Interpretable and Generalizable Strategies for Stably Following Hydrodynamic Trails

Haotian Hang[1], Yusheng Jiao[1], Sina Heydari[1], Feng Ling[1], Josh Merel[2], and Eva Kanso[1,*]

[1] *Department of Aerospace and Mechanical Engineering, University of Southern California, 854 Downey way, Los Angeles, California 90089, USA*
[2] *Meta Reality Labs, Redmond, Washington, USA*

Abstract

Aquatic organisms offer compelling evidence that local flow sensing alone, without vision, is sufficient to guide them to the source of a vortical flow field, be it a swimming or stationary object. However, the feedback mechanisms that allow a flow-sensitive follower to track hydrodynamic trails remain opaque. Here, using high-fidelity fluid simulations and Reinforcement Learning (RL), we discovered two equally effective policies for trail following. While not apriori obvious, the RL policies led to parsimonious response strategies, analogous to Braitenberg's simplest vehicles, where a follower senses local flow signals and turns away from or towards the direction of stronger signal. We analyzed the stability of the RL-inspired strategies in ideal and simulated flows and demonstrated their robustness in tracking unfamiliar flows using diverse types of sensors. Our findings uncovered a surprising connection between the stability of hydrodynamic trail following and sense-to-response time delays, akin to those observed in the sensorimotor systems of aquatic organisms, and could guide future designs of flow-responsive autonomous robots.

Introduction

Most aquatic species have flow sensing abilities that supplement or even substitute vision, allowing them to negotiate conditions with reduced or no visibility [1]. The type and distribution of flow sensitive mechano-receptors vary by species, from the lateral line system in fish [2–6] to the whiskers of harbor seals [7, 8] and the setae and antennae of copepods [9, 10]. The principles that allow these mechano-receptors to detect flows [11, 12] have direct relevance to designing bio-inspired sensors [13–17]. However, how animals use and respond to detected flow signals for underwater navigation is far less understood [18, 19]. Experimental evidence in fish [1, 20, 21], harbor seals [7, 8], and copepods [22, 23] suggests that aquatic animals use information contained in flows generated by predators, prey, conspecifics, and even abiotic sources, for underwater navigation not only over short distances, as in the final stage of mate or prey localization [1, 9, 10], but also to locate distant objects by following hydrodynamic trails over considerable length and time spans [20, 24]. Yet, the strategies that these animals use to respond on the fly to local hydrodynamic cues remain opaque. Inspired by these remarkable trail following abilities, we seek to discover minimal flow sensing and response strategies that mimic these abilities.

Hydrodynamic trails, whether generated by a swimming fish or a stubborn rock, manifest clear patterns of alternating-signed swirls or vortices that contain information about the sources generating them [25–27]. Much effort has been dedicated to cataloging wake types [28, 29], mostly to uncover the thrust and drag forces exerted on the bodies generating the wakes [30, 31], but far less is known about the response strategies that allow a follower to track a hydrodynamic trail to its generating source by relying solely on local flow sensing [18, 24, 32]. Such understanding, besides its immediate relevance to naval and robotic applications for ocean monitoring and marine life surveillance and maintenance [33–37], can be applied more broadly in the context of navigating changeable environments [19, 32], and could guide future explorations of the neuronal mechanisms underlying trail following in aquatic animals [1, 7, 20, 38, 39].

This problem falls under the broader context of locally sensing a signal field and inferring unknown information about the field and its source [25, 26, 40–43], including steering the sensor to locate the source [18, 32, 44–47] or other target points [19, 48]. When the signal field is transported by laminar and diffusive processes, gradient based methods perform admirably in locating the source [47, 49, 50] and identifying optimal sensor placements [27, 43, 51]. Even when spatial gradients are not directly measurable, extremum

*Corresponding Author: Kanso@usc.edu

seeking algorithms allow a dynamic seeker to "perturb and observe" a slowly varying signal field over a period of fast oscillations to estimate the direction of changing signal [45, 52]. On the opposite end of the spectrum, when the signal field is advected by a turbulent flow, Bayesian inference methods, notably "infotaxis," are employed to continuously update an internal spatial map of the likelihood of the source location and to move in the direction that locally maximizes the expected rate of information [32, 53]. In this work, we focus on the in-between scenario, where the signal is transported by a flow field that is neither fully laminar nor fully turbulent, but characterized by coherent frequencies and geometric patterns traveling downstream from the generating source. In this scenario, the Reynolds number, defined as the ratio of inertial to viscous forces, is moderate and both inertial and viscous effects are at play. This flow regime, though encompassing the wakes of nearly all intermediate sized organisms and robots, is far less explored [18, 19].

Two notable exceptions emerged recently for navigation at intermediate Reynolds numbers featuring a handcrafted policy [18] and a reinforcement learning (RL) approach [19]. RL is an optimization framework that, unlike classic trajectory optimization, seeks to learn feedback mechanisms based on experience [54–58]. It has been successfully applied in robotic [59, 60] and simulated [61] bodies to identify sensing and control laws that produce desired behavior through interactions with the physical environment [62–64]. While the RL policy in [19] represents progress for underwater point-to-point navigation, the work lacked any exploration of the interpretability, stability, or generalizability of the RL policy [19]. Ensuring motion stability and robustness to unfamiliar flow environments are fundamental requirements for implementing feedback laws in real-world robotic applications, and failure to provide such guarantees limits the applicability of the RL policy in [19]. In contrast, the handcrafted policy in [18] offered stability guarantees and robustness in diverse flow environments but suffered from inefficiency due to its reliance on a time history of sensory information.

Here, we use Reinforcement Learning (RL) as a tool to discover policies for hydrodynamic trail following in the intermediate Reynolds number range that rely only on instantaneous and local flow sensing. We show that RL converges to one of two policies where a swimmer equipped with on-board flow sensors tends to respond to local flows by predominantly *turning towards* or *away* from the direction of stronger flow signal in a way reminiscent of Braitenberg's vehicles [65]. From these interpretable policies, we abstracted two handcrafted *RL-inspired strategies*, amenable to rigorous stability analysis and efficient exploration of the parameter space of observations and actions. Our investigations assert the generalizability and robustness of the RL-inspired strategies to unfamiliar flows and diverse types of flow sensors. We conclude by discussing the relevance of these strategies to biological and robotic systems.

Results

Model of Hydrodynamic Trail Follower. The hydrodynamic trail follower is modeled as a self-propelled particle $\mathbf{x}(t) = (x(t), y(t))$, with no inertia, moving at a constant speed V at an angle $\theta(t)$ from the x-axis. The unit vectors $\mathbf{t} = (\cos\theta, \sin\theta)$ and $\mathbf{n} = (-\sin\theta, \cos\theta)$ represent the body-fixed heading and normal directions (Fig. 1A). The trail follower is equipped with two sets of flow sensors, located at ℓ_{head} and ℓ_{tail} from the center, each providing one sensory measurement s_{head} and s_{tail}, respectively. Inspired by the lateral line system and functional evidence suggesting that flow receptors are optimized to correlate with differential flow signals [4, 5, 27, 51], we considered $s_{\text{head}} = (\mathbf{n} \cdot \nabla|\mathbf{u}|)_{\text{head}}$ and $s_{\text{tail}} = (\mathbf{n} \cdot \nabla|\mathbf{u}|)_{\text{tail}}$ that measure the local gradients of the fluid velocity normal to the follower's swimming direction. Save these two sensory measurements, the follower is at all time ignorant of all other features of the flow field and blind to its position relative to the source generating the flow. To respond to this limited and localized information, with no memory of past measurements, we provided the follower only control over its turning rate $\dot{\theta}$ (the dot denotes time derivative), as opposed to direct control over its heading angle θ [19]. The follower's motion is thus described by the nonholonomic unicycle model [45, 47, 66],

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

We aimed to learn control policies $\pi(s_{\text{head}}, s_{\text{tail}})$ that guide the follower to track a vortical wake to its source.

