

11 Supporting Information Text

12 **Mean squared displacement.** To characterize the diffusiveness of movements, we calculated the mean squared displacement
 13 (MSD) as defined in the Materials and Methods section of the main text for the simulated trajectories. Moreover, how MSD
 14 increases with time t can characterize the movements (1) as follows:

$$15 \quad \text{MSD} = \langle (\mathbf{X}_t - \mathbf{X}_0)^2 \rangle \sim t^\alpha$$

16 where $\alpha \in [0, 2]$ is an exponent for MSD. Brownian motions and straight lines have $\alpha = 1$ and $\alpha = 2$, respectively. The fact
 17 that the exponent α for Lévy walks is larger than 1 suggests that Lévy walks are superdiffusive.

18 For $\varepsilon = 0.1, 0.22, 0.3$, an ensemble of 100 trajectories was obtained from different initial conditions, x_0 and y_0 , which were
 19 randomly chosen. The result shown in Fig. S1 shows that MSD and indicates that the slope α of $\varepsilon = 0.22$ is between those
 20 corresponding to $\varepsilon = 0.1$ and $\varepsilon = 0.3$. The relationship between time and MSD for $\varepsilon = 0.22$ does not seem linear over the
 21 whole range. The reason is that the step length distribution is not a pure power law distribution; instead, it a truncated power
 22 law distribution. Therefore, the slope of the log(MSD) versus log(time) relationship can converge to 1 after a sufficiently long
 23 time. However, we can conclude that the movement patterns for $\varepsilon = 0.22$ are highly diffusive; they are more diffuse than those
 24 for $\varepsilon = 0.1$, and they are less diffuse than those for $\varepsilon = 0.3$, considering that biologically realistic situations are not the limit of
 25 long time. This pattern is consistent with the argument that the movement patterns can be classified as Lévy walks.

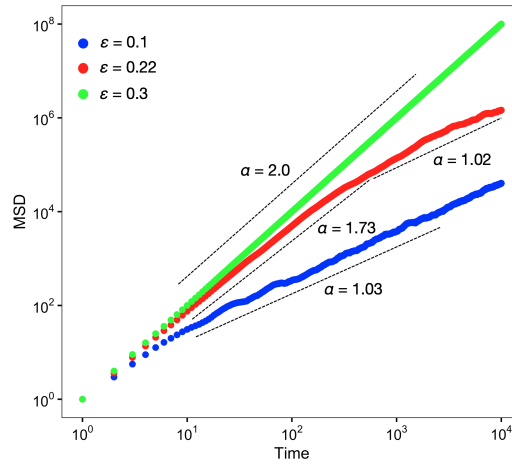


Fig. S1. Mean squared displacement (MSD). Blue dots represent MSD for $\varepsilon = 0.1$, red dots represent MSD for $\varepsilon = 0.22$, and green dots represent MSD for $\varepsilon = 0.3$. The dashed lines represent the regression lines with the slope α .

26 **Characteristics of dynamics.** We showed that the movement trajectories of our model (Eqs. 1 and 2 in the main text) can be
 27 classified into Lévy walks, Brownian motions, and ballistic movements (see main text). To intuitively grasp how the dynamics
 28 of the internal states x_t, y_t produce such variable movement patterns, we show the attractors of x_t, y_t (Fig. S2). The diagonal
 29 line in Fig. S2 represents $x = y$, namely, synchronous states that lead to no turnings in the movements. Thus, if $x = y$, namely,
 30 $\Delta\theta = 0$ is a (globally) stable attractor, the movement patterns are equivalent to straight (ballistic) movements ($\varepsilon = 0.3$; Fig.
 31 S2C), although x and y fluctuate in a chaotic manner. In contrast, if $x = y$ is unstable, $x - y$ exhibits random fluctuations (see
 32 insets of Fig. 1BC in the main text). In particular, the case of $\varepsilon = 0.1$ covers a large range of x and y (Fig. S2A). Consequently,
 33 the movement patterns become a normal diffusion process, that is, Brownian motions. Between these two patterns ($\varepsilon = 0.22$),
 34 we can see the mixture of $x \simeq y$ and random fluctuations, which produces intermittent patterns in the dynamics of $x - y$ and
 35 Lévy walks as movement trajectories (Fig. S2B).

