A hybrid versatile method for state estimation and feature extraction from the trajectory of animal behavior

Authors: Shuhei J. Yamazaki[1,2], Kazuya Ohara[3], Kentaro Ito[4], Nobuo Kokubun[4,5], Takuma Kitanishi[6,7,8], Daisuke Takaichi[9], Yasufumi Yamada[10], Yosuke Ikejiri[1,2], Fumie Hiramatsu[1], Kosuke Fujita[1,14], Yuki Tanimoto[1,15], Akiko Yamazoe-Umemoto[1], Koichi Hashimoto[11], Katsufumi Sato[12], Ken Yoda[13], Akinori Takahashi[4,5], Yuki Ishikawa[9], Azusa Kamikouchi[9], Shizuko Hiryu[10], Takuya Maekawa[3], Koutarou D. Kimura[1,2*]

Affiliations

[1]Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan

[2]Graduate School of Natural Sciences, Nagoya City University, Nagoya, Aichi, 467-8501, Japan

[3]Graduate School of Information Science and Technology, Osaka University, Suita, Osaka 565-0871, Japan

[4]Department of Polar Science, SOKENDAI, Tachikawa, Tokyo 190-8518, Japan

[5]National Institute of Polar Research, Tachikawa, Tokyo 190-8518, Japan

[6]Department of Physiology, Osaka City University Graduate School of Medicine, Osaka 545-8585, Japan

[7]Center for Brain Science, Osaka City University Graduate School of Medicine, Osaka 545-8585, Japan

[8]PRESTO, Japan Science and Technology Agency (JST), Kawaguchi, Saitama 332-0012, Japan.

[9]Graduate School of Science, Nagoya University, Chikusa, Nagoya, Aichi, 464-8602, Japan

[10]Faculty of Life and Medical Sciences, Doshisha University, Kyotanabe, Kyoto 6100321, Japan

[11]Graduate School of Information Sciences, Tohoku University, Sendai, Miyagi 980-8579, Japan

[12]Atmosphere and Ocean Research Institute, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8564, Japan

[13]Graduate School of Environmental Studies, Nagoya University, Nagoya, Aichi 464-8601, Japan

[14]Present address: Department of Ophthalmology, Graduate School of Medicine, Tohoku University, Sendai, Miyagi 980-8574, Japan

[15]Present address: RIKEN Center for Brain Science, Wako, Saitama 351-0198, Japan

*Correspondence: kokimura-lab@umin.ac.jp or kokimura@nsc.nagoya-cu.ac.jp

Manuscript information:

Abstract: 224 words

Main body: 8995 words

Figures & Tables: 12

Supplementary file: 1

ABSTRACT

Animal behavior is the final and integrated output of the brain activity. Thus, recording and analyzing behavior is critical to understand the underlying brain function. While recording animal behavior has become easier than ever with the development of compact and inexpensive devices, detailed behavioral data analysis requires sufficient previous knowledge and/or high content data such as video images of animal postures, which makes it difficult for most of the animal behavioral data to be efficiently analyzed to understand brain function. Here, we report a versatile method using a hybrid supervised/unsupervised machine learning approach to efficiently estimate behavioral states and to extract important behavioral features only from low-content animal trajectory data. As proof of principle experiments, we analyzed trajectory data of worms, fruit flies, rats, and bats in the laboratories, and penguins and flying seabirds in the wild, which were recorded with various methods and span a wide range of spatiotemporal scales—from mm to 1000 km in space and from sub-seconds to days in time. We estimated several states during behavior and comprehensively extracted characteristic features from a behavioral state and/or a specific experimental condition. Physiological and genetic experiments in worms revealed that the extracted behavioral features reflected specific neural or gene activities. Thus, our method provides a versatile and unbiased way to extract behavioral features from simple trajectory data to understand brain function.

