Behavioral decoding of *Drosophila* locomotion in a circular arena

Running title: Locomotion decoding in *Drosophila*

Shuang Qiu[†] and Chengfeng Xiao[†,*]

[†] Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada

[*] Correspondence: xiao.c@queensu.ca (C.X)

Abstract

Motion detection and position tracking of animal behavior over a period of time produce massive amount of information, but analysis and interpretation of such huge datasets are challenging. Here we describe statistical methods to extract major movement structures of *Drosophila* locomotion in a circular arena, and examine the effects of pulsed light stimulation on these identified locomotor structures. *Drosophila* adults performed exploratory behavior when restrained individually in the circular arenas (1.27 cm diameter 0.3 cm depth). Measures of the distance to the center of the arena followed a gumbel distribution with mixed components. Representation learning of distance to center from 177,000 observations (from 63 controls and 55 flies with pulsed light stimulation) revealed three major movement components, indicating three locomotor structures characterized as: side-wall walk, angle walk and cross of the central region. Pulsed flies showed reduced cross of the central region compared with controls. We also showed that counter-clockwise walk and clockwise walk were the two major rotation behaviors with equal time proportion. There was a peak relative turning angle at 25.6° for counter-clockwise walk, and 334.4° for clockwise walk. Regression analysis of relative turning angle as a function of distance to center indicated that as distance to center increased, flies switched turning from directions for perimeter-returning to directions for major rotation. Pulsed flies reduced trajectories that had irregular circle-shape and increased trajectories with regular circle-shape during rotation. Taken together, we present a feasible approach to extract major locomotor structures of *Drosophila* locomotion in a circular arena, and demonstrate how pulsed light stimulation increased the regularity of locomotor trajectory.

Keywords:
animal behavior; locomotor structure; modeling; gumbel distribution; rotation behavior; locomotor regularity; *Drosophila melanogaster*

Author Summary

This work describes computational methods for the analysis and interpretation of *Drosophila* locomotion in an experimental setting. We present the good fit of gumbel distribution, a non-gaussian-based model, to the observations. We provide a method to extract locomotor structures based on individual parameters. In addition, we show that an external light stimulation increases the regularity of movement trajectories in adult *Drosophila*. We applied currently leading statistical techniques, including representation learning, modeling with finite mixture distribution, and non-parametric linear-circular regression. Our data provide a novel approach in the interpretation of behavioral structures of movement in *Drosophila*.

Introduction

The complexity of animal behavior has multiple causes, including genetic makeup, evolutionary constraints, and environmental impact. Cracking the codes of behavior provides insights in understanding animal internal physiological states as well as adaptive responses to external influence (Kays *et al*. 2015). The understanding of behavioral complexity in the biological perspective is hampered due to a lack of effective tools for data analysis and interpretation (Urbano *et al*. 2010). The situation is exacerbated when massive movement data are collected, deposited and shared worldwide using advanced mobile sensing technology (e.g. Global Positioning System) (Wikelski & Kays 2010). In behavioral studies using a small insect, *Drosophila melanogaster* (fruitfly), ethologists have developed advanced techniques and software for behavioral tracking (Fry *et al*. 2003; Branson *et al*. 2009; Uhlmann *et al*. 2016; Kumar *et al*. 2016; Gomez-Marin *et al*. 2012; Colomb *et al*. 2012; Geurten *et al*. 2014). These tools allow rapid and high-throughput data collection of flying or walking behavior of a group or individuals (Branson *et al*. 2009), movement of a body part (e.g. wings) (Fry *et al*. 2003; Geurten *et al*. 2014) or activity of a single leg segment (Uhlmann *et al*. 2016). However, there is an urgent need of developing statistical tools for sufficient data interpretation.

The information richness and biological significance of animal behavior are well-described. Animal behavior is often self-explanatory. Army ants *Eciton burchelli* form three traffic lanes with a home-inbound lane flanked by two outbound lanes to maximize traffic flow for food-hunting in a large population (Couzin & Franks 2003). Collective movement and coherent turning of bird flocks (Heppner 1974) and fish shoals (Pitcher 1986) have evolutionary advantages in avoiding predators and saving energy during migration. Caged laboratory mice consistently develop stereotypical wire-gnawing, a clear reflection of frustration and behavioral suffering (Würbel *et al*. 1996). In addition, behavior of wild animals are natural indicators of ecological and environmental fluctuations. Barnacle geese migrate by following a green wave of vegetation phenology for high quality food (Shariatinajafabadi *et al*. 2014). Non-migratory European butterflies shift their distributional ranges northwards due to regional warming (Parmesan *et al*. 1999). The ranging behavior of elephants closely match the patterns of greening and senescing of vegetation in their home area (Bohrer *et al*. 2014). Furthermore, the behaviors of laboratory animals are indicative of internal physiological states and the responsiveness to external stimulation. Phototaxis, geotaxis (Carpenter 1905), circadian rhythms of locomotor activity (Konopka & Benzer 1971) and rover/sitter foraging behavior (Osborne *et al*. 1997) are among the most prominent behavioral characteristics in fruitflies. *Drosophila* larvae leave food burrows under hypoxic conditions (Vermehren-Schmaedick *et al*. 2010). Adult fly increases boundary preference during locomotion in a circular arena after pulsed light stimulation (Qiu *et al*. 2016). Collective escape behavior can be induced by CO_2 exposure, and can be conveyed by a single touch of leg appendage between *Drosophila* individuals (Ramdya *et al*. 2015).

Whereas massive movement data have been already obtained (Wikelski & Kays 2010), behavioral statistics and computational biology are just about starting (Kays *et al*. 2015; Markowetz 2017). A working strategy is that complex animal behavior could be cracked down to identifiable and representative elements termed movement modules. A behavior is thus deemed the combination of sequential movement modules with predictable transition probabilities. This strategy has been applied in mapping sub-second scale of 3D pose structures in mouse behavior (Wiltschko *et al*. 2015). A similar method is used to decode diving behavior of an Imperial Cormorant (Grundy

et al. 2009) and seals (Bestley *et al*. 2015), the activities of cows (Schwager *et al*. 2007), and behavioral patterns of albatross movement (Torres *et al*. 2017). *Drosophila* wild-type larvae when placed in an agar arena (5 cm diameter) apparently alternate between straight move and pause (Suster *et al*. 2003). These two locomotor patterns can be sufficiently modeled and interpreted using a single parameter of changes of direction (Holzmann *et al*. 2006). Freely walking adult flies show three main behavioral patterns (resting, translation and rotation) (Geurten *et al*. 2014). However, statistical methods for robust modeling and classification of *Drosophila* movement are lacking.

We have recently developed software for tracking fly positions in circular arenas (Xiao & Robertson 2015), and collected a large amount of movement data. In this study, we focus on behavioral analysis and interpretation of walking activity of adult *Drosophila*. By combining the techniques of fly tracking, representation learning and advanced statistical modelling, we show here that *Drosophila* movement in a circular arena could be structurally extracted and interpreted by single or multiple parameters. In addition, using these methods we analyze how pulsed light stimulation increases the regularity of locomotor trajectories.

Results

Locomotor trajectory of adult fly in a circular arena

A five-day-old male fly of w1118 performed exploratory locomotion when loaded into a circular arena (1.27 cm diameter and 0.3 cm depth). The 60 s walking trajectory, represented as connected time-series of position every 0.2 s, showed a coil-like shape with irregularity. Similar features were observed for different individuals (Figure 1a). In contrast, the 60 s trajectories of flies which had been subjected to pulsed light stimulation showed coiled shape with visually improved regularity (Figure 1b). 2D plots of x, y-positions indicated that, whereas control flies performed substantial center crossing in addition to perimeter preference, pulsed flies greatly reduced center crossing during locomotion (Figure S1). These observations were consistent with a previous report of increased boundary preference induced by pulsed light stimulation in w1118 males (Qiu *et al*. 2016).

Distance to center followed gumbel distribution

Each walking trajectory contained complex information about locomotion including, for example, the time-series of distance to center, rotation orientation and relative turning angle. We first examined the distribution of time-series of distance to center, because this parameter could be indicative of location preference or walking mode. A two-step procedure, model selection and parameter estimation (Stasinopoulos *et al*. 2017), was conducted for appropriate and adequate modeling of distance to center.

A total of 1500 observations (the values of distance to center every 0.2 s) for each fly were used. We considered several models of continuous distributions: gumbel, weibull, gamma, reverse gumbel, logistic, normal and *t* family distributions (Rigby *et al*. 2014). A generalized Akaike information criterion (GAIC) (Akaike 1983) was applied for model selection. By fitting each of the

selected models to data, we found that a gumbel distribution was favored by GAIC for better description of distance to center in a control (Figure 2a) or a pulsed fly (Figure 2b). The appropriate fit of gumbel compared with weibull or gamma distribution was supported by a diagnostic tool of worm plot (Buuren & Fredriks 2001). Modeled with gumbel distribution, the deviations of residuals were well distributed within 95 $\%$ confident intervals but not apparently tilted, U-shaped, inverted U-shaped or S-shaped (Figure 2c and 2d), indicating adequate modeling of parameters (location and scale), skewness and kurtosis of data.

The same process of model selection was applied to multiple individuals. Gumbel distribution was favored for modeling the distance to center in 92.1 $\%$ (58/63) controls and 80.0 $\%$ (44/55) pulsed flies (Table S1). The percentages for controls and pulsed flies were statistically the same ($P > 0.05$, Fisher's exact test). Therefore, the distance to center of either control or pulsed flies could be sufficiently modeled and interpreted by gumbel distribution.

Main patterns of locomotion in a circular arena

We next asked whether there were identifiable components or structures for the observations of distance to center. Similar to the process of representation learning (Bengio *et al*. 2013), a gumbel distribution with 1 - 3 mixed components was fitted to a relatively large dataset, which contained 177,000 observations from 63 controls and 55 pulsed flies. Parameters from the model giving lowest GAIC (Stasinopoulos *et al*. 2017) were extracted.

