Flow currents support simple and versatile trail-tracking strategies

Haotian Hang [1], Yusheng Jiao [1], Josh Merel,[2] and Eva Kanso [1,3,*]

[1]*Department of Aerospace and Mechanical Engineering, University of Southern California, Los Angeles, California 90089, USA*

[2]*Fauna Robotics, New York City, New York, USA*

[3]*Department of Physics and Astronomy, University of Southern California, Los Angeles, California 90089, USA*

(Received 30 March 2025; accepted 23 July 2025; published 12 January 2026)

Aquatic animals offer compelling evidence that flow sensing alone, without vision, is sufficient to guide a swimming organism to the source of an unsteady hydrodynamic trail. However, the sensory feedback strategies that allow these remarkable trail-tracking abilities remain opaque. Here, by integrating mechanistic flow simulations with reinforcement learning techniques, we discovered two simple and equally effective strategies for hydrodynamic trail following. Though not *a priori* obvious, these strategies possess parsimonious interpretations, analogous to Braitenberg's simplest vehicles, where the agent senses local flow signals and turns away from or toward the direction of stronger signals. A rigorous stability analysis shows that the effectiveness of these strategies in robustly tracking flow currents is independent of the type of sensor but depends on sensor placement and the traveling nature of the flow signal. Importantly, these results inform a suite of versatile strategies for hydrodynamic trail following applicable to both vortical and turbulent flows. These insights support the future design and implementation of adaptive real-time sensory feedback strategies for autonomous robots in dynamic flow environments.

DOI: 10.1103/qj9q-p7rg

I. INTRODUCTION

The physical environment plays a pivotal role in shaping the sensory, behavioral, and survival strategies of living organisms [1–4]. Chemotactic bacteria inhabit a viscous flow environment, where chemical concentrations are governed by diffusion (see Refs. [5,6] and Fig. 1(a)), and employ a run-and-tumble strategy to navigate toward nutrient sources and away from toxins [7–9]. Minute insects and crustaceans, including drosophilas [10–12], moths [13], and copepods [14–17], inhabit a turbulent flow environment. To find potential mates, they detect pheromones carried by turbulent wind or water current, where encounters with odor patches are random and intermittent; see Ref. [18] and Fig. 1(b). Under such information-scarce conditions, trail following requires elaborate sensory and movement strategies that rely on constructing spatial maps [18] or remembering flow direction and temporal history of the detected signal [19]. Larger animals, such as fish [20–22] and harbor seals [23,24], do not experience the world as merely laminar or turbulent. They encounter a rich array of flow patterns and coherent vortex structures [25–27] created by conspecifics, predators, prey, and even abiotic sources [Fig. 1(c)]. These flows contain, in principle, useful information about the source generating them that can be deciphered by an aspiring follower [26–28]. However, effective feedback strategies for tracking coherent vortical flows remain largely unexplored.

Hydrodynamic trail following is a critical task for many aquatic organisms [20,24], with direct implications for developing navigation strategies in underwater robotic vehicles [29–32]. Various organisms detect flow signals [20–22] to negotiate conditions with reduced or no visibility [33] or to reduce the cost of locomotion [34–36]. The flow-sensitive mechanoreceptors vary by species, from the lateral line system in fish [37–41] to the whiskers of harbor seals [24,42]. The underlying principles that allow flow detection [43,44] have direct implications to designing bioinspired flow sensors [45–47]. Here, we are not concerned with sensor design, but with the navigation ability of flow-sensitive swimmers. Indeed, fish [20,22,33] and harbor seals [24,42] follow hydrodynamic trails for underwater navigation not only over short distances, as in the final stage of mate or prey localization [33], but also to track and locate distant objects over considerable length and time spans [20,21,24,42]. Additionally, olfaction is often integrated with flow sensing [22], as seen in the homing behavior of the Pacific salmon [48]. Yet, the minimal sensory feedback laws required for a swimmer to respond to local flow signals and follow hydrodynamic trails remain obscure.

Early efforts for designing control strategies to navigate water currents date back to the work of Zermelo [49], where a swimmer is required to navigate toward a target while being influenced by a well-defined laminar flow field. More recently, techniques from optimal control [50,51], model predictive control [52], and reinforcement learning (RL) [50,51,53] were applied to solve this problem in unsteady flow fields, including

*Contact author: kanso@usc.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

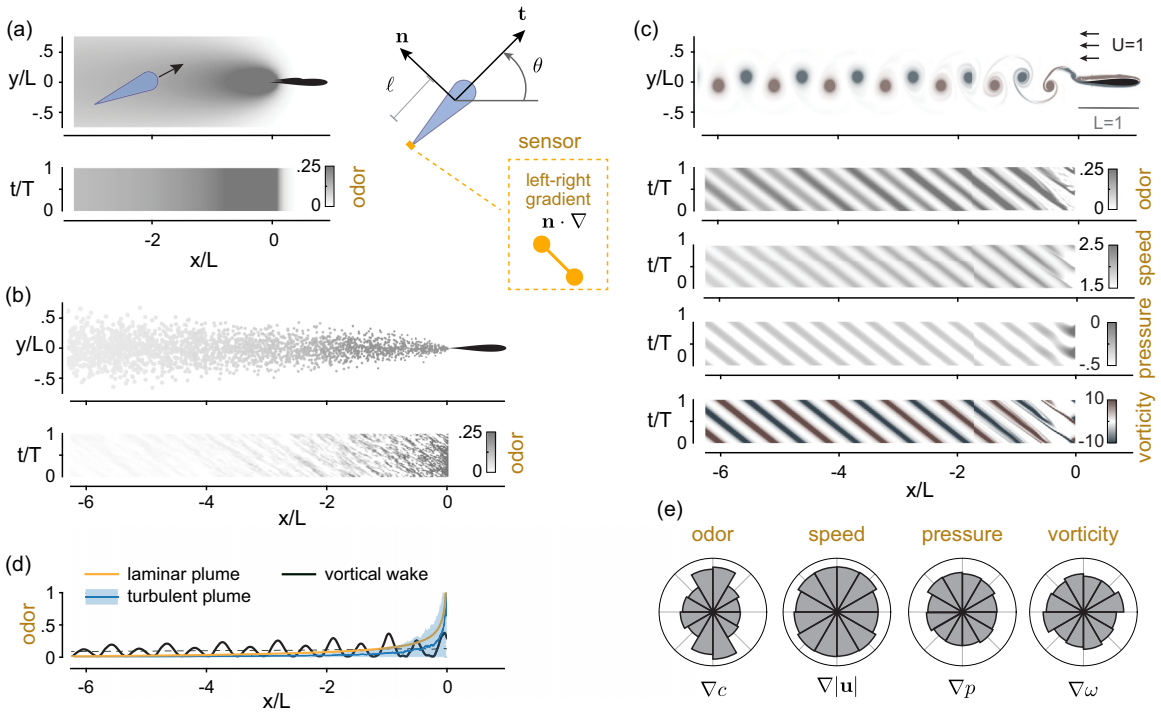

FIG. 1. Scented and odorless trails at different scales. (a) In laminar plumes, odor is released from a source at a fixed emission rate $R = 1$ and is advected by a uniform background flow while undergoing molecular diffusion at $\mathrm{Pe} = 10$: (top row) steady-state solution; (bottom row) concentration along the midline $y/L = 0$. An agent unaware of the source location is modeled as a swimmer with body-fixed frame $(\mathbf{t}, \mathbf{n})$, moving at a constant speed V and with control over its rate of orientation $\dot{\theta}$, where θ is the angle between its heading direction $\mathbf{t}$ and the inertial x axis. (b) To simulate turbulent plumes, odor packets are released stochastically from the source and transported by the turbulent wind: (top row) odor packets of decaying strength undergoing random cross-stream perturbations; (bottom row) odor concentration along the midline $y/L = 0$. (c) Hydrodynamic trail created by a Joukowsky foil undergoing pitching oscillations in uniform background flow $U = 1$ with tailbeat amplitude $A = 0.2$ and frequency $f = 1.25$, at $\mathrm{Re} = 5000$ and $\mathrm{St} = 0.25$. Odor released from the airfoil's trailing edge is carried by the flow field and undergoes diffusion at $\mathrm{Pe} = 10\,000$. (Bottom rows) Physical quantities along the midline $y/L = 0$, including odor, flow speed, pressure, and vorticity, exhibit traveling-wave characteristics. (d) Decay rates of the concentration fields in panels (a), (b), and (c) along the midline $y/L = 0$: Vortical flow shows the weakest decay. (e) Statistics of gradient directions of each of the physical quantities—odor, speed, pressure, and vorticity—evaluated along a regular grid in the flow field of panel (c). Gradients have no preference for upstream or downstream directions.

turbulent [53] and vortical flows [50–52]. Interestingly, local [50] and egocentric [51] flow sensing abilities are sufficient to enable the agent to perform efficient navigation, with performance comparable to time-optimal control utilizing full knowledge of the flow field. Behavioral strategies relying only on local sensory information were also devised for soaring birds [54,55], schooling fish [56], and vertical migration of copepod in turbulent flows [17].

Hydrodynamic trail tracking is a related yet different problem. Instead of moving toward a target whose location is independent of the flow source, the navigator is tasked to move toward the source that generates the flow. This navigation problem is often considered in the context of tracking turbulent plumes, where odor patches are transported randomly [57,58]. For example, infotaxis uses Bayesian-based methods to build and update a spatial map of the possible location of the source [18]. Alternative strategies for tracking turbulent plumes rely on reinforcement learning [19,59] or use a finite-state machine approach with limited memory [60]. However, turbulent odor tracking is distinct from tracking hydrodynamic trails characterized by long-lasting coherent flow structures [25,61,62]. To track such trails, strategies where the agent collects a time history of the local flow signal and acts accordingly were shown to function successfully in a variety of vortical flows [27,63]. While promising, these handcrafted strategies are suboptimal in terms of sensory requirements.

In this study, we use deep reinforcement learning to uncover efficient tracking strategies that rely purely on local and instantaneous flow sensing. Specifically, we consider a trail-following task with two main features: (1) The swimmer can only sense local flow information and (2) the swimmer can only sense gradients of the local flow, not the flow itself. That is, we consider a scenario with a completely blind swimmer with no prior or developed knowledge of the source location, inertial coordinates, or overall flow direction. These stringent constraints pose significant challenges to robustly following hydrodynamic trails, whether by locally sensing the flow itself or by sensing the odor concentration carried by the flow. We thus explore a variety of odor and flow sensors. We find parsimonious strategies for trail following, where success depends on sensor placement and the traveling nature of the flow signal, but is largely independent of the sensor type. Importantly, we demonstrate that these strategies can easily translate from coherent to turbulent flows. To conclude, we compare the

TABLE I. Comparison to existing trail-following strategies. Sensory feedback strategies employed by an autonomous agent for tracking a signal field to its source. We proposed two strategies that track turbulent plumes and vortical wakes, respectively. In turbulent plumes, our strategy does not require knowledge of the mean flow direction, unlike in Refs. [11,19,58,64], and uses less memory and storage than existing strategies designed for the turbulent regime [18,60,65]. In vortical flows, our strategies respond instantaneously to the spatial gradient with no memory requirements, unlike in Ref. [27], and its performance is optimized using deep reinforcement learning.

Strategy	Flow environment	Signal	Signal gradient	Nonlocal information	Memory	Method	Related organism
Gradient ascent [9]	Laminar	No	Yes	No	No	Bioinspired	Bacteria
Run and tumble [9]	Laminar	No	No	No	Yes	Bioinspired	Bacteria
Infotaxis [18]	Turbulent	No	No	Mean flow	Spatial map	Bayesian	–
Finite-state controller [60]	Turbulent	No	No	No	Latent states	RL	–
[65]	Turbulent	Yes	No	No	Few time steps	RL	Mouse
Surge and cast [58]	Turbulent	No	No	Mean flow	Yes	Bioinspired	Moth
RL with RNN[a] [19]	Turbulent with mean flow	Yes	No	Mean flow	RNN latent states	RL	–
[12]	Turbulent with mean flow	No	No	Mean flow	One time step	Handcrafted	*Drosophila* larva
[11,64]	Turbulent with mean flow	Yes	No	Ground and air speed	–	Handcrafted	*Drosophila*
This study	Turbulent with mean flow	Yes	No	No	One time step	Handcrafted	–
Phase gradient [27]	Vortical wake	No	Yes	No	Entire time period	Handcrafted	–
This study	Vortical wake	No	Yes	No	No	RL	–

[a]RNN represents Recurrent neural network.

strategies developed in this work to existing approaches where the navigator has knowledge of an external reference frame [32,50], overall flow direction [19], vision-based information [50,51,56], or past sensory information [18,19,59] (Table I), and we discuss the implications of these findings for future underwater robotic applications.

II. RESULTS AND DISCUSSION

A. Laminar, turbulent, and vortical hydrodynamic trails

Consider a source of size L moving in a fluid at a constant speed U with a constant odor emission rate R. The characteristics of the odor concentration field depend on the length and speed of the source, as well as the fluid environment, such as water or air. At the micron scale in water, the Reynolds number $\mathrm{Re} = UL/\nu$, where ν is the kinematic viscosity of the fluid, describing the ratio of inertial forces to viscous forces, is nearly zero. The Péclet number $\mathrm{Pe} = UL/D$, where D is the molecular diffusivity of the odor in fluid, describes the ratio between advective and diffusive timescales and is typically small (less than 100) [5,6,66,67]. Thus, the odor concentration diffuses into a laminar plume, governed by the advection-diffusion equation, with a single maximum at the source location [Fig. 1(a); Appendix A 1]. For a millimeter-scale insect moving in air, Re and Pe are both large. The signal field is transported by a turbulent flow [Fig. 1(b); Appendix A 2]. Here, we employed a particle-based two-dimensional (2D) plume model to simulate the random release of odor packets from the source [19,68]: The odor packets get advected by a mean downstream flow and random cross-stream perturbations, while they decrease in intensity and increase in size over time.

