Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Mechanical Threshold and Divergence Law for Collective Transport: Closed-Form Theory, Rigid-Body Simulation, and Decentralized Reinforcement Learning

Milad Tahavor

miladtahavor69@gmail.com

Iran University of Science and Technology

Mohsen Rabdoost Motlagh

Imam Hossein University

Research Article

Keywords:

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11415783/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Mechanical Threshold and Divergence Law for Collective Transport: Closed-Form Theory, Rigid-Body Simulation, and Decentralized Reinforcement Learning

Milad Tahavor*[1] and Mohsen Rabdoost Motlagh[2]

[1]Department of Mathematics, Iran University of Science and Technology, Tehran, Iran
[2]Department of Artificial Intelligence, Imam Hossein University, Tehran, Iran.

Abstract

We study collective transport, in which N simple robots jointly push a rigid body toward a goal. From elementary rigid-body dynamics and Coulomb friction we derive a closed-form mechanical threshold $N^* = \mu mg/F_{\max}$ below which the body cannot leave rest, and show that the time to reach a goal at distance d diverges as $T(N) \sim (N - N^*)^{-1/2}$. A Newton–Euler simulation with stick–slip friction recovers $\alpha = 0.500$ to within 0.1% in five physical regimes, with errors that decrease roughly in proportion to the time step. We then test whether the law survives decentralized, communication-free control with noisy local sensing, and derive a closed-form noise-coherence correction $N^*_{\text{eff}} = N^* e^{\sigma^2/2}$ for controllers that do not compensate their sensing errors. In the simple setting, the central methodological finding is a control: an untrained *identity* policy (applied force direction equal to the sensed goal direction) already attains 100% success, $\alpha \approx 0.49$ and $N_0 \approx 29.8$. Consequently, in that setting success and exponent recovery by a trained policy are not evidence of learning. With the goal direction randomized, 20 independently trained policies converge to this identity map (sensitivity slope 1.00 ± 0.00) and the correction matches the observed threshold within 0.5% at every noise level tested ($\sigma_{\text{step}} \in [0, 0.5]$ rad); with a fixed goal, trained policies instead partly memorize the goal and fall below the correction. Masked-input negative controls behave as controls only when the goal is randomized. Rotational dynamics do not alter the translational law. Finally, where learning has something to learn, i.e. in a harder task with large sensing error and an offset centre of mass, we train decentralized policies with MAPPO (residual-angle action, threshold curriculum, recurrent memory); they far exceed standard MARL baselines and slightly exceed tuned designed controllers at the training condition (success 0.813 ± 0.025 over 10 seeds versus 0.758 and 0.781), but are worse than those controllers at low noise, and the margin over the strongest learned baselines is not significant on fresh seeds. All results are simulation-only.

1 Introduction

Collective transport – multiple simple agents jointly moving a payload too large for any one of them – is a canonical benchmark for multi-agent robotics and a close analogue of collective transport by ant colonies. A substantial literature addresses *how* to control such systems, from centralized

*Corresponding author: miladtahavor69@gmail.com

planning to fully decentralized, communication-free policies trained with multi-agent reinforcement learning (MARL); see the review by Farivarnejad and Berman [2] for a survey spanning three decades of control strategies. A prominent recent example is Heuthe et al. [1], who trained up to 200 individually controlled microrobots with MARL using counterfactual credit assignment, and reported that the resulting swarm is *empirically robust to variations in group size*, without giving a quantitative account of why, or of what happens as group size approaches the point where transport becomes impossible.

This is the gap we address. Two simple facts about collective transport appear to be absent from the literature in combined form: (i) there is an exact, closed-form minimum number of agents N^* required to move a body against friction at all, derivable from elementary force balance; and (ii) as N approaches N^* from above, the time required to complete a transport task diverges as a power law with a specific, predictable exponent. We derive this law, verify it in a rigid-body simulation, and then test whether it survives the transition from an idealized centralized controller to a decentralized, noise-corrupted, learned one.

That last test turns out to require care. A decentralized policy whose output is parametrized as a bounded correction to its own sensed goal direction already satisfies the law *without any learning*, so a high success rate and an exponent near 1/2 do not by themselves show that a learned controller has acquired anything. We therefore report every decentralized result against an untrained identity baseline, randomize the goal direction so that memorization is impossible, and state which conclusions survive these controls. The finding is specific to that setting: when sensing error is small relative to what the identity map can absorb, there is nothing left to learn. We therefore also study a harder task (Section 9), with large sensing error and an offset centre of mass, where a controller can gain by inferring its own bias and by cancelling torque jointly; there, a MAPPO policy with a residual-angle action, curriculum and memory clearly outperforms standard MARL baselines and marginally outperforms tuned designed controllers at the training condition, while designed controllers remain better at low noise.

Table 1: Positioning of this work against the closest prior work, along the dimensions studied here. ✓: addressed; ∼: partly addressed; ×: not addressed (for prior work: not reported in the cited work as we read it, which does not mean it is impossible); –: not applicable or not assessed. Entries for prior work follow the cited abstracts and our reading; they are not an exhaustive audit. Our own ∼ (learned-versus-designed comparison: small and partly not significant on fresh seeds, Section 9) is intentional. The table covers simulation-level dimensions only; this work contains no physical experiments (Section 10 is a second simulator).

Dimension	Survey [2]	MARL swarm [1]	Ants [5]	Mechanical [10]	This work
Exact force-balance threshold N^*	×	×	∼	×	✓
Predicted divergence exponent of completion time	×	×	×	×	✓
Analytical effect of sensing noise on the threshold	×	×	×	×	✓
Decentralized, communication-free control	∼	✓	–	✓	✓
Learned (MARL) policy	×	✓	×	×	✓
Untrained control to test whether learning matters	×	×	×	×	✓
Learned vs. designed controllers, with statistics	×	×	×	×	∼
Dependence on group size studied	×	✓	✓	–	✓
Cross-check in an independent physics engine	–	–	–	–	✓

Table 1 summarizes where this work sits relative to prior work; it is a positioning aid and not a claim of superiority, since the compared works pursue different goals (the experimental studies establish what our simulation-only study cannot, namely physical validation).

Figure 1 illustrates the setting: N robots attached to the boundary of a payload push it toward a goal, each along its own noisy estimate of the goal direction, and the payload accelerates uniformly once N exceeds N^*.

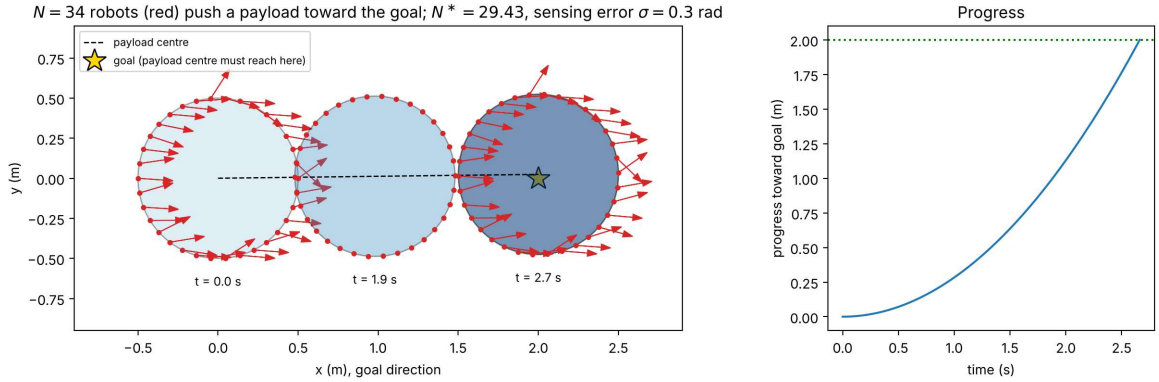


Figure 1: The setting, from one run of the physics-engine check of Section 10 ($N = 34$ robots, $N^* = 29.43$, sensing error $\sigma = 0.3$ rad). Left: snapshots of the payload (disc) and the robots (red dots) with the direction of each robot's force (red arrows; the scatter reflects the sensing error); the star marks the goal, which the payload centre must reach. Right: progress toward the goal, which grows as t^2 because the net force is constant.

Our contributions are:

1. A closed-form threshold N^* and divergence exponent $\alpha = 1/2$ under Coulomb friction (Section 3); a noise-coherence correction $N^*_{\text{eff}} = N^* e^{\sigma^2/2}$ for controllers that do not compensate sensing error; and a proposition showing that, for world-frame forces, rotational dynamics decouple from the translational law (Section 3.2).

2. Verification in a Newton–Euler simulation with stick–slip friction across five physical regimes ($\alpha = 0.500$ to 0.1%), a time-step convergence study, and extensions to anisotropic and Stribeck friction, where the power law is replaced by a logarithmic divergence (Section 4).

3. An untrained identity-policy control for decentralized transport, which attains the success (1.000), exponent ($\alpha \approx 0.49$) and threshold location ($N_0 \approx 29.8$) of the trained random-goal policies (Section 5.3).

4. A 20-seed study with randomized goal direction in which trained policies converge to the identity map and the noise-coherence correction predicts the fitted threshold within 0.5% for all 20 policies at all five noise levels (Sections 5.4–6); and the contrasting fixed-goal result, where policies partly memorize and beat the correction.

5. Ablations (5 seeds) in which, with randomized goal, masked-input controls behave as negative controls and no training-offset effect is observed (Section 7), together with a full per-seed table (Section 8).

6. A MAPPO study in a task where learning has something to learn (large sensing error, offset centre of mass), with a residual-angle action, threshold curriculum and recurrent memory, ten training seeds, a pre-registered confirmatory comparison, and an explicit account of where it wins, ties and loses against learned baselines and tuned designed controllers (Section 9).

7. An independent physics-engine check (PyBullet) of the threshold, the divergence exponent and the sensing-error shift, which reproduces the first two to 0.2% and the third to first order, with a $\sim 2\%$ finite-N and rotation-coupling deviation at the largest sensing error (Section 10).

Implementation and provenance of results. Sections 3 and 4 (closed-form theory and centralized-control rigid-body verification, Tables 2–7) use the original full Newton–Euler simulator. The numerical-convergence study (Section 4.4), the additional simulation figures, and all decentralized-control results of Sections 5–8 use a second, independently written implementation (batched PyTorch; translational and full Newton–Euler variants), written from the model description given here; open design choices (loss function, correction bound, success window, friction relaxation) are stated in Section 5.2. The learning study of Section 9 uses a third environment (batched PyTorch; rigid-body disc with an offset centre of mass and large sensing error), and Section 10 reports an independent physics-engine check. Earlier decentralized-control results (fixed goal direction; ES and PPO comparisons) are kept in Appendix B with a discussion of why they are not used as primary evidence.

2 Related Work

Collective transport control. Farivarnejad and Berman [2] survey control strategies for collective transport, categorizing them by whether sensing, computation, and communication are centralized or autonomous to each robot. Heuthe et al. [1] demonstrate that MARL with counterfactual rewards produces a decentralized swarm-transport policy for up to 200 microrobots that is empirically robust to group-size variation, malfunctioning units, and environmental noise, but the robustness claim there is observational, not tied to a mechanical threshold or a predicted functional form. To our knowledge no prior work derives an exact force-balance threshold N^* together with a predicted divergence exponent and tests whether a decentralized, communication-free policy reproduces it.

Collective transport in biological and physical systems. Our mechanical-threshold framing has a close analogue in the study of cooperative transport by ants. Feinerman et al. [5] show that ants cooperatively hauling a load can be mapped onto an Ising-like model, with the group residing near a transition between an uncoordinated and a coordinated collective state controlled by group size, conceptually parallel to our N^* though derived from a very different (statistical-mechanical, not force-balance) argument, and without an explicit divergence-time exponent. Heckenthaler et al. [6] connect this phenomenon to self-propelled particle physics, and Chatterjee et al. [7] provide direct experimental evidence that ant collectives show maximal responsiveness at a critical group size. McCreery and Breed [9] review the proximate behavioral mechanisms ants use to coordinate such transport, and Berman et al. [8] bridge this literature to swarm robotics by modeling and experimentally validating ant-inspired group retrieval strategies. Most directly relevant to our mechanical framing, Arbel et al. [10] showed that cooperative transport can emerge in swarms of robots with *no* sensing or communication at all, purely from the mechanical design of individual agents, demonstrating experimentally that mechanics alone can drive collective transport. Our work is complementary: where Arbel et al. show mechanics alone can produce transport, we derive the exact mechanical threshold and divergence law governing when and how fast that transport occurs, and test how a decentralized controller interacts with it.

Multi-agent reinforcement learning and parameter sharing. Parameter sharing across homogeneous agents, the design we use throughout Sections 5–8, was established as an effective MARL baseline by Gupta et al. [12] and is used (often implicitly) by decentralized actor-critic

methods including MADDPG [13]; Terry et al. [11] formalize why and when parameter sharing converges to optimal policies. For policy optimization in Sections 5–8 we use direct gradient-based optimization through a differentiable simulator rather than PPO [15] or black-box Evolution Strategies [14]; earlier comparisons against these two methods are reported in Appendix B. The harder task of Section 9, where the dynamics are not used for gradients, is trained with PPO-based MARL (IPPO/MAPPO).

Differentiable simulation for control. Our training procedure exploits the fact that rigid-body dynamics with Coulomb friction is differentiable almost everywhere, following work on differentiable physics engines such as DiffTaichi [3] and Brax [4], which showed that backpropagating through simulated trajectories can give lower-variance gradients than black-box or sampling-based estimators when the dynamics are known.

