PNAS
www.pnas.org

Supplementary Information for

Supplementary Document for: Evaluating Evasion Strategies in Zebrafish larvae

Yusheng Jiao[1], **Brendan Colvert**[2], **Yi Man**[1], **Matthew McHenry**[3], **Eva Kanso**[1]*

Eva Kanso
E-mail: kanso@usc.edu

This PDF file includes:

Figs. S1 to S17
Tables S1 to S2
SI References

*Corresponding Author

Table S1. List of notation used in the framework to evaluate evasion strategies (top) and in the mechanical model of three-link fish (bottom)

variable	meaning
n	dummy variable, index of evasion strategies, $n = 1, 2, 3, 4, 5, 6$
θ	response angle, change in prey orientation induced by C-start response
d	distance between predator and prey at the onset of a C-start evasion
ϕ	angular position of predator in the prey frame of reference
ψ	predator heading in the prey frame of reference
λ	deviation between angle of predator heading ψ and its angular position ϕ
U	prey speed
V	predator speed
χ	$\chi = \arccos(U/V)$
$\mathbf{s}$	actual variables of the predator at the onset of (can be any subset of (ϕ, ψ, λ)) as measured in experiment
$\mathbf{r}$	actual response of prey (equal to θ in this work) as measured in experiment
$\hat{\mathbf{s}}$	sensory cues perceived by prey, input to evasion strategy
$\hat{\mathbf{r}}$	evasion response intended by prey, output of evasion strategy
$f^{(n)}(\mathbf{s})$	deterministic form of evasion strategy n that maps sensory cues $\mathbf{s}$ to response$\mathbf{r}$
$p^{(n)}(\mathbf{s}, \mathbf{r})$	probabilistic form of evasion strategy n, a joint PDF of sensory cues $\mathbf{s}$ and response $\mathbf{r}$
$\boldsymbol{\sigma} \equiv (\sigma_\Phi, \sigma_\Psi, \sigma_\Lambda, \sigma_\Theta)$	noise parameters for sensing and response variables corresponding to the standard deviation of the von Mises distribution

variable	meaning
α_1, α_2	relative rotation angles of three-link fish, α_1 angle between head link and middle link, and α_2 angle between middle link and tail link
β	orientation of the middle link, $\beta(t) = \beta(0) + \theta$
m_s	mass of each segment in three-link fish
J_s	moment of inertia of each segment in three-link fish
m_{1a}	added mass of each link in the direction of the major axis
m_{2a}	added mass of each link in the direction of the minor axis
J_a	added moment of inertia of each link
m_1	total mass of each link in the major axis, $m_1 = m_s + m_{1a}$
m_2	total mass of each link in the minor axis, $m_2 = m_s + m_{2a}$
J	total moment of inertia of each link, $J = J_s + J_a$
(x, y)	position of center of three-link fish
(v_1, v_2)	velocity of center of three-link fish in body-fixed frame $v_1 = \dot{x}\cos\beta + \dot{y}\sin\beta$ and $v_2 = -\dot{x}\sin\beta + \dot{y}\cos\beta$

11

Yusheng Jiao[1], Brendan Colvert[2], Yi Man[1], Matthew McHenry[3], Eva Kanso[1]*

"Everything should be made as simple as possible, but not simpler" –Albert Einstein

1. Data Collection and Data Processing Protocols

A. Experimental set-up. We consider the experimental data partially published in (1), where zebrafish larvae were exposed to a robotic predator. All zebrafish larvae were bred from wild-type (AB line) colonies in a flow-through tank system (Aquatic Habitats, Apopka, FL, USA) that was maintained at 28.5°C on a 14L:10D cycle. The robotic predator consisted of a sacrified adult zebrafish controlled to move through an aquarium of otherwise still water through the action of a linear servo-motor (fig. S1A) (1, 2). The motion of the robotic predator was set to a constant speed $V = 2, 11, 20$ cm s^{-1} to reflect the entire range of speeds of a foraging predator (3). The servo-motor also propelled two high-speed (250 frames s^{-1}) cameras that were mounted above the predator to record the responses of larval zebrafish. The prey were generally motionless until exhibiting a fast start in response to the robot.

B. Processing of video recordings. Since the predator was rigid and moved forward at a constant speed in all the experiments, we conveniently defined an inertial frame of reference attached to the rostrum of the predator with the x-axis along its moving direction (fig. S1B). Based on the kinematic analysis in (1), we obtained for each recorded response, the time of the onset and the end of the prey C-start response t_1 and t_2, as well as the positions of the rostrum $\mathbf{x}_{\text{head},1}, \mathbf{x}_{\text{head},2}$ and center of mass $\mathbf{x}_{\text{com},1},, \mathbf{x}_{\text{com},2}$ of the prey at t_1 and t_2 in the predator frame of reference (fig. S1B). All original measurements are three dimensional. Here, we considered the azimuthal response and projected all kinematic quantities onto the two-dimensional plane where the predator was operated. In Fig. 2 of the main text, we show, relative to the predator frame of reference, the prey position and orientation at t_1 and t_2 for each predator speed, $V = 2, 11, 20$ cm s^{-1}.

C. Transformation to prey-centric frame of reference. We then introduced a prey frame of reference attached to the center of mass of the prey to determine the state of the predator at the onset of evasion in a prey-centric frame (fig. S1C). The predator state at the onset of an evasion event in the prey's frame of reference consisted of the predator relative distance d, predator angular position ϕ with respect to the prey heading, the predator heading ψ, and the deviation λ of the predator heading ψ from its angular position ϕ,

$$\begin{aligned} d &= ||\mathbf{x}_{\text{com},1}||, \qquad \phi = \arctan2(-\mathbf{x}_{\text{com},1}) - \arctan2(\mathbf{x}_{\text{head},1} - \mathbf{x}_{\text{com},1}), \\ \psi &= -\arctan2(\mathbf{x}_{\text{head},1} - \mathbf{x}_{\text{com},1}), \qquad \lambda = -\arctan2(\mathbf{x}_{\text{com},1}). \end{aligned} \tag{1}$$

Note that (d, ϕ) represent the polar coordinates of the predator in the prey's frame of reference.

The prey's response consisted of the direction of the escape response, here parameterized by the change in orientation θ relative to the prey's initial direction, and the average escape speed U,

$$\begin{aligned} \theta &= \arctan2(\mathbf{x}_{\text{head},2} - \mathbf{x}_{\text{com},2}) - \arctan2(\mathbf{x}_{\text{head},1} - \mathbf{x}_{\text{com},1}), \\ U &= \frac{\|\mathbf{x}_{\text{com},2} - \mathbf{x}_{\text{com},1}\|}{t_2 - t_1}. \end{aligned} \tag{2}$$

Note that the orientation of the prey is defined as the direction from the center of mass to the head. In (1), the response angle was defined as the angle $\arctan2(\mathbf{x}_{\text{com},2} - \mathbf{x}_{\text{com},1}) - \arctan2(\mathbf{x}_{\text{head},1} - \mathbf{x}_{\text{com},1})$ which is the angle that the displacement vector $(\mathbf{x}_{\text{com},2} - \mathbf{x}_{\text{com},1})$ of the fish center of mass makes during fast start relative to the fish heading $(\mathbf{x}_{\text{head},1} - \mathbf{x}_{\text{com},1})$ prior to evasion. Here, we define the response angle θ as the angle the prey heading direction of the prey makes between pre- and post-evasion.

D. Statistical analysis. We separated the experimental data into three sub-datasets based on the predator approaching speed: slow predator ($V = 2$ cm s^{-1}), mid-speed predator ($V = 11$ cm s^{-1}) and fast predator ($V = 20$ cm s^{-1}). The univariate distribution of the predator's state d, ϕ, ψ and λ at the onset of evasion are displayed in fig. S2(A-D) and the univariate distribution of the prey change in orientation θ at evasion are displayed in fig. S2E. Bivariate distributions between (ϕ, λ), (ϕ, θ), (ψ, θ) are shown in Figs. fig. S5–fig. S7.

For the univariate distributions of d, ϕ, ψ, λ and θ in fig. S2, we calculated the mean and standard deviation (using the Circular Statistic Toolbox in MATLAB for ϕ, ψ, and λ (4)); values reported in the caption of fig. S2 and in table S2. For the bivariate distributions in figs S3–S7, we calculated the correlations between the corresponding two variables and the p-values. The p-value indicates the probability that the true correlation coefficient is zero: smaller values of the p-value indicate higher confidence that a correlation between two (or more) variables exists. Conventionally, people use 0.05 or 0.01 as the threshold, lower p-values indicate a significant correlation. The correlation coefficient $\mathcal{C}$ ranges between -1 and 1, indicating the strength of the linear relationship between two variables and whether they are positively or negatively correlated. We also calculated the linear regression between the pairs of variables in figs. S5-S7. For example, in fig. S5, we set $\theta = c_1\phi + c_2$ and we sought values of c_1 and c_2 that best fit the data. To deal with the inherent periodicity of the variables θ and ϕ, we used a customized loss function $\mathfrak{L} = \sum_i [\sin(c_1\phi_i + c_2) - \sin(\theta_i)]^2 + \sum_i [\cos(c_1\phi_i + c_2) - \cos(\theta_i)]^2$, and we employed a gradient descent approach to solve for the coefficients c_1 and c_2; values are reported in the captions of figs S3–S7 and in table S2.

The predator relative distance d at which the prey was reactive extended up to 55 mm (the prey's length is about 4 mm) (fig. S2A). However, when plotting the bivariate histograms between the reactive distance d and prey's response θ (fig. S3), we

saw no marked correlation between the two, implying that the prey change in orientation θ at evasion is independent of the reactive distance d.

The predator's angular position ϕ at the onset of evasion covered all directions between 0 and 2π (fig. S2B), indicating that evasion responses were triggered at any relative angular position of the predator. However, although the prey was randomly placed in the tank, the angular positions ϕ that triggered evasion were not uniformly distributed: the prey were more alert to predatory signals approaching from behind, especially at slow predator speed. The prey perception of the predator is influenced by both vision (5) and flow sensing (2, 3, 6). Vision has a "blind zone" when the predator is approaching behind the prey ($\phi \to 180°$)(1, 7), but flow sensing can still be triggered. The prey were more responsive as the deviation $\lambda = \psi - \phi - \pi$ between the predator heading and angular position increase.

When plotting the bivariate histograms between ϕ and λ for which an evasion event is triggered (fig. S4), we observed two trends. First, both vision and flow sensing seemed to be compromised when the predator is approaching symmetrically behind the prey at $\phi = 180°$ and $\lambda = 0°$. Second, at slow predator speed, although λ and ϕ are mathematically independent, the bivariate histogram showed a weak correlation between the predator's angular position ϕ and relative heading λ. That is, at slow predator speed, the prey was more likely to escape when the predator is headed towards the front side of prey ($\lambda > 0$ for $0° < \phi < 180°$ and $\lambda < 0$ for $180° < \phi < 360°$), as opposed to when the predator was headed towards the back side of prey. This correlation got weaker for faster predator. These results suggest that the prey is more likely to initiate an escape response when it is more strongly stimulated.

The distribution of the predator heading direction ψ at the onset of evasion (fig. S2C) is consistent with these observations: evasion instances were less likely to be triggered when the predator heading ψ was close to 180° in magnitude. Importantly, when plotting the deviation $\lambda = \psi - \phi - \pi$ of the predator's heading away from the angular position where it appears in the prey's visual field (fig. S2D), we observed that evasion was rarely triggered when λ was close to 180° in magnitude, that is, when $\phi \approx \psi$ and the predator was heading away from the prey. Evasion was mostly triggered around $\lambda = 0$, for which $\psi = \phi + \pi$, when the predator was heading straight towards the prey.

To tease out how the prey's evasion response depends on the predator's state at evasion, we plotted the bivariate histograms between ϕ and θ (fig. S5), and between ψ and θ (fig. S6A) as well as between λ and θ (fig. S6B), for each predator speed and for the combined dataset. The results in fig. S5 show a linear correlation between θ and ϕ with a constant of proportionality that is approximately equal to 5/9. That is, when $\phi \to 0°$, the prey's change in heading could be as large as 100°. The change in heading is mediated by the C-start response of the prey; it is known that at about 100° angle, the efficiency of the C-start response drops, most likely due to the prey's mechanical constraints and wake interception (8).

Compared to the correlation between θ and ϕ, the results in fig. S6 show a weaker correlation between θ and ψ, and correspondingly, between θ and λ, especially at slow predator speed. There is also a large variance in these plots that is not explained by linear regression. However, when at each predator speed, we separated the data based on whether the stimulus is sinistral or dextral, that is, based on whether the predator approach heading at the onset of evasion was to the left or to the right of where it appeared in the predator field of vision (fig. S1C), we found a strong and consistent correlation between θ and ψ, especially at intermediate and fast predator speeds fig. S7.

E. Summary of insights based on statistical analysis. The statistical analysis in fig. S2–S7 and table S2 leads to the following general observations:

- The prey's change in orientation θ at evasion is independent of the reactive distance d (fig. S3).
- The prey's evasion response θ correlates linearly with ϕ, consistent with (9) (fig. S5); it also correlates linearly with ψ, especially when the data is arranged based on λ values to distinguish between sinistral and dextral stimuli (fig. S7). In all cases, the dependence of θ on ϕ (fig. S5) and on ψ (fig. S7) is qualitatively independent of the predator speed.
- Large reorientation angles θ exceeding about 100° occur at lower frequency, independent of the predator state (fig. S2E and fig. S5–fig. S7).

These observations imply that the evasion behavior is independent of d and that the predator speed V does not change qualitatively the trend of how θ depends on ϕ, ψ, λ. However, while insightful, these observations do not disambiguate which of the predator states at evasion ϕ, ψ, λ are most relevant for determining the evasion direction θ. To investigate these questions, we present six theoretical evasion models. In these models, the escape heading depends on one or more of the variables ϕ, ψ, λ, whereas V (or any quantity that depends on V) is included as a model parameter. We then develop a probabilistic approach for comparing the theoretical evasion models to experimental data, as explained in the following sections.

2. Theoretical Evasion Strategies

A. Evasion Strategies. We examined six candidate evasion strategies, or mappings from predator state to prey response. Four strategies are based on existing theories: pure-protean, distance-optimal, parallel, and contralateral. The orthogonal strategy is postulated based on intuition that the prey should escape in a direction that makes it challenging for the predator to change its approach course. The antipodal strategy is postulated based on present (fig. S5) and prior experimental observations (9). The contralateral strategy is also based on empirical observations (1). We described all strategies in a mathematically consistent manner assuming that, at evasion, the prey turns instantaneously and swims in the direction defined by θ, and that the prey and the predator are point masses equipped with heading directions.

Distance-optimal strategy According to classic game theory (10), the optimal strategy for the prey is to turn in a direction that maximizes its distance of closest approach to the predator. This requires the prey to estimate both the trajectory of the predator and its own trajectory after the escape response. Using $\mathbf{b}_1, \mathbf{b}_2$ to denote the prey-centric frame, the prey estimate of its own trajectory $\mathbf{x}_{\text{prey}}$ and that of the predator $\mathbf{x}_{\text{predator}}$ are given by

$$\begin{aligned}\mathbf{x}_{\text{prey}} &= Ut\cos\theta\mathbf{b}_1 + Ut\sin\theta\mathbf{b}_2,\\ \mathbf{x}_{\text{predator}} &= (d_o\cos\phi + Vt\cos\psi)\,\mathbf{b}_1 + (d_o\sin\phi + Vt\sin\psi)\,\mathbf{b}_2,\end{aligned} \tag{3}$$

where d_o is the distance between predator and prey right before the evasive response, and the respective speeds of the prey and predator U and V are both assumed to be constant. The estimated distance between predator and prey $d(t) = \|\mathbf{x}_{\text{prey}} - \mathbf{x}_{\text{predator}}\|$ is

$$d^2(t) = (d_o\cos\phi + Vt\cos\psi - Ut\cos\theta)^2 + (d_o\sin\phi + Vt\sin\psi - Ut\sin\theta)^2. \tag{4}$$

Taking the derivative of $d^2(t)$ with respect to t and finding its zero ($\partial(d^2)/\partial t = 0$) yield the time of closest approach

$$t_{\min} = \frac{d_o\left[U\cos(\theta-\phi) - V\cos(\psi-\phi)\right]}{U^2 + V^2 - 2UV\cos(\psi-\theta)}. \tag{5}$$

We substitute $t_{\min}$ into Eq. (4) to get the distance of closest approach

$$d^2_{\min} = \frac{d_o^2\left[U\sin(\theta-\phi) - V\sin(\psi-\phi)\right]^2}{U^2 + V^2 - 2UV\cos(\psi-\theta)}. \tag{6}$$

Next, we take the derivative of $d^2_{\min}$ with respect to θ and find its root ($\partial d^2_{\min}/\partial\theta = 0$) to get the optimal escape angle θ that maximizes the distance of closest approach for $V \geq U$ (predator faster than prey) and $\chi = \cos^{-1}(U/V)$,

$$\theta = f^{(1)}(\psi,\lambda;\chi) = \begin{cases}\psi - \chi & \lambda\in(0,\pi) \quad \text{(sinistral stimulus)},\\ \psi + \chi & \lambda\in(-\pi,0) \quad \text{(dextral stimulus)}.\end{cases} \tag{7}$$

When $U > V$ (prey faster than predator), the closest approach happens at $t_{\min} = 0$ and can be maximized by a range of escape angles. In this case we choose $\chi = 0, \theta = \psi$ to maximize the distance between predator and prey at all $t > 0$.

Here, the escape direction depends on ψ, λ, and on U/V, or more precisely on $\chi = \cos^{-1}(U/V)$. Note that $\psi = \phi + \lambda + \pi$, thus, in our analysis in section 3–4, we treat ϕ and λ as independent variables and χ as a parameter.

Orthogonal strategy If we take the speed ratio to the limit $U/V \to 0$ (fast predator), we obtain a strategy where the prey orients itself at $\chi = \pi/2$ from the heading angle of the predator. The predator will want to turn *away* from the predicted trajectory of the predator.

$$\theta = f^{(2)}(\psi,\lambda) = \begin{cases}\psi - \dfrac{\pi}{2}, & \lambda\in(0,\pi) \quad \text{(sinistral stimulus)},\\ \psi + \dfrac{\pi}{2}, & \lambda\in(-\pi,0) \quad \text{(dextral stimulus)}.\end{cases} \tag{8}$$

Here, the escape direction depends on ψ and λ; the prey does not need an exact estimate of the relative speed U/V.