For a source of moderate size and speed, such as a swimming fish [20,24,25], the wake is characterized by coherent vortical structures [Fig. 1(c); Appendix A 3]. Here, we solved the Navier-Stokes equation, using an open-source computational fluid dynamics (CFD) solver [69–71], to obtain the flow field past a pitching airfoil following sinusoidal oscillations in a uniform background flow, where we set the Reynolds number to $\mathrm{Re} = 5000$ and the Strouhal number $\mathrm{St} = fL/U$ to $\mathrm{St} = 0.25$, with $f = 1/T$ being the flapping frequency and T the flapping period. This Strouhal number is comparable to those observed in swimming fish, linking the frequency of tailbeat oscillations to the swimming speed and body length (Appendix A 3). We solved for the concentration field numerically, considering the advection-diffusion equation at Péclet number $\mathrm{Pe} = 10^4$. Considering the odor concentration as the signal field detectable by an aspiring follower, we find that it resembles a traveling wave characterized by wavelength λ and frequency f. In fact, this traveling-wave characteristics emerges in all physical quantities, odor, speed, pressure, and vorticity, in the wake of the source [Fig. 1(c)].

Whereas the signal field in the laminar plume [Fig. 1(a)] has a single local maximum at the source, in the turbulent [Fig. 1(b)] and vortical [Fig. 1(c)] trails, an aspiring follower encounters multiple local maxima and minima. Importantly, the signal decays downstream at different rates [Fig. 1(d)]: In the laminar and turbulent regimes, the signal decays exponentially downstream because of molecular and turbulent diffusion [72], whereas in comparison, the signal lasts much longer in the vortical wake [Fig. 1(d)]. For an aspiring trail follower, the dynamics of the signal fields are intimately connected to which sensing and response strategies are viable. In the laminar flow regime, it is feasible to measure the gradient of the signal field and use gradient ascent methods to reach the source, but spatial gradients are not well defined in the turbulent regime, and gradient ascent tends to get stuck in local maxima in the vortical flow regime. We address these challenges first in the intermediate coherent flow regime, where we develop trail-following strategies that rely on flow or odor sensing using reinforcement learning methods. We then discuss extensions

TABLE II. Training instances of RL policies based on different sensory cues. We conducted multiple training instances for four different signals transported by the flow field. The cumulative reward is shown in Figs. 2 and 8 based on representative training instances for each sensory cue.

	Sensory cue	Number of training instances	Convergence fraction	Excitatory fraction	Inhibitory fraction
Speed	$\boldsymbol{n}\cdot\nabla\lvert\boldsymbol{u}\rvert$	20	17/20	8/17	9/17
Vorticity	$\boldsymbol{n}\cdot\nabla\lvert\omega\rvert$	15	0/15	–	–
Pressure	$\boldsymbol{n}\cdot\nabla p$	11	9/11	0/9	9/9
Odor	$\boldsymbol{n}\cdot\nabla c$	9	9/9	9/9	0/9

of these strategies to the turbulent regime based on odor sensing.

B. Learning to track vortical trails

We modeled the trail follower [Fig. 1(a)] as a self-propelled artificial agent, with no inertia, moving at a constant speed V and heading at a time-dependent angle $\theta(t)$ from the x axis [73–75]. We first considered an agent with mechanosensing abilities, where inspired by the fish lateral line system [21,39,40] and functional evidence suggesting that flow receptors are optimized to correlate with differential flow signals [76,77], we considered two sets of flow sensors located at ℓ_{head} and ℓ_{tail} from the center, which measure the local gradients of flow speed normal to the agent's swimming direction: $s_{\text{head}} = (\mathbf{n}\cdot\nabla|\mathbf{u}|)_{\text{head}}$ and $s_{\text{tail}} = (\mathbf{n}\cdot\nabla|\mathbf{u}|)_{\text{tail}}$. Aside from these two sensory measurements, the agent is at all times ignorant of all other information about the flow field and blind to its position relative to the source generating the flow. To respond to this limited and localized information without memory of past measurements, we provided the agent with control over only its angular velocity $\dot\theta$, as opposed to direct control over its heading angle θ [50]. The agent's motion is thus described by the nonholonomic unicycle model [73–75],

$$\dot x = V\cos\theta, \quad \dot y = V\sin\theta, \quad \dot\theta = \pi(s_{\text{head}}, s_{\text{tail}}). \tag{1}$$

We aimed to learn a policy $\dot\theta = \pi(s_{\text{head}}, s_{\text{tail}})$ that guides the agent to track the vortical wake to its source from trial-and-error interactions with the flow environment. We employed model-free, deep reinforcement learning based on the Proximal Policy Optimization algorithms [78,79] (Appendix B). Actions taken by the policy are restricted to the range $\dot\theta = [-\Omega, \Omega]$ to avoid unrealistically sharp turns. Each training session consisted of 30 000 episodes, with parameter values $\Omega = 3U/L$, $V = 0.25U$, and $\ell_{\text{head}} = \ell_{\text{tail}} = 0.25L$. In each episode, the agent was assigned a random initial position, orientation, and phase in the period of the unsteady wake.

Inspired by Monte Carlo approaches to random processes, we performed 20 independent training instances, each starting from a random policy (Table II). Surprisingly, these 20 training instances consistently converged to two classes of policies, as evident from the evolution of the cumulative reward in Fig. 2(b): Nearly half the training instances converged to policies with reward shown in red and nearly half converged to policies with reward shown in blue (Table II). In Fig. 2(c), we show representative policies from each of the two classes of policies with colormaps representing the RL action $\dot\theta$ over the observation space $(s_{\text{head}}, s_{\text{tail}})$. The two classes of policies exhibit striking features: Both are largely independent of the measurement at the head, and, although both converge during training, they exhibit opposite behaviors. The red policy, which we call hereafter *excitatory*, instructs the agent to turn in the same direction as the flow gradient at tail, that is, in the direction of increasing flow speed. The blue policy, which we denote *inhibitory*, instructs the agent to turn in the opposite direction to the flow gradient at tail, that is, in the direction of decreasing flow speed. Deploying either policy, the agent successfully tracks the wake to its generating source (Fig. 2(d); Movie S1 in the Supplemental Material [80]). When applying the excitatory strategy, it ventures more into the wake, slaloming between vortices, similar to experimental observations in live animals [20,24,34,42].

C. Learned policies are interpretable and analogous to Braitenberg's vehicles

These features of the RL policies—independence of s_{head} and nearly binary action [Fig. 2(c)]—suggest that they can be approximated analytically as bang-bang controllers [81] whose behavior depends only on the sensory measurement at tail s_{tail}. Thus, without loss of significant properties, we discounted the sensory measurement at head s_{head} entirely and used ℓ and s to refer to the location and measurement of a single sensor, with negative ℓ implying sensor location at tail. These simplifications lead to analytic expressions of the two classes of RL policies,

$$\text{excitatory:}\quad \dot\theta = \Omega\,\text{sgn}(s), \qquad \text{inhibitory:}\quad \dot\theta = -\Omega\,\text{sgn}(s). \tag{2}$$

When implementing these simplified policies, the agent exhibited similar behavior to that dictated by the full RL policies (Figs. 2(d) and 2(e); Movie S1 in the Supplemental Material [80]), while being readily interpretable and allowing for a rigorous stability analysis. These policies evoke Braitenberg's vehicles, originally proposed as thought experiments for constructing complex behavior from simple sensorimotor responses to environmental stimuli [82]. Braitenberg's vehicles have been applied to model animals' tropotaxic behaviors [83], including rheotaxis [84], phototaxis [85], and chemotaxis [86]. In the simplest vehicle, two motors that drive the right and left wheels of a vehicle are either directly or inversely connected to two sensors, and each motor turns more as the stimulus to which the sensor is tuned increases. As in our strategies, the vehicle's behavior depends on the differential signal between the two sensors. At zero difference, the vehicle moves straight ahead. When the motors are connected to sensors on the same side, the vehicle is aversive to stimulus and *turns away*, and when connected to opposite sensors, the

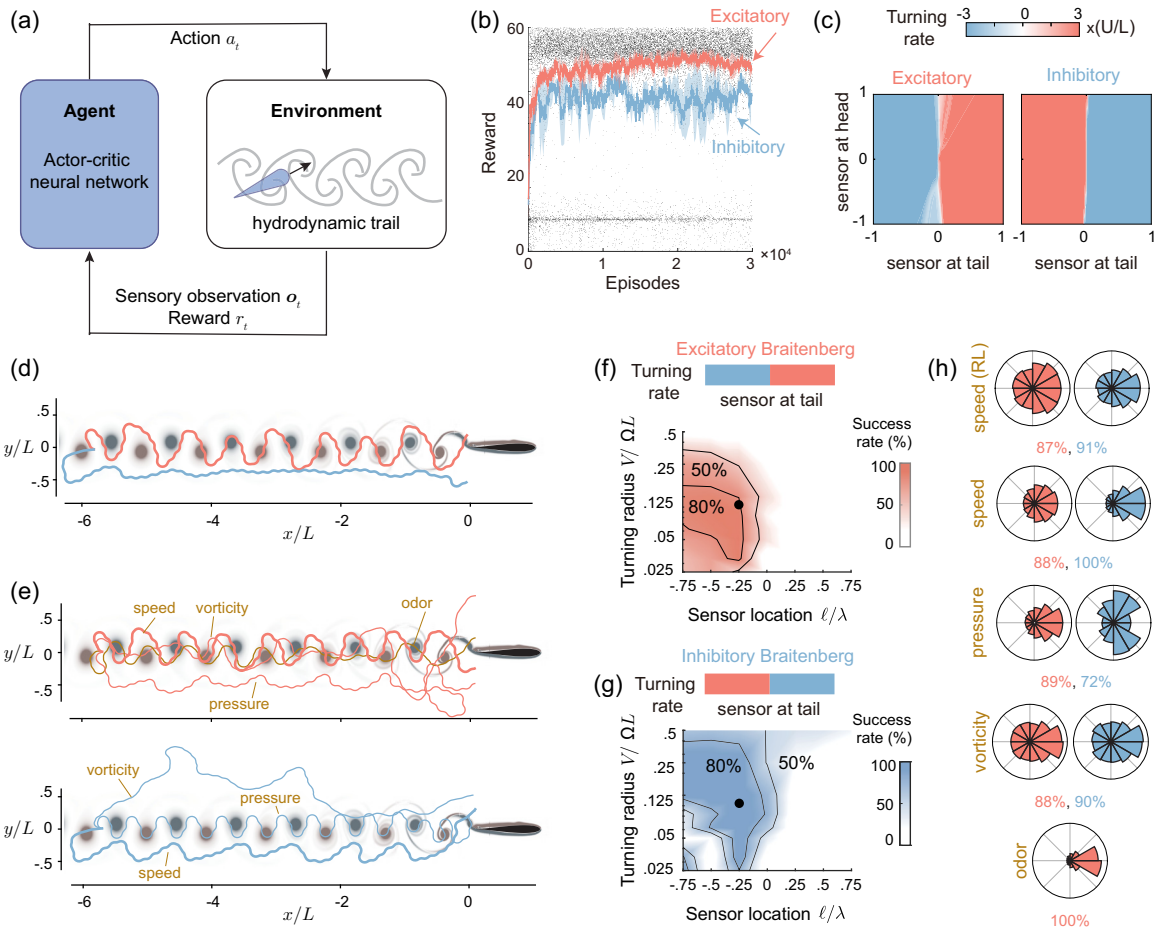

FIG. 2. RL policies and simpler Braitenberg-like policies for tracking hydrodynamic trails. (a) To follow vortical trails, we train, using Deep RL, a swimmer that senses the flow gradients at its head s_{head} and tail s_{tail} (observations $\mathbf{o}_t$) and responds by modulating its angular velocity $\dot{\theta}$ (action a_t). (b) Learning curves corresponding to 20 training instances, where nearly half converged to one of two classes of policies, respectively; moving average of the total reward is shown in red and blue; total reward for randomly chosen 10% of the training episodes is shown as black dots. Convergence rate of different training instances is reported in Table II. (c) Representatives of the two classes of converged policies shown as colormaps over the observation space (s_{tail}, s_{head}), normalized by the maximal value of the flow signal. Dominant features of these policies are captured by parsimonious Braitenberg-like strategies with excitatory and inhibitory responses to the flow signal measured only at tail [Eq. (2)] and associated depiction at the top of panels (f) and (g). (d) Sample trajectories based on the excitatory (red) and inhibitory (blue) RL policies. (e) Sample trajectories of the parsimonious representation of the excitatory (red) and inhibitory (blue) policies are superimposed based on different sensory cues. The same parameter values are used in RL training: $\ell = -0.25L$ and $\Omega = 3U/L$. These trajectories are also compared in Movie S3 in the Supplemental Material [80]. Success rate as a function of sensor location ℓ/λ and turning radius $V/\Omega L$ calculated over 76 500 test cases for each set of parameter values $(\ell/\lambda, V/\Omega L)$ for (f) excitatory and (g). inhibitory policies in Eq. (2), with flow speed as sensory cue. Parametric study performed using the other sensory cues—pressure, vorticity, and odor—is reported in Fig. 9. (h) Statistics on the gradient direction of the physical quantities measured by the onboard sensors along trajectories based on the RL or associated Braitenberg policies. The overall success rates over 76 500 test cases are listed for each sensory cue in red (excitatory) and blue (inhibitory).

vehicle is responsive to stimulus and *turns toward*. Likewise, in our excitatory policy, the swimmer turns toward larger flow speeds, while in the inhibitory policy, it turns away. In uniform flows, the swimmer swims straight.

D. Robustness to sensory limitations

To evaluate the performance of the RL policies and their simplified counterparts in Eq. (2), we systematically tested them in a domain defined by a Cartesian grid starting downstream of the source. At each grid point, we considered equally spaced initial orientation between 0 and 2π and phase within the pitching period. In total, we tested 76 500 initial conditions for each policy (Appendix B). The excitatory and inhibitory RL policies achieved success rates of 87% and 91%, respectively, while the simplified counterparts in Eq. (2) achieved success rates of 88% and 100%, respectively, all tested in the same wake used during training. When testing the RL policies, we employed the deterministic version by taking its mean value instead of employing a stochastic policy.

