# Supplementary Materials: Supporting Information: BioTraj: Information-Geometric Mapping of Spatial and Temporal Heterogeneity in Biological Trajectories

**Danilo Roccatano** [1]

## Supporting Information

This Supporting Information provides additional conventional trajectory descriptors, simulation details, finite-sampling diagnostics, annotation-aware analyses, and reproducibility information for the BioTraj analyses reported in the main article. The supplementary analyses are intended to document robustness and implementation details without changing the descriptive interpretation of the experimental recordings adopted in the main text.

## 1. Conventional Trajectory Analysis

The information-geometric analysis presented in the main article forms one component of the broader BioTraj framework. For completeness, this section reports conventional geometric and dynamical descriptors for the experimental trajectory datasets. These descriptors characterize complementary properties of the recordings and are not used to rank the biological datasets.

### *1.1. Trajectory distributions and explored domains*

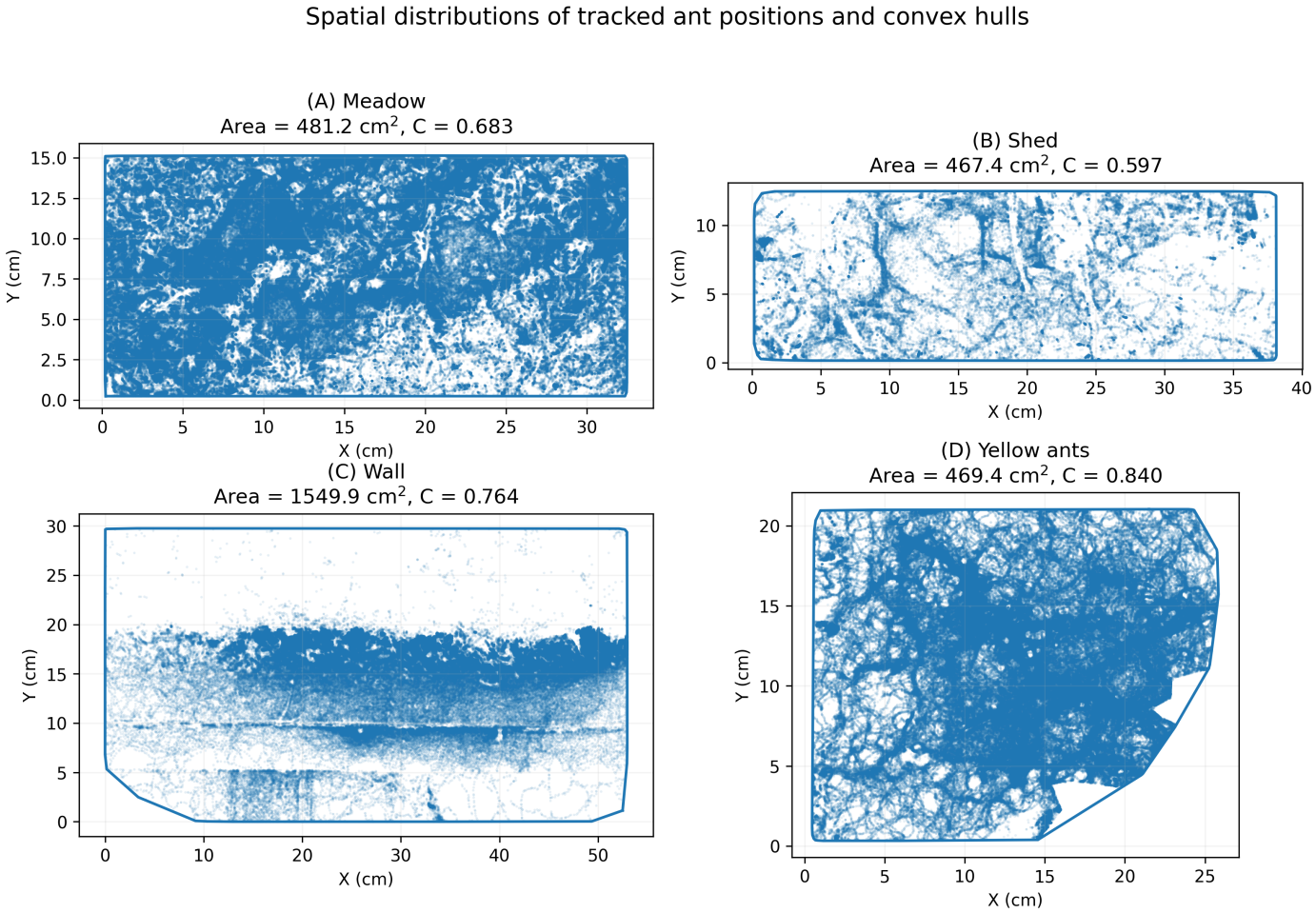



**Figure S1. Spatial distributions of tracked positions in the four experimental recordings.** Tracked ant positions are shown together with the corresponding convex hull for (A) Meadow, (B) Shed, (C) Wall, and (D) Yellow-ant recordings. The convex-hull area and circularity C are reported for each dataset. The Wall panel corresponds to the final full-frame recording.