Parallel strategy A simple strategy in the limit of slow predator $U/V > 1$ is for the prey to reorient itself along the predator heading: the predator is bound to stay behind the prey and never catch it. This strategy is equivalent to the distance-optimal strategy for $\chi = 0$. That is,

$$\theta = f^{(3)}(\psi) = \psi, \tag{9}$$

and the escape direction depends only on ψ. This strategy has a serious disadvantage compared to other strategies: it could place the predator in the prey's blind zone.

Antipodal strategy we suggest a strategy in which the prey reorients itself antipodal to the angular position of the predator, consistent with the observations in fig. S5,

$$\theta = f^{(4)}(\phi) = \begin{cases}\phi + \pi, & \phi\in(0,\pi) \quad \text{(left stimulus)}\\ \phi - \pi, & \phi\in(\pi,2\pi) \quad \text{(right stimulus)}.\end{cases} \tag{10}$$

Contralateral strategy In (1), the authors suggested that the prey escapes contralateral to the predator when approached from the side. This strategy predicts that the prey turns either 'left' or 'right' by $\pi/2$ depending on the predator's angular position.

$$\theta = f^{(5)}(\phi) = \begin{cases}-\frac{\pi}{2}, & \phi\in(0,\pi) \quad \text{(left stimulus)},\\ \frac{\pi}{2}, & \phi\in(\pi,2\pi) \quad \text{(right stimulus)}.\end{cases} \tag{11}$$

This is the simplest of all deterministic strategies: it requires only knowledge of whether the predator appears to the left or right of the prey. Compared to this strategy, the antipodal strategy in Eq. (10) requires the prey to have more precise sensory information about the predator's angular position.

Pure-protean strategy An evasion strategy by which an animal flees in a random direction may be optimal for survival (11): the high variability in evasion directions makes it difficult for the predator to anticipate the prey's heading. Mathematically, the pure-protean strategy consists of a completely random change in orientation, which requires no precise measurement of the predator state. However, the change of orientation θ in the experiment (fig. S2E) does not conform to a uniform distribution and exhibits strong correlations with the predator angular position ϕ (fig. S5) and heading ψ (fig. S7). The pure-protean strategy is unlikely to be the dominant strategy implemented by larval zebrafish, and we thus eliminate it from further consideration.

3. Probabilistic Modeling of Evasion

To emphasize the generality of the probabilistic approach presented here and in the main document, we write our approach in terms of stimulus $\mathbf{s}$ and response $\mathbf{r}$, without specifying apriori the degrees of freedom that they encompass. In the zebrafish experiments, $\mathbf{r} = \{\theta\}$ whereas $\mathbf{s}$ could include ϕ, ψ, λ; d is discounted based on the statistical analysis in section 1 and fig. S3, and V is treated separately as a model parameter (see section 7). Theoretically, the vector $\mathbf{s}$ depends on the evasion model (see section 2), and each evasion model can be written as a mapping from stimulus $\mathbf{s}$ to response $\mathbf{r}$ of the form

$$\mathbf{r} = \mathbf{f}(\mathbf{s}; n) \equiv \mathbf{f}^{(n)}(\mathbf{s}), \tag{12}$$

where n is a dummy index to distinguish between strategies.

A. Probabilistic interpretation of experimental data. We view the experimental data points $(\mathbf{s}_i, \mathbf{r}_i)$, $i = 1, \dots N$, as sampled from a joint probability density function (PDF), denoted $p_o(\mathbf{s}, \mathbf{r})$, whose exact form is unknown. An evasion behavior follows a conditional probability $p_o(\mathbf{r}|\mathbf{s})$, related to the joint probability $p_o(\mathbf{s}, \mathbf{r})$ via $p_o(\mathbf{s}, \mathbf{r}) = p_o(\mathbf{r}|\mathbf{s})p_o(\mathbf{s})$, where $p_o(\mathbf{s})$ is the distribution of stimuli that elicit an escape response (fig. S8A). Only discrete samples of the PDFs $p_o(\mathbf{s}, \mathbf{r})$ and $p_o(\mathbf{s})$ are available from experiments.

It is worth noting that in probability theory, the random variable is often denoted by a capital letter and a realization of the random variable is denoted by a lower case letter. For example, the random variable associated with the stimulus vector would be $\mathbf{S}$, whereas $\mathbf{s}$ would refer to a specific state occupied by the predator. It is also customary to write the probability density function associated with $\mathbf{S}$ as $p_{o_\mathbf{S}}(\mathbf{s})$. Here, we simplify the notation, with the understanding that, in the mathematical models, both $\mathbf{s}$ and $\mathbf{r}$ are random variables and $\mathbf{s}_i$ and $\mathbf{r}_i$ are realizations.

B. Stimulus and response without and with noise. We distinguish between the actual predator state $\mathbf{s}$ and the prey's perception $\hat{\mathbf{s}}$ of the predator state. Similarly, we distinguish between the actual escape response $\mathbf{r}$ and the prey's desired escape response $\hat{\mathbf{r}}$. If the prey's perception and motor response are precise, the stimulus perceived by the prey $\hat{\mathbf{s}}$ is the same as the actual predator state $\mathbf{s}$ obtained from video recording, and the prey's desired escape response $\hat{\mathbf{r}}$ is the same as the recorded prey escape response $\mathbf{r}$. However, the sensorimotor modalities underlying evasion are often noisy: the prey may perceive a noisy version $\hat{\mathbf{s}}$ of the predator's state $\mathbf{s}$ and its desired response $\hat{\mathbf{r}}$ may be altered by noisy execution or environmental conditions to yield $\mathbf{r}$. We treat the two cases, without and with perception and motor/environmental noise, separately (fig. S8B,C),

$$\begin{cases} \textbf{Without noise:} & \hat{\mathbf{s}} = \mathbf{s}, \qquad \hat{\mathbf{r}} = \mathbf{r}, \\ \textbf{With noise:} & p(\hat{\mathbf{s}}|\mathbf{s}) = \mathcal{S}(\hat{\mathbf{s}}; \mathbf{s}, \boldsymbol{\sigma}_\mathbf{S}), \qquad p(\mathbf{r}|\hat{\mathbf{r}}) = \mathcal{R}(\mathbf{r}; \hat{\mathbf{r}}, \boldsymbol{\sigma}_\mathbf{R}) \end{cases} \tag{13}$$

Here, $\mathcal{S}$ and $\mathcal{R}$ denote the probability distribution functions of the noise in sensing/perception and response, parameterized by the mean $\mathbf{s}$ and standard deviation $\boldsymbol{\sigma}_\mathbf{S}$, and the mean $\mathbf{r}$ and standard deviation $\boldsymbol{\sigma}_\mathbf{R}$, respectively. That is, in the perception model, we take $\hat{\mathbf{s}}$ to be normally-distributed around the actual state of the predator $\mathbf{s}$, with dispersion $\boldsymbol{\sigma}_\mathbf{S}$, and in the response model, we take $\mathbf{r}$ to be normally-distributed around the desired response $\hat{\mathbf{r}}$ of the prey, with dispersion $\boldsymbol{\sigma}_\mathbf{R}$. For the components of $\mathbf{s}$ and $\hat{\mathbf{r}}$ that are angles, we use the von Mises distribution (normal distribution on the circle), as discussed in section 4.

C. Conditional probability based on theoretical evasion strategies. In section 2, each strategy n defines a desired escape response $\hat{\mathbf{r}}$ given a perceived predatory stimulus $\hat{\mathbf{s}}$. An evasion strategy n can be expressed as a conditional probability using the Dirac-delta function

$$p^{(n)}(\hat{\mathbf{r}}|\hat{\mathbf{s}}) = \delta\left(\hat{\mathbf{r}} - \mathbf{f}^{(n)}(\hat{\mathbf{s}})\right). \tag{14}$$

D. Probabilistic evasion strategies without noise. In the absence of noise, we have $\hat{\mathbf{s}} = \mathbf{s}$ and $\mathbf{r} = \hat{\mathbf{r}}$. Using the probabilistic representation of evasion strategy Eq. (14), the joint probability $p^{(n)}(\mathbf{s}, \mathbf{r})$ following the theoretical evasion model n is given by

$$p^{(n)}(\mathbf{s}, \mathbf{r}) = p^{(n)}(\mathbf{r} = \hat{\mathbf{r}}|\hat{\mathbf{s}} = \mathbf{s})p_o(\mathbf{s}) = \delta\left(\mathbf{r} - \mathbf{f}^{(n)}(\mathbf{s})\right)p_o(\mathbf{s}). \tag{15}$$

E. Probabilistic evasion strategies with noise. In the presence of perception and response noise, the joint probability $p^{(n)}(\mathbf{s}, \mathbf{r})$ following the theoretical evasion model n is given by the Law of Total Probability (12)

$$p^{(n)}(\mathbf{s}, \mathbf{r}) = \iint p(\mathbf{r}|\hat{\mathbf{r}}; \boldsymbol{\sigma}_\mathbf{R})p^{(n)}(\hat{\mathbf{r}}|\hat{\mathbf{s}})p(\hat{\mathbf{s}}|\mathbf{s}; \boldsymbol{\sigma}_\mathbf{S})p_o(\mathbf{s})\mathrm{d}\hat{\mathbf{s}}\,\mathrm{d}\hat{\mathbf{r}}. \tag{16}$$

4. Explicit Expressions of Probabilistic Evasion Strategies With Perception and Response Noise

Our goal here is to rewrite the noisy probabilistic evasion model presented in Eq. (16) in component form for each evasion strategy n.

A. Perception noise. The predator's actual state is described by its angular position ϕ relative to the prey and heading ψ in a prey-centric frame. The deviation angle $\lambda = \psi - \phi - \pi$ depends on ϕ and ψ. Alternatively, one could consider ϕ and λ as independent variables, and $\psi = \lambda + \phi + \pi$ as the dependent variable. This distinction may not be important for the prey, or for how the escape heading depends on these angles, but it is important for how noise is incorporated in the model. Without loss of generality, we chose to apply normally-distributed noise to ϕ, λ and let the noise on ψ follow from $\psi = \phi + \lambda + \pi$.

Note that for completeness and to ensure robustness of our results to how noise is incorporated in the model, we also used ϕ and ψ as the independent variables where noise is applied and calculated the noise on λ from $\lambda = \psi - \phi - \pi$. Qualitatively, we arrived at the same conclusions. For simplicity of exposition, we present only the latter case where noise is applied to ϕ and λ for all strategies.

Mediated by the prey's ability to perceive the predator, the noisy estimate of the predator state is denoted as $\hat{\phi}$ and $\hat{\lambda}$. This noisy perception is modeled by the conditional distribution

$$p((\hat{\phi}, \hat{\lambda})|(\phi, \lambda)) = p(\hat{\phi}|\phi)p(\hat{\lambda}|\lambda), \tag{17}$$

where each state is perceived independently. Specifically, the noise follows the von Mises distrubtions, which approximates a normal distribution on a circle, centered around the actual state ϕ and λ of the predator,

$$p(\hat{\phi}|\phi) = \frac{1}{2\pi I_0(\sigma_\Phi^{-2})} \exp\left[\frac{\cos(\hat{\phi} - \phi)}{\sigma_\Phi^2}\right], \qquad p(\hat{\lambda}|\lambda) = \frac{1}{2\pi I_0(\sigma_\Lambda^{-2})} \exp\left[\frac{\cos(\hat{\lambda} - \lambda)}{\sigma_\Lambda^2}\right]. \tag{18}$$

Here, I_0 is the modified bessel function of order zero, and $\sigma_\Phi, \sigma_\Lambda$ are noise parameters representing the dispersion of the von Mises distributions over ϕ and λ.

B. Response noise. By means of its motor control and through interaction with the fluid environment, the prey engages in an escape maneuver which is characterized by its actual change in orientation θ. The actual escape heading θ might be different from the desired escape heading $\hat{\theta}$ due to motor control or environmental noise. The noise in the underlying biomechanical process and environmental conditions is modeled by the conditional distribution $p(\theta|\hat{\theta})$,

$$p(\theta|\hat{\theta}) = \frac{1}{2\pi I_0(\sigma_\Theta^{-2})} \exp\left[\frac{\cos(\hat{\theta} - \theta)}{\sigma_\Theta^2}\right], \tag{19}$$

where I_0 is the modified bessel function of order zero, and σ_Θ represents the dispersion in the actual response θ around the desired response $\hat{\theta}$.

C. Conditional probability based on theoretical evasion strategies. The prey's desired response $\hat{\theta}$ is modeled as one of several strategies described by the conditional distribution introduced in Eq. (15) and expressed here in component form

$$p^{(n)}(\hat{\theta}|(\hat{\phi}, \hat{\lambda})) = \delta\left(\hat{\theta} - f^{(n)}(\hat{\phi}, \hat{\lambda})\right). \tag{20}$$

Distance-optimal strategy conditional probability We rewrite Eq. (20) for the distance-optimal strategy described in Eq. (7). The desired escape heading $\hat{\theta}$ depends on the perceived heading of the predator $\hat{\psi}$ and perceived deviation $\hat{\lambda} = \hat{\psi} - \hat{\phi} - \pi$ between $\hat{\psi}$ and perceived angular position $\hat{\phi}$. The conditional probability distribution associated with Eq. (7) is given by

$$p^{(1)}\left(\hat{\theta}|(\hat{\psi}, \hat{\lambda}); \chi\right) = \delta\left(\hat{\theta} - \hat{\psi} + \chi\right) \mathrm{rect}_{(0,\pi)}(\hat{\lambda}) + \delta\left(\hat{\theta} - \hat{\psi} - \chi\right) \mathrm{rect}_{(-\pi,0)}(\hat{\lambda}), \tag{21}$$

where $\delta(\cdot)$ is the Dirac delta function and $\mathrm{rect}_{(a,b)}(\cdot)$ is the rectangular window function, equal to 1 between a and b and 0 otherwise.

Orthogonal strategy conditional probability The conditional probability distribution describing the orthogonal strategy in Eq. (8) is given by

$$p^{(2)}\left(\hat{\theta}|(\hat{\psi}, \hat{\lambda})\right) = \delta\left(\hat{\theta} - \hat{\psi} + \frac{\pi}{2}\right) \mathrm{rect}_{(0,\pi)}(\hat{\lambda}) + \delta\left(\hat{\theta} - \hat{\psi} - \frac{\pi}{2}\right) \mathrm{rect}_{(-\pi,0)}(\hat{\lambda}). \tag{22}$$

Parallel strategy conditional probability The conditional probability distribution describing the parallel strategy in Eq. (9) is given by

$$p^{(3)}\left(\hat{\theta}|\hat{\psi}\right) = \delta\left(\hat{\theta} - \hat{\psi}\right). \tag{23}$$

Antipodal strategy conditional probability According to Eq. (10), the conditional probability distribution describing the antipodal strategy is given by

$$p^{(4)}(\hat{\theta}|\hat{\phi}) = \delta\left(\hat{\theta} - \hat{\phi} - \pi\right) \mathrm{rect}_{(0,\pi)}(\hat{\phi}) + \delta\left(\hat{\theta} - \hat{\phi} + \pi\right) \mathrm{rect}_{(\pi,2\pi)}(\hat{\phi})$$

Using the periodicity condition $\phi + 2\pi \to \phi$, we get that

$$p^{(4)}(\hat{\theta}|\hat{\phi}) = \delta\left(\hat{\theta} - \hat{\phi} - \pi\right) \tag{24}$$

Contralateral strategy conditional probability According to Eq. (11), the conditional probability distribution associated with the contralateral strategy is

$$p^{(5)}(\hat{\theta}|\hat{\phi}) = \delta\left(\hat{\theta} + \frac{\pi}{2}\right) \mathrm{rect}_{(0,\pi)}(\hat{\phi}) + \delta\left(\hat{\theta} - \frac{\pi}{2}\right) \mathrm{rect}_{(\pi,2\pi)}(\hat{\phi}). \quad [25]$$

D. Probabilistic evasion models with perception and response noise. Using the Law of Total Probability (12), we rewrite Eq. (16) in component form and perform the cascading operations

$$p^{(n)}(\theta|(\phi,\lambda)) = \iiint p(\theta|\hat{\theta}) p^{(n)}(\hat{\theta}|(\hat{\phi},\hat{\lambda})) p((\hat{\phi},\hat{\lambda})|(\phi,\lambda)) \mathrm{d}\hat{\theta}\,\mathrm{d}\hat{\phi}\,\mathrm{d}\hat{\lambda}. \quad [26]$$

Given $p(\theta|(\phi,\lambda))$ and using the definition of conditional probability, the joint distribution of predator state and prey response can be computed using

$$p^{(n)}(\theta,\phi,\lambda) = p^{(n)}(\theta|(\phi,\lambda)) p_o(\phi,\lambda). \quad [27]$$

The joint distribution $p^{(n)}(\theta,\phi,\lambda)$ forms a *probabilistic model* with perception and response noise that can then be compared to measured experimental data.

In the following, we derive the conditional distribution $p^{(n)}(\theta|(\phi,\lambda))$ in Eq. (26) for each of the five strategies.