We then evaluated the distributions of the terminal lateral locations y/L from the source [Fig. 3(a)] and the associated arrival time t_f [Fig. 3(b)]. The excitatory policy is more accurate in honing in on the source [Fig. 3(a)], while the

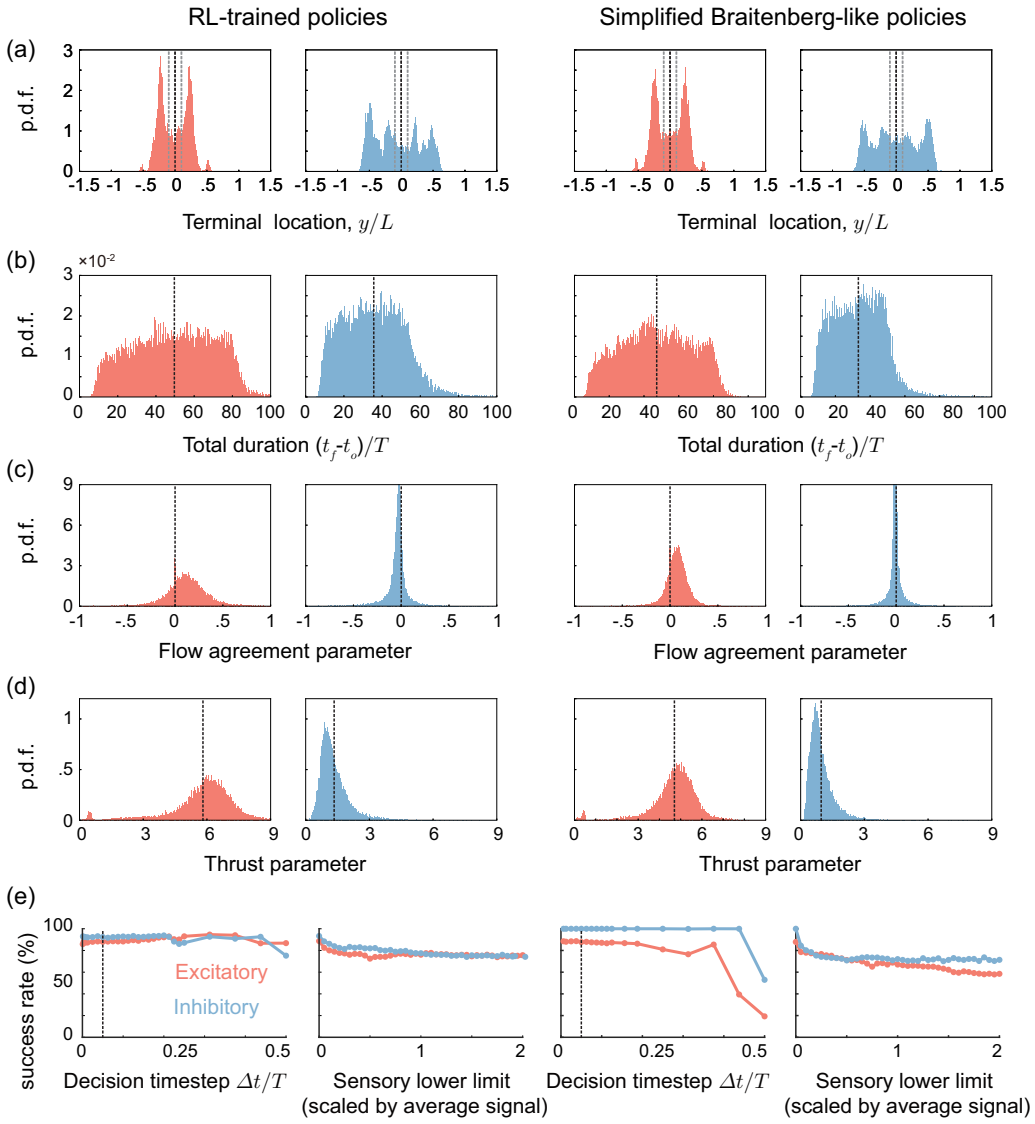

FIG. 3. Performance of RL-trained policies and simpler counterparts. Each policy is tested on 76 500 test cases downstream of the source. Statistics are based on all trajectories, out of the 76 500 test cases, that succeed in moving upstream. The overall success rates for RL-trained policies and their simplified counterparts are 86.82% and 87.5%, respectively, for the excitatory policies (red), and 91.47% and 100%, respectively, for the inhibitory policies (blue). (a) Distribution of terminal location y/L at t_f defined as the time when $x/L = 0$. Dashed gray lines represent the tailbeat amplitude of the pitching foil. (b) Distribution of total duration $t_f - t_o$ to reach the source normalized by the period $T = 1/f$ of the wake. The RL-trained policies are, on average, slower in reaching the terminal location than their simpler counterparts. (c) Flow agreement parameter and (d) thrust parameter along the successful trajectories. (e) Success rate as a function of sensory limitations for the excitatory (red) and inhibitory (blue) policies: (leftmost) RL-trained policies are robust to increasing decision time step Δt up to 50% of pitching period T, that is, up to the agent changing direction only twice during a flapping period of the source. Dashed vertical line indicates the decision time step used during training $\Delta t = 0.05T$; (left middle) RL-trained policies are robust to sensory constraints where the agent cannot discern and respond to flow gradients below a lower limit. Success rate of the RL-trained policies decreases more gradually compared to their simpler counterparts (right middle and rightmost).

inhibitory strategy locates the source faster [Fig. 3(b)] because the corresponding trajectories meander less in the wake [Figs. 2(d) and 2(e)]. This explains why the inhibitory policies are suboptimal compared to the excitatory ones in Fig. 2. By the same token, because it causes the agent to move inside the wake and slalom between vortices, the excitatory policy provides the agent with more opportunities to utilize the flow for efficient navigation [Figs. 3(c) and 3(d)]. To assess the potential for efficient navigation, we calculated two metrics along each test trajectory. First, the flow agreement parameter $[1/(t_f - t_0)\Omega^2] \int_{t_0}^{t_f} \dot{\theta}(t)\omega(\mathbf{x}(t), t)dt$ [87–89] measures the agreement between the agent's angular velocity $\dot{\theta}$ and the vorticity $\omega(\mathbf{x}, t)$ of the ambient flow field at the agent's location. Second, the thrust parameter $[1/(t_f - t_0)V^2] \int_{t_0}^{t_f} (\dot{y}(t) - v(\mathbf{x}(t), t))^2 dt$ [89–92] evaluates how well the agent can utilize the transverse flow velocity $v(\mathbf{x}, t)$ to generate thrust. Compared to the inhibitory strategy, the trajectories instructed by the excitatory strategy have a

more positive flow agreement and larger thrust parameters on average [Figs. 3(c) and 3(d)]. These trends are consistent between the RL policies and their simpler counterparts. Importantly, these findings show that by moving inside the wake, the excitatory strategy has more opportunities to harness the flow to swim more efficiently, consistently with empirical observations in fish [34].

Finally, to test the robustness of these policies under noisy and corrupt conditions, we challenged the agent in two major ways: (1) We decreased the rate at which it collects observations and responds to flow signals by increasing the decision time step Δt and (2) we imposed sensory limits $s_{\min}$ below which the agent is unable to sense the flow signal [Fig. 3(e)]. Both the RL policies and the simpler versions in Eq. (2) exhibited remarkable robustness to these reduced abilities, with stronger robustness exhibited by the RL policies.

E. Robustness to agent design

We next asked whether the performance of the controller is sensitive to the specific agent design, including the choice of turning rate Ω and sensor placement ℓ. During RL training, the agent's turning rate was fixed at $\Omega = 3U/L$ and its sensor was placed at $\ell = -0.25L$. Here, we set out to test the performance of each simplified policy in Eq. (2) over a broad range of sensor locations $\ell/L \in [-0.75, 0.75]$ and turning rates $\Omega L/U \in [0.5, 10]$, corresponding to dimensionless turning radii $V/\Omega L \in [0.025, 0.5]$. Using the same wake employed during training, we conducted 76 500 tests for each combination of parameters $(\ell/\lambda, V/\Omega L)$. The resulting success rates are shown in Figs. 2(f) and 2(g), with the RL policy parameters indicated by black dots. Both the excitatory (red) and inhibitory (blue) policies achieved success rates of over 80% throughout nearly the entire parameter space, provided that sensory measurements were taken at the tail ($\ell < 0$). Importantly, because the performance of the RL-inspired strategies is insensitive to parameter values, an agent does not need to fine-tune its sensor location ℓ (provided $\ell < 0$) and turning radius $V/\Omega L$ to a specific wake type; thus, we expect the agent's performance to be also robust to variations in the wake itself, a point that we revisit later.

F. Robustness to sensor type

How do these trail-following strategies perform when different types of flow information are available to the agent? For example, if instead of measuring the difference in flow speed, the agent measures the difference in flow quantities such as pressure, vorticity, or the odor concentration carried by the flow field, how would that affect its ability to track hydrodynamic trails?

To investigate this question, we repeated the RL training using three distinct types of sensory cues: pressure, vorticity, and odor. For pressure sensors, the RL training converged to only inhibitorylike policies, while the odor sensors converged to only excitatorylike policies. Because the pressure field is on average weaker inside the wake, an inhibitory policy that instructs the agent to move to lower pressure is needed to stay inside the wake. Meanwhile, odor concentration is on average stronger inside the wake; thus, an excitatory policy keeps the agent in the wake. The RL training did not converge using vorticity measurements (Table II and Fig. 8). In the converged cases, the RL policies resemble bang-bang controllers similar to those proposed in Eq. (2).

In Fig. 2(e), we show sample trajectories, starting from the same initial conditions, based on direct implementation of Eq. (2) using pressure, odor, and vorticity sensors (see also Movie S3 in the Supplemental Material [80]). Surprisingly, although vorticity measurements led to poor convergence in RL training, directly implementing this sensory cue in Eq. (2) allowed the agent to reorient and move upstream.

Importantly, when tested on the 76 500 sets of initial conditions in the wake of the source, we found that all four types of sensors—speed, pressure, vorticity, and odor sensors—exhibited high success rates in tracking the unsteady wake. This is surprising at first because when examining the distributions of gradient directions in the wake of the source [Fig. 1(e)], these distributions show no bias toward the upstream direction. However, when computing the direction of the spatial gradient of the signal collected along each test trajectory [Fig. 2(h)], we found that for all sensors, the gradient directions exhibited a bias toward the upstream direction. This indicates that the motion of the agent shapes the signal it acquires.

We next probed the robustness of these results to variations in the agent's design parameters $(\ell/\lambda, V/\Omega L)$; Fig. 9 shows the success rates across the parameter space for each sensory cue. Both the excitatory and inhibitory strategies achieved high success rates of over 80%, over large ranges of the parameter space, provided that the sensory measurement is trailing the agent's position ($\ell < 0$). Consistent with RL training, odor sensing functions only with the excitatory strategy. That is, it requires turning toward higher concentration. All other signals functioned well using either the excitatory or inhibitory policies, even the vorticity, for which the RL training did not converge. This is because the reward used in RL training requires the agent to accurately locate the source, whereas the success rate in Fig. 9 is based on the ability of the agent to move upstream, even if it ends at a lateral offset y/L from the source [trajectories in Fig. 2(e)]. That is, vorticity measurements are viable to enable the agent to reorient and move upstream, as suggested in Ref. [93], but compromise the agent's accuracy in locating the source. However, even when tested based only on reorienting and moving upstream, the vorticity sensor underperformed compared to the other sensors—speed, odor, and pressure—in terms of range and robustness to agent design parameters and maximum success rate (Fig. 9). This, together with the failed training of the vorticity-sensitive agent, aligns with recent reports of the inefficacy of sensing vorticity in point-to-point underwater navigation [50,51] and challenges approaches that advocate for sensing vorticity [93]. Sensing flow speed, pressure, or odor performs better in tracking flows. Because pressure sensors are most readily available commercially [31,84,94], the pressure-based strategy would be most amenable for immediate implementation in a robotic demonstration. Combining different types of sensors at once will be the topic of future studies and may lead to more robust trail-following behavior.

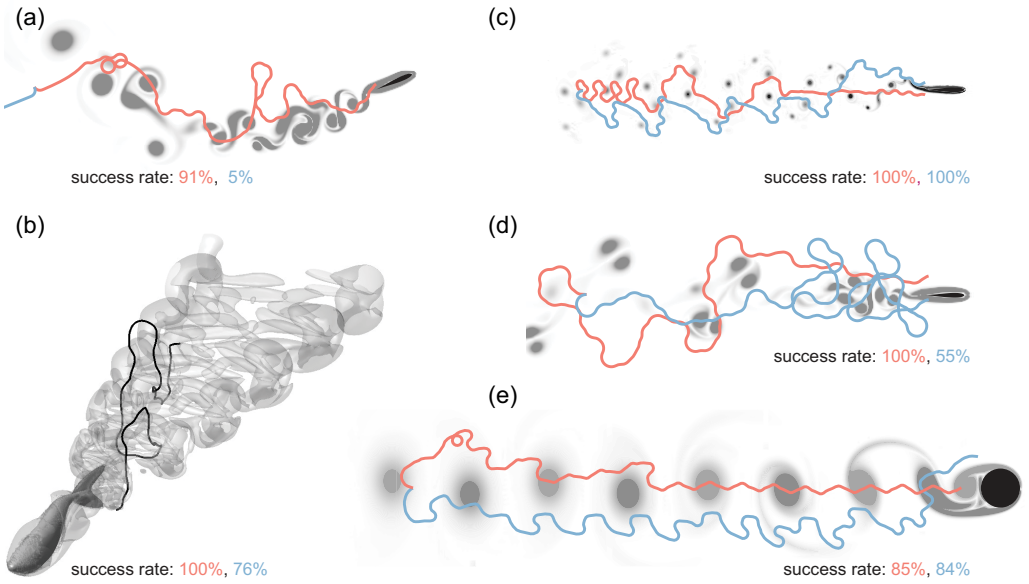

FIG. 4. Generalization to unseen flows. Sample trail-following trajectories obtained by applying Eq. (2) to unseen wakes; agent successfully follows the wake of (a) a pitching foil at Re = 1000 and St = 0.25 that changes direction of motion; (b) 3D swimming fish [95]; (c) pitching airfoil at Re = 5000 and St = 0.1; (d) burst-and-cost pitching airfoil at Re = 1000, St = 0.25, and duty cycle 0.6; and (e) fixed cylinder at Re = 400. Success rates out of 76 500 test trajectories are listed for both the excitatory and inhibitory policies and reported in Table III for all hydrodynamic trails. Parameter values are set to $\Omega = \pm 3$ and $\ell = -0.25\lambda$. Sample movies are given in Movie S2 in the Supplemental Material [80].