### *1.2. Occupancy maps*

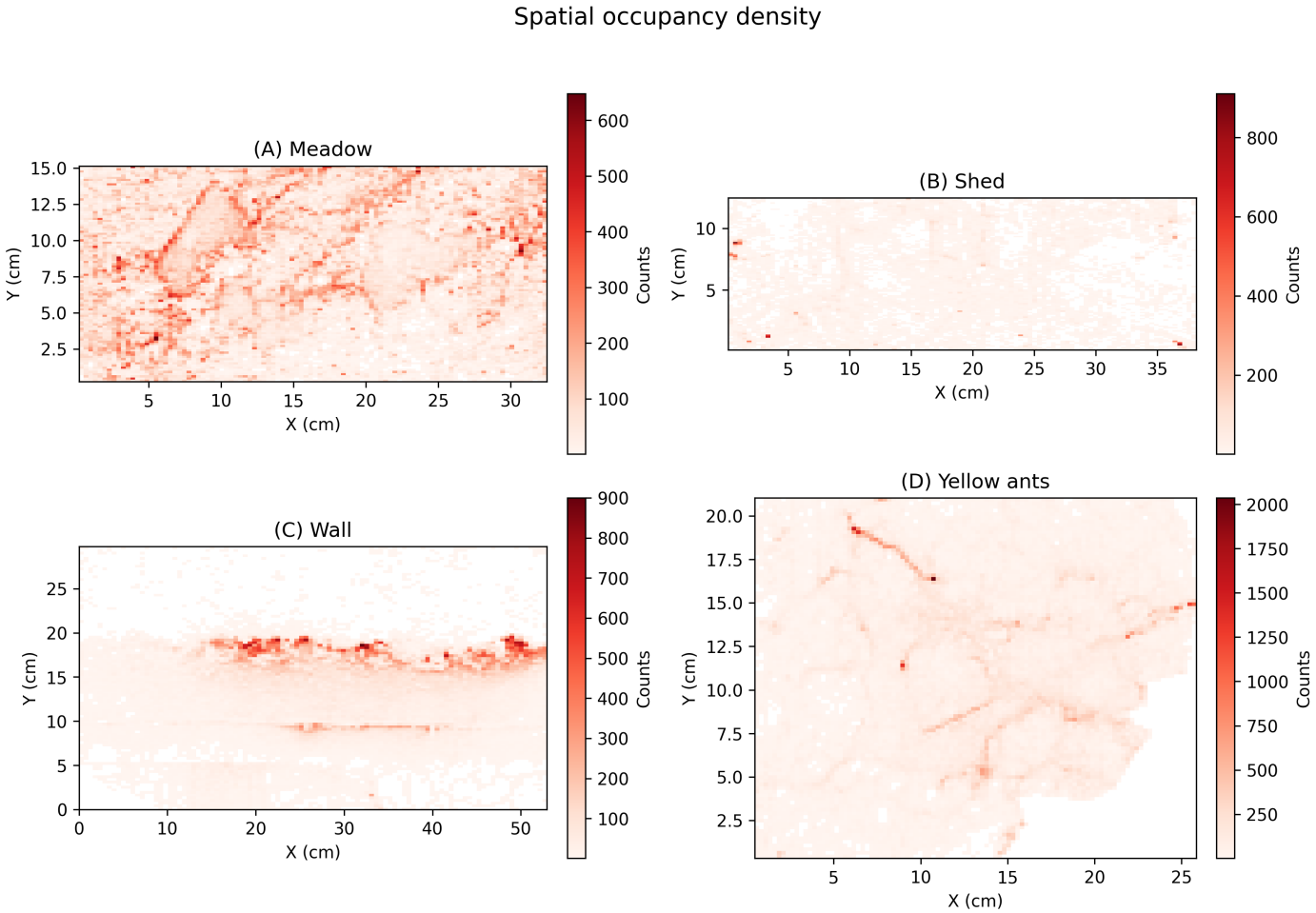



**Figure S2. Spatial occupancy distributions of the four experimental recordings.** Two-dimensional occupancy histograms for (A) Meadow, (B) Shed, (C) Wall, and (D) Yellow-ant recordings. Colour represents the number of tracked positions within each spatial bin. Colour scales are shown independently for each recording because the datasets contain different numbers of observations; colour intensity should therefore be interpreted within, rather than quantitatively between, panels. The Wall panel corresponds to the final full-frame recording.

### *1.3. Conventional trajectory descriptors*

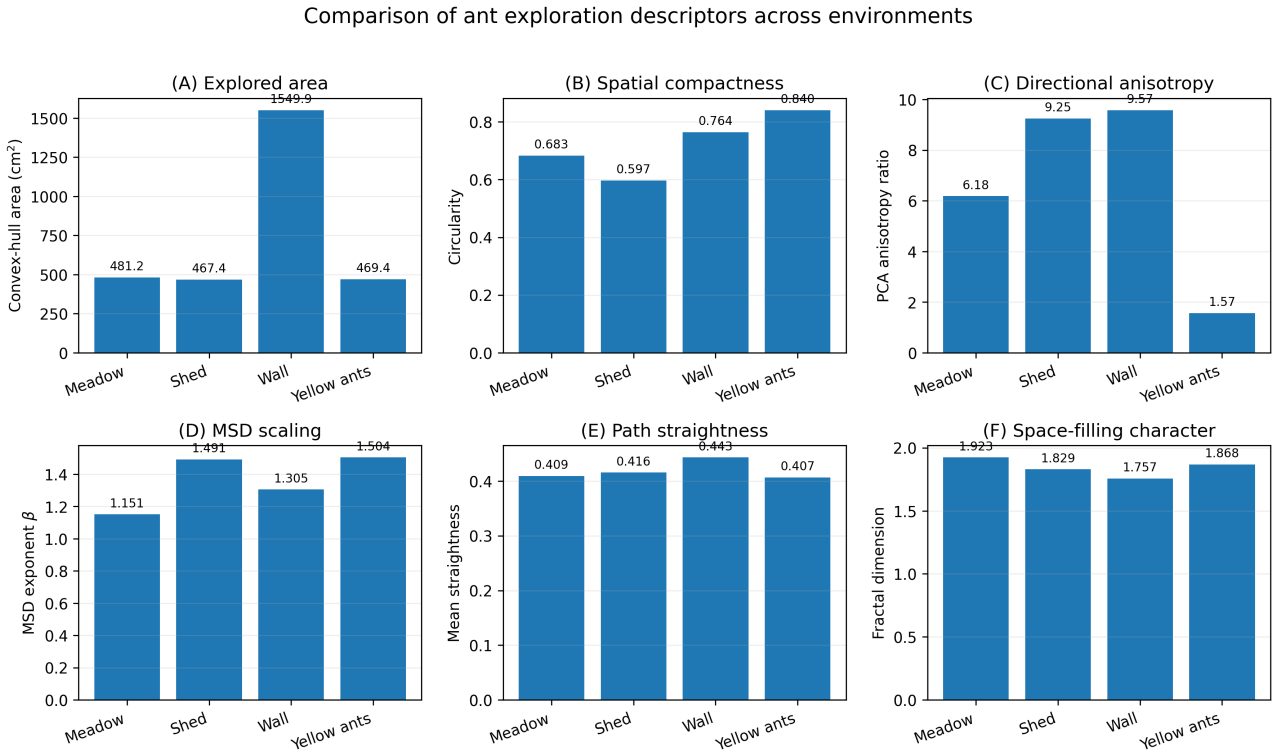



**Figure S3. Comparison of conventional trajectory descriptors across the four experimental recordings.** (A) Convex-hull explored area, (B) spatial circularity, (C) PCA anisotropy ratio, (D) mean-squared-displacement scaling exponent $\beta$ (E) mean tracklet straightness, and (F) box-counting fractal dimension. Values were obtained using the conventional trajectory-analysis components of BioTraj. The Wall results correspond to the final full-frame recording. These descriptors provide global summaries complementary to the spatially resolved information-geometric analyses presented in the main text.