3 Theoretical Framework

3.1 Setup

A rigid body of mass m rests on a surface with Coulomb friction coefficients μ_s (static) and $\mu_k \leq \mu_s$ (kinetic). N robots, each capable of exerting a force of magnitude at most $F_{\max}$, are attached to points on the body's boundary and steer their forces toward a common goal direction $\hat{g}$. We ask: (i) what is the minimum N for which the body can leave rest, and (ii) how does the time to reach a goal at distance d behave as $N \to N^{*+}$.

Theorem 1 (Mechanical threshold)**.** *Let N robots each apply force of magnitude $F_{\max}$ in a common direction $\hat{g}$ to a body of mass m resting on a surface with static friction coefficient μ_s. The body remains at rest for all N such that*

$$NF_{\max} \leq \mu_s mg,$$

and can leave rest once $N > N^* := \dfrac{\mu_s mg}{F_{\max}}$.

Proof. By Newton's second law applied to the body's center of mass, the net applied force is $F_{\text{applied}} = NF_{\max}$ in direction $\hat{g}$. Static (Coulomb) friction opposes the applied force up to a maximum magnitude $\mu_s mg$ (normal force mg on a horizontal surface). If $F_{\text{applied}} \leq \mu_s mg$, friction exactly cancels the applied force ($F_{\text{net}} = 0$) and the body remains at rest by definition of static friction. If $F_{\text{applied}} > \mu_s mg$, static friction cannot supply enough opposing force, the body begins to slip, and kinetic friction (with coefficient $\mu_k \leq \mu_s$) applies thereafter. Solving $NF_{\max} = \mu_s mg$ for N gives the stated threshold. □

Theorem 2 (Divergence law)**.** *Suppose $\mu_s = \mu_k = \mu$ (no static/kinetic friction gap) and $N > N^* = \mu mg/F_{\max}$. Under constant net force $F_{net}(N) = NF_{\max} - \mu mg$, the time to travel distance d from rest is*

$$T(N) = \sqrt{\frac{2d\,m}{F_{\max}\,(N - N^*)}},$$

so that as $N \to N^{*+}$, $T(N) \sim (N - N^*)^{-\alpha}$ with $\alpha = 1/2$.

Proof. With $\mu_s = \mu_k = \mu$, once $N > N^*$ the body accelerates uniformly with $a(N) = F_{\text{net}}(N)/m = F_{\max}(N - N^*)/m$, a positive constant independent of time (since kinetic friction has fixed magnitude μmg once moving, matching the value used to define N^*). Starting from rest, $d = \frac{1}{2}a(N)T(N)^2$, so $T(N) = \sqrt{2d/a(N)} = \sqrt{2dm/(F_{\max}(N - N^*))}$, which is exactly of the claimed form with $\alpha = 1/2$. □

Corollary 1. *The exponent $\alpha = 1/2$ is independent of m, μ, $F_{\max}$, and d; these parameters only set the location N^* of the threshold and the prefactor $C = \sqrt{2dm/F_{\max}}$, not the divergence exponent itself.*

Theorem 2 is, mechanically, an elementary consequence of constant-acceleration kinematics; its interest is not in analytical depth but in (a) its exactness as a falsifiable, parameter-independent prediction, and (b) whether it survives the less idealized settings of rigid-body simulation and decentralized, noisy control, which we test in the remainder of the paper. Because Theorem 2 is nearly an identity in a deterministic constant-force simulation, the checks in Section 4 test the *implementation* (integration, friction switching, fitting) rather than the physics.

3.2 Extensions: noisy force alignment and rotational decoupling

Theorem 1 assumes every robot applies its force exactly along the goal direction. We now treat robots that sense the goal direction only noisily.

Proposition 1 (Noise-corrected mechanical threshold)**.** *Suppose each of N robots applies force of magnitude $F_{\max}$ at an angle θ_i from the true goal direction, where the θ_i are independent and identically distributed $\mathcal{N}(0, \sigma^2)$ and are* not *compensated by the controller (the applied direction equals the sensed direction). Then*

$$\mathbb{E}\big[F_{\parallel}\big] = NF_{\max}\,\mathbb{E}[\cos\theta] = NF_{\max}\,e^{-\sigma^2/2}, \qquad \mathbb{E}[F_{\perp}] = 0, \quad \mathrm{Var}(F_{\perp}) = NF_{\max}^2\,\tfrac{1-e^{-2\sigma^2}}{2},$$

where $F_{\parallel}$ and $F_{\perp}$ are the net force components along and transverse to the goal direction. Equating the expected parallel force to the friction threshold μmg gives the noise-corrected threshold

$$N^*_{\mathit{eff}}(\sigma) = N^*\,e^{\sigma^2/2},$$

which exceeds N^* for any $\sigma > 0$ and recovers N^* as $\sigma \to 0$.

Proof. $\mathbb{E}[\cos\theta] = \mathrm{Re}\;\mathbb{E}[e^{i\theta}] = e^{-\sigma^2/2}$ for $\theta \sim \mathcal{N}(0, \sigma^2)$ is exact. Likewise $\mathbb{E}[\sin\theta] = 0$ and $\mathbb{E}[\sin^2\theta] = (1 - \mathbb{E}[\cos 2\theta])/2 = (1 - e^{-2\sigma^2})/2$; independence gives the variance of the transverse sum. The threshold follows from Theorem 1 applied to the expected parallel force. □

Remark 1 (Scope)**.** *Proposition 1 describes controllers that do not compensate sensing error. It is* not *a lower bound on what a learned controller can achieve: if a policy can infer the goal direction from information other than the noisy reading (for instance, by memorizing a fixed goal), it can cancel the angular error and approach N^*; Section 6 shows this occurs in our fixed-goal experiments. The statement is also an expectation; for finite N the net force fluctuates, which we do not analyze. In our setting $\sigma^2 = \sigma^2_{\mathrm{bias}} + \sigma^2_{\mathrm{step}}$ ($\sigma_{\mathrm{bias}} = 0.10$ rad, fixed per scenario, and per-step jitter σ_{step}), treating the two sources as independent and each robot's force as sampled from the combined distribution.*

Proposition 2 (Rotational decoupling)**.** *Suppose every robot applies a force whose direction is fixed in the world frame (a function of the goal direction and its sensing error only), friction depends only on the translational velocity of the center of mass, and rotation is resisted by at most a viscous torque. Then the center-of-mass trajectory $x(t)$ is independent of the body's angle $\theta(t)$ and angular velocity $\omega(t)$, so Theorems 1–2 and Proposition 1 hold unchanged in the full Newton–Euler dynamics.*

Proof. The center-of-mass equation $m\ddot{x} = \sum_i F_i + F_{\text{fric}}(\dot{x})$ contains θ and ω only through the F_i, which by assumption do not depend on them. The rotational equation $I\dot{\omega} = \sum_i r_i(\theta) \times F_i - c\,\omega$ is then driven by x-independent terms. □

Rotation can therefore influence the translational law only if the policy observes quantities that depend on the body angle (such as its attachment angle) or if the physical model couples rotation and translation (body-fixed forces, friction torque). We test the first possibility empirically in Section 5.4; the second would require a different physical model than the one studied here.

4 Rigid-Body Simulation

4.1 Physics model

We simulate full 2D Newton–Euler dynamics: a rigid body with mass m, moment of inertia I, and N robots attached at points evenly spaced on a boundary circle of radius r. Each robot applies a force of magnitude $F_{\max}$ toward the goal. We use a stick–slip Coulomb friction model: while the body's speed is below a numerical threshold, friction exactly cancels the applied force if the applied force is below $\mu_s mg$ (the body remains stuck), and otherwise kinetic friction $\mu_k mg$ opposes the direction of the applied force; once moving, kinetic friction opposes the velocity direction. Rotational dynamics include torque from off-center forces and a small viscous damping term to prevent runaway spin. We integrate with explicit Euler steps at $dt = 2 \times 10^{-3}\,\text{s}$. (For evenly spaced robots applying identical forces the net torque vanishes, so in the ideal centralized setting rotation plays no role; see Proposition 2.)

4.2 Verification of the divergence law

Figure 2 shows time-to-goal as a function of N for the baseline parameter set ($m = 5\,\text{kg}$, $I = 2\,\text{kg}\,\text{m}^2$, $r = 0.5\,\text{m}$, $\mu_s = \mu_k = 0.6$, $F_{\max} = 1\,\text{N}$, $d = 5\,\text{m}$, giving $N^* = 29.43$). Simulated times match the closed-form prediction almost exactly; fitting $T(N) = C(N - N^*)^{-\alpha}$ against the *exact continuous* threshold recovers $\alpha = 0.500$. Fitting against the *integer-rounded* $\lceil N^* \rceil$ instead introduces a systematic bias ($\alpha_{\text{fit}} = 0.413$ in our data), not because the physics is wrong but because the true singularity sits at the continuous value. We flag this because it is an easy way to be misled by an otherwise-correct simulation.

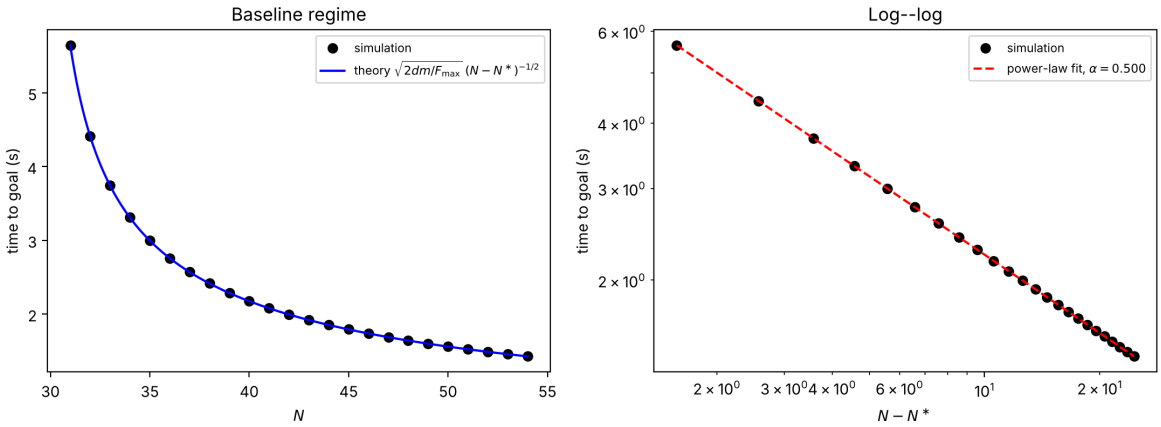


Figure 2: Left: time to goal vs. N, theory (line) vs. simulation (points), baseline regime. Right: log–log plot with power-law fit, recovering $\alpha = 0.500$.

Figure 3 shows the underlying trajectories. For every N above threshold, position and speed follow $at^2/2$ and at with $a = F_{\max}(N - N^*)/m$; a body pushed by $N = 29 < N^*$ robots never moves (static friction cancels the applied force), whereas $N = 30$ slips immediately. Figures 2, 3, 4, 5, 6 and 7 were produced with the independently written re-implementation of the simulator (see the provenance note); their fitted exponents differ from the original simulator's values in the fourth decimal (Table 2).

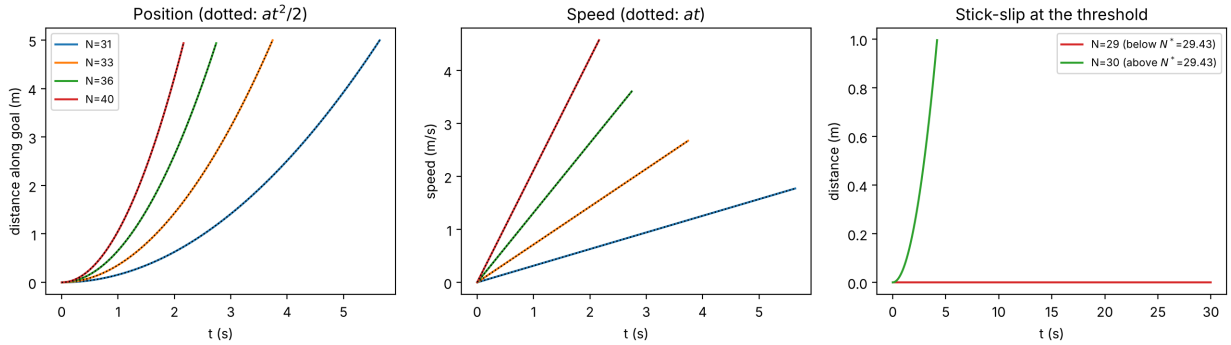


Figure 3: Trajectories in the baseline regime (re-implemented simulator). Left and centre: distance and speed along the goal direction for $N = 31, 33, 36, 40$ (solid) against the constant-acceleration solution (dotted). Right: below threshold ($N = 29$) the body stays at rest; just above ($N = 30$) it slips.

4.3 Generalization across physical regimes

We repeated the sweep across five regimes spanning roughly an order of magnitude in mass, friction, robot strength, and body size (Table 2). In every regime the fitted exponent is $\alpha = 0.500$ to within the reported precision, confirming Corollary 1. The reported SE is the standard error of the least-squares fit; the simulation is deterministic, so it measures fit quality, not variability across runs.

Table 2: Divergence exponent across five independent physical regimes (ideal, centralized control, no sensing noise).

