BARS-RL: BIMANUAL ARTICULATED-OBJECT MANIPULATION UNDER RESISTANCE WITH STABILIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Robots in human environments frequently interact with articulated objects that rest unanchored on a surface (e.g., microwaves, toaster ovens), and opening these objects requires overcoming an initial opening resistance. A pulling force below this resistance leaves the moving part closed, whereas one above it can drag the object body, so one arm must pull while the other holds the object body in place. Existing methods that assume anchored objects or negligible opening resistance do not capture this coupling between opening and stabilization. To address these limitations, we propose **BARS-RL**, a reinforcement learning framework in which a single policy allocates force between a task arm and a support arm without observing the resistance or friction. A stabilization reward penalizes contact forces near the base-slip boundary, providing a training signal before the object body moves. Training follows a curriculum organized by the resistance-to-friction ratio, which moves from an anchored object body to an unanchored one and finally mixes episodes that base friction alone can hold with those that require support. On three simulated objects with revolute and prismatic joints, BARS-RL achieves an average success rate of 83.5%, outperforming baselines while approaching the performance of an oracle that observes the resistance and friction.

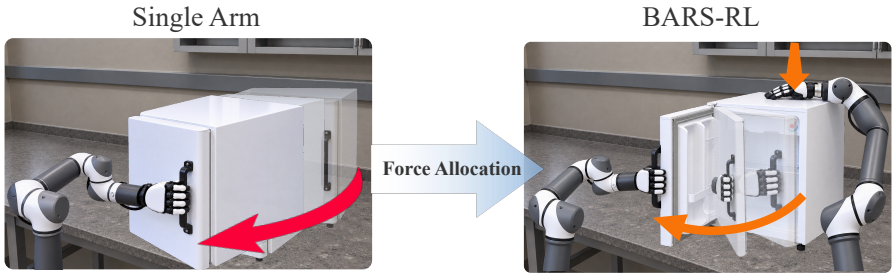

Figure 1: **Conceptual overview of BARS-RL.** With a single arm, pulling on the door drags the unanchored object across the table (left). BARS-RL allocates force between a task arm that opens the door and a support arm that presses the object to hold it in place, even when the resistance is unknown (right).

1 INTRODUCTION

Robots that work in household and industrial settings must manipulate articulated objects such as doors and drawers. Learning-based methods (Mo et al., 2021; Eisner et al., 2022; Wang et al., 2025a) and classical controllers (Jain & Kemp, 2010; Rizzi et al., 2023) have made steady progress on this problem. However, these methods rarely consider articulated objects that rest unanchored on a surface and stay closed until the pull exceeds a breakaway force. In this work, we study how to open such objects (e.g., microwaves, mini fridges, and air fryers) with two arms, one pulling the door open while the other presses the object body to hold it in place.

An important question in this problem is how the two arms should share the force. Existing methods avoid this question through two assumptions. They treat the object body as fixed to the environment

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

or heavy enough not to move, so that the environment absorbs the reaction to the opening force (Rizzi et al., 2023; Mo et al., 2021; Wang et al., 2025a). They also either ignore the opening resistance (Mo et al., 2021; Eisner et al., 2022) or model it as a lock released by a separate action, such as rotating a handle or pressing a button (Li et al., 2024b; Wang et al., 2025b). Without these assumptions, pulling the door can drag the whole object toward the robot before the door opens. A successful policy must therefore satisfy two requirements simultaneously. First, the task arm must pull with a force exceeding the breakaway force. Second, the support arm must press the object body so that base friction can resist this pull.

Both requirements are difficult to satisfy because the robot cannot observe the physical parameters (e.g., resistance and friction) of the object. Our key insight is that a single quantity, the ratio of the opening resistance to the friction, determines whether the support arm is required. Building on this insight, we propose **BARS-RL**, a bimanual reinforcement learning framework in which a single policy allocates force between the task arm and the support arm from force/torque (F/T) and tactile feedback. BARS-RL combines a stabilization reward, which penalizes forces that cause base slip, with a curriculum organized by the resistance-to-friction ratio.

We evaluate BARS-RL in simulation on three articulated objects with revolute and prismatic joints. Without observing the physical parameters, BARS-RL achieves an average success rate of 83.5%, outperforming both scripted and learned baselines and approaching an oracle that observes these parameters.

The contributions of this paper are as follows:

- **Stability quantitative formulation.** We represent latches, detents, and magnetic seals by a single breakaway force and formulate the base-slip margin of an unanchored object body. From this formulation, we derive the resistance-to-friction ratio, which determines whether the support arm is required and is comparable across objects of different mass and friction.
- **Bimanual force allocation.** We propose BARS-RL, in which a single policy allocates force between the task arm and the support arm from proprioceptive, F/T, tactile, and object-state feedback. A stabilization reward penalizes contact forces near the base-slip boundary, and the physical parameters are given only to the critic during training.
- **Physics-organized curriculum and validation.** We organize training by the resistance-to-friction ratio, moving from an anchored object body to an unanchored one before sampling the full range of the ratio. On three simulated articulated objects with revolute and prismatic joints, BARS-RL outperforms scripted and learned baselines, and our ablations show that the curriculum is necessary and that organizing it by the ratio outperforms organizing it by the opening resistance.

2 Related Work

Articulated object manipulation. Prior work on articulated object manipulation predicts actionable regions from visual input (Mo et al., 2021; Wu et al., 2022; Geng et al., 2023), estimates articulation flow (Eisner et al., 2022; Zhang et al., 2023), reasons about parts with foundation models (Geng et al., 2024; Li et al., 2024a), scales policy learning in simulation (Wang et al., 2025a), or opens doors and drawers with classical controllers under kinematic uncertainty (Jain & Kemp, 2010; Karayiannidis et al., 2016; Rizzi et al., 2023). These methods are typically developed in simulators whose articulated objects are fixed to the world and whose joints move without resistance once actuated (Xiang et al., 2020). Opening resistance has been addressed by discovering opening strategies through interaction (Wang et al., 2022; Zhang et al., 2025), or by modeling the closure as a latch with friction (Rajeswaran et al., 2018), a lock released by a preceding motion (Li et al., 2024b; Wang et al., 2025b), or a handle unlatched under force feedback (Wang et al., 2025c; Kang et al., 2024). Measurements on everyday objects also show that breakaway forces vary widely (Jain & Kemp, 2013). In all of these works, however, the object body is anchored or treated as stable, so the opening force does not move the object. Tac-Man (Zhao et al., 2025) handles unexpected part motion through tactile feedback, but it does not model when the pull makes an unanchored object body slide. In contrast, we open unanchored objects under an unknown breakaway force, which requires opening the door and holding the object body in place at the same time (Sections 3 and 4.1).

Force-aware and bimanual manipulation. Contact-rich manipulation methods regulate interaction forces through impedance or hybrid force/position control (Do et al., 2025; Fang et al., 2026; Tahmaz et al., 2025) or learn from force-informed demonstrations (Liu et al., 2025b; Ablett et al., 2025). Because the object body is fixed in these works, the applied force only has to move the joint. Bimanual policies assign complementary roles to the two arms (Fu et al., 2025; Wang & Johns, 2025; Lin et al., 2023; Huang et al., 2025), and the closest to ours are those in which one arm stabilizes the object while the other acts on it (Grannen et al., 2023; Liu et al., 2025a; Lin et al., 2025). Cooperative manipulation regulates internal forces between two arms that rigidly grasp a known load (Uchiyama & Dauchez, 1988; Caccavale et al., 2008; Smith et al., 2012). In both lines of work, the stabilized object is grasped or rests on a fixture, so the stabilizing force does not determine whether the task succeeds. In our problem, the support arm presses on an unanchored object body, and this force raises the friction the base can sustain. Using an external contact to raise friction relates our work to extrinsic dexterity (Chavan-Dafle et al., 2014), fixturing under friction limits (Holladay et al., 2024), friction-cone analysis in grasping (Murray et al., 1994), and incipient-slip detection (Veiga et al., 2015; Chen et al., 2018). Hidden physical parameters are often handled through teacher–student distillation or online adaptation (Chen et al., 2020; Lee et al., 2020; Kumar et al., 2021); we instead use an asymmetric actor–critic (Pinto et al., 2018), which gives the resistance and friction only to the critic. To our knowledge, no prior work uses base friction to coordinate two arms that open an articulated object under unknown resistance.

Curriculum learning. Curricula for RL (Bengio et al., 2009; Narvekar et al., 2020; Portelas et al., 2020b) order episodes by start-state or goal difficulty (Florensa et al., 2017; 2018), learning progress (Portelas et al., 2020a), expanding randomization ranges (OpenAI et al., 2019), or regret- and replay-based level selection (Dennis et al., 2020; Jiang et al., 2021). Staged training has also been effective in locomotion (Rudin et al., 2022) and contact-rich manipulation (Huang et al., 2026; Wang et al., 2026). These methods mostly treat difficulty as a generic variable, without tying it to the physics of the task. Our curriculum instead schedules episodes by the resistance-to-friction ratio, which determines whether the support arm is required (Section 4.3).