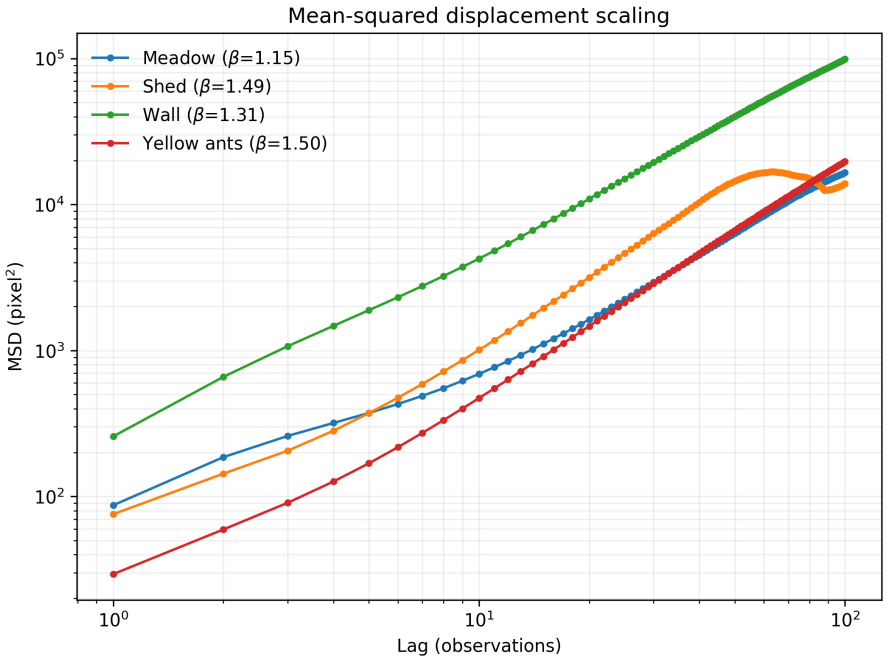



**Figure S4. Mean-squared-displacement scaling of the experimental trajectories.** Mean-squared displacement (MSD) as a function of observation lag for the Meadow, Shed, Wall, and Yellow-ant recordings. Displacements were calculated within individual tracklets and pooled at each lag. The exponent $\beta$ was obtained by fitting $\log(\mathrm{MSD}) = \beta \log(\mathrm{lag}) + c$ over the first half of the available lag range, as implemented in BioTraj. The Wall curve corresponds to the final full-frame dataset.

**Table S1. Conventional trajectory descriptors for the experimental recordings.** Values were obtained using the conventional trajectory-analysis components of BioTraj. The Wall results correspond to the final full-frame recording. Uncertainties are bootstrap standard errors.

| Metric | Wall (full frame) | Shed | Meadow | Yellow |
|---|---|---|---|---|
| Number of positions | 293,298 | 77,360 | 763,036 | 597,417 |
| Convex hull area ($cm^2$) | 1549.93 | 467.4 | 481.2 | 469.4 |
| Circularity index | 0.764 | 0.597 | 0.683 | 0.840 |
| Monte Carlo area estimate ($cm^2$) | 1547.43 | 467.33 | 481.06 | 469.28 |
| Anisotropy ratio | 9.57 | 9.25 | 6.18 | 1.57 |
| MSD exponent $\beta$ | $1.305 \pm 0.009$ | $1.492 \pm 0.045$ | $1.151 \pm 0.046$ | $1.504 \pm 0.036$ |
| Mean straightness | $0.443 \pm 0.002$ | $0.416 \pm 0.007$ | $0.410 \pm 0.003$ | $0.407 \pm 0.004$ |
| Fractal dimension $D$ | $1.757 \pm 0.024$ | $1.829 \pm 0.076$ | $1.923 \pm 0.058$ | $1.868 \pm 0.065$ |