Distance-optimal strategy with perception and response noise Substitute $\hat{\psi} = \hat{\lambda} + \hat{\phi} + \pi$ into Eq. (21) to obtain the conditional probability based on the distance-optimal strategy

$$p^{(1)}(\hat{\theta}|\hat{\phi},\hat{\lambda}) = \delta\left(\hat{\theta} - \hat{\lambda} - \hat{\phi} - \pi + \chi\right) \mathrm{rect}_{(0,\pi)}(\hat{\lambda}) + \delta\left(\hat{\theta} - \hat{\lambda} - \hat{\phi} - \pi - \chi\right) \mathrm{rect}_{(-\pi,0)}(\hat{\lambda}). \quad [28]$$

Now, substitute Eq. (28), Eq. (18) and Eq. (19) into Eq. (26) and simplify the resulting integrals to arrive at

$$p^{(1)}(\theta|(\phi,\lambda)) = \frac{1}{8\pi^3 I_0(\sigma_\Lambda^{-2}) I_0(\sigma_\Phi^{-2}) I_0(\sigma_\Theta^{-2})} \left[\int_0^{2\pi}\int_0^{\pi} \exp\left(\frac{\cos(\hat{\theta}-\theta)}{\sigma_\Theta^2} + \frac{\cos(\hat{\lambda}-\lambda)}{\sigma_\Lambda^2} + \frac{\cos(\hat{\theta}-\hat{\lambda}-\pi+\chi-\phi)}{\sigma_\Phi^2} \right) \mathrm{d}\hat{\lambda}\mathrm{d}\hat{\theta} \right. \quad [29]$$
$$\left. + \int_0^{2\pi}\int_{-\pi}^{0} \exp\left(\frac{\cos(\hat{\theta}-\theta)}{\sigma_\Theta^2} + \frac{\cos(\hat{\lambda}-\lambda)}{\sigma_\Lambda^2} + \frac{\cos(\hat{\theta}-\hat{\lambda}-\pi-\chi-\phi)}{\sigma_\Phi^2} \right) \mathrm{d}\hat{\lambda}\mathrm{d}\hat{\theta} \right].$$

Orthogonal strategy with perception and response noise The computation is similar to that of the distance-optimal strategy, with $\chi = \cos^{-1}(U/V) = \pi/2$. We get

$$p^{(2)}(\theta|(\phi,\lambda)) = \frac{1}{8\pi^3 I_0(\sigma_\Lambda^{-2}) I_0(\sigma_\Phi^{-2}) I_0(\sigma_\Theta^{-2})} \left[\int_0^{2\pi}\int_0^{\pi} \exp\left(\frac{\cos(\hat{\theta}-\theta)}{\sigma_\Theta^2} + \frac{\cos(\hat{\lambda}-\lambda)}{\sigma_\Lambda^2} + \frac{\cos(\hat{\theta}-\hat{\lambda}-\frac{\pi}{2}-\phi)}{\sigma_\Phi^2} \right) \mathrm{d}\hat{\lambda}\mathrm{d}\hat{\theta} \right. \quad [30]$$
$$\left. + \int_0^{2\pi}\int_{-\pi}^{0} \exp\left(\frac{\cos(\hat{\theta}-\theta)}{\sigma_\Theta^2} + \frac{\cos(\hat{\lambda}-\lambda)}{\sigma_\Lambda^2} + \frac{\cos(\hat{\theta}-\hat{\lambda}-\frac{3\pi}{2}-\phi)}{\sigma_\Phi^2} \right) \mathrm{d}\hat{\lambda}\mathrm{d}\hat{\theta} \right].$$

Parallel strategy with perception and response noise Here also we have a similar computation with $\chi = \cos^{-1}(U/V) = 0$,

$$p^{(3)}(\theta|(\phi,\lambda)) = \frac{1}{8\pi^3 I_0(\sigma_\Lambda^{-2}) I_0(\sigma_\Phi^{-2}) I_0(\sigma_\Theta^{-2})} \int_0^{2\pi}\int_{-\pi}^{\pi} \exp\left(\frac{\cos(\hat{\theta}-\theta)}{\sigma_\Theta^2} + \frac{\cos(\hat{\lambda}-\lambda)}{\sigma_\Lambda^2} + \frac{\cos(\hat{\theta}-\hat{\lambda}-\pi-\phi)}{\sigma_\Phi^2} \right) \mathrm{d}\hat{\lambda}\mathrm{d}\hat{\theta}. \quad [31]$$

Antipodal strategy with perception and response noise Substitute Eq. (18), Eq. (19) and Eq. (24) into Eq. (26) and simplify the resulting integrals to arrive at

$$p^{(4)}(\theta|\phi) = \frac{1}{4\pi^2 I_0(\sigma_\Phi^{-2}) I_0(\sigma_\Theta^{-2})} \int_0^{2\pi} \exp\left(\frac{\cos(\hat{\theta}+\pi-\phi)}{\sigma_\Phi^2} + \frac{\cos(\hat{\theta}-\theta)}{\sigma_\Theta^2} \right) \mathrm{d}\hat{\theta}. \quad [32]$$

Contralateral strategy with perception and response noise Substitute Eq. (18), Eq. (19) and Eq. (25) into Eq. (26) and simplify the resulting integrals to arrive at

$$p^{(5)}(\theta|\phi) = \frac{1}{2\pi I_0(\sigma_\Theta^{-2})} \left[\exp\left(\frac{\cos(\theta+\frac{\pi}{2})}{\sigma_\Theta^2} \right) \xi(\phi) + \exp\left(\frac{\cos(\theta-\frac{\pi}{2})}{\sigma_\Theta^2} \right) (1-\xi(\phi)) \right], \quad [33]$$

where

$$\xi(\phi) = \frac{1}{2\pi I_0(\sigma_\Phi^{-2})} \int_0^{\pi} \exp\left(\frac{\cos(\hat{\phi}-\phi)}{\sigma_\Phi^2} \right) \mathrm{d}\hat{\phi}. \quad [34]$$

E. Interpretation of noisy strategies in the limit of large noise parameters. The noise parameters $\sigma_\phi, \sigma_\lambda, \sigma_\theta$ in our probabilistic models act approximately as the standard deviations for the von Mises distributions, whereas mean values of the perception noise are determined from experimental measurements of the predator state and mean values of response noise are based on the evasion strategies $\hat{\theta} = f^{(n)}(\hat{\phi},\hat{\psi},\hat{\lambda})$. To interpret our probabilistic models, it is important to first understand the limiting behavior of a von Mises distribution: when noise is zero, the von Mises distribution converges to a Dirac-delta function at the mean value; when the noise level is high, the distribution resembles a circular uniform distribution, where the PDF becomes $1/(2\pi)$ everywhere.

Limit of small perception and response noise At zero or small noise $\sigma_\Theta, \sigma_\Phi, \sigma_\Lambda \to 0$, the probabilistic models in Eqs. (29)–(33) behave similar to their deterministic counterparts in Eqs. (21)–(25) and predict any prey response other than $\hat{\mathbf{r}} = \mathbf{r} = f^{(n)}(\hat{\mathbf{s}} = \mathbf{s})$ with probability almost 0.

Limit of large response noise When the response noise is large $\sigma_\Theta \to \infty$, the probabilistic models predict any prey response with equal probability $1/(2\pi)$, irrespective of the strategy or the perception noise, thus all strategies become essentially equivalent to the pure-protean strategy.

Limit of large perception noise Consider no response noise, $\sigma_\Theta = 0$. Depending on the specific strategy, the two perception noise parameters σ_Φ and σ_Λ play different roles. For the contralateral strategy, high level of noise in predator angular position ϕ leads to equal probability of 0.5 in predicting $\pm\pi/2$, that is, the probabilistic model becomes that of a 'split' pure-protean strategy with equal probability of turning 90° to the left or right of the prey. For the other four models, large noise in predator heading σ_Φ works equivalently to the response noise σ_Θ, due to the symmetry in Eqs. (29)–(32); that is, in the limit of large perception noise $\sigma_\Phi \to \infty$, all four strategies become essentially equivalent to the pure-protean strategy.

Now assume that the perception noise in ϕ is negligible $\sigma_\Phi \to 0$, but the perception noise in λ can be large $\sigma_\Lambda \to \infty$. Since the antipodal and contralateral strategies are independent of λ, their behavior remains the same as that of their deterministic counterparts. For the distance-optimal strategy with $\chi > 0$, the noise in predator heading deviation σ_Λ itself does not lead to a pure-protean strategy, no matter how large it gets. Assuming zero noise in $\phi = \hat{\phi}$ and $\theta = \hat{\theta}$ and infinite noise in λ, the probability of any $\hat{\lambda}$ is equal to $1/(2\pi)$. Now, given a particular value of ϕ and considering the perceived stimulus is sinistral $\hat{\lambda} > 0$, according to Eq. (7), the evasion direction θ would lie in the range $[\phi + \pi - \chi, \phi + 2\pi - \chi]$ with a probability $1/\pi$. The probability of an escape direction in the complementary half-plane is zero. This is clearly illustrated in fig. S9A for the example case $\phi = \pi/2$. Similarly, given a particular value of ϕ and considering the perceived stimulus is dextral $\hat{\lambda} < 0$, the evasion direction θ would lie in the range $[\phi + \chi, \phi + \pi + \chi]$ with a probability $1/\pi$, as illustrated in fig. S9B. That is, the corresponding PDF gives twice the probability than average for the range of angles $[\phi + \pi - \chi, \phi + \pi + \chi]$ centered around the opposite direction of the predator angular position $(\phi + \pi)$, and zero probability for the range of angles $[\phi - \chi, \phi + \chi]$ centered around the predator position; see fig. S9C. That is, for a uniformly-distributed $\hat{\lambda}$, the PDF associated with the distance-optimal strategy is given by

$$p(\theta \in [\phi + \pi - \chi, \phi + \pi + \chi]) = \frac{1}{\pi}, \qquad p(\theta \in [\phi - \chi, \phi + \chi]) = 0, \qquad p(\theta \in [\phi + \chi, \phi + \pi - \chi] \cup [\phi + \pi + \chi, \phi - \chi]) = \frac{1}{2\pi}. \quad [35]$$

Note that both the range of angles for which the probability is zero and the range of angles for which the probability is $1/\pi$ span a sector of 2χ. For the orthogonal strategy $\chi = \pi/2$, the PDF gives probability of $1/\pi$ on the half-plane opposite to the predator position and zero probability on the half-plane on the same side as the predator. Mathematically, the orthogonal strategy with large noise on λ converges to the antipodal strategy with uniform noise spanning a range of π on either ϕ or θ.

5. Evaluation of Probabilistic Evasion Models with No Perception and Response Noise

A. Theoretical predictions without noise. The joint probability $p^{(n)}(\mathbf{s}, \mathbf{r})$ of following strategy n is given by Eq. (15) in the absence of perception and response noise. To obtain predictions of the evasion response given the recorded predatory stimuli, we use as input the distribution of the empirically-observed stimuli $\mathbf{s}_i$, and we construct a dataset $(\mathbf{s}_i, \mathbf{r}_i^{(n)})$, where $\mathbf{r}_i^{(n)} = \mathbf{f}^{(n)}(\mathbf{s}_i)$, for each strategy. Each dataset forms discrete samples from the corresponding joint probability $p^{(n)}(\mathbf{s}, \mathbf{r})$. We arrive at five datasets representing theoretical predictions of the prey's evasion behavior: one dataset for each evasion strategy. Our goal in this section is to evaluate these predictions in comparison to experimental observations.

B. Comparison to experimental observations using the Kullback-Leibler (K-L) Divergence. We use the Kullback-Leibler divergence $\mathcal{D}_{\mathrm{KL}}$ to quantify the goodness of fit between model prediction and experimental data. Namely, $\mathcal{D}_{\mathrm{KL}}$ measures the difference between the predicted probability distribution $p^{(n)}(\mathbf{s}, \mathbf{r})$ and the experimental probability distribution $p_o(\mathbf{s}, \mathbf{r})$. The expression of the Kullback-Leibler divergence $\mathcal{D}_{\mathrm{KL}}$ for two l-dimensional continuous distributions $p_o(\mathbf{x})$ and $p^{(n)}(\mathbf{x})$, where $\mathbf{x} \equiv (\mathbf{s}, \mathbf{r}) \in \mathbb{R}^l$, is given by

$$\mathcal{D}_{\mathrm{KL}}(p_o || p^{(n)}) = \int_{\mathbb{R}^l} p_o(\mathbf{x}) \log \frac{p_o(\mathbf{x})}{p^{(n)}(\mathbf{x})} \mathrm{d}\mathbf{x} \geq 0. \quad [36]$$

The K-L divergence is not a distance metric because it is not symmetric, namely, $\mathcal{D}_{\mathrm{KL}}(p_o || p^{(n)}) \neq \mathcal{D}_{\mathrm{KL}}(p^{(n)} || p_o)$. An evaluation of the K-L divergence from the expression in Eq. (36) is not directly feasible because we have no access to analytic expressions of $p_o(\mathbf{x})$ and $p^{(n)}(\mathbf{x})$. Instead, we have representative samples of these distributions for each predator speed V. We thus resort to numerical estimates of the K-L divergence.

C. Numerical estimates of the K-L divergence. To estimate the K-L divergence directly from samples, we implemented the numerical estimator proposed in (13). Given a dataset with N data points, we let $\mathbf{x}_i^{\mathrm{exp}} = \{\phi_i, \lambda_i, \theta_i\}$, where $i = 1, 2...N$, denote the samples from the experiment, and $\mathbf{x}_i^{\mathrm{model}(n)} = \{\phi_i, \lambda_i, \theta_i^{(n)}\}$, where $i = 1, 2...N$, denote the corresponding predictions from evasion model n. The discrete estimate of the K-L divergence is given by

$$\hat{\mathcal{D}}_{KL}(p_o || p^{(n)}) = \frac{3}{N} \sum_{i=1}^{N} \log \frac{k^{\mathrm{th}} \min_m ||\mathbf{x}_i^{\mathrm{exp}} - \mathbf{x}_m^{\mathrm{model}(n)}||}{k^{\mathrm{th}} \min_{m \neq i} ||\mathbf{x}_i^{\mathrm{exp}} - \mathbf{x}_m^{\mathrm{exp}}||} + \log \frac{N}{N-1}, \quad [37]$$

where k^{th} min means the k^{th} smallest value of the argument and k is a discretization parameter to be tuned.

The parameter k shows up because the estimator follows the concept of k-nearest-neighbor for density estimate. In density estimation the choice of k is related to the sample size N. Conceptually, it is guaranteed that, given a sufficiently large sample size, the estimate converges to the actual value of the K-L divergence with $k = 1$ (13). Practically, the choice of k is problem-dependent: a smaller k means that noise has a higher influence on the results while a larger k requires larger sample size for the estimate to converge to the accurate value. Here, we tested the effect of different k by computing the K-L divergence estimate from the predictions of distance-optimal strategy to all four experimental datasets. For $k = 1$, the noise had an observable influence on the result. Through trial-and-error, we found that $k = 5$ worked well. Hereafter, we set $k = 5$ for all calculation of the estimates.

In fig. S10 we demonstrate the method in Eq. (37) for estimating the K-L divergence to assess the antipodal strategy using the fast-predator dataset $V = 20$ cm·s^{-1} and $N_{\text{fast}} = 215$. For clarity of exposition, we omitted the variable λ from the graphical demonstration in fig. S10, but λ is included in the numerical estimation of the K-L divergence. In fig. S10A, we plot the density of occurrence of a joint θ and ϕ as a bivariate histogram: the histrogram on the left is constructed from experimental data, and the histogram on the right is constructed based on the predictions $(\phi_i, \theta_i^{(4)} = f^{(4)}(\phi_i))$, $i = 1, \ldots, N_{\text{fast}}$, of the antipodal strategy. In fig. S10B, the experimental data points (ϕ_i, θ_i) are plotted on the (ϕ, θ) plane as circles and the predictions $(\phi_i, \theta_i^{(4)})$ from the antipodal strategy are plotted as blue diamonds. Considering an experimental data point, say the data point highlighted in green in the zoomed-in view, our algorithm identifies the five nearest neighbors from the experimental dataset and the five nearest neighbors from the predicted results, marked in gray and black, respectively. The distances between the green circle and its fifth neighbors from both groups give us the values of $5^{\text{th}} \min_m ||\mathbf{x}_i^{\text{exp}} - \mathbf{x}_m^{\text{model}(n)}||$ and $5^{\text{th}} \min_{m \neq i} ||\mathbf{x}_i^{\text{exp}} - \mathbf{x}_m^{\text{exp}}||$ in Eq. (37).

Following the same process demonstrated here, we applied the K-L divergence estimation to all four datasets with predictions from all five evasion strategies without noise, and arrived at the twenty estimate values shown in fig.3A of the main document. It is important to note that the values of the K-L divergence estimate obtained from different datasets are not comparable to one another due to the difference in sample size. Only within one dataset the results from different strategies can be compared and ranked; see main document.

6. Evaluation of Probabilistic Evasion Models with Perception and Response Noise

In section 4, we derived five expressions for the conditional probability $p^{(n)}(\theta|(\phi, \lambda)$, for $n = 1, \ldots, 5$, listed in Eqs. (29–33). In each case, perception and response noise are accounted for using the von Mises distributions in Eq. (18) and Eq. (19), thus introducing three unknown noise parameters $\boldsymbol{\sigma} = \{\sigma_\Phi, \sigma_\Lambda, \sigma_\Theta\}$ that describe the dispersion in perceived stimulus around the actual state of the predator and the dispersion in actual response around the desired response. Our goal in this section is to find optimal values for $\boldsymbol{\sigma} = \{\sigma_\Phi, \sigma_\Lambda, \sigma_\Theta\}$ given experimental data, then to rank the optimized probabilistic models based on their parsimony and closeness to experimental observations.

