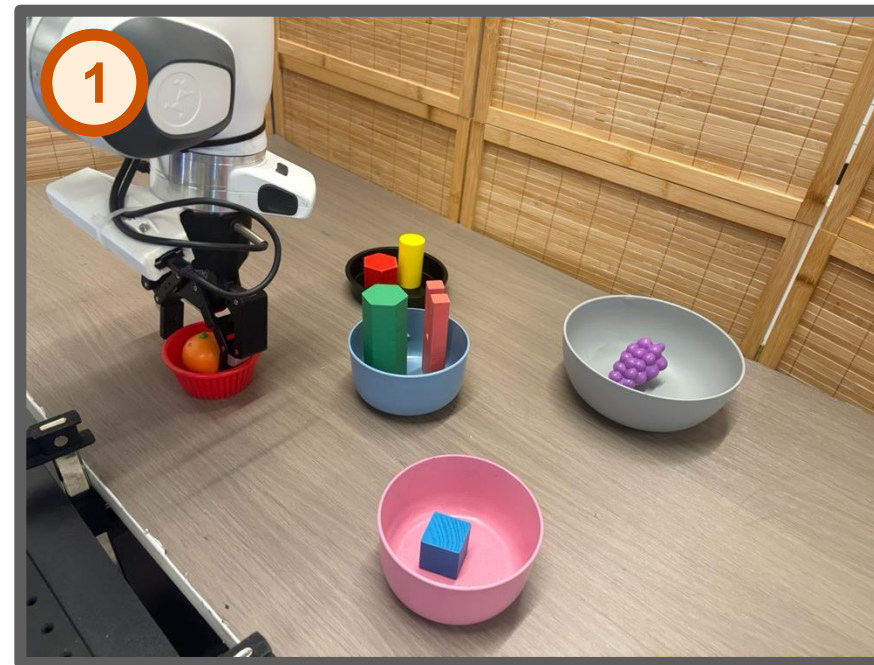*"Empty two of the bowls."*

High-level VLM

Empty the red bowl
1. Toy orange → grey bowl

Empty the blue bowl
2. Pink toy → grey bowl
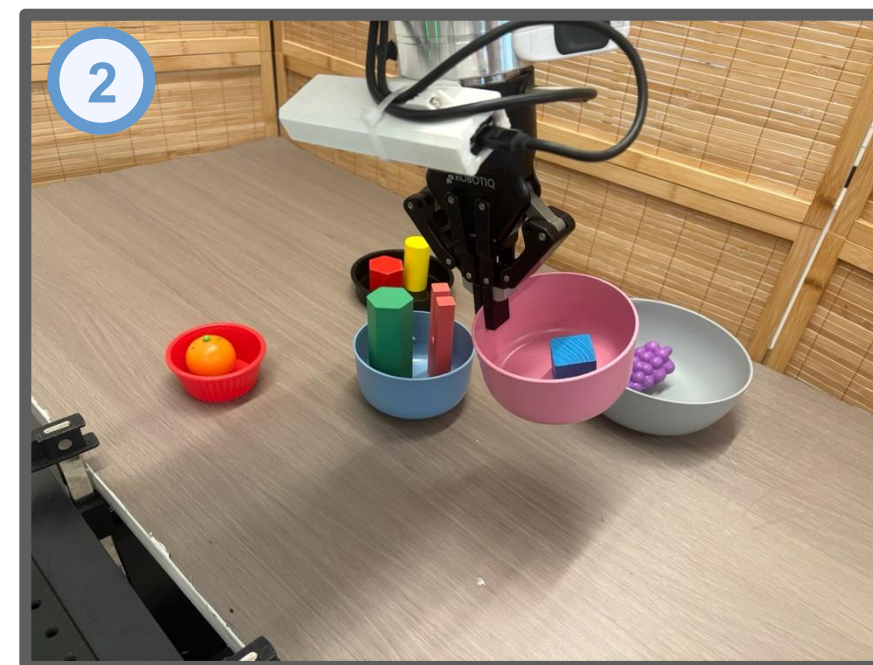3. Green block → grey bowl