## 2. Simulation-Based Validation

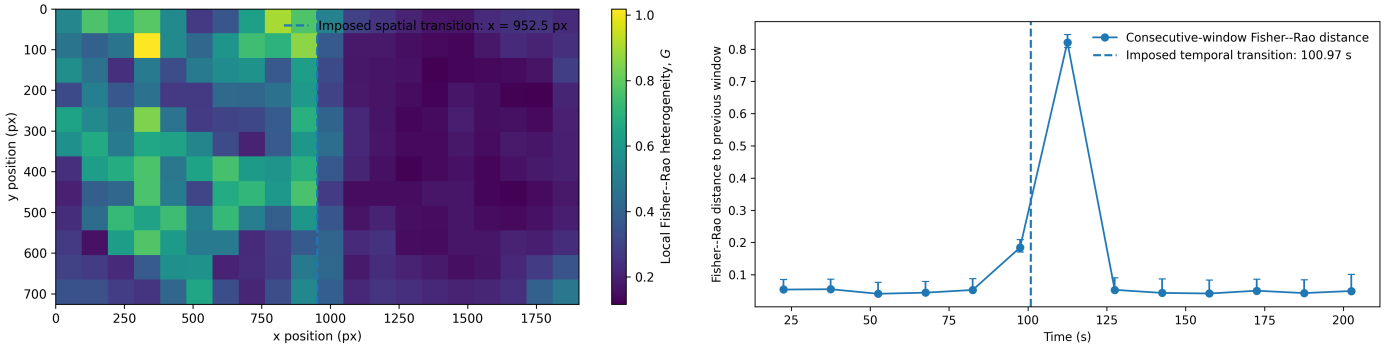
### *2.1. Simulation design*

Four simulated trajectory scenarios were used: stationary, spatial two-domain, temporal two-state, and combined spatiotemporal structure. Each simulation contained 20 agents propagated for more than 12,000 frames, producing more than 240,000 recorded positions. A common random seed was used so that differences among scenarios reflected the deliberately imposed statistical structure rather than unrelated random realizations.

**Table S2. Principal settings used for simulation-based validation and information-geometric analysis.**

| Parameter | Value |
|---|---|
| Number of simulated agents | 20 |
| Frames | 12,116 |
| Recorded positions | 242,320 |
| Frame rate | 60 frames $s^{-1}$ |
| Arena size | 1905 × 726 px |
| Spatial grid | 20 × 12 cells |
| Minimum turning angles per spatial cell | 30 |
| Turning-angle bins | 18 |
| Spatial bootstrap replicates | 200 |
| Temporal window | 900 frames (15 s) |
| Temporal step | 900 frames (15 s) |
| Minimum observations per temporal window | 100 |
| Temporal bootstrap replicates | 200 |
| Random seed | 12345 |
| Spatial transition | $x_c \approx 953$ px (between pixel columns 952 and 95 |
| Temporal transition | frame 6058 ($t_c \approx 101.0$ s) |

*2.2. Combined spatiotemporal benchmark*



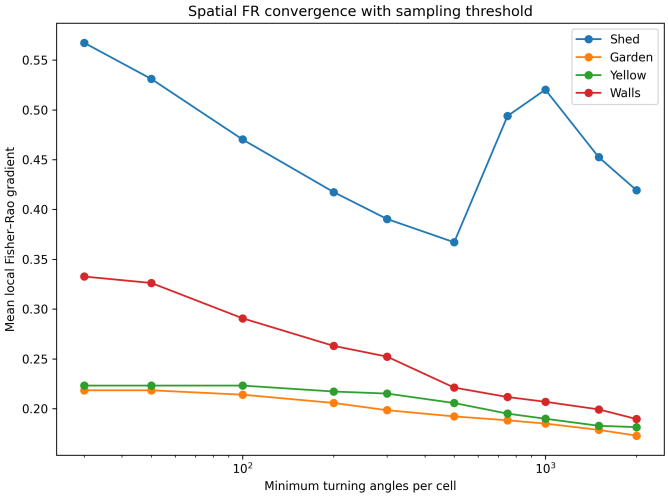

(**a**) Spatial Fisher–Rao heterogeneity.