CFD Simulations of Unsteady Vortical Flows. Fig. 1B shows a hydrodynamic wake created by a pitching foil of chord length L, oscillating at a frequency f and peak-to-peak tailbeat amplitude A in an oncoming uniform flow of speed U and kinematic viscosity ν. The Reynolds number $\text{Re} = UL/\nu = 5000$

and Strouhal number $\mathrm{St} = fA/U = 0.25$ are comparable to those of swimming fish [30, 67]. The unsteady wake of alternating vortices has a wavelength $\lambda \approx L$, where λ is defined as the streamwise distance between neighboring replicate vortical structures. In addition to the flow field in Fig. 1B, we simulated a gallery of hydrodynamic trails encompassing both drag and thrust wakes, at a range of Reynolds numbers Re=200-5000, generated by uniform flows past flapping foils and fixed cylinders [68, 69]. We also used published results of three-dimensional flow past an undulating body [70] (Table S.1, §S.2).

Learning to Track Flows using RL. To train the follower to track the hydrodynamic trail from repeated experiences with the flow environment, we employed model-free deep reinforcement learning based on an actor-critic approach and a Proximal Policy Optimization algorithm [71] (§S.3). The RL policy took as observations the two sensory measurements s_{head} and s_{tail} and provided as action a noisy turning rate $\dot{\theta}$. To limit the action space, we imposed a threshold value $\Omega_{\max}$ on the maximal angular rotation rate $|\dot{\theta}| \leq \Omega_{\max}$. This, in turn, constrained the minimum turning radius $V/|\dot{\theta}| \geq V/\Omega_{\max}$ of the follower. When it reached the source in the x-direction, the follower received a sparse reward of $50e^{-y^2}$, decaying with the lateral distance y from the source. Additionally, to encourage initial learning, the follower received a 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 total shaping and sparse reward $r_t = -d(t) + d(t - \Delta t) + 50e^{-y^2}$.

We repeated the training twenty times in the flow field of Fig. 1. Each training session consisted of 30,000 episodes, with parameter values $\Omega_{\max} = 3\,U/L$, $V = 0.25U$, and $\ell_{\mathrm{head}} = \ell_{\mathrm{tail}} = 0.25L$. In each episode, the follower was assigned random initial position, orientation, and starting time relative to the motion of the unsteady wake. Out of the twenty training instances, eight converged to one class of policies, nine to another, and three did not converge. In Fig. 2, we report the evolution of the total reward for ten training instances, five from each of the two converged classes of policies, shown in pink and green; solid lines show average values and transparent envelopes show standard deviations. The two classes of policies exhibit striking features (Fig. 2B): both are largely independent of the sensory measurement at the head, and although both converge during training, they exhibit opposite behavior. The pink *amplification* policies instruct the follower to mainly turn in the same direction as the flow signal s_{tail} at the tail, that is, in the direction of increasing flow speed. The green *attenuation* policies instruct the follower to turn in the opposite direction to flow signal s_{tail} at the tail, that is, in the direction of decreasing flow speed. In both policies, the follower successfully tracks the wake to its generating source (Fig. 1B), but when following the amplification policy, it ventures more into the wake, slaloming between vortices, similar to experimental observations in live animals [7, 8, 20, 38].

RL-inspired Strategies. The two key features of the RL policies in Fig. 2B – the action being relatively independent of sensor measurement s_{head} at ℓ_{head} and dictating that the follower turns towards (amplification) or away (attenuation) from stronger flow signals at the maximal allowed turning rate $\Omega_{\max}$ – suggest that these policies can be approximated analytically as bang-bang controllers [72] whose behavior depends only on the sensory measurement s_{tail} at ℓ_{tail}. Starting from the colormaps in Fig. 2B that express the RL action $\dot{\theta}$ over the observation space $(s_{\mathrm{head}}, s_{\mathrm{tail}})$, without loss of significant features, we discounted the sensory measurement s_{head} at ℓ_{head} entirely and used ℓ and s to refer to the location and measurement of the tail sensor. These simplifications lead to the analytic expressions

$$\text{Amplification:} \quad \dot{\theta} = \Omega\,\mathrm{sgn}(s), \qquad \text{Attenuation:} \quad \dot{\theta} = -\Omega\,\mathrm{sgn}(s), \tag{2}$$

where $\Omega = \Omega_{\max}$. It is remarkable that RL robustly converged to two trail-following policies that can be readily interpretable and approximated by the simple analytic expressions in (2). These RL-inspired strategies exhibit similar behavior to the RL policies they approximate (Fig. 1, Movie S.1), while being more amenable to analysis.

Analogy to Braitenberg's Vehicles. The RL policies and RL-inspired strategies in Fig. 2B evoke Braitenberg's vehicles, originally proposed as thought experiments for constructing complex behavior from simple sensorimotor responses to environmental stimuli [65]. 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 if there is more of the stimulus to which the sensor is tuned. As in our RL and RL-inspired policies, 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 vehicle is responsive to stimulus and *turns towards*. Likewise, in the amplification policy, the swimmer turns towards larger flow speeds, while in the attenuation policy, it turns away. In uniform flow, the swimmer swims straight.

Performance of the RL Policies. To test the performance of the RL policies and RL-inspired strategies of Fig. 2B, and to standardize these tests across different wakes and policies, 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 follower. 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 $[0, 1/f]$. That is, for each wake and each policy, we run $765 \times 20 \times 5 = 76,500$ tests.

The amplification and attenuation RL policies achieved 86.82% and 91.47% success rates, while the RL-inspired counterparts achieved 87.5% and 100%, all when tested in the same wake used during training. Here, success rate is measured in terms of the ability of the follower to reorient and move upstream, independent of its final lateral distance y from the source generating the wake, whereas the lateral distance y enters explicitly in the reward value during training. Statistical properties of the $76,500$ tests for each policy showed that the amplification policy is more accurate in honing in on the source generating the flow while the attenuation strategy locates the source of the wake faster because the corresponding trajectories meander less in the wake (see quantitative analysis in §S.4 and Fig. S.1). By the same token, because it causes the follower to move inside the wake and slalom between vortices, the amplification policy provides the follower with more opportunities for utilizing the flow for efficient navigation. These trends are consistent in the original RL policies and simpler RL-inspired strategies.

We further challenged the follower in two major ways: (i) we decreased the rate at which it collected observations and responded to flows by increasing the decision timestep Δt, and (ii) we imposed sensory limits $s_{\min}$ below which it was not able to sense the flow signal (Fig. S.2). All policies exhibited remarkable robustness to these reduced abilities, with robustness exhibited by the RL-trained policies.

Equivalence between Sensor Placement at Tail and Time-Delayed Response. The fact that the RL policies ignore the sensory measurements at the head is puzzling. In nature, most sensors are located at the head. The setae and whiskers of copepods and harbor seals precede the organism, and the lateral line sensors, while distributed along the entire fish body, occur in higher density at the head [27, 51]. Why does our optimal RL solution seem to contradict these observations? It doesn't. It simply reflects a time delay between sensing and response. For example, consider a seal with flow sensors at its head and time delay between sensory input and response output. If the point-particle model is used to describe how the seal orients to the sensed signal, it would seem that its response depends on a past signal, now in the flow behind its current location, as if instantaneously responding to a sensor at its tail.