Include prior interactions in-context

High-level VLM

Empty the grey bowl *because it is large*
1. Toy grapes → pink bowl

Empty the black bowl *because its contents are tall*
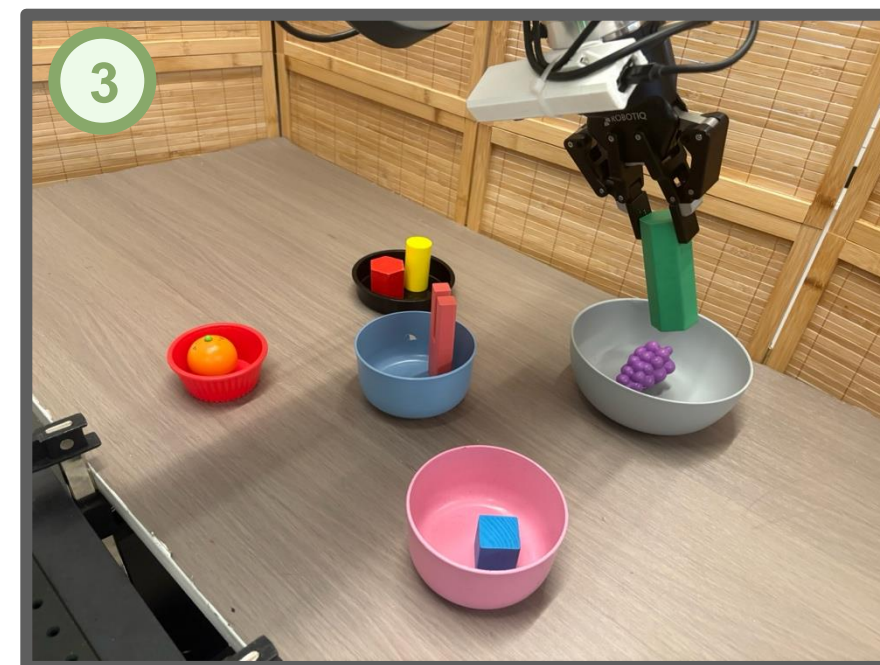2. Red block → pink bowl
3. Yellow cylinder → pink bowl