36 **Stability condition for ballistic movements.** To understand how the phase transition shown in Fig. 2 of the main text occurs,
 37 we derive the stability condition for ballistic movements, that is, $\Delta\theta = 0$. First, we set the new variables to $H = (x + y)/2$ and
 38 $K = (x - y)/2$. From Eqs. 1 and 2 in the main text, we obtain the difference equation for the new variables as follows:

$$39 \quad \begin{aligned} H(t+1) &= \frac{1}{2}(f(H(t) + K(t)) + f(H(t) - K(t))) \\ 40 \quad K(t+1) &= \frac{1-2\varepsilon}{2}(f(H(t) + K(t)) - f(H(t) - K(t))) \end{aligned}$$

41 where f is a nonlinear function that produces chaotic dynamics. Here, as mentioned in the main text, we used a tent map. The
 42 equations of a small perturbation h, k in the vicinity of $G = 0$ are obtained by linearization as follows:

$$43 \quad \begin{aligned} h(t+1) &= f'(H(t))h(t) \\ 44 \quad k(t+1) &= (1-2\varepsilon)f'(H(t))k(t) \end{aligned}$$

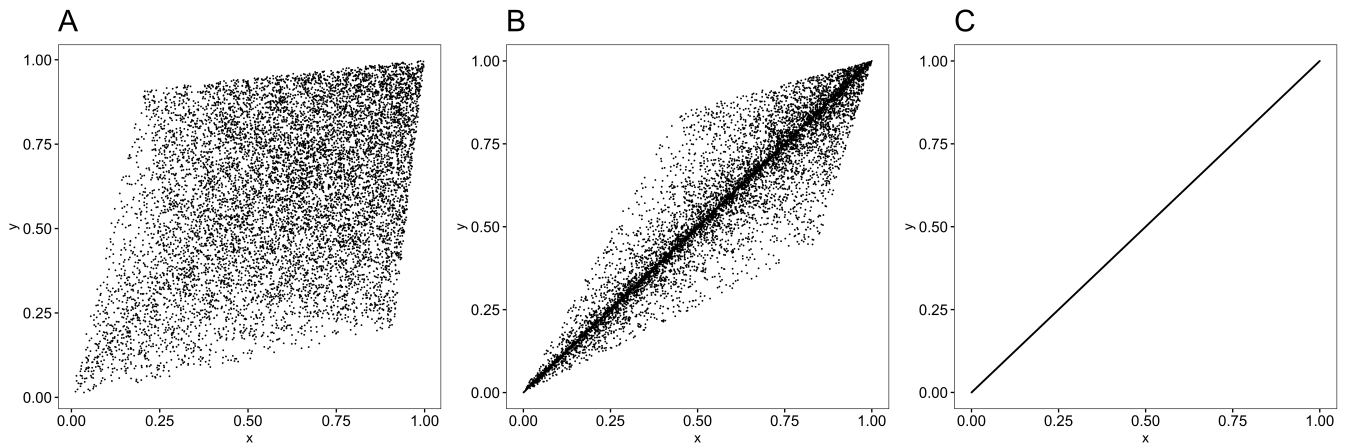


Fig. S2. Attractors depending on the parameter ε . The horizontal and vertical axes represent the internal states x and y of the agent, respectively. The black dots represent the value at a certain time. (A) $\varepsilon = 0.1$, that is, attractor of Brownian motions. (B) $\varepsilon = 0.22$, that is, attractor of Lévy walks. (C) $\varepsilon = 0.3$, attractor of straight movement. The other parameters were set to $r = 0.7$ and $t_{\max} = 10,000$.