To support this proposition, we examined the RL policies in a simpler representation of the signal field that captures two salient features of the more complex wake: its (quasi) periodicity and traveling-wave character. Namely, we considered a policy $\pi(s(x(t-\tau), t-\tau))$ where the response depends on a delayed signal measured at time $t-\tau$, where τ is the time delay, in a 1D signal field that satisfies $g(x,t) = g(2\pi(x+\lambda f t)/\lambda)$, where λf is the speed of the traveling signal. Taking the Taylor series expansion of $x(t-\tau)$ about time t, assuming small τ and that the follower moves faster than the signal, we get that

$$\dot{\theta}(t) = \pi\left(s(x(t-\tau), t-\tau)\right) \approx \pi\left(s(x - \ell\cos\theta, t)\right), \qquad \text{where} \quad \ell = V\tau \tag{3}$$

That is, a time-delayed response is equivalent to having a sensor located at a distance $V\tau$ behind the follower's location. The follower with time-delayed response (dashed lines in Fig. 1B), like its counterpart with instantaneous response to a sensor at the tail (solid lines in Fig. 1B), successfully tracks the wake to its generating source. Hereafter, we focus on the instantaneous response because it is simpler to analyze, with the understanding that, when sensors are placed at the tail, it is equivalent to a time-delayed response.

Stability of RL-inspired Strategies in Traveling Wave Signal. To emphasize the importance of the sensor placement on the ability to follow hydrodynamic trails, we analyzed the stability of the RL-inspired strategies in a 1D traveling-wave signal of the form $g(x,t) = A\cos(2\pi(x+\lambda f t)/\lambda)$. The amplification and attenuation strategies are related by a change of sign from Ω to $-\Omega$, which, in the traveling wave signal, 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 amplification strategy with the understanding that the asymptotic behavior of both strategies is similar.

Trail following in this 1D traveling wave signal amounts to moving upstream in the direction of $\theta = 0$ opposite to the traveling wave direction. Numerical simulations show that upstream motion depends on the location ℓ of the sensor (Fig. 3A). Here, we let ℓ vary in sign: positive values correspond to sensor located ahead of the follower's center and negative values correspond to values behind the follower's center. We next introduced the coordinate transformation $z = x + \lambda f t$, where z represents a material point moving with the traveling signal, and substituted z and $g(z) = g(x + \lambda f t)$ into (1) and (2) to eliminate the explicit dependence on time,

$$\dot{z} = V\cos\theta + \lambda f, \qquad \dot{\theta} = \Omega\,\mathrm{sgn}\left[\frac{2\pi A}{\lambda}\sin\theta\sin\left(\frac{2\pi}{\lambda}\left(z + \ell\cos\theta\right)\right)\right]. \tag{4}$$

The system in (4) 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. In Fig. 3B are depictions of the phase portraits of (4) over the phase space (z, θ). Dashed black lines indicate level sets where $\dot{\theta}$ is 0. The phase velocity between these level sets alternate directions such that, for $\ell < 0$, a trajectory travels longer distance in the regions where the phase velocity points towards $\theta = 0$; the upstream equilibrium is thus stable. For $\ell > 0$, trajectories travel longer in the regions where the phase velocity points towards $\theta = \pm\pi$; the downstream equilibrium is stable. At the bifurcation point $\ell = 0$, both equilibria are neutrally stable.

Optimal Parameters of RL-inspired Strategies. The parameter values of the controllers in (2) were inherited from the RL policies, with angular speed $\Omega = \Omega_{max}$ and $\ell = -0.25L$. Could the performance of the trail follower be optimized further for a specific wake by proper choice of the turning radius and sensor placement? To address this question, we tested the performance of each RL-inspired strategies in (2) over an entire range of sensor locations $\ell/L \in [-0.75, 0.75]$ and angular speeds $\Omega L/U \in [0.5, 10]$, amounting to turning radii $V/\Omega L \in [0.025, 0.5]$. We used the same wake employed during RL training for which $\lambda/L \approx 1$ and we performed $76,500$ tests for each combination of parameters $(\ell/\lambda, V/\Omega L)$. The resulting success rates are shown in Fig. 2C. The parameter values of the RL policies are indicated in black dots. Both amplification (pink) and attenuation (green) policies are successful, with over 80% success rates, for an entire range of parameters provided that sensory measurements are taken at the tail. These results are consistent with the stability analysis in Fig. 3 and can be viewed as a nonlinear analysis of the stability of upstream motion in the 2D wake of Fig. 1. Importantly, because the performance of the RL-inspired strategies is insensitive to parameter values, a follower need not fine tune its sensor location ℓ (provided $\ell < 0$) and turning radius $V/\Omega L$ to a specific wake type; its performance should be robust to variations in the wake itself.

Generalization to Unfamiliar Flows. We next tested the performance of the strategies in (2) in wakes not seen during training of the "parent" RL policies. Specifically, we tested in four wakes generated by flows past pitching foils spanning Re=500-5000 and St=0.1-0.25; four wakes generated by flows past a fixed cylinder spanning Re=200-500; and one three-dimensional (3D) wake of past an undulating body at Re= 2400 from [70]. For the latter, we extended the simplified policies to control both yaw and pitch angles of the follower (§ S.5). Because each wake has a different wavelength, we used the wavelength λ of the wake to scale the corresponding sensor location $\ell = -0.25\lambda$ relative to the wake in Fig. 1. With no other parameter tuning, both amplification and attenuation strategies were equally successful, with success rates nearing 100%, in guiding the follower to reorient upstream and track the hydrodynamic trails (Fig. 4 and Suppl. Movies S.2). Despite their parsimony and simplicity, the RL-inspired strategies are remarkably generalizable, allowing the swimmer to follow hydrodynamic trails of various characteristics, from thrust to drag jets to 3D wake behind undulating swimmers (Fig. 4).

Robustness to Diverse Sensory Cues. How does the performance of these trail-following strategies depend on the type of flow information available to the follower? For example, if instead of measuring the difference in flow speed, the follower was able to measure pressure difference, how would that affect its ability to follow hydrodynamic trails?

To illustrate this point, we considered five types of sensors, each measuring a different signal carried by the wake, including differential measurements of vorticity, pressure, and tangential and normal components of fluid velocity. Additionally, because chemo- and mechano-sensing are often combined in biological behavior [21], with olfactory signals playing important roles in behaviors ranging from the homing of the Pacific salmon [73] to the mating of copepods and moth [74], we considered differential measurements of concentration of dissolved chemicals injected at the tail of the leader (flapping foil) and passively advected by its unsteady wake. RL training based on these five types of sensory cues converged in all but the policy based on vorticity measurements (Table S.2 and Fig. S.3). For tangential and normal flow velocities, two policies analogous to the amplification and attenuation policies in Fig. 2 emerged during training (Fig. S.3). For pressure and concentration, RL training led to only amplification-like policies. Just as the RL policies in Fig. 2 inspired simpler strategies in (2), the RL policies in Fig. S.3 also inspire strategies of the same form.

In Fig. 5, we show results based on direct implementation of the simpler RL-inspired strategies in (2) for the same parameter values used during RL-training $\ell = -0.25L$ and $\Omega = 3U/L$. Sample trajectories, starting from the same initial conditions and following the amplification strategy are plotted in Fig. 5A and Suppl. Movies S.3. Average success rates over the parameter space $(\ell/\lambda, V/\Omega L)$ are shown in Fig. 5B for both amplification and attenuation strategies, exhibiting high success rates, over 80%, over large ranges of the parameter space, provided that the sensory measurement is trailing the follower's position ($\ell < 0$). Consistent with RL-training, chemo-sensing functioned only under the amplification strategy, that is, it required turning in the direction of higher concentration. All other signals functioned well in amplification and attenuation policies, even the vorticity-based for which the training did not converge. This is because the reward used in RL-training required the follower to reach the source generating the wake, whereas the success rate in Fig. 5B is based on the ability of the follower to move upstream, even if it ends at a lateral offset y from the source (green trajectory in Fig. 5A). But even by this measure of success, the performance of all sensory measurements exceeded those of the vorticity sensor, both in range and robustness to parameter values and in maximum success rate. These results, failure of vorticity-based RL training and suboptimal performance of associated RL-inspired strategies, are consistent with recent reports of the inefficacy of sensing vorticity in point-to-point underwater navigation [19] and challenge approaches that advocate for sensing vorticity [75]. Sensing flow velocity, pressure, or concentration fare better in tracking flows. Indeed, since pressure sensors are most readily available commercially [36, 76], the strategy based on sensing pressure 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.