3 Force Allocation Formulation

We consider opening an unanchored articulated object under opening resistance. The task arm grasps the object handle, the support arm presses on the object body, and the object body rests on the table through a frictional base contact. Let q_t denote the articulation state of the moving part at time t and q_g the target state; the goal is to drive q_t to q_g while the object body stays in place on the table. Each task instance is characterized by a physical parameter vector $\zeta = (F_{\text{res}}, \mu_{\text{b}}, m)$ that the robot cannot observe. Here F_{res} is the breakaway force of the closure that holds the moving part shut, such as a latch or a magnetic seal, measured along the opening direction; μ_{b} is the friction coefficient of the base contact; and m is the total object mass. We define this task as a *bimanual force allocation* problem: the task arm must apply a force that exceeds a resistance of unknown magnitude, while the support arm presses on the object body to hold it in place.

Base-slip margin and the ratio ρ. While the closure is engaged, the task arm pulls on the handle with force F_{pull} along the opening direction, and the closure transmits this force to the object body. The support arm presses down on the top of the object body with normal force F_{s}^{n}, which raises the normal load on the base contact to $N_{\text{b}} = mg + F_{\text{s}}^{n}$. Let F_T denote the component of the net force that the two arms exert on the object along the opening direction; this is the load the base contact must resist. The task arm contributes F_{pull} to F_T, and as long as the support arm does not push the object body along the opening direction, any friction under the pressing palm opposes it, so $F_T \leq F_{\text{pull}}$. Neglecting tipping and off-axis force, both of which are penalized in the reward (Section 4.2), we define the *base-slip margin* as the difference between the friction capacity of the base contact and this

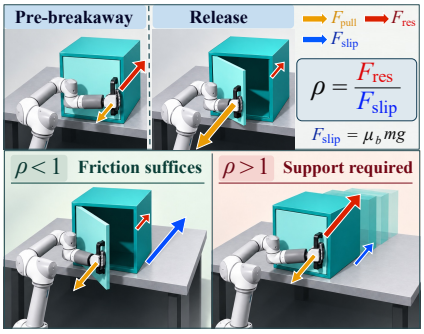

Figure 2: **Resistance-to-friction ratio.** Base friction alone resists the pulling force when $\rho < 1$, whereas the support arm must press on the object body when $\rho > 1$.

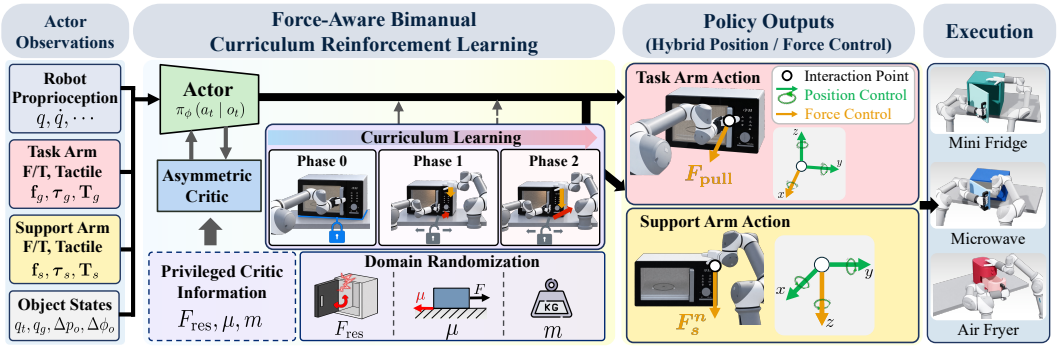

Figure 3: **Pipeline of BARS-RL.** Single policy maps proprioceptive, F/T, tactile, and object-state observations to hybrid position/force commands for both arms. During training, only the critic receives the opening physical parameters, and a curriculum organized by the ρ schedules the episodes.

load,

$$\delta_{\text{b}} = \mu_{\text{b}} N_{\text{b}} - F_T. \tag{1}$$

A positive margin means that base friction has capacity to spare, a margin near zero places the object body at the slip boundary, and a negative margin means that the object body slides.

Breakaway occurs once F_{pull} exceeds F_{res} (Section 4.1). Since $F_T \leq F_{\text{pull}}$, the margin stays positive until breakaway if

$$F_{\text{res}} < \mu_{\text{b}} \left(mg + F_{\text{s}}^n \right), \tag{2}$$

which is a sufficient force-allocation condition for the two arms: the task arm must reach F_{res}, and the support arm must press so that base friction can resist this force. Without support ($F_{\text{s}}^n = 0$), the right-hand side reduces to the slip threshold $F_{\text{slip}} = \mu_{\text{b}} mg$, which defines the resistance-to-friction ratio $\rho = F_{\text{res}}/F_{\text{slip}}$. When $\rho < 1$, base friction alone can hold the object body, and single-arm opening is feasible in principle. When $\rho > 1$, F_{pull} exceeds F_{slip} before breakaway, and equation 2 holds if the support arm presses with

$$F_{\text{s}}^n > mg\,(\rho - 1). \tag{3}$$

This bound depends on the unobserved parameters ζ and is conservative, since it ignores the palm friction that lowers F_T. Because ρ summarizes how much support a task instance requires in a single quantity comparable across objects of different mass and friction, we use $\rho = 1$ as the boundary between the two regimes and organize the curriculum around it (Section 4.3).

4 Method

Overview. We propose BARS-RL (Figure 3), a reinforcement learning framework for opening unanchored articulated objects under opening resistance. Section 4.1 introduces the opening-resistance model, which reduces every closure to a single breakaway force, and Section 4.2 adds stabilization terms built around the base-slip margin to the task reward. For $\rho < 1$, this reward is sufficient on its own. For $\rho > 1$, however, the policy must coordinate both arms, and this coordination is difficult to learn directly; we therefore order training by ρ using the curriculum of Section 4.3.

4.1 Opening Resistance Model

Prior simulation environments model a closure as a kinematic lock that is released by turning a handle or moving a preceding joint, after which the moving part moves freely (Li et al., 2024b; Wang et al., 2025b). These models omit the force that must be overcome before the moving part can move, and it is this force that makes the support arm necessary. We therefore represent latches, detents, and magnetic seals by a single breakaway force F_{res}. While the closure is engaged, a fixed joint attaches the moving part to the object body, so the full pulling force is transmitted to the object body, as assumed in equation 1. At each control step, the wrist force of the task arm is projected onto the opening direction,

$$F_{\text{pull}}(t) = \left\langle \bar{\mathbf{f}}_{\text{w}}(t),\ \hat{\mathbf{a}}(q_t) \right\rangle, \tag{4}$$

where $\bar{\mathbf{f}}_{\mathrm{w}}(t)$ is the wrist F/T force of the task arm averaged over a short moving window and $\hat{\mathbf{a}}(q_t)$ is the unit opening direction. Once $F_{\mathrm{pull}}(t)$ exceeds F_{res}, the fixed joint is released, and the moving part moves under the stiffness and damping of its own joint. Breakaway is thus modeled as a discrete event, consistent with the abrupt release of most household closures. In addition, F_{res} can be set independently in every episode, which the curriculum in Section 4.3 relies on.

4.2 Manipulation under Resistance with Stabilization

Task reward. The task reward drives the moving part to the target and stops it there,

$$r_{\mathrm{task}} = \lambda_{\mathrm{prog}}\, r_{\mathrm{prog}} + \lambda_{\mathrm{succ}}\, r_{\mathrm{succ}} + \lambda_{\mathrm{pull}}\, r_{\mathrm{pull}} + \lambda_{\mathrm{settle}}\, r_{\mathrm{settle}} - \lambda_{\mathrm{time}} - \lambda_{\mathrm{act}}\, c_{\mathrm{act}}, \tag{5}$$

where r_{prog} is the normalized reduction in distance to the target, credited only while the task arm holds the handle, and r_{succ} is a bonus given when the moving part rests at the target with the grasp held and the object body in place. The shaping terms r_{pull} and r_{settle} reward pulling along the opening direction before the closure releases and coming to rest at the target, respectively; the last two terms penalize elapsed time and action rate. However, none of the dense terms can distinguish opening the moving part from dragging the object body toward the robot.

Base-slip cost. The central stabilization term acts on the base-slip margin δ_{b} of equation 1. We express this margin as the fraction of the base friction capacity demanded by the load, $u_{\mathrm{b}} = F_T/(\mu_{\mathrm{b}} N_{\mathrm{b}})$, and penalize it beyond a safety factor $\kappa < 1$,

$$c_{\mathrm{base}} = \mathrm{clip}\left(\frac{u_{\mathrm{b}} - \kappa}{1 - \kappa},\ 0,\ 1\right). \tag{6}$$

Without support, u_{b} reaches ρ at breakaway; the cost is therefore active whenever $\rho > \kappa$ unless the support arm presses on the object body, which raises N_{b} as in equation 2. We also include a displacement cost c_{body} for translating and rotating the object body. While this cost responds only after the object body moves, c_{base} responds to the force state of the contact before any displacement accumulates. Because c_{base} requires μ_{b} and the base-contact force, it is used only during training, and neither quantity is given to the actor.