A. Optimization of noise parameters. We used maximum likelihood estimation to optimize each stochastic strategy based on the experimental data. The likelihood function for a candidate model n can be written as (14)

$$\mathcal{L}(\boldsymbol{\sigma}|(\theta, \phi, \lambda); n) = \prod_i p^{(n)}(\theta_i, \phi_i, \lambda_i; \boldsymbol{\sigma}). \quad [38]$$

Here i is the index of data point. We seek to calculate the set of noise parameters $\boldsymbol{\sigma} = \{\sigma_\Phi, \sigma_\Lambda, \sigma_\Theta\}$ that maximizes the total likelihood, or equivalently, the sum of its logarithm

$$\boldsymbol{\sigma}^* = \arg\max_{\boldsymbol{\sigma}} \ln \mathcal{L}(\boldsymbol{\sigma}|(\theta, \phi, \lambda); n). \quad [39]$$

Since the distribution $p_o(\mathbf{s})$ of predator states does not depend on the noise parameters, we can thus ignore this part from the expression for the joint probability $p^{(n)}(\mathbf{s}, \mathbf{r}) = p^{(n)}(\mathbf{r}|\mathbf{s})p_o(\mathbf{s})$ and simplify the optimization problem to maximize the noise parameters using the modified likelihood function

$$\mathcal{L}(\boldsymbol{\sigma}|(\theta|(\phi, \lambda)); n) = \prod_i p^{(n)}(\theta_i|(\phi_i, \lambda_i); \boldsymbol{\sigma}). \quad [40]$$

The equivalence between the original and simplified maximization objectives is straightforward,

$$\begin{aligned} \boldsymbol{\sigma}^* &= \arg\max_{\boldsymbol{\sigma}} \sum_i \ln p^{(n)}(\theta_i, \phi_i, \lambda_i; \boldsymbol{\sigma}), \\ &= \arg\max_{\boldsymbol{\sigma}} \sum_i \left[\ln p^{(n)}(\theta_i|(\phi_i, \lambda_i); \boldsymbol{\sigma}) + \ln p_o(\phi_i, \lambda_i)\right], \\ &= \arg\max_{\boldsymbol{\sigma}} \sum_i \ln p^{(n)}(\theta_i|(\phi_i, \lambda_i); \boldsymbol{\sigma}). \end{aligned} \quad [41]$$

Yusheng Jiao[1], Brendan Colvert[2], Yi Man[1], Matthew McHenry[3], Eva Kanso[1‡‡]

This optimization problem can be rewritten in terms of the negative log-likelihood function

$$\mathrm{NLL} = -\ln \mathcal{L}(\boldsymbol{\sigma}|(\theta|(\phi,\lambda)); n) = -\sum_i \ln p^{(n)}(\theta_i|(\phi_i,\lambda_i);\boldsymbol{\sigma}), \quad [42]$$

such that the task of calculating $\boldsymbol{\sigma}^* = \{\sigma_\Phi^*, \sigma_\Lambda^*, \sigma_\Theta^*\}$ becomes that of minimizing the NLL,

$$\boldsymbol{\sigma}^* = \arg\min_{\boldsymbol{\sigma}} \mathrm{NLL}. \quad [43]$$

We used *fmincon* function in MATLAB to numerically find the optimal noise parameters and the likelihood of the optimized probabilistic models. Specifically, for each dataset (slow, mid, fast, and all combined), we computed $\boldsymbol{\sigma}^*$ for each evasion strategy within the range of $(0, \pi)$.

We illustrate the optimization process in fig. S11 for all five evasion strategies. In fig. S11A, using the combined dataset for all predator speeds, we varied the noise parameters in perception and motor response and we plotted the corresponding values of NLL (Eq. (42)) as a colormap on the $(\sigma_\Lambda, \sigma_\Theta)$ or $(\sigma_\Phi, \sigma_\Theta)$ plane. Because of the symmetry in Eq. (29)–Eq. (31) where σ_Φ and σ_Θ have the same impact on the likelihood for the distance-optimal, orthogonal, and parallel strategies, we show σ_Θ and σ_Φ on the same axis. We verified that imposing this condition has a negligible effect on the optimized parameters and likelihoods, whilst improves the optimization speed and stability. For every probabilistic evasion model, we select the combination of noise parameters that results in the lowest NLL (the darkest spot in the colormap) as the optimal noise parameters for that model; meanwhile, the lowest NLL value is used to represent the performance of the strategy given the dataset.

In fig. S11B, we show the optimized noise parameters obtained from the optimization procedure using *fmincon* in MATLAB. The evaluation results presented in Fig. 3 of the main text are based on this optimization procedure. We used the open-loop evaluation in fig. S11A to double-check our optimization procedure and to verify that our MATLAB code converges to the correct optima. We found excellent agreement, up to machine precision, between the results of the open-loop approach in fig. S11A and the optimization routine in fig. S11B.

B. Akaike Information Criterion. For model selection, we used the Akaike Information Criterion (AIC) to rank the optimized models and seek a parsimonious model for the experimental data. The AIC score is defined as (15)

$$\mathrm{AIC} = 2K - 2\log \mathcal{L}(\boldsymbol{\sigma}|(\theta|(\phi,\lambda)); n). \quad [44]$$

It considers both the goodness of fit represented by the likelihood function, and the complexity of the model. Here, M denotes the number of estimated parameters used in the model. The criterion favors a simpler model if two models have the same ability to explain the data.

In our models, the coefficient for each perception variable is pre-defined and the noise parameters $\boldsymbol{\sigma} = \{\sigma_\Phi, \sigma_\Lambda, \sigma_\Theta\}$ are estimated using maximum likelihood estimation, so we have $K = 3$ for distance-optimal strategy, orthogonal strategy and parallel strategy, and $K = 2$ for antipodal strategy and contralateral strategy.

C. Bootstrapping. To ensure accuracy and assess the variability in the result, we used the bootstrapping method in evaluating the probabilistic models. Starting from one dataset and one evasion model, say the dataset for the fast predator and antipodal strategy, we constructed 200 different datasets of equal size to the original dataset by random sampling with repetition. For each newly constructed dataset, we solved for the optimal noise parameters $\boldsymbol{\sigma}^* = \{\sigma_\Phi^*, \sigma_\Lambda^*, \sigma_\Theta^*\}$ by following the same maximum likelihood estimation process introduced above. Then we calculated 200 corresponding AIC scores and used the mean and the standard deviation to show the performance of the antipodal model on the fast predator dataset. We repeated this process for all five evasion strategies and all four datasets, thus solving a total of $200 \times 5 \times 4 = 4,000$ optimization problems.

D. Model performance based on AIC. For the purpose of ranking the probabilistic models by the order of their fit to experimental data, we use the AIC scores and report the difference $\Delta\mathrm{AIC} = \mathrm{AIC} - \mathrm{AIC}_{\min}$ between the average AIC and minimum AIC among all models for the same dataset. Evaluation results are shown in Fig. 4B and Fig. 6B of the main text.

In addition, we quantitatively interpret ΔAIC by computing the quantity $\exp(-\Delta\mathrm{AIC}/2)$, which lies in the interval $[0, 1]$ because ΔAIC is non-negative. This quantity represents the relative *strength of evidence* for each model. Empirically, ΔAIC less than 2 suggests the model is comparable to the best model, and ΔAIC over 10 suggests the model can be omitted from further consideration (16). In fig. S12, we plot $\exp(-\Delta\mathrm{AIC}/2)$ as a function of ΔAIC (black lines), and plot the ΔAICs for each evasion model in each dataset from Fig. 4B of the main text. Instead of considering one ΔAICs as commonly done, We retain the entire distribution of AIC scores obtained based on bootstrapping. Focusing only on the mean value of each distribution, all models, save the orthogonal strategy, have $\Delta\mathrm{AIC} \gtrsim 10$ in all four datasets. This indicates that the orthogonal strategy consistently explains the experimental data substantially better than all the other models. Examining the entire distribution of AIC scores, the AIC scores of the distance-optimal strategy always overlaps with the AIC scores of the orthogonal strategy within one standard deviation. The AIC scores of antipodal strategy also overlaps with the orthogonal strategy in all but the fast dataset. The AIC scores of parallel strategy and contralateral strategy are always substantially higher than the AIC scores of the orthogonal strategy in all datasets, showing that they clearly are not supported by the experimental data. Therefore, we exclude the parallel and contralateral strategies from further analysis and do not show their results in fig. 6B of the main text, where we incorporate the physical constraint in turning.

E. Levels of perception and motor noise can be used to further discriminate between evasion models. From a biological perspective: if a sensory cue is too noisy, it is not wise to make evasion decisions based on that sensory cue; it is better to rely on robust sensory cues. In turn, an evasion strategy is reasonable if the predicted mean of prey response $\hat{\theta}$ is close to the most likely value of θ given the predator state (ϕ, λ), such that the response noise σ_Θ merely adds dispersion to the model prediction $\hat{\theta}$. Mathematically, if a noise parameter needs to be very large to best fit the data, it probably indicates that the corresponding strategy is not a reasonable evasion strategy.

In most evasion models, a clear global minimum at moderate noise levels seems to exist (fig. S11). However, in the distance-optimal strategy, the NLL changes slowly along the σ_Λ axis and the optimal noise σ_Λ^* occasionally converges to π, implying that the optimal noise level in λ is not unique. Based on the above discussion, this suggests the predator heading deviation λ is not very helpful. In comparison, the NLL always converges to one minimum at relatively small σ_Λ^* for the orthogonal strategy, suggesting that λ plays a more reliable role in the orthogonal strategy.

7. Further Analysis of the Distance-Optimal Strategy

The distance-optimal strategy has an important parameter $\chi = \cos^{-1}(U/V)$ that depends on the ratio between prey speed U and predator speed V and that takes values between 0° and 90° (fig. S13). In the case of a very fast predator $U/V \to 0$, we have $\chi = 90^\circ$ and the distance-optimal strategy converges to the orthogonal strategy. At the other extreme, when the prey is faster than the predator, the distance-optimal strategy converges to the parallel strategy.

One challenge in the assessment of distance-optimal strategy is that we do not know whether prey sense the speed ratio U/V in real time and adjust the angle χ accordingly, or whether the prey responds with a prior knowledge of χ that is hardwired in the prey. The predator speed was controlled at three constant values $V = 2, 11, 20$ cm s^{-1}, whereas the zebrafish larvae used in all experiments were almost identical, implying that the speed ratio U/V could vary drastically between evasion instances: for the fast predator, it could be up to 10 fold that of the slow predator. If the prey were to sense and use the real-time speed ratio to implement the distance-optimal strategy, we would expect the best performance of the distance optimal strategy to be at different value of χ depending on the dataset (slow, mid, or fast predator). In this section, we test this proposition by treating χ as a model parameter and evaluating the performance of the distance optimal strategy as a function of $\chi \in [0, 90^\circ]$.

A. Side note. Another major consideration in evaluating the distance optimal strategy is that the actual speed ratio U/V is ambiguous during the C-start response: the prey starts from almost zero speed and abruptly accelerates while its orientation also changes rapidly. This is different from the assumption adopted in mathematically deriving the distance-optimal model (10), where the two involved parties (predator and prey) swim at constant speed and constant orientation. Although in general, the predator has a higher free-swimming speed, the prey can reach an exceptionally high instantaneous speed during the short time of a C-start response (17), and it is unclear which speed should be used in optimizing the minimal distance in Eq. (7). In the evaluation presented in fig. 3 of the main text, we chose $U/V = 1/2$ and $\chi = 60^\circ$.

B. Performance of optimized strategy as a function of speed ratio between predator and prey. We evaluated the distance-optimal strategy for each of the four datasets (slow, mid, fast and all data combined) as a function of $\chi \in [0, 90^\circ]$. In fig. S14 we show the change in evaluation results of distance-optimal strategy as we increase χ incrementally by 5° from 0° to 90°. Specifically, in fig. S14A, we show the K-L divergence estimate for the probabilistic models without noise and in fig. S14B, we give the NLL values for the model with optimized noise, where for each value of χ, we performed the optimization 200 times for each dataset, using the bootstrapping described in section C. That is, we performed a total of $19 \times 200 \times 4 = 15,200$ optimizations.

If the prey were to sense and use the real-time speed ratio to implement the distance-optimal strategy, we would expect the lowest K-L divergence estimate and NLL to appear at different values of χ in different datasets. Instead, in fig. S14A we see the K-L divergence estimates decrease as χ increases and reach the minimum near $\chi = 75^\circ$ for all four datasets, and they stay relatively low as χ increases further to 90°. In fig. S14B, we see a similar trend that the NLLs for all datasets drop as χ increases until they reach a minimum at or close to $\chi = 90^\circ$. These results suggest that prey do not calculate the angle χ by sensing the real-time predator speed.

C. Re-examining optimal noise levels as a function of speed ratio between predator and prey. The same conclusion can be strengthened by examining the values of the optimized noise parameters. At χ near zero, the optimizer avoids large values of σ_Λ because they cause the distance-optimal strategy to reduce to the pure-protean strategy, for which NLL$= -N\ln(1/2\pi) = 1.8379N$ (where N is the sample size). Thus, the optimizer is challenged to find noise values for which NLL $\leq 1.8379N$. As χ increases, the distance-optimal strategy resembles a variant of the antipodal strategy, which has lower NLLs than the pure-protean strategy even at large σ_Λ. In the range $20^\circ \lesssim \chi \lesssim 75^\circ$, the optimizer often chooses the largest possible value of $\sigma_\Lambda^* \to \pi$ to produce the optimized NLLs. Taken together, the larger values of NLL at $\chi \lesssim 20^\circ$ and the large values of optimal σ_Λ in the range $20^\circ \lesssim \chi \lesssim 75^\circ$ indicate that λ is not an effective sensory cue for $\chi \lesssim 75^\circ$. As χ gets closer to 90°, lower values of σ_Λ lead to lower NLL values (see fig. S14B). Therefore, the prey is unlikely to use the distance-optimal strategy at small values of χ; this strategy is best used for $\chi \to 90^\circ$, thus favoring the orthogonal strategy.

8. Fish Evasion Mechanics: C-start Maneuver

Larval zebrafish execute a "C-start" or "fast start" to reorient themselves rapidly in response to predators (18). The prey bends its body into a "C" shape from a straight configuration within the order of 10ms (17). Computational work found that pressure-driven action-reaction forces are dominant in C-start maneuvers, whereas vorticity generated during this short process has negligible effects on fish motion (19). Thus, to investigate the evasion mechanics, we adopt a simplified 3-link fish model in potential flow (20, 21) which ignores the effects of viscosity and vortex shedding, but accounts for reactive hydrodynamic forces through the added mass effect.

A. Fish kinematics. To describe planar motions of the three-link fish, we let $\mathbf{x} = (x, y)$ denote the position of the center of mass G of the middle link, and β denote the orientation of the fish relative to a fixed inertial frame, here taken to be the angle between the x-axis and the major axis of symmetry of the middle link. Let α_1 and α_2 be the rotation angles of the front link relative to the middle link and the middle link relative to the rear link; that is to say, (α_1, α_2) represents the shape of the three-link fish.

It is convenient for the following development to introduce a body-fixed frame $(\mathbf{b}_1, \mathbf{b}_2)$, attached at G and co-rotating with the middle link. This body-fixed frame is related to the inertial frame $(\mathbf{e}_1, \mathbf{e}_2)$ via a rigid-body rotation such that $\mathbf{e}_1 = \cos\beta\mathbf{b}_1 - \sin\beta\mathbf{b}_2$ and $\mathbf{e}_2 = \sin\beta\mathbf{b}_1 + \cos\beta\mathbf{b}_2$. The velocity $(\dot{x}, \dot{y})$ of the center of mass of the middle link, when expressed in the body-fixed frame, is given by

$$\mathbf{v} = v_1\mathbf{b}_1 + v_2\mathbf{b}_2 = (\dot{x}\cos\beta + \dot{y}\sin\beta)\mathbf{b}_1 + (-\dot{x}\sin\beta + \dot{y}\cos\beta)\mathbf{b}_2 \quad [45]$$

Assuming all three links are made of identical ellipsoids of major axis a, and minor axes $b = 0.2a, c = a$, the velocities of the centers of mass G_1 and G_2 of the front and rear link, expressed in the body-fixed frame of the middle link, are given by ($i = 1, 2$, denote the front and rear links, respectively)

$$\mathbf{v}_i = (v_1 \mp a\dot{\alpha}_i \sin\alpha_i)\mathbf{b}_1 + (v_2 \pm a\dot{\beta} + a\dot{\alpha}_i\cos\alpha_i)\mathbf{b}_2. \quad [46]$$

The angular velocities of the middle, front, and rear links are given by $\dot{\beta}$, $\dot{\beta} + \dot{\alpha}_1$, and $\dot{\beta} - \dot{\alpha}_2$ respectively.

B. Kinetic energy of the three-link fish. The kinetic energy of the fish is given by

$$T_{\text{prey}} = \frac{1}{2}m_s\mathbf{v}\cdot\mathbf{v} + \frac{1}{2}J_s\dot{\beta}^2 + \frac{1}{2}\sum_{i=1,2}\left[m_s\mathbf{v}_i\cdot\mathbf{v}_i + J_s(\dot{\beta} \pm \dot{\alpha}_i)^2\right] \quad [47]$$

where $m_s = \frac{4}{3}abc\rho_s$ and $J_s = \frac{1}{5}(a^2 + b^2)m_s$ are the mass and moment of inertia of each solid link with ρ_s the density of the links.