INTRODUCTION

The brain receives, integrates, and processes a range of ever-changing environmental information to produce relevant behavioral outputs. Therefore, understanding salient behavioral features can augment our understanding of important aspects of environmental information as well as of brain activity, which links the environmental information to behavior. Recent technological development of compact and inexpensive cameras and/or global positioning system (GPS) devices has facilitated convenient monitoring and recording of animal behavior (Brown and de Bivort, 2018; Dell et al., 2014; Egnor and Branson, 2016). However, the behavioral data generated through these approaches are frequently represented as a few simple measures, such as velocity, migratory distance, or the probability of reaching a particular goal, due to the challenges related to identification of specific aspects of behavior to be analyzed; in other words, it is still difficult to figure out how we can describe an animal behavior meaningfully (Berman, 2018). Owing to poor description of behavior, dynamic neural activity, for example, is not sufficiently interpreted even though simultaneous optical monitoring can measure a large number of time-series neural activities (Alivisatos et al., 2012; Landhuis, 2017). This large asymmetry in data richness between neural activity and behavior has emerged as one of the most significant issues in modern neuroscience (Anderson and Perona, 2014; Gomez-Marin et al., 2014; Krakauer et al., 2017).

One way to overcome the challenges in the appropriate descriptions of behavior is to describe its salient features via comprehensive analysis through an approach such as machine learning. Machine learning involves extracting latent patterns and uncovering knowledge from a large amount of data (Bishop, 2006). In fact, multiple behavioral analysis methods based on machine learning have been reported in the last decade (Baek et al., 2002; Branson et al., 2009; Brown et al., 2013; Dankert et al., 2009; Kabra et al., 2013; Mathis et al., 2018; Robie et al., 2017; Stephens et al., 2008; Vogelstein et al., 2014; Wiltschko et al., 2015). Most of these studies

have classified behavioral states based on detailed analyses of animal postures as observed in video images (Dell et al., 2014); the classification of behavioral states into classes, such as foraging, sleeping, chasing, or fighting, is considered to be critical for efficient behavioral analysis, as each of behavioral features varies differently across different behavioral states (Egnor and Branson, 2016; Jonsen et al., 2013; Patterson et al., 2008). Although these methods have worked successfully for the analysis of behavioral videos of worms, fruit flies, and rodents in laboratories, they have some limitations. First, these methods are not suitable for analyzing relatively long-distance navigation given their requirement of recording reasonably large and detailed images of animals in the video frame. Second, the extraction of behavioral features from a state, as opposed to just state classification, is more critical in understanding how environmental information and/or brain activities trigger transitions among states for behavioral response.

To analyze relatively long-distance navigation behavior comprehensively, we developed a method for the estimation of behavioral states and extraction of relevant behavioral features based only on the trajectories of animals. For estimating behavioral states, we used an unsupervised learning method involving the expectation maximization (EM) algorithm (Dempster et al., 1977) because it is difficult for the human eye to classify behavior into distinct states without using posture images. For extracting salient behavioral features, we used information gain, an index used for a supervised learning method (the decision tree analysis) (Quinlan, 1992), and compared the features between two different experimental conditions (e.g., with or without certain stimulus). It is because supervised learning is considered advantageous in the extraction of characteristic behavioral features and comparing them among multiple conditions. We named this hybrid supervised/unsupervised machine learning approach as the *st*ate *e*stimation and *f*eature ex*tr*action (STEFTR) method (Fig. 1).

Because the STEFTR method only uses trajectory information for the analysis, it becomes possible to analyze movement behavior of various animals regardless of the spatiotemporal scale of movement. As proof-of-principle experiments, we analyzed the trajectories of worms, flies, rats, and bats in laboratories and those of penguins and flying seabirds in the wild; these experiments involved a spatiotemporal scale ranging from mm to 1000 km in space and from sub-seconds to days in time. The behavioral states of worms and penguins estimated by the STEFTR method were in reasonable conformation with the ones described in previous literature, supporting the reliability of our method. We further extracted learning-dependent behavioral features from a behavioral state of worms, in which one of the behavioral features is correlated with learning-dependent changes in neural activities. We also analyzed the behavioral features of mutant strains of worms and found that the patterns of features are correlated with gene function, suggesting that comprehensive feature extraction may enable us to estimate unknown functions of a gene product. We were also able to extract learning-dependent features from bats and pheromone-dependent features from fruit flies. Taken together, our findings indicate that the STEFTR method allows us to estimate internal state, neural activity, and gene function related to animal behavior only from movement trajectories, regardless of the recording method or the spatiotemporal scales.