Regime	m (kg)	μ	$F_{\max}$ (N)	r (m)	N^*	α_{fit}	SE
Baseline	5.0	0.6	1.0	0.5	29.43	0.5004	0.0001
Heavy / low friction	10.0	0.4	1.5	0.7	26.16	0.5004	0.0001
Light / high friction	2.0	0.8	0.4	0.3	39.24	0.5002	0.0001
Large body	20.0	0.5	2.0	1.0	49.05	0.5004	0.0001
Low friction	5.0	0.3	1.0	0.5	14.71	0.5004	0.0001

Figure 4 repeats the five-regime sweep with the re-implemented simulator ($I = 2\,\text{kg}\,\text{m}^2$ in all regimes; $N = \lceil N^* \rceil + 1, \ldots, \lceil N^* \rceil + 16$): the fitted exponents are 0.4995–0.4999, the largest deviation of any simulated T from the closed form is 0.10–0.17%, and after rescaling by $\sqrt{2dm/F_{\max}}$ all five regimes fall on a single $-1/2$ line. Figure 5 sweeps N and μ at fixed m and $F_{\max}$: the time to goal grows without bound along the line $N^* = \mu mg/F_{\max}$.

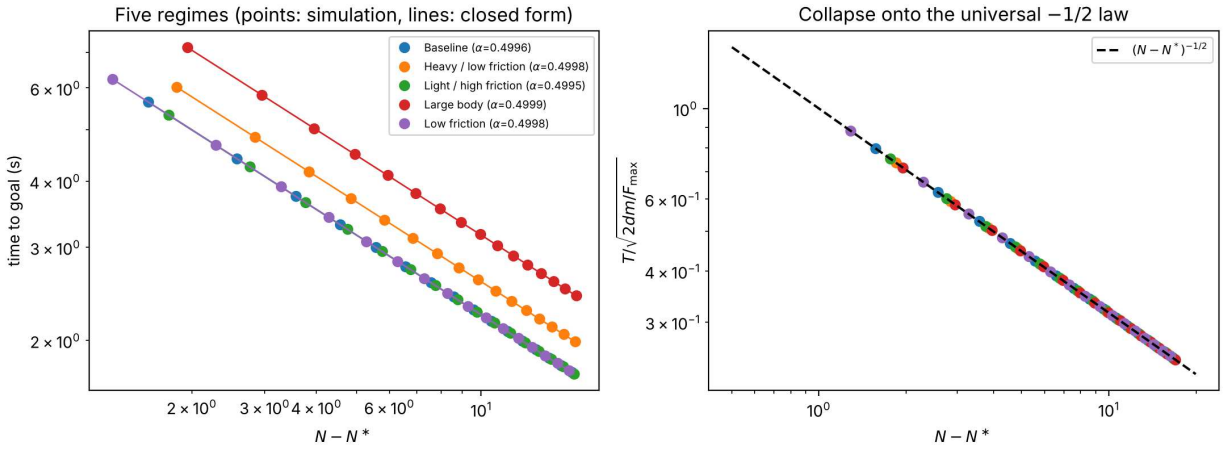


Figure 4: Left: time to goal against $N - N^*$ in the five regimes of Table 2 (points: re-implemented simulator; lines: closed form); fitted α in the legend. Right: the same data divided by the prefactor $\sqrt{2dm/F_{\max}}$, collapsing onto $(N - N^*)^{-1/2}$.

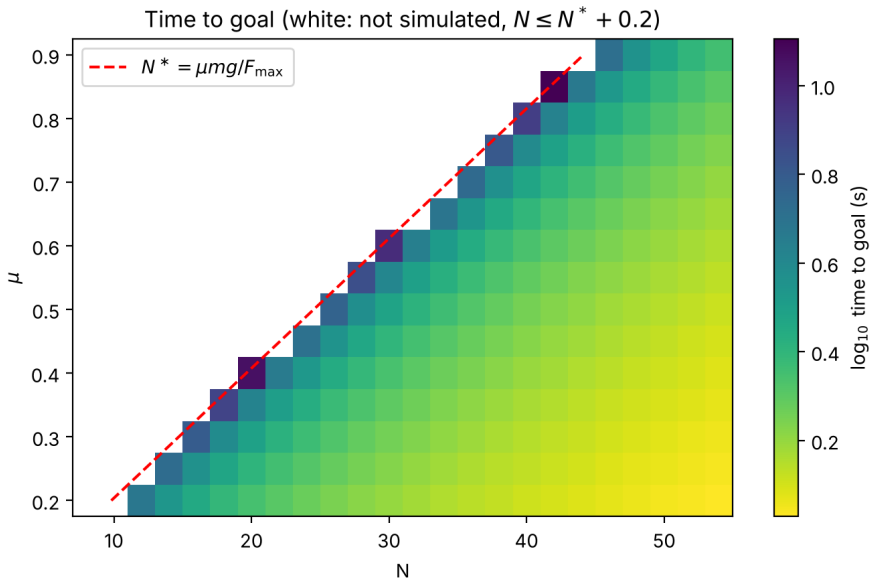


Figure 5: Time to goal (colour, $\log_{10}$ seconds) as a function of N and μ ($m = 5\,\mathrm{kg}$, $F_{\max} = 1\,\mathrm{N}$, $d = 5\,\mathrm{m}$, $\mu_s = \mu_k = \mu$). Cells with $N \le N^* + 0.2$ (white) were not simulated; the stuck behaviour below threshold is shown in Figure 3. Dashed: $N^* = \mu mg/F_{\max}$.

4.4 Numerical convergence

Near the threshold the acceleration is small and discretization error could in principle distort the exponent. We repeated the baseline sweep ($N = 30, \ldots, 46$) with the independently written implementation of the full Newton–Euler simulator at five time steps (Table 3). The relative error of T against the closed form decreases with dt, roughly in proportion to it as expected for explicit Euler (Figure 6); the errors at 4×10^{-3} and 2×10^{-3} s are identical, which we attribute to the quantization of the goal-crossing time to a multiple of dt but did not verify. The exponent converges to 0.5000; fitting the threshold as a free parameter returns $N_0 = 29.430$ at every dt. At the step used in Table 2 (2×10^{-3} s) this implementation gives $\alpha = 0.4997$, within 0.1% of the tabulated 0.5004; the small difference reflects unspecified details of integration and goal-crossing detection.

Table 3: Time-step convergence of the full Newton–Euler simulator (baseline regime, $N = 30$–46). α_{fit} uses the exact N^*; "free" fits N^* as a third parameter. Relative errors compare simulated and closed-form T.

dt (s)	α_{fit}	α_{fit} (free N_0)	N_0 (free)	Rel. err. T ($N = 30$)	Rel. err. T ($N = 46$)
4.0×10^{-3}	0.4994	0.4994	29.430	+0.023%	+0.167%
2.0×10^{-3}	0.4997	0.4996	29.430	+0.023%	+0.167%
1.0×10^{-3}	0.4999	0.4998	29.430	+0.012%	+0.052%
5.0×10^{-4}	0.4999	0.4999	29.430	+0.007%	+0.023%
2.5×10^{-4}	0.5000	0.4999	29.430	+0.002%	+0.009%

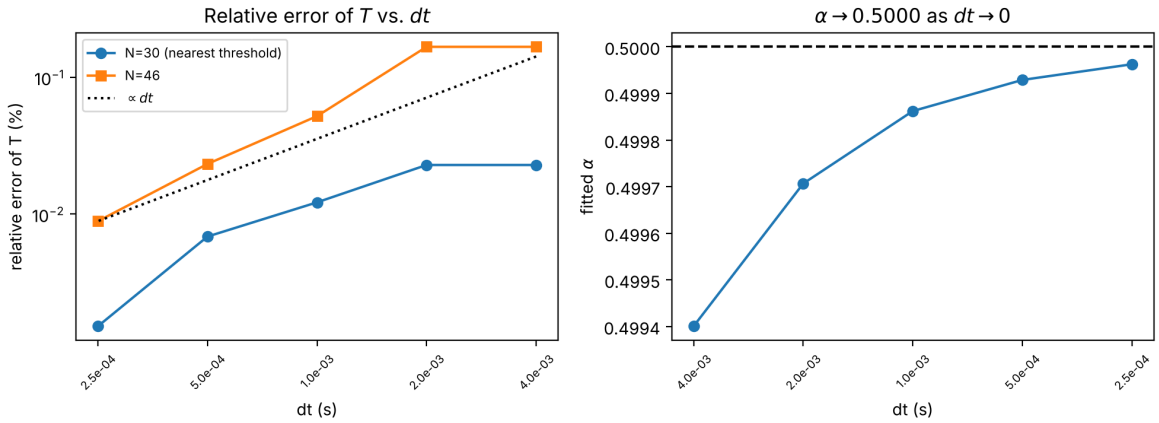


Figure 6: Left: relative error of the simulated time to goal against the closed form versus time step, at the smallest ($N = 30$) and largest ($N = 46$) robot counts; the dotted line has slope one. Right: fitted exponent versus dt.

4.5 Friction model realism

Theorems 1–2 assume isotropic Coulomb friction with $\mu_s = \mu_k$. We tested two extensions.

Anisotropic friction. We let the friction coefficient depend on direction of motion, $\mu_\parallel$ along the direction of travel and $\mu_\perp$ transverse to it, and re-ran the sweep holding $\mu_\parallel = 0.6$ fixed. Since transport in our setup is a straight-line task, only $\mu_\parallel$ enters the force balance; as expected, the fitted exponent was unchanged to four decimal places ($\alpha = 0.5004$) across eight ratios $\mu_\perp/\mu_\parallel \in [0.1, 10]$ in the baseline regime (Table 4, top), and in all five regimes of Table 2 with $\mu_\perp = 2\mu_\parallel$ (bottom). This outcome follows directly from the force balance and confirms that N^* and α depend only on the friction coefficient along the direction of travel.

Table 4: Anisotropic friction: only $\mu_\parallel$ affects the law. Top: ratio sweep in the baseline regime. Bottom: $\mu_\perp = 2\mu_\parallel$ in all five regimes.

Configuration	$\mu_\parallel$	$\mu_\perp$	α_{fit}
Ratio 0.1 (baseline)	0.6	0.06	0.5004 ± 0.0001
Ratio 0.5 (baseline)	0.6	0.3	0.5004 ± 0.0001
Ratio 1.0, isotropic (baseline)	0.6	0.6	0.5004 ± 0.0001
Ratio 1.5 (baseline)	0.6	0.9	0.5004 ± 0.0001
Ratio 2.0 (baseline)	0.6	1.2	0.5004 ± 0.0001
Ratio 3.0 (baseline)	0.6	1.8	0.5004 ± 0.0001
Ratio 5.0 (baseline)	0.6	3.0	0.5004 ± 0.0001
Ratio 10.0 (baseline)	0.6	6.0	0.5004 ± 0.0001
Baseline ($\mu_\perp = 2\times$)	0.6	1.2	0.5004 ± 0.0001
Heavy / low friction ($\mu_\perp = 2\times$)	0.4	0.8	0.5004 ± 0.0001
Light / high friction ($\mu_\perp = 2\times$)	0.8	1.6	0.5002 ± 0.0001
Large body ($\mu_\perp = 2\times$)	0.5	1.0	0.5004 ± 0.0001
Low friction ($\mu_\perp = 2\times$)	0.3	0.6	0.5004 ± 0.0001

Velocity-dependent (Stribeck) friction. Replacing the constant kinetic coefficient by a

Stribeck-style term [16] $\mu(v) = \mu_k + (\mu_s - \mu_k)\exp(-|v|/v_{\text{ref}})$, with $\mu_k < \mu_s$, breaks the clean power law: a forced power-law fit against $N^*_{\text{static}} = \mu_s mg/F_{\max}$ gives $\alpha_{\text{fit}} = 0.230$, 0.198 and 0.165 in the three regimes of Table 6. We derive the expected behavior.

Asymptotic analysis. Let $\varepsilon = F_{\max}(N - N^*_{\text{static}}) > 0$. Once moving, $m\dot{v} = \varepsilon + (\mu_s - \mu_k)mg\,[1 - \exp(-v/v_{\text{ref}})]$. For small v, $1 - \exp(-v/v_{\text{ref}}) \approx v/v_{\text{ref}}$, giving $\dot{v} = \varepsilon/m + \lambda v$ with $\lambda = (\mu_s - \mu_k)g/v_{\text{ref}}$ and solution $v(t) = (\varepsilon/(m\lambda))(e^{\lambda t} - 1)$: velocity grows exponentially because decreasing friction at higher speed is a negative-damping effect. The time T_1 for v to reach v_{ref} is

$$T_1 \approx \frac{1}{\lambda} \ln\!\left(\frac{m\lambda v_{\text{ref}}}{\varepsilon}\right),$$

a **logarithmic** divergence as $\varepsilon \to 0^+$. In the baseline regime a fit of $T(N) = A\ln(B/\varepsilon)$ achieves a sum of squared residuals of 0.011, about 18× lower than the best forced power-law fit (0.205; Figure 7). The fitted rate $A = 0.711$ s is of the same order as, but 39% above, the predicted $1/\lambda = 0.510$ s. Across all cases the ratio $A_{\text{fit}}/A_{\text{pred}}$ ranges from 0.95 to 2.6 (Tables 6 and 5); the logarithmic form is therefore supported as the better functional description, but the predicted rate is only an order-of-magnitude guide. A likely origin of the discrepancy is that the linearization is valid only for $v \ll v_{\text{ref}}$ while the fitted data span speeds comparable to v_{ref}; we have not verified this. The fitted prefactor B is weakly constrained and is not reported. Two caveats apply to the model comparison: each fit has two free parameters, but the power-law fit is forced to the static threshold (leaving N^* free could improve it), and each case has only about a dozen points spanning one decade of ε, over which a logarithm and a small power are hard to distinguish. We did not compute information criteria. The result is a qualitative change in universality class (exponentially weaker divergence) driven by a negative-damping instability absent from idealized Coulomb friction.