Discussion

Using RL, we discovered two policies for following hydrodynamic trails at intermediate Reynolds numbers. The policies rely only on local and instantaneous flow sensing. From these policies, we extracted two parsimonious, interpretable, and generalizable response strategies, where the swimmer measures locally a differential flow signal and responds by turning either towards or away from the direction of stronger signal (Fig. 2). These remarkably simple response strategies depend only on two parameters – the minimum turning radius $R = V/\Omega L$ of the follower and the sensor location ℓ. Through rigorous stability analyses, we proved that the RL-inspired strategies are unconditionally stable in signal fields with radial symmetry (§S.7 and Fig. S.4) and stable, provided posterior sensor placement, in signal fields with traveling wave character (Fig. 3). We showed that this requirement for sensor placement ℓ at the tail is equivalent to a time delay τ between sensing and response. Using Monte Carlo simulations, we demonstrated that these rigorous results carry over to unfamiliar wakes (Fig. 4) and to sensors that probe different types of flow signals (Fig. 5).

To assess the broader significance of our results, we gathered experimental data of the turning radius R, swimming speed V, and time delays τ between sensory input to motor response output recorded in fish [77, 78], harbor seals [79], and copepods [80] (Fig. 6A-C, § S.6 and Supplemental Excel Table). We chose these organisms because of the strong experimental evidence indicating their ability to follow hydrodynamic

trails [1, 7, 8, 20, 39, 81]. For example, a predator catfish, rather than directly heading towards a prey guppy, was reported to track the prey's sinuous trajectory, demonstrating a response to the prey's wake that strongly depended on the predator's lateral line mechanoreceptors [20, 81]. Harbor seals were shown to track the wakes of small submarines and other seals, even with blocked visual and audio cues [7, 8]. Male copepods track females over distances that far exceed their perceptive distance of odour cues [39]. Except for the copepod which inhabits a more viscous flow regime at Re $\approx$ 1-10, the Reynolds number in the experiments on fish and harbor seals ranged from 200 to 2000, comparable to the range explored in our simulations.

The turning radius and swimming speed of all surveyed aquatic organisms are shown in Fig. 6A in dimensional scale and in Fig. 6B in units of body length. In both representations, the copepod datapoint is an outlier. The time delays between mechanosensory input and motor response are shown Fig. 6C. Such neuronal delays are often considered detrimental for sensorimotor control [82], but our analysis suggests that they are beneficial for the stability of hydrodynamic trail following. In Fig. 6D, we expressed the data on the parameter space $(\ell/L, V/\Omega L)$, where $\ell = -V\tau$, and we superimposed the colormaps from Fig. 2C indicating the success rates of the RL-inspired strategies. Organisms, such as the copepod, that live outside the flow regime explored in our study do not, and are not expected to, lie within the predicted optimal range of parameter values. Remarkably, organisms that inhabit and track hydrodynamic trails at comparable Reynolds numbers [7, 8, 20, 81] all lie within the optimal parameter range. These findings provide strong evidence that our RL-inspired strategies are achievable in biological systems that inhabit the same flow regime and encourage future experimental studies to test this prospect.

In addition to biological data, we collected data from bioinspired robotic systems at different levels of biomimicry [83–86]. Robotic systems capable of large ranges of body deformations exhibited turning radii comparable to those of swimming organisms [83–85]. Using the results in Fig. 2D, we predicted ranges of sensor locations (shown in yellow in Fig. 6D) for which these robotic systems would exhibit success rates exceeding 95%. To probe these predictions numerically, we employed a three-link fish model [64, 87] and developed, based on the RL-inspired amplification and attenuation strategies, controllers that mapped local flow signals to body deformations, much like Braitenberg's vehicles that linked signal intensity to wheel rotation. Numerical tests indicate success of the three-link fish in tracking a traveling wave signal (Suppl. Movie S.4). Further analysis and experimental validation of these predictions will be the topic of future research.

Lastly, we comment on the relation to extremum seeking [45, 52] and infotaxis [32]. Extremum seeking is an algorithm for estimating the direction of changing signal [52]. This feature for measuring local gradients could be combined with our RL-inspired strategies, which exhibit simpler response and unconditional stability in similar situations to those considered in [45] (§S.7 and Fig. S.4). Infotaxis [32] is designed to work when information is sparse, such as in turbulent conditions where detectable signals are distributed in random and disconnected patches, but suffers from a high cost of computation and memory requirements. Our trail following policies are much more efficient, and although were obtained under information-rich conditions, performed admirably when tested under limited sensory abilities (Fig. S.2). These results encourage future work for evaluating the performance of these strategies in random and patchy signal fields and mapping their efficacy to admissible levels of signal stochasticity and sparsity.

Acknowledgement

This work is supported by the Office of Naval Research (ONR) Grants N00014-22-1-2655, N00014-19-1-2035, N00014-17-1-2062, and N00014-14-1-0421, and the National Science Foundation (NSF) Grants RAISE IOS-2034043 and CBET-2100209. We thank Dr. Rajat Mittal and Dr. Jung-Hee Seo of Johns Hopkins University for providing us with the 3D simulation dataset.

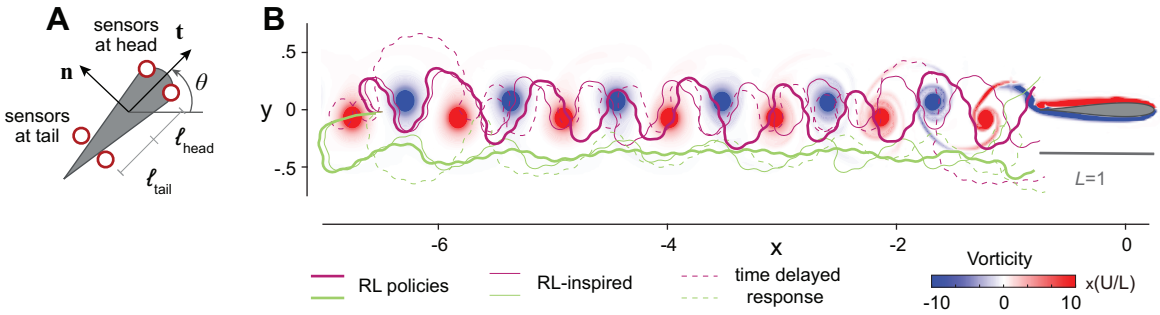

Figure 1: **Following hydrodynamic trails. A.** The follower is modeled as a particle with body-fixed frame $(\mathbf{t}, \mathbf{n})$, moving at a speed V and able to control its heading direction $\mathbf{t}$. The heading direction is given by its angle θ relative to an inertial x-axis. Two pairs of sensors, each measuring the normal gradient of the flow signal, are placed on the head and tail of the swimmer at distances ℓ_{head} and ℓ_{tail} from the center. **B.** Hydrodynamic trail created by a Joukowsky airfoil undergoing pitching oscillations in uniform background flow $U = 1$ with tailbeat amplitude $A = 0.2$ and frequency $f = 1.25$, at Re=5000 and St=0.25, comparable to swimming fish [30, 67]. Example trajectories of amplification (pink) and attenuation (green) policies are superimposed; RL-policies in thick solid lines, RL-inspired strategies in thin solid lines, and time-delayed response in dashed lines. See also Movies S.1.