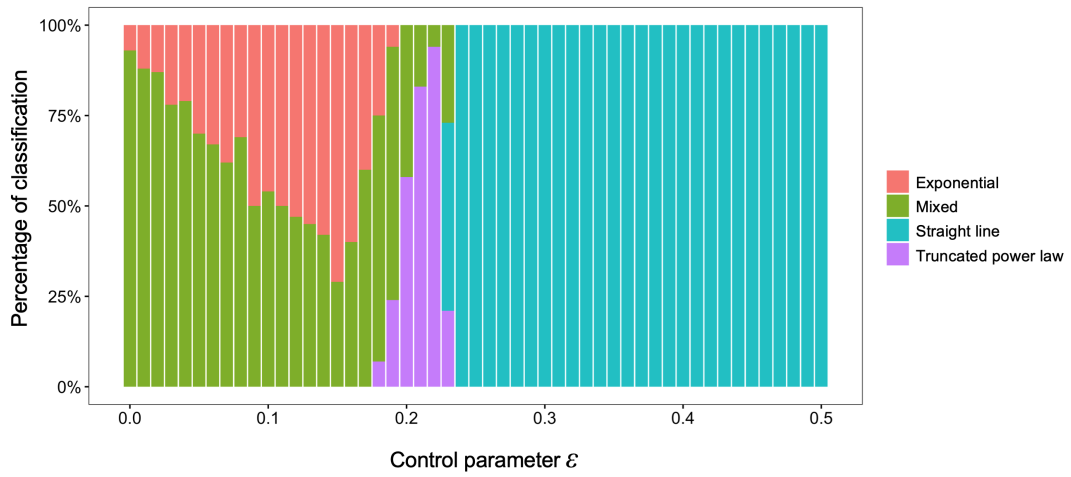


Fig. S3. Detail of the classifications based on statistical analysis. The data are the same as shown in Fig. 2 in the main text.

45 Note that we can separate $k(t)$ from $h(t)$. Because $H(t+1) = f(H(t))$ exhibits chaotic dynamics, the linear stability for
 46 $x - y = 0$ can be characterized by two exponents:

$$47 \quad \lambda_h = \langle \ln |f'(H)| \rangle$$

$$48 \quad \lambda_k = \ln |1 - 2\varepsilon| + \langle \ln |f'(H)| \rangle.$$

49 where ε is the coupling strength in the system, and $\langle \ln |f'(H)| \rangle$ denotes a Lyapunov exponent of the nonlinear map f because
 50 of the ergodicity of chaos (2). Because a variable in a non-coupled tent map f is distributed uniformly from 0 to 1 (2), we can
 51 analytically derive the Lyapunov exponents $\lambda_{tent} = \lambda_h$ of a single tent map as follows:

$$52 \quad \lambda_{tent} = -r \ln r - (1 - r) \ln (1 - r)$$

53 where r is a parameter of the tent map. From the above, the stability index λ_c of $x - y$ is described as

$$54 \quad \lambda_c = \ln |1 - 2\varepsilon| + \lambda. \quad [1]$$

55 Therefore, the stability condition for $\Delta\theta = c(x-y) = 0$ is $\ln |1 - 2\varepsilon| + \lambda < 0$. For the case of $r = 0.7$, we obtain $\varepsilon > \varepsilon_c \simeq 0.228$
 56 as the stability condition, which is consistent with the main result.

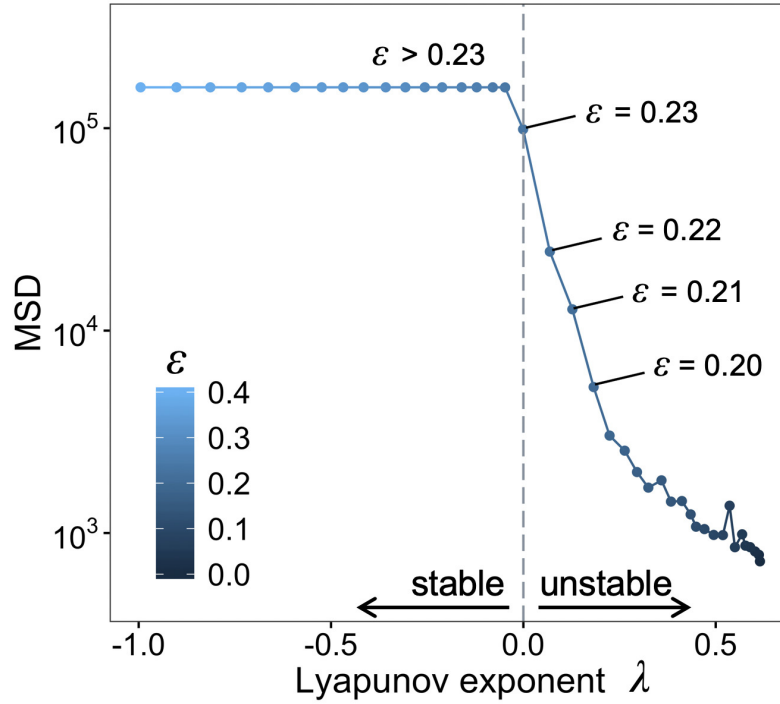


Fig. S4. A trade-off relationship between the instability index (Lyapunov exponent λ) of $x_t - y_t$ and the diffusiveness (MSD) of the agents. The Lyapunov exponent was obtained analytically using equation 1. The MSD was calculated for $t = 400$. The color of the plots represents the control parameter ε . Note that the results for $\varepsilon > 0.4$ are not shown to improve the visibility. The gray dashed line corresponds to a Lyapunov exponent of 0, which signifies a boundary between the stable and unstable phases.