MATERIAL AND METHODS

Estimation of behavioral states

For the analysis of trajectory information of an animal obtained from video images or from the GPS device attached to an animal, approximately 1/1,000 and 1/100 of the median recording time across animals were used as an unit for time frame and the time window for moving average, respectively (Table 1). These values were used to draw the 8 histograms of the averages (*Ave*) and the variances (*Var*) of velocity (*V*), bearing (*B*), time-differential of *V* (*dV*) and *B* (*dB*) as the basic behavioral features. The time window for moving average was critical to

reduce noise and to detect relatively long trends of behavior. From the histograms, a basic behavioral feature appeared to include multiple normal distributions was selected, and the number of clusters and boundaries of each cluster were automatically determined by EM algorithm (see below). In the case of worms, the cluster analysis was performed with specifying the maximum number cluster of 20. In other cases, maximum cluster number 5 was predetermined based on the knowledge that the number of basic behavioral states are several in general (Patterson et al., 2008).

The EM algorithm assigns a cluster label to each time frame although the clustering results, which reflect behavioral states of an animal, should be smooth in time. To smooth out the clustering results and removing outlying results, moving average was again applied to the cluster labels, which resulted in clusters resemble to the human-labeled behavioral states.

When the value of a behavioral feature changes suddenly and largely, the influence of the change may extend over a wide range. For example, if an animal moving straightly initiates local search suddenly, *dB* value will be 0°, 0°, 0°, 0°, 0°, 0°, 180°, 0°, 90°, 0°, 270°, etc. If moving average with ±5 time frame was applied, the value change occur from -5 time frame of the sudden value change, which should be compensated. Because worm's clusters 0 and 1 corresponded to this case, the beginning and the end of each cluster 0 was extended by the half of time window.

The cluster labels obtained as described above were mapped to the corresponding trajectory position with colors. We used a custom-made python program for calculating basic behavioral features, Weka data mining software (the University of Waikato, New Zealand) (Frank et al., 2016) for EM calculation, and Excel (Microsoft) for others.

EM algorithm for cluster analysis

A set of values of the *i*th basic behavioral feature F_i, which were extracted from trajectories of interest, and the number of clusters *N* were given. We employed the EM algorithm to cluster F_i into *N* clusters, i.e., a mixture of *N* Gaussians. The probability distribution of the Gaussian mixtures M_N is represented as follows:

$$p(f_i|M_N) = \sum_{n=1}^{N} \pi_n N(f_i, \mu_n, \sigma_n),$$

where f_i is a feature value of the *i*th feature, π_n is the mixture weight of the *n*th Gaussian, μ_n is the mean of the *n*th Gaussian, and σ_n is the standard deviation of the *n*th Gaussian. Thus, the EM algorithm was used to estimate the cluster parameters: π_n, μ_n, and σ_n.

Determination of cluster number using information criteria

The method evaluates a set of clusters (model M_N) obtained by the EM algorithm using information criteria, and finds the best *N* using the following criteria.

$$AIC(M_N) = -2 \ln L\big(p(F_i|M_N)\big) + 2N,$$

where $L\big(p(F_i|M_N)\big)$ is the maximum likelihood of F_i under model M_N and *N* is the penalty term that measures model complexity.

Feature extraction with information gain

We leverage information gain to evaluate the classification ability of each feature, i.e., its ability to identify a characteristic of a state (cluster). Information entropy is used to compute the ambiguity of a set of data points according to the following formula:

$$H = -\sum_{n=1}^{N} p_n \log_2 p_n,$$

where p_n is the number of data points belonging to the *n*th class (cluster). Given that we classify all the data points into two groups (i.e. two experimental conditions) using a particular threshold related to a specific feature, the feature is considered to be a characteristic feature (in that it classifies the data points well) if the ambiguity within the two groups is lower than that of

the original data set. Thus, we evaluated features in terms of their ability in ambiguity reduction upon classification (information gain).

For worm odor avoidance behavior, we extracted behavioral features that have positive information gain in naive versus pre-exposed worms or in mock-treated versus pre-exposed worms. Next, we chose the extracted features that were common for both comparisons; these features were termed as "features modulated in a learning-dependent manner." For flies, behavioral features were compared between with or without pheromone tapping. For bats, behavioral features were compared between unfamiliar flights (1st-3rd) and familiar (10th-12th) flights. The Weka software was used for these calculations.