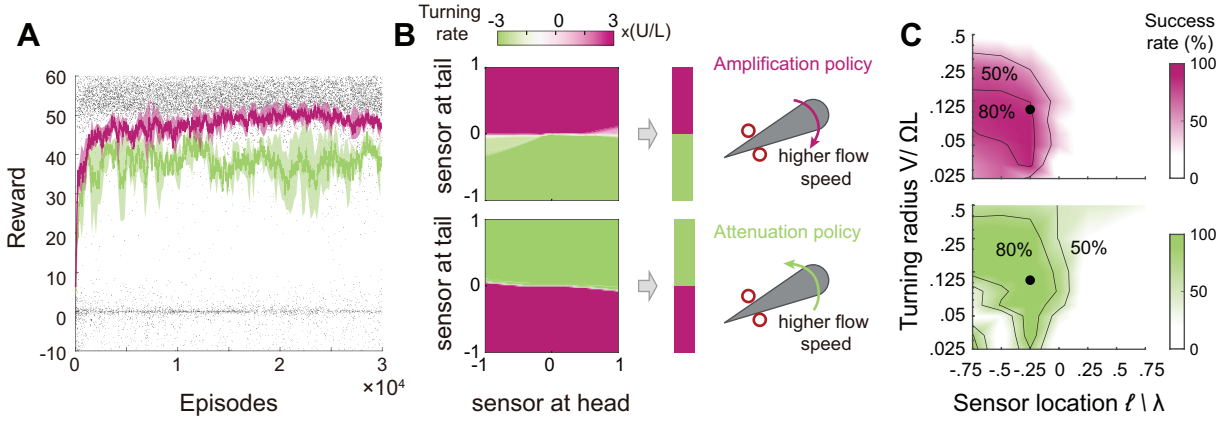

Figure 2: **RL-policies for tracking hydrodynamic trails. A.** Learning curves corresponding to 10 different training instances, where half converged to one of two classes of policies; moving average of the total reward is shown in pink and green, total reward for randomly-chosen 10% of the training episodes from all training instances is shown as black dots. **B.** Representative policies of the two classes of converged policies represented as colormaps over the observation space $(s_{\text{tail}}, s_{\text{head}})$, normalized by maximal signal value in the flow field. Dominant features of these policies are captured by parsimonious Braitenberg-like strategies with amplification and attenuation responses to the flow signal at the tail. Evaluation of the follower's behavior based on the RL policies and simplified counterparts (see Supplemental § S.4 and Figs. S.1 and S.2) show no major difference in performance. **C.** Parametric study, using the simplified policies in panel B, of the success rate of the follower over the space of sensor location ℓ/λ and turning radius $V/\Omega L$.

References

[1] H. Bleckmann, J. Mogdans, and S. L. Coombs. Flow sensing in air and water. *Berlin, Germany*, 976, 2014.

[2] S. Dijkgraaf. The functioning and significance of the lateral-line organs. *Biological reviews*, 38(1):51–105, 1963.

[3] J. Blaxter. Structure and development of the lateral line. *Biological Reviews*, 62(4):471–514, 1987.

[4] A. Kroese and N. Schellart. Velocity-and acceleration-sensitive units in the trunk lateral line of the trout. *Journal of Neurophysiology*, 68(6):2212–2221, 1992.

[5] J. Engelmann, W. Hanke, J. Mogdans, and H. Bleckmann. Hydrodynamic stimuli and the fish lateral line. *Nature*, 408(6808):51–52, 2000.

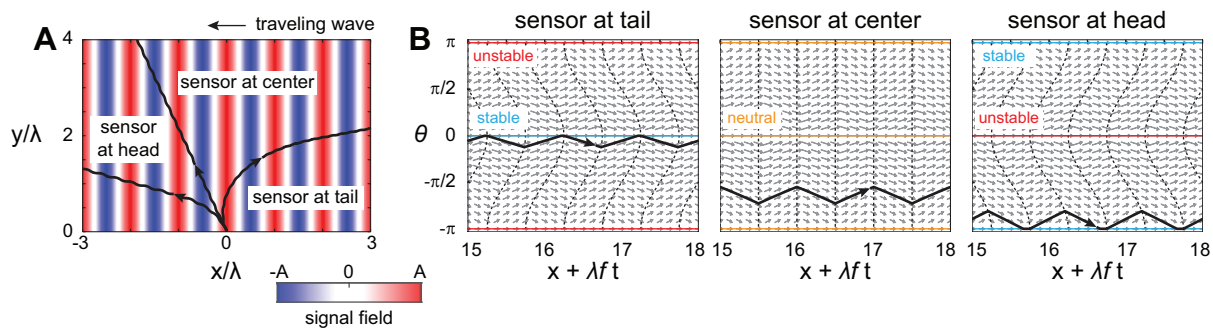

Figure 3: **Stability analysis in 1D traveling signal field. A.** Sample trajectories of follower with sensor placements at tail ($\ell = -0.25$), center ($\ell = 0$), and head ($\ell = 0.25$) and initial conditions $(x, \theta, t)|_0 = (0, 0.55\pi, 0)$ in a simplified signal field $A\cos(2\pi x/\lambda + 2\pi f t)$. **B.** Phase portraits on the phase space $(x + \lambda f t, \theta)$ phase space for sensor at tail ($\ell = -0.25$), center ($\ell = 0$), and head ($\ell = 0.25$). Trajectories are superimposed after they have reached steady state. Upstream motion is stable only for sensor placement behind the follower's position. Motion towards the source of a diffusive signal field is unconditionally stable (See Supplemental §. S.7 and Fig. S.4). Swimming speed $V = 0.25$ and turning rate $\Omega = 1$.

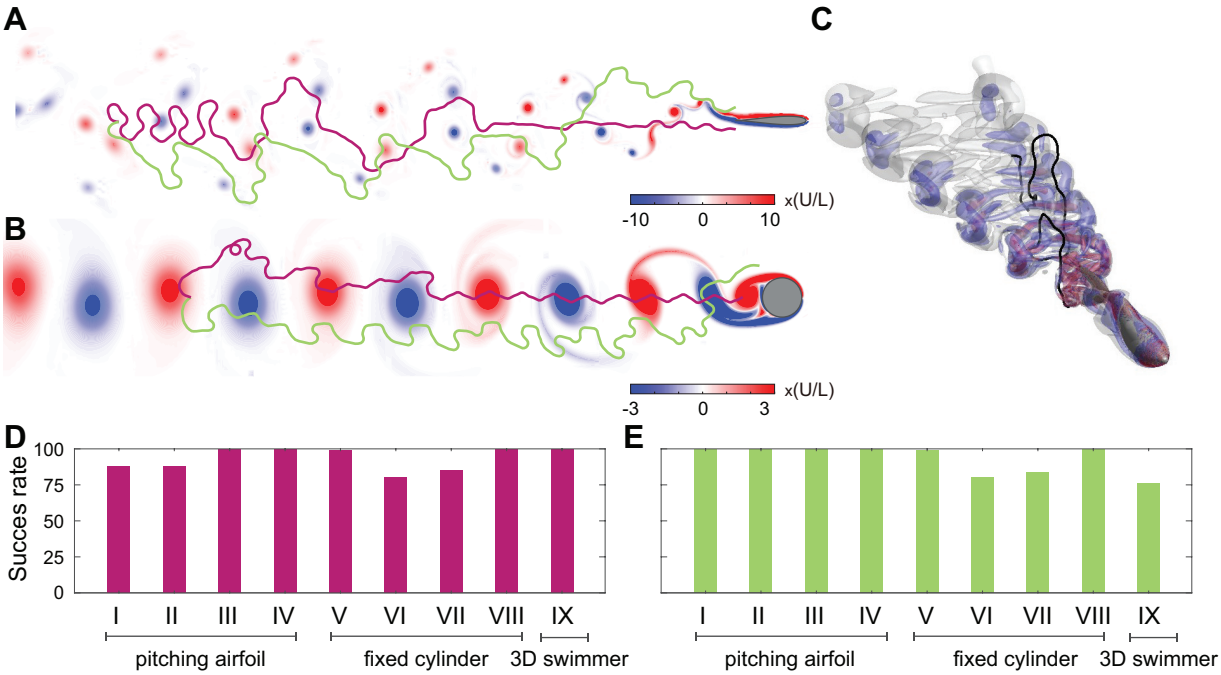

Figure 4: **Generalization to unfamiliar flows.** Sample trajectories based on employing the RL-inspired strategies in unseen wakes **A.** pitching airfoil at Re= 5000, St= 0.1, **B.** fixed cylinder at Re= 400, and **C.** 3D swimming fish [70] showing the isosurfaces of the Q-criterion = 0.01, 1, 10 in light grey, blue and red. **D, E.** Success rates of amplification (pink) and attenuation (green) strategies over 76,500 test cases in each of the eight 2D wakes and 90,000 test cases in the 3D wake. Namely, we tested in four wakes behind pitching airfoils: (I) Re= 5000, St= 0.25, (II) Re= 500, St= 0.25, (III) Re= 5000, St= 0.1, (IV) Re= 500, St= 0.1; four wakes behind fixed cylinders: (V) Re= 200, (VI) Re= 300, (VII) Re= 400, (VIII) Re= 500; and (IX) one 3D wake behind a swimming fish at Re= 5000. See Supplemental Table S.1 for detailed wake parameters. Sensor location is set to $\ell = -0.25\lambda$ with respect to the wavelength λ of the corresponding wake.