57 **Other models.** We assessed the robustness of the results for emergence conditions of Lévy walks as shown in Fig. 2 of the main
 58 text for different nonlinear functions, the number of elements, and stochastic noise.

59 **Model with different nonlinear functions.** First, we explored the effect of the parameter r in the tent map on the condition for the
 60 emergence of Lévy walks. Fig. S5 shows that the parameter r does not qualitatively affect the result. In the case of $r = 0.5$, we
 61 used $r = 0.501$ because the rounding error of the values caused a falling to a fixed point. The result suggests that the tendency
 62 of Lévy walks to emerge near a critical point was insensitive to r . Second, we used different nonlinear functions, here logistic
 63 maps $f(x) = ax(1 - x)$, where $a \in [0, 4]$ is a parameter, to determine the attractor (3). The result of the model with $a = 3.9$,
 64 which produces chaotic dynamics, is shown in Fig. S6 and also supports the main conclusion.

65 **Model with large degree of freedom.** In the main analysis, we adopted a model system composed of only two elements. However,
 66 one might wonder if the number of elements affects the result. To investigate this question, we analyzed globally coupled maps

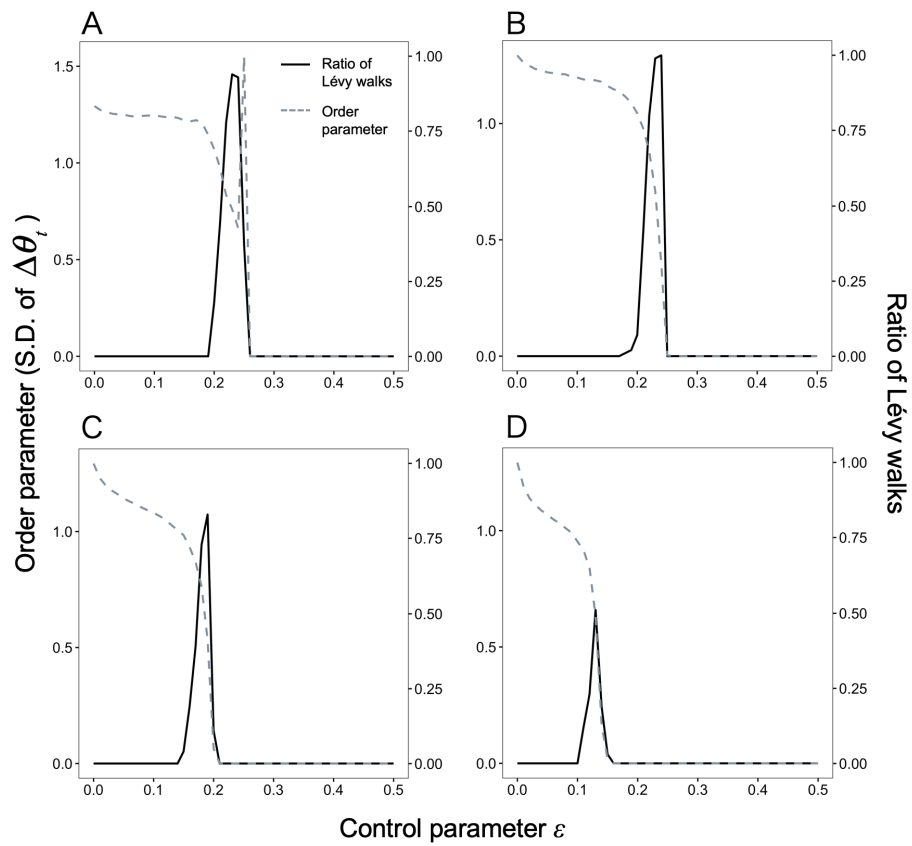


Fig. S5. Phase diagram for the model with different r . (A) $r = 0.501$, (B) $r = 0.6$, (C) $r = 0.8$, (D) $r = 0.9$. The other settings are the same as for the results shown in Fig. 2 in the main text.