Support reward. The support reward couples the two arms explicitly,

$$r_{\mathrm{sup}} = \mathbf{1}_{\mathrm{pulling}}\, \mathbf{1}_{\mathrm{pressing}}\, \exp\left[-\left(\|\Delta p_o\|/\sigma_o\right)^2\right], \tag{7}$$

where Δp_o is the translation of the object body since reset, $\mathbf{1}_{\mathrm{pulling}}$ is active while the task arm holds the handle and pulls along the opening direction, and $\mathbf{1}_{\mathrm{pressing}}$ is active while the support arm presses above a threshold. Because both indicators must be active, an idle policy receives no support reward for an object that is merely stationary.

Overall reward. The per-step reward is

$$\begin{aligned} r_t = r_{\mathrm{task}} + w_{\mathrm{sup}}\, \lambda_{\mathrm{sup}}\, r_{\mathrm{sup}} - \lambda_{\mathrm{body}}\, c_{\mathrm{body}} \\ - w_{\mathrm{stab}}\left(\lambda_{\mathrm{base}}\, c_{\mathrm{base}} + \lambda_{\mathrm{overgrip}}\, c_{\mathrm{overgrip}} + \lambda_{\mathrm{offaxis}}\, c_{\mathrm{offaxis}}\right), \end{aligned} \tag{8}$$

where c_{overgrip} penalizes excess grip force, c_{offaxis} penalizes wrist force that does not contribute to opening, and $w_{\mathrm{stab}} \in [0, 1]$ and $w_{\mathrm{sup}} \in \{0, 1\}$ are stage-dependent weights (Section 4.3). The off-axis force and the tipping that the base-slip margin neglects are covered by c_{offaxis} and by the rotation part of c_{body}, respectively. All term definitions, weights, and thresholds are given in Appendix C.2.

4.3 Physics-Organized ρ-Curriculum

Motivation. Because most episodes in the final task distribution have $\rho > 1$, optimizing equation 8 directly on this distribution requires the policy to discover pulling and pressing at the same time; until it does, pulling only drags the object body and the closure never releases. The ratio ρ separates episodes by this difficulty at $\rho = 1$: below one, base friction absorbs the resistance and a single arm suffices; above one, the object body starts to slide before the pulling force reaches the breakaway force, so the closure does not release unless the support arm presses on the object body. It is not, however, a property of the resistance alone: the same F_{res} is a single-arm task on a heavy

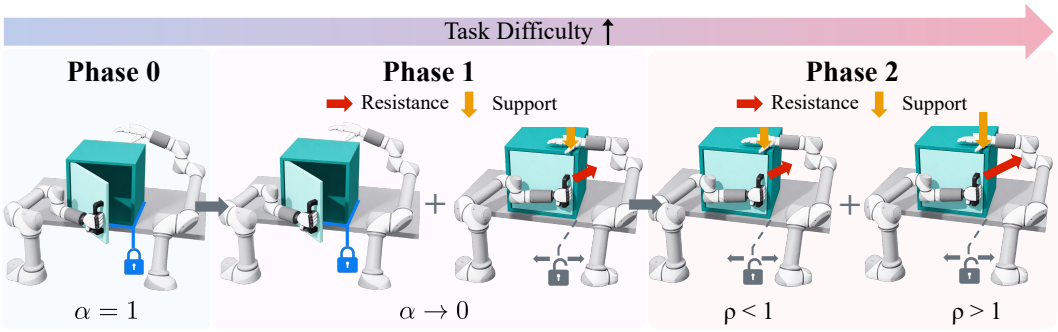

Figure 4: **Curriculum phases.** Training proceeds from anchored articulation (Phase 0) to a mixture of anchored and unanchored episodes with $\rho < 1$ (Phase 1), and then to a single stage that samples ρ uniformly over an interval straddling $\rho = 1$, in which episodes with $\rho > 1$ dominate (Phase 2).

object with a rough base and a two-arm task on a light, slippery one. We therefore organize the curriculum by ρ itself. The mass m and base friction μ_b are sampled per environment, ρ is sampled per episode from a stage-dependent interval $\mathcal{R}_s$, and the resistance is set to the value that realizes the drawn ratio,

$$F_{\text{res}} = \rho\, F_{\text{slip}}(m, \mu_b). \tag{9}$$

Curriculum phases. Each stage s fixes the probability α_s that the object body is anchored, together with the interval $\mathcal{R}_s$; the stages are grouped into the three phases of Figure 4.

Phase 0: anchored articulation. The object body is rigidly anchored ($\alpha_0 = 1$), the closure offers no resistance, and the contact costs and the support reward are switched off ($w_{\text{stab}} = w_{\text{sup}} = 0$). The policy learns to move the moving part and stop at q_g before it is asked to stabilize anything.

Phase 1: from anchored to unanchored. Each episode is anchored with probability α_s and unanchored otherwise, and α_s is lowered stage by stage to zero. The support reward is enabled ($w_{\text{sup}} = 1$) and the contact costs enter at half weight ($w_{\text{stab}} = 0.5$), while $\mathcal{R}_s$ is held fixed and lies entirely below one. The policy therefore learns to pull without moving an unanchored object body before it encounters a resistance that base friction cannot absorb.

Algorithm 1 ρ-curriculum

Require: Stages $\{(\alpha_s, \mathcal{R}_s, \tau_s)\}_{s=0}^{S}$
$s \leftarrow 0$, $\mathcal{B} \leftarrow \emptyset$
for each PPO iteration **do**
  for each environment where episode is done **do**
    if episode began under stage s **then**
      Append success indicator to $\mathcal{B}$
    end if
    Anchor the object body with probability α_s
    Sample $\rho \sim \mathcal{U}(\mathcal{R}_s)$
    Set $F_{\text{res}} \leftarrow \rho \cdot F_{\text{slip}}(m, \mu_b)$
  end for
  if $|\mathcal{B}| \geq$ number of environments **then**
    if $\text{mean}(\mathcal{B}) \geq \tau_s$ **and** $s < S$ **then**
      $s \leftarrow s + 1$
      Resample m, μ_b
    end if
    $\mathcal{B} \leftarrow \emptyset$
  end if
end for

Phase 2: full ρ range. All episodes are unanchored and the contact costs are at full weight ($w_{\text{stab}} = 1$). A single final stage samples ρ uniformly from an interval $\mathcal{R}_S$ that spans both $\rho < 1$ and $\rho > 1$; the upper end of the μ_b range is clipped so that the largest resistance stays within the wrench limit of the task arm. Phase 2 thus retains the episodes that base friction can hold and adds those that require support, so the policy learns from one distribution both when support is needed and how much force to apply.

5 Experiments

In this section, we evaluate whether BARS-RL can open unanchored articulated objects under unknown opening resistance while holding the object body in place. Section 5.1 describes the objects, robot, and baselines. Section 5.2 compares BARS-RL with these baselines, and Section 5.3 analyzes when the support arm is needed, why the curriculum is defined by ρ, and how ρ should be scheduled in the final phase. Unless stated otherwise, we trained each method with 5 random seeds and evaluated the checkpoint with the highest training success rate of each seed on 1,024 parallel envi-

Table 1: **Main results.** Success rate (%) on three unanchored articulated objects.

Method	Microwave	Mini fridge	Air fryer	Avg.
Fixed-force bimanual controller	15.7±0.8	38.1±2.4	17.7±1.7	23.8
Single-arm	39.5±2.5	31.8±1.2	17.1±1.3	29.5
BARS-RL w/o curriculum	0	0	0	0
BARS-RL	**85.3**±2.0	**68.4**±3.3	**96.7**±0.6	**83.5**
Oracle-physics actor (ζ observed)	94.7±2.8	77.2±5.5	98.2±1.4	90.0

ronments, with ρ sampled uniformly from $[1.25, 2.0]$, a range in which base friction alone cannot hold the object body. An episode succeeds when Eq. equation 12 holds, and we report success rates (%) with standard deviations across seeds.

5.1 Experimental Setup

Environment. We evaluated on three unanchored articulated objects in Isaac Lab (NVIDIA et al., 2025): a microwave and a mini fridge with revolute doors, and an air fryer with a prismatic basket. Each object rested on a table without being fixed, and its closure was equipped with the opening-resistance model of Section 4.1. The robot consisted of two RB5-850E (Rainbow Robotics, 2026) arms; the task arm carries an XHand1 (ROBOTERA, 2025) that grasps the handle, and the support arm carries an Inspire hand (Beijing Inspire-Robots Technology Co., Ltd., 2025) that presses on the object body. We trained one policy per object with PPO (Schulman et al., 2017; Schwarke et al., 2025) and an asymmetric actor–critic (Pinto et al., 2018). The actor observed only proprioception, wrist F/T, tactile signals, and the object state, whereas the physical parameters ζ were randomized during training and provided to the critic alone. Observation and action spaces, reward weights, hyperparameters, and randomization ranges are given in Appendix C.