Table 5: Stribeck log-law versus crossover speed v_{ref} (baseline regime, deterministic parameter sweep).

v_{ref} (m/s)	$A_{\text{predicted}} = 1/\lambda$	A_{fit}	$A_{\text{fit}}/A_{\text{pred}}$	SSR ratio (log/power)
0.2	0.204	0.525	2.57	0.314
0.3	0.306	0.589	1.92	0.204
0.5	0.510	0.711	1.39	0.054
0.7	0.714	0.822	1.15	0.016
1.0	1.019	0.971	0.95	0.171

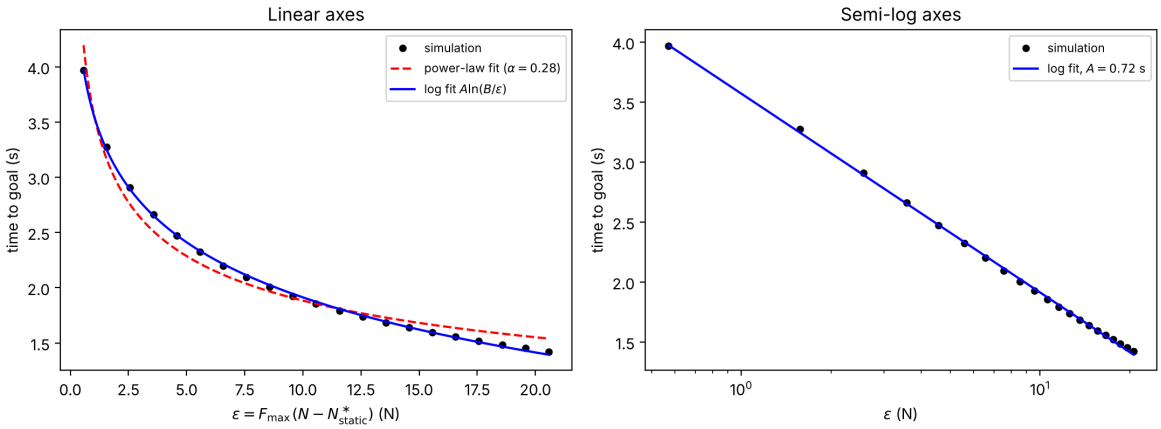


Figure 7: Left: power-law fit (dashed) vs. the logarithmic-divergence fit derived from the linearized Stribeck dynamics (solid), against simulated data. Right: the same data and logarithmic fit on semi-log axes, appearing linear as predicted. Produced with the re-implemented simulator (baseline Stribeck regime: fitted $\alpha = 0.28$ and $A = 0.72$ s); the original-simulator values quoted in the text and in Tables 6–7 are $\alpha = 0.230$ and $A = 0.711$ s.

Table 6: Velocity-dependent (Stribeck) friction breaks the idealized $\alpha = 1/2$ power law across three physical regimes.

Regime	μ_s	μ_k	v_{ref}	α_{fit} (power-law)	A_{fit} (log)	$A_{\text{predicted}} = 1/\lambda$
Baseline (Coulomb, no Stribeck)	0.6	0.6	–	0.500 ± 0.000	–	–
Baseline (Stribeck)	0.6	0.5	0.5	0.230	0.711	0.510
Heavy / low friction (Stribeck)	0.4	0.3	0.4	0.198	0.632	0.408
Light / high friction (Stribeck)	0.8	0.65	0.3	0.165	0.416	0.204

Table 7: SSR of the logarithmic fit relative to the forced power-law fit (lower = log fit better).

Regime	SSR (power-law)	SSR (log fit) [ratio to power-law]
Baseline	0.2048	0.0111 [0.054×]
Heavy / low friction	0.2060	0.0338 [0.164×]
Light / high friction	0.1377	0.0498 [0.362×]

Independent reproduction. We re-ran the Stribeck study with the re-implemented simulator (integer N from just above $\lceil N^*_{\text{static}} \rceil$ to 50; two-parameter power-law fit against the static threshold and two-parameter logarithmic fit; Figure 8). The qualitative conclusions agree: the logarithmic fit has a lower SSR than the power law in all three regimes (SSR ratios 0.024, 0.027, 0.270) and at all five v_{ref} (ratios 0.128, 0.064, 0.024, 0.133, 0.609 for $v_{\text{ref}} = 0.2, \ldots, 1.0$), and the fitted rate exceeds $1/\lambda$ at small v_{ref} ($A_{\text{fit}}/A_{\text{pred}} = 2.9, 2.1, 1.4, 1.1$) and falls below it at $v_{\text{ref}} = 1.0$ (0.88). The numbers differ from Tables 6–5 (for example baseline $A_{\text{fit}} = 0.720$ against 0.711, and a power-law exponent of 0.28 against 0.230), as expected given different point sets and fitting details, and the advantage of the logarithm is modest at $v_{\text{ref}} = 1.0$.

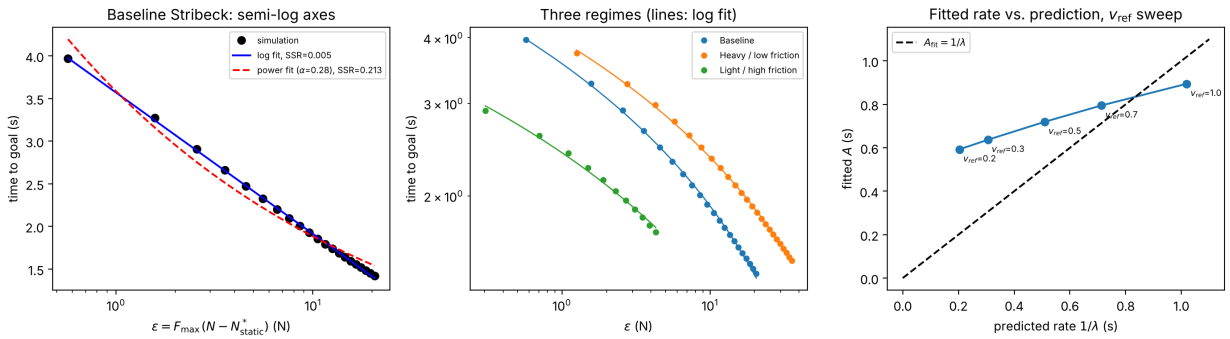


Figure 8: Stribeck friction, re-implemented simulator. Left: baseline regime ($\mu_s = 0.6$, $\mu_k = 0.5$, $v_{\text{ref}} = 0.5$), time to goal against ε on a logarithmic axis with the logarithmic fit (solid) and the power-law fit (dashed). Centre: log–log view for the three regimes of Table 6 with logarithmic fits. Right: fitted logarithmic rate A against the predicted $1/\lambda$ for $v_{\text{ref}} = 0.2$–1.0.

5 Decentralized Control

The theory and simulation above assume every robot is perfectly informed and aligned with the goal direction. Real swarm-robotics settings instead use decentralized, locally sensed, communication-free control. We ask whether the $\alpha = 1/2$ law survives when robots (i) share no information, (ii) sense the goal direction only locally with a persistent per-robot bias plus per-step noise, and (iii) are controlled by a policy with no knowledge of the closed-form law. Because the answer depends on what the controller can in principle learn, every result is reported against an untrained baseline.

5.1 Policy and observation model

We use a single multilayer perceptron whose parameters are shared across all N robots. Each robot's observation is strictly local: its own attachment angle $(\cos, \sin)$, its own noisy reading of the goal direction (a fixed per-robot angular bias, $\sigma_{\text{bias}} = 0.10$ rad, plus per-step jitter, $\sigma_{\text{step}} = 0.15$ rad at training), and its own instantaneous speed along its sensed goal direction. No robot observes any other robot, and no communication channel exists. The output is parametrized as an angle: the network predicts a bounded (tanh-limited, $c_{\max} = 0.5$ rad) correction to the robot's *sensed goal angle*, and the applied force is $F_{\max}(\cos, \sin)$ of the corrected angle. This avoids the unstable gradient of a raw-vector-plus-normalization output but has a structural consequence used throughout: with zero correction the controller applies its sensed direction.

5.2 Training

We implement a batched differentiable simulator in PyTorch and train by backpropagating through simulated trajectories. For training only, the hard stick–slip switch is replaced by a smooth sigmoid blend; *evaluation always uses exact stick–slip physics.* Choices not fixed by the model description are: the loss is the mean over time of the remaining-distance fraction $\max(0, 1 - p_t/d)$ (p_t = progress along the true goal direction); the MLP is $5 \to h \to h \to 1$ with $h = 8$ and tanh activations, with the output layer initialized near zero; success means reaching the goal within 7.2 s ($= 90 \times 0.08$ s, the training horizon). Each batch uses $N = \lceil N^* \rceil + 1 = 31$ robots (the hardest count used in the study), 320 gradient steps, batch size 24, horizon 90 steps at $dt = 0.08$ s, Adam with 10% linear warmup then cosine decay (peak 3×10^{-3}), gradient-norm clipping at 2.0, and a noise curriculum ramping

both noise sources linearly from 0 to their final level over the first half of training. Evaluation uses $dt = 0.02$ s, robot counts $N = 31, \ldots, 46$ (offsets 1–16 above $\lceil N^* \rceil = 30$; a 15-robot range beyond the single training count), five noise trials per count, and a three-parameter fit $T = C(N - N_0)^{-\alpha}$ to the mean time per count (counts at which any trial failed to reach the goal within 30 s are excluded from the fit). We study two goal conditions: *fixed* (goal at world angle 0, every episode) and *random* (goal angle drawn uniformly from $[0, 2\pi)$ each episode). A *sensitivity probe* sweeps each policy's sensed goal angle by ± 0.3 rad and reports the slope of the applied angle and of the correction alone; slope 1 of the applied angle means the policy tracks its sensed input one-to-one, slope 0 means a constant output.

5.3 The identity baseline

Setting the correction to zero gives the *identity policy*: no training, no network, applied direction equal to sensed direction. Its evaluation (first two rows of Table 8) gives success 1.000, $\alpha = 0.487$ (fixed goal) and 0.492 (random goal), and $N_0 = 29.81$ and 29.78. Its $T(N)$ curve follows the noise-corrected prediction closely (Figure 9). These values are close to every quantity that a trained near-identity policy can produce, so *achieving* 100% *success with* $\alpha \approx 0.5$ *is not evidence of learning in this architecture*. This follows from the architecture's structure, not from training details.

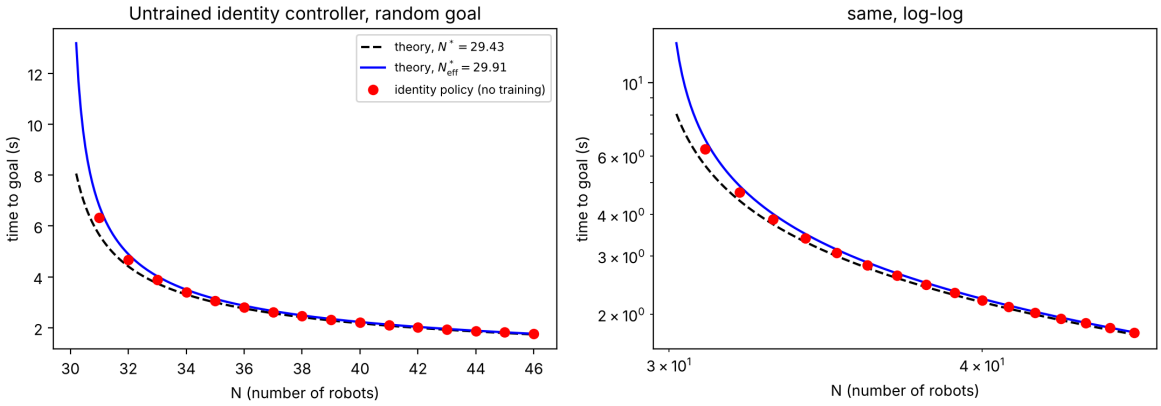


Figure 9: Untrained identity controller with randomized goal direction ($\sigma_{\text{bias}} = 0.10$, $\sigma_{\text{step}} = 0.15$; 10 noise trials per N). Points: mean time to goal. Dashed: closed form with the noiseless N^*. Solid: closed form with the noise-corrected N^*_{eff} of Proposition 1.

5.4 Results with trained policies

Table 8: Decentralized control, baseline regime, evaluated at offsets 1–16 ($N = 31$–46), five noise trials per count. Mean ± std across seeds. Slopes are from the sensitivity probe (applied angle, and correction alone, versus sensed goal angle).

Controller	Seeds	Success	α_{fit}	N_0	Applied slope	Correction slope
Identity, fixed goal (no training)	–	1.000	0.487	29.81	1.00	0.00
Identity, random goal (no training)	–	1.000	0.492	29.78	1.00	0.00
Trained, fixed goal (translational)	20	1.000 ± 0.000	0.475 ± 0.006	29.40 ± 0.07	0.23 ± 0.09	-0.77 ± 0.09
Trained, random goal (translational)	20	1.000 ± 0.000	0.493 ± 0.001	29.77 ± 0.01	1.00 ± 0.00	0.00 ± 0.00
Trained, fixed goal (full Newton–Euler)	5	1.000 ± 0.000	0.476 ± 0.003	29.42 ± 0.07	0.26 ± 0.07	-0.74 ± 0.07
Trained, random goal (full Newton–Euler)	5	1.000 ± 0.000	0.493 ± 0.001	29.78 ± 0.01	1.00 ± 0.00	0.00 ± 0.00

Random goal. Twenty independently initialized policies reach success 1.000 ± 0.000, $\alpha = 0.493 \pm 0.001$ and $N_0 = 29.77 \pm 0.01$ (Table 8; per-seed values in Table 11). The sensitivity slope is 1.00 ± 0.00 and the correction slope is 0.00: the trained policies *are* the identity map to within the resolution of the probe, and their results are indistinguishable from the untrained control.