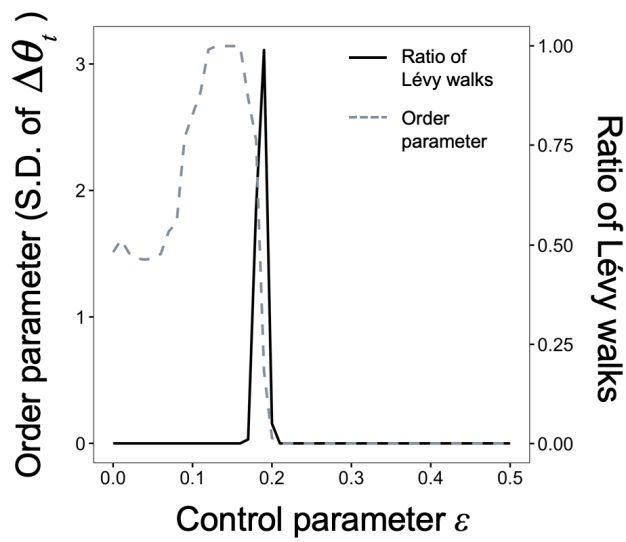


Fig. S6. Phase diagram for the logistic map model with $\alpha = 3.9$. The other settings are the same as those shown in the results in Fig. 2 in the main text.

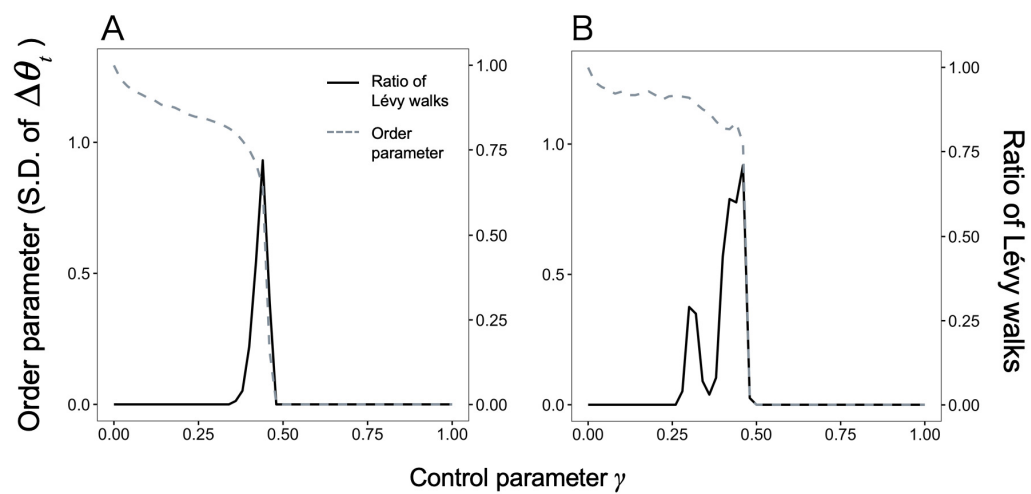


Fig. S7. Phase diagram for the model with large degrees of freedom. (A) $N = 10$, (B) $N = 100$. The other settings are the same as those shown in Fig. 2 in the main text.

67 (GCMs) with N elements, which have been used for modeling complex systems such as neural networks (4). The definition of
 68 GCMs is as follows:

$$69 \quad x_i(t+1) = (1-\gamma)f(x_i(t)) + \frac{\gamma}{N} \sum_{j=1}^N f(x_j(t))$$

70 where x_i is a value of element i and $\gamma \in [0, 1]$ is the coupling strength between the elements. Here, as we did in the main
 71 analysis, we used a tent map with $r = 0.7$ as the nonlinear function f . Note that our model in the main analysis was a
 72 special case for $N = 2$ when $\gamma = 2\varepsilon$. We randomly chose two elements as the output nodes and used them for producing the
 73 movements. We used the same procedure for the two-elements model (Fig. 1A in the main text). The results for $N = 10$ and
 74 100 are shown in Fig. S7. The results suggest that even for systems with a relatively high number of elements, Lévy walks can
 75 emerge near the critical point.

76 **Model with noise.** When one considers a biologically realistic situation, there is noise in the system. To verify the robustness
 77 of the result of the emergence of Lévy walks near the critical point, we added an inherently stochastic noise to the systems.
 78 At each time step, white noise following a normal distribution with $\mu = 0, \sigma = \eta$ was added to the internal states x_t and y_t
 79 independently. Fig. S8 shows that although the stable area for large ε is likely to break down with increasing size of the noise,
 80 the tendency for the emergence of Lévy walks does not change qualitatively.