(**b**) Temporal Fisher–Rao response.

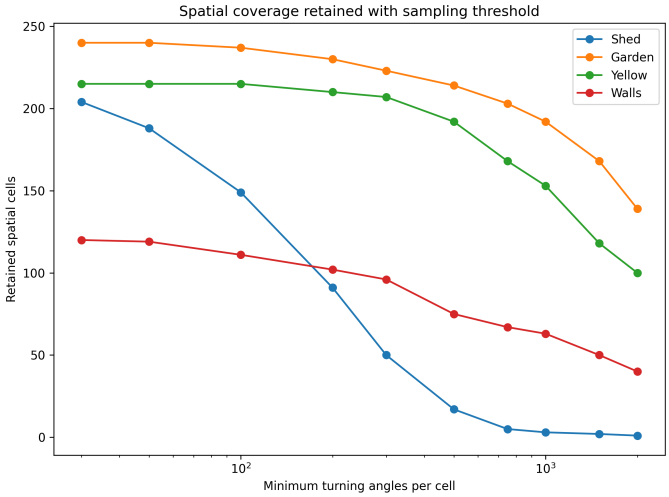
**Figure S5. Combined spatiotemporal simulation benchmark.** Unlike the separate spatial and temporal validation cases shown in Figure 3 of the main article, this simulation contains both deliberately imposed spatial and temporal changes within the same trajectory dataset. (a) Spatial distribution of local Fisher–Rao heterogeneity for the combined simulation. The dashed vertical line indicates the deliberately imposed spatial transition at $x_c \approx 953$ px (between pixel columns 952 and 953). (b) Fisher–Rao distance between consecutive 15-s temporal windows. The dashed vertical line indicates the deliberately imposed temporal transition at frame 6058 ($t_c \approx 101.0$ s at 60 frames $s^{-1}$). The temporal response increases when the transition enters the consecutive-window comparison and reaches its largest value for the subsequent comparison of strongly differing movement distributions. Together, the panels demonstrate sensitivity of the spatial and temporal information-geometric analyses to deliberately imposed distributional structure within the same simulated trajectory dataset.

## 3. Finite-Sampling Diagnostics

Empirical Fisher–Rao distances are affected by the number of observations used to estimate each histogram. The following supplementary analyses provide the detailed diagnostics underlying the finite-sampling discussion in the main article.




(**a**) Fisher–Rao heterogeneity.




(**b**) Number of retained spatial cells.

**Figure S6. Sensitivity to the minimum spatial sampling threshold.** (a) Local Fisher–Rao heterogeneity as the minimum number of turning-angle observations required per spatial cell is increased. (b) Corresponding number of spatial cells retained for analysis. The analysis illustrates the trade-off between reducing finite-sampling effects and retaining sufficient spatial coverage.

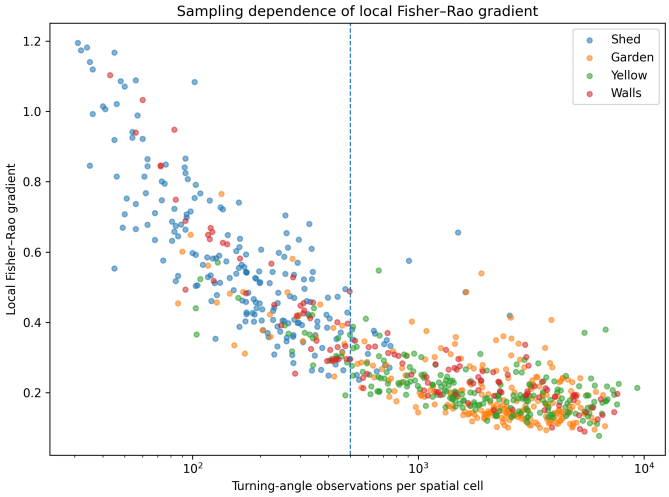



**Figure S7. Relationship between spatial sampling and local Fisher–Rao heterogeneity.** Local Fisher–Rao heterogeneity is shown as a function of the number of turning-angle observations contributing to each spatial cell for the experimental datasets. The relationship demonstrates that empirical Fisher–Rao estimates depend on local sample size, motivating the equal-sampling control shown in Figure S8.