C. Kinetic energy of the fluid. The three-link fish is submerged in an unbounded domain of incompressible and irrotational fluid, such that the fluid velocity $\mathbf{u} = \nabla\Phi$ can be expressed as the gradient of a potential function Φ. It is a standard result in potential flow theory that the kinetic energy of the fluid can be expressed in terms of the variables of the submerged solid (20–22). In the case of a single ellipsoid, the kinetic energy of the fluid is given by $T_{\text{fluid}} = [(m_{1a}v_1^2 + m_{2a}v_2^2) + J\dot{\beta}^2]/2$, where m_{1a}, m_{2a} and J_a are the added mass and the added moment of inertia due to the presence of the fluid, expressed in a body-fixed frame that coincides with the major and minor axes of the ellipsoid. The added mass and moment depend on the geometric shape a, b, c of the submerged ellipsoid and the fluid density ρ_{f} (23). For a non-spherical body, the added masses m_{1a}, m_{2a} depend on the direction of motion: the added mass is larger when moving in the direction of the minor axis of symmetry of the ellipsoid, that is to say, in the transverse direction; hence $m_1 \le m_2$. For the three-link fish, the kinetic energy of the fluid is of the form

$$\begin{aligned} T_{\text{fluid}} =& \frac{1}{2}m_{1a}v_1^2 + \frac{1}{2}m_{2a}v_2^2 + \frac{1}{2}J_a\dot{\beta}^2 + \frac{1}{2}\sum_i J_a\left(\dot{\beta} \pm \alpha_i\right)^2 \\ &+ \frac{1}{2}\sum_i m_{1a}\left(v_1\cos\alpha_i \pm v_2\sin\alpha_i + a\dot{\beta}\sin\alpha_i\right)^2 \\ &+ \frac{1}{2}\sum_i m_{2a}\left(\mp v_1\sin\alpha_i + v_2\cos\alpha_i \pm a\dot{\beta}\cos\alpha_i \pm a\dot{\beta} + a\dot{\alpha}_i\right)^2. \end{aligned} \quad [48]$$

D. Kinetic energy of the fish-fluid system. The kinetic energy of the fish-fluid system is obtained by taking the sum of Eq. (47) and Eq. (48), which can be expressed in matrix form as follows:

$$T = T_{\text{prey}} + T_{\text{fluid}} = \frac{1}{2}\begin{bmatrix} v_1 \\ v_2 \\ \dot{\beta} \\ \dot{\alpha}_1 \\ \dot{\alpha}_2 \end{bmatrix}^{\mathsf{T}} \left[\begin{array}{cc:cc} & & & \\ & \mathbb{I}_{\text{lock}} & & \mathbb{I}_{\text{couple}} \\ & & & \\ \hdashline & & & \\ & \mathbb{I}_{\text{couple}}^{\mathsf{T}} & & \mathbb{I}_{\text{shape}} \\ & & & \end{array}\right] \begin{bmatrix} v_1 \\ v_2 \\ \dot{\beta} \\ \dot{\alpha}_1 \\ \dot{\alpha}_2 \end{bmatrix}, \quad [49]$$

Here, $\mathbb{I}_{\text{shape}}$ is a 2×2 matrix associated with shape deformation,

$$\mathbb{I}_{\text{shape}} = \begin{bmatrix} J + m_2 a^2 & 0 \\ 0 & J + m_2 a^2 \end{bmatrix}. \tag{50}$$

where $m_1 = m_s + m_{1a}$, $m_2 = m_s + m_{2a}$, and $J = J_s + J_a$. $\mathbb{I}_{\text{couple}}$ is a 3×2 matrix that couples rigid body motion with shape deformation,

$$\mathbb{I}_{\text{couple}} = \begin{bmatrix} -m_2 a \sin\alpha_1 & m_2 a \sin\alpha_2 \\ m_2 a \cos\alpha_1 & m_2 a \cos\alpha_2 \\ J + m_2 a^2(1+\cos\alpha_1) & -J - m_2 a^2(1+\cos\alpha_2) \end{bmatrix}. \tag{51}$$

Finally, $\mathbb{I}_{\text{lock}}$ is a 3×3 locked mass matrix, function of α_1 and α_2,

$$\mathbb{I}_{\text{lock}} = \left[\begin{array}{cc:c} \multicolumn{2}{c:}{\multirow{2}{*}{$\mathbb{M}$}} & \multirow{2}{*}{$\mathbb{D}$} \\ \multicolumn{2}{c:}{} & \\ \hdashline \mathbb{D}^{\mathsf{T}} & & \mathbb{J} \end{array}\right], \tag{52}$$

where $\mathbb{M}$ is a 2×2 mass matrix given by

$$\mathbb{M} = \begin{bmatrix} m_1\left(1 + \sum_i \cos^2\alpha_i\right) + m_2 \sum_i \sin^2\alpha_i & \frac{1}{2}(m_1 - m_2)(\sin 2\alpha_1 - \sin 2\alpha_2) \\ \frac{1}{2}(m_1 - m_2)(\sin 2\alpha_1 - \sin 2\alpha_2) & m_2\left(1 + \sum_i \cos^2\alpha_i\right) + m_1 \sum_i \sin^2\alpha_i \end{bmatrix}, \tag{53}$$

$\mathbb{J}$ is a moment-of-inertia scalar given by

$$\mathbb{J} = 3J + m_1 a^2 \sum_i \sin^2\alpha_i + m_2 a^2 \sum_i (1+\cos\alpha_i)^2, \tag{54}$$

and $\mathbb{D}$ couples the translational and rotational motion of the fish. For a single ellipsoid, $\mathbb{D}$ is identically zero. For the three-link fish, it is given by

$$\mathbb{D} = \begin{bmatrix} \frac{1}{2}(m_1 - m_2) a \sum_i \sin 2\alpha_i - m_2 a \sum_i \sin\alpha_i \\ \frac{1}{2}(m_1 - m_2) a (\cos 2\alpha_2 - \cos 2\alpha_1) + m_2 a (\cos\alpha_1 - \cos\alpha_2) \end{bmatrix}. \tag{55}$$

E. Local connection matrix. We define a *local connection matrix* $\mathbb{A} = -\mathbb{I}_{\text{lock}}^{-1}\mathbb{I}_{\text{couple}}$, where the entries A_{xi}, A_{yi}, A_i $(i = 1, 2)$ of the connection matrix $\mathbb{A}$ are nonlinear functions of α_1, α_2, (see (20, 21, 24, 25))

$$\mathbb{A} = -\mathbb{I}_{\text{lock}}^{-1}\mathbb{I}_{\text{couple}} = \begin{bmatrix} A_{x1} & A_{x2} \\ A_{y1} & A_{y2} \\ A_1 & A_2 \end{bmatrix}. \tag{56}$$

The local connection matrix $\mathbb{A}$ is a function of the variables (α_1, α_2) that define the shape of the three-link fish. Specifically, $\mathbb{A}$ defines three vector fields $\mathbf{A}_x \equiv (A_{x1}, A_{x2})$, $\mathbf{A}_y \equiv (A_{y1}, A_{y2})$, and $\mathbf{A} \equiv (A_1, A_2)$ over the shape space (α_1, α_2) as shown in fig. S15(A).

F. Equations of motion. In the absence of external forces and torques, the equations of motion of the fish are obtained from the conservation of linear and angular momentum. Starting from rest, at zero total momentum, one gets

$$\begin{bmatrix} \cos\beta & \sin\beta & 0 \\ -\sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\beta} \end{bmatrix} = \begin{bmatrix} A_{x1} & A_{x2} \\ A_{y1} & A_{y2} \\ A_1 & A_2 \end{bmatrix} \begin{bmatrix} \dot{\alpha}_1 \\ \dot{\alpha}_2 \end{bmatrix}. \tag{57}$$

Eq. (57) provides a set of first-order differential equations that can be integrated numerically to obtain the state (x, y, β) for prescribed cyclic shape changes in the (α_1, α_2) plane. Translations motions (x, y) and rotational motions β of the fish are dictated by the connection matrix $\mathbb{A}$. Translations are coupled to rotations via $\mathbb{A}$.

Rotations β, are directly proportional to the line integral of the vector field $\mathbf{A} = (A_1, A_2)$. Considering cyclic shape changes in the (α_1, α_2) plane and using Green's Theorem, we get, considering the last row of Eq. (57),

$$\beta(T) - \beta(0) = \oint_C \mathrm{d}\beta = \oint_C (A_1 \mathrm{d}\alpha_1 + A_2 \mathrm{d}\alpha_2) = \iint_S \mathrm{curl}_2(\mathbf{A}) \mathrm{d}\alpha_1 \mathrm{d}\alpha_2 = \iint_S \left(\frac{\partial A_2}{\partial \alpha_1} - \frac{\partial A_1}{\partial \alpha_2}\right) \mathrm{d}\alpha_1 \mathrm{d}\alpha_2. \tag{58}$$

Here, T is the time required to go around the closed trajectory in the shape space once; note that a re-scaling of time does not affect how the motion of the fish depends on (α_1, α_2), only the speed at which the fish completes these cyclic shape changes.

G. Turning versus swimming gaits. In fig. S15A, we show the three scalar fields defined by $\text{curl}_2\mathbf{A}_x$, $\text{curl}_2\mathbf{A}_y$, and $\text{curl}_2\mathbf{A}$, where the normalized curls are visualized as a colormap (from blue to orange) over the (α_1, α_2) plane. The scalar fields $\text{curl}_2\mathbf{A}_x$, $\text{curl}_2\mathbf{A}_y$, and $\text{curl}_2\mathbf{A}$ are informative of the net motions (x, y, β) of the fish, but whereas rotations β are directly proportional to the area integral of $\text{curl}_2\mathbf{A}$, translational motions (x, y) are not directly proportional to the area integrals of $\text{curl}_2\mathbf{A}_x$ and $\text{curl}_2\mathbf{A}_y$, due to their coupling with rotations β, as evident from Eq. (57).

To illustrate the utility of the scalar fields $\text{curl}_2\mathbf{A}_x$, $\text{curl}_2\mathbf{A}_y$, and $\text{curl}_2\mathbf{A}$ shown in fig. S15A, we show two examples of cyclic shape changes depicted in black and red. Mathematically, these trajectories are closed circles of radius A such that

$$\begin{aligned} \text{Turning gait:} \quad & \alpha_1(t) = \frac{1}{\sqrt{2}}\left[A\left(1-\cos\frac{2\pi t}{T}\right) + A\sin\frac{2\pi t}{T}\right], \quad \alpha_2(t) = \frac{1}{\sqrt{2}}\left[A\left(1-\cos\frac{2\pi t}{T}\right) - A\sin\frac{2\pi t}{T}\right], \\ \text{Swimming gait:} \quad & \alpha_1(t) = \frac{1}{\sqrt{2}}\left[A\cos\frac{2\pi t}{T} - A\sin\frac{2\pi t}{T}\right], \quad \alpha_2(t) = \frac{1}{\sqrt{2}}\left[A\cos\frac{2\pi t}{T} + A\sin\frac{2\pi t}{T}\right]. \end{aligned} \tag{59}$$

For both examples, we chose $A = \pi/3$. For the shape deformations following the red trajectory, the corresponding maximum bending angle is given by $\alpha_{\max} = \sqrt{2}A$. At $t = 0$, the fish is straight and $\alpha_1 = \alpha_2 = 0$ (point T1 in fig. S15A). For $0 < t < T/2$, the fish bends its body to one side; α_1 increases first and α_2 catches up till they meet at the same value $\alpha_1 = \alpha_2$, where their sum is maximal (point T2 in fig. S15A). For $T/2 < t < T$, the fish unfurls its body; α_1 and α_2 decrease back to zero. At $t = T$, the fish recovers to a straight line (point T4 in fig. S15A). In the swimming (black curve in fig. S15A), the three-link fish starts in a deformed shape (point S1 in in fig. S15A) and undulates its body cyclically to both sides.

For each trajectory, we numerically integrate Eq. (57) to obtain the translational motion $x(t)$ and $y(t)$ and orientation $\beta(t)$ of the fish model. Note that $\beta(t)$ can be solved analytically following the expression in Eq. (58) and substituted back into Eq. (57) to numerically solve for $x(t)$ and $y(t)$. A fish changing its shape following the black curve in fig. S15A will undergo zero net rotations and zero net displacement in the y-direction, but it will swim forward in the x-direction as shown in fig. S15B. A fish following the red line in fig. S15A will undergo a positive net rotation and a negligible displacement in both the x- and y-directions as shown in fig. S15C. We emphasize that time is inconsequential in these motions: a re-scaling of time neither affects the fish shape deformations (α_1, α_2) nor its locomotion (x, y, β), only the speed at which the fish follows these trajectories.

9. Comparison of C-start Model to Experimental Data

A. Experimental measurements of C-start kinematics. In (26), the authors measured the excursions of the fish body during an evasive response. They divided the fish body into 4 sections of equal length and recorded the orientation of each section relative the initial orientation of the fish as a function of time. To use these experimental measurements in our 3-link fish model, we averaged the excursion of the middle two sections and considered that to be the change in orientation θ. We computed the difference in excursion between the first section and θ and the difference in excursion between θ and the last section, and we considered these differences to be the shape angles α_1, α_2 in the three-link model. The shape trajectory (α_1, α_2) based on the experimental measurements is shown in S15A in grey dots. The values of θ from experimental measurements is shown in Fig. 4 of the main text.

B. Comparison between experimental measurements and three-link fish model. We fitted the experimental data (α_1, α_2) with a third-order Fourier series; results shown in light blue line in S15A. We then used this fitted trajectory (α_1, α_2) as input to the 3-link fish model in Eq. (58) and compared the results obtained from the three-link model to the experimental values of θ; results are shown in Fig. 4 of the main text.

Remarkably, the experimentally-constructed shape motion (α_1, α_2) (light blue line in fig. S15A) resembles the model-based shape trajectories (red and black circles in fig. S15A) if we start from the origin, follow the red line, then join the black line at the location where the red and black trajectories intersect, indicated by a green star in S15A.

We numerically integrated Eq. (57) using the synthetic shape motion (with α_1 and α_2 tracing the red then black trajectories in S15A) and the experimentally-obtained C-start maneuver (with α_1 and α_2 tracing the light blue trajectory in S15A). Note that for the synthetic shape changes (tracing the red then black lines in fig. S15A), we chose the time scale to match the time scale in the experimentally-observed C-start (blue line in fig. S15A). The fish first turns then begins to swim forward in a post C-start direction. Snapshots of the fish body shape are shown in fig. S15D and time evolution of x, y, and β are shown in fig. S15E; the displacements x and y are shown in solid and dashed lines in the top panel of S15E, with no corresponding experimental data available for comparison, while the rotational motion β is shown in the bottom panel of fig. S15E. The experimentally-obtained change in heading θ (grey dots) is superimposed for comparison. Remarkably, the rotational motion obtained based on the three-link fish model accurately matches the body rotation observed experimentally in the C-start response.

C. Mechanical constraints on prey reorientation. Inspired by the fact that the synthetic turning gait (red curve in fig. S15A) accurately predicts the rotational response observed experimentally during a C-start maneuver, we use the three-link model to probe the mechanical constraints on the reorientation angle $\theta = \beta(T) - \beta(0)$ during evasion. To this end, we construct a family of elliptic trajectories in the shape space α_1, α_2 that are adapted from Eq. (59) to produce net rotations,

$$\alpha_1(t) = \frac{1}{\sqrt{2}}\left[A\left(1-\cos\frac{2\pi t}{T}\right) + B\sin\frac{2\pi t}{T}\right], \qquad \alpha_2(t) = \frac{1}{\sqrt{2}}\left[A\left(1-\cos\frac{2\pi t}{T}\right) - B\sin\frac{2\pi t}{T}\right]. \tag{60}$$

The corresponding maximum bending angle is given by $\alpha_{\max} = \sqrt{2}A$. We varies $\alpha_{\max}$ from 0 to 180°. For each $\alpha_{\max}$, we solved for the value of B that maximizes the response angle θ. Examples of the optimized trajectories are given in fig. S16A and B. In fig. S16A, we show $\text{curl}_2\mathbf{A}$ for a three-link fish of negligible mass, whereas in fig. S16B, we show $\text{curl}_2\mathbf{A}$ for a neutrally-buoyant three-link fish. The difference between fig. S16A and B stems from how $\mathbb{A}$ in Eq. (56), consequently $\mathbf{A}$, changes with fish body density ρ. The optimized trajectories enclose as much as possible of the orange regions where $\text{curl}_2\mathbf{A}$ is strictly positive. The net reorientation θ resulting from tracing each of cycle of shape deformations is shown in fig. S16C as a function of the maximum bending angle $\alpha_{\max}$, thus establishing a mapping from the maximum bending angle $\alpha_{\max}$ to the response angle θ. We find that even at $\alpha_{\max} = 180°$, where the fish model is fully folded, the fish cannot accomplish a 180° turn. This is a reflection of the mechanical constraints imposed by the mechanics of the fish and fish-fluid interactions on the turning angle θ. At $\alpha_{\max} = 120°$ which is the largest experimentally-observed value of bending angle in (27), the turning angle θ is barely 100°.

D. Effort in performing cyclic shape deformations. We assume the total time T to trace each of the shape trajectories in fig. S16A and fig. S16B is constant, and we evaluate the total energy associated with these shape deformations

$$E = \int_0^T T_{\text{shape}}(t)\mathrm{d}t = \frac{1}{2}(J + m_2 a^2) \int_0^T (\dot{\alpha}_1^2(t) + \dot{\alpha}_2^2(t))\mathrm{d}t. \quad [61]$$

In fig. S17, we show the energy associated as a function of the response angle θ for the family of shape changes $\alpha_1(t), \alpha_2(t)$ in Eq. (60) parameterized by A.

E. Statistical evidence of the physical constraint in C-start response.. Now that we have established that the C-start mechanics fish imposes constraints and limits on large turning motions, we look at our experimental dataset again to probe the scarcity of large prey response. In fig. S17A, we show the frequencies of the magnitude $|\theta|$ of prey responses recorded experimentally and of those predicted by the distance-optimal strategy and the orthogonal strategy without noise. We calculate a 'prediction error' by subtracting the frequencies of the experimental data from the predicted frequencies (fig. S17B). Superimposed on this plot, we show the actuation effort computed following Eq. (61). Apart one exception point in the distance-optimal strategy, both strategies predict higher frequencies of evasion angles θ that exceed 100° and lower frequencies of evasion angles below 100° than observed experimentally. For the orthogonal strategy, the prediction error rises sharply around 100°, where the actuation effort increase sharply and non-linearly. These discrepancies between model predictions and experimental data further support our findings based on the physics model that zebrafish larvae undergoing a C-start evasion response have difficulty performing turns that are larger than 100°.

References

1. Nair A, Changsing K, Stewart WJ, McHenry MJ (2017) Fish prey change strategy with the direction of a threat. *Proceedings of the Royal Society B: Biological Sciences* 284(1857):20170393.
2. Stewart WJ, Nair A, Jiang H, McHenry MJ (2014) Prey fish escape by sensing the bow wave of a predator. *Journal of Experimental Biology* 217(24):4328–4336.
3. Stewart WJ, Cardenas GS, McHenry MJ (2013) Zebrafish larvae evade predators by sensing water flow. *Journal of Experimental Biology* 216(3):388–398.
4. Berens P (2009) Circstat: A matlab toolbox for circular statistics. *Journal of Statistical Software, Articles* 31(10):1–21.
5. Blaxter J, Fuiman L (1990) The role of the sensory systems of herring larvae in evading predatory fishes. *J. Mar. Biol. Assoc. U. K.* 70(2):413–427.
6. McHenry M, Feitl K, Strother J, Van Trump W (2009) Larval zebrafish rapidly sense the water flow of a predator's strike. *Biol. Ltrs.* 5(4):477–479.
7. Easter Jr SS, Nicola GN (1996) The development of vision in the zebrafish (danio rerio). *Developmental biology* 180(2):646–663.
8. Li G, Müller UK, van Leeuwen JL, Liu H (2014) Escape trajectories are deflected when fish larvae intercept their own c-start wake. *Journal of The Royal Society Interface* 11(101):20140848.
9. Eaton RC, Emberley DS (1991) How stimulus direction determines the trajectory of the mauthner-initiated escape response in a teleost fish. *Journal of Experimental Biology* 161(1):469–487.
10. Weihs D, Webb P (1984) Optimal avoidance and evasion tactics in predator-prey interactions. *Journal of Theoretical Biology* 106(2):189 – 206.
11. Humphries DA, Driver PM (1970) Protean defence by prey animals. *Oecologia* 5(4):285–302.
12. Schervish MJ (2012) *Theory of statistics*. (Springer Science & Business Media).
13. Perez-Cruz F (2008) Kullback-leibler divergence estimation of continuous distributions in *2008 IEEE International Symposium on Information Theory*. pp. 1666–1670.
14. Rossi RJ (2018) *Mathematical statistics: an introduction to likelihood based inference*. (John Wiley & Sons).
15. Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6):716–723.
16. Burnham KP, Anderson DR, eds. (2002) *Information and Likelihood Theory: A Basis for Model Selection and Inference*. (Springer New York, New York, NY), pp. 49–97.