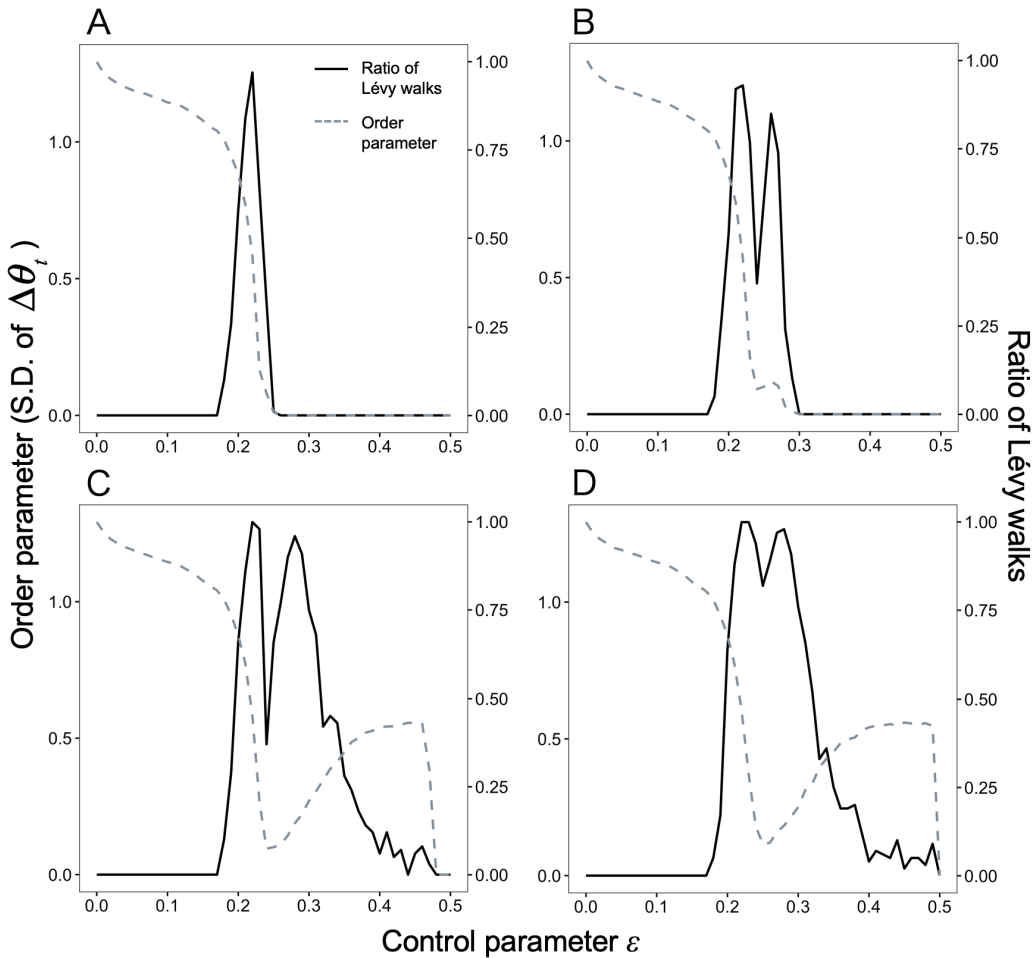


Fig. S8. Phase diagram for the model with stochastic noise. (A) $\eta = 1.0 \times 10^{-10}$, (B) $\eta = 1.0 \times 10^{-8}$, (C) $\eta = 1.0 \times 10^{-6}$, (D) $\eta = 1.0 \times 10^{-4}$. The other settings are the same as those shown in the results in Fig. 2 in the main text.

81 **Data-based simulation of responses to perturbations.** To examine responses to a perturbation in the empirical time series
 82 z_1, \dots, z_N of turning angles (Fig. 5B in the main text) where N is the number of data points, we developed a method based
 83 on a twin surrogate algorithm proposed by Thiel et al. (5). The twin surrogate algorithm allowed us to produce surrogate
 84 data with different initial states while preserving the properties of a nonlinear dynamical system of the original time series. In
 85 addition to this algorithm, we virtually observed the response to a perturbation by specifying the initial value corresponding to

