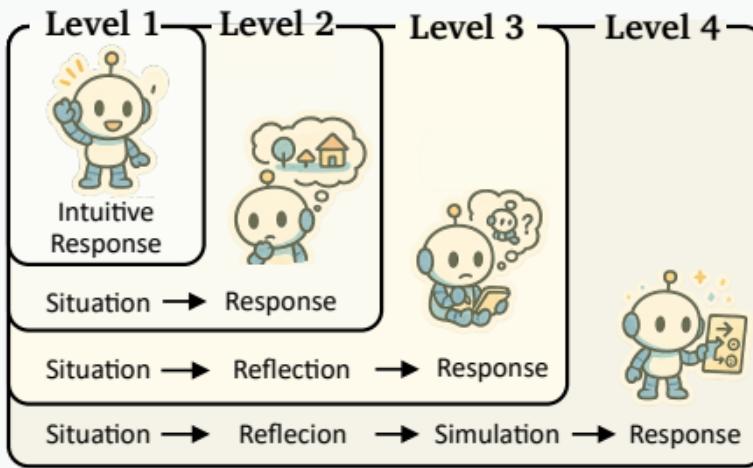
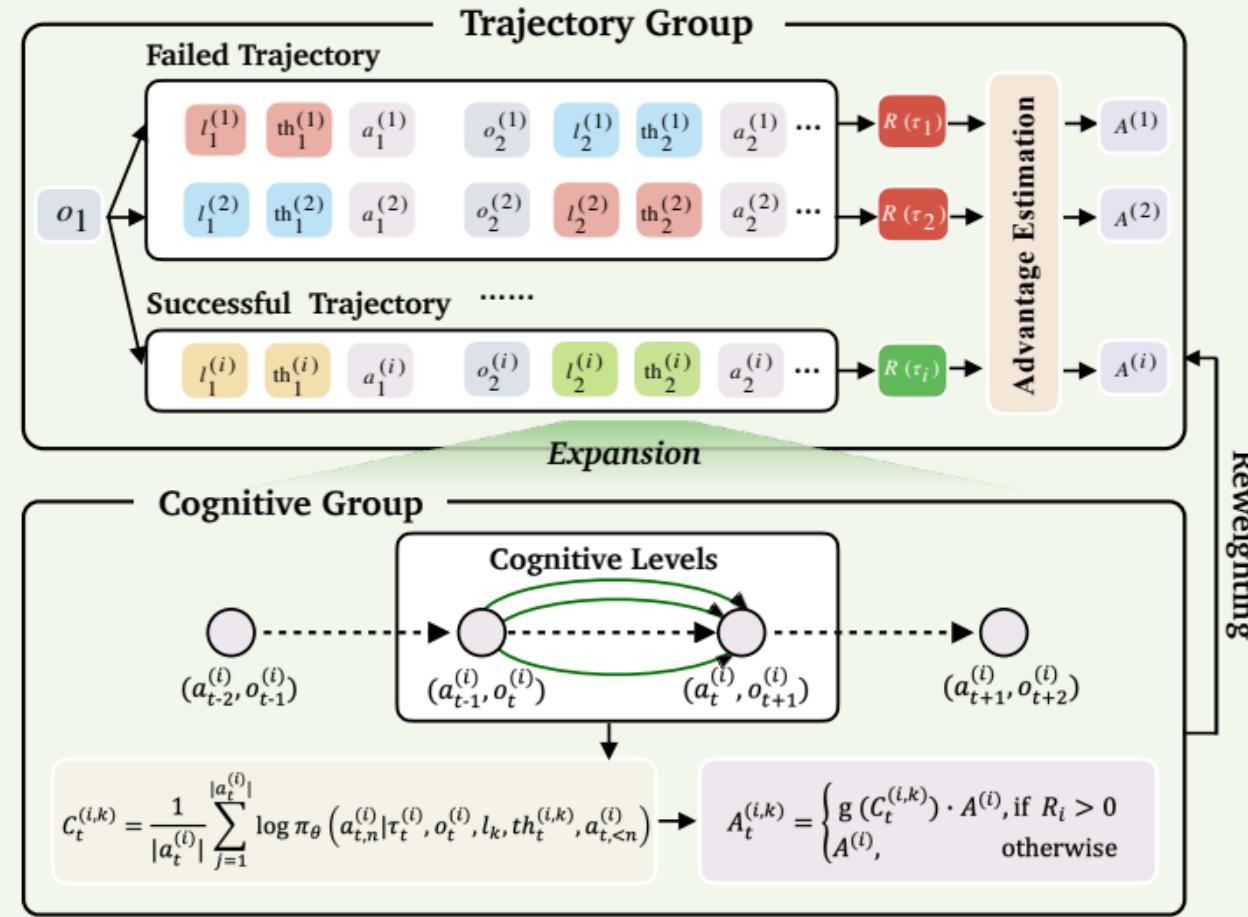


(A) Cognitive Level Design



(C) Cognition-Aware Policy Optimization



(B) Cognition-Aware SFT

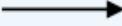
Initialize Balanced Distribution

Randomly Sample Cognitive Level



Expert Trajectory
 $o_1, a_1, \dots, o_T, a_T$

Expert Annotation th_t



Annotated Trajectory
 $\dots, o_t, l_t, th_t, a_t, \dots$