Baselines. We compared against (i) a *fixed-force bimanual controller*, a scripted controller that pulled and pressed with constant force targets, (ii) a *single-arm* policy trained with the same objective and curriculum but without the support arm, and (iii) *BARS-RL w/o curriculum*, trained directly on the final-phase distribution. As an upper bound, we also report (iv) an *oracle-physics actor* that additionally observes ζ. The fixed-force controller pulled with the largest breakaway force in the final-phase range and presses with the force that equation 3 requires for this breakaway force, plus a 25% margin, within the force limit of the support arm.

5.2 Main Results

Table 1 reports the success rate of each method on the three objects, and Figure 5 compares rollouts of each method on the mini fridge. BARS-RL achieves an average success rate of 83.5%, outperforming all baselines by more than 50 points and reaching about 93% of the oracle-physics actor (90.0%), which directly observes ζ. On the air fryer, BARS-RL (96.7%) nearly matches the oracle (98.2%), whereas the gap is largest on the mini fridge (68.4% vs. 77.2%). Since the oracle differs from BARS-RL only in observing ζ, this gap quantifies the benefit of direct access to the physical parameters over acting from feedback alone.

Why the baselines fail. The *fixed-force bimanual controller* always pulls with the largest breakaway force in the final-phase range, even though the actual breakaway force varies across episodes with the object mass, base friction, and ρ. It therefore pulls harder than necessary in most episodes, and once the closure releases, the door stops before the target (Figure 5, first row). The *single-arm* policy succeeds only in episodes that base friction alone can hold. Beyond this range, the pulling force is transmitted to the object body, which slides out of tolerance before the closure releases (Figure 5, second row; Section 5.3). *BARS-RL w/o curriculum*, trained directly on the final-phase distribution, fails on every object. In this distribution, the policy must discover pulling and pressing at the same time, and until it learns to press, pulling only drags the object body (Figure 5, third row). The closure therefore never releases, and learning stalls without a success signal (Section 4.3). The curriculum avoids this problem: the policy first learns to open the moving part on an anchored object

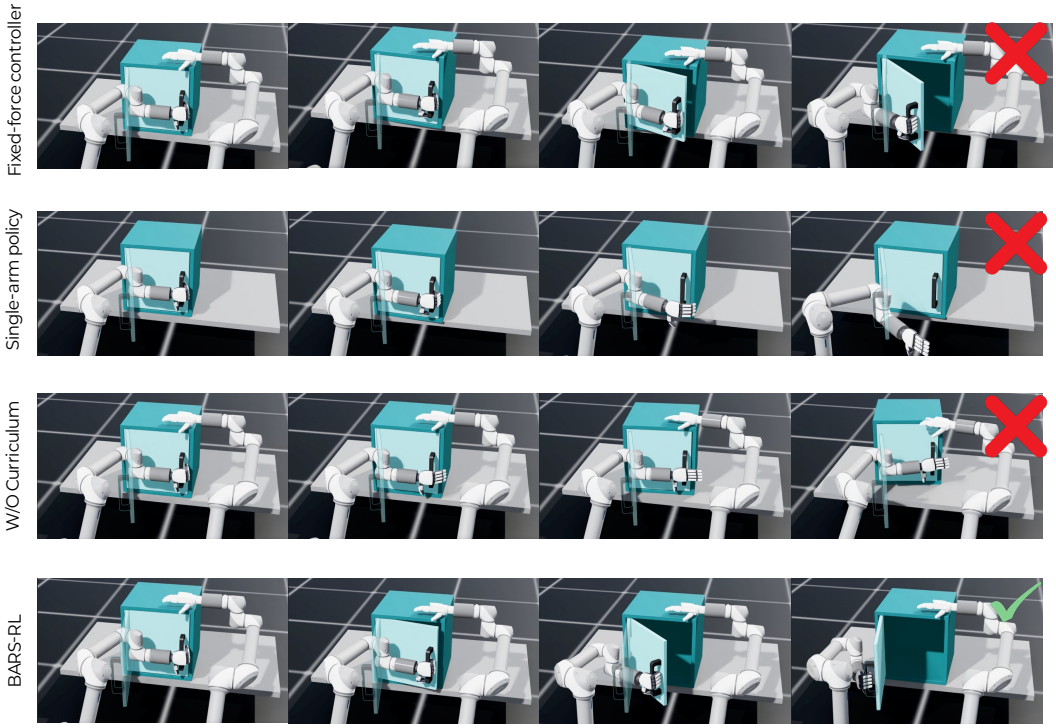

Figure 5: **Rollouts on the mini fridge.** Each row shows one method from left to right over time. The Fixed-force bimanual controller exceeds the required force and does not reach the target. A Single-arm policy and BARS-RL w/o curriculum drag the object body while the door stays closed. BARS-RL presses on the object body with the support arm and opens the door to the target.

body and then on an unanchored one that base friction can hold, before it trains on a range of ρ that includes episodes requiring support.

5.3 Q&As

We examine the formulation and the curriculum design of BARS-RL through the following Q&As.

Q: Is the support arm necessary?

A: In our formulation (Section 3), ρ is the ratio of the opening resistance to the friction the base can sustain on its own, so the object body begins to slide once $\rho > 1$ and the support arm is needed to hold it in place. To test this, we trained a single-arm policy for each object under the same settings as BARS-RL and evaluated both policies over $\rho \in [0.75, 2.0]$ (Figure 6). At low ρ, the single-arm policy performs better (e.g., 98% vs. 76% on the mini fridge at $\rho = 0.75$). Since a single arm suffices in this range and the success rate of BARS-RL changes little with ρ, we believe this gap is a side effect of using both arms even when support is unnecessary. As ρ increases, the two policies become comparable at $\rho = 1.25$, where the object body already slides but its displacement mostly stays within the success tolerance. The gap becomes clear at $\rho = 1.5$, where the displacement exceeds the tolerance and the single-arm policy drops to 11% on the microwave and 0% on the other two objects, while BARS-RL maintains its success rate over the whole range. All single-arm failures in this regime are drag failures, which correspond to the event $\delta_b < 0$ that ρ predicts, and the same pattern holds across objects with different mass, friction, and joint type.

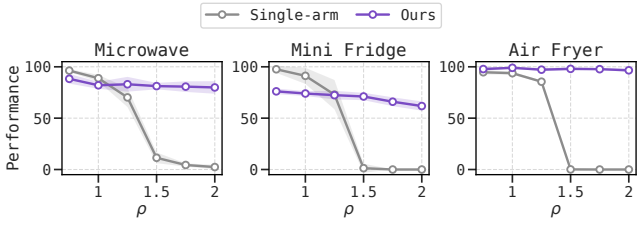

Figure 6: **Support arm becomes necessary at high ρ.**

The support arm therefore becomes necessary once the object body slides beyond the task tolerance, which occurs between $\rho = 1.25$ and $\rho = 1.5$ for all three objects.

Q: Is a curriculum necessary, and why define it in terms of ρ?

A: BARS-RL trains on anchored objects, then on unanchored objects with $\rho < 1$, and finally on a range of ρ that spans both $\rho < 1$ and $\rho > 1$ (Section 4.3). To test this design, we compared it with two variants: *w/o curriculum*, which was trained directly on the final-phase distribution, and an F_{res} *curriculum*, which shared Phases 0 and 1 but raised the opening resistance F_{res} stage by stage in the final phase instead of scheduling by ρ. Table 2 shows the results. Without the curriculum, the policy fails on every object, because it must learn to pull and press at the same time from the start. The F_{res} curriculum learns the task but falls behind BARS-RL on all three objects (68.6% vs. 83.5% on average). We believe this is because the same F_{res} corresponds to different values of ρ under the randomized mass and friction, so each stage mixes episodes that need no support with episodes that need strong support. In contrast, ρ directly measures how much support an episode requires, as shown by the consistent transition across objects in Figure 6.

Table 2: **Curriculum defined by ρ is necessary.**

Training	Microwave	Mini fridge	Air fryer	Avg.
w/o curriculum	0	0	0	0
F_{res} curriculum	53.9±4.8	64.2±2.5	87.7±6.6	68.6
BARS-RL	**85.3**±2.0	**68.4**±3.3	**96.7**±0.6	**83.5**

Q: Should ρ be increased progressively in the final phase?