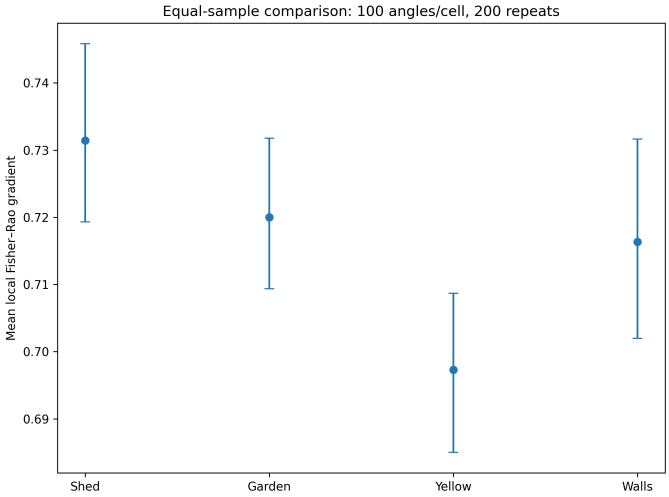



**Figure S8. Equal-sampling control of local Fisher–Rao heterogeneity.** Fisher–Rao heterogeneity was recalculated by randomly sampling 100 turning-angle observations from each eligible spatial cell, with the procedure repeated 200 times. The resulting distributions quantify the finite-sampling contribution to the observed Fisher–Rao values and should not be interpreted as biological effect sizes.

## 4. Yellow-Ant Annotation-Aware Analysis

The former rock footprint and candidate nest openings were defined from the recording scene independently of the Fisher–Rao map. Because elevated local heterogeneity was observed near the tracking-domain boundary, the regional analysis reported in the main article excludes cells within one grid-cell width (approximately 88 px) of that boundary.

**Table S3. Edge-controlled Yellow-ant regional summaries.** Spatial cells are descriptive units and are not treated as independent biological replicates.

| **Region** | ***n* cells** | **Mean *G*** | **Median *G*** |
|---|---|---|---|
| Inside former rock footprint | 93 | 0.181 | 0.177 |
| Outside former rock footprint | 46 | 0.215 | 0.214 |
| Near candidate nest openings | 30 | 0.204 | — |
| Away from candidate nest openings | 63 | 0.196 | — |

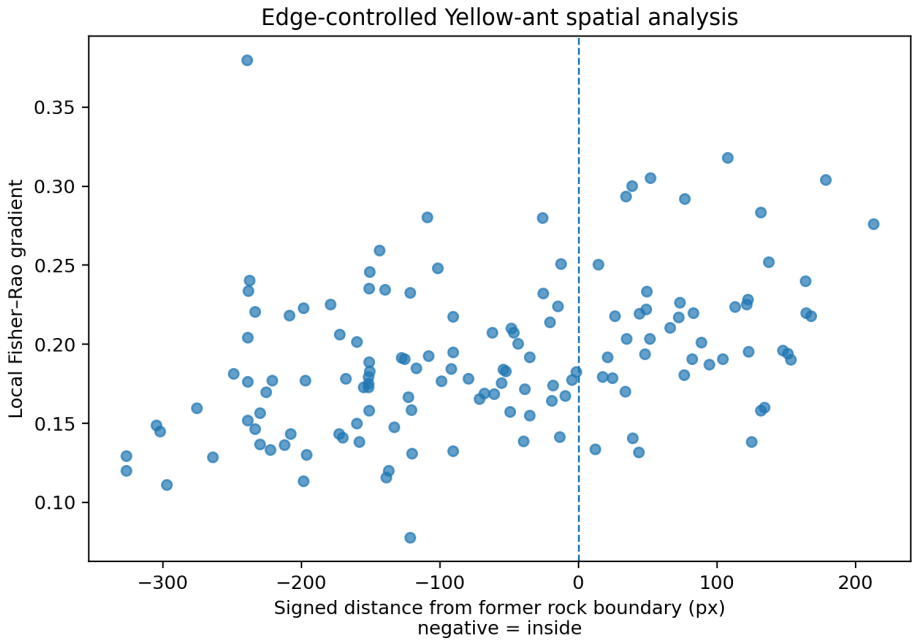



**Figure S9. Yellow-ant edge-control diagnostic.** Supplementary visualization of the tracking-domain exclusion used to assess broad rock-footprint differences while reducing the influence of elevated Fisher–Rao values close to the observational boundary.