Behavioral parameters included in a feature vector

For the machine learning analysis of worm' odor avoidance behavior, the following behavioral features were calculated for each cluster 0 segment from the coordinates of the centroid of the trajectory: velocity (*V*), bearing (*B*), odor concentration the worm experienced during the run (*C*), the time differential values for these (*dV*, *dB*, and *dC*), directedness (*Dir*) (Gorelik and Gautreau, 2014), curvature (called weathervane; *WV*) (Iino and Yoshida, 2009), and durations of cluster 0 and 1 (*Clst0Dur* and *Clst1Dur*, respectively). For *V*, *dV*, *B*, *dB*, *C*, *dC*, and *Dir*, the average (*Ave*) and median (*Med*) values for at the initiation (*Ini*), middle (*Mid*), termination (*Ter*), and all (*All*) periods of a cluster 0 segment were calculated. A total of 333 features was calculated by combining all these features.

For analyzing changes in the flight of bats, the following behavioral features in each flight were calculated from the coordinates of the bats and obstacles: three-dimensional flight velocity (*V*), horizontal and vertical bearings of the flight (*B_hori* and *B_vert*, respectively), distance (*R_obs*) and bearing (*B_obs*) of the bat to the nearest edge point of the obstacle chain array, longitudinal

215 directional distance to the frontal chain array (*R_x*), and lateral directional distance to the inside
216 pitch of the chain array (*R_y*). Time-differential values were calculated for *V* (*dV*), *B* (*dB*), *dB*
217 (*ddB*), and the flight height (*dH*), which were calculated with frame units of the high-speed
218 video cameras (1/125 s). All flight trajectories were divided into three segments: early, middle,
219 and late terms. The time window for the analysis of each behavioral feature was 0.1 s, 0.2 s, or
220 0.3 s before or while ($t = 0$) passing through the chain array. A total of 42 features was
221 calculated by combining all these features.
222
223 Excel and Visual C# (Microsoft) were used for the calculations, while the Beeswarm package
224 for R (The R Project) was used to obtain a scatter plot of the data.
225
226 **Worms**
227 The culture and handling of *Caenorhabditis elegans* strains were performed according to
228 techniques described previously (Brenner, 1974). Wild-type Bristol strain RRID:WB-
229 STRAIN:N2_Male and mutant strains RRID:WB-STRAIN:MT1219 *egl-3(n589)*, RRID:WB-
230 STRAIN:VC671 *egl-3(ok979)*, RRID:WB-STRAIN:CX4544 *ocr-2(ak47),* RRID:WB-
231 STRAIN:JC1636 *osm-9(ky10)*, and RRID:WB-STRAIN:FK127 *tax-4(p678)*, RRID:WB-
232 STRAIN:MT6308 *eat-4(ky5)*, and RRID:WB-STRAIN:IK105 *pkc-1(nj1)* were obtained from
233 the Caenorhabditis Genetics Center at the University of Minnesota, USA. The RRID:WB-
234 STRAIN:KDK1 *dop-3(tm1356)* strain was originally obtained from the National BioResource
235 Project (Japan) and back-crossed with the wild-type N2 strain five times.
236
237 A 2-nonanone avoidance assay was performed according to the protocol described previously
238 (Kimura et al., 2010; Yamazoe-Umemoto et al., 2015). Briefly, 2-3 young adult hermaphrodite
239 worms grown synchronously were placed in the center of a 9-cm nematode growth media
240 (NGM) plate. Worm behavior was recorded for 12 min after 2-μL of 30% 2-nonanone (cat. no.