A: A common way to build a curriculum is to raise the difficulty step by step (Bengio et al., 2009; Narvekar et al., 2020). We therefore compared BARS-RL, which samples ρ uniformly in the final phase, with *progressive* ρ, which divided the same range into stages from low to high ρ and advanced to the next stage once the success rate passed a threshold. The two variants share Phases 0 and 1 and differed only in the schedule of the final phase. Table 3 shows that ordering does not help: the two perform comparably on the microwave and air fryer, and BARS-RL is better on the mini fridge (68.4% vs. 58.7%). The reason is that progressive ρ reaches the high-ρ stages only after passing the success threshold at every earlier stage. It reaches the final stage within the training budget on the microwave and air fryer, but on the mini fridge it remains at a stage covering $\rho \leq 1.35$ in all seeds, so most of the evaluated range lies beyond what it has trained on. Uniform sampling has no such threshold and exposes the policy to the full range of ρ from the start of the final phase.

Table 3: **Ordering ρ does not help.**

Final phase	Microwave	Mini fridge	Air fryer	Avg.
Progressive ρ	82.2±2.4	58.7±4.7	**97.4**±0.1	79.4
Uniform ρ (BARS-RL)	**85.3**±2.0	**68.4**±3.3	96.7±0.6	**83.5**

6 Conclusion

We present BARS-RL, a reinforcement learning framework in which a single bimanual policy allocates force between a task arm and a support arm to open unanchored articulated objects under unknown opening resistance. Through the base-slip margin, we show that the resistance-to-friction ratio determines when the support arm is required, and we use this ratio to organize the training curriculum. Our experiments on three simulated objects with revolute and prismatic joints demonstrate that BARS-RL outperforms scripted and learned baselines and approaches an oracle that observes the physical parameters.

Limitations. Although we represent latches, detents, and magnetic seals by a single breakaway force, real closures may release gradually or depend on the pulling direction, and all of our experiments were conducted in simulation. Transferring BARS-RL to a physical robot will require closure models identified from real objects (Jain & Kemp, 2013). In addition, we trained one policy per object, and at low ρ BARS-RL performs worse than a single-arm policy because it engages both arms even when support is unnecessary (Section 5.3). Learning a single policy across objects that engages the support arm only when needed is a natural next step.

REFERENCES

T. Ablett, O. Limoyo, A. Sigal, A. Jilani, J. Kelly, K. Siddiqi, F. Hogan, and G. Dudek. Multimodal and force-matched imitation learning with a see-through visuotactile sensor. *IEEE Transactions on Robotics*, 41:946–959, 2025.

Beijing Inspire-Robots Technology Co., Ltd. Inspire hand official website. https://inspire-robots.store/collections/the-dexterous-hands, 2025. Accessed: 2025-03-17.

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *International Conference on Machine Learning (ICML)*, 2009.

Fabrizio Caccavale, Pasquale Chiacchio, Alessandro Marino, and Luigi Villani. Six-DOF impedance control of dual-arm cooperative manipulators. *IEEE/ASME Transactions on Mechatronics*, 13(5): 576–586, 2008.

Nikhil Chavan-Dafle, Alberto Rodriguez, Robert Paolini, Bowei Tang, Siddhartha S. Srinivasa, Michael Erdmann, Matthew T. Mason, Ivan Lundberg, Harald Staab, and Thomas Fuhlbrigge. Extrinsic dexterity: In-hand manipulation with external forces. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1578–1585, 2014.

Dian Chen, Brady Zhou, Vladlen Koltun, and Philipp Krähenbühl. Learning by cheating. In *Conference on Robot Learning (CoRL)*, volume 100 of *Proceedings of Machine Learning Research*, pp. 66–75, 2020.

Wei Chen, Heba Khamis, Ingvars Birznieks, Nathan F. Lepora, and Stephen J. Redmond. Tactile sensors for friction estimation and incipient slip detection—toward dexterous robotic manipulation: A review. *IEEE Sensors Journal*, 18(22):9049–9064, 2018.

Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (ELUs). In *International Conference on Learning Representations (ICLR)*, 2016.

Michael Dennis, Natasha Jaques, Eugene Vinitsky, Alexandre Bayen, Stuart Russell, Andrew Critch, and Sergey Levine. Emergent complexity and zero-shot transfer via unsupervised environment design. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

T.-D. Do, N. Gireesh, J. Wang, and H. Wang. Watch less, feel more: Sim-to-real RL for generalizable articulated object manipulation via motion adaptation and impedance control. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4806–4812, 2025.

B. Eisner, H. Zhang, and D. Held. FlowBot3D: Learning 3D articulation flow to manipulate articulated objects. In *Robotics: Science and Systems (RSS)*, 2022.

H. Fang, S. Tang, M. Mei, H. Qin, Z. He, J. Chen, Y. Feng, C. Wang, W. Liu, Z. He, C. Lu, and S. Wang. Force Policy: Learning hybrid force-position control policy under interaction frame for contact-rich manipulation. In *Robotics: Science and Systems (RSS)*, 2026.

Carlos Florensa, David Held, Markus Wulfmeier, Michael Zhang, and Pieter Abbeel. Reverse curriculum generation for reinforcement learning. In *Conference on Robot Learning (CoRL)*, 2017.

Carlos Florensa, David Held, Xinyang Geng, and Pieter Abbeel. Automatic goal generation for reinforcement learning agents. In *International Conference on Machine Learning (ICML)*, 2018.

Zipeng Fu, Tony Z Zhao, and Chelsea Finn. Mobile ALOHA: Learning bimanual mobile manipulation using low-cost whole-body teleoperation. In *Conference on Robot Learning (CoRL)*, volume 270 of *Proceedings of Machine Learning Research*, pp. 4066–4083, 2025.

H. Geng, Z. Li, Y. Geng, J. Chen, H. Dong, and H. Wang. PartManip: Learning cross-category generalizable part manipulation policy from point cloud observations. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2978–2988, 2023.

H. Geng, S. Wei, C. Deng, B. Shen, H. Wang, and L. J. Guibas. SAGE: Bridging semantic and actionable parts for generalizable articulated-object manipulation under language instructions. In *Robotics: Science and Systems (RSS)*, 2024.

Jennifer Grannen, Yilin Wu, Brandon Vu, and Dorsa Sadigh. Stabilize to act: Learning to coordinate for bimanual manipulation. In *Conference on Robot Learning (CoRL)*, volume 229 of *Proceedings of Machine Learning Research*, pp. 563–576, 2023.

Rachel Holladay, Tomás Lozano-Pérez, and Alberto Rodriguez. Robust planning for multi-stage forceful manipulation. *The International Journal of Robotics Research*, 43(3):330–353, 2024.

B. Huang, J. Xu, I. Akinola, W. Yang, B. Sundaralingam, R. O'Flaherty, D. Fox, X. Wang, A. Mousavian, Y.-W. Chao, and Y. Li. VT-Refine: Learning bimanual assembly with visuo-tactile feedback via simulation fine-tuning. In *Conference on Robot Learning (CoRL)*, volume 305 of *Proceedings of Machine Learning Research*, pp. 2484–2500, 2025.

C.-Y. Huang, A. Efendi, and Y.-C. Wang. Curriculum-guided deep reinforcement learning with fuzzy rewards for autonomous push-grasp manipulation in cluttered environments. *Robotics and Autonomous Systems*, 198:105325, 2026.

Advait Jain and Charles C. Kemp. Pulling open doors and drawers: Coordinating an omni-directional base and a compliant arm with equilibrium point control. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2010.

Advait Jain and Charles C. Kemp. Improving robot manipulation with data-driven object-centric models of everyday forces. *Autonomous Robots*, 35(2–3):143–159, 2013.

Minqi Jiang, Edward Grefenstette, and Tim Rocktäschel. Prioritized level replay. In *International Conference on Machine Learning (ICML)*, 2021.

G. Kang, H. Seong, D. Lee, and D. H. Shim. A versatile door opening system with mobile manipulator through adaptive position-force control and reinforcement learning. *Robotics and Autonomous Systems*, 180:104760, 2024.

Yiannis Karayiannidis, Christian Smith, Francisco Eli Viña Barrientos, Petter Ögren, and Danica Kragic. An adaptive control approach for opening doors and drawers under uncertainties. *IEEE Transactions on Robotics*, 32(1):161–175, 2016.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

Ashish Kumar, Zipeng Fu, Deepak Pathak, and Jitendra Malik. RMA: Rapid motor adaptation for legged robots. In *Robotics: Science and Systems (RSS)*, 2021.

Joonho Lee, Jemin Hwangbo, Lorenz Wellhausen, Vladlen Koltun, and Marco Hutter. Learning quadrupedal locomotion over challenging terrain. *Science Robotics*, 5(47):eabc5986, 2020.

X. Li, M. Zhang, Y. Geng, H. Geng, Y. Long, Y. Shen, R. Zhang, J. Liu, and H. Dong. ManipLLM: Embodied multimodal large language model for object-centric robotic manipulation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18061–18070, 2024a.