G. Generalization to diverse hydrodynamic trails

We next tested the performance of the simplified policies in Eq. (2) in wakes not seen during RL training. Specifically, we tested the policies in wakes generated by flow past pitching foils mimicking turning gait [Fig. 4(a)], intermittent swimming [Fig. 4(d)], and at different Strouhal and Reynolds numbers spanning St = 0.1–0.25 and Re = 500–5000 [Fig. 4(c); Table III]. We also tested the policy in wakes generated by flow past a fixed cylinder, with Re = 200−500 [Fig. 4(e); Table III] and in a three-dimensional (3D) wake, adapted from Ref. [95], generated by an undulating body at Re = 2400 [Fig. 4(b)]. For the 3D wake, we extended the simplified policies to control both yaw and pitch angles of the agent (Appendix C 3). In Table III, we report a summary of the performance of each policy in each of these wakes. Because each wake has a different wavelength, we used the flow wavelength λ to scale the corresponding sensor location $\ell = -0.25\lambda$ in Fig. 1. Without any additional parameter

TABLE III. Success rate of excitatory and inhibitory strategies in CFD simulations of flows past stationary and moving bodies. Wake type depends on how many vortices are shed in a period and their spatial relationship [96,97]. When two single vortices are shed in one period, such as Fig. 1(c), the wake is called a 2S wake. If two pairs of vortices and two single vortices are shed during one period, such as in Fig. 4(c), the wake is categorized as a 2P + 2S wake. A von Kármán vortex street that generates a jet pointing upstream is a drag wake [Fig. 4(e)], while a reverse von Kármán vortex street is a thrust wake [Fig. 1(c)]. The success rates of excitatory and inhibitory strategies of $\ell = -0.25\lambda$ and $\Omega = 3U/L$ are reported in the last two columns for different wake types.

								Success rate (%)	
Source	f	U	v	A	Re	St	Wake type	Excitatory	Inhibitory
Flapping foil	1.25	1	2×10^{-4}	0.2	5000	0.25	2S thrust	88	100
Flapping foil	1.25	1	2×10^{-3}	0.2	500	0.25	2S balanced	88	100
Flapping foil	0.5	1	2×10^{-4}	0.2	5000	0.1	2P + 2S	100	100
Flapping foil	0.5	1	2×10^{-3}	0.2	500	0.1	2S drag	100	100
Flapping foil	1.25	1	1×10^{-3}	0.2	1000	0.25	Intermittent	100	55
Flapping foil	1.25	1	1×10^{-3}	0.2	1000	0.25	Curved	91	5
Fixed cylinder	–	1	5×10^{-3}	–	200	0.203	2S drag	99	99
Fixed cylinder	–	1	3.33×10^{-3}	–	300	0.216	2S drag	80	80
Fixed cylinder	–	1	2.5×10^{-3}	–	400	0.222	2S drag	85	84
Fixed cylinder	–	1	2×10^{-3}	–	500	0.229	2S drag	100	100
3D Undulating fish	1	0.48	2×10^{-4}	0.32	2400	0.42	2S balanced	100	76

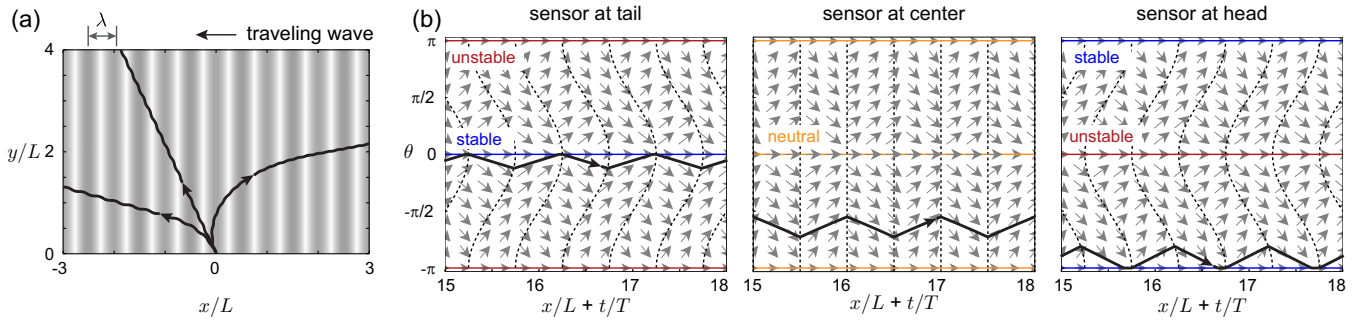

FIG. 5. Stability analysis. (a) Sample trajectories in 1D traveling signal field $A\cos(2\pi x/\lambda + 2\pi ft)$ by three agents with sensor placements at tail ($\ell = -0.25$), center ($\ell = 0$), and head ($\ell = 0.25$), all starting from the same initial condition $(x, \theta, t)|_0 = (0, 0.55\pi, 0)$. (b) Phase portraits on the phase space $(x + \lambda ft, \theta)$ for sensor at tail ($\ell = -0.25$), center ($\ell = 0$), and head ($\ell = 0.25$). Trajectories are superimposed after reaching the quasisteady state. Parameter values are set to swimming speed $V = 0.25$ and turning rate $\Omega = 1$.

tuning, the RL-inspired strategies achieved success rates of nearly 100% in guiding the agent to reorient upstream and track most of these hydrodynamic trails. In challenging scenarios, e.g., when direction of background flow is changing [Fig. 4(a)] or during intermittent flapping [Fig. 4(d)], in which the traveling-wave structure is not always present in the entire time series, the excitatory strategy performed notably better than the inhibitory strategy by navigating inside the wake, where the traveling-wave structure is more discernible. Thus, in addition to their parsimony and simplicity, the RL-inspired strategies are remarkably generalizable, allowing the agent to track a wide range of hydrodynamic trails, from thrust to drag jets to 3D wake behind undulatory, self-propelled swimmers (Fig. 4; Table III; Movie S2 in the Supplemental Material [80]).

H. Stability analysis and sensor placement

The most salient feature among all tested wakes is the quasiperiodic and traveling-wave nature of the signal field [Fig. 1(c)]. But how does this feature contribute to the success of the RL policies and their simpler Braitenberg-like analogs? More importantly, why is the sign of control gain not important, with inhibitory or excitatory strategies working equally well, while the sensor placement is critical? To address these questions, we analyzed the stability of the RL-inspired strategies in a one-dimensional (1D) traveling-wave signal field of the form $g(x, t) = A\cos(2\pi(x + \lambda ft)/\lambda)$, which is a general solution of the wave equation $g_{tt} = \lambda^2 f^2 g_{xx}$. This 1D wave solution is the simplest representation of the flow signal left by a fishlike swimmer [89,92].

The excitatory and inhibitory strategies are related by a change of sign from Ω to $-\Omega$, which, in the traveling-wave signal field, is equivalent to a translation in either space $x \to x + \lambda/2$ or time $t \to t + 1/(2f)$. It thus suffices to analyze, say, the excitatory strategy with the understanding that the asymptotic behavior of both strategies is similar.

Trail tracking in this 1D traveling-wave signal field amounts to moving upstream in the direction of $\theta = 0$, opposite to the direction of wave propagation. Numerical simulations show that upstream motion in this signal field depends on the sensor location ℓ [Fig. 5(a)]. To analyze the problem analytically, we introduced a coordinate transformation $z = x + \lambda ft$, where z represents a Lagrangian frame moving with the traveling wave, and substituted z and $g(z) = g(x + \lambda ft)$ into Eqs. (1) and (2) to eliminate explicit dependence on time,

$$\dot{z} = V\cos\theta + \lambda f,$$
$$\dot{\theta} = \Omega\,\mathrm{sgn}\left[\frac{2\pi A}{\lambda}\sin\theta\sin\left(\frac{2\pi}{\lambda}(z + \ell\cos\theta)\right)\right]. \quad (3)$$

The system in Eq. (3) has two moving equilibria at $\theta = 0$ (moving upstream) and $\theta = \pm\pi$ (moving downstream). The stability of these equilibria depends on the sensor location. Figure 5(b) depicts the phase portraits of Eq. (3) over the phase space (z, θ). Dashed black lines indicate level sets where $\dot{\theta}$ is 0, with phase velocity $\dot{\theta}$ alternating directions between adjacent level sets. The linear stability at these equilibria $\theta = 0, \pm\pi$ varies periodically with z. The shape of the level sets ensures that when $\ell < 0$, trajectories travel longer distance in the regions where the phase velocity points toward $\theta = 0$; the upstream equilibrium is thus stable. For $\ell > 0$, trajectories travel longer in the regions where the phase velocity points toward $\theta = \pm\pi$; the downstream equilibrium is stable. At the bifurcation point $\ell = 0$, both equilibria are neutrally stable.

I. Linking sensor placement to sensing-to-response time delays

Our finding that trail following is stable when sensors are placed at the tail is intriguing given that biological evidence indicates that flow sensors are mostly located at the head, such as in the whiskers of harbor seals [24] and in the lateral line system of fish where flow-sensitive neuromasts are denser at the head [76,77]. Indeed, physics-based [76,77] and data-driven [98,99] models suggest that the flow signal measured at the head is more informative in decoding flow information compared to the flow signal measured at the tail. However, recent behavioral experiments found that disabling the posterior portion of the lateral line is more harmful than disabling the anterior lateral line for certain behaviors that are essential for navigation such as fish rheotaxis [93,100]. Our study offers a potential hydrodynamic explanation for why a sensor at the tail is crucial for these behaviors.

Another perspective is that a trailing sensor is mathematically equivalent to a sensor at the head with a time delay

TABLE IV. Survey of relevant biological and robotic data from existing literature.

Biological systems						
	Common name	Body length (m)	Swimming speed (m/s)	Turning radius (m)	Reynolds number	Reference
Fish	Pacific trout	0.33	0.165	0.13	5.4×10^5	Weihs [106]
	Dolphinfish	2 ± 0.3	1.5 ± 0.5	0.33 ± 0.08	3.0×10^6	Domenici and Blake [105]
	Knitefish	0.113	0.95 ± 0.17	0.0062	1.1×10^6	Domenici and Blake [105]
	Anglefish	0.073	0.79 ± 0.068	$(4.7 \pm 0.15) \times 10^{-3}$	5.7×10^5	Domenici and Blake [105]
	Pike	0.51 ± 0.044	0.221 ± 0.035	0.046 ± 0.004	1.1×10^6	Domenici and Blake [105]
	Smallmouth bass	0.31 ± 0.07	0.8 ± 0.2	0.034 ± 0.014	2.5×10^6	Domenici and Blake [105]
	Rainbow trout	0.12 ± 0.0035	0.35 ± 0.01	0.021 ± 0.003	4.1×10^5	Domenici and Blake [105]
	Yellowtail	0.19	0.77 ± 0.094	0.043	1.5×10^6	Domenici and Blake [105]
	Yellowfin tuna	0.85 ± 0.048	0.68 ± 0.11	0.40 ± 0.17	5.7×10^5	Domenici and Blake [105]
Pinniped	Sea lion	1.8 ± 0.085	2.89 ± 0.77	0.41 ± 0.14	5.2×10^7	Fish *et al.* [118]
Underwater robotic systems						
Underwater robots	Multilink fish	0.45	0.25	0.09	1.1×10^6	Li *et al.* [36]
	Multilink fish	0.34	0.063	0.1	2.1×10^5	Hirata *et al.* [125]
	Multilink fish	0.65	0.19	0.22	1.2×10^6	Yu *et al.* [124]
	Driven by caudal fin	1.23	1.1	1.8	1.3×10^7	Liang *et al.* [126]
	Driven by propeller	1.6	1.1	4.2	1.8×10^7	Liang *et al.* [126]

Response time in biological systems		
	Response time (ms)	Reference
Fish	5–150	Domenic and Hale [103]; Roberts *et al.* [104]
Seal	188–982	Kastelein *et al.* [102]

τ between sensing and response. Indeed, in aquatic animals, including fish, harbor seals, and copepods, there is a time delay in the neural system between mechanosensory input and behavioral response [101–104]. To explore this hypothesis, we collected biological data of animals' locomotion capabilities (Table IV; [105–119]), mapped their time delay τ to sensor placement $\ell \approx V\tau$, and evaluated our predictions of the success rate of the RL-inspired policies over the parameter space of turning radius $V/\Omega L$ and sensor location ℓ/L expressed in terms of time delay τ between sense and response, $\ell/L \approx V\tau/L$. Remarkably, organisms that inhabit and track hydrodynamic trails at comparable Reynolds numbers [20,21,24,42] lie within a parameter range compatible with our model predictions for stable and robust trail following (Fig. 6).

The last perspective is that having a sensor at the head makes moving downstream in the wake a stable behavior, which links to animals escaping from predators [121].

J. Sensing longitudinal gradients

Our trail-tracking policies relied on sensing the transverse gradient of the signal field $\mathbf{n} \cdot \nabla$. Do these policies work when sensing the longitudinal gradient of the signal field $\mathbf{t} \cdot \nabla$? In Fig. 7(a), we show in simulations that measuring the longitudinal gradient of the flow speed allows effective trail tracking in vortical flows. Through a rigorous stability analysis similar to that presented above, we found that sensing the longitudinal gradient of the flow signal results in stable upstream motion when the sensor is placed at the tail (Fig. 11). When evaluating other sensory cues such as odor, pressure, and vorticity, we found that their performance was suboptimal compared to flow speed (Fig. 10). We next placed the longitudinal sensor at the center $\ell = 0$ and allowed a time delay τ between sensing and response [Fig. 7(b)]. Again, we tested this sensory strategy across all 76 500 initial conditions; the overall success rate of the corresponding policy was nearly as good as that with the sensor placed at the tail.

K. Bridging trail-tracking strategies from vortical to turbulent trails

To extend these policies to turbulent plumes where measuring spatial gradients of the signal field is not viable [18,19], we considered an agent that measures the odor signal itself at time t: $s(t) = g(\mathbf{x}(t), t)$ and compares it to that measured at a previous time $t - \tau$: $s(t - \tau) = g(\mathbf{x}(t - \tau), t - \tau)$. Following one strategy, the agent rotates with angular velocity Ω when $s(t) > s(t - \tau)$ and $-\Omega$ when $s(t) < s(t - \tau)$; the other strategy does the opposite (C2). In Fig. 7(c), we show that both strategies achieve similar performance with over 80% success in tracking the turbulent plume. The success of these simple strategies in a turbulent plume stems from the fact that, despite the absence of a traveling-wave structure with well-defined frequency and/or wavelength, odor patches are consistently advected downstream by the mean flow; see Fig. 1(b) and Ref. [12].