Fixed goal. With the goal always at angle 0, the same procedure produces policies that partly memorize the goal: applied-angle slope 0.23 ± 0.09 (correction slope -0.77 ± 0.09), $N_0 = 29.40 \pm 0.07$ – lower than the identity value 29.81 and close to the noiseless $N^* = 29.43$ – and $\alpha = 0.475 \pm 0.006$. These policies cancel part of their sensing error by exploiting the fixed goal, a solution that is unavailable when the goal is randomized and that would not transfer to a setting where the goal is not known in advance.

Rotation. Repeating both conditions with the full Newton–Euler dynamics (torque, rotating attachment points; 5 seeds each) changes nothing: random goal $\alpha = 0.493 \pm 0.001$, $N_0 = 29.78 \pm 0.01$; fixed goal $\alpha = 0.476 \pm 0.003$, $N_0 = 29.42 \pm 0.07$ (Table 8), as predicted by Proposition 2. This addresses the earlier limitation that the decentralized study used translational dynamics only, under the stated assumption that forces are world-frame and friction is translational.

Interpretation. The law survives decentralized, noisy, communication-free control, but with this parametrization and sensing-noise model the surviving behavior is the identity map's behavior, which Proposition 1 describes exactly. The experiments do not show that a learned controller discovers the law; they show that the law holds for a controller that does not need to. Making the learning problem non-trivial requires an observation from which a robot can infer its own bias (for example, its velocity direction rather than a scalar speed), which we leave open.

5.5 Negative controls

With the goal *fixed*, masking the sensed goal input to zero (so that the base direction $\operatorname{atan2}(0, 0) = 0$ coincides with the true goal) leaves success at 1.000 ($\alpha = 0.484 \pm 0.004$, $N_0 = 29.34 \pm 0.03$; 5 seeds), but the same policies reach the goal in only 0.39 ± 0.02 of random-goal episodes. With the goal *randomized*, masking to zero gives success 0.360 ± 0.027, and replacing the input by pure noise gives 0.000 (5 seeds each). The masked input therefore acts as a negative control only when the goal is randomized; the fixed-goal version passes silently because the default output happens to equal the answer (Section 7). Note that in this architecture the control is nearly guaranteed to fail once the goal varies, because the base angle is the sensed input; it tests that the evaluation can detect a signal-blind policy, not that the policy has learned to sense.

6 Noise Coherence: Validation of Proposition 1

We re-evaluated every trained policy (20 per goal condition) and the identity control without retraining at five levels of per-step sensing noise ($\sigma_{\text{step}} \in \{0, 0.05, 0.15, 0.30, 0.50\}$ rad, training level 0.15) and compared the fitted threshold N_0 with $N^*_{\text{eff}} = N^* e^{(0.10^2 + \sigma^2_{\text{step}})/2}$, using only the known noise parameters and $N^* = 29.43$ (no fitted constants).

Table 9: Noise-coherence prediction versus fitted threshold. "Max rel. err." is the largest relative deviation of any of the 20 policies from the prediction. Success for random-goal policies is averaged over seeds.

		Identity	Random goal (20 policies)			Fixed goal (20 policies)	
σ_{step}	N^*_{eff}	N_0	N_0	Max rel. err.	Success	N_0	Max rel. err.
0.00	29.578	29.455	29.452 ± 0.004	0.45%	1.000	29.365 ± 0.046	1.10%
0.05	29.615	29.488	29.485 ± 0.005	0.47%	1.000	29.372 ± 0.049	1.16%
0.15	29.912	29.779	29.774 ± 0.006	0.49%	1.000	29.400 ± 0.073	2.29%
0.30	30.939	30.866	30.860 ± 0.008	0.31%	0.938	29.572 ± 0.130	5.30%
0.50	33.516	33.403	33.418 ± 0.032	0.44%	0.750	30.415 ± 0.233	10.42%

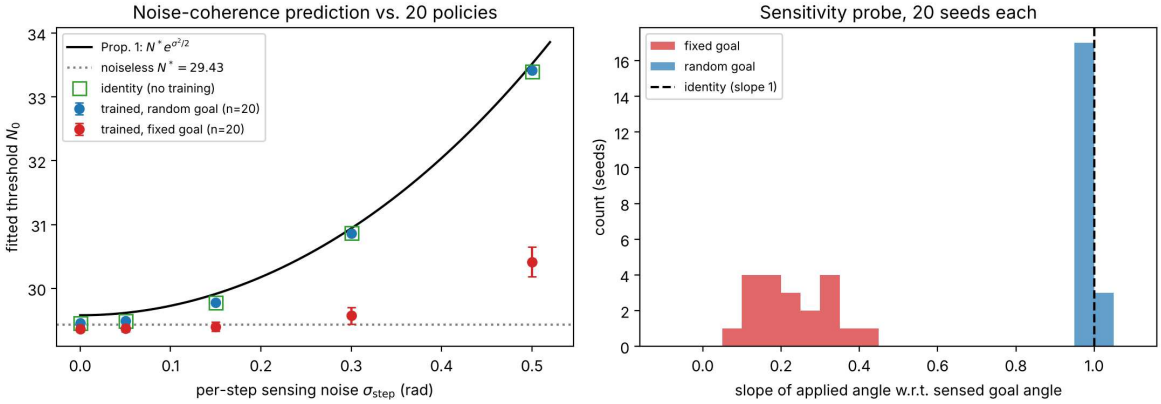


Figure 10: Left: Proposition 1 (line) against fitted thresholds of the identity control (squares) and of 20 trained policies per goal condition (mean ± std). Right: sensitivity-probe slopes of the applied angle for 20 fixed-goal and 20 random-goal policies; the dashed line marks the identity map.

Random goal. For all 20 policies and all five noise levels the fitted threshold agrees with the prediction within 0.49% (Table 9), including at $\sigma_{\text{step}} = 0.50$, more than three times the training noise. The same agreement holds for the identity control, as expected. A small systematic offset remains: for the random-goal policies the fitted N_0 lies 0.26–0.46% *below* the prediction at every noise level (e.g. 29.452 against 29.578 at $\sigma_{\text{step}} = 0$). We have not isolated its origin; the three-parameter fit and the finite $N = 31$–46 range are candidates, but we did not test them. Success degrades with noise, to 0.938 at $\sigma_{\text{step}} = 0.30$ and 0.750 at 0.50.

Fixed goal. The prediction fails for the fixed-goal policies, with deviations growing with noise (largest single-policy deviation 10.4% at $\sigma_{\text{step}} = 0.50$), because these policies partly cancel their sensing error; this is the situation excluded in Proposition 1. Proposition 1 is thus supported for controllers that follow their sensed direction, and is not a bound on what a controller exploiting a fixed goal can achieve.

7 Ablation Studies

We retrain with individual factors changed, using the baseline procedure (fixed offset $N = 31$) and 5 seeds per variant (the baseline rows use the same 5 seeds, 0–4), in both goal conditions where informative (Table 10). The number of seeds is smaller than in Section 5.4; reported spreads are across seeds and do not capture variability across evaluation noise draws.

Table 10: Ablations (baseline regime, retrained from scratch per variant, 5 seeds). Left block: fixed goal; right block: random goal. Dashes: not run, or fit undefined because the policy did not reach the goal at the counts needed for a fit.

	Fixed goal			Random goal		
Variant	Success	α_{fit}	N_0	Success	α_{fit}	N_0
Full observation (baseline)	1.000 ± 0.000	0.475 ± 0.003	29.42 ± 0.07	1.000 ± 0.000	0.493 ± 0.001	29.78 ± 0.01
No attachment angle	1.000 ± 0.000	0.478 ± 0.006	29.42 ± 0.07	1.000 ± 0.000	0.492 ± 0.001	29.78 ± 0.01
No speed-toward-goal	1.000 ± 0.000	0.488 ± 0.000	29.42 ± 0.00	1.000 ± 0.000	0.493 ± 0.000	29.78 ± 0.00
Goal input masked to zero	1.000 ± 0.000	0.484 ± 0.004	29.34 ± 0.03	0.360 ± 0.027	–	–
Goal input replaced by noise	–	–	–	0.000 ± 0.000	–	–
Hidden width 4	1.000 ± 0.000	0.480 ± 0.004	29.46 ± 0.02	–	–	–
Hidden width 16	1.000 ± 0.000	0.480 ± 0.008	29.31 ± 0.08	–	–	–
Mixed offsets 1–12	1.000 ± 0.000	0.481 ± 0.001	29.38 ± 0.01	1.000 ± 0.000	0.492 ± 0.001	29.78 ± 0.01
Fixed offset 2	1.000 ± 0.000	0.479 ± 0.003	29.38 ± 0.01	–	–	–

Observation features. Removing the attachment angle or the speed feature changes nothing, which is expected when the policy converges to the identity map and uses neither. **Capacity.** Hidden widths 4, 8 and 16 give similar results in the fixed-goal condition (success 1.000; N_0 between 29.31 and 29.46). **Training offset.** Training on a mixture of offsets 1–12 gives success 1.000 in both goal conditions, and training at offset 2 (fixed goal only) also gives 1.000, with no difference from fixed-hardest-offset training in success or threshold (random goal: $N_0 = 29.78 \pm 0.01$ for mixed versus 29.78 ± 0.01 for fixed). We therefore find no training-offset effect. **Negative controls** are discussed in Section 5. The fixed-goal masked-input result is the architectural confound noted above: $\text{atan2}(0, 0) = 0$ equals the fixed goal, so masking silently supplies the answer; randomizing the goal removes it.

8 Training Reliability

Table 11 gives the per-seed outcome of the 20 random-goal training runs underlying Table 8. Because all policies converge to the identity map, the small seed-to-seed spread reflects convergence to a common solution rather than diversity of learned behavior.

Table 11: Twenty independent training runs, randomized goal, fixed-hardest-offset procedure (baseline regime, translational simulator).

Seed	Final loss	Success rate	α_{fit}	N_0	Applied slope
0	0.5101	1.000	0.491	29.79	1.002
1	0.5099	1.000	0.493	29.77	0.996
2	0.5108	1.000	0.493	29.77	0.997
3	0.5102	1.000	0.493	29.77	0.998
4	0.5101	1.000	0.493	29.77	0.997
5	0.5090	1.000	0.492	29.78	0.997
6	0.5091	1.000	0.494	29.77	0.992
7	0.5107	1.000	0.493	29.77	0.999
8	0.5083	1.000	0.494	29.77	0.998
9	0.5114	1.000	0.493	29.77	0.996
10	0.5087	1.000	0.493	29.78	0.999
11	0.5116	1.000	0.493	29.77	0.998
12	0.5089	1.000	0.491	29.79	1.002
13	0.5090	1.000	0.493	29.77	0.990
14	0.5102	1.000	0.493	29.78	1.000
15	0.5123	1.000	0.493	29.77	0.995
16	0.5100	1.000	0.492	29.78	1.000
17	0.5098	1.000	0.493	29.77	0.996
18	0.5105	1.000	0.492	29.78	1.000
19	0.5085	1.000	0.493	29.77	0.997
All 20: mean ± std	0.5100 ± 0.0011	1.000 ± 0.000	0.493 ± 0.001	29.77 ± 0.01	0.997 ± 0.003

9 Learning Cooperative Transport with MAPPO

Section 5.3 showed that a controller built as a correction to the sensed goal direction already satisfies the divergence law, and that in that setting learning adds nothing. We now ask whether multi-agent reinforcement learning (MARL) can outperform designed controllers in a regime where something remains to be learned. Two effects are made strong. First, per-robot sensing error (persistent bias plus jitter) is large, so an unprocessed controller loses coherent force: with $\sigma^2 = \sigma^2_{\text{bias}} + \sigma^2_{\text{step}} = 0.34$ the noise-coherence threshold of Proposition 1 moves from $N^* = 17.66$ to $N^* e^{\sigma^2/2} = 20.9$. Second, the centre of mass is offset from the geometric centre, so pushing toward the goal produces a disturbance torque that the team must cancel to keep the payload's heading while still carrying it. Both effects can only be reduced by robots acting together through the payload: the payload's motion carries information about the true goal direction and about the net torque, and the cancelling torque must be shared across robots.

9.1 Task

A disc of mass $m = 3\,\text{kg}$, inertia $I = 0.5\,\text{kg}\,\text{m}^2$ and radius $r = 0.35\,\text{m}$ is carried by N robots attached at evenly spaced boundary points (random global phase) over a surface with Coulomb friction $\mu = 0.6$ and exact stick–slip contact ($\mu_s = \mu_k$), with viscous rotational damping, integrated with explicit Euler at $dt = 0.1\,\text{s}$ for at most 140 steps. Each robot applies a force of fixed magnitude $F_{\max} = 1\,\text{N}$ at a world-frame angle it chooses. The centre of mass is offset from the geometric centre by a random vector of magnitude in $[0.01, 0.04]$ m. The goal direction is drawn uniformly at random in every episode, at distance $d = 4\,\text{m}$. An episode succeeds if the payload reaches the goal with $|\text{heading error}| < 0.35$ rad. Robot counts are $N \in \{19, \ldots, 28\}$ ($N^* = 17.66$).

Each robot's observation is local and noisy: its attachment angle relative to its sensed goal direction $(\cos, \sin)$, the payload velocity expressed in its sensed-goal frame, the payload heading, and the angular velocity (gyro). The sensed goal angle equals the true goal plus a persistent per-

robot bias ($\sigma_{\text{bias}} = 0.5$ rad) plus per-step jitter ($\sigma_{\text{step}} = 0.3$ rad). There is no communication. The team reward is $8\,\Delta p/d - (1-\cos\theta)\,dt + 4\cdot\mathbf{1}$[arrival with admissible heading], where p is progress along the true goal direction. During training only, the critic observes a privileged global state (progress, lateral offset, payload velocity, heading, angular velocity, centre-of-mass offset, goal direction, N, time, and summary statistics of the biases).