Yu Li, Xiaojie Zhang, Ruihai Wu, Zilong Zhang, Yiran Geng, Hao Dong, and Zhaofeng He. UniDoorManip: Learning universal door manipulation policy over large-scale and diverse door manipulation environments. *arXiv preprint arXiv:2403.02604*, 2024b.

Toru Lin, Zhao-Heng Yin, Haozhi Qi, Pieter Abbeel, and Jitendra Malik. Twisting lids off with two hands. In *Conference on Robot Learning (CoRL)*, volume 270 of *Proceedings of Machine Learning Research*, pp. 5220–5235, 2025.

Y. Lin, A. Church, M. Yang, H. Li, J. Lloyd, D. Zhang, and N. F. Lepora. Bi-Touch: Bimanual tactile manipulation with sim-to-real deep reinforcement learning. *IEEE Robotics and Automation Letters*, 8(9):5472–5479, 2023.

I.-C. A. Liu, S. He, D. Seita, and G. S. Sukhatme. VoxAct-B: Voxel-based acting and stabilizing policy for bimanual manipulation. In *Conference on Robot Learning (CoRL)*, volume 270 of *Proceedings of Machine Learning Research*, pp. 4354–4370, 2025a.

W. Liu, J. Wang, Y. Wang, W. Wang, and C. Lu. ForceMimic: Force-centric imitation learning with force-motion capture system for contact-rich manipulation. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1105–1112, 2025b.

K. Mo, L. J. Guibas, M. Mukadam, A. Gupta, and S. Tulsiani. Where2Act: From pixels to actions for articulated 3D objects. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6813–6823, 2021.

R. M. Murray, Z. Li, and S. S. Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, Boca Raton, FL, 1994.

Sanmit Narvekar, Bei Peng, Matteo Leonetti, Jivko Sinapov, Matthew E. Taylor, and Peter Stone. Curriculum learning for reinforcement learning domains: A framework and survey. *Journal of Machine Learning Research*, 21(181):1–50, 2020.

NVIDIA, Mayank Mittal, Pascal Roth, James Tigue, Antoine Richard, Octi Zhang, Peter Du, Antonio Serrano-Muñoz, Xinjie Yao, René Zurbrügg, Nikita Rudin, Lukasz Wawrzyniak, Milad Rakhsha, Alain Denzler, Eric Heiden, Ales Borovicka, Ossama Ahmed, Iretiayo Akinola, Abrar Anwar, Mark T. Carlson, Ji Yuan Feng, Animesh Garg, Renato Gasoto, Lionel Gulich, Yijie Guo, M. Gussert, Alex Hansen, Mihir Kulkarni, Chenran Li, Wei Liu, Viktor Makoviychuk, Grzegorz Malczyk, Hammad Mazhar, Masoud Moghani, Adithyavairavan Murali, Michael Noseworthy, Alexander Poddubny, Nathan Ratliff, Welf Rehberg, Clemens Schwarke, Ritvik Singh, James Latham Smith, Bingjie Tang, Ruchik Thaker, Matthew Trepte, Karl Van Wyk, Fangzhou Yu, Alex Millane, Vikram Ramasamy, Remo Steiner, Sangeeta Subramanian, Clemens Volk, CY Chen, Neel Jawale, Ashwin Varghese Kuruttukulam, Michael A. Lin, Ajay Mandlekar, Karsten Patzwaldt, John Welsh, Huihua Zhao, Fatima Anes, Jean-Francois Lafleche, Nicolas Moënne-Loccoz, Soowan Park, Rob Stepinski, Dirk Van Gelder, Chris Amevor, Jan Carius, Jumyung Chang, Anka He Chen, Pablo de Heras Ciechomski, Gilles Daviet, Mohammad Mohajerani, Julia von Muralt, Viktor Reutskyy, Michael Sauter, Simon Schirm, Eric L. Shi, Pierre Terdiman, Kenny Vilella, Tobias Widmer, Gordon Yeoman, Tiffany Chen, Sergey Grizan, Cathy Li, Lotus Li, Connor Smith, Rafael Wiltz, Kostas Alexis, Yan Chang, David Chu, Linxi Fan, Farbod Farshidian, Ankur Handa, Spencer Huang, Marco Hutter, Yashraj Narang, Soha Pouya, Shiwei Sheng, Yuke Zhu, Miles Macklin, Adam Moravanszky, Philipp Reist, Yunrong Guo, David Hoeller, and Gavriel State. Isaac Lab: A GPU-accelerated simulation framework for multi-modal robot learning. *arXiv preprint arXiv:2511.04831*, 2025.

OpenAI, Ilge Akkaya, Marcin Andrychowicz, Maciek Chociej, Mateusz Litwin, Bob McGrew, Arthur Petron, Alex Paino, Matthias Plappert, Glenn Powell, Raphael Ribas, Jonas Schneider, Nikolas Tezak, Jerry Tworek, Peter Welinder, Lilian Weng, Qiming Yuan, Wojciech Zaremba, and Lei Zhang. Solving Rubik's cube with a robot hand. *arXiv preprint arXiv:1910.07113*, 2019.

Lerrel Pinto, Marcin Andrychowicz, Peter Welinder, Wojciech Zaremba, and Pieter Abbeel. Asymmetric actor critic for image-based robot learning. In *Robotics: Science and Systems (RSS)*, 2018.

Rémy Portelas, Cédric Colas, Katja Hofmann, and Pierre-Yves Oudeyer. Teacher algorithms for curriculum learning of deep RL in continuously parameterized environments. In *Conference on Robot Learning (CoRL)*, volume 100 of *Proceedings of Machine Learning Research*, pp. 835–853, 2020a.

Rémy Portelas, Cédric Colas, Lilian Weng, Katja Hofmann, and Pierre-Yves Oudeyer. Automatic curriculum learning for deep RL: A short survey. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2020b.

Rainbow Robotics. RB5-850 collaborative robot official website. https://rainbow-robotics.com/en/products/rb5-850/, 2026. Accessed: 2026-09-21.

A. Rajeswaran, V. Kumar, A. Gupta, G. Vezzani, J. Schulman, E. Todorov, and S. Levine. Learning complex dexterous manipulation with deep reinforcement learning and demonstrations. In *Robotics: Science and Systems (RSS)*, 2018.

Giuseppe Rizzi, Jen Jen Chung, Abel Gawel, Lionel Ott, Marco Tognon, and Roland Siegwart. Robust sampling-based control of mobile manipulators for interaction with articulated objects. *IEEE Transactions on Robotics*, 39(3):1929–1946, 2023.

ROBOTERA. XHAND1 official website. https://www.robotera.com/en/goods1/4.html, 2025. Accessed: 2025-03-17.

Nikita Rudin, David Hoeller, Philipp Reist, and Marco Hutter. Learning to walk in minutes using massively parallel deep reinforcement learning. In *Conference on Robot Learning (CoRL)*, volume 164 of *Proceedings of Machine Learning Research*, pp. 91–100, 2022.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Clemens Schwarke, Mayank Mittal, Nikita Rudin, David Hoeller, and Marco Hutter. RSL-RL: A learning library for robotics research. *arXiv preprint arXiv:2509.10771*, 2025.

Christian Smith, Yiannis Karayiannidis, Lazaros Nalpantidis, Xavi Gratal, Peng Qi, Dimos V. Dimarogonas, and Danica Kragic. Dual arm manipulation—a survey. *Robotics and Autonomous Systems*, 60(10):1340–1353, 2012.

A. B. Tahmaz, R. Prakash, and J. Kober. Impedance primitive-augmented hierarchical reinforcement learning for sequential tasks. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 10973–10979, 2025.

Masaru Uchiyama and Pierre Dauchez. A symmetric hybrid position/force control scheme for the coordination of two robots. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 350–356, 1988.

Filipe Veiga, Herke van Hoof, Jan Peters, and Tucker Hermans. Stabilizing novel objects by learning to predict tactile slip. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5065–5072, 2015.

Y. Wang and E. Johns. One-shot dual-arm imitation learning. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5660–5668, 2025.

Y. Wang, R. Wu, K. Mo, J. Ke, Q. Fan, L. J. Guibas, and H. Dong. AdaAfford: Learning to adapt manipulation affordance for 3D articulated objects via few-shot interactions. In *European Conference on Computer Vision (ECCV)*, pp. 90–107, 2022.

Y. Wang, Z. Wang, M. Nakura, P. Bhowal, C.-L. Kuo, Y.-T. Chen, Z. Erickson, and D. Held. ArticuBot: Learning universal articulated object manipulation policy via large scale simulation. In *Robotics: Science and Systems (RSS)*, 2025a.

Yanzhe Wang, Wei Yu, Hao Wu, Haotian Guo, and Huixu Dong. SA-DEM: Dexterous extrinsic robotic manipulation of non-graspable objects via stiffness-aware dual-stage reinforcement learning. *IEEE Transactions on Automation Science and Engineering*, 23:347–362, 2026.