## 5. Wall Annotation-Aware Analysis

Four visible surface regions (concrete wall, horizontal wood, vertical fence, and grass/ground) and three surface transitions were annotated independently of the Fisher–Rao field. Only three valid cells were assigned to grass/ground; this region is therefore retained for visualization but not used for quantitative regional interpretation.

**Table S4. Local Fisher–Rao heterogeneity by annotated surface in the Wall recording.**

| **Surface** | ***n* cells** | **Mean *G*** | **Median *G*** |
|---|---|---|---|
| Concrete wall | 80 | 0.270 | 0.238 |
| Horizontal wood | 40 | 0.512 | 0.486 |
| Vertical fence | 25 | 0.648 | 0.672 |
| Grass/ground | 3 | — | — |

**Table S5. Local Fisher–Rao heterogeneity within one spatial-cell width of annotated surface transitions.**

| **Annotated transition** | **Mean local *G*** |
|---|---|
| Horizontal wood / concrete | 0.359 |
| Concrete / grass | 0.336 |
| Vertical fence / horizontal wood | 0.621 |

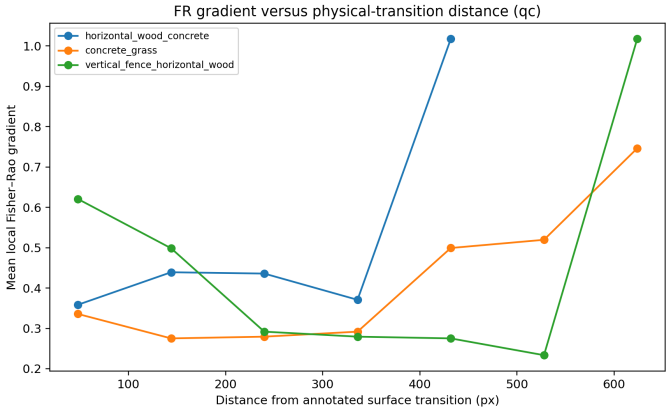



**Figure S10. Local Fisher–Rao heterogeneity as a function of distance from independently annotated surface transitions in the Wall recording.** Distance-binned values are shown for the annotated surface transitions after quality control. The curves are retained as a diagnostic because no universal monotonic relationship between local Fisher–Rao heterogeneity and distance from a surface transition was observed.

## 6. Reproducibility and BioTraj v1.2 Settings

The analyses in the main article were performed with BioTraj v1.2. The principal information-geometric command-line settings were:

```
python3 AntTrajectoryAnalysis.py ant_tracks.csv \
    --output BioTraj_V1_2_revision2 \
    --info-geometry \
    --info-grid-x 20 \
    --info-grid-y 12 \
    --info-angle-bins 18 \
    --info-min-samples 30 \
    --info-clusters 4 \
    --info-pseudocount 0 \
    --info-bootstrap 200 \
    --info-seed 12345 \
    --temporal-info-geometry \
    --temporal-window-frames 900 \
    --temporal-step-frames 900 \
    --temporal-min-samples 100 \
    --temporal-bootstrap 200 \
    --fps 60
```

The spatial and temporal information-geometric analyses generate numerical outputs including cell-level summaries, temporal-window summaries, and Fisher–Rao distance matrices. The BioTraj repository provides the software, documentation, configuration information, and example data required to reproduce and test the analysis workflow. The complete experimental trajectory datasets and original video recordings used in the present study are not included in full in the public repository and are available from the corresponding author upon reasonable request, as stated in the Data Availability Statement of the main article.

## 7. Interpretation of Supplementary Analyses

The supplementary analyses follow the same interpretation adopted in the main article. Local Fisher–Rao heterogeneity identifies spatial or temporal changes in estimated

movement distributions; it does not by itself determine their biological or physical cause. Differences may reflect biological responses, physical constraints, surface geometry, projection, tracking quality, finite sampling, or combinations of these factors. The experimental scene comparisons are therefore descriptive and the spatial cells are not treated as independent biological replicates.