A three-component gumbel distribution was supported. This model gave rise to location parameter μ at 5.30, 5.04 and 3.05, scale parameter σ at 0.18, 0.36 and 0.79, mixture probability π at 0.43, 0.49 and 0.08 (Figure S2), and posterior probabilities by which every actual observation could be classified to the most probable component. These parameters were essential in defining potential locomotor patterns common to control and pulsed flies.

We aligned the modeled values along with actual observations of distance to center, and extracted the corresponding video frames for visual validation. The modeled values with lowest μ (at 3.05) were the precise representations of observation with small distance to center (Figure 3a). Relevant fly positions extracted from video were indicative of the activities crossing over the central region of arena (Figure 3b). The activity of cross typically last around 1 s, starting from and ending at the perimeter of arena. We used a term "cross" to describe such activities classified by the model component with lowest μ.

Three additional locomotor structures: "side-wall walk", "angle walk" and "stop", were classified. "Side-wall walk" was descriptive of walking activities with highest μ (at 5.30), corresponding to fly positions on the extreme: the side wall of arena. In each of these cases fly was positioned laterally to the camera (Figure 3b). For a conservative estimate, a criterion of at least five consecutive positions on the side wall was applied to define a side-wall walk. "Angle walk" was descriptive of walking activities with μ at 5.04, representing two situations: (1) a fly was near the edge of arena and on the floor or ceiling but not on the side wall, and (2) body main axis often formed an acute angle to the side wall during walking. "Stop" was a description of locomotion with a step size no greater than a threshold value (1.41 pixels or 0.27 mm) and which persisted for at least 1 s (5 steps). Most stops last for less than 5 s (Figure 3b). These four structures ("cross", "side-wall walk", "angle walk" and "stop") together were sufficient to explain most observations of locomotion, leaving < 10 $\%$ data unclassified (Figure 4).

Reduced cross in flies with pulsed light stimulation

We examined how classified locomotor structures were modified by pulsed light stimulation. The activities of cross of individual flies from controls (n = 63) and pulsed flies (n = 55) are illustrated (Figure 4a). Within 300 s locomotion, time proportion for cross in controls (median 0.19, interquartile range (IQR) 0.14 – 0.24) was higher than that in pulsed flies (median 0.09, IQR 0.06 – 0.13) ($P < 0.05$, Mann-Whitney test). The median number of cross in controls (median 61, IQR 43 – 72) was greater than that in pulsed flies (median 29, IQR 19 – 38) ($P < 0.05$, Mann-Whitney test). Thus, pulsed light illumination reduced time proportion for cross and reduced number of cross during 300 s locomotion (Figure 4b). Time proportion for side-wall walk in pulsed flies (median 0.31, IQR 0.24 - 0.44) was comparable with that in controls (median 0.30, IQR 0.25 - 0.35) with insignificant difference ($P > 0.05$, Mann-Whitney test). Time proportion for angle walk in pulsed flies (median 0.44, IQR 0.36 - 0.52) was also comparable with that in controls (median 0.42, IQR 0.36 - 0.46) ($P > 0.05$, Mann-Whitney test). There was no significant difference of time proportion for stop between controls (median 0.08, IQR 0.04 - 0.11) and pulsed flies (median 0.07, IQR 0.04 - 0.11) ($P > 0.05$, Mann-Whitney test). Additionally, time proportions for unclassified activities in controls (median 0.10, IQR 0.09 - 0.12) and pulsed flies (median 0.08, IQR 0.07 - 0.10) were comparable ($P > 0.05$, Mann-Whitney test) and relatively small (Figure 4c).

Counter-clockwise walk and clockwise walk were two main rotation behavior

Flies circled at the perimeter while walking in the arenas. The parameter distance to center was insufficient to characterize circling behavior, although it provided information for modelling location distribution. We next examined rotation orientation by analyzing ω, the angular displacement per 0.2 s.

We computed ω from 1500 observations (corresponding to 300 s locomotion) of each fly. The rotation orientation of 24 controls and 24 pulsed flies were illustrated (Figure 5a). It was observed that counter-clockwise walk and clockwise walk were the main rotation behavior in both groups. Time proportion for counter-clockwise walk and clockwise walk were examined. In control flies, time proportion for counter-clockwise walk (median 0.37, IQR 0.34 – 0.42) was comparable to that for clockwise walk (median 0.37, IQR 0.34 – 0.40) ($P > 0.05$, Mann-Whitney test). In pulsed flies, time proportion for counter-clockwise walk (median 0.41, IQR 0.35 – 0.47) and clockwise walk (median 0.41, IQR 0.37 – 0.49) were also similar ($P > 0.05$, Mann-Whitney test) (Figure 5b). We thus observed no preference of rotation orientation in either controls or pulsed flies.

Time proportion for both rotations (counter-clockwise and clockwise walk) was 0.74 (IQR 0.71 – 0.78) in controls and 0.85 (IQR 0.80 – 0.88) in pulsed flies. There was a significant increase of time proportion for both rotations in pulsed flies compared to controls ($P < 0.05$, Mann-Whitney test) (Figure 5c). We further scored the numbers of rotation in controls and pulsed flies. Within 300 s locomotion, median number of counter-clockwise walk was 39 (IQR 34 – 43) in controls and 29 (IQR 21 – 35) in pulsed flies. Median number of clockwise walk was 38 (IQR 33 – 43) in controls and 27 (IQR 22 – 37) in pulsed flies. Numbers of counter-clockwise walk and numbers of clockwise walk were both reduced in pulsed flies (Figure 5d and 5e).

Together, counter-clockwise walk and clockwise walk were the two main rotation behaviors in both control and pulsed flies. Pulsed light stimulation induced increased time proportion for

rotation behavior and decreased numbers of counter-clockwise and clockwise walks.

Increased density of rotation at peak turning angle in pulsed flies

To further characterize how flies turned during counter-clockwise walk and clockwise walk, we analyzed relative turning angle θ, a parameter showing instant change of turning direction at current position based on three observations: current (*i*), immediately past (*i-1*) and immediately future (*i+1*) relocations (Calenge 2006).

Nonparametric circular methods (Oliveira *et al*. 2014) were applied for modelling θ. We observed that during counter-clockwise walk flies turned left most of the time. There was a peak θ at 27.4° for a control and 23.8° for a pulsed fly. During clockwise walk flies mostly turned right. There was a peak θ at 336.2° for a control and 336.9° for a pulsed fly (Figure 6a). The schemed relationships between ω and θ for counter-clockwise walk and clockwise walk were illustrated (Figure 6b).

During counter-clockwise walk, there was no significant difference of turning angle between controls (median 26.0°, IQR 23.8 – 28.1°) and pulsed flies (median 25.3°, IQR 23.1 – 28.1°) ($P >$ 0.05, Mann-Whitney test) (Figure 6c). During clockwise walk, there was no significant difference of turning angle between controls (median 334.8°, IQR 331.9 – 336.9°) and pulsed flies (median 334.0°, IQR 332.6 – 336.2°) ($P >$ 0.05, Mann-Whitney test) (Figure 6d). Therefore, pulsed light stimulation had no effect on relative turning angle θ. The modeled values of peak θ, 25.6° for counter-clockwise walk and 334.4° for clockwise walk, for overall observations from controls and pulsed flies are provided (Figure 6b).

Relative density at peak θ was increased in pulsed flies (median 1.53, IQR 1.41 – 1.68) compared with controls (median 1.18, IQR 1.02 – 1.32) during counter-clockwise walk ($P <$ 0.05, Mann-Whitney test) (Figure 6e). Relative density at peak θ was also increased in pulsed flies (median 1.49, IQR 1.32 – 1.64) compared with controls (median 1.17, IQR 1.04 – 1.33) during clockwise walk ($P <$ 0.05, Mann-Whitney test) (Figure 6f). Thus, pulsed light stimulation increased relative densities of rotation behavior at peak turning angles.

Turning behavior as a response to distance to center

Classification of locomotor structures based on single parameter was straightforward, but the explanatory power was limited. Unraveling the dependence between multiple parameters, for example, turning behavior as a function of distance to center, would be critical to understand the dynamic changes of locomotor structures from one to another. The nature of high boundary preference during locomotion (Liu *et al*. 2007; Xiao & Robertson 2015) implied that *Drosophila* rapidly adapted to the arena and decided the walking direction as a response to the location. We modeled the relation between relative turning angle and distance to center with respect to rotation behavior. The estimation was performed using nonparametric linear-circular regression with smoothing parameter computed by plug-in rule (Oliveira *et al*. 2014).

During counter-clockwise walk, as distance to center increased, flies changed turning directions from a tendency of right-turn to a concentrated left-turn. In contrast, during clockwise walk, as distance to center increased, flies changed turning directions from a tendency of left-turn to a concentrated right-turn. Both controls and pulsed flies showed such relationships between turn-

ing direction and distance to center (Figure 7a). Distances further to arena center were clearly associated with concentrated turning, in particular, left-turn for counter-clockwise walk and right-turn for clockwise walk. Further distances to center (with gumbel location parameter at 5.30 or 5.04) were also associated with side-wall walk and angle walk (see Figure 3). Hence, during side-wall walk and angle walk flies were concentrated at major turning directions, either left or right. Distances closer to center were associated with the trend of opposite turning. Closer distance to center was indicative of a locomotor structure cross. A cross with turning direction opposite to concentrated turning would promote quick return of fly location to the perimeter.

We therefore schematically illustrated the turning behavior as a response to fly location with visual proofs of actual video frames. We depicted these turning behavior: counter-clockwise walk with right-turning cross; clockwise walk with left-turning cross; counter-clockwise walk without cross; clockwise walk without cross. Trajectory of counter-clockwise walk with right-turning cross and clockwise walk with left-turning cross were irregularly circle-shaped, whereas trajectory of counter-clockwise walk without cross and clockwise walk without cross appeared regular in circle-shape (Figure 7b).