Yuanfei Wang, Xiaojie Zhang, Ruihai Wu, Yu Li, Yan Shen, Mingdong Wu, Zhaofeng He, Yizhou Wang, and Hao Dong. AdaManip: Adaptive articulated object manipulation environments and policy learning. In *International Conference on Learning Representations (ICLR)*, 2025b.

Z. Wang, Y. Mo, S. Jin, and W. Yuan. DoorBot: Closed-loop task planning and manipulation for door opening in the wild with haptic feedback. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 16058–16064, 2025c.

R. Wu, Y. Zhao, K. Mo, Z. Guo, Y. Wang, T. Wu, Q. Fan, X. Chen, L. Guibas, and H. Dong. VAT-Mart: Learning visual action trajectory proposals for manipulating 3D articulated objects. In *International Conference on Learning Representations (ICLR)*, 2022.

Fanbo Xiang, Yuzhe Qin, Kaichun Mo, Yikuan Xia, Hao Zhu, Fangchen Liu, Minghua Liu, Hanxiao Jiang, Yifu Yuan, He Wang, Li Yi, Angel X. Chang, Leonidas J. Guibas, and Hao Su. SAPIEN: A SimulAted Part-based Interactive ENvironment. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

H. Zhang, B. Eisner, and D. Held. FlowBot++: Learning generalized articulated objects manipulation via articulation projection. In *Conference on Robot Learning (CoRL)*, 2023.

X. Zhang, Y. Wang, R. Wu, K. Xu, Y. Li, L. Xiang, H. Dong, and Z. He. Adaptive articulated object manipulation on the fly with foundation model reasoning and part grounding. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 13032–13042, 2025.

Z. Zhao, Y. Li, W. Li, Z. Qi, L. Ruan, Y. Zhu, and K. Althoefer. Tac-Man: Tactile-informed prior-free manipulation of articulated objects. *IEEE Transactions on Robotics*, 41:538–557, 2025.

A LLM USAGE STATEMENT

We used large language models for proofreading and correcting grammatical errors, assisting with translation, and refactoring the implementation code. All other aspects of this work, including the research ideas, methodology, experiments, and analysis, were carried out entirely by the authors. All LLM-assisted text and code were carefully reviewed and verified by the authors, and we take full responsibility for the final content of this paper.

B QUALITATIVE ROLLOUTS

Figure 7 shows simulated rollouts of BARS-RL on the three objects.

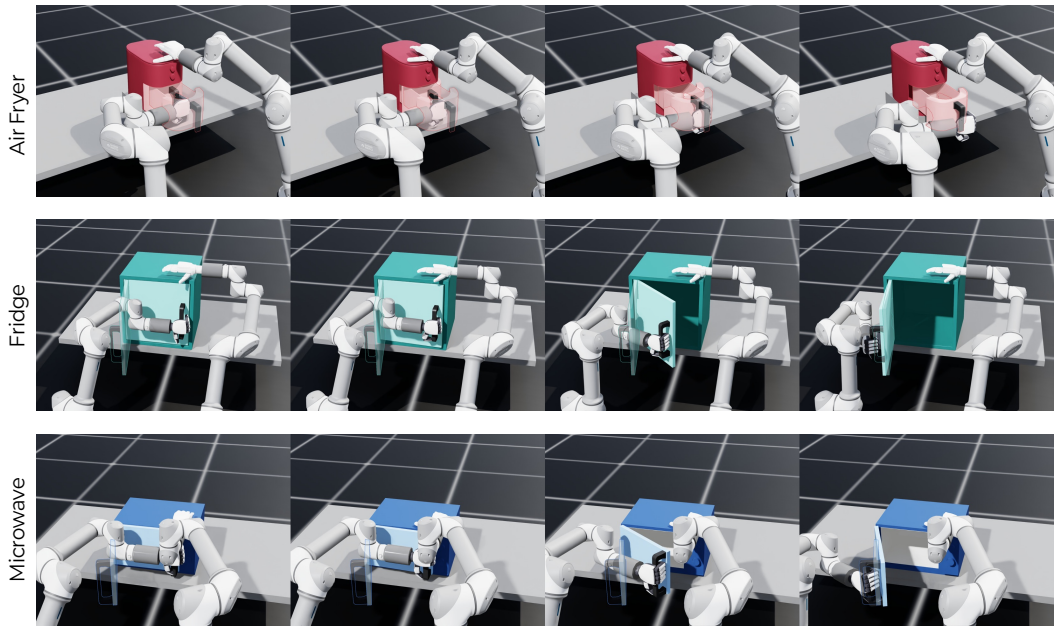

Figure 7: **Rollouts of BARS-RL on the three objects.** Each row shows one object (air fryer, mini fridge, and microwave from top to bottom), with frames ordered from left to right over time.

C IMPLEMENTATION DETAILS

C.1 OBSERVATION AND ACTION SPACES

The actor observes only quantities that are measurable or estimable on the physical robot (Table 4). The object component of the observation is the 10-dimensional vector

$$o_t^{\mathrm{obj}} = [\, q_t,\ q_g,\ q_g - q_t,\ \Delta p_o,\ \Delta\theta_o,\ v_o \,], \tag{10}$$

where $\Delta p_o \in \mathbb{R}^3$ is the translation of the object body since reset, $\Delta\theta_o \in \mathbb{R}$ is its rotation since reset, and $v_o \in \mathbb{R}^3$ is its translational velocity. The critic receives the actor observation together with the physical parameters $\zeta = (F_{\mathrm{res}}, \mu_{\mathrm{b}}, m)$, the base-slip margin δ_{b} of equation 1, and the base-contact force. These quantities are used only during training and never enter the actor.

The policy outputs a 25-dimensional action at 30 Hz (Table 5). Each arm receives a 12-dimensional operational-space command consisting of a relative end-effector pose target and a wrench target, and the hand of the task arm receives one additional grip command. A fixed hybrid force/position selection matrix assigns one translational axis to force control and the remaining five Cartesian directions to position control. For the task arm, the force-controlled axis follows the opening direction $\hat{\mathbf{a}}(q_t)$, and for the support arm, it follows the support-palm normal. A low-level operational-space controller converts the policy commands into joint-torque commands at 120 Hz.

Table 4: Actor observation.

Component	Dim.
Arm and hand proprioception (both arms)	48
End-effector poses (both arms)	14
Six-axis wrist F/T (both arms)	12
Per-pad tactile signals (both hands)	104
Previous action	25
Object vector, equation 10	10
Total	213

Table 5: Policy action.

Component	Dim.
Task arm relative pose target	6
Task arm wrench target	6
Support arm relative pose target	6
Support arm wrench target	6
Task hand grip command	1
Total	25

C.2 Reward

This section gives the closed forms, weights, and thresholds of the reward in equation 8.

Task reward. The progress term is the normalized reduction in articulation error, credited only while the task arm holds the handle,

$$r_{\text{prog}} = \mathbf{1}_{\text{grasp}} \frac{|q_g - q_{t-1}| - |q_g - q_t|}{|q_g - q_0|}, \tag{11}$$

where $\mathbf{1}_{\text{grasp}}$ is active while the task arm maintains its grasp on the handle, which we detect from the fingertip contact sensors: at least one finger reports a contact force above $1\,\text{N}$ (debounced over three control steps, $100\,\text{ms}$, to bridge single-frame contact dropouts) and the nearest fingertip lies within $40\,\text{mm}$ of the handle axis. The success bonus r_{succ} is awarded once, at the step when

$$\mathbf{1}_{\text{succ}} = \mathbf{1}[\,|q_g - q_t| \le \epsilon_q\,]\cdot\mathbf{1}[\,|\dot{q}_t| \le \epsilon_{\dot{q}}\,]\cdot\mathbf{1}_{\text{grasp}}\cdot\mathbf{1}\left[\max_{t' \le t} \|\Delta p_o(t')\| \le \epsilon_p\right]\cdot\mathbf{1}[\,|\Delta\theta_o| \le \epsilon_\theta\,]\cdot\mathbf{1}[\,\delta_{\text{b}} \ge 0\,], \tag{12}$$

and the episode terminates at that step. No hold duration is required; the velocity clause rules out a moving part that coasts through the goal, the displacement clause uses the peak displacement of the object body over the episode, the rotation clause bounds its rotation, and the last clause requires the base-slip margin of equation 1 to be non-negative. For success, $\mathbf{1}_{\text{grasp}}$ uses the stricter, non-debounced contact test (all four fingers loaded on the handle for the microwave and air fryer; thumb plus at least one finger for the mini fridge). The pull term rewards force applied along the opening direction before breakaway, the settling term rewards small articulation error together with low articulation velocity, and the action cost penalizes changes between consecutive actions,

$$r_{\text{pull}} = \mathbf{1}_{\text{grasp}} \cdot \mathbf{1}[\,q_t \text{ short of } q_g\,] \cdot \tanh\left(\frac{\max(F_{\text{pull}}, 0)}{\alpha_{\text{sat}} F_{\text{slip}}}\right), \tag{13}$$

$$r_{\text{settle}} = \mathbf{1}_{\text{grasp}} \cdot \exp\left(-\frac{(q_g - q_t)^2}{\sigma_q^2}\right) \exp\left(-\frac{\dot{q}_t^2}{\sigma_{\dot{q}}^2}\right), \tag{14}$$

$$c_{\text{act}} = \frac{1}{|\mathcal{A}|} \|a_t - a_{t-1}\|_2^2, \tag{15}$$

where F_{pull} is the pulling force of equation 4, $F_{\text{slip}} = \mu_{\text{b}}\, m\, g$ is the per-episode slip threshold of Section 3, and $|\mathcal{A}|$ is the action dimension.