III. CONCLUSIONS

Using deep reinforcement learning, we discovered two policies for tracking biologically relevant hydrodynamic trails

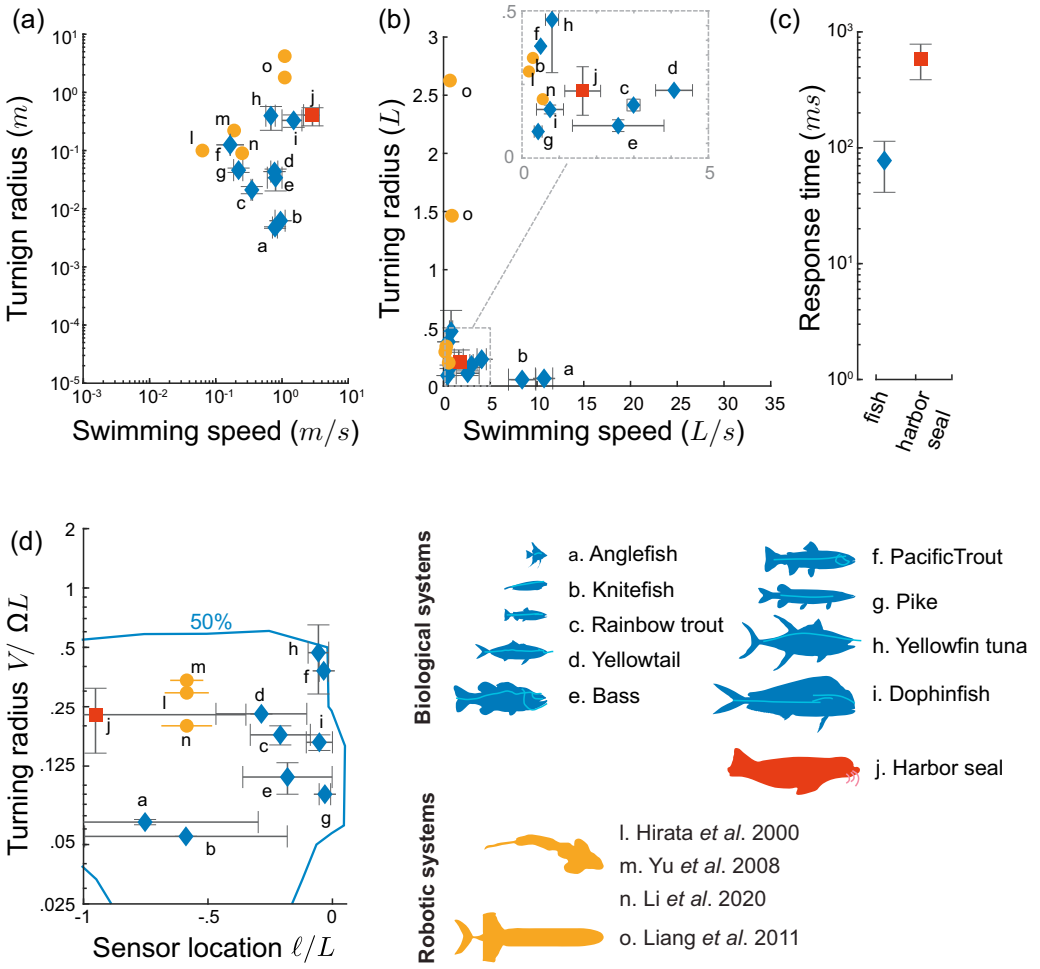

FIG. 6. Linking sensor location to sensing-to-response time delays in biological and robotic systems. We gathered biological data from fish, harbor seals, and their sea lion relatives, as well as from underwater robotic systems (Table S1 in the Supplemental Material [80]). Turning radius versus swimming speed (a) in log-log scale and dimensional units, and (b) in units of body length L. (c) Motor response time to mechanosensory input. Error bars represent the standard deviation of biological data. (d) Data mapped onto the parameter space $(\ell/L, V/\Omega L)$ in semilog scale with sensor location $\ell = -V\tau$ based on the distance traveled during the neuronal response time τ. For robotic fish, we obtained the minimum turning rate from the literature (Table S1 in the Supplemental Material [80]) and calculated, based on our model, the range of sensor locations ℓ/L for achieving 95% success rates.

at intermediate Reynolds numbers. The policies rely only on local and instantaneous flow sensing, and can be expressed in terms of two parsimonious, interpretable, and generalizable strategies, where the swimmer measures locally a differential flow signal and responds by turning either toward or away from the direction of stronger signal (Fig. 2). These remarkably simple strategies depend only on two dimensionless parameters—the minimum turning radius $R = V/\Omega L$ of the agent and the sensor location ℓ/L. Through rigorous stability analysis, we proved that these strategies are stable in signal fields with traveling-wave characteristics, provided posterior sensor placement $\ell/L < 0$ (Fig. 5), and we offered both biological and hydrodynamic explanations for these findings. Using Monte Carlo simulations, we demonstrated that these strategies carry over to unfamiliar wakes (Fig. 4; Table III) and to sensors that probe different types of flow signals [Figs. 2(d), 2(e), and 9]. Moreover, inspired by these strategies, we handcrafted simple strategies that respond to the temporal difference of the signal field to track turbulent plumes, where spatial gradients of the signal are well defined [Fig. 7(b)]. Our findings underscore an important but often overlooked fact: Sensing and control strategies are intimately intertwined with the characteristic of the environment in which they operate. Here, the directionality of the signal's traveling motion plays an active role in enabling simple, yet effective, sensing and control strategies for trail following.

Our strategies outperform existing sensory and response mechanisms for trail tracking (Table I) in three assessment metrics—interpretability, versatility, and adaptability to different flow currents. Importantly, our strategies are remarkable for their minimal sensory requirements. Sensing is expensive! Sensory limitations are thought to be the greatest challenge preventing robotic systems from achieving performance on par with their animal counterparts [122]. Our findings show that trail tracking of flow currents can be

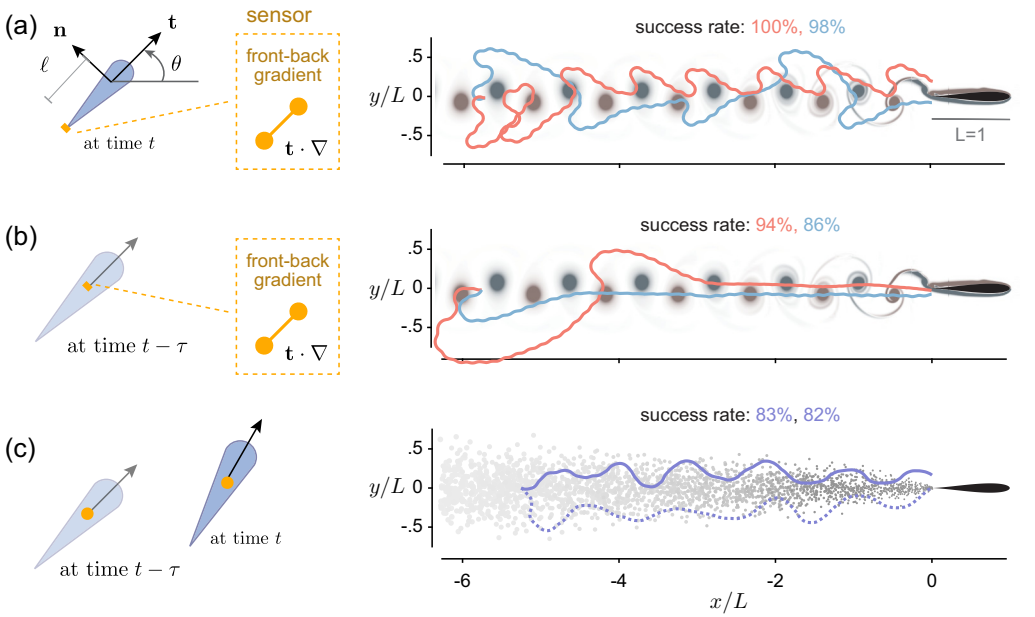

FIG. 7. Simple and versatile strategies for tracking vortical and turbulent wakes. Sample trajectories and success rates of simple yet adaptable trail-following strategies: (a) An agent measuring longitudinal gradient of flow speed at tail, with $\ell = -0.25$ and $\Omega = 3$, is equally successful in tracking vortical wakes as the agent measuring lateral gradient in Fig. 2. (b) An agent, with sensor collocated at the agent's position, measuring longitudinal gradient at a time delay $\tau = 0.1$, is also equally successful in tracking vortical wakes. (c) An agent measuring temporal difference in the signal itself, not the signal gradient, is successful in tracking a turbulent wake. Here, $\tau = 1$ and $\Omega = 2$. The trajectories show the classic "cast and surge" behavior [57,58]. As a baseline for comparison, the success rate in Ref. [19] of RNN agent tracking signal field of constant direction is about 75%. Movie of the trajectory is shown in Movie S2 in the Supplemental Material [80].

robustly achieved with minimal requirements on sensory abilities and memory.

Our work offers a powerful framework for designing robust and adaptable underwater navigation systems. Case in point, we employed a three-link fish model [121,123] and developed, based on the RL-inspired excitatory and inhibitory strategies, controllers that directly mapped local flow signals to body deformations, much like Braitenberg's vehicles that linked signal intensity to wheel rotation [82]. Numerical tests in traveling signal fields confirmed the success of the three-link fish in tracking a traveling-wave signal (Movie S1 in the Supplemental Material [80]). Further analysis and experimental validation of these predictions will be the topic of future research [29,31,32].

Our findings also guide the strategic placement of sensors in underwater robotic systems. Surveying existing systems [36,124–126] (Table IV), we found that robots capable of large ranges of body deformations exhibited turning radii comparable to those of swimming organisms [36,124,125]. Then, using our results [Figs. 2(f) and 2(g)] that quantify the success in trail following over the space of sensor location and turning radius, we predicted ranges of sensor locations [Fig. 6(d)] for which these robotic systems would exhibit success rates exceeding 95%. These predictions provide valuable tools for designing and optimizing the sensory layout of future underwater robotic systems and testing them in real-world scenarios, thus advancing the development of underwater technologies [29,31,32].

ACKNOWLEDGEMENTS

Funding support provided by the NSF Grants No. RAISE IOS-2034043 and No. CBET-2100209, ONR Grants No. N00014-22-1-2655 and No. N00014-19-1-2035, and NIH Grant No. R01-HL153622 (all to E.K.).

DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX A: SIMULATIONS OF FLOW ENVIRONMENT

1. Laminar plume

An odor-emitting source with continuous emission rate R is located at point $\mathbf{x}^*$. Time evolution of the odor concentration field is composed of advection of background flow $\mathbf{u}(\mathbf{x}, t)$ and diffusion at constant diffusive constant D:

$$\frac{\partial c}{\partial t} + \mathbf{u} \cdot \nabla c = D\Delta c + R\delta(\mathbf{x} - \mathbf{x}^*). \tag{A1}$$

In laminar plume, the flow is a uniform background flow $\mathbf{u}(\mathbf{x}, t) = U = 1$. Thus, Péclet number is defined as the ratio between the advection rate and diffusion rate $Pe = LU/D$. The solution of Eq. (A1) is given by

$$c(x, y) = \frac{R}{2\pi D} \exp\left(-\frac{U(x - x^*)}{2D}\right) K_0\left(\frac{U||\mathbf{x} - \mathbf{x}^*||}{2D}\right), \tag{A2}$$

where K_0 is the modified Bessel function of the second kind of order zero [18]. The parameters are chosen as odor emission rate $R = 0.25$, uniform background flow velocity $U = 1$, and diffusion rate $D = 0.1$. These result in a Péclet number Pe $= 10$.

2. Turbulent plume

We employed a particle-based two-dimensional plume model [19,68]. The plume model simulates the random release of odor packets from a source with a fixed initial radius r_0 and initial strength c_0. A turbulent wind transported the odor packets and spread out in all directions due to radial diffusion. The concentration decreases with the increases in packet size:

$$c_t = c_0 \left(\frac{r_0}{r_t} \right)^3. \tag{A3}$$

The turbulent wind is simplified to the combination of a uniform background flow of velocity $U = 1$ and random perturbations with strength σ. These parameters are chosen as emission rate $R = 1$, initial radius and concentration of the plume $r_0 = 0.01$ and $c_0 = 1$, random perturbation crosswind $\sigma = 0.005$, and diffusion rate 0.001.

3. Vortical trail

Hydrodynamic trails were generated from CFD simulations and used in this study (Table III). These simulations included 2D wakes generated by flows past an oscillating airfoil, 2D wakes generated by flows past a fixed cylinder, and a 3D wake generated by flow past an undulating fishlike body.

In the CFD simulations, the spatial-temporal evolution of the flow field [velocity $\mathbf{u}(\mathbf{x}, t)$, pressure $p(\mathbf{x}, t)$, and vorticity $\omega(\mathbf{x}, t) = \nabla \times \mathbf{u}$] is governed by the incompressible Navier-Stokes equations, given in dimensionless form as

$$\nabla \cdot \mathbf{u} = 0, \quad \frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\nabla p + \frac{1}{\mathrm{Re}} \Delta \mathbf{u}. \tag{A4}$$

The boundary conditions are uniform velocity inlet and no-slip boundary conditions on the surface of the structure. The flow fields are solved using immersed boundary method to deal with the moving boundary [70,71,127–129].

The 2D simulations were conducted using an open-source fluid solver, IBAMR [69–71]. The airfoil is a symmetric Joukowsky airfoil of chord length L and maximum thickness $0.12L$, undergoing pitch oscillations $\alpha(t) = \alpha_0 \sin(2\pi f t)$ about a pivot point located at $0.25L$ from its leading edge, where α_0 is the amplitude and f is the pitch frequency. The peak-to-peak tailbeat amplitude is $A = 1.5 f \sin \alpha_0$. The parameter values are summarized in Table III.

We chose a rectangular computational domain of size $[-24, 8] \times [-8, 8]$ and placed the pivot point of the airfoil at the origin $(0, 0)$ of the domain. The coarsest Eulerian mesh is a uniform 128×64 Cartesian grid. There are three layers of refinement mesh, with refinement ratio $= 4$ for each layer. The refinement criterion is based on the vorticity of the wake. The simulation time step is $\Delta t = 1 \times 10^{-4}$. The types of simulated wakes are listed in Table III. The 3D wake generated by flow past the undulating body is based on ViCar3D [95,128,130]. Details of the computational setup are reported in Ref. [95].