[6] S. Coombs, P. Görner, and H. Münz. *The mechanosensory lateral line: neurobiology and evolution.* Springer Science & Business Media, 2012.

[7] G. Dehnhardt, B. Mauck, W. Hanke, and H. Bleckmann. Hydrodynamic trail-following in harbor seals (phoca vitulina). *Science*, 293(5527):102–104, 2001.

[8] N. Schulte-Pelkum, S. Wieskotten, W. Hanke, G. Dehnhardt, and B. Mauck. Tracking of biogenic

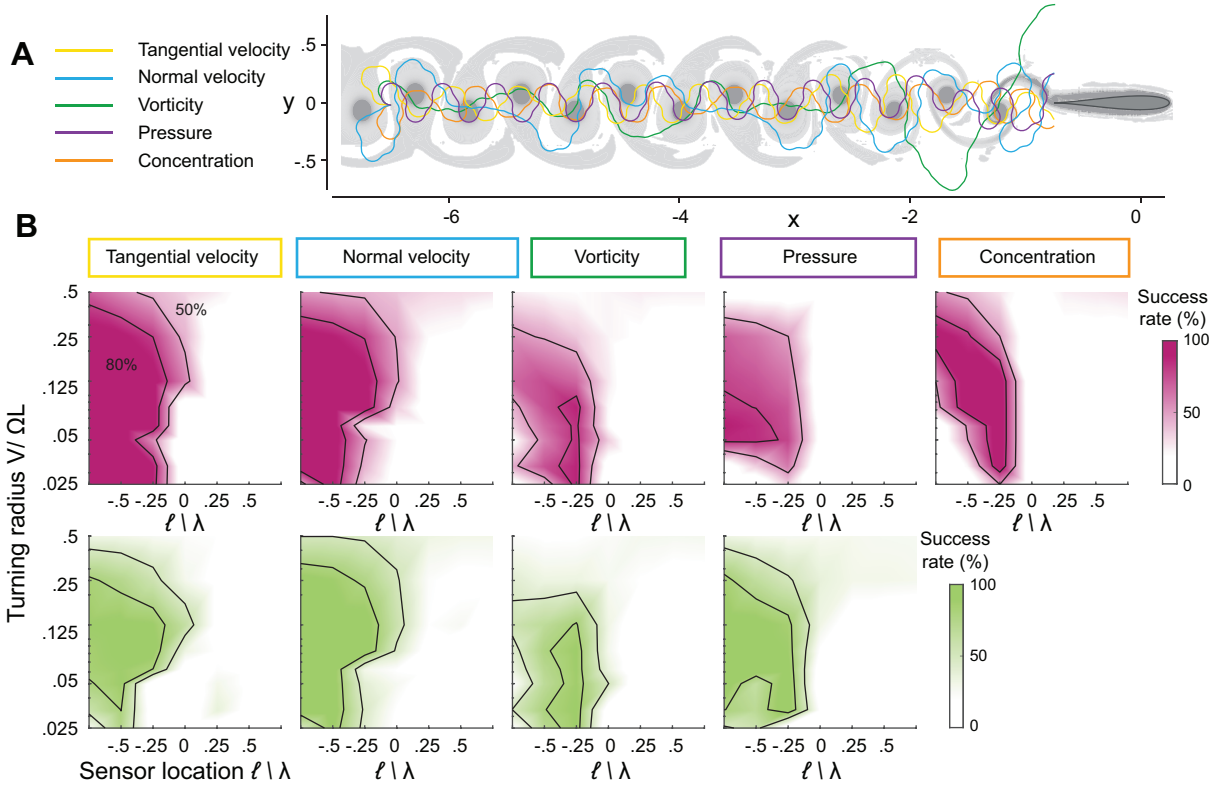

Figure 5: **Diversity of sensory cues. A.** Trajectories based on five different sensory cues starting from the same initial location and orientation are superimposed on grey-scale colormap of the vorticity field in Fig. 1B. Sensory signals consist of differential measurements of tangential velocity component $\mathbf{n}\cdot\nabla(\mathbf{u}\cdot\mathbf{t})$, normal velocity component $\mathbf{n}\cdot\nabla(\mathbf{u}\cdot\mathbf{n})$, vorticity $\mathbf{n}\cdot\nabla|\omega|$, pressure $\mathbf{n}\cdot(-\nabla p)$, and concentration $\mathbf{n}\cdot\nabla c$. **B.** Success rate of the swimmer as a function of sensor location ℓ/λ and turning radius $R = V/\Omega L$ for different sensory cues with amplification (pink) and attenuation (green) RL-inspired strategies.

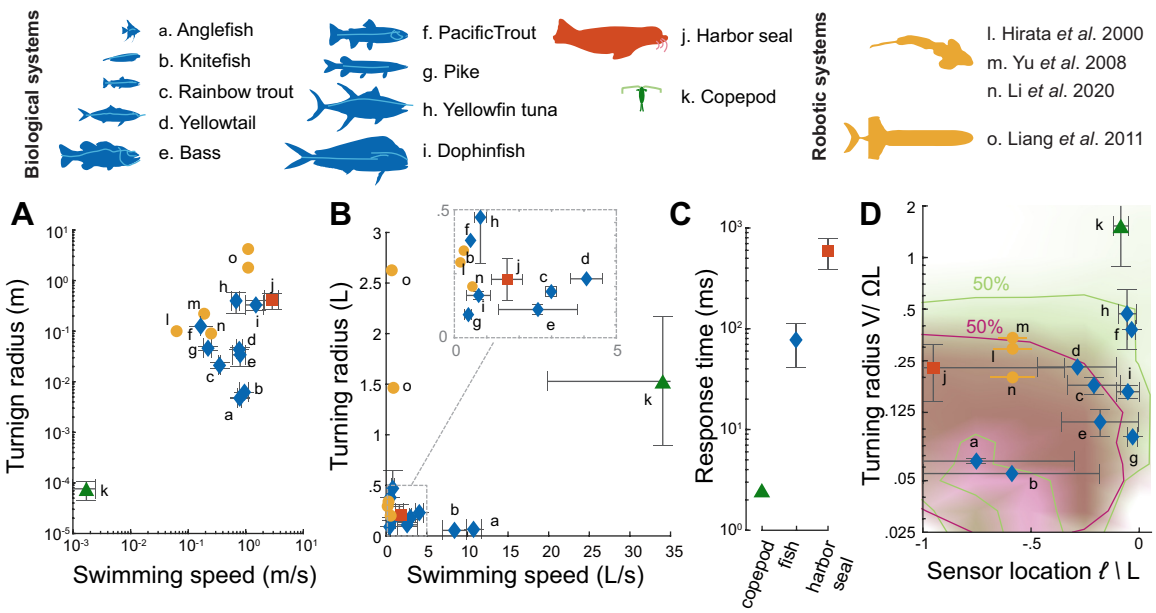

Figure 6: **Comparison to biological and robotic systems.** We gathered biological data from fish, harbor seals and their sea lion relatives, and copepods, as well as from underwater robotic systems. Complete dataset available in Supplemental Methods § S.6 and Supplemental Excel sheet. **A.** Turning radius versus swimming speed in log-log scale and dimensional units, and **B.** in units of body length L . **C.** Motor response time to mechano-sensory input. Error bars represent the standard deviation of biological data. **D.** Data mapped onto the parameter space $(\ell/L, V/\Omega L)$ in semi-log scale with sensor location $\ell = -V\tau$ based on the distance travelled during the neuronal response time τ. For robotic fish data, we obtained the minimum turning rate from the literature and calculated the optimal value of sensor location ℓ/L based on our model.

hydrodynamic trails in harbour seals (phoca vitulina). *Journal of Experimental Biology*, 210(5):781–787, 2007.