17. Müller UK, van Leeuwen JL (2004) Swimming of larval zebrafish: ontogeny of body waves and implications for locomotory development. *Journal of Experimental Biology* 207(5):853–868.
18. Weihs D (1973) The mechanism of rapid starting of slender fish. *Biorheology* 10(3):343–350.
19. Gazzola M, Van Rees WM, Koumoutsakos P (2012) C-start: optimal start of larval fish. *Journal of Fluid Mechanics* 698:5–18.
20. Kanso E, Marsden JE (2005) Optimal motion of an articulated body in a perfect fluid in *Proceedings of the 44th IEEE Conference on Decision and Control*. (IEEE), pp. 2511–2516.
21. Kanso E, Marsden JE, Rowley CW, Melli-Huber JB (2005) Locomotion of articulated bodies in a perfect fluid. *Journal of Nonlinear Science* 15(4):255–289.
22. Lamb H (1945) *Hydrodynamics*, Dover Books on Physics. (Dover publications).
23. Leonard NE (1997) Stability of a bottom-heavy underwater vehicle. *Automatica* 33(3):331–346.
24. Hatton RL, Choset H (2010) Connection vector fields and optimized coordinates for swimming systems at low and high reynolds numbers in *ASME 2010 Dynamic Systems and Control Conference*. (American Society of Mechanical Engineers Digital Collection), pp. 817–824.
25. Jiao Y, et al. (2021) Learning to swim in potential flow. *Physical Review Fluids* 6(5):050505.
26. Nair A, Azatian G, McHenry MJ (2015) The kinematics of directional control in the fast start of zebrafish larvae. *Journal of Experimental Biology* 218(24):3996–4004.
27. Voesenek CJ, Pieters RPM, Muijres FT, van Leeuwen JL (2019) Reorientation and propulsion in fast-starting zebrafish larvae: an inverse dynamics analysis. *Journal of Experimental Biology* 222(14).
28. Jammalamadaka SR, Sengupta A (2001) *Topics in circular statistics*. (world scientific) Vol. 5.

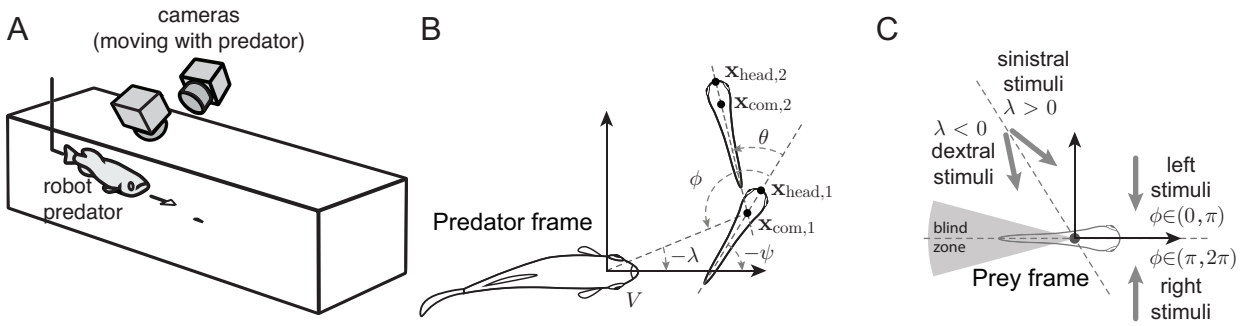

Fig. S1. Experimental measurement of evasion behavior in zebrafish larvae. (A) Experiment setup. The preys were initially motionless in the water tank. The robot predator was controlled to approach the preys at a certain speed and two cameras were used to record the evasion response. (B) Locations of the fish center-of-mass and fish head were measured right before and after the evasion response in the predator's frame of reference. The measurements were then processed to derive the perceived predator state ϕ, ψ, λ in the prey's frame of reference, as well as the prey response θ. (C) Categories of stimuli considered in the candidate strategies based on the sign of ϕ or λ.

Yusheng Jiao[1], Brendan Colvert[2], Yi Man[1], Matthew McHenry[3], Eva Kanso[1,17]

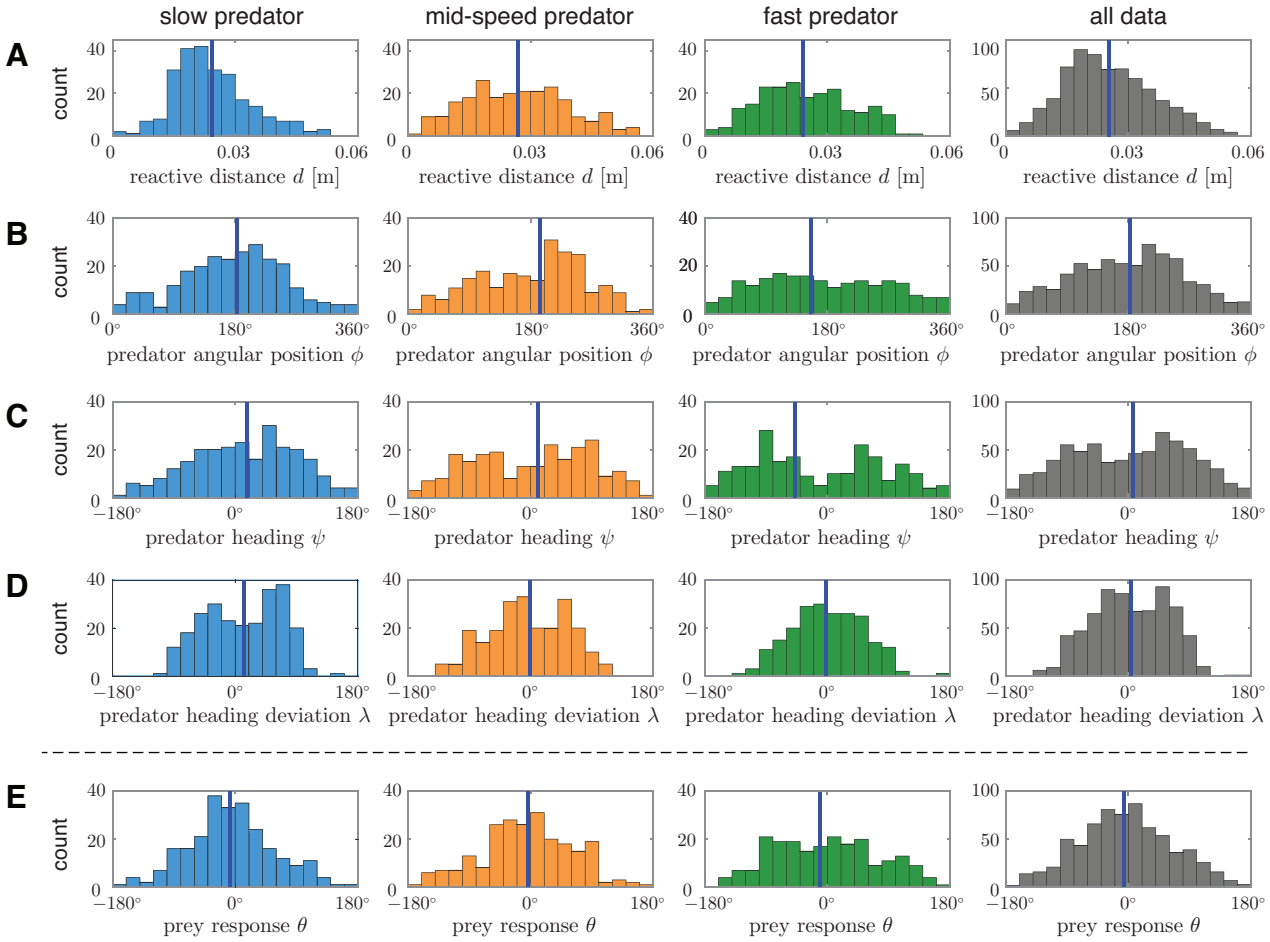

Fig. S2. Univariate histograms of experimental data. (A)-(D) Predator state at the onset of evasion. (E) Prey response, i.e., change in orientation during evasion. Three left columns show three datasets organized by predator speed; last column shows all data combined. In each histogram, the mean value is marked by blue line. For the angular variables ϕ, ψ, λ, we computed circular means using the Matlab Circular Statistics Toolbox (4, 28) and defined the proper range of angles based on the means.

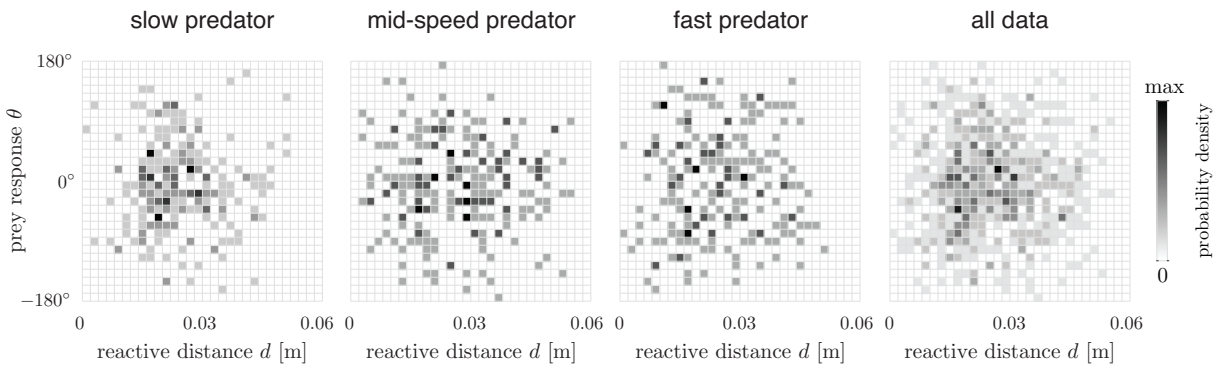

Fig. S3. Bivariate histograms between reactive distance d and prey response θ at evasion. The intensity of gray represents the data occurrence density. From left to right, the linear-circular correlation coefficient is $\mathcal{C} = 0.04809, 0.09034, 0.07005, 0.06974$ and the p-value is $p = 0.7481, 0.3864, 0.5901, 0.1827$. That is, there is no significant correlation between the prey's escape direction θ and the reaction distance d at the onset of evasion.

Yusheng Jiao[1], Brendan Colvert[2], Yi Man[1], Matthew McHenry[3], Eva Kanso[1]19

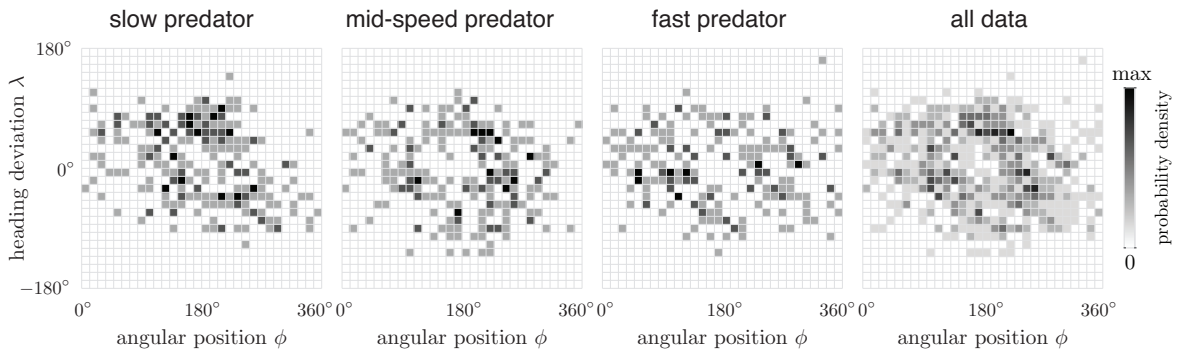

Fig. S4. Bivariate histograms between predator's angular position ϕ and relative heading λ at the onset of evasion. We notice two trends: First, the data density is low in the center of each histogram ($\phi \approx 180^\circ, \lambda \approx 0$). This is when the predator appears in the 'blind zone' of the zebrafish larvae vision and flow sensing modalities; see fig. S1C. However, even for ϕ equal or close to 180°, evasion is triggered for larger values of the predator's heading deviation λ at all predator speeds. Second, at slow predator speed, there is a weak correlation between the magnitude of the two measurements of predator's angular position ϕ and relative heading λ, as evidenced by the circular correlation coefficient is $\mathcal{C} = -0.2571, -0.1278, -0.0736$ and the p-value $p = 0.0000, 0.0346, 0.2472$ for slow, mid-speed, and fast predators, respectively. That is, at slow predator speed, the prey is more likely to escape when the predator is heading towards the front side of prey ($\lambda > 0$ for $0^\circ < \phi < 180^\circ$ and $\lambda < 0$ for $180^\circ < \phi < 360^\circ$), as opposed to when the predator is heading towards the back of from the prey. This correlation gets weaker for faster predator. Here, and in all subsequent statistical analysis, the circular correlation coefficients were calculated using the MATLAB Circular Statistics Toolbox (4, 28).

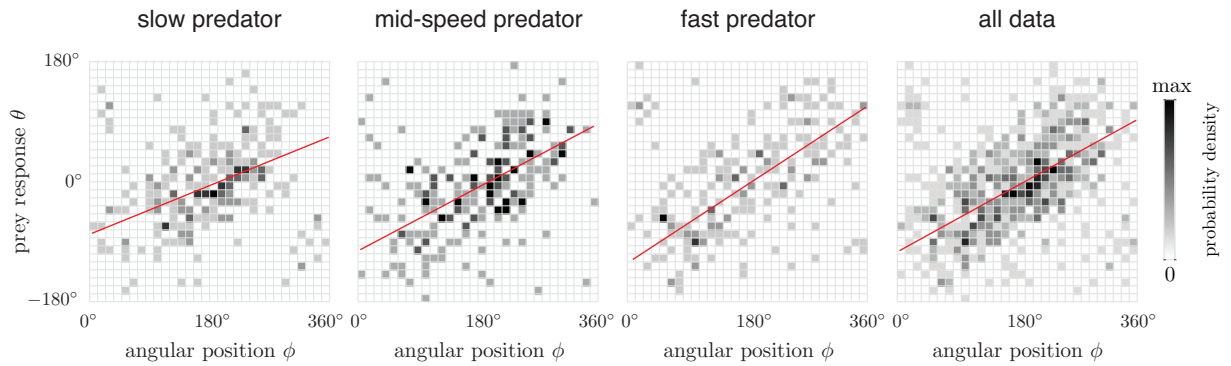

Fig. S5. Bivariate histograms between predator angular position ϕ and prey response θ. The intensity of gray represents the data occurrence density. From left to right, the circular-circular correlation coefficient is $\mathcal{C} = 0.3890, 0.5330, 0.5097, 0.4870$, and the p-value is $p = 0.0000, 0.0000, 0.0000, 0.0000$. That is there is significant linear correction between ϕ and θ. The circular regression gives, from left to right, $\theta = 0.5290\phi - 79^\circ$ for slow predator, $\theta = 0.4057\phi - 104^\circ$ for mid-speed predator, $\theta = 0.6537\phi - 122^\circ$ for fast predator, and $\theta = 0.5513\phi - 105^\circ$ for combined data; regression lines are superimposed in red.

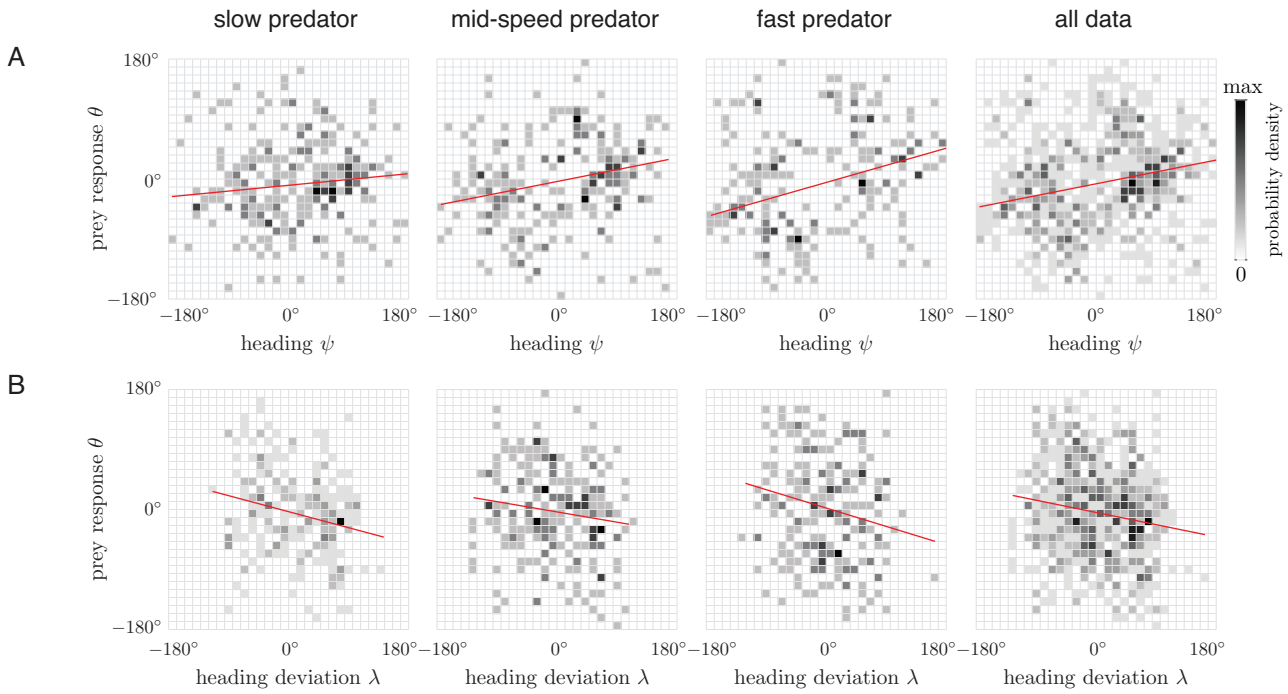

Fig. S6. Bivariate histograms between predator heading ψ and prey response θ and between predator relative heading λ and prey response θ. (A) From left to right, the circular-circular correlation coefficient is $\mathcal{C} = 0.13429, 0.28031, 0.21852, 0.24764$ and the p-value is $p = 0.025880.00000.0009, 0.0000$. There is consistent but not strong positive correlation between ψ and θ. The circular regression gives, from left to right, $\theta = 0.0967\psi - 9^\circ$ for slow predator, $\theta = 0.1975\psi - 4^\circ$ for mid-speed predator, $\theta = 0.2852\psi - 5^\circ$ for fast predator, and $\theta = 0.1975\psi - 7^\circ$ for combined data; regression lines are superimposed in red. The variance unexplained by linear regression is large. (B) Bivariate histograms between predator heading deviation λ and prey response θ. From left to right, the circular-circular correlation coefficient is $\mathcal{C} = -0.2009, -0.1377, -0.0762, -0.1419$, and the p-value is $p = 0.0016, 0.0333, 0.2474, 0.0001$. The circular regression gives, from left to right, $\theta = -0.2671\lambda - 3^\circ$ for slow predator, $\theta = 0.1709\lambda - 4^\circ$ for mid-speed predator, $\theta = -0.3061\lambda + 2^\circ$ for fast predator, and $\theta = -0.2038\lambda - 4^\circ$ for combined data; regression lines are superimposed in red. That is, there is consistent but weak negative correlation between λ and θ. The weak negative correlation is largely due to the relationship $\lambda = \psi - \phi - \pi$ the with predator angular position ϕ.

