.40 ^(±0.2)	0.80 ^(±0.2)	0.60 ^(±0.2)	0.40 ^(±0.2)
(D) take the item out of the drawer	0.60 ^(±0.2)	0.00 ^(±0.0)	0.00 ^(±0.0)	0.20 ^(±0.2)	0.00 ^(±0.0)	0.40 ^(±0.2)	0.00 ^(±0.0)	0.00 ^(±0.0)	0.20 ^(±0.2)	0.00 ^(±0.0)
(B) put the spoons on towels	1.00 ^(±0.0)	0.40 ^(±0.2)	0.20 ^(±0.2)	0.00 ^(±0.0)	0.20 ^(±0.2)	1.00 ^(±0.0)	0.20 ^(±0.2)	0.60 ^(±0.2)	0.40 ^(±0.2)	0.60 ^(±0.2)
(B) put the spoons on the plates	0.80 ^(±0.2)	0.20 ^(±0.2)	0.20 ^(±0.2)	0.20 ^(±0.2)	0.00 ^(±0.0)	1.00 ^(±0.0)	0.00 ^(±0.0)	0.40 ^(±0.2)	0.00 ^(±0.0)	0.80 ^(±0.2)
(C) put the corn and sushi on plate	0.90 ^(±0.1)	0.00 ^(±0.0)	0.40 ^(±0.2)	0.00 ^(±0.0)	0.50 ^(±0.2)	0.70 ^(±0.1)	0.00 ^(±0.0)	0.20 ^(±0.2)	0.00 ^(±0.0)	0.30 ^(±0.1)
(C) sushi and mushroom in bowl	0.80 ^(±0.2)	0.00 ^(±0.0)	0.60 ^(±0.2)	0.20 ^(±0.2)	0.60 ^(±0.2)	0.60 ^(±0.2)	0.00 ^(±0.0)	0.20 ^(±0.2)	0.40 ^(±0.2)	0.60 ^(±0.2)
(C) corn, banana, and sushi in bowl	0.80 ^(±0.1)	0.00 ^(±0.0)	0.00 ^(±0.0)	0.00 ^(±0.0)	0.20 ^(±0.1)	0.50 ^(±0.2)	0.00 ^(±0.0)	0.00 ^(±0.0)	0.40 ^(±0.2)	0.50 ^(±0.2)
(D) corn on plate then sushi in pot	0.70 ^(±0.1)	0.00 ^(±0.0)	0.40 ^(±0.2)	0.60 ^(±0.2)	0.20 ^(±0.2)	0.30 ^(±0.1)	0.20 ^(±0.2)	0.00 ^(±0.0)	0.00 ^(±0.0)	0.00 ^(±0.0)
(A) sweep to the right	0.80 ^(±0.1)	0.20 ^(±0.2)	0.40 ^(±0.2)	0.40 ^(±0.2)	0.00 ^(±0.0)	0.70 ^(±0.1)	0.40 ^(±0.2)	0.00 ^(±0.0)	0.80 ^(±0.2)	0.00 ^(±0.0)
(B) fold cloth into the center	1.00 ^(±0.0)	0.20 ^(±0.2)	0.40 ^(±0.2)	0.40 ^(±0.2)	0.40 ^(±0.2)	1.00 ^(±0.0)	0.00 ^(±0.0)	0.00 ^(±0.0)	0.60 ^(±0.2)	0.00 ^(±0.0)
(B) move bell pepper and sweep towel	0.50 ^(±0.2)	0.00 ^(±0.0)	0.00 ^(±0.0)	0.20 ^(±0.2)	0.00 ^(±0.0)	0.60 ^(±0.2)	0.20 ^(±0.2)	0.20 ^(±0.2)	0.40 ^(±0.2)	0.00 ^(±0.0)

(A) One step tasks
(C) Semantic generalization

(B) Task concatenation
(D) Tasks with dependency

†The best-performing method(s) up to statistical significance are **highlighted**