Stabilization terms. The base-slip cost c_{base} is defined in equation 6 with safety factor κ, and the support reward r_{sup} in equation 7 with length scale σ_o. The indicators in equation 7 compare the applied forces with fixed thresholds,

$$\mathbf{1}_{\text{pulling}} = \mathbf{1}_{\text{grasp}} \cdot \mathbf{1}\left[F_{\text{pull}} \geq \bar{F}_{\text{pull}} \ \vee \ \dot{q}_t \geq \bar{\dot{q}}\right], \qquad \mathbf{1}_{\text{pressing}} = \mathbf{1}\left[F_{\text{s}}^{n} \geq \bar{F}_{\text{press}}\right], \tag{16}$$

where F_{s}^{n} is the normal force of the support arm in equation 2. The pulling-force threshold is a fraction of the per-episode slip threshold, $\bar{F}_{\text{pull}} = \alpha_{\text{pull}} F_{\text{slip}}$, and the velocity clause keeps the indicator active through the release of the closure, when the moving part starts to move and the wrist load drops to joint friction. The displacement cost penalizes translation and rotation of the object body, the overgrip cost penalizes grip force beyond what the handle load requires, and the off-axis cost penalizes wrist force that does not contribute to opening,

$$c_{\text{body}} = \frac{\|\Delta p_o\|}{\sigma_p} + \frac{|\Delta \theta_o|}{\sigma_\theta}, \tag{17}$$

$$c_{\text{overgrip}} = \frac{1}{F_{\text{grip}}^{\max}} \max\left(F_{\text{grip}} - \frac{F_{t,\text{handle}}}{\mu_{\text{tip}}} - \delta_{\text{safe}},\, 0\right), \tag{18}$$

$$c_{\text{offaxis}} = \frac{\|\bar{\mathbf{f}}_{\text{w}}(t) - F_{\text{pull}}\, \hat{\mathbf{a}}(q_t)\|}{\sigma_{\text{off}}}, \tag{19}$$

where F_{grip} and $F_{t,\text{handle}}$ are the summed normal and tangential contact forces between the fingertips of the task-arm hand and the handle, so $F_{t,\text{handle}}/\mu_{\text{tip}}$ is the grip the current handle load requires and δ_{safe} is a free band above it; $\bar{\mathbf{f}}_{\text{w}}(t)$ and $\hat{\mathbf{a}}(q_t)$ are defined in equation 4. Each cost is clipped at its full scale (c_{body} at 2, c_{act} at 4, the rest at 1), which removes the tail of diverged simulation states without changing the terms inside the operating range.

Weights and parameters. Table 6 lists the weights of equation 8; the stage-dependent weights w_{stab} and w_{sup} are given in Appendix C.4. Table 7 lists the remaining parameters.

Table 6: Reward weights.

Weight	Applies to	Value
λ_{prog}	r_{prog}	2.0
λ_{succ}	r_{succ}	10
λ_{pull}	r_{pull}	0.02
λ_{settle}	r_{settle}	0.02
λ_{sup}	r_{sup}	0.02
λ_{time}	per step	0.003
λ_{act}	c_{act}	0.01
λ_{body}	c_{body}	0.02
λ_{base}	c_{base}	0.04
$\lambda_{\text{overgrip}}$	c_{overgrip}	0.04
λ_{offaxis}	c_{offaxis}	0.04

Table 7: Reward parameters. Angle-based values apply to the revolute doors.

Symbol	Description	Value
κ	Safety factor of the base-slip cost, equation 6	0.7
σ_o	Length scale of the support reward, equation 7	30 mm
α_{pull}	Pulling-force threshold $\bar{F}_{\text{pull}} = \alpha_{\text{pull}} F_{\text{slip}}$, equation 16	0.4
$\bar{\dot{q}}$	Articulation-velocity clause of $\mathbf{1}_{\text{pulling}}$, equation 16	0.1 rad/s
$\bar{F}_{\text{press}}$	Support normal-force threshold for $\mathbf{1}_{\text{pressing}}$, equation 16	5 N
ϵ_q	Articulation tolerance for success, equation 12	5°
$\epsilon_{\dot{q}}$	Articulation velocity tolerance for success, equation 12	0.1 rad/s
ϵ_p	Object-body displacement tolerance for success, equation 12	30 mm
ϵ_θ	Object-body rotation tolerance for success, equation 12	5°
α_{sat}	Saturation of r_{pull} in units of F_{slip}, equation 13	1.4
σ_q	Angle scale of r_{settle} ($= 2\epsilon_q$), equation 14	10°
$\sigma_{\dot{q}}$	Velocity scale of r_{settle} ($= 4\epsilon_{\dot{q}}$), equation 14	0.4 rad/s
σ_p	Translation scale of c_{body}, equation 17	50 mm
σ_θ	Rotation scale of c_{body}, equation 17	5°
$F_{\text{grip}}^{\max}$	Grip-force normalization of c_{overgrip}, equation 18	50 N
δ_{safe}	Free grip band above the required grip, equation 18	5 N
σ_{off}	Force scale of c_{offaxis}, equation 19	30 N

C.3 Training

We optimize the actor and the critic of the asymmetric actor–critic with Proximal Policy Optimization (PPO) (Schulman et al., 2017). Both networks are multilayer perceptrons. The actor receives the observation of Appendix C.1, and the critic additionally receives the privileged inputs listed there (Pinto et al., 2018). Table 8 lists the training hyperparameters.

Table 8: PPO hyperparameters.

Hyperparameter	Value
Parallel environments	2048
Rollout length (steps per environment)	32
Epochs per update	5
Minibatch size	16384
Learning rate	1e-3
Optimizer	Adam (Kingma & Ba, 2015)
Activation	ELU (Clevert et al., 2016)
Discount γ	0.99
GAE λ	0.95
Clip ratio ϵ	0.2
Entropy coefficient	0.005
Value loss coefficient	1.0
Actor MLP hidden sizes	[512, 256, 128]
Critic MLP hidden sizes	[512, 256, 128]

C.4 Curriculum

Table 9 lists the settings of each curriculum stage in Algorithm 1. Observation noise (Table 10) is enabled only in the final phase.

Table 9: Curriculum stages. α_s is the probability that the object body is anchored, $\rho \sim \mathcal{U}(\mathcal{R}_s)$, and τ_s is the success threshold for advancing to the next stage.

Phase	Stage	α_s	$\mathcal{R}_s$	w_{stab}	w_{sup}	Obs. noise	τ_s
0	0	1	$\{0\}$	0	0	off	0.7
1	1	0.75	$[0.1, 0.5]$	0.5	1	off	0.3
1	2	0.50	$[0.1, 0.5]$	0.5	1	off	0.5
1	3	0.25	$[0.1, 0.5]$	0.5	1	off	0.7
1	4	0	$[0.1, 0.5]$	0.5	1	off	0.8
2	5	0	$[0.4, 2.0]$	1	1	on	none

C.5 Domain Randomization

Table 10 lists the randomized parameters in the final phase.

Table 10: Randomization parameters.

Group	Parameter	Distribution	Operation	Sampled
Object	Mass m	$\mathcal{U}(0.85, 1.15)$	Scaling	per env
Object	Base friction μ_{b}	$\mathcal{U}(0.15, \mu_{\text{b}}^{\max})$	Absolute	per env
Object	Breakaway force F_{res}	$\rho\, F_{\text{slip}}$, $\rho \sim \mathcal{U}(0.4, 2.0)$	Sampling	per episode
Object	Goal q_g (revolute)	$\mathcal{U}(25^\circ, 80^\circ)$	Sampling	per episode
Object	Goal q_g (prismatic)	$\mathcal{U}(0.28, 0.89)$ of stroke	Sampling	per episode
Robot	Fingertip friction	$\mathcal{U}(0.8, 1.2)$	Scaling	per env
Robot	Palm friction	$\mathcal{U}(0.8, 1.2)$	Scaling	per env
Observation	Wrist F/T	$\mathcal{U}(-1.0, 1.0)$ N	Additive	per step
Observation	Tactile (task arm)	$\mathcal{U}(-0.2, 0.2)$	Additive	per step
Observation	Tactile (support arm)	$\mathcal{U}(-0.5, 0.5)$	Additive	per step