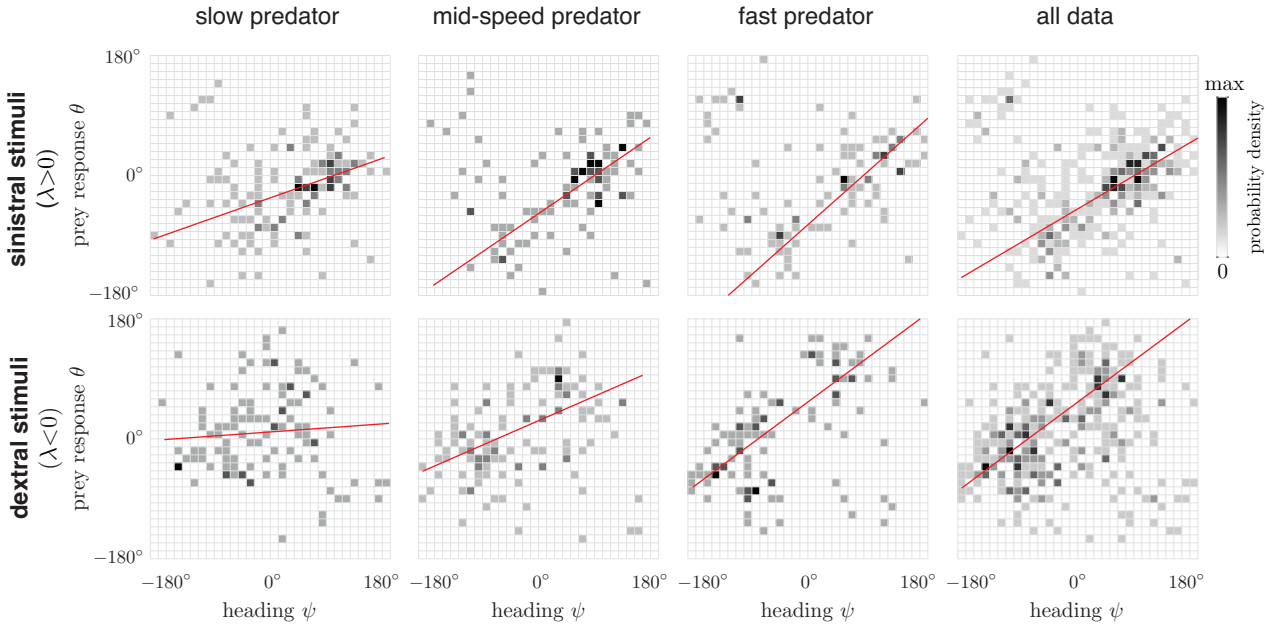

Fig. S7. Bivariate histograms between predator heading ψ and prey response θ organized based on predator relative heading λ. Dataset is split based on the sign of relative heading λ: In each case, sinistral stimulus for $\lambda > 0$ and dextral stimulus for $\lambda < 0$, the joint distribution of predator heading ψ and prey response θ is plotted. From left to right, for $\lambda > 0$, the circular-circular correlation coefficient is $\mathcal{C} = 0.4147, 0.5626, 0.5731, 0.5129$, and the p-value is $p = 0.0000, 0.0000, 0.0000, 0.0000$ for $\lambda > 0$; whereas for $\lambda < 0$, $\mathcal{C} = 0.1959, 0.4857, 0.5377, 0.4537$ and $p = 0.03423, 0.0000, 0.0000, 0.0000$. The circular regression gives, for $\lambda > 0$, $\theta = 0.3520\psi - 34^\circ$ for slow predator, $\theta = 0.6812\psi - 57^\circ$ for mid-speed predator, $\theta = 0.8876\psi - 74^\circ$ for fast predator, and $\theta = 0.5899\psi - 50^\circ$ for combined data, whereas for $\lambda < 0$, $\theta = 0.0721\psi + 10^\circ$ for slow predator, $\theta = 0.4361 + 27^\circ$ for mid-speed predator, $\theta = 0.7406\psi + 55^\circ$ for fast predator, and $\theta = 0.4970\psi + 32^\circ$ for combined data; regression lines are superimposed in red. Compared to the bivariate histograms in fig. S6(A), in each case, there is more significant correlation between ψ and θ, with the exception of the slow predator dataset when $\lambda < 0$.

Yusheng Jiao[1], Brendan Colvert[2], Yi Man[1], Matthew McHenry[3], Eva Kanso[1 2 3]

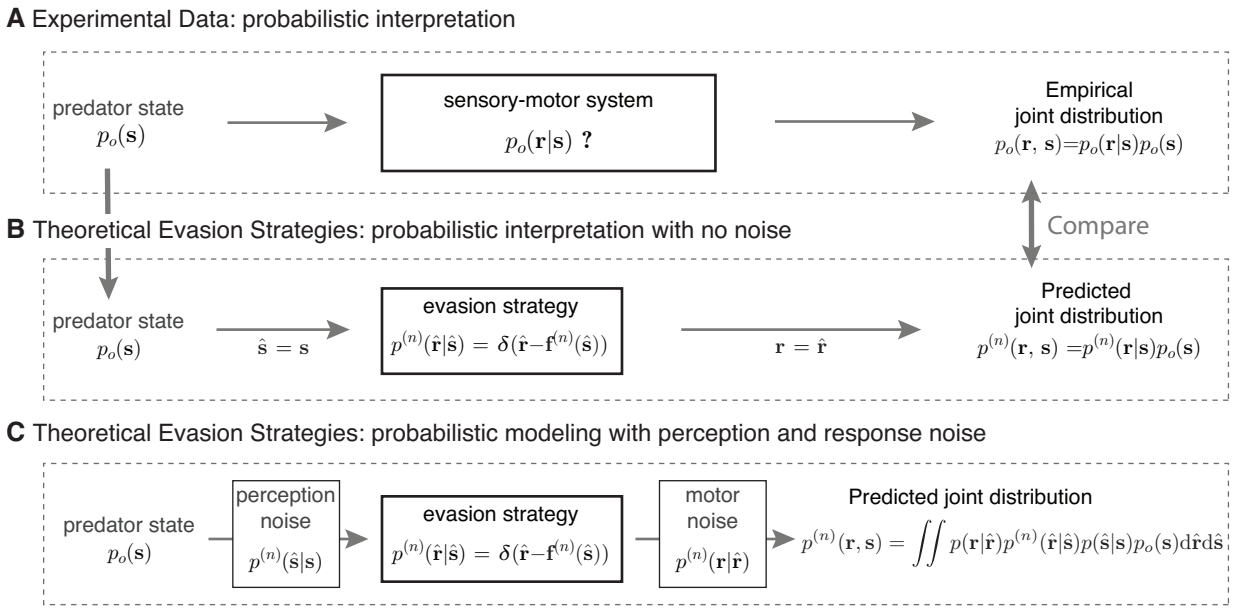

Fig. S8. Probabilistic interpretation of experimental data and probabilistic formulation of theoretical evasion models (A) Predator state $p_o(\mathbf{s})$ and joint distribution $p_o(\mathbf{s}, \mathbf{r})$ of predator state and prey response at evasion are accessible experimentally, however, the conditional probability $p_o(\mathbf{r}|\mathbf{s})$ of an evasion response given a predatory stimulus is not readily accessible and remains unknown. (B) Theoretical evasion models $\mathbf{r}^{(n)} = \mathbf{f}^{(n)}(\mathbf{s})$ that map predatory stimulus $\mathbf{s}$ into evasion response $\mathbf{r}$ can be expressed in probabilistic form; the experimentally-observed distribution of predatory stimuli $p_o(\mathbf{s})$ can be used as input to generate theoretical predictions of the joint distribution $p^{(n)}(\mathbf{s}, \mathbf{r})$ for each theoretical model n. (C) Perception and response noise are accounted for in the probabilistic modeling of evasion.

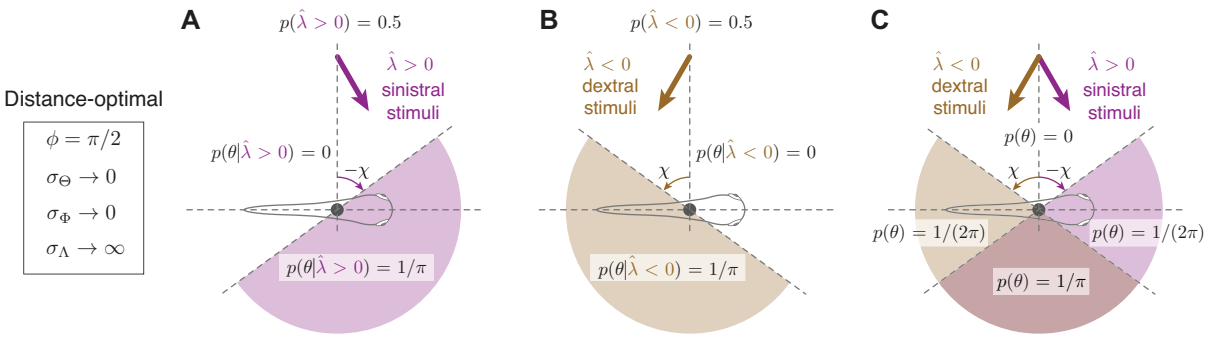

Fig. S9. Limiting behavior of distance-optimal strategy when predator heading deviation λ is very noisy $\sigma_\Lambda \to \infty$. High level of noise means uniformly random perception of λ, but it alone does not cause random prediction in prey response θ. The predicted response is distributed away from the predator based on predator angular position ϕ and parameter χ. (A) probability distribution of predicted prey response in the case of sinistral stimuli $\hat{\lambda} > 0$. The prediction is equally distributed in the purple half circle. (B) probability distribution of predicted prey response in the case of dextral stimuli $\hat{\lambda} < 0$. The prediction is equally distributed in the brown half circle. (C) the overall probability distribution of predicted prey response with large noise in predator heading. The probability distribution density $p(\theta)$ is calculated by the law of total probability.

Yusheng Jiao[1], Brendan Colvert[2], Yi Man[1], Matthew McHenry[3], Eva Kanso[1,25]

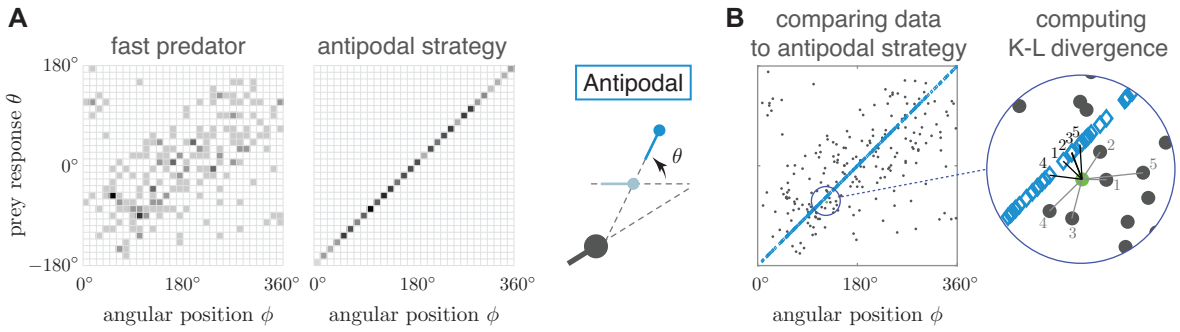

Fig. S10. K-L divergence estimation method. (A) Distribution of (ϕ, θ) based on experimental data from the fast-predator dataset (left) as compared to the predicted distribution of (ϕ, θ) from the antipodal strategy (right), shown schematically to the far right. (B) Dark gray dots represent the experimental data and blue diamonds represent antipodal model prediction, here superimposed and shown as data points rather than occurrence density as in panel A. On the right is a zoom-in demonstration of the K-L divergence estimate method using $K = 5$. From a data point from the experiment (green dot), the nearest five neighbors from the experimental data and from the antipodal strategy prediction are highlighted.

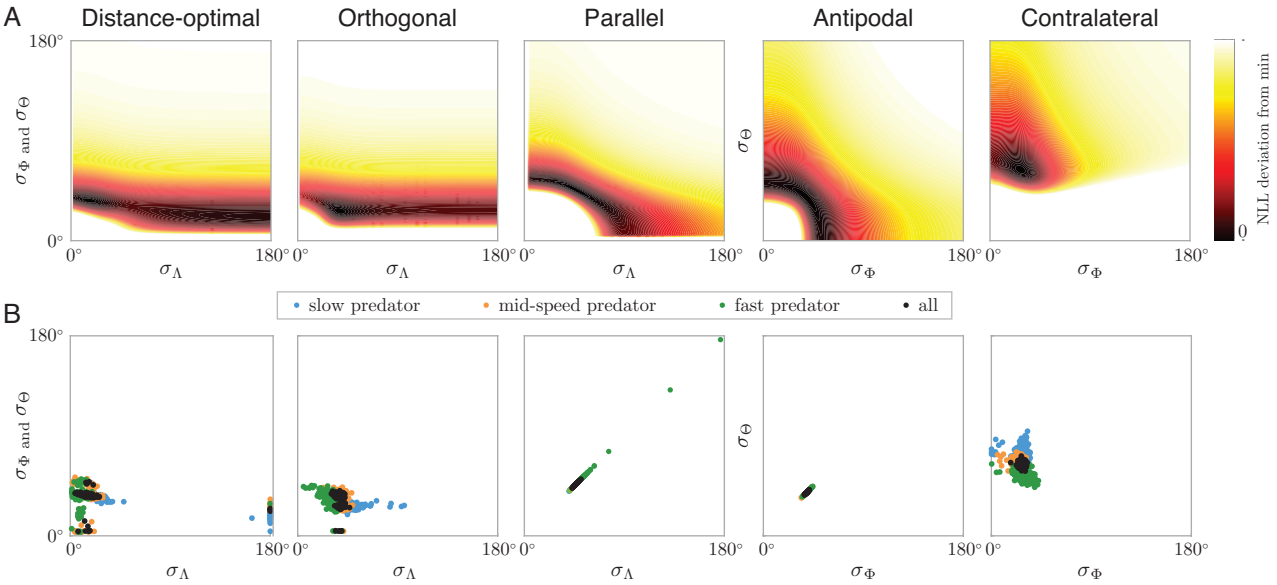

Fig. S11. Optimization of the noise parameters. (A) Values of the NLL (negative log-likelihood) over the noise parameter space of each evasion model using all data from all 699 evasion instances. The color indicates the relative NLL compared the minimum NLL. The darkest point gives the optimal noise parameters that best fit the experimental data (lowest NLL values or largest likelihood values). In the first three models, the 3D parameter space is reduced to 2D, by utilizing the mathematical symmetry between σ_Φ and σ_Θ. (B) The optimized noise parameters of all bootstrapping evaluations. Two hundred sets of optimal parameters are plotted using resampled data from each dataset, represented by different colors, per candidate strategy. Each resampled dataset is of the same size as the original dataset.