[9] J. Yen and J. R. Strickler. Advertisement and concealment in the plankton: what makes a copepod hydrodynamically conspicuous? *Invertebrate Biology*, pages 191–205, 1996.

[10] J. Yen, D. W. Murphy, L. Fan, and D. R. Webster. Sensory-motor systems of copepods involved in their escape from suction feeding. *Integrative and comparative biology*, 55(1):121–133, 2015.

[11] 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. *Journal of Comparative Physiology A*, 194:795–810, 2008.

[12] 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. *Journal of Experimental Biology*, 213(22):3819–3831, 2010.

[13] M. E. McConney, N. Chen, D. Lu, H. A. Hu, S. Coombs, C. Liu, and V. V. Tsukruk. Biologically inspired design of hydrogel-capped hair sensors for enhanced underwater flow detection. *Soft Matter*, 5(2):292–295, 2009.

[14] H. Beem, M. Hildner, and M. Triantafyllou. Calibration and validation of a harbor seal whisker-inspired flow sensor. *Smart Materials and Structures*, 22(1):014012, 2012.

[15] A. G. P. Kottapalli, M. Asadnia, J. Miao, and M. Triantafyllou. Touch at a distance sensing: lateral-line inspired mems flow sensors. *Bioinspiration & biomimetics*, 9(4):046011, 2014.

[16] Y. Zhai, X. Zheng, and G. Xie. Fish lateral line inspired flow sensors and flow-aided control: A review. *Journal of Bionic Engineering*, 18(2):264–291, 2021.

[17] 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 Journal*, pages 1–26, 2023.

[18] B. Colvert, G. Liu, H. Dong, and E. Kanso. Flowtaxis in the wakes of oscillating airfoils. *Theoretical and Computational Fluid Dynamics*, 34(4):545–556, 2020.

[19] P. Gunnarson, I. Mandralis, G. Novati, P. Koumoutsakos, and J. O. Dabiri. Learning efficient navigation in vortical flow fields. *Nature communications*, 12(1):1–7, 2021.

[20] K. Pohlmann, F. W. Grasso, and T. Breithaupt. Tracking wakes: the nocturnal predatory strategy of piscivorous catfish. *Proceedings of the National Academy of Sciences*, 98(13):7371–7374, 2001.

[21] J. M. Gardiner and J. Atema. Sharks need the lateral line to locate odor sources: rheotaxis and eddy chemotaxis. *Journal of Experimental Biology*, 210(11):1925–1934, 2007.

[22] M. J. Weissburg, M. Doall, and J. Yen. Following the invisible trail: kinematic analysis of mate-tracking in the copepod temora longicornis. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353(1369):701–712, 1998.

[23] J. Yen, J. K. Sehn, K. Catton, A. Kramer, and O. Sarnelle. Pheromone trail following in three dimensions by the freshwater copepod hesperodiaptomus shoshone. *Journal of plankton research*, 33(6):907–916, 2011.

[24] G. R. Spedding. Wake signature detection. *Annual review of fluid mechanics*, 46:273–302, 2014.

[25] B. Colvert, M. Alsalman, and E. Kanso. Classifying vortex wakes using neural networks. *Bioinspiration & biomimetics*, 13(2):025003, 2018.

[26] M. Alsalman, B. Colvert, and E. Kanso. Training bioinspired sensors to classify flows. *Bioinspiration & biomimetics*, 14(1):016009, 2018.

[27] L. Ristroph, J. C. Liao, and J. Zhang. Lateral line layout correlates with the differential hydrodynamic pressure on swimming fish. *Physical Review Letters*, 114(1):018102, 2015.

[28] C. H. Williamson and A. Roshko. Vortex formation in the wake of an oscillating cylinder. *Journal of fluids and structures*, 2(4):355–381, 1988.

[29] T. Schnipper, A. Andersen, and T. Bohr. Vortex wakes of a flapping foil. *Journal of Fluid Mechanics*, 633:411, 2009.

[30] G. S. Triantafyllou, M. S. Triantafyllou, and M. A. Grosenbaugh. Optimal thrust development in oscillating foils with application to fish propulsion. *Journal of Fluids and Structures*, 7(2):205–224, 1993.

[31] E. D. Tytell and G. V. Lauder. The hydrodynamics of eel swimming: I. wake structure. *Journal of Experimental Biology*, 207(11):1825–1841, 2004.

[32] M. Vergassola, E. Villermaux, and B. I. Shraiman. 'infotaxis' as a strategy for searching without gradients. *Nature*, 445(7126):406–409, 2007.

[33] J. A. Farrell, S. Pang, and W. Li. Chemical plume tracing via an autonomous underwater vehicle. *IEEE Journal of Oceanic Engineering*, 30(2):428–442, 2005.

[34] H. Ishida, G. Nakayama, T. Nakamoto, and T. Moriizumi. Controlling a gas/odor plume-tracking robot based on transient responses of gas sensors. *IEEE Sensors Journal*, 5(3):537–545, 2005.

[35] 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)*, pages 91–96. IEEE, 2012.

[36] T. Salumäe, I. Ranó, O. Akanyeti, and M. Kruusmaa. Against the flow: A braitenberg controller for a fish robot. In *2012 IEEE International Conference on Robotics and Automation*, pages 4210–4215. IEEE, 2012.

[37] X. Ai, K. You, and S. Song. A source-seeking strategy for an autonomous underwater vehicle via on-line field estimation. In *2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pages 1–6. IEEE, 2016.

[38] J. C. Liao, D. N. Beal, G. V. Lauder, and M. S. Triantafyllou. Fish exploiting vortices decrease muscle activity. *Science*, 302(5650):1566–1569, 2003.

[39] J. Yen, M. J. Weissburg, and M. H. Doall. The fluid physics of signal perception by mate-tracking copepods. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353(1369):787–804, 1998.

[40] B. Colvert, K. Chen, and E. Kanso. Local flow characterization using bioinspired sensory information. *Journal of Fluid Mechanics*, 818:366, 2017.

[41] B. Colvert, K. K. Chen, and E. Kanso. Bioinspired sensory systems for shear flow detection. *Journal of Nonlinear Science*, 27(4):1183–1192, 2017.

[42] S. Verma, C. Papadimitriou, N. Lüthen, G. Arampatzis, and P. Koumoutsakos. Optimal sensor placement for artificial swimmers. *Journal of Fluid Mechanics*, 884, 2020.

[43] P. Weber, G. Arampatzis, G. Novati, S. Verma, C. Papadimitriou, and P. Koumoutsakos. Optimal flow sensing for schooling swimmers. *Biomimetics*, 5(1):10, 2020.

[44] S. H. Singh, F. van Breugel, R. P. Rao, and B. W. Brunton. Emergent behaviour and neural dynamics in artificial agents tracking odour plumes. *Nature Machine Intelligence*, 5(1):58–70, 2023.

[45] J. Cochran and M. Krstic. Nonholonomic source seeking with tuning of angular velocity. *IEEE Transactions on Automatic Control*, 54(4):717–731, 2009.