9.2 Compared methods

Designed controllers (same observations; no learning). *Identity*: apply the sensed direction. *Tuned PD*: sensed direction plus a heading-and-rate correction weighted by each robot's lever arm, $\delta_i = -K(\theta + k_d\,\omega)\cos(\text{attachment angle relative to the goal})$. *Tuned PD + bias correction*: additionally subtracts a running estimate of the difference between the sensed direction and the payload's velocity direction. Gains were chosen by grid search on a separate tuning seed and evaluated on separate seeds.

Learned baselines. *IPPO* and *MAPPO* with a parameter-shared actor and a two-dimensional direction-vector action, in the standard form (initial mean zero) and in a variant centred on the sensed-goal unit vector; *IPPO + residual prior* (local critic) and *MAPPO + residual prior* (centralized critic), whose action is a bounded correction of the sensed angle (below).

RCR-MAPPO (Residual–Curriculum–Recurrent MAPPO). MAPPO with (i) a *residual-angle action*, $\psi_i = s_i + c_{\max}\tanh\mu_\theta(o_i)$ with $c_{\max} = 1.2$ rad, where s_i is the sensed goal angle, a Gaussian exploration noise with state-independent standard deviation (initially 0.3 rad), and a near-zero output initialization so that training starts from the Identity controller; (ii) a *threshold-annealing curriculum*, in which training robot counts are drawn from $\{n_{\text{lo}}, \ldots, 28\}$ with n_{lo} decreasing linearly from 25 to 19 over the first half of training; and (iii) a *GRU memory* (64 units) giving each robot a filtered history of its own observations, which allows it to estimate its own sensing bias from the payload's motion. Variants add a *counterfactual (CF) credit*: the per-agent advantage is $Q(s,a) - Q(s,(a_{-i},\bar{a}_i))$, obtained by a one-step counterfactual simulation with the known dynamics in which agent i is replaced by its mean action. All learned methods use the same budget: 450 PPO iterations of 24 episodes, 3 epochs, clip 0.2, $\gamma = 0.99$, GAE $\lambda = 0.95$, learning rate 10^{-3}, entropy coefficient 0.003.

9.3 Protocol and statistics

Each learned method is trained from independent random seeds (seeds 10–14 for all arms except the two vector-action arms without goal prior, which have seeds 10–12; plus seeds 15–19 for the four arms of the confirmatory comparison) and evaluated on identical environment draws (evaluation seed 5000, 150 episodes per N). The primary metric, fixed before the confirmatory runs, is nominal-noise success averaged over $N = 19, \ldots, 28$. Secondary metrics are N_{90} (smallest N from which success is at least 0.9 for every larger N, with the number of seeds that reach it), arrival time at the ample counts $N = 25$–28 (a fixed set of counts, so that methods that also succeed at hard counts are not penalized), and success under conditions never seen in training (sensing noise $\times 0.5$ and $\times 1.5$, centre-of-mass offset $\times 2$). Designed controllers are deterministic given the environment draw; their spread is over five evaluation seeds. The pre-registered confirmatory test (Welch t-test against learned baselines, one-sample t-test against the mean of the designed controllers, Holm correction over five comparisons) was written before seeds 15–19 were run. **Disclosure.** RCR-MAPPO was identified as the reference *after* inspecting the first five seeds; the variant with CF credit was the originally intended method and did not outperform it. Earlier in this study a different MARL method (local neighbour averaging) was tried in the setting of Section 5 and did not outperform

plain MAPPO or a designed neighbour-average controller; those results are in Appendix A. The design of the present task and method followed that negative result.

9.4 Results

Table 12: Nominal noise ($\sigma_{\text{bias}} = 0.5$, $\sigma_{\text{step}} = 0.3$). Success is averaged over $N = 19$–28; N_{90} is the median over the seeds that reach it, with the number reaching it in parentheses; time is the mean arrival time of successful episodes at $N = 25$–28. Mean ± standard deviation across training seeds. †Designed controllers: five evaluation seeds. The two "vector action" rows without goal prior never produced a single success in training.

Method	Seeds	Success	N_{90}	Time (s)
Identity (sensed direction, no learning)	5†	0.092 ± 0.006	– (0/5)	2.29 ± 0.03
Tuned PD heuristic	5†	0.758 ± 0.002	22 (5/5)	2.31 ± 0.01
Tuned PD + bias-correction heuristic	5†	0.781 ± 0.005	22 (5/5)	1.96 ± 0.00
IPPO, vector action	3	0.000 ± 0.000	– (0/3)	–
MAPPO, vector action	3	0.000 ± 0.000	– (0/3)	–
MAPPO, vector action + goal prior	5	0.187 ± 0.097	– (0/5)	2.59 ± 0.19
IPPO + residual-angle prior	10	0.686 ± 0.116	22 (5/10)	2.01 ± 0.11
MAPPO + residual-angle prior	10	0.610 ± 0.166	22 (3/10)	1.93 ± 0.09
+ threshold curriculum	10	0.750 ± 0.040	22 (4/10)	1.91 ± 0.05
+ GRU memory (RCR-MAPPO)	10	0.813 ± 0.025	22 (8/10)	1.95 ± 0.05

Table 13: Success averaged over $N = 19$–28 under conditions never seen in training (trained at the nominal noise level and centre-of-mass offset).

Method	Noise ×0.5	Noise ×1.5	COM offset ×2
Identity (sensed direction, no learning)	0.132 ± 0.015	0.045 ± 0.004	0.119 ± 0.004
Tuned PD heuristic	0.996 ± 0.003	0.279 ± 0.007	0.732 ± 0.008
Tuned PD + bias-correction heuristic	0.998 ± 0.002	0.362 ± 0.013	0.627 ± 0.010
IPPO, vector action	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
MAPPO, vector action	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
MAPPO, vector action + goal prior	0.192 ± 0.099	0.119 ± 0.037	0.191 ± 0.066
IPPO + residual-angle prior	0.873 ± 0.171	0.290 ± 0.057	0.443 ± 0.162
MAPPO + residual-angle prior	0.695 ± 0.255	0.276 ± 0.086	0.353 ± 0.112
+ threshold curriculum	0.913 ± 0.089	0.345 ± 0.025	0.466 ± 0.093
+ GRU memory (RCR-MAPPO)	0.918 ± 0.103	0.428 ± 0.052	0.612 ± 0.112

Table 14: Pre-registered confirmatory comparisons of RCR-MAPPO against the strongest learned baselines and the designed controllers on the primary metric. Welch t-test (learned) or one-sample t-test (designed controllers); Holm correction over the five comparisons. Seeds are pooled over all ten, five of which were used to select RCR-MAPPO; on the five fresh seeds alone no comparison survives Holm correction (see text).

Comparison method	Data	Diff. (RCR − method)	Raw p	Holm p
+ threshold curriculum	10 seeds	+0.063	0.000734	0.00294
IPPO + residual-angle prior	10 seeds	+0.127	0.00697	0.00906
MAPPO + residual-angle prior	10 seeds	+0.203	0.00374	0.00906
Tuned PD heuristic	heuristic, 5 eval seeds	+0.055	6.26e-05	0.000313
Tuned PD + bias-correction heuristic	heuristic, 5 eval seeds	+0.032	0.00302	0.00906

Table 15: Counterfactual-credit (CF) variants of RCR-MAPPO, reported separately because they have five seeds each (the other arms have ten) and were the originally intended method. Success is averaged over $N = 19$–28; time is arrival time at $N = 25$–28. Compare with the RCR-MAPPO row of Tables 12 and 13.

Variant	Seeds	Nominal	Noise ×0.5	Noise ×1.5	COM ×2	Time (s)
+ CF credit, no GRU	5	0.823 ± 0.028	0.933 ± 0.127	0.495 ± 0.018	0.773 ± 0.055	4.60 ± 0.99
+ GRU + CF credit	5	0.817 ± 0.038	0.954 ± 0.042	0.521 ± 0.023	0.774 ± 0.070	3.98 ± 1.37
GRU + CF credit, no curriculum	5	0.697 ± 0.252	0.812 ± 0.332	0.482 ± 0.147	0.676 ± 0.251	6.36 ± 1.98

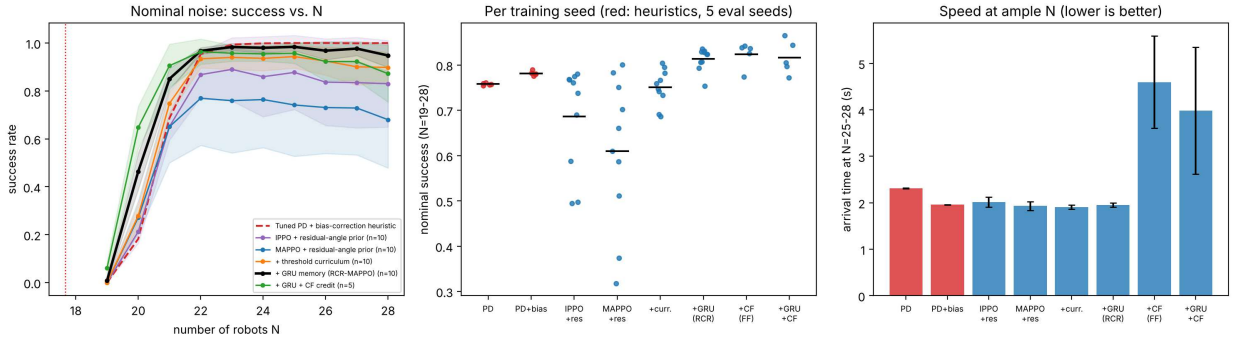


Figure 11: Left: success versus N at nominal noise (dashed: designed controller; solid: learned, mean ± std over seeds; dotted vertical line: N^*). Centre: nominal success per training seed (horizontal bar: mean; red: designed controllers, five evaluation seeds). Right: arrival time at $N = 25$–28.

Main outcome. RCR-MAPPO reaches nominal success 0.813 ± 0.025 ($n = 10$ seeds), the highest mean and the smallest across-seed spread among the learned methods trained with ten seeds (Table 12); the CF variants, trained with five seeds, have statistically indistinguishable nominal success (Table 15). Vector-action MAPPO/IPPO without a goal prior never succeed (0.000); with a goal prior MAPPO reaches only 0.187 ± 0.097. With the residual-angle action, IPPO reaches 0.686 ± 0.116 and MAPPO 0.610 ± 0.166; the curriculum raises MAPPO to 0.750 ± 0.040 and the GRU to 0.813 ± 0.025. The ordering of means is thus residual action $\gg$ vector action, and each of curriculum and GRU adds to success, with the GRU mainly reducing seed-to-seed variance (Levene test against MAPPO+residual: $p = 0.002$; against IPPO+residual: $p = 0.049$; against the curriculum variant: $p = 0.17$).

Pre-registered confirmatory comparison (Table 14). Pooling all ten seeds, RCR-MAPPO is significantly better than all five comparison methods after Holm correction (Holm $p \leq 0.009$; differences +0.032 to +0.203). This pooled test is *not* clean, because five of the ten seeds are those on which RCR-MAPPO was selected as the reference. On the five fresh seeds alone (seeds 15–19, RCR-MAPPO 0.800), the differences keep the same sign and similar size (+0.044 over the curriculum variant, +0.129 over IPPO+residual, +0.244 over MAPPO+residual, +0.042 over the tuned PD controller, +0.019 over PD with bias correction), but none survives Holm correction (Holm $p = 0.16$–0.24; raw $p = 0.032$–0.23). The confirmatory evidence is therefore suggestive and not conclusive; the largest margins are over the weakest learned baselines.

Compared with the designed controllers. The advantage over the tuned designed controllers at nominal noise is small (+0.03 to +0.06). The designed controllers are better when sensing is cleaner or the centre-of-mass offset is larger than in training (Table 13): at noise ×0.5 they reach 0.996–0.998 against 0.918 for RCR-MAPPO, and at COM offset ×2 the tuned PD controller reaches 0.732 against 0.612. RCR-MAPPO is better at noise ×1.5 (0.428 against 0.279 and 0.362). Hence there is no regime in which the learned method dominates.

Counterfactual credit (Table 15). CF credit does not improve nominal success (FF variant 0.823 ± 0.028, GRU variant 0.817 ± 0.038, $n = 5$ each, against 0.813 ± 0.025) and slows arrival at ample N by roughly a factor two (3.98–4.60 s against 1.95 s). Exploratorily, it gives the best robustness to unseen conditions (noise ×1.5: 0.521 ± 0.023; COM ×2: 0.774 ± 0.070 for the GRU variant), where it matches or exceeds the designed controllers at COM ×2. These unseen-condition comparisons were not pre-registered and use $n = 5$.

Centralized critic. A centralized critic does not help: IPPO with the same residual action is better on average than MAPPO with it (0.686 against 0.610), though the difference is within seed variance.

Threshold. N_{90} is reached by only 8/10 seeds for RCR-MAPPO and by fewer for the other learned baselines (median 22 where reached, against 22 for the designed controllers, which reach it in every evaluation); we therefore regard it as unreliable here and do not base claims on it.