Pulsed light stimulation increased regularity of rotation behavior

During 300 s locomotion, the number of counter-clockwise walk with right-turning cross in pulsed flies (median 5, IQR 3 - 8) was lower than that in controls (median 12, IQR 9 - 15) ($P < 0.0001$, Man-Whitney test). The number of clockwise walk with left-turning cross in pulsed flies (median 6, IQR 3 - 8) was also lower than that in controls (median 12, IQR 8 - 16) ($P < 0.0001$, Man-Whitney test) (Figure 7c). Thus, pulsed light stimulation reduced rotations with irregular circle-shape. Within 300 s, time proportion for counter-clockwise walk without cross in pulsed flies (median 0.23, IQR 0.18 - 0.32) was higher than that in controls (median 0.19, IQR 0.16 - 0.23) ($P < 0.001$, Man-Whitney test). Time proportion for clockwise walk without cross in pulsed flies (median 0.26, IQR 0.20 - 0.33) was also higher than that in controls (median 0.18, IQR 0.15 - 0.22) ($P < 0.0001$, Man-Whitney test) (Figure 7d). Therefore, pulsed light stimulation led to increased time proportion for rotations with regular circle-shape.

Discussion

Large amounts of movement data of *Drosophila* walking and flying behavior have been collected, but the tools for robust analysis and interpretation of such big datasets are underdeveloped. We describe statistical methods for modeling data distribution and extracting movement structures of *Drosophila* walking activities. We show how pulsed light stimulation modifies locomotor structures and increases the regularity of walking trajectory of adult fly in a circular arena.

A gumbel distribution with mixed components is most sufficient to describe the observations of distance to center. Gumbel distribution belongs to a family of generalized extreme value distribution, which differs from normal or bell-shaped Gaussian distributions. Gumbel distribution has been applied to model extreme or rare events, for examples, extreme floods (Gumbel 1941), or fastest record of men's 100-meter sprint within the next 100 years (Baum & McKelvey 2006). In R and MATLAB, gumbel distribution has negative skewness, suitable for modeling minimum extreme. The good fit of gumbel distribution to the observations of distance to center raises a critical

issue: the observations without any exclusion do not violate gumbel properties of negative skewness and asymptotic behavior. It has been common that data treatment technique such as block maxima/minima, *r*-largest/smallest or threshold exceedence is applied to exclude non-extreme values that might disfavor asymptotic model (Baum & McKelvey 2006). There is no literature previously documenting that without data exclusion the complete time-series of observations support a generalized extreme value distribution.

Unlike Lévy or Brownian foraging movement of open-ocean predators (Humphries *et al*. 2010), fly movement in a circular arena shows a clear preference to the boundary with around 60 $\%$ of the time spent on the perimeter (Xiao & Robertson 2015). Percent time on perimeter is increased if flies are pre-conditioned with pulsed light stimulation (Qiu *et al*. 2016). Both controls and pulsed flies spend less than a half of exploratory time to relocate off the perimeter. This could explain the good fit of gumbel distribution because locations off the perimeter could serve as negatively-tailed, rare or extreme events. Normal, logistic and *t* family distributions have zero skewness, thus are inappropriate for modeling skew data. Gamma and reverse gumbel distributions are for positive skew data, thus also inappropriate. We found that weibull distribution is favored for modeling the observations from only a small proportion of flies.

Although application of gumbel distribution has been successful in approximating the extreme event, little attention has been paid to take full advantage of gumbel location parameter μ and scale parameter σ. Additionally, by breaking down to several homogeneous gumbel components, the complexity of distance to center could be better described than by single-component distribution. We report here three movement representations generated by finite-mixture modeling (Stasinopoulos *et al*. 2017) from a relative large dataset, including those from both controls and pulsed flies. Each representation is approximated by a gumbel component with unique μ, σ and mixing probability π. These movement representations are indicative of three locomotor structures: side-wall walk, angle walk and cross. Together with stop, these locomotor structures are sufficient to classify around 90 $\%$ of the observations. Such a locomotor classification leads us to understand the subtle changes of locomotor structure induced by pulsed light stimulation: pulsed flies have decreased time proportion for cross and decreased numbers of cross during the observation time.

Drosophila rotation orientation and instant turning activity in a circular arena are relatively straightforward to be modeled and described. We report that counter-clockwise walk and clockwise walk are the two major rotation behavior, which together account for a large time proportion throughout the observation duration. The equal time-proportions for counter-clockwise walk and clockwise walk indicate that there is no preference of rotation orientation of *Drosophila* locomotion in a circular arena. Additionally, there is a peak turning angle at 25.6° for counter-clockwise walk, and 334.4° for clockwise walk. These peaks are unaffected by pulsed light stimulation. The symmetry of peak turning angles along straight direction indicates the co-existence of two equal-chanced opposite rotation behavior.

Pulsed light stimulation induces increased relative density at the peaks of turning angle without changing the peaks. It suggests that pulsed flies become concentrated on counter-clockwise walk or clockwise walk and reduce the turning at angles other than the peaks. This is consistent with the reduction of cross of arena center in pulsed flies, because a fly might have high freedom to turn at any angle during the cross while there is not much option to change turning angle during side-wall walk and angle walk.

Modeling the relation between relative turning angle and distance to center has provided

further explanation how flies turn in relation to locations. There are two peak turning directions when a fly relocates on the perimeter: turning left for counter-clockwise walk and turning right for clockwise walk. However, a fly tends to turn at directions opposite to perimeter-associated peak directions when it is near the center. Therefore, during counter-clockwise walk, a departure from perimeter, rendering small distance to center, is associated with a trend of turning right. This behavior facilitates return to perimeter, although it causes irregular circle-shape of walking trajectory. Similarly, during clockwise walk, a departure from perimeter associated with a trend of turning left promotes quick return to perimeter, leaving also an irregular trajectory. Pulsed flies reduce numbers of rotation with irregular circle-shape and increase time proportion for rotation with regular circle-shape. Thus, pulsed flies show increased regularity of locomotor trajectories in the circular arenas.

Methods

Flies

The white-eyed strain w1118 (L. Seroude laboratory) was used in the current study. This strain has been widely used as a genetic background to construct transgenic flies in the *Drosophila* laboratories. Flies were maintained with standard medium (cornmeal, agar, molasses and yeast) at room temperatures of 21-23 $^{\circ}$C with 60-70 $\%$ relative humidity. An illumination of light$/$dark (12$/$12 hr) cycle was provided with three light bulbs (Philips 13 W compact fluorescent energy saver) in a room around 133 square feet. Male flies were collected for experiments within 0-2 days after emergence. We used pure nitrogen gas to anesthetize flies during collection. Collected flies were raised in food vials at a density of 20-25 flies per vial for at least three additional days. A minimum of three days free of nitrogen exposure was guaranteed before testing. The ages of tested flies were 3-9 days old. To avoid natural peak activities in the mornings and evenings (Grima *et al*. 2004), experiments were performed during the light time with three hours away from light on$/$off switch.

Pulsed light stimulation

The procedures of pulsed light stimulation has been previously described (Qiu *et al*. 2016). Briefly, freshly laid embryos of w1118 were subject to pulsed light illumination (continuous cycle of 5 s light on and 15 s light off during the light time) throughout the life cycle. Pulsed light was provided in a light box with LED light strip (DC 12V Rxment 5050 SMD RGB LED) powered by an electrical pulse generator (Grass S88 Stimulator). Male flies were collected at the age of 0-2 days, and raised under pulsed light illumination for at least three additional days. Flies received pulsed light stimulation were named ”pulsed” flies. The remaining raising conditions (e.g. food recipe, temperature and humidity) and collection procedures were the same as flies under regular conditions (”control” flies). A period of 1 h before test was allotted for the adaptation of pulsed flies to constant light illumination. Pulsed and control flies were examined simultaneously during the light time.

Locomotor assay

Locomotor assay was performed by following a protocol (Xiao & Robertson 2015). In general, individual flies were loaded into small circular arenas (1.27 cm diameter and 0.3 cm depth). The depth of 0.3 cm was considered to allow flies to turn around but suppress vertical movement. We machined 128 arenas (8 $\times$ 16) in an area of 31 $\times$ 16 cm^2 Plexiglas. The bottom side of arena was covered with thick filter paper allowing air circulation. The top was covered by a slidable Plexiglas with holes (0.3 cm diameter) at one end for fly loading. The Plexiglas with arenas was secured in a large chamber (48.0 $\times$ 41.5 $\times$ 0.6 cm^3). A flow of room air (2 L/min) was provided to remove the effect of dead space (Bouhuys 1964). Locomotor activities were video-captured for post analysis. Fly positions (the locations of center of mass) with 0.2 s interval were computed by custom-written fly tracking software (Xiao & Robertson 2015). Subsequent movement analysis, including trajectory reconstruction and data modeling, were based on the values of center of mass, regardless of fly posture (e.g. dorsal, ventral or lateral view to camera). For each fly, a data containing 1500 position information, corresponding to 300 s locomotion, was used for statistical analysis.

Modeling the distribution of distance to center

The statistical modeling process, including model selection and parameter estimation, was conducted as suggested (Stasinopoulos *et al*. 2017). The main procedures were as below.

(1) Calculate the distance to center at each time point using x, y-positions. The x, y-positions were calibrated according to the center of arena (0, 0), so that the distance to center was calculated by a simplified Euclidean distance formula $d = \sqrt{x^2 + y^2}$.

(2) Select the ”best” model from candidate distributions using generalized Akaike information criterion (GAIC) (Akaike 1983; Stasinopoulos *et al*. 2017). Because flies preferred boundary to center region of arena (Liu *et al*. 2007; Xiao & Robertson 2015), a distribution with negative skewness (e.g. Gumbel distribution) was preferably considered. We initially compared the fit of Gumbel, Weibull, Gamma, Reverse Gumble, Logistic, Normal and *t* family distributions, all of which have been fully described (Rigby *et al*. 2014). We fitted three-component homogeneous mixtures of each selected distribution to the data by gamlssMX() function from an R package (Rigby & Stasinopoulos 2005). The distribution giving lowest GAIC was chosen as the best model. By following these procedures we selected the best model using a standard criterion, thus avoided a guess or a blind assumption of data distribution.