[46] J. Cochran, E. Kanso, S. D. Kelly, H. Xiong, and M. Krstic. Source seeking for two nonholonomic models of fish locomotion. *IEEE Transactions on Robotics*, 25(5):1166–1176, 2009.

[47] Y. Huang, J. Yen, and E. Kanso. Detection and tracking of chemical trails in bio-inspired sensory systems. *European Journal of Computational Mechanics*, 26(1-2):98–114, 2017.

[48] Y. Jiao, H. Hang, J. Merel, and E. Kanso. Evaluating egocentric and geocentric sensing for efficient underwater navigation using deep reinforcement learning. *(in prep)*, 2023.

[49] R. E. Goldstein. Traveling-wave chemotaxis. *Physical review letters*, 77(4):775, 1996.

[50] J. A. Farrell, S. Pang, W. Li, and R. Arrieta. Chemical plume tracing experimental results with a remus auv. In *Oceans 2003. Celebrating the Past... Teaming Toward the Future (IEEE Cat. No. 03CH37492)*, volume 2, pages 962–968. IEEE, 2003.

[51] B. Colvert and E. Kanso. Fishlike rheotaxis. *Journal of Fluid Mechanics*, 793:656–666, 2016.

[52] M. Krstic and H.-H. Wang. Stability of extremum seeking feedback for general nonlinear dynamic systems. *Automatica-Kidlington*, 36(4):595–602, 2000.

[53] G. Reddy, V. N. Murthy, and M. Vergassola. Olfactory sensing and navigation in turbulent environments. *Annual Review of Condensed Matter Physics*, 13:191–213, 2022.

[54] T. Degris, P. M. Pilarski, and R. S. Sutton. Model-free reinforcement learning with continuous action in practice. In *2012 American Control Conference (ACC)*, pages 2177–2182. IEEE, 2012.

[55] A. M. Haith and J. W. Krakauer. Model-based and model-free mechanisms of human motor learning. In *Progress in motor control: Neural, computational and dynamic approaches*, pages 1–21. Springer, 2013.

[56] A. Cully, J. Clune, D. Tarapore, and J.-B. Mouret. Robots that can adapt like animals. *Nature*, 521(7553):503–507, 2015.

[57] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

[58] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[59] D. P. Ashmos, D. Duchon, and R. R. McDaniel Jr. Participation in strategic decision making: The role of organizational predisposition and issue interpretation. *Decision Sciences*, 29(1):25–51, 1998.

[60] J. Kober, J. A. Bagnell, and J. Peters. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 32(11):1238–1274, 2013.

[61] N. Heess, D. TB, S. Sriram, J. Lemmon, J. Merel, G. Wayne, Y. Tassa, T. Erez, Z. Wang, S. Eslami, et al. Emergence of locomotion behaviours in rich environments. *arXiv preprint arXiv:1707.02286*, 2017.

[62] G. Reddy, J. Wong-Ng, A. Celani, T. J. Sejnowski, and M. Vergassola. Glider soaring via reinforcement learning in the field. *Nature*, 562(7726):236–239, 2018.

[63] S. Verma, G. Novati, and P. Koumoutsakos. Efficient collective swimming by harnessing vortices through deep reinforcement learning. *Proceedings of the National Academy of Sciences*, 115(23):5849–5854, 2018.

[64] Y. Jiao, F. Ling, S. Heydari, N. Heess, J. Merel, and E. Kanso. Learning to swim in potential flow. *Phys. Rev. Fluids*, 6:050505, May 2021.

[65] V. Braitenberg. *Vehicles: Experiments in synthetic psychology*. MIT press, 1986.

[66] 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)*, volume 3, pages 2968–2973. IEEE, 2001.

[67] G. V. Lauder, J. Lim, R. Shelton, C. Witt, E. Anderson, and J. L. Tangorra. Robotic models for studying undulatory locomotion in fishes. *Marine Technology Society Journal*, 45(4):41–55, 2011.

[68] 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. *Journal of computational physics*, 223(1):10–49, 2007.

[69] IBAMR. Ibamr:an adaptive and distributed-memory parallel implementation of the immersed boundary (ib) method. https://ibamr.github.io.

[70] J.-H. Seo and R. Mittal. Improved swimming performance in schooling fish via leading-edge vortex enhancement. *Bioinspiration & Biomimetics*, 17(6):066020, 2022.

[71] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[72] T. Seyde, I. Gilitschenski, W. Schwarting, B. Stellato, M. Riedmiller, M. Wulfmeier, and D. Rus. Is bang-bang control all you need? solving continuous control with bernoulli policies. *Advances in Neural Information Processing Systems*, 34, 2021.

[73] A. D. Hasler and A. T. Scholz. *Olfactory imprinting and homing in salmon: Investigations into the mechanism of the imprinting process*, volume 14. Springer Science & Business Media, 2012.

[74] R. T. Cardé and A. Mafra-Neto. Mechanisms of flight of male moths to pheromone. In *Insect pheromone research*, pages 275–290. Springer, 1997.

[75] P. Oteiza, I. Odstrcil, G. Lauder, R. Portugues, and F. Engert. A novel mechanism for mechanosensory-based rheotaxis in larval zebrafish. *Nature*, 547(7664):445–448, 2017.

[76] R. Venturelli, O. Akanyeti, F. Visentin, J. Ježov, L. D. Chambers, G. Toming, J. Brown, M. Kruusmaa, W. M. Megill, and P. Fiorini. Hydrodynamic pressure sensing with an artificial lateral line in steady and unsteady flows. *Bioinspiration & biomimetics*, 7(3):036004, 2012.

[77] A. Roberts, R. Borisyuk, E. Buhl, A. Ferrario, S. Koutsikou, W.-C. Li, and S. R. Soffe. The decision to move: response times, neuronal circuits and sensory memory in a simple vertebrate. *Proceedings of the Royal Society B*, 286(1899):20190297, 2019.

[78] P. Domenici and M. E. Hale. Escape responses of fish: a review of the diversity in motor control, kinematics and behaviour. *Journal of Experimental Biology*, 222(18):jeb166009, 2019.

[79] 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. *The Journal of the Acoustical Society of America*, 129(1):488–495, 2011.

[80] P. Lenz and D. Hartline. Reaction times and force production during escape behavior of a calanoid copepod, undinula vulgaris. *Marine Biology*, 133:249–258, 1999.

[81] K. Pohlmann, J. Atema, and T. Breithaupt. The importance of the lateral line in nocturnal predation of piscivorous catfish. *Journal of Experimental Biology*, 207(17):2971–2978, 2004.

[82] M. S. Madhav and N. J. Cowan. The synergy between neuroscience and control theory: the nervous system as inspiration for hard control challenges. *Annual Review of Control, Robotics, and Autonomous Systems*, 3:243–267, 2020.

[83] J. Yu, L. Liu, L. Wang, M. Tan, and D. Xu. Turning control of a multilink biomimetic robotic fish. *IEEE Transactions on Robotics*, 24(1):201–206, 2008.

[84] K. Hirata, T. Takimoto, and K. Tamura. Study on turning performance of a fish robot. In *First International Symposium on Aqua Bio-Mechanisms*, pages 287–292. Mitaka, 2000.

[85] L. Li, M. Nagy, J. M. Graving, J. Bak-Coleman, G. Xie, and I. D. Couzin. Vortex phase matching as a strategy for schooling in robots and in fish. *Nature Communications*, 11(1):5408, 2020.

[86] J. Liang, T. Wang, and L. Wen. Development of a two-joint robotic fish for real-world exploration. *Journal of Field Robotics*, 28(1):70–79, 2011.

[87] Y. Jiao, B. Colvert, Y. Man, M. J. McHenry, and E. Kanso. Evaluating evasion strategies in zebrafish larvae. *Proceedings of the National Academy of Sciences*, 120(7):e2218909120, 2023.