Yusheng Jiao[1], Brendan Colvert[2], Yi Man[1], Matthew McHenry[3], Eva Kanso[1]27

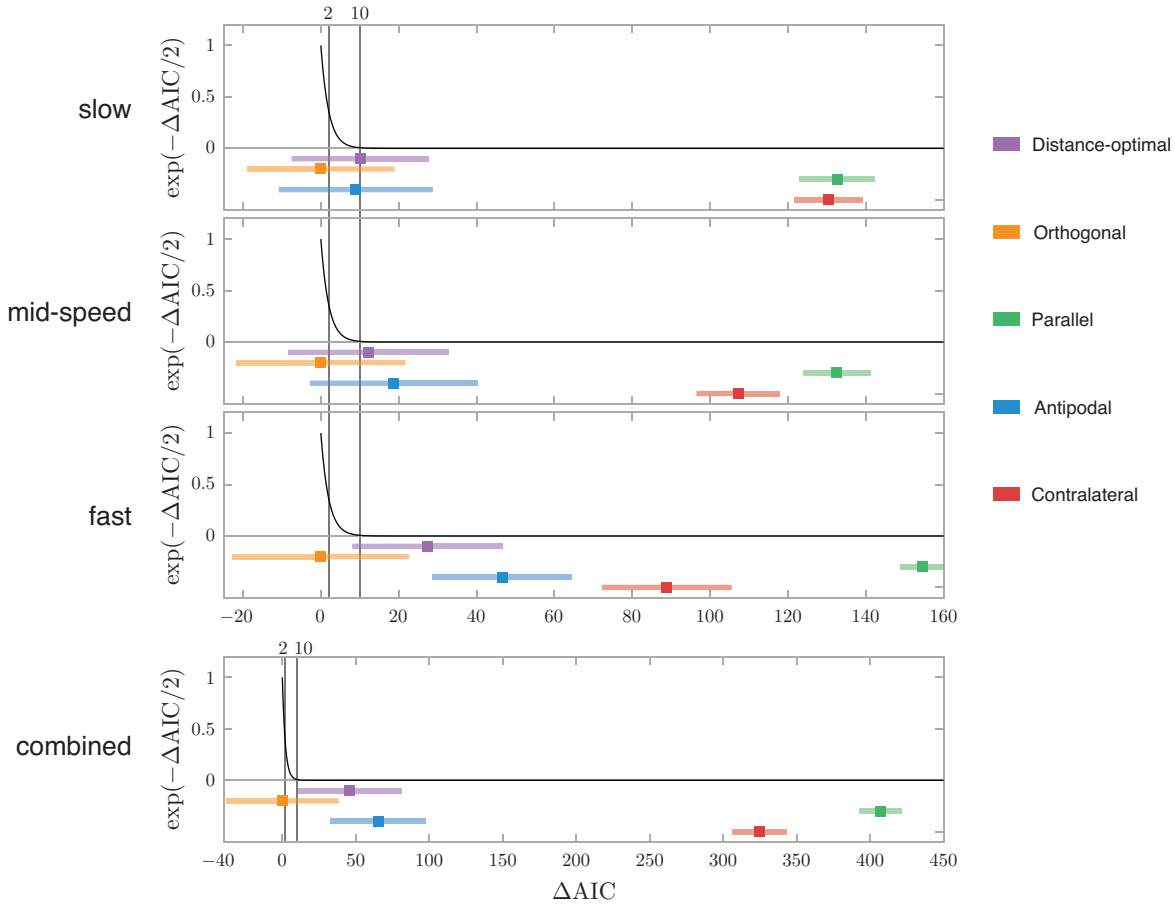

Fig. S12. Interpretation of AIC differences. The quantity $\exp(-\Delta\mathrm{AIC}/2)$ (black lines) can be interpreted as the relative likelihood, or the probability that the model minimizes information loss relative to the best-ranking model. The mean values and the range of one standard deviation of AICs for the noisy strategy models (same as Fig.4B of the main text) are highlighted along the horizontal axis. The two empirical thresholds $\Delta\mathrm{AIC} = 2$ and $\Delta\mathrm{AIC} = 10$ are marked by gray vertical lines. Compared to the orthogonal strategy, the probability that other models minimize information loss is almost nil. Considering the distribution of AICs, the distance-optimal strategy and the antipodal strategy overlap with the orthogonal strategy within one standard deviation, except for antipodal strategy in fast predator dataset. In all datasets, the parallel and contralateral strategies are consistently further away from the range of one standard deviation of the orthogonal strategy.

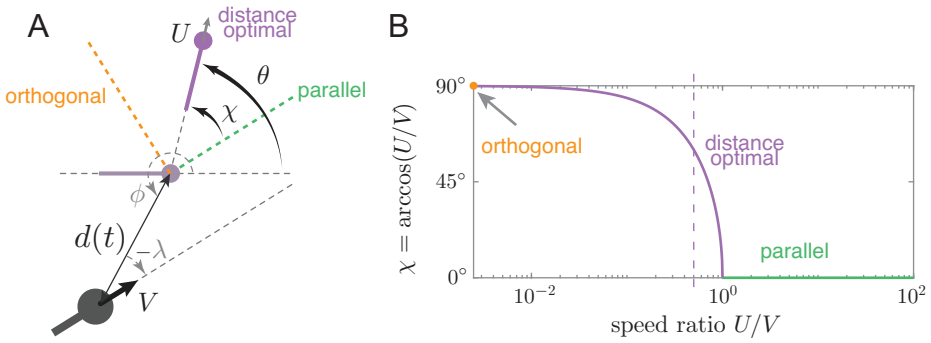

Fig. S13. Comparison between distance-optimal, orthogonal and parallel strategies.. (A) The distance-optimal strategy by Weihs and Webb (10) maximizes the minimum distance between the predator and the prey during the evasion. The angle χ is determined by the speed ratio between the prey and the predator U/V. In the orthogonal strategy, the prey always heads at 90° angle away from the predator's heading direction. In the parallel strategy, the prey moves in the same direction as the predator. (B) The relationship between χ and the speed ratio U/V. The distance-optimal strategy converges to the orthogonal strategy $\chi = 90^\circ$ when $U/V \to 0$ (fast predator) and to the parallel strategy $\chi = 0^\circ$ when $U/V \geq 1$ (slow predator).

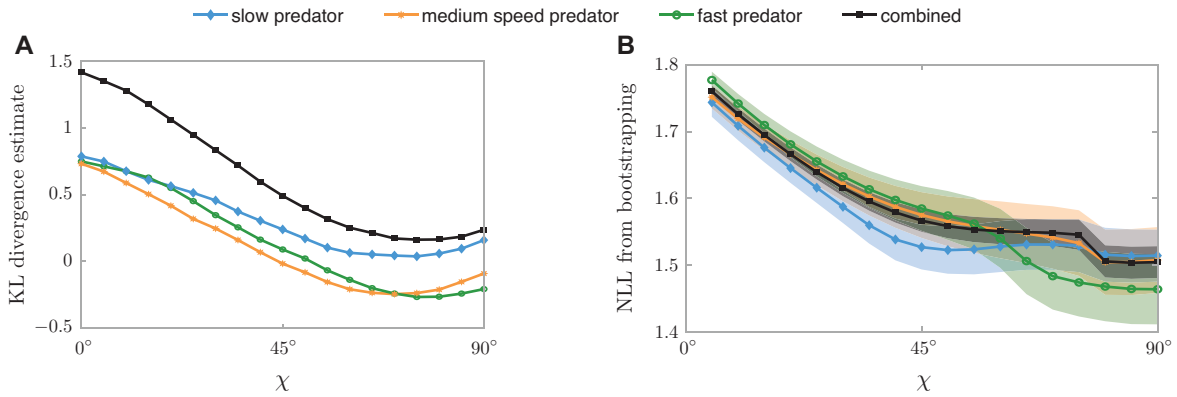

Fig. S14. Evaluation of distance-optimal strategy as a function of $\chi = \cos^{-1}(U/V)$. The distance-optimal strategy uses the speed ratio U/V as a parameter. (A) The K-L divergence estimate for the distance-optimal strategy compared to experimental data, using $k = 5$. We vary the value of χ from 0° to 90° for each dataset. (B) The average NLLs for the distance-optimal strategy as a function of χ. Again, we vary the value of χ from 0° to 90° and computed the mean and standard deviation for each χ and each dataset using bootstrapping, where we randomly drew the same number of samples with repetition from the actual experimental data. In every evaluation, all noise parameters are optimized within the range $(0, \pi)$ (the same process as in fig. S11) to get the lowest possible NLL value.

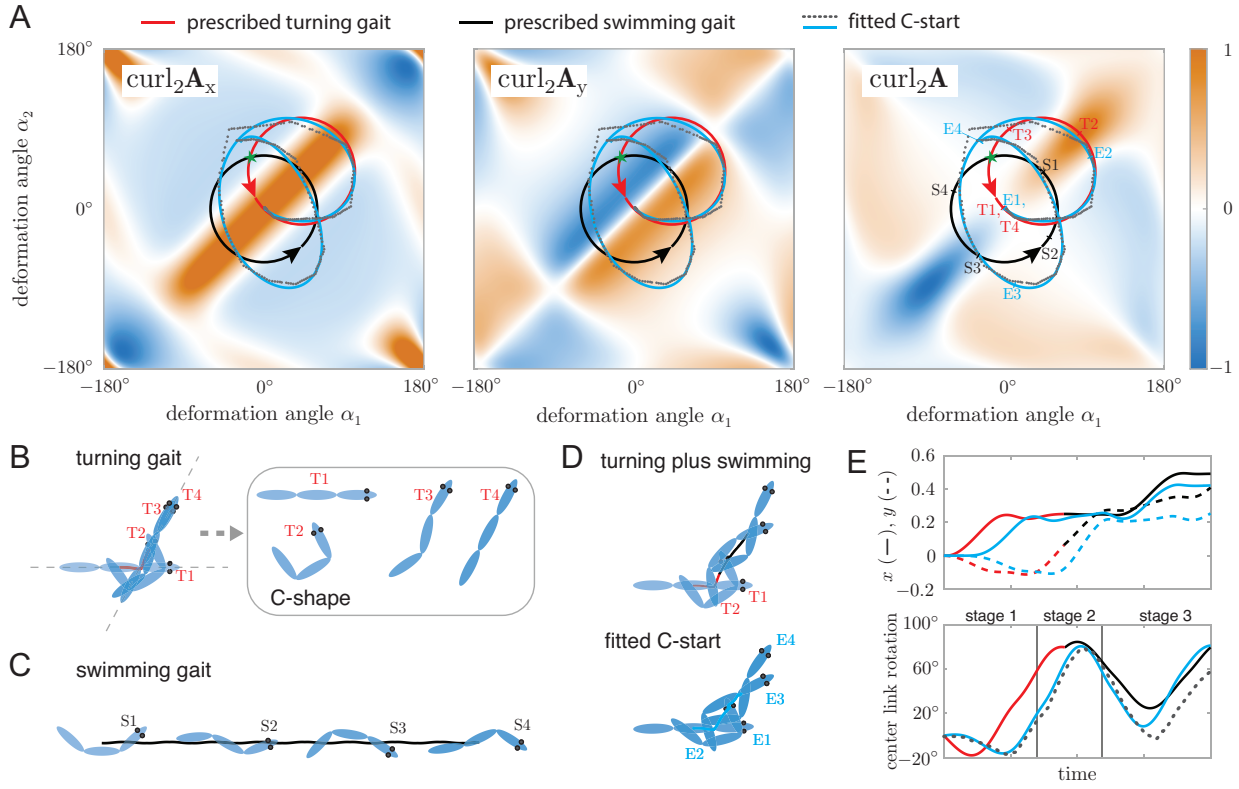

Fig. S15. Simulation of C-start response with the three-link fish model. (A) The scalar fields formed from the geometric mechanics analysis. The magnitude has been normalized to be within the range of $[-1, 1]$. The net motion of the three-link fish from following a closed trajectory on the plane can be inferred from the color enclosed by the trajectory. Examples of trajectories that produce swimming gait and turning gait are given by the black and the red circles, respectively. The grey dots show the trajectory of a C-start response extracted from experiment. This trajectory is then fitted by third-order Fourier series, shown in light blue. (B) Snapshots of the turning gait following the red circular trajectory in A. The corresponding locations on the deformation trajectory are marked by numbers 1,2,3,4 in red. (C) Snapshots of the swimming gait following the black circular trajectory in A for multiple times. The corresponding locations on the deformation trajectory for the four snapshots are marked by numbers 1,2,3,4 in black. Snapshots of the turning plus swimming by following the red circle first and then joining the black circle at the green star, alongside the snapshots of the C-start motion by following the light blue trajectory in A. (E) (above) The net displacement (x, y) of three-link fish generated from the prescribed turning-plus-swimming motion and from the fitted C-start motion (D). (below) The rotation in the center link generated from the prescribed turning-plus-swimming motion, from the fitted C-start motion, and the rotation measured from experiment (26).

Yusheng Jiao[1], Brendan Colvert[2], Yi Man[1], Matthew McHenry[3], Eva Kanso[1]31

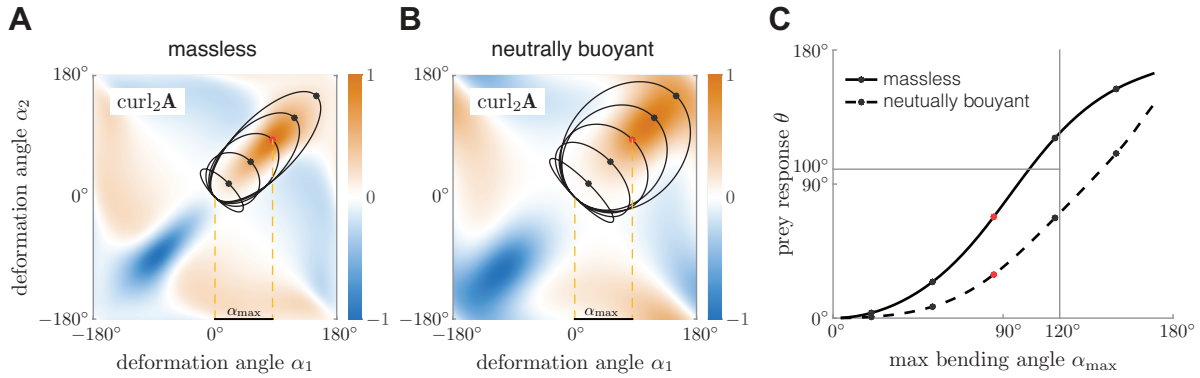

Fig. S16. Predicting prey response from its bending angle. A family of elliptic trajectories are parameterized by the max bending angle $\alpha_{\max}$. The other axis length is selected by maximizing the turn angle achieved in one cycle. (A) A few representative trajectories are shown for the massless three-link fish model. The black and red dots mark the point where the fish curls to a 'C' shape and where the max bending angle $\alpha_{\max}$ is defined. The distance between the two dashed lines gives $\alpha_{\max}$ for the trajectory marked by the red dot. (B) Representative trajectories with same max bending angles for the neutrally buoyant three-link fish model. (C) The constructed mapping from max bending angle $\alpha_{\max}$ to prey response θ for models with different mass parameters.The head and tail of fish model start crossing each other when $\alpha_{\max}$ exceeds 120°, where the predicted prey response reaches around 100° depending on the mass parameter.

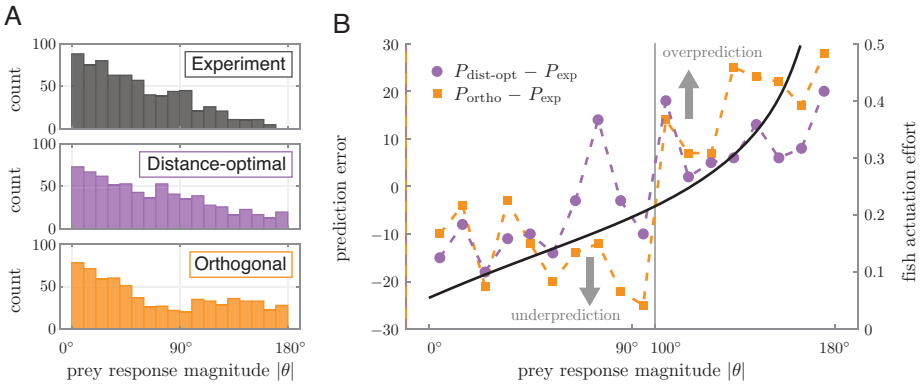

Fig. S17. Comparison between model predictions and experiment measurements of prey response magnitude $|\theta|$. (A) The histogram of $|\theta|$ from experiment dataset, distance-optimal strategy and orthogonal strategy. (B) The colored markers with dotted lines show the difference in the frequency of prey response θ predicted by distance-optimal and orthogonal strategies compared to the experiment data. Both errors change from negative to positive as response angle θ increases above 100 degrees. Here prey responses from all datasets are shown together with no distinction. The black solid line shows the actuation effort required to turn a certain angle computed from the 3-link fish model, which exhibits the same trend as the prediction errors.

Yusheng Jiao[1], Brendan Colvert[2], Yi Man[1], Matthew McHenry[3], Eva Kanso[1]33

Table S2. Results from statistical analysis of experimental data

Univariate analysis								
Variable	Slow predator		Mid-speed predator		Fast predator		Combined	
	mean	std	mean	std	mean	std	mean	std
reactive distance d	0.02434	0.00934	0.02685	0.01232	0.02470	0.01115	0.02529	0.01100
predator angular position ϕ	4.4742°	62.3731°	3.2134°	64.7275°	1.8999°	73.9732°	5.0989°	67.2255°
predator heading ψ	5.2485°	64.1633°	4.4574°	71.5149°	0.17955°	75.8059°	0.76637°	71.1338°
predator heading deviation λ	0.43071°	51.7531°	4.663°	52.333°	0.83638°	48.5776°	4.3195°	51.2418°
prey response θ	4.8889°	54.1421°	3.1013°	56.9645V	5.3878°	65.2675°	0.31316°	58.7074°

* The mean and the std above refer to circular mean and circular standard deviation for all variables except reactive distance d.

Bivariate analysis between predator state and prey response θ												
Predator state	Slow predator			Mid-speed predator			Fast predator			Combined		
	Correlation		LR	Correlation		LR	Correlation		LR	Correlation		LR
	coef. $\mathcal{C}$	p-value	slope	coef. $\mathcal{C}$	p-value	slope	coef. $\mathcal{C}$	p-value	slope	coef. $\mathcal{C}$	p-value	slope
reactive distance d	0.0481	0.7481		0.0903	0.3864		0.0701	0.5901		0.0697	0.1827	
angular position ϕ	0.3890	0.0000	0.406	0.5330	0.0000	0.529	0.5097	0.0000	0.654	0.4870	0.0000	0.551
heading ψ	0.1343	0.0259	0.097	0.2803	0.0000	0.198	0.2185	0.0009	0.285	0.2476	0.0000	0.198
heading deviation λ	-0.2009	0.0016	-0.267	-0.1377	0.0333	-0.171	-0.0762	0.2474	-0.306	-0.1419	0.0001	-0.204
heading $\psi\|_{\lambda>0}$	0.4147	0.0000	0.352	0.5626	0.0000	0.681	0.5731	0.0000	0.888	0.5129	0.0000	0.590
heading $\psi\|_{\lambda<0}$	0.1959	0.03423	0.072	0.4857	0.0000	0.436	0.5377	0.0000	0.741	0.4537	0.0000	0.497

* The correlation above refers to circular-linear correlation between d and θ and circular-circular correlation for the other rows.

* The LR slope above denotes the slope coefficient of the linear regression between two circular variables.