Initial rollout fails

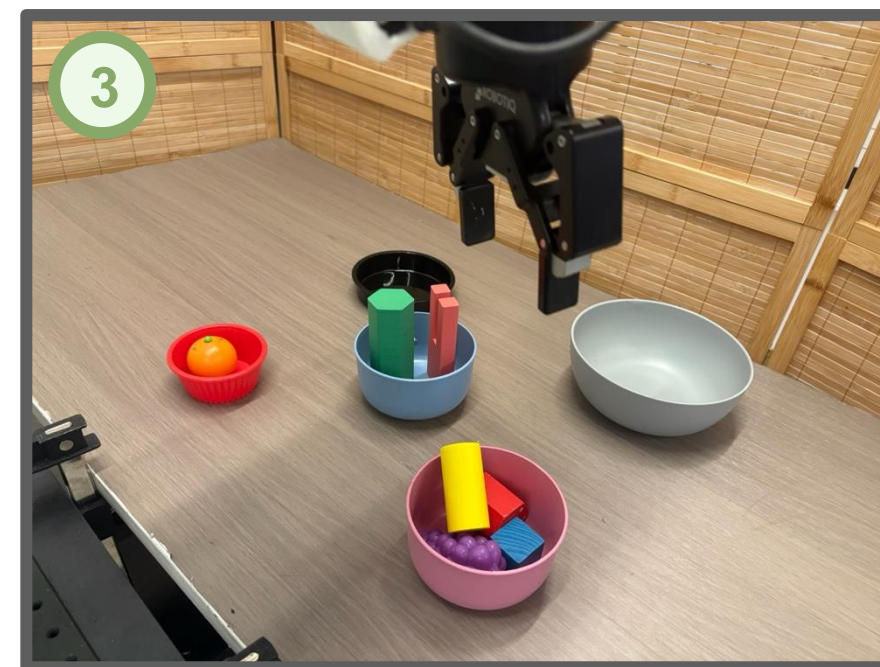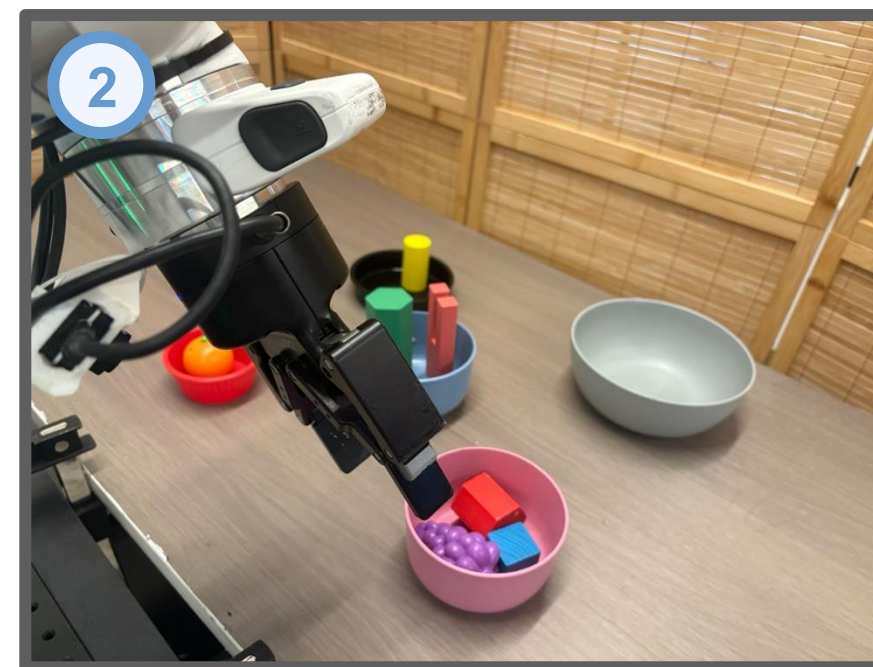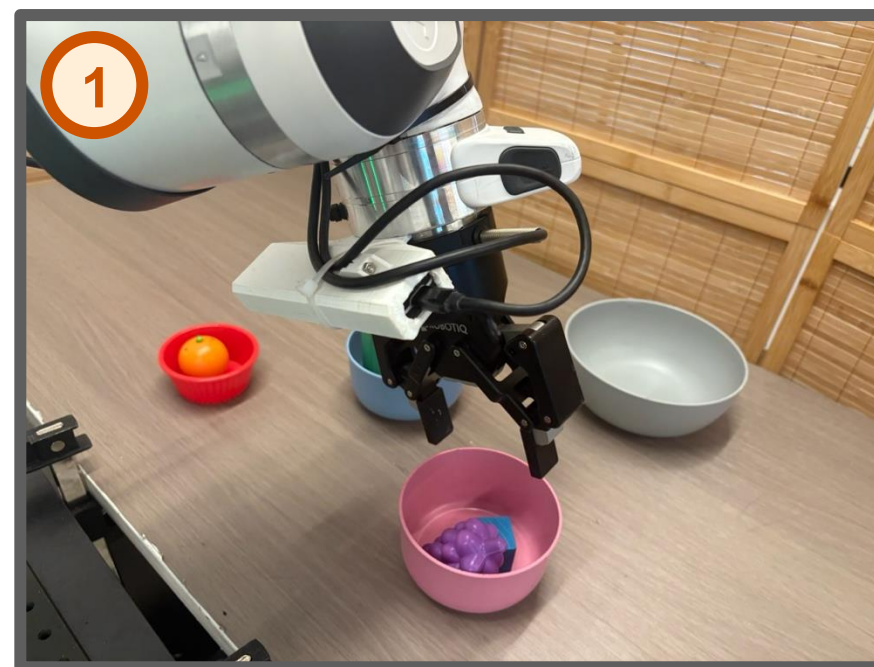① Fingers can't fit in small bowls

② Mistook pink *bowl* for pink *toy*

③ Correctly picked tall green block

Subsequent rollout succeeds by *learning from past experience*

Grey and black bowls cleared, **overall success!**