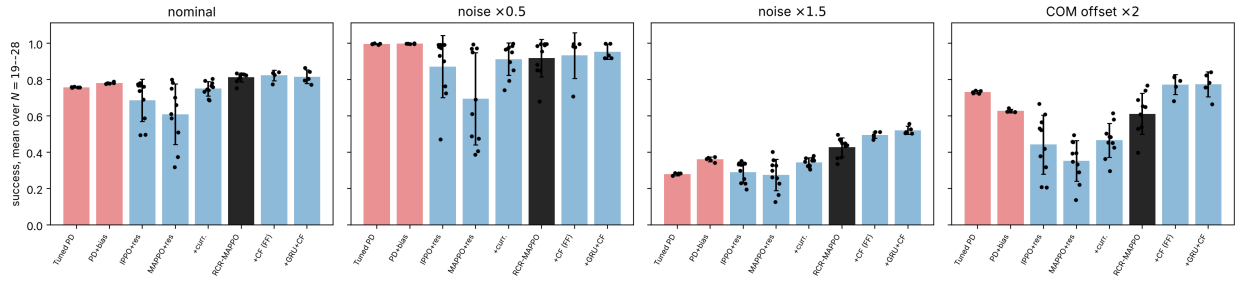


Figure 12: Mean success over $N = 19$–28 under the nominal condition and the three conditions never seen in training. Bars: mean ± std over training seeds (red: designed controllers, five evaluation seeds); dots: individual seeds. The designed controllers lead at low noise and, for the PD controller, at doubled COM offset; the learned methods lead at high noise.

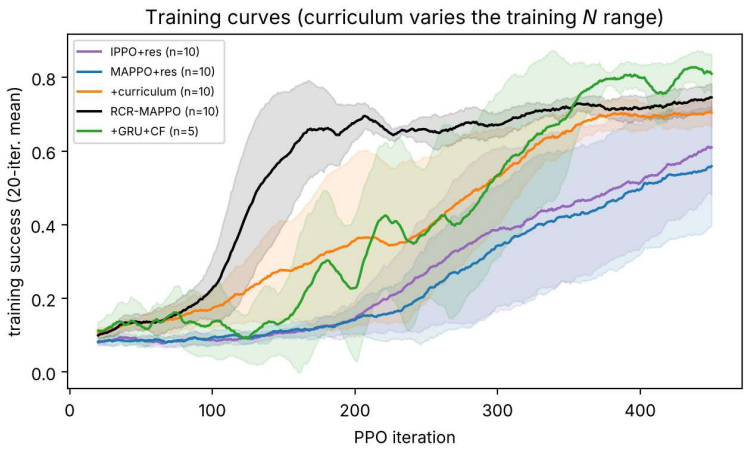


Figure 13: Training-time success (20-iteration moving average, mean ± std over seeds). Methods with the curriculum train on a range of robot counts that widens during the first half of training, so their training success is not directly comparable with that of methods trained on $N = 19$–28 throughout.

9.5 What this does and does not show

The evidence supports the following and nothing stronger. (i) The residual-angle parametrization is necessary: with a direction-vector action, MAPPO and IPPO either fail to learn at all or learn far less, even when the vector is centred on the sensed direction. (ii) Among the learned methods with ten seeds, RCR-MAPPO has the highest and most consistent nominal success (the five-seed CF variants are indistinguishable from it); on fresh seeds alone the margin over the other learned methods is not statistically significant after multiplicity correction. (iii) Against tuned designed controllers it is marginally better at nominal noise, better at high noise, and worse at low noise and at a doubled centre-of-mass offset. (iv) A centralized critic did not help, and CF credit trades speed for robustness. We do not claim that MARL beats well-designed controllers in this task; we claim that a residual-angle MAPPO with curriculum and memory matches or slightly exceeds them at the training condition, using only local observations and no hand-designed correction.

Limitations. All results are simulation-only and in a single physical regime. The designed controllers are two members of a design space we did not exhaust; a better-designed controller could close the gap. Baseline tuning was less extensive than method development: the learning rate and curriculum were chosen on development runs of the residual-action configurations, whereas the vector-action baselines received a single configuration (initial exploration noise, one learning rate), so their weakness is partly a statement about that parametrization and not about MARL in general. The reference method was selected post hoc from the first five seeds, which is why the confirmatory seeds and the Holm correction are reported. Success is a coarse metric near the threshold, and N_{90} is an integer-valued summary that several seeds do not reach.

10 Independent Physics-Engine Check

The simulator of Section 4 was written by us. To test the mechanical claims against an independent implementation of contact and friction we repeated them in a general-purpose rigid-body engine (PyBullet/Bullet, API version 202010061, 500 Hz, 100 solver iterations). This is a check by a second simulator and *not* a physical experiment; it can expose errors in our integrator and contact

handling, but it shares with the theory the modelling of friction as Coulomb-type, and it does not replace hardware.

10.1 Setup and protocol

A cylinder (height 0.1 m) rests on a plane. Plane friction is 1 and body friction is μ, so the effective Coulomb coefficient is μ (the engine multiplies the two); rolling and spinning friction are zero, and the contact stiffness and damping (10^5, 10^3) were set once and not tuned. N equal forces F act at evenly spaced boundary points in the world frame toward the goal at distance $d = 2$ m, with a limit of 10 s (60 s for divergence runs). The protocol was written before the runs and is included with the code. **E1** (threshold): for each of the five regimes of Table 2, $N = \lfloor N^* \rfloor - 2, \ldots, \lceil N^* \rceil + 2$; "moves" means a displacement of at least 1 cm along the goal; the prediction is that the body moves if and only if $N > N^* = \mu mg/F$. **E2** (divergence): $N = \lceil N^* \rceil + 1, \ldots, \lceil N^* \rceil + 10$ in each regime; $T = C(N - N^*)^{-\alpha}$ is fitted with N^* fixed (primary) and free (secondary); criterion $|\alpha - 0.5| < 0.01$. **E3** (sensing error): baseline regime, per-puller angle errors $\theta_i \sim \mathcal{N}(0, \sigma^2)$ drawn independently per trial, $\sigma \in \{0, 0.15, 0.30, 0.50\}$ rad, 20 draws per (σ, N); the threshold is the N at which the fraction of draws that move crosses 0.5, compared with $N^* e^{\sigma^2/2}$; criterion: relative error below 2%. **E4**: time-step dependence. The engine is deterministic, so repeated runs of one configuration do not vary; the draws of E3 are the only source of spread.

10.2 Results

Threshold and divergence (E1, E2). In all five regimes the body moves for every tested $N > N^*$ and does not move for every tested $N < N^*$ (Table 16); The fitted exponent with N^* fixed is 0.4984–0.4996, within the criterion in every regime, and the free-threshold fit recovers N^* to within 0.01 in every regime. The simulated times deviate from the closed form by at most 0.19%. The exponents are close to, but not identical with, those of Table 2, as expected from different integrators.

Table 16: Physics-engine check, E1 and E2. "Last/first" are the largest N that did not move and the smallest that did. α fitted with $N^* = \mu mg/F$ fixed or free; N_0 is the free threshold. Max. err.: largest relative deviation of the simulated T from the closed form.

Regime	N^*	Last/first	α (fixed)	α (free)	N_0 (free)	Max. err.
Baseline	29.43	29/30	0.4994	0.4983	29.439	0.18%
Heavy / low friction	26.16	26/27	0.4996	0.4992	26.164	0.11%
Light / high friction	39.24	39/40	0.4993	0.4994	39.239	0.19%
Large body	49.05	49/50	0.4984	0.4993	49.043	0.16%
Low friction	14.71	14/15	0.4996	0.4992	14.718	0.16%

Time step (E4). In the baseline regime the exponent is 0.4989, 0.4993 and 0.5001 at $dt = 4$, 2 and 1 ms, and the relative error of T is below 0.5% at all three (Table 17).

Table 17: Physics-engine check, E4: dependence on the time step (baseline regime, $N = 30$–46).

Rate (Hz)	dt (ms)	α	Err. T ($N = 30$)	Err. T ($N = 46$)
250	4	0.4989	-0.06%	+0.49%
500	2	0.4993	-0.09%	+0.12%
1000	1	0.5001	+0.14%	+0.12%

Sensing error (E3). At $\sigma = 0$ and 0.15 rad the fraction of draws that move jumps from 0 at $N = 29$ to 1 at $N = 30$ (Figure 14, centre), so the 50% point is identified only to within one robot; the predicted thresholds (29.43 and 29.76) lie inside that interval. The logistic fit specified in the protocol is degenerate for such step data, so the 50% point in Table 18 is obtained by linear interpolation of the fraction that move (a deviation from the protocol); its bootstrap interval is undefined where the data are a step. At $\sigma = 0.30$ the engine threshold is 30.53 against 30.79 predicted (-0.8%). At $\sigma = 0.50$ the engine threshold is 32.58 (bootstrap 95% interval 32.38–32.89) against 33.35 predicted, a deviation of -2.3%, which *fails* the pre-specified 2% criterion; the prediction is outside the interval.

Table 18: Physics-engine check, E3: threshold versus sensing error. Engine: linearly interpolated 50% point over 20 draws per N (interval: bootstrap over draws, undefined where the data are a step). Last two columns: engine-free Monte Carlo (2×10^5 draws per N) of the threshold criterion with the parallel force $\sum \cos\theta_i > N^*$ and with the full net-force magnitude $|\sum F_i| > N^*$.

σ	$N^* e^{\sigma^2/2}$	Engine	95% interval	Rel. err.	MC, parallel	MC, full magnitude
0.00	29.43	29.50	–	+0.2%	29.50	29.50
0.15	29.76	29.50	–	-0.9%	29.50	29.50
0.30	30.78	30.53	[30.50, 30.59]	-0.8%	30.67	30.64
0.50	33.35	32.58	[32.37, 32.89]	-2.3%	33.31	33.17

Why the threshold falls below the prediction at large σ. Proposition 1 equates the *expected parallel* force with μmg. Two finite-N effects lower the threshold of an individual draw relative to that expectation. First, the body slips when the full net-force magnitude exceeds μmg, not only its component along the goal; an engine-free Monte Carlo of this criterion gives a 50% point of 33.17 at $\sigma = 0.5$, against 33.31 for the parallel criterion. Second, an off-centre net torque makes the body spin, which lowers the translational friction capacity in the engine's contact model: with rotation suppressed by a large inertia (exploratory, 40 draws per N, no interval), the engine threshold at $\sigma = 0.5$ rises from 32.73 to 33.06, within 0.4% of the full-magnitude value. We interpret the remaining gap as the effect of rotation coupling to friction, which Proposition 2 excludes by assumption, and we have not tested this beyond the diagnostic above. The prediction $N^* e^{\sigma^2/2}$ is therefore a good first-order estimate (within 1% for $\sigma \leq 0.3$ rad) and overestimates the engine threshold by about 2% at $\sigma = 0.5$.

10.3 Scope

The engine reproduces the threshold, the $-1/2$ exponent and its collapse across five regimes to within 0.2% in α and T, which supports the correctness of our integrator and contact handling and of the closed forms. The sensing-error result supports Proposition 1 to first order and identifies a $\sim 2\%$ finite-N and rotation-coupling effect at $\sigma = 0.5$. Bullet's friction is Coulomb-type with $\mu_s = \mu_k$ by construction, so the check does not test the Stribeck and static–kinetic effects of Section 4.5. No claim is made about physical robots or about decentralized control.

11 Discussion and Limitations

Established. The closed-form threshold and divergence law (Theorems 1–2) are exact within the stated idealization and are reproduced by the simulator in five regimes to 0.1%, with discretization error decreasing roughly in proportion to dt (Section 4.4). Under centralized control the checks test the implementation more than the physics. Under decentralized control with sensing error that is

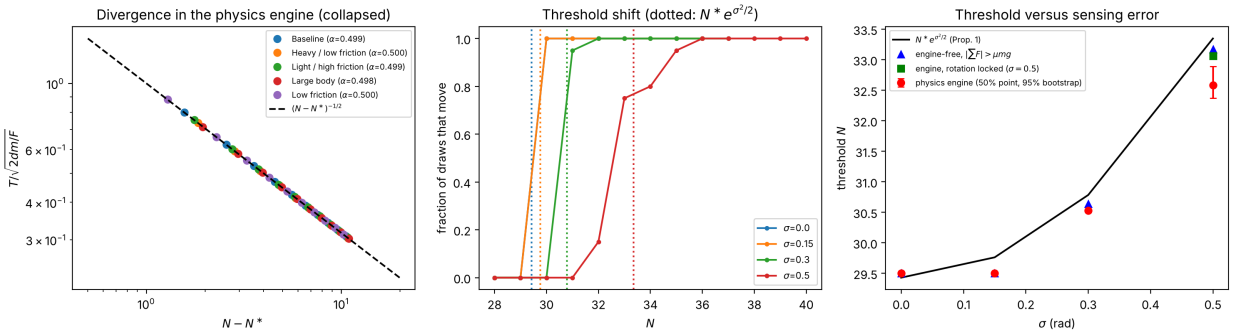


Figure 14: Physics-engine check. Left: $T/\sqrt{2dm/F}$ against $N - N^*$ in five regimes, collapsing onto $(N - N^*)^{-1/2}$ (dashed). Centre: fraction of draws that move against N for four sensing-error levels (dotted: $N^* e^{\sigma^2/2}$). Right: threshold versus σ: Proposition 1 (line), the engine (points, bootstrap interval), the engine-free full-magnitude criterion (triangles) and the engine with rotation suppressed at $\sigma = 0.5$ (square).

not compensated, the expected threshold shift $N^*_{\text{eff}} = N^* e^{\sigma^2/2}$ matches the fitted threshold of 20 trained policies and an untrained control within 0.5% over a range of noise levels (random goal). For world-frame forces and translational friction, rotation does not change the translational law (Proposition 2), and the simulation agrees.