To consider the concentration field advected by the flow field, we placed an odor-emitting source at the trailing edge of the airfoil and solved the advection-diffusion equation for the concentration field advected by the flow field (A1). The Péclet number in Fig. 1(c) is 10 000.

APPENDIX B: REINFORCEMENT LEARNING

1. Training the agent using deep reinforcement learning

We implemented the clipped-advantage proximal policy optimization (PPO) method proposed by Ref. [78] for our RL training as in Ref. [123] (Fig. 8). PPO maximizes a surrogate objective that clips off unwanted changes when the policy deviates too much from the policy of the previous experience to ensure faster and more robust convergence. Two neural networks, actor and critic (both 64×32), which approximate the mean of the policy and value function of the state, respectively, are being updated during the training process. We refer readers to the original reference cited above as well as the OpenAI's documentation of the PPO algorithm [131], their baseline implementations [132], and previous paper from our group [123] for a thorough explanation of the theory and details behind this method. In this work, both actor and critic are represented and fitted by 64×32 Deep Neural Networks.

We trained the agent to track the hydrodynamic trail from repeated experiences interacting with the flow environment. The RL agent took as observations $o = (s_{\mathrm{head}}, s_{\mathrm{tail}})$ two sensory measurements taken at head and tail, and provided as action a noisy turning rate $a = \dot{\theta} \sim \pi(s_{\mathrm{head}}, s_{\mathrm{tail}})$.

To limit the action space, we imposed a threshold value Ω on the maximal angular velocity $|\dot{\theta}| \leqslant \Omega$. This, in turn, constrained the minimum turning radius $V/|\dot{\theta}| \geqslant V/\Omega$ of the agent. The decision time step Δt is chosen as $0.05T$, where T is the period of pitching. When it reached the source in the streamwise direction, the agent received a sparse reward of $50e^{-y^2}$, decaying with the lateral distance y from the source. Additionally, to encourage initial learning, the agent received a dense shaping reward at each decision time step Δt, $-d(t) + d(t - \Delta t)$, smaller in magnitude, that depended on its relative change in distance $d(t)$ to the trailing edge of the source generating the flow, yielding a reward $r_t = -d(t) + d(t - \Delta t) + 50e^{-y^2}$. The reinforcement learning algorithm is trying to maximize total discounted return, $R_t = \sum_{t' > = t} \gamma^{t'-t} r_t$. Thus, the algorithm tries to find a time-optimal policy by itself.

2. Evaluating policy success rate

To test the performance of the strategies, and to standardize these tests across different wakes and sensory cues, we designed a test domain consisting of a rectangular grid starting at a distance L downstream from the trailing edge of the airfoil and spanning a $9L \times 3.2L$ range in the x and y directions, with grid size $0.2L$. This grid provides $45 \times 17 = 765$ distinct initial positions $[(x(0), y(0)]$ of the agent. At each grid point, we explored 20 equally spaced initial orientations $\theta(0)$ between $[0, 2\pi)$ and 5 equally spaced initial phases in a pitching period

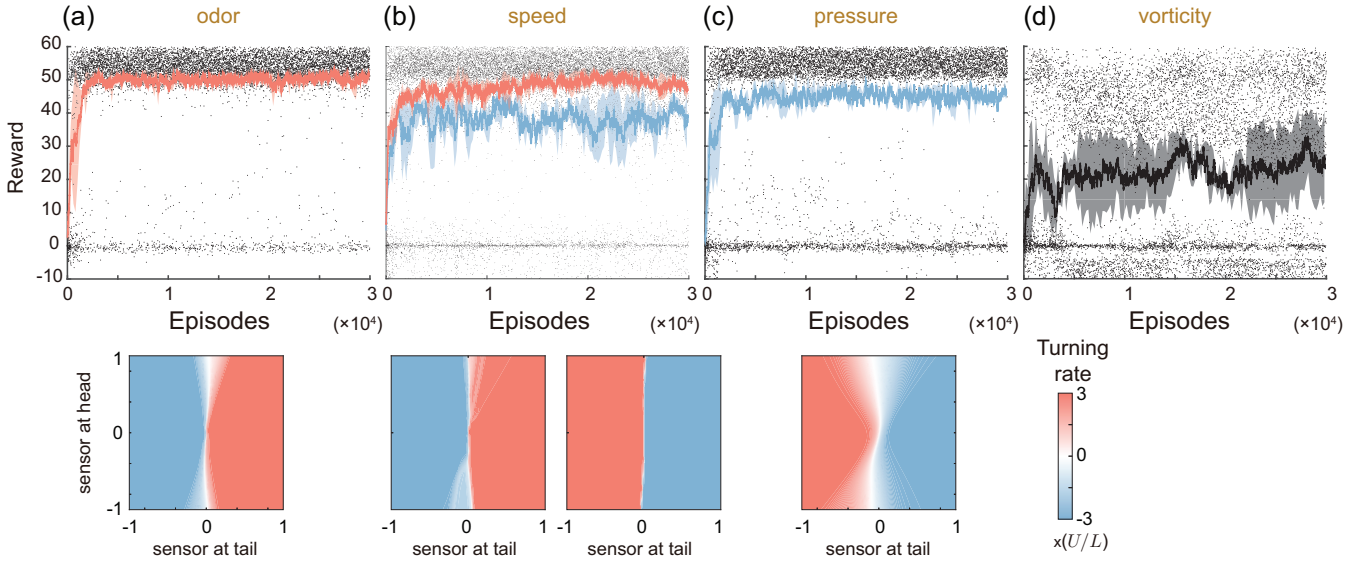

FIG. 8. RL-training for different sensory cues. (a) Lateral gradient of odor concentration. (b) Lateral gradient of flow speed. (c). Lateral gradient of pressure. (d). Lateral gradient of vorticity magnitude. The convergence rates over multiple training instances are reported in Table II. In panels (a) and (c), RL training converges to excitatory and inhibitory strategies, respectively. In panel (b), different training instances converge to either excitatory or inhibitory strategies. In panel (d), the training does not converge. The insets show the converged RL policy by plotting the action $\dot{\theta}$ as a colormap over the observation space $(s_{\text{tail}}, s_{\text{head}})$. These colormaps exhibit similar features, though not as pronounced, to those obtained in Fig 2(c): To first-order approximation, the action $\dot{\theta}$ is mostly independent of the signal s_{head}. We thus approximate these strategies following Eq. (2). Parameter values: $\ell_{\text{tail}} = -0.25L$, $\ell_{\text{head}} = 0.25L$, and $\Omega = 3U/L$. The flow field used during training is the same as that in Fig. 2.

$[0, 1/f)$. That is, for each wake and each policy, we run 76 500 tests.

In the three-dimensional flow field [Fig. 4(b)], the initial condition is composed of six independent variables $[x(0), y(0), z(0), \theta(0), \psi(0), t(0)]$. Doing a grid search is infeasible under this scenario, so we generated 90 000 random initial conditions to evaluate the performance.

APPENDIX C: HANDCRAFTED STRATEGIES FOR HYDRODYNAMIC TRAIL FOLLOWING

1. Simplified strategies based on local sensing of signal gradient

We devised Braitenberg-like strategies where the agent simply responds to the local gradient of the signal field $g(\mathbf{x}, t)$.

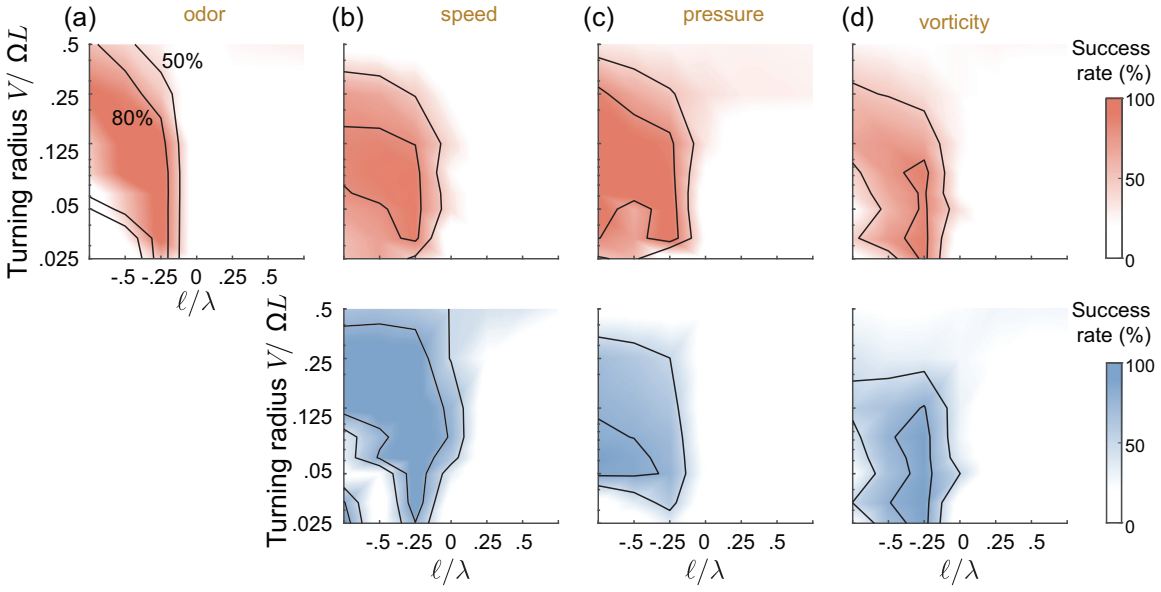

FIG. 9. Parametric study for different sensory cues. Success rate of the agent as a function of sensor location ℓ/λ and turning radius $R = V/\Omega L$ for different sensory cues with excitatory (red) and inhibitory (blue) RL-inspired strategies. Four different sensory cues are tested: (a) odor, (b) flow speed, (c) pressure, and (d) vorticity.

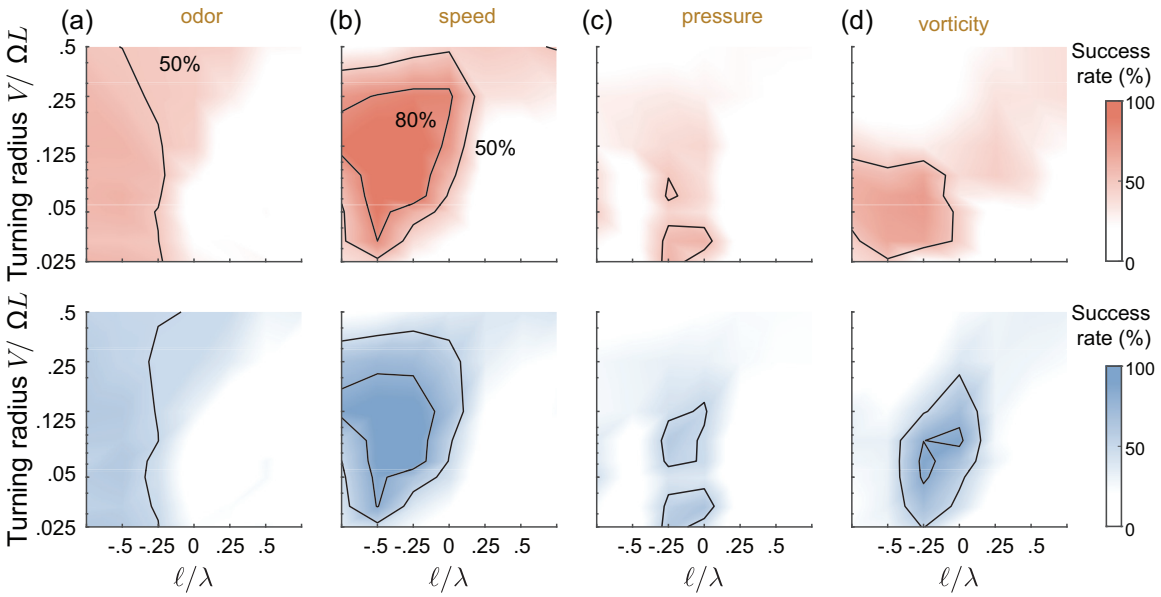

FIG. 10. Parametric study for different sensory cues using gradient in longitudinal direction. Success rate of the agent as a function of sensor location ℓ/λ and turning radius $R = V/\Omega L$ for different sensory cues with excitatory (red) and inhibitory (blue) RL-inspired strategies. Four different sensory cues are tested: (a) odor, (b) flow speed, (c) pressure, and (d) vorticity.

The gradient can be taken in either the normal or tangential direction, namely,

$$\text{transverse gradient: } \dot{\theta}(t) = \pm\Omega\,\mathrm{sgn}(\mathbf{n}\cdot\nabla g) \quad \text{or}$$
$$\text{longitudinal gradient: } \dot{\theta}(t) = \pm\Omega\,\mathrm{sgn}(\mathbf{t}\cdot\nabla g). \tag{C1}$$

An excitatory strategy responding to the transverse gradient of concentration is similar to a gradient ascent; it works perfectly in laminar plumes (Fig. 9). In turbulent plumes, measuring spatial gradients is not feasible given the sporadic and random distribution of the signal. In vortical wake, multiple local minima and maxima exist in the concentration field, and thus measuring normal and longitudinal gradients both work as long as the sensor is not collocated with the agent's position, but placed at the tail, behind the agent.