(3) Identify latent components and estimate parameters for each component. The function of gamlssMX() was applied to fit finite mixtures (K = 1-3) of selected distribution to observations. The Expectation-Maximization (EM) algorithm (Dempster *et al*. 1977) was used to maximize the likelihood function for a distribution with given K (number of component). The best K was determined by GAIC (Stasinopoulos *et al*. 2017). Parameters for each component, for example, location, scale and estimated probabilities, were output by gamlssMX().

Analysis of rotation orientation

The parameter ω (angular displacement per 0.2 s) was calculated by trigonometric function atan2(y, x), where x, y were the calibrated position coordinations from locomotor assay. To

avoid the big jump of ω value due to radian rotation, we calculated ω twice using radian interval (0, 2π] and (-π, π], and chose one with smaller absolute value. Rotation orientation of a fly at each location was determined by the reference of immediate next location, and was classified as counter-clockwise ($\omega > 0$) or clockwise ($\omega < 0$) rotation.

We defined "a counter-clockwise walk" as at least five consecutive relocations with $\omega > 0$, and "a clockwise walk" at least five consecutive relocations with $\omega < 0$. Classified data were further treated to allow single hesitation step (i.e. pausing or backward walking with radian $<$ 0.05) without apparent turning around. In a pretest, a control fly had a mixture of many episodes of counter-clockwise walk or clockwise walk during 300 s locomotion.

Analysis of relative turning angle θ

Relative turning angle θ was an indication of turning direction (left or right) from current location (Calenge 2006). Turning left could be mathematically described as $\theta \in$ (0, π), and turning right $\theta \in$ (π, 2π). θ was obviously different from ω, the latter an angular displacement between two polar angles. A fly may choose counter-clockwise walk ($\omega > 0$) with the freedom of turning left or right. A combination of θ and ω would indicate locomotor complexity of rotation orientation and turning direction. The function as.ltraj() from an R package "adehabitatLT" was used to compute θ (Calenge *et al*. 2009).

Modeling relative turning angle θ as a function of distance to center

The common phenomenon of boundary preference (Liu *et al*. 2007; Xiao & Robertson 2015) indicated that flies chose specific turning behavior according to location information. For example, when located on the boundary of an open-field arena, flies performed wall-following activities. When located in the open region, flies performed activities with directional persistence (Soibam *et al*. 2012). Here we treated θ as a response variable and distance to center an explanatory variable, and modeled the relation between θ and distance to center using nonparametric linear - circular statistical methods (Oliveira *et al*. 2014). In addition, to reduce data dimension we split actual observations by rotation orientation. The reasons were, (1) that counter-clockwise walk and clockwise walk were the two major rotation behavior in the circular arena from preliminary analysis; (2) that flies had equal time proportion to perform counter-clockwise walk and clockwise walk; and (3) to prevent the neutralization of relationship because of the co-exist of two equal-chanced opposite rotation behavior.

Statistics

We used statistical software R (R Core Team 2014) and the following R packages: gamlss and gamlss.mx (Stasinopoulos *et al*. 2007), gdata, lattice (Sarkar 2008), adehabitatLT (Calenge *et al*. 2009), circular (Jammalamadaka & Sengupta 2001), and NPCirc (Oliveira *et al*. 2014). Data normality was examined by D'Agostino $\&$ Pearson omnibus normality test. Nonparametric tests were performed for the comparison of medians. Statistical data were illustrated as scatter dot plot. A $P < 0.05$ was considered significant difference.

References

Akaike, H. (1983). Information measures and model selection. *Bulletin of the International Statistical Institute* **50**, 277–291.

Baum, J. A. & McKelvey, B. (2006). Analysis of extremes in management studies. *in* 'Research methodology in strategy and management'. Emerald Group Publishing Limited. pp. 123–196.

Bengio, Y., Courville, A. & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* **35**, 1798–1828.

Bestley, S., Jonsen, I. D., Hindell, M. A., Harcourt, R. G. & Gales, N. J. (2015). Taking animal tracking to new depths: synthesizing horizontal–vertical movement relationships for four marine predators. *Ecology* **96**, 417–427.

Bohrer, G., Beck, P. S., Ngene, S. M., Skidmore, A. K. & Douglas-Hamilton, I. (2014). Elephant movement closely tracks precipitation-driven vegetation dynamics in a kenyan forest-savanna landscape. *Movement Ecology* **2**, 2.

Bouhuys, A. (1964). Respiratory dead space. Section 3: Respiration. *Handbook of Physiology* **1**, 699.

Branson, K., Robie, A. A., Bender, J., Perona, P. & Dickinson, M. H. (2009). High-throughput ethomics in large groups of *Drosophila*. *Nat Methods* **6**, 451–457.

Buuren, S. v. & Fredriks, M. (2001). Worm plot: a simple diagnostic device for modelling growth reference curves. *Statistics in medicine* **20**, 1259–1277.

Calenge, C. (2006). The package "adehabitat" for the r software: a tool for the analysis of space and habitat use by animals. *Ecol Model* **197**, 516–519.

Calenge, C., Dray, S. & Royer-Carenzi, M. (2009). The concept of animals' trajectories from a data analysis perspective. *Ecol Inform* **4**, 34–41.

Carpenter, F. W. (1905). The reactions of the pomace fly (*Drosophila ampelophila* Loew) to light, gravity, and mechanical stimulation. *Am Nat* **39**, 157–71.

Colomb, J., Reiter, L., Blaszkiewicz, J., Wessnitzer, J. & Brembs, B. (2012). Open source tracking and analysis of adult *Drosophila* locomotion in buridan's paradigm with and without visual targets. *PLoS One* **7**, e42247.

Couzin, I. D. & Franks, N. R. (2003). Self-organized lane formation and optimized traffic flow in army ants. *Proceedings of the Royal Society of London B: Biological Sciences* **270**, 139–146.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Series B Stat Methodol* pp. 1–38.

Fry, S. N., Sayaman, R. & Dickinson, M. H. (2003). The aerodynamics of free-flight maneuvers in drosophila. *Science* **300**, 495–498.

Geurten, B. R., Jähde, P., Corthals, K. & Göpfert, M. C. (2014). Saccadic body turns in walking drosophila. *Frontiers in behavioral neuroscience* **8**, 365.

Gomez-Marin, A., Partoune, N., Stephens, G. J. & Louis, M. (2012). Automated tracking of animal posture and movement during exploration and sensory orientation behaviors. *PloS one* **7**, e41642.

Grima, B., Chélot, E., Xia, R. & Rouyer, F. (2004). Morning and evening peaks of activity rely on different clock neurons of the *Drosophila* brain. *Nature* **431**, 869–873.

Grundy, E., Jones, M. W., Laramee, R. S., Wilson, R. P. & Shepard, E. L. (2009). Visualisation of sensor data from animal movement. *in* 'Computer Graphics Forum'. Vol. 28. Wiley Online Library. pp. 815–822.

Gumbel, E. J. (1941). Probability-interpretation of the observed return-periods of floods. *Eos, Transactions American Geophysical Union* **22**, 836–850.

Heppner, F. H. (1974). Avian flight formations. *Bird-Banding* **45**, 160–169.

Holzmann, H., Munk, A., Suster, M. & Zucchini, W. (2006). Hidden markov models for circular and linear-circular time series. *Environ Ecol Stat* **13**, 325–347.

Humphries, N. E., Queiroz, N., Dyer, J. R., Pade, N. G., Musyl, M. K., Schaefer, K. M., Fuller, D. W., Brunnschweiler, J. M., Doyle, T. K., Houghton, J. D. *et al.* (2010). Environmental context explains lévy and brownian movement patterns of marine predators. *Nature* **465**, 1066–1069.

Jammalamadaka, S. R. & Sengupta, A. (2001). *Topics in circular statistics*. Vol. 5. World Scientific.

Kays, R., Crofoot, M. C., Jetz, W. & Wikelski, M. (2015). Terrestrial animal tracking as an eye on life and planet. *Science* **348**, aaa2478.

Konopka, R. J. & Benzer, S. (1971). Clock mutants of drosophila melanogaster. *Proceedings of the National Academy of Sciences* **68**, 2112–2116.

Kumar, S. S., Sun, Y., Zou, S. & Hong, J. (2016). 3d holographic observatory for long-term monitoring of complex behaviors in drosophila. *Sci Rep* **6**, 33001.

Liu, L., Davis, R. L. & Roman, G. (2007). Exploratory activity in *Drosophila* requires the kurtz nonvisual arrestin. *Genetics* **175**, 1197–212.

Markowetz, F. (2017). All biology is computational biology. *PLoS Biology* **15**, e2002050.

Oliveira, M., Crujeiras, R. M. & Rodríguez-Casal, A. (2014). NPCirc: An R package for nonparametric circular methods. *J Stat Softw* **61**, 1–26.

Osborne, K. A., Robichon, A., Burgess, E., Butland, S., Shaw, R. A., Coulthard, A., Pereira, H. S., Greenspan, R. J. & Sokolowski, M. B. (1997). Natural behavior polymorphism due to a cGMP-dependent protein kinase of *Drosophila*. *Science* **277**, 834–6.

Parmesan, C., Ryrholm, N., Stefanescu, C., Hill, J. K. *et al.* (1999). Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature* **399**, 579.

Pitcher, T. J. (1986). Functions of shoaling behaviour in teleosts. *in* 'The behaviour of teleost fishes'. Springer. pp. 294–337.

Qiu, S., Xiao, C. & Robertson, R. M. (2016). Pulsed light stimulation increases boundary preference and periodicity of episodic motor activity in *Drosophila melanogaster*. *PloS one* **11**, e0163976.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Ramdya, P., Lichocki, P., Cruchet, S., Frisch, L., Tse, W., Floreano, D. & Benton, R. (2015). Mechanosensory interactions drive collective behaviour in drosophila. *Nature* **519**, 233–236.