Not established. (i) That a learned controller *acquires* the law: an untrained identity policy matches trained ones, and with randomized goal training converges to it. (ii) Any benefit of training in the setting of Sections 5–8: we found no training-offset effect there, and the ES and PPO comparisons of Appendix B were made in a regime where the identity baseline already attains the reported accuracy. In the harder task of Section 9, the benefit of learning over tuned designed controllers is small, restricted to nominal and high noise, and not significant on fresh seeds against the strongest baselines. (iii) The behavior of controllers with richer observations (e.g. velocity direction), which can infer and cancel their own bias and drop below $N^* e^{\sigma^2/2}$; Proposition 1 does not bound such controllers, and Section 9 studies them only empirically and in one regime. (iv) The physical model: all results are simulation-only (an independent physics engine, Section 10, reproduces the mechanical results but is not a physical experiment), friction is Coulomb-type, and Proposition 2 relies on world-frame forces and translational friction; body-fixed forces or friction torque would couple rotation to translation and need separate study. (v) The decentralized results use 5 noise trials per count and 5 seeds in the ablations and full-dynamics runs; the origin of the 0.3–0.5% offset between fitted and predicted thresholds is not isolated (the finite-N and rotation-coupling effects of Section 10 are of the same kind but were not tested for these policies). (vi) Implementation provenance: the decentralized results are from an independent implementation written from the model description, and earlier results from a different implementation (Appendix B) could not be reproduced (fixed-goal sensitivity slope, threshold location, and the training-offset effect differ); the difference likely stems from unspecified details of the loss and parametrization, but we have not identified it. (vii) The Stribeck rate prediction $1/\lambda$ is only order-of-magnitude accurate (ratio 0.95–2.6), and the power-law comparison used a fixed threshold.

12 Conclusion

Collective transport exhibits an exact mechanical threshold $N^* = \mu mg/F_{\max}$ and a parameter-independent divergence $T(N) \sim (N - N^*)^{-1/2}$, reproduced in rigid-body simulation across five regimes with convergence in the time step, and replaced by a logarithmic divergence under Stribeck friction. For decentralized controllers that follow their noisy sensed goal direction, sensing error shifts the threshold to $N^* e^{\sigma^2/2}$, matching 20 trained policies and an untrained control within 0.5%. The law survives decentralized control because a controller built as a correction to its sensed direction satisfies it without learning; trained policies with randomized goals converge to that controller, while fixed-goal policies exploit the fixed goal to do better. Section 9 then gives robots observations that allow inferring their own bias: an MAPPO policy with residual action, curriculum and memory reaches nominal success 0.813 ± 0.025 (10 seeds), far above standard MARL baselines and marginally above tuned designed controllers, but below them at low noise. Future work should (a) establish whether a better-designed controller or a better learner decides this comparison, (b) test body-fixed forces and friction torque, where rotation and translation couple, and (c) validate on physical hardware.

A An earlier negative result: local communication

Before the task above we tried MARL with local neighbour communication in the setting of Section 5 (translational dynamics, fixed $N^* = 29.43$, goal in a fixed direction; velocity-noise settings A and B). Using five seeds per method, MAPPO with neighbour-averaged messages ($k = 2, 4$) matched but did not exceed plain MAPPO or IPPO at the moderate and harder sensing-noise levels, and was beaten at the highest noise by a hand-designed neighbour-average controller with $k = 4$ (success 0.933 versus 0.938 for the designed controller at the hardest regime of setting B, and 0.990 versus 1.000 at the intermediate regime). When velocity observations were noisy (setting B), the communicating variants clearly beat the non-communicating learned methods. These runs were exploratory; their code and raw results accompany this paper.

B Results from the earlier decentralized-control implementation

Before the study in Section 5, decentralized control was studied with a different implementation (translational dynamics, goal direction fixed, 20 seeds, baseline regime), with these results: success 1.000 ± 0.000, $\alpha = 0.505 \pm 0.009$ (range $[0.481, 0.516]$), $N_0 = 30.00 \pm 0.08$, and a sensitivity slope 1.330 ± 0.002 for the applied angle with respect to the sensed goal angle. We retain these results for completeness (Tables 19–24).

Why they are not used as primary evidence. (1) An untrained identity policy gives success 1.000, $\alpha = 0.487$ and $N_0 = 29.81$ (Section 5.3), so the reported accuracy does not distinguish trained from untrained controllers. (2) The goal direction was fixed, and a policy could satisfy the task by memorizing a constant; the slope-based sensitivity probe cannot rule this out in this architecture, because the output is the sensed angle plus a correction (slope 1 by construction), and the masked-input control fails to behave as a control for the same reason (Section 5). (3) The re-implementation in Section 5 does not reproduce these numbers: with the same fixed goal it yields a slope of 0.23 ± 0.09 and $N_0 = 29.40 \pm 0.07$, and it finds no training-offset effect (success 1.000 for mixed offsets versus 0.812 ± 0.006 below). The earlier comparison with ES and PPO also used a loss and parametrization that we cannot reconstruct. Readers should treat the tables below as results of that implementation under its own settings.

Evolution Strategies and PPO (earlier implementation). ES (population 16, $\sigma = 0.05$, rank-based shaping) and PPO (GAE $\gamma = 0.99$, $\lambda = 0.95$, clip 0.2, stochastic Gaussian head) were trained under the same budget of 691,200 scenario-timesteps as gradient-based training, from 20 seeds each.

Table 19: Earlier implementation: three training algorithms, matched compute (691,200 scenario-timesteps), fixed goal, 20 seeds each.

Method	Success rate	α_{fit}	N_0	Final loss
Gradient-based	1.000 ± 0.000	0.505 ± 0.009	30.00 ± 0.08	0.0024 ± 0.0002
Evolution Strategies	1.000 ± 0.000	0.512 ± 0.027	30.00 ± 0.19	0.0145 ± 0.0173
PPO	1.000 ± 0.000	0.503 ± 0.009	29.93 ± 0.07	–

In that implementation ES showed larger seed-to-seed spread than gradient-based training (exponent std 0.027 versus 0.009) and PPO matched it (0.009); hyperparameters were not tuned extensively. Given point (1) above, these comparisons show that all three optimizers reach a controller with the identity baseline's accuracy; they do not show that the law was learned.

Noise sensitivity, one policy (earlier implementation).

Table 20: Earlier implementation: one trained policy (seed 0) re-evaluated at five sensing-noise levels, versus $N^* e^{\sigma^2/2}$ with $\sigma^2 = 0.10^2 + \sigma_{\text{step}}^2$.

σ_{step} (rad)	Success rate	α_{fit}	N_0 (observed)	N_{eff}^* (predicted)	Relative error
0.00 (noiseless)	1.000	0.507	29.597	29.578	0.07%
0.05	1.000	0.506	29.643	29.615	0.10%
0.15 (training level)	1.000	0.507	29.960	29.912	0.16%
0.30 (2× training)	0.938	0.502	31.089	30.939	0.48%
0.50 (3.3× training)	0.812	0.488	33.865	33.516	1.04%

Ablations (earlier implementation). Observation-feature and width ablations were run from a mixed-offset baseline (2–3 seeds per variant), not the fixed-hardest-offset procedure; the offset comparison used 20 seeds for fixed and mixed offsets and 2 seeds for offset 2.

Table 21: Earlier implementation: observation-feature ablation (mixed-offset baseline, fixed goal).

Variant	Seed	Final loss	Success rate	α_{fit}
Full observation, mixed-offset baseline	0–19	2.31 ± 0.11	0.812 ± 0.006	0.503 ± 0.050
No attachment-angle	0	2.194	0.825	0.488
No attachment-angle	1	2.205	0.800	0.449
No attachment-angle	2	2.135	0.812	0.510
No speed-toward-goal	0	2.201	0.825	0.526
No speed-toward-goal	1	2.209	0.787	0.494
No speed-toward-goal	2	2.135	0.812	0.507
No sensed goal direction	0	2.546	0.875	0.507

Table 22: Earlier implementation: hidden-layer width (mixed-offset baseline, two seeds per width).

Hidden width	Seed	Final loss	Success rate	α_{fit}
4	0	2.262	0.825	0.494
4	1	2.274	0.812	0.448
16	0	2.178	0.800	0.528
16	1	2.220	0.812	0.573

Table 23: Earlier implementation: fixed hardest offset versus mixed offsets (success rate and α across offsets 1–16).

Training condition	Seed(s)	Final loss	Success rate	α_{fit}
Fixed hardest offset	0–19	0.0024 ± 0.0002	1.000 ± 0.000	0.505 ± 0.009
Mixed offsets 1–12	0–19	2.31 ± 0.11	0.812 ± 0.006	0.503 ± 0.050
Fixed offset 2	0	0.0085	0.975	0.563
Fixed offset 2	1	0.0104	0.988	0.545

Table 24: Earlier implementation: 20 independent runs, fixed goal, fixed-hardest-offset procedure.

Seed	Final loss	Success rate	α_{fit}	N_0	Sensitivity slope
0	0.0023	1.000	0.507	29.96	1.333
1	0.0026	1.000	0.510	29.98	1.329
2	0.0022	1.000	0.514	29.89	1.329
3	0.0020	1.000	0.508	29.93	1.332
4	0.0024	1.000	0.515	29.91	1.329
5	0.0025	1.000	0.497	30.06	1.330
6	0.0026	1.000	0.500	30.06	1.332
7	0.0021	1.000	0.498	30.02	1.330
8	0.0026	1.000	0.487	30.12	1.328
9	0.0024	1.000	0.512	29.99	1.326
10	0.0025	1.000	0.513	29.95	1.330
11	0.0026	1.000	0.516	29.91	1.332
12	0.0022	1.000	0.508	30.01	1.330
13	0.0020	1.000	0.506	29.97	1.333
14	0.0024	1.000	0.504	30.05	1.331
15	0.0021	1.000	0.513	29.95	1.332
16	0.0022	1.000	0.505	30.01	1.332
17	0.0023	1.000	0.496	30.09	1.325
18	0.0027	1.000	0.481	30.19	1.331
19	0.0024	1.000	0.511	29.98	1.328
All 20: mean ± std	0.0024 ± 0.0002	1.000 ± 0.000	0.505 ± 0.009	30.00 ± 0.08	1.330 ± 0.002

In the earlier implementation we observed that individual seed outcomes were somewhat sensitive to functionally inert code changes (plausibly because trajectories near the stick/slip transition amplify floating-point differences), so exact point estimates above might shift under a bitwise-identical rerun.

Funding Declaration

The authors declare that no funding was received for the research, authorship, and publication of this article.

References

[1] V.-L. Heuthe, E. Panizon, H. Gu, C. Bechinger. *Counterfactual rewards promote collective transport using individually controlled swarm microrobots.* Science Robotics, 9(97):eado5888, 2024.

[2] H. Farivarnejad, S. Berman. *Multirobot Control Strategies for Collective Transport.* Annual Review of Control, Robotics, and Autonomous Systems, 5:205–219, 2022.

[3] Y. Hu, L. Anderson, T.-M. Li, Q. Sun, N. Carr, J. Ragan-Kelley, F. Durand. *DiffTaichi: Differentiable Programming for Physical Simulation.* International Conference on Learning Representations (ICLR), 2020.

[4] C. D. Freeman, E. Frey, A. Raichuk, S. Girgin, I. Mordatch, O. Bachem. *Brax – A Differentiable Physics Engine for Large Scale Rigid Body Simulation.* arXiv:2106.13281, 2021.

[5] O. Feinerman, I. Pinkoviezky, A. Gelblum, E. Fonio, N. S. Gov. *The Physics of Cooperative Transport in Groups of Ants.* Nature Physics, 14:683–693, 2018.

[6] T. Heckenthaler, T. Holder, A. Amir, O. Feinerman, E. Fonio. *Connecting Cooperative Transport by Ants with the Physics of Self-Propelled Particles.* PRX Life, 1:023001, 2023.

[7] A. Chatterjee, T. Tzook, N. Gov, O. Feinerman. *Maximal Response to a Mechanical Leader at Critical Group Size in Ant Collectives.* Nature Communications, 16, 2025.

[8] S. Berman, Q. Lindsey, M. S. Sakar, V. Kumar, S. C. Pratt. *Experimental Study and Modeling of Group Retrieval in Ants as an Approach to Collective Transport in Swarm Robotic Systems.* Proceedings of the IEEE, 99(9):1470–1481, 2011.

[9] H. F. McCreery, M. D. Breed. *Cooperative Transport in Ants: A Review of Proximate Mechanisms.* Insectes Sociaux, 61:99–110, 2014.

[10] E. Arbel, L. L. K. M. Buise, C. C. R. M. M. van Waes, N. Oppenheimer, Y. Lahini, M. Y. Ben Zion. *A Mechanical Route for Cooperative Transport in Autonomous Robotic Swarms.* Nature Communications, 2025 (arXiv:2402.05659, 2024).

[11] J. K. Terry, N. Grammel, A. Hari, L. Santos, B. Black. *Revisiting Parameter Sharing in Multi-Agent Deep Reinforcement Learning.* arXiv:2005.13625, 2020.

[12] J. K. Gupta, M. Egorov, M. Kochenderfer. *Cooperative Multi-agent Control Using Deep Reinforcement Learning.* International Conference on Autonomous Agents and Multiagent Systems (AAMAS), 2017.

[13] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, I. Mordatch. *Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments.* Advances in Neural Information Processing Systems (NeurIPS), 30, 2017.

[14] T. Salimans, J. Ho, X. Chen, S. Sidor, I. Sutskever. *Evolution Strategies as a Scalable Alternative to Reinforcement Learning.* arXiv:1703.03864, 2017.

[15] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov. *Proximal Policy Optimization Algorithms.* arXiv:1707.06347, 2017.

[16] B. Armstrong-Hélouvry, P. Dupont, C. Canudas de Wit. *A Survey of Models, Analysis Tools and Compensation Methods for the Control of Machines with Friction.* Automatica, 30(7):1083–1138, 1994.