2. Strategies based on temporal difference of detected signal

When measuring spatial gradient is not applicable, the sensory control strategies are generalized by substituting spatial gradient with finite temporal difference. Namely, at each time step t, it compares the current signal $g(\mathbf{x}(t), t)$ with a previous the signal at previous time $g(\mathbf{x}(t-\tau), t-\tau)$ and turns based on whether it is bigger or smaller than the current signal:

$$\dot{\theta}(t) = \Omega\,\mathrm{sgn}\{g[\mathbf{x}(t), t] - g[\mathbf{x}(t-\tau), t-\tau]\}. \tag{C2}$$

The temporal difference of signal is an approximation of the longitudinal gradient of the signal field plus a time delay (Fig. 10). Although the strategy does not work in the laminar plume and vortical wake, it achieves a high success rate in the turbulent plume as illustrated in Fig. 7(c). In turbulent plume, $\Omega > 0$ and $\Omega < 0$ generate similar success rates (Fig. 7), but do not intuitively map to excitatory or inhibitory strategy

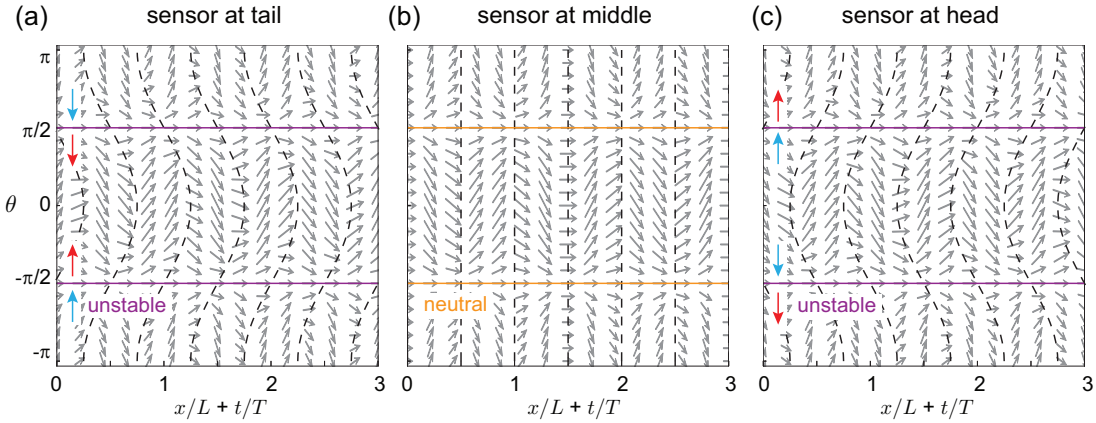

FIG. 11. Phase portraits for responding to longitudinal flow gradient. Phase portraits on the phase space $(x + \lambda ft, \theta)$ for (a) sensor at tail ($\ell = -0.25$), (b) center ($\ell = 0$), and (c) head ($\ell = 0.25$). Swimming speed $V = 0.25$ and turning rate $\Omega = 1$.

since they are responding to temporal difference instead of lateral gradient.

3. Generalization to 3D wakes

In 3D, the agent's position vector $\mathbf{x}$ is expressed in terms of the coordinates $\mathbf{x} = (x, y, z)$ in the fixed inertial frame $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$. We affixed a body frame $(\mathbf{t}, \mathbf{n}, \mathbf{b})$ to the agent, where $\mathbf{t}$ is a tangential unit vector in the agent's swimming direction, and $\mathbf{n}$ and $\mathbf{b}$ are the normal and binormal unit vectors, respectively. We used Euler angles $(\theta,\ \psi,\ \phi)$ to describe the orientation of the agent. We excluded spinning around $\mathbf{t}$ by setting $\phi = 0$ and controlled the agent's heading angle θ and ψ based on the gradient of the signal field in the $\mathbf{b}$ and $\mathbf{n}$ directions, respectively. Thus, in laboratory frame $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$, the body frame is expressed as $\mathbf{t} = (\cos\psi\cos\theta, \sin\psi\cos\theta, -\sin\theta)$, $\mathbf{n} = (-\sin\psi, \cos\psi, 0)$, $\mathbf{b} = (\cos\psi\sin\theta, \sin\psi\sin\theta, \cos\theta)$ (Fig. 11). In this case, the agent's equations of motion are written as

$$\dot{\mathbf{x}} = (\dot{x}, \dot{y}, \dot{z}) = V\mathbf{t}, \quad \dot{\theta} = -\Omega\,\mathrm{sgn}(\mathbf{b}\cdot\nabla g)|_{\mathbf{x}+\ell\mathbf{t}},$$
$$\dot{\psi} = \Omega\,\mathrm{sgn}(\mathbf{n}\cdot\nabla g)|_{\mathbf{x}+\ell\mathbf{t}}, \tag{C3}$$

where $g(\mathbf{x}, t)$ is the signal field. In the case shown in Fig. 4(b), the sensory cue is chosen to be the flow speed $|\mathbf{u}|$.

[1] E. M. Purcell, Life at low Reynolds number, in *Physics and Our World: Reissue of the Proceedings of a Symposium in Honor of Victor F. Weisskopf* (American Institute of Physics, Melville, New York, 1977), pp. 47–67.

[2] D. Roszak and R. Colwell, Survival strategies of bacteria in the natural environment, Microbiol. Rev. **51**, 365 (1987).

[3] G. M. Viswanathan, M. G. Da Luz, E. P. Raposo, and H. E. Stanley, *The Physics of Foraging: An Introduction to Random Searches and Biological Encounters* (Cambridge University Press, Cambridge, 2011).

[4] National Academies of Sciences, Engineering, and Medicine, *Physics of Life* (The National Academies Press, Washington, DC, 2022), https://nap.nationalacademies.org/catalog/26403/physics-of-life.

[5] H. C. Berg, *Random Walks in Biology* (Princeton University Press, Princeton, New Jersey, 1993).

[6] H. C. Berg and E. M. Purcell, Physics of chemoreception, Biophys. J. **20**, 193 (1977).

[7] J. Adler, Chemoreceptors in bacteria: Studies of chemotaxis reveal systems that detect attractants independently of their metabolism, Science **166**, 1588 (1969).

[8] E. F. Keller and L. A. Segel, Model for chemotaxis, J. Theor. Biol. **30**, 225 (1971).

[9] H. C. Berg and D. A. Brown, Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking, Nature (London) **239**, 500 (1972).

[10] A. Gomez-Marin, G. J. Stephens, and M. Louis, Active sampling and decision making in *Drosophila* chemotaxis, Nat. Commun. **2**, 441 (2011).

[11] K. J. Leitch, F. V. Ponce, W. B. Dickson, F. van Breugel, and M. H. Dickinson, The long distance flight behavior of *Drosophila* supports an agent-based model for wind-assisted dispersal in insects, Proc. Natl. Acad. Sci. USA **118**, e2013342118 (2021).

[12] N. Kadakia *et al.*, Odour motion sensing enhances navigation of complex plumes, Nature (London) **611**, 754 (2022).

[13] A. Mafra-Neto and R. T. Cardé, Fine-scale structure of pheromone plumes modulates upwind orientation of flying moths, Nature (London) **369**, 142 (1994).

[14] R. T. Cardé and A. Mafra-Neto, Neuroethology of pheromone-mediated responses, *Insect Pheromone Research*, edited by R. T. Cardé and A. K. Minks (Springer, New York, 1997), pp. 275–290.

[15] M. J. Weissburg, M. Doall, and J. Yen, Following the invisible trail: Kinematic analysis of mate-tracking in the copepod *Temora longicornis*, Philos. Trans. R. Soc. London B **353**, 701 (1998).

[16] J. Yen, J. K. Sehn, K. Catton, A. Kramer, and O. Sarnelle, Pheromone trail following in three dimensions by the freshwater copepod *Hesperodiaptomus shoshone*, J. Plankton Res. **33**, 907 (2011).

[17] R. Monthiller, A. Loisy, M. A. Koehl, B. Favier, and C. Eloy, Surfing on turbulence: A strategy for planktonic navigation, Phys. Rev. Lett. **129**, 064502 (2022).

[18] M. Vergassola, E. Villermaux, and B. I. Shraiman, 'Infotaxis' as a strategy for searching without gradients, Nature (London) **445**, 406 (2007).

[19] S. H. Singh, F. van Breugel, R. P. Rao, and B. W. Brunton, Emergent behaviour and neural dynamics in artificial agents tracking odour plumes, Nat. Mach. Intell. **5**, 58 (2023).

[20] K. Pohlmann, F. W. Grasso, and T. Breithaupt, Tracking wakes: The nocturnal predatory strategy of piscivorous catfish, Proc. Natl. Acad. Sci. USA **98**, 7371 (2001).

[21] K. Pohlmann, J. Atema, and T. Breithaupt, The importance of the lateral line in nocturnal predation of piscivorous catfish, J. Exp. Biol. **207**, 2971 (2004).

[22] J. M. Gardiner and J. Atema, Sharks need the lateral line to locate odor sources: Rheotaxis and eddy chemotaxis, J. Exp. Biol. **210**, 1925 (2007).

[23] G. Dehnhardt, B. Mauck, and H. Bleckmann, Seal whiskers detect water movements, Nature (London) **394**, 235 (1998).

[24] G. Dehnhardt, B. Mauck, W. Hanke, and H. Bleckmann, Hydrodynamic trail-following in harbor seals (*Phoca vitulina*), Science **293**, 102 (2001).

[25] G. R. Spedding, Wake signature detection, Annu. Rev. Fluid Mech. **46**, 273 (2014).

[26] B. Colvert, K. Chen, and E. Kanso, Local flow characterization using bioinspired sensory information, J. Fluid Mech. **818**, 366 (2017).

[27] B. Colvert, G. Liu, H. Dong, and E. Kanso, Flowtaxis in the wakes of oscillating airfoils, Theor. Comput. Fluid Dyn. **34**, 545 (2020).

[28] S. D. Boie *et al.*, Information-theoretic analysis of realistic odor plumes: What cues are useful for determining location? PLoS Comput. Biol. **14**, e1006275 (2018).

[29] J. Ježov, O. Akanyeti, L. D. Chambers, and M. Kruusmaa, Sensing oscillations in unsteady flow for better robotic swimming efficiency, in *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (Seoul, Korea, 2012), pp. 91–96.

[30] N. Muhammad *et al.*, Map-based localization and loop-closure detection from a moving underwater platform using flow features, Auton. Robots **43**, 1419 (2019).

[31] P. Gunnarson and J. O. Dabiri, Fish-inspired tracking of underwater turbulent plumes, Bioinspir. Biomim **19**, 056024 (2024).

[32] P. Gunnarson and J. O. Dabiri, Surfing vortex rings for energy-efficient propulsion, Proc. Natl. Acad. Sci. USA **4**, pgaf031 (2025).

[33] H. Bleckmann, J. Mogdans, and S. L. Coombs, *Flow Sensing in Air and Water* (Springer-Verlag, Berlin, Germany, 2014), Vol. 976.

[34] J. C. Liao, D. N. Beal, G. V. Lauder, and M. S. Triantafyllou, Fish exploiting vortices decrease muscle activity, Science **302**, 1566 (2003).

[35] D. N. Beal, F. S. Hover, M. S. Triantafyllou, J. C. Liao, and G. V. Lauder, Passive propulsion in vortex wakes, J. Fluid Mech. **549**, 385 (2006).

[36] L. Li *et al.*, Vortex phase matching as a strategy for schooling in robots and in fish, Nat. Commun. **11**, 5408 (2020).

[37] S. Dijkgraaf, The functioning and significance of the lateral-line organs, Biol. Rev. **38**, 51 (1963).

[38] J. Blaxter, Structure and development of the lateral line, Biol. Rev. **62**, 471 (1987).

[39] A. Kroese and N. Schellart, Velocity-and acceleration-sensitive units in the trunk lateral line of the trout, J. Neurophysiol. **68**, 2212 (1992).

[40] J. Engelmann, W. Hanke, J. Mogdans, and H. Bleckmann, Hydrodynamic stimuli and the fish lateral line, Nature (London) **408**, 51 (2000).

[41] S. Coombs, P. Görner, and H. Münz, *The Mechanosensory Lateral Line: Neurobiology and Evolution* (Springer Science and Business Media, New York, 2012).

[42] N. Schulte-Pelkum, S. Wieskotten, W. Hanke, G. Dehnhardt, and B. Mauck, Tracking of biogenic hydrodynamic trails in harbour seals (*Phoca vitulina*), J. Exp. Biol. **210**, 781 (2007).

[43] S. P. Windsor, S. E. Norris, S. M. Cameron, G. D. Mallinson, and J. C. Montgomery, The flow fields involved in hydrodynamic imaging by blind Mexican cave fish (*Astyanax fasciatus*). Part I: Open water and heading towards a wall, J. Exp. Biol. **213**, 3819 (2010).

[44] M. J. McHenry, J. A. Strother, and S. M. Van Netten, Mechanical filtering by the boundary layer and fluid–structure interaction in the superficial neuromast of the fish lateral line system, J. Comp. Physiol. A **194**, 795 (2008).

[45] M. E. McConney *et al.*, Biologically inspired design of hydrogel-capped hair sensors for enhanced underwater flow detection, Soft Matter **5**, 292 (2009).

[46] H. Beem, M. Hildner, and M. Triantafyllou, Calibration and validation of a harbor seal whisker-inspired flow sensor, Smart Mater. Struct. **22**, 014012 (2013).

[47] A. C. Hollenbeck, R. Grandhi, J. H. Hansen, and A. M. Pankonien, Bioinspired artificial hair sensors for flight-by-feel of unmanned aerial vehicles: A review, AIAA J. **61**, 5206 (2023).

[48] A. D. Hasler and A. T. Scholz, *Olfactory Imprinting and Homing in Salmon: Investigations into the Mechanism of the Imprinting Process* (Springer Science and Business Media, New York, 2012).

[49] E. Zermelo, Über das Navigationsproblem bei ruhender oder veränderlicher Windverteilung, ZAMM—J. App. Math. Mech./Z. Angew. Math. Mech. **11**, 114 (1931).

[50] P. Gunnarson, I. Mandralis, G. Novati, P. Koumoutsakos, and J. O. Dabiri, Learning efficient navigation in vortical flow fields, Nat. Commun. **12**, 7143 (2021).

[51] Y. Jiao, H. Hang, J. Merel, and E. Kanso, Sensing flow gradients is necessary for learning autonomous underwater navigation, Nat. Commun. **16**, 3044 (2025).

[52] K. Krishna, Z. Song, and S. L. Brunton, Finite-horizon, energy-efficient trajectories in unsteady flows, Proc. R. Soc. A **478**, 20210255 (2022).

[53] L. Biferale, F. Bonaccorso, M. Buzzicotti, P. Clark Di Leoni, and K. Gustavsson, Zermelo's problem: Optimal point-to-point navigation in 2D turbulent flows using reinforcement learning, Chaos **29**, 103138 (2019).

[54] G. Reddy, A. Celani, T. J. Sejnowski and M. Vergassola, Learning to soar in turbulent environments, Proc. Natl. Acad. Sci. USA **113**, 4877 (2016).

[55] G. Reddy, J. Wong-Ng, A. Celani, T. J. Sejnowski, and M. Vergassola, Glider soaring via reinforcement learning in the field, Nature (London) **562**, 236 (2018).

[56] S. Verma, G. Novati, and P. Koumoutsakos, Efficient collective swimming by harnessing vortices through deep reinforcement learning, Proc. Natl. Acad. Sci. USA **115**, 5849 (2018).

[57] C. David, J. Kennedy, and A. Ludlow, Finding of a sex pheromone source by gypsy moths released in the field, Nature (London) **303**, 804 (1983).