Rigby, B., Stasinopoulos, M., Heller, G. & Voudouris, V. (2014). The distribution toolbox of gamlss. *The GAMLSS Team* .

Rigby, R. A. & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **54**, 507–554.

Sarkar, D. (2008). *Lattice: multivariate data visualization with R*. Springer Science & Business Media.

Schwager, M., Anderson, D. M., Butler, Z. & Rus, D. (2007). Robust classification of animal tracking data. *Computers and Electronics in Agriculture* **56**, 46–59.

Shariatinajafabadi, M., Wang, T., Skidmore, A. K., Toxopeus, A. G., Kölzsch, A., Nolet, B. A., Exo, K.-M., Griffin, L., Stahl, J. & Cabot, D. (2014). Migratory herbivorous waterfowl track satellite-derived green wave index. *PloS One* **9**, e108331.

Soibam, B., Goldfeder, R. L., Manson-Bishop, C., Gamblin, R., Pletcher, S. D., Shah, S., Gunaratne, G. H. & Roman, G. W. (2012). Modeling *Drosophila* positional preferences in open field arenas with directional persistence and wall attraction. *PLoS One* **7**, e46570.

Stasinopoulos, D. M., Rigby, R. A. *et al.* (2007). Generalized additive models for location scale and shape (gamlss) in r. *J Stat Softw* **23**, 1–46.

Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z., Voudouris, V. & De Bastiani, F. (2017). *Flexible Regression and Smoothing: Using GAMLSS in R*. CRC Press. chapter Model selection techniques, pp. 253–305.

Suster, M. L., Martin, J.-R., Sung, C. & Robinow, S. (2003). Targeted expression of tetanus toxin reveals sets of neurons involved in larval locomotion in drosophila. *Journal of neurobiology* **55**, 233–246.

Torres, L. G., Orben, R. A., Tolkova, I. & Thompson, D. R. (2017). Classification of animal movement behavior through residence in space and time. *PloS One* **12**, e0168513.

Uhlmann, V., Ramdya, P., Delgado-Gonzalo, R., Benton, R. & Unser, M. (2016). Flylimbtracker: an active contour based approach for leg segment tracking in unmarked, freely behaving *Drosophila*. *bioRxiv* p. 089714.

Urbano, F., Cagnacci, F., Calenge, C., Dettki, H., Cameron, A. & Neteler, M. (2010). Wildlife tracking data management: a new vision. *Philos Trans R Soc Lond B Biol Sci* **365**, 2177–2185.

Vermehren-Schmaedick, A., Ainsley, J. A., Johnson, W. A., Davies, S. A. & Morton, D. B. (2010). Behavioral responses to hypoxia in *Drosophila* larvae are mediated by atypical soluble guanylyl cyclases. *Genetics* **186**, 183–96.

Wikelski, M. & Kays, R. (2010). Movebank: archive, analysis and sharing of animal movement data. Hosted by the Max Planck Institute for Ornithology. www.movebank.org.

Wiltschko, A. B., Johnson, M. J., Iurilli, G., Peterson, R. E., Katon, J. M., Pashkovski, S. L., Abraira, V. E., Adams, R. P. & Datta, S. R. (2015). Mapping sub-second structure in mouse behavior. *Neuron* **88**, 1121–1135.

Würbel, H., Stauffacher, M. & Holst, D. (1996). Stereotypies in laboratory mice—quantitative and qualitative description of the ontogeny of 'wire-gnawing'and 'jumping'in zur: Icr and zur: Icr nu. *Ethology* **102**, 371–385.

Xiao, C. & Robertson, R. M. (2015). Locomotion induced by spatial restriction in adult *Drosophila*. *PloS One* **10**, e0135825.

Author contributions

S.Q. contributed to data collection and manuscript preparation; C.X. contributed to experimental design, data collection, statistical analysis and manuscript preparation. The authors declare no conflict of interest.

Figure legends

Figure 1: **Coil-like locomotor trajectories of** *Drosophila* **locomotion in the circular arenas.** (**a**) 60 s walking trajectories of control flies. (**b**) 60 s walking trajectories of pulsed flies. Individual flies were loaded into the circular arenas (1.27 cm diameter 0.3 cm depth). Fly positions were tracked once per 0.2 s. Time-series of x, y-positions during 60 s locomotion were re-constructed as 3D locomotor trajectories.

Figure 2: **Distance to center followed a gumbel distribution with finite mixtures.** 300 s time-series of distance to center were plotted and fitted by gumbel, weibull or gamma distribution with finite mixtures in (**a**) a control fly, and (**b**) a pulsed fly. Diagnostic worm plots of residual deviations vs unit normal quantile for gumbel, weibull or gamma modeling in (**c**) the control, and (**d**) the pulsed fly.

Figure 3: **Locomotor structures of** *Drosophila* **movement in a circular arena.** (**a**) Alignment of actual observations (black open circles and linking lines) and modeled values (red open circles) of distance to center during 60 s locomotion. (**b**) Visual proofs of extracted locomotor structures: cross, side-wall walk, angle walk and stop. Time series of distance to center corresponding to several crosses were illustrated (in blue) in **a**. Black number on each image indicates starting time (s) and red number duration (s).

Figure 4: **Analysis of locomotor structures during 300 s movement.** (**a**) Illustrations of cross activities (blue blocks) in the circular arenas from 63 controls and 55 pulsed flies. (**b**) Time proportion for cross and numbers of cross during 300 s movement in controls and pulsed flies. ***, $P < 0.0001$ by Mann-Whitney test. (**c**) Time proportion for side-wall walk, angle walk, stop and unclassified activities in controls and pulsed flies. n.s., non-significant difference.

Figure 5: *Drosophila* **counter-clockwise walk and clockwise walk in a circular arena.** (**a**) Illustrations of counter-clockwise walk (red) and clockwise walk (grey) during 300 s movement in controls and pulsed flies. (**b**) Time proportion for counter-clockwise walk and clockwise walk in controls (left panel), and pulsed flies (right panel). (**c**) Time proportion for both rotations in controls and pulsed flies. (**d**) Number of counter-clockwise walk in controls and pulsed flies. (**e**) Number of clockwise walk in controls and pulsed flies. ***, $P < 0.0001$ by Mann-Whitney test.

Figure 6: **Analysis of** *Drosophila* **relative turning angles in a circular arena.** (**a**) Rose diagram plot and modeling of relative turning angle during counter-clockwise walk (red) and clockwise walk (grey). (**b**) Schematic relation between ω and θ during counter-clockwise walk (red square) and clockwise walk (grey square). The modeled values of θ for overall observations from controls and pulsed flies were provided. Black dots (with labels *i-1, i* and *i+1*) indicate sequential relocations. Black arrows indicates instant walking direction. **X** represents arena center, and grey curves arena edges. (**c**) Peak turning angle θ during counter-clockwise walk in controls and pulsed flies. (**d**) Peak turning angle θ during clockwise walk in controls and pulsed flies. (**e**) Relative densities at peak θ during counter-clockwise walk in controls and pulsed flies. (**f**) Relative densities at peak θ during clockwise walk in controls and pulsed flies. ***, $P < 0.0001$ by Mann-Whitney test.

Figure 7: **A relation between turning direction and distance to center.** (**a**) Relationship between distance to center and relative turning angle in controls and pulsed flies. Analysis was performed separately for counter-clockwise walk (red) and clockwise walk (grey). We rotated angle interval from (180°, 360°] to (-180°, 0°] for zero-centered demonstration. Each panel contained observations from ten flies. (**b**) Schematic movement structures for rotation behavior and their visual proofs. Black dots represent centers of mass and arrows rotation orientations. Red circle-shaped trajectories indicate counter-clockwise walk with or without right-turning cross. Grey circle-shaped trajectories indicate clockwise walk with or without left-turning cross. (**c**) Numbers of classified rotations in controls and pulsed flies. (**d**) Time proportion of classified rotations in controls and pulsed flies. ***, $P < 0.0001$ by Mann-Whitney test.

Figure S1: *Drosophila* **walking activities in the circular arenas.** (**a**) Control flies' walking activities in the circular arenas. (**b**) Pulsed flies' walking activities in the circular arenas. A single fly was loaded into each arena (1.27 cm diameter 0.3 cm depth). Each panel illustrates activities from 8 flies by 5 min. Walking activity is shown as blue lines connecting fly positions, which are the centers of mass computed once per 0.2 s. A visual difference is that pulsed flies reduce the chance of crossing central region of arenas compared with controls.

Figure S2: **Modeled gumbel components of distance to center.** (**a**) Three representative gumbel components (in black, red and blue) with respective parameters (μ, σ and π) modeled by gamlss.MX() function. A total of 177,000 observations from 63 controls and 55 pulsed flies were used for computing representative gumbel components. The bottom panel shows overlap of components. (**b**) Classified observations of distance to center from a control and a pulsed fly by posterior probabilities of gumbel components. Each color (blue, red or black) represents a unique gumbel identity.