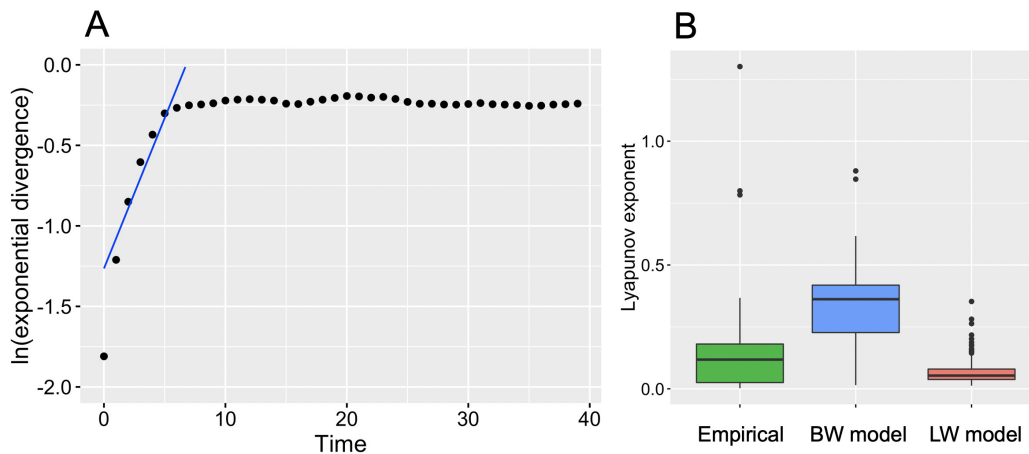


Fig. S9. Lyapunov exponents in empirical data. (A) The Lyapunov exponent is estimated as the initial slope of the log-transformed exponential divergence versus time, as calculated by linear regression (blue line) for a linear increasing region. In our analysis, a regression line was obtained for a value $d(\geq 1)$ up to 95% of the maximum value of the vertical axis. We changed d and selected the regression line with R^2 values greater than 0.9 that fit the longest region of 3 or less were excluded. (B) Estimated Lyapunov exponents in empirical data ($n = 158$), Brownian walk model (BW; $\varepsilon = 0.0$, $n = 246$), and Lévy walk model (LW; $\varepsilon = 0.22$, $n = 253$). The result of the multiple comparison test showed that all the differences between them were significant ($p < 10^{-5}$).

86 the perturbation size and stimuli type in the theoretical analysis (Figs. 3–4 in the main text) and observing the subsequent
87 dynamics.

88 **Twin surrogate algorithm.** First, we will explain the twin surrogate algorithm. This method shuffles an embedded vector sequence
89 $\mathbf{v}(1), \dots, \mathbf{v}(M)$ according to certain rules, where M is the number of vectors. Here, $\mathbf{v}(i)$ is $\{z_i, z_{i-1}, \dots, z_{i-(E-1)}\}$, and E is an
90 embedding dimension obtained from a simplex projection (6). Specifically, this method consisted of the following four steps:

1. For the M vectors $\mathbf{v}(1), \dots, \mathbf{v}(M)$ created by embedding the original time series data, create a recurrence matrix with
 $i, j = 1, \dots, M$

$$R_{i,j} = \Theta(\delta - \|\mathbf{v}(i) - \mathbf{v}(j)\|)$$

91 where $\Theta(\cdot)$ is a Heaviside function (i.e., $\Theta(\cdot) = 0$ for $\cdot < 0$ and $\Theta(\cdot) = 1$ for $\cdot \geq 0$) and $\|\mathbf{v}(i) - \mathbf{v}(j)\|$ denotes the maximum
92 norm (the maximum absolute value of each component of $\mathbf{v}(i) - \mathbf{v}(j)$). This recurrence matrix indicates whether any two
93 vectors $\mathbf{v}(i)$ and $\mathbf{v}(j)$ are farther from the threshold δ by 0 or 1. Therefore, the recurrence matrix contains the same
94 information as the original nonlinear dynamical system attractor (7), although with less resolution.

2. If $R_{i,k} = R_{j,k}$ for all $k = 1, \dots, M$, the pair $\mathbf{v}(i)$ and $\mathbf{v}(j)$ is called a twin. On a recurrence matrix of 0 and 1, it is
95 impossible to distinguish between two vectors that are twins.
3. The i -th surrogate datum is denoted by $\mathbf{v}_s(i)$, and the original vector \mathbf{v} is randomly selected as the first one, $\mathbf{v}_s(1) = \mathbf{v}$.
4. If $\mathbf{v}_s(j) = \mathbf{v}(m)$, then $\mathbf{v}_s(j+1) = \mathbf{v}(m+1)$ if $\mathbf{v}(m)$ does not have a twin. In contrast, if $\mathbf{v}(m)$ has $\mathbf{v}(n)$ as a twin, then
96 $\mathbf{v}_s(j+1) = \mathbf{v}(m+1)$ or $\mathbf{v}_s(j+1) = \mathbf{v}(n+1)$ with equal probability, even if there are more than two twins.