[58] J. Murlis, J. S. Elkinton, and R. T. Carde, Odor plumes and how insects use them, Annu. Rev. Entomol. **37**, 505 (1992).

[59] A. Loisy and C. Eloy, Searching for a source without gradients: How good is infotaxis and how to beat it, Proc. R. Soc. A **478**, 20220118 (2022).

[60] K. V. B. Verano, E. Panizon, and A. Celani, Olfactory search with finite-state controllers, Proc. Natl. Acad. Sci. USA **120**, e2304230120 (2023).

[61] W. Hanke, C. Brucker, and H. Bleckmann, The ageing of the low-frequency water disturbances caused by swimming goldfish and its possible relevance to prey detection, J. Exp. Biol. **203**, 1193 (2000).

[62] K. Taira *et al.*, Modal analysis of fluid flows: Applications and outlook, AIAA J. **58**, 998 (2020).

[63] B. Colvert, K. K. Chen, and E. Kanso, Bioinspired sensory systems for shear flow detection, J. Nonlinear Sci. **27**, 1183 (2017).

[64] F. van Breugel and M. H. Dickinson, Plume-tracking behavior of flying *Drosophila* emerges from a set of distinct sensory-motor reflexes, Curr. Biol. **24**, 274 (2014).

[65] G. Reddy, B. I. Shraiman, and M. Vergassola, Sector search strategies for odor trail tracking, Proc. Natl. Acad. Sci. USA **119**, e2107431118 (2022).

[66] T. Redaelli, C. Eloy, and E. Kanso, Efficient strategy to track a planktonic target from its hydrodynamic signature, in *APS Division of Fluid Dynamics Meeting Abstracts* (American Physical Society, Phoenix, Arizona, 2021), Q11-009.

[67] J. Liu, J. H. Costello, and E. Kanso, Flow physics of nutrient transport drives functional design of ciliates, Nat. Commun. **16**, 4154 (2025).

[68] J. A. Farrell, J. Murlis, X. Long, W. Li, and R. T. Cardé, Filament-based atmospheric dispersion model to achieve short time-scale structure of odor plumes, Environ. Fluid Mech. **2**, 143 (2002).

[69] IBAMR: An adaptive and distributed-memory parallel implementation of the immersed boundary (IB) method (2018), https://ibamr.github.io.

[70] A. P. S. Bhalla, R. Bale, B. E. Griffith, and N. A. Patankar, A unified mathematical framework and an adaptive numerical method for fluid–structure interaction with rigid, deforming, and elastic bodies, J. Comput. Phys. **250**, 446 (2013).

[71] B. E. Griffith, R. D. Hornung, D. M. McQueen, and C. S. Peskin, An adaptive, formally second order accurate version of the immersed boundary method, J. Comput. Phys. **223**, 10 (2007).

[72] G. T. Csanady, *Turbulent Diffusion in the Environment* (Springer Science and Business Media, New York, 2012).

[73] N. E. Leonard and E. Fiorelli, Virtual leaders, artificial potentials and coordinated control of groups, in *Proceedings of the 40th IEEE Conference on Decision and Control (Cat. No. 01CH37228)* (IEEE, Orlando, Florida, 2001), Vol. 3, pp. 2968–2973.

[74] J. Cochran and M. Krstic, Nonholonomic source seeking with tuning of angular velocity, Trans. Autom. Control, **54**, 717 (2009).

[75] Y. Huang, J. Yen, and E. Kanso, Detection and tracking of chemical trails in bio-inspired sensory systems, Eur. J. Comput. Mech. **26**, 98 (2017).

[76] L. Ristroph, J. C. Liao, and J. Zhang, Lateral line layout correlates with the differential hydrodynamic pressure on swimming fish, Phys. Rev. Lett. **114**, 018102 (2015).

[77] B. Colvert and E. Kanso, Fishlike rheotaxis, J. Fluid Mech. **793**, 656 (2016).

[78] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, Proximal policy optimization algorithms, arXiv:1707.06347.

[79] Y. Jiao *et al.*, Deep Dive into Model-free Reinforcement Learning for Biological and Robotic Systems: Theory and Practice, arXiv:2405.11457.

[80] See Supplemental Material at http://link.aps.org/supplemental/10.1103/qj9q-p7rg for Movies S1–S3 presents representative hydrodynamic trail-following trajectories for the RL-trained and RL-inspired control policies. It includes examples from the training domain as well as generalization to unseen flow conditions, spanning 2D and 3D wakes and turbulent plumes. Performance is further demonstrated across multiple sensing modalities, including flow speed, pressure, vorticity, and odor concentration.

[81] T. Seyde *et al.*, Is bang-bang control all you need? solving continuous control with Bernoulli policies, Adv. Neural Inf. Proc. Syst. **34**, 27209 (2021).

[82] V. Braitenberg, *Vehicles: Experiments in Synthetic Psychology* (MIT Press, Cambridge, Massachusetts, 1986).

[83] D. Shaikh and I. Rañó, Braitenberg vehicles as computational tools for research in neuro-science, Front. Bioeng. Biotechnol. **8**, 565963 (2020).

[84] T. Salumäe, I. Rañó, O. Akanyeti, and M. Kruusmaa, Against the flow: A Braitenberg controller for a fish robot, in *2012 IEEE International Conference on Robotics and Automation* (IEEE, Saint Paul, Minnesota, 2012), pp. 4210–4215.

[85] Y. Zhu, A. Nern, S. L. Zipursky, and M. A. Frye, Peripheral visual circuits functionally segregate motion and phototaxis behaviors in the fly, Curr. Biol. **19**, 613 (2009).

[86] A. Lilienthal and T. Duckett, Experimental analysis of gas-sensitive Braitenberg vehicles, Adv. Robotics **18**, 817 (2004).

[87] S. Heydari and E. Kanso, School cohesion, speed and efficiency are modulated by the swimmers flapping motion, J. Fluid Mech. **922**, A27 (2021).

[88] G. Arranz, O. Flores, and M. Garcia-Villalba, Flow interaction of three-dimensional self propelled flexible plates in tandem, J. Fluid Mech. **931**, A5 (2022).

[89] S. Heydari, H. Hang, and E. Kanso, Mapping spatial patterns to energetic benefits in groups of flow-coupled swimmers, elife **2024**, 02 (2024).

[90] G. S. Triantafyllou, M. S. Triantafyllou, and M. A. Grosenbaugh, Optimal thrust development in oscillating foils with application to fish propulsion, J. Fluids Struct. **7**, 205 (1993).

[91] D. Floryan, T. Van Buren, C. W. Rowley, and A. J. Smits, Scaling the propulsive performance of heaving and pitching foils, J. Fluid Mech. **822**, 386 (2017).

[92] J. W. Newbolt, J. Zhang, and L. Ristroph, Flow interactions between uncoordinated flapping swimmers give rise to group cohesion, Proc. Natl. Acad. Sci. USA **116**, 2419 (2019).

[93] P. Oteiza, I. Odstrcil, G. Lauder, R. Portugues, and F. Engert, A novel mechanism for mechanosensory-based rheotaxis in larval zebrafish, Nature (London) **547**, 445 (2017).

[94] R. Venturelli *et al.*, Hydrodynamic pressure sensing with an artificial lateral line in steady and unsteady flows, Bioinspiration Biomimetics **7**, 036004 (2012).

[95] J.-H. Seo and R. Mittal, Improved swimming performance in schooling fish via leading-edge vortex enhancement, Bioinspiration Biomimetics **17**, 066020 (2022).

[96] C. H. Williamson and A. Roshko, Vortex formation in the wake of an oscillating cylinder, J. Fluids Struct. **2**, 355 (1988).

[97] T. Schnipper, A. Andersen, and T. Bohr, Vortex wakes of a flapping foil, J. Fluid Mech. **633**, 411 (2009).

[98] S. Verma, C. Papadimitriou, N. Lüthen, G. Arampatzis, and P. Koumoutsakos, Optimal sensor placement for artificial swimmers, J. Fluid Mech. **884**, A24 (2020).

[99] P. Weber *et al.*, Optimal flow sensing for schooling swimmers, Biomimetics **5**, 10 (2020).

[100] G. Vanwalleghem, K. Schuster, M. A. Taylor, I. A. Favre-Bulle, and E. K. Scott, Brain-wide mapping of water flow perception in zebrafish, J. Neurosci. **40**, 4130 (2020).

[101] M. S. Madhav and N. J. Cowan, The synergy between neuroscience and control theory: The nervous system as inspiration for hard control challenges, Annu. Rev. Control Rob. Auton. Syst. **3**, 243 (2020).

[102] R. A. Kastelein, P. J. Wensveen, J. M. Terhune, and C. A. de Jong, Near-threshold equal-loudness contours for harbor seals (*Phoca vitulina*) derived from reaction times during underwater audiometry: A preliminary study, J. Acoust. Soc. Am. **129**, 488 (2011).

[103] P. Domenici and M. E. Hale, Escape responses of fish: A review of the diversity in motor control, kinematics and behaviour, J. Exp. Biol. **222**, jeb166009 (2019).
[104] A. Roberts *et al.*, The decision to move: response times, neuronal circuits and sensory memory in a simple vertebrate, Proc. R. Soc. B **286**, 20190297 (2019).
[105] P. Domenici and R. W. Blake, The kinematics and performance of fish fast-start swimming, J. Exp. Biol. **200**, 1165 (1997).
[106] D. Weihs, The mechanism of rapid starting of slender fish, Biorheology **10**, 343 (1973).
[107] K. Sato *et al.*, Stroke frequency, but not swimming speed, is related to body size in free-ranging seabirds, pinnipeds and cetaceans, Proc. R. Soc. B **274**, 471 (2007).
[108] J. Brett, The relation of size to rate of oxygen consumption and sustained swimming speed of sockeye salmon (*Oncorhynchus nerka*), J. Fish. Board Can. **22**, 1491 (1965).
[109] M. Kasapi, P. Domenici, R. Blake, and D. Harper, The kinematics and performance of escape responses of the knifefish *Xenomystus nigri*, Can. J. Zool. **71**, 189 (1993).
[110] P. Domenici and R. W. Blake, The kinematics and performance of the escape response in the angelfish (*Pterophyllum eimekei*), J. Exp. Biol. **156**, 187 (1991).
[111] D. G. Harper and R. W. Blake, Fast-start performance of rainbow trout *Salmo gairdneri* and northern pike *Esox lucius*, J. Exp. Biol. **150**, 321 (1990).
[112] S. J. Peake and A. P. Farrell, Locomotory behaviour and post-exercise physiology in relation to swimming speed, gait transition and metabolism in free-swimming smallmouth bass (*Micropterus dolomieu*), J. Exp. Biol. **207**, 1563 (2004).
[113] J. D. Stieglitz, E. M. Mager, R. H. Hoenig, D. D. Benetti, and M. Grosell, Impacts of deepwater horizon crude oil exposure on adult mahi-mahi (*Coryphaena hippurus*) swim performance, Environ. Toxicol. Chem. **35**, 2613 (2016).
[114] J. D. Kieffer, D. Alsop, and C. M. Wood, A respirometric analysis of fuel use during aerobic swimming at different temperatures in rainbow trout (*Oncorhynchus mykiss*), J. Exp. Biol. **201**, 3123 (1998).
[115] N. C. Wegner, M. A. Drawbridge, and J. R. Hyde, Reduced swimming and metabolic fitness of aquaculture-reared California yellowtail (*Seriola dorsalis*) in comparison to wild-caught conspecifics, Aquaculture **486**, 51 (2018).
[116] B. Block *et al.*, Environmental preferences of yellowfin tuna (*Thunnus albacares*) at the northern extent of its range, Mar. Biol. **130**, 119 (1997).
[117] P. W. Webb and R. S. Keyes, Division of labor between median fins in swimming dolphin (Pisces: Coryphaenidae), Copeia **1981**, 901 (1981).
[118] F. E. Fish, J. Hurley, and D. P. Costa, Maneuverability by the sea lion *Zalophus californianus*: Turning performance of an unstable body design, J. Exp. Biol. **206**, 667 (2003).
[119] Some of these data were available in the min-max range (e.g., [102,103]), which we converted to mean and standard deviation using the range rule std $=$ (max $-$ min)/4 [120].
[120] G. Van Belle, *Statistical Rules of Thumb* (John Wiley & Sons, Hoboken, New Jersey, 2011).
[121] Y. Jiao, B. Colvert, Y. Man, M. J. McHenry, and E. Kanso, Evaluating evasion strategies in zebrafish larvae, Proc. Natl. Acad. Sci. USA **120**, e2218909120 (2023).
[122] S. A. Burden, T. Libby, K. Jayaram, S. Sponberg, and J. M. Donelan, Why animals can outrun robots, Sci. Rob. **9**, eadi9754 (2024).
[123] Y. Jiao *et al.*, Learning to swim in potential flow, Phys. Rev. Fluids **6**, 050505 (2021).
[124] J. Yu, L. Liu, L. Wang, M. Tan, and D. Xu, Turning control of a multilink biomimetic robotic fish, IEEE Trans. Rob. **24**, 201 (2008).
[125] K. Hirata, T. Takimoto, and K. Tamura, Study on turning performance of a fish robot, in *First International Symposium on Aqua Bio-Mechanisms* (Society of Aero-Aqua Bio-Mechanism, Honolulu, Hawaii, 2000), pp. 287–292.
[126] J. Liang, T. Wang, and L. Wen, Development of a two-joint robotic fish for real-world exploration, J. Field Rob. **28**, 70 (2011).
[127] C. S. Peskin, Numerical analysis of blood flow in the heart, J. Comput. Phys. **25**, 220 (1977).
[128] R. Mittal *et al.*, A versatile sharp interface immersed boundary method for incompressible flows with complex boundaries, J. Comput. Phys. **227**, 4825 (2008).
[129] A. Vargas, R. Mittal, and H. Dong, A computational study of the aerodynamic performance of a dragonfly wing section in gliding flight, Bioinspiration Biomimetics **3**, 026004 (2008).
[130] R. Mittal and G. Iaccarino, Immersed boundary methods, Annu. Rev. Fluid Mech. **37**, 239 (2005).
[131] OpenAI Spinning up documentation—Proximal policy optimization (2018), https://spinningup.openai.com/en/latest/algorithms/ppo.html.
[132] OpenAI Github baseline—Proximal policy optimization (2017), https://github.com/openai/baselines/tree/master/baselines/ppo2.