Table S1: Model selection for the observations of distance to center

	Number of flies	
Distributions	**control**	**pulsed**
Gumbel	58	44
Weibull	5	8
Gamma	0	0
Reverse gumbel	0	0
t family	0	3
Logistic	0	0
Normal	0	0
Total	**63**	**55**

Figure 1

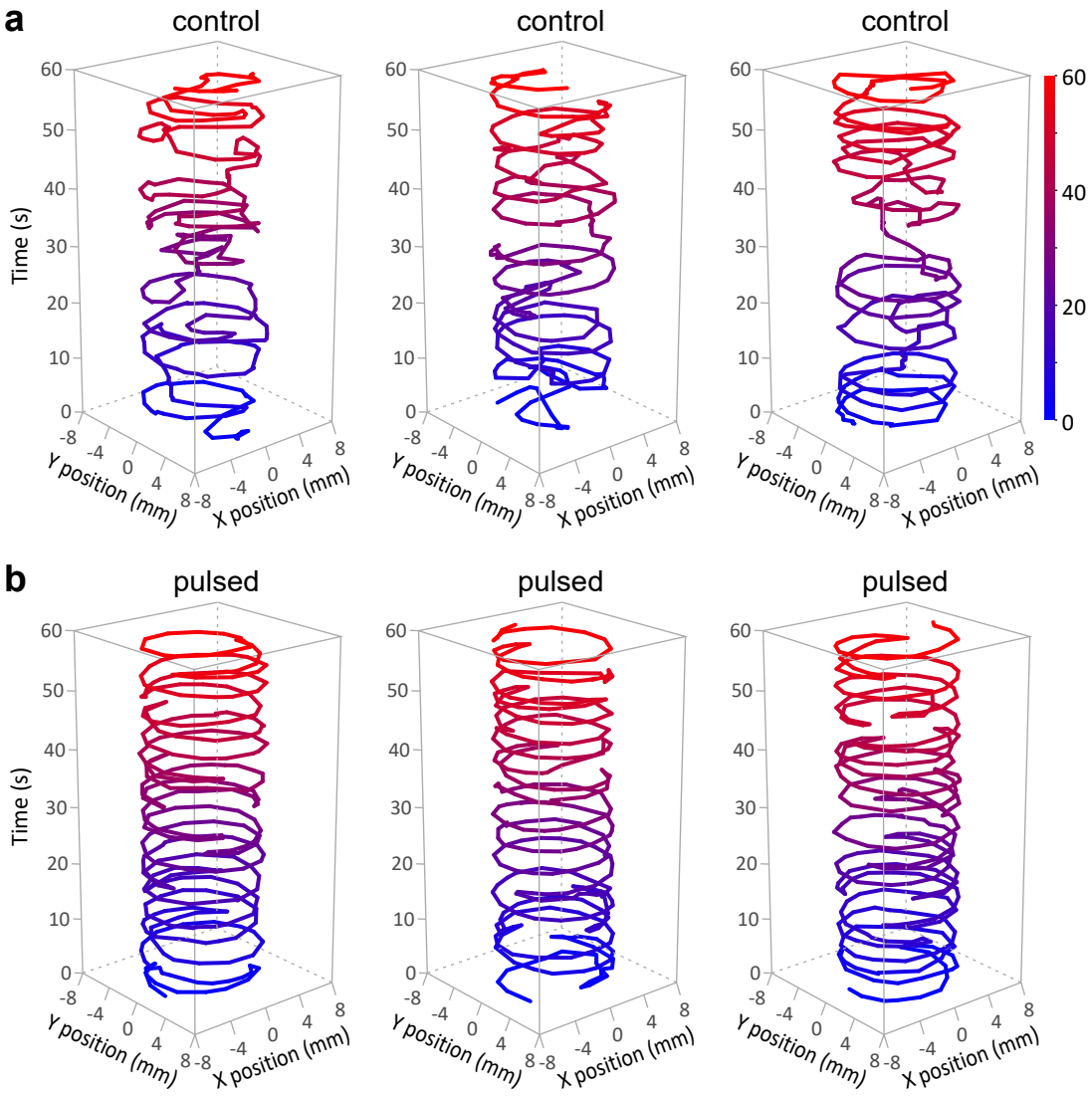