100 Repeat Step 4 to create surrogate data with the same length as the original. The surrogate data shall be rejected if they reach
101 the end of the original vector (i.e., $\mathbf{v}(M)$) before it reaches the same length as the original, and the end vector $\mathbf{v}(M)$ does not
102 have a twin.

103 Although this method requires the parameter δ , the dependence on δ has been shown to be small (5), and the best results
104 seem to be obtained when the ratio of 1 in the recurrence matrix $R_{i,j}$ is $0.05 \sim 0.2$. For the analysis in this paper, we adopted a
105 δ that produced a ratio of 1 in $R_{i,j} = 0.125$. It is known that the surrogate data created by this method preserve the properties
106 of the original time series. In particular, mutual information, which is a nonlinear property, is known to be conserved by twin
107 surrogates, but it is not conserved by other surrogate methods (5).

108 **Responses to perturbations.** Furthermore, to examine responses to perturbations, we improved the procedure of determining the
109 initial value in Step 3 in the following four cases:

- (a) The case of a small perturbation (corresponding to small S). We considered a new, original time series starting from
110 a randomly chosen vector $\mathbf{v}(i)$ ($1 \leq i \leq M - \tau$), where τ is the time after the perturbation. We defined the initial vector
111 of the perturbed time series to be $\mathbf{v}_s(1) = \mathbf{v}(i)$. Then, we repeated Step 4. Finally, the new, original time series was
112 $\mathbf{v}(i), \dots, \mathbf{v}(i + \tau)$, and the perturbed series was $\mathbf{v}_s(1), \dots, \mathbf{v}_s(1 + \tau)$. These two time series started from the same values,
113 but they exhibited different dynamics via twins. Therefore, $\mathbf{v}_s(1), \dots, \mathbf{v}_s(1 + \tau)$ corresponded to a time series starting
114 from a slightly different initial state.
- (b) The case of a large perturbation (corresponding to large S). The new, original time series was the same as in case (a),
115 namely $\mathbf{v}(i), \dots, \mathbf{v}(i + \tau)$. The initial value $\mathbf{v}_s(1)$ of the perturbed time series was randomly chosen from vectors that
116 were in the top 5% of the distribution of $\|\mathbf{v}(i)\|$ where $\|\cdot\|$ means the Euclidean norm. Then, we repeated Step 4. The
117 perturbed time series $\mathbf{v}_s(1), \dots, \mathbf{v}_s(1 + \tau)$ corresponded to a time series starting from a drastically changed initial state.
- (c) The case of stimuli that evoke straight movements (corresponding to small S'). To examine behavioral switching
118 according to the type of stimulus, we randomly selected an initial value from vectors that were in the bottom 5% of the
119 distribution of $\|\mathbf{v}(i)\|$. Then we generated subsequent time series with the twin surrogate method. Finally, we compared
120 the positions constructed from $\mathbf{v}_s(1)$ and $\mathbf{v}_s(1 + \tau)$.
- (d) The case of stimuli evoking turns (corresponding to large S'). We randomly selected an initial value from vectors
121 that were in the top 5% of the distribution of $\|\mathbf{v}(i)\|$. Otherwise, this case was the same as case (c).

122 Note that if we could not create the perturbed time series due to the absence of twins in entire time series or a shortage of
123 original time series, we did not use the time series for analysis. The final sample size was $n = 263$.

128 References

1. F Bartumeus, MGE da Luz, GM Viswanathan, J Catalan, Animal search strategies: a quantitative random-walk analysis.
129 *Ecology* **86**, 3078–3087 (2005).
2. A Pikovsky, J Kurths, M Rosenblum, J Kurths, *Synchronization: a universal concept in nonlinear sciences*. (Cambridge
130 university press) Vol. 12, (2003).
3. RM May, Simple mathematical models with very complicated dynamics. *Nature* **261**, 459 (1976).

- 134 4. K Kaneko, Clustering, coding, switching, hierarchical ordering, and control in a network of chaotic elements. *Phys. D:*
135 *Nonlinear Phenom.* **41**, 137–172 (1990).
- 136 5. M Thiel, MC Romano, J Kurths, M Rolf, R Kliegl, Twin surrogates to test for complex synchronisation. *EPL (Europhysics*
137 *Lett.* **75**, 535 (2006).
- 138 6. G Sugihara, RM May, Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*
139 **344**, 734–741 (1990).
- 140 7. M Thiel, MC Romano, J Kurths, How much information is contained in a recurrence plot? *Phys. Lett. A* **330**, 343–349
141 (2004).