Figure 2

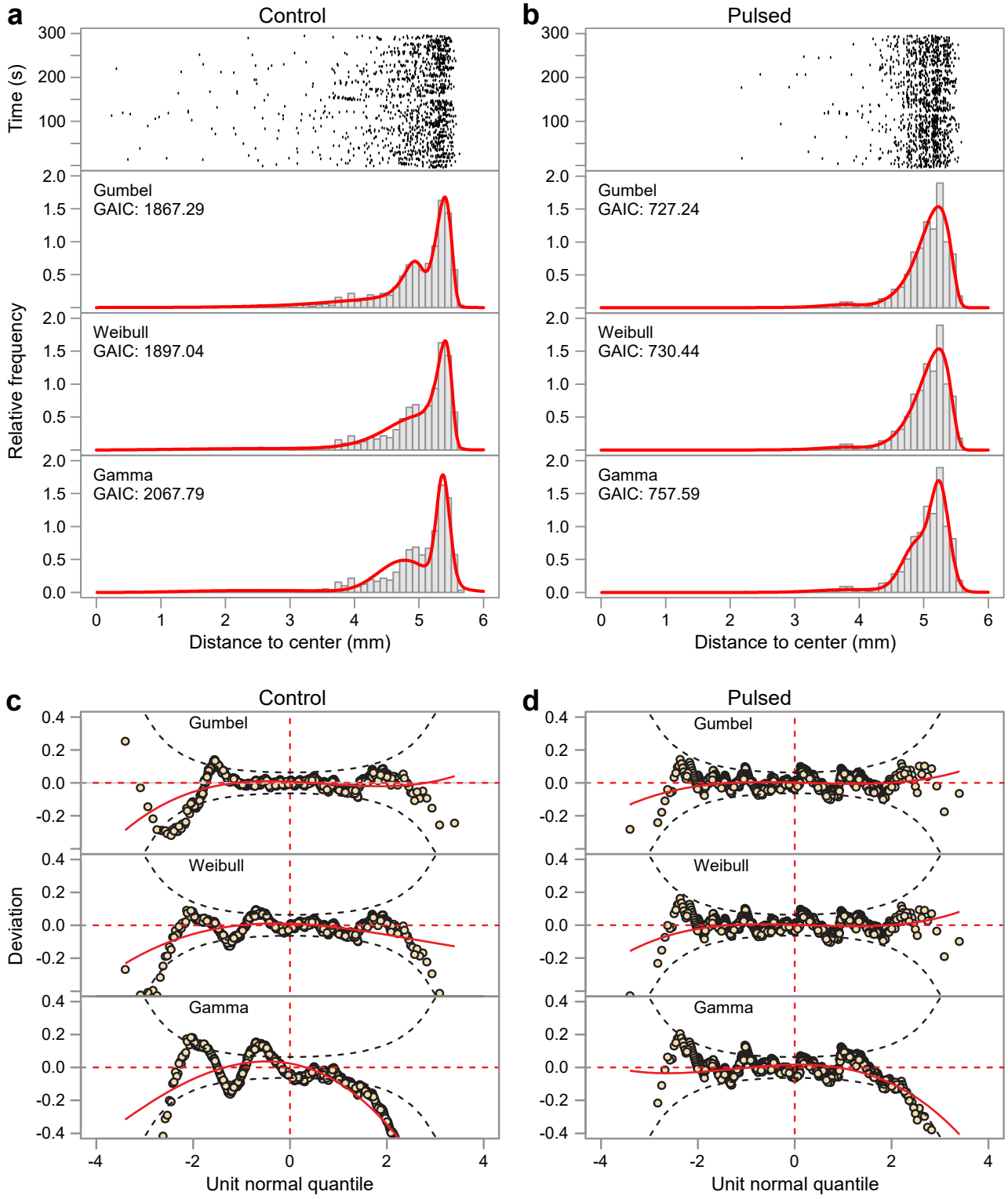


Figure 3

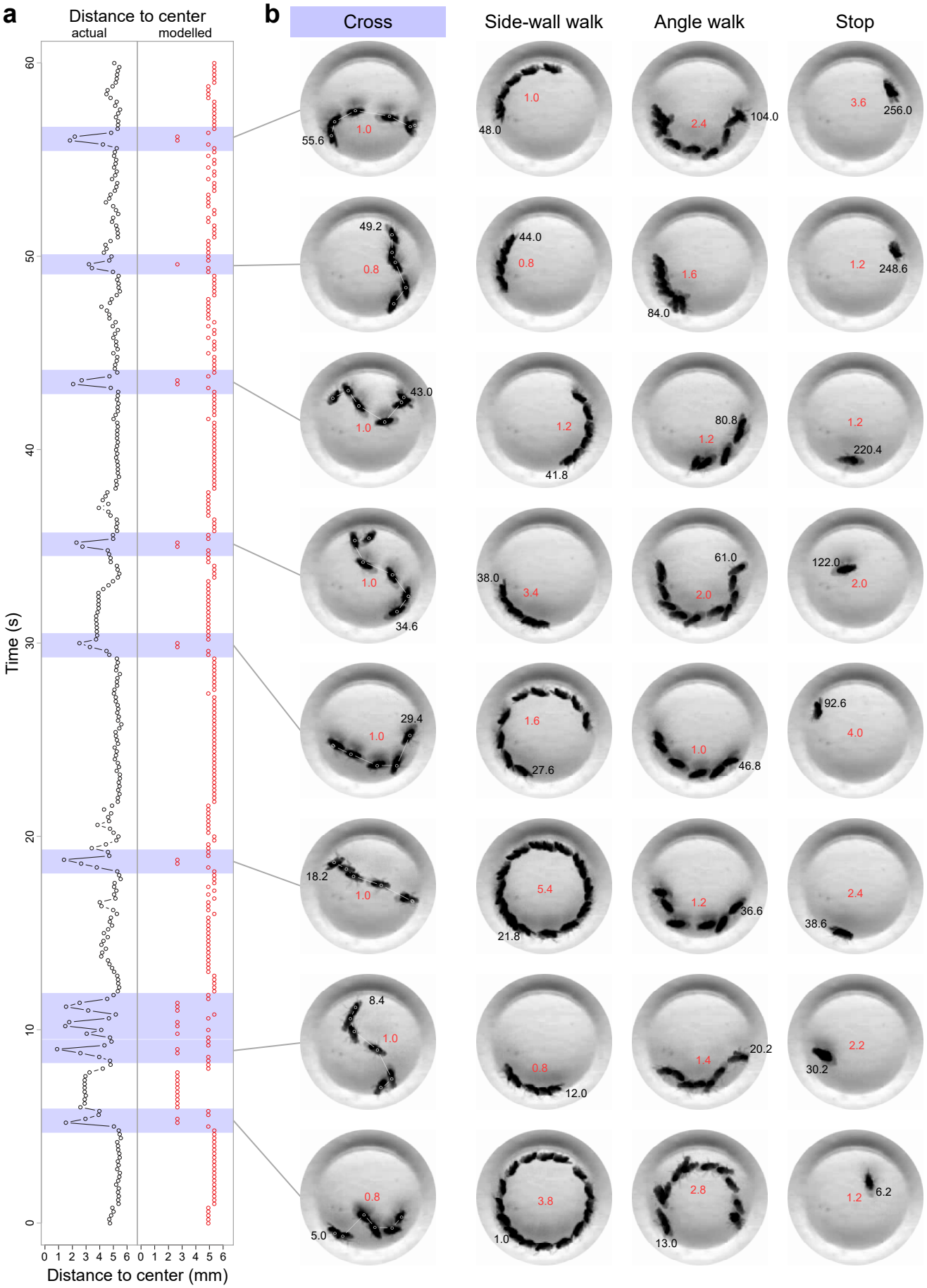


Figure 4

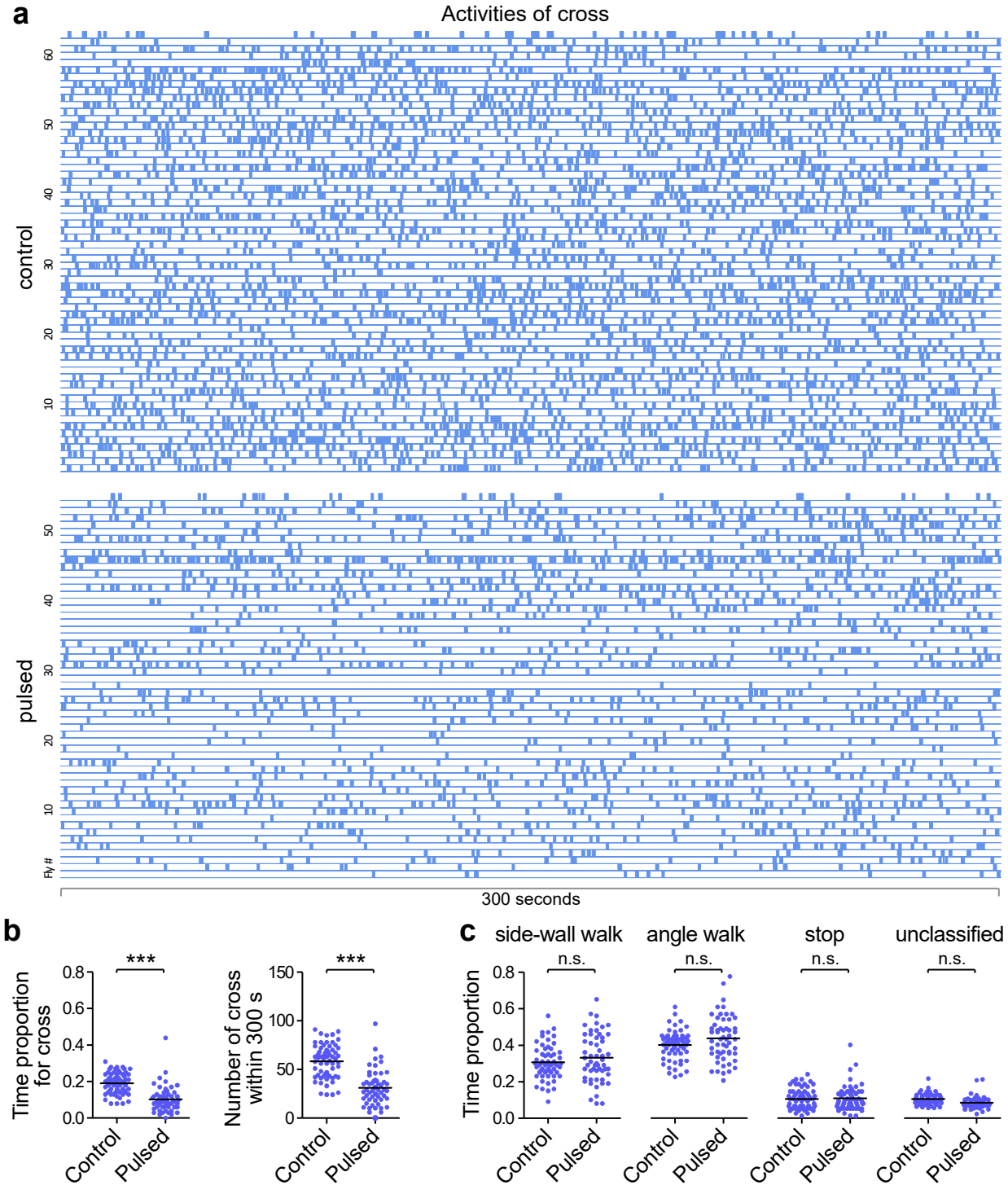


Figure 5

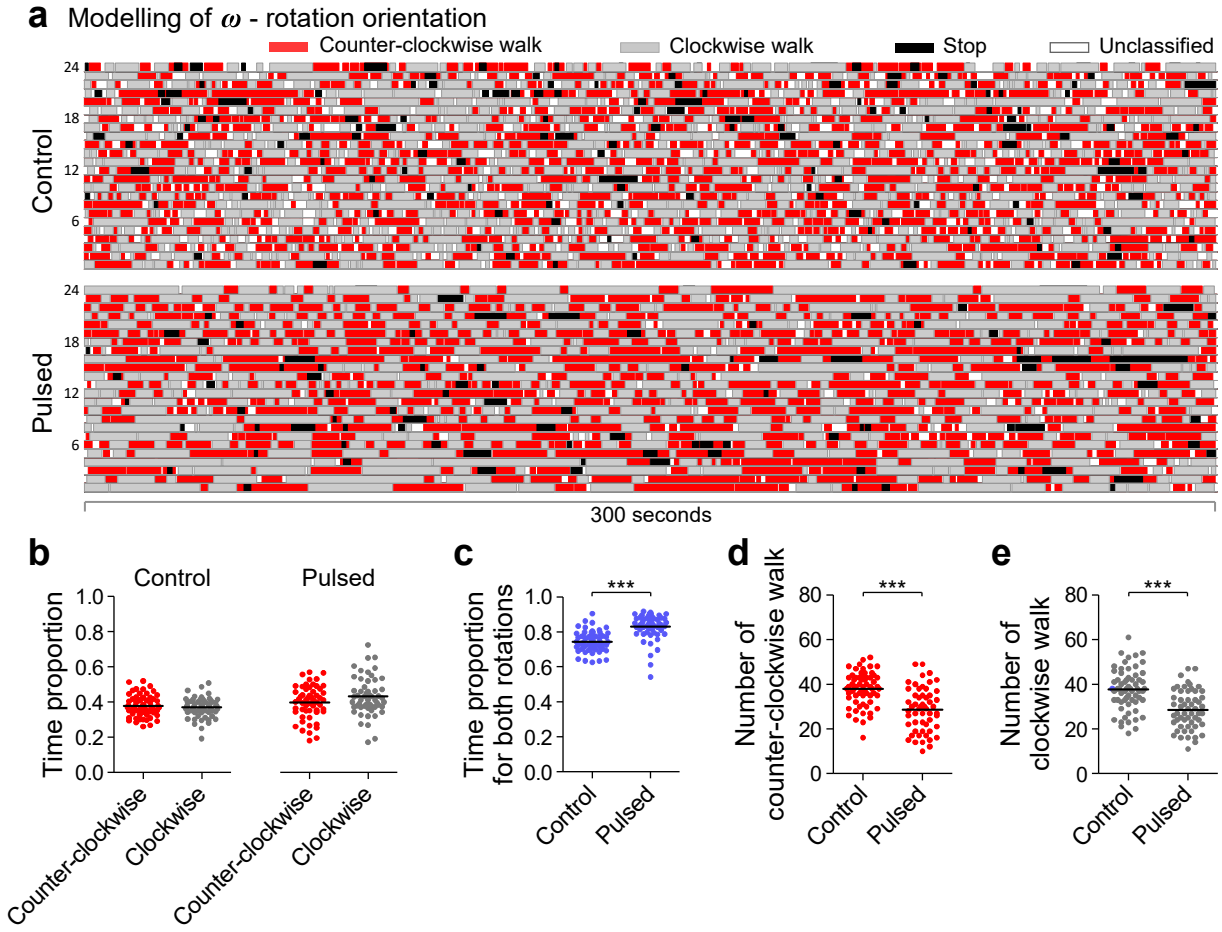
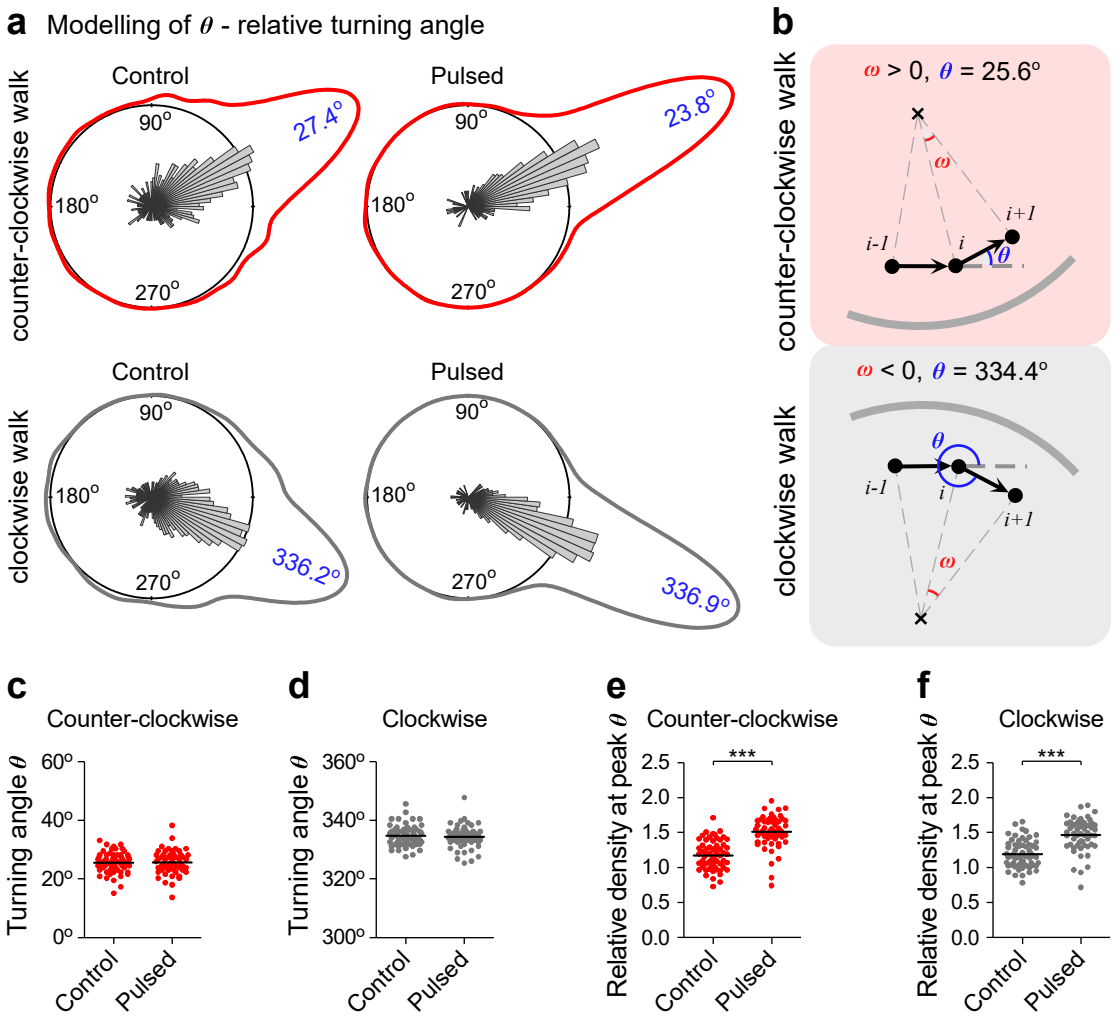


Figure 6



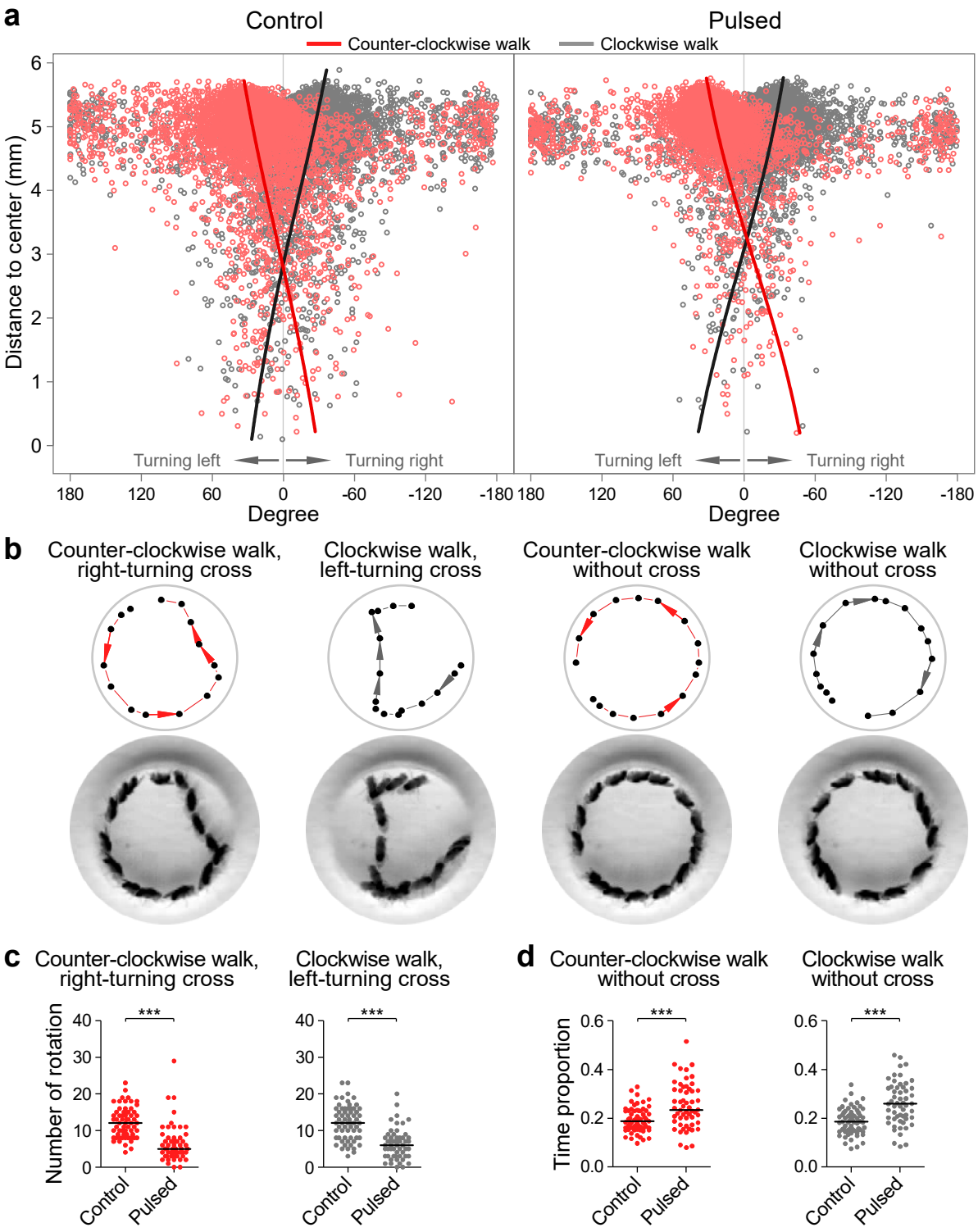


Figure S1

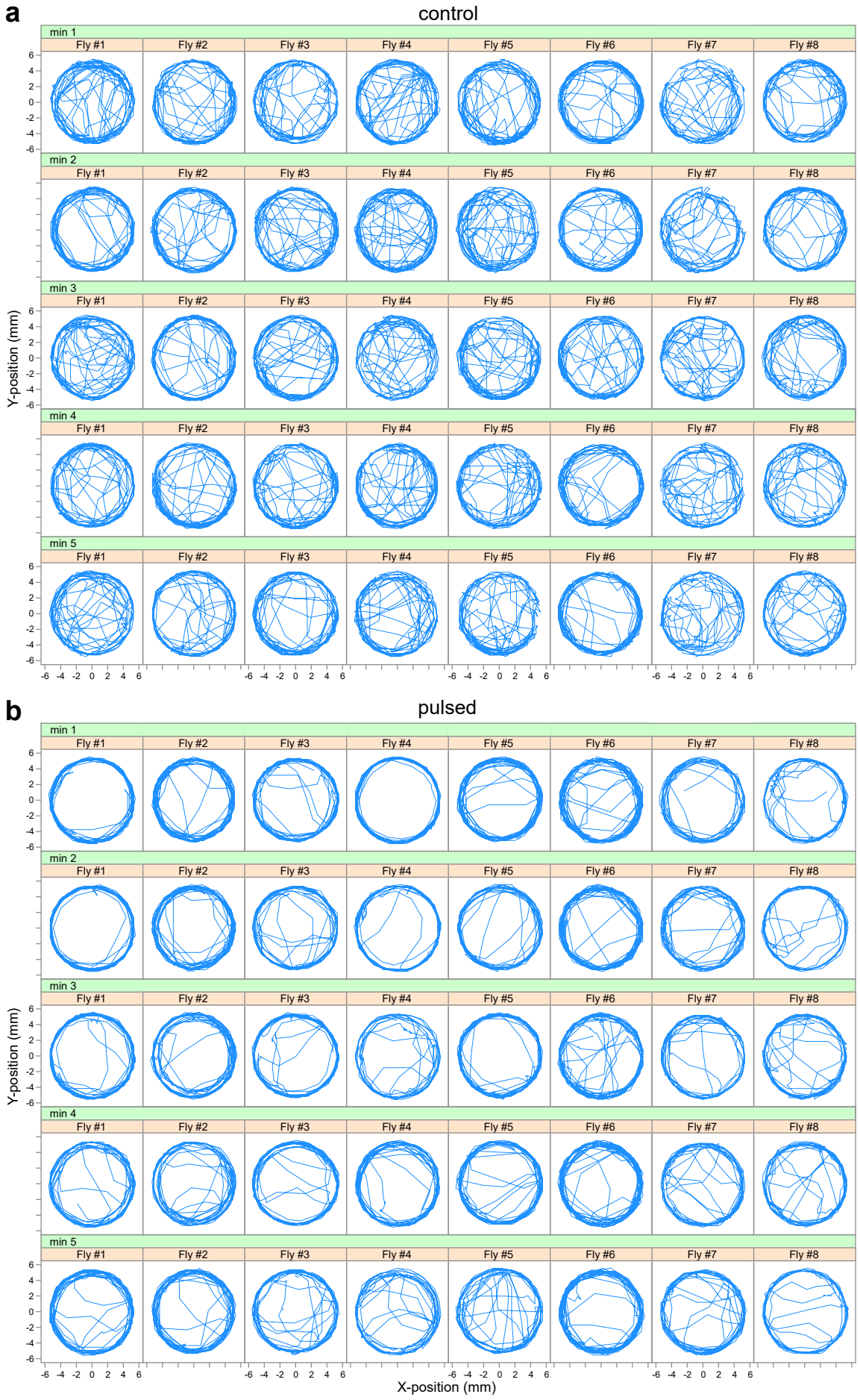


Figure S2