241 132-04173; Wako, Japan) diluted in 99.5% ethanol (cat. no. 0057-00456; Wako, Japan) were
242 placed at two spots on the surface of the NGM plate. This assay was performed under the
243 following three conditions: (1) Naive—the worms cultivated on 6-cm NGM plates with the
244 RRID:WB-STRAIN:OP-50 bacteria as food were briefly washed with NGM buffer and
245 subjected to the assay; (2) Pre-exposed—the worms were subjected to the assay after being pre-
246 exposed to 0.6 μL of 15% 2-nonanone spotted on the lid of a 6-cm NGM plate for 1 h without
247 food; and (3) Mock—the worms were subjected to the assay after being pre-exposed to ethanol
248 similarly to the pre-exposed condition. We added the mock-treated control group to ensure that
249 the starvation itself did not affect the odor avoidance behavior of worms and to extract
250 behavioral features modulated by odor pre-exposure compared to the naive and mock-treated
251 control groups. Images of worms on the 9-cm NGM plate during the odor avoidance assay were
252 acquired by a high-resolution USB camera (DMK 72AUC02; The Imaging Source, USA) with a
253 lens (LM16JC5MW; Kowa, Japan) at 1 Hz for 12 min. The coordinates of individual animals'
254 centroids were acquired from the recorded images using the Move-tr/2D software (Library Co.,
255 Ltd., Tokyo, Japan) and used for the STEFTR analysis.
256
257 Similar to the other sensory behaviors of the worms, trajectories in the 2-nonanone avoidance
258 behavior can be divided into two states: (1) "run"—a relatively long period of straight
259 movement, and (2) "pirouette"—a period of short movements interrupted by frequent reversals
260 and turns (Kimura et al., 2010; Pierce-Shimomura et al., 1999). The angular change per second
261 was calculated from the centroid coordinates, and movements of 1 s with angular changes larger
262 than 90° were classified as a turn. The histogram of turn intervals could be fitted to two
263 exponentials, suggesting that the turn intervals are regulated by two probabilistic mechanisms
264 (Pierce-Shimomura et al., 1999; Yamazoe-Umemoto et al., 2015). The time point at which the
265 two exponentials intersected was defined as t_{crit}, and turn intervals longer or shorter than the t_{crit}
266 were classified as runs or included in pirouettes, respectively. The t_{crit} was calculated for the

control (i.e., naive and mock-treated) condition for wild-type and mutant strains. Excel (Microsoft) was used for the above calculations. The odor concentrations that the worms experienced at specific spatiotemporal points were calculated according to the dynamic odor gradient model based on the measured odor concentration (Tanimoto et al., 2017; Yamazoe-Umemoto et al., 2018).

Statistical analyses were performed with Prism ver. 5.0 for Mac OSX (GraphPad Software, CA, USA) and R (The R Project). Sample size was determined based on the previous report (Yamazoe-Umemoto et al., 2015). A part of the original data used in this study had already been analyzed and published previously (Yamazoe-Umemoto et al., 2015), and re-analyzed with the STEFTR method.

Penguins

Fieldwork was performed on chick-rearing Adélie penguins at Hukuro Cove colony (69°13′ S, 39°38′ E) in Lützow-Holm Bay, East Antarctica. GPS-depth loggers (GPL380-DT or GPL400-D3GT, weighing 55-85 g; Little Leonardo, Japan) were deployed among 11 penguins during the period between 27 December 2016 and 10 January 2017 and recovered from all the birds after 1-2 days. While the loggers were set to record GPS positions and depth every second, they could not record GPS positions when the penguins were diving. Therefore, we linearly interpolated the data, when necessary, to obtain GPS positions every 1 minute before further analysis. See Kokubun et al. for methodological details (Kokubun et al., 2015). This fieldwork was carried out in accordance with the recommendations of the Law relating to Protection of the Environment in Antarctica. The protocol was approved by the Ministry of the Environment, Government of Japan. Sample size was not predetermined.

Flying seabirds

293 Fieldwork was performed on *Calonectris leucomelas* at Funakoshi-Ohshima Island (39°24'N,
294 141°59'E) between August and September in 2011, 2012, 2013, and 2015. We attached GPS
295 loggers (GiPSy-2, 37 × 16 × 4 mm or GiPSy-4, 37 × 19 × 6 mm; TechnoSmArt, Roma, Italy) to
296 the back feathers of chick-rearing Streaked Shearwaters with Tesa® tape (Beiersdorf AG;
297 GmbH, Hamburg, Germany) and cyanoacrylate glue (Loctite®401; Henkel Ltd., Hatfield, UK).
298 The loggers were housed in waterproof heat-shrink tubing and set to record one fix per minute.
299 The total weight of the unit was 25 g, which was less than 5% of the mean mass of the birds in
300 accordance with the suggested load limit for flying seabirds. After approximately two weeks of
301 deployment, we recaptured and retrieved the loggers. See Yoda et al. for methodological details
302 (Yoda et al., 2014). The study was carried out in accordance with the recommendations of the
303 guidelines of the Animal Experimental Committee of Nagoya University. The protocol was
304 approved by the Animal Experimental Committee of Nagoya University. Sample size was not
305 predetermined. A part of the original data used in this study had already been analyzed and
306 published previously (Yoda et al., 2014), and re-analyzed with the STEFTR method.
307
308 **Rats**
309 Locomotion data of an adult male Long Evans rat were obtained from the Collaborative
310 Research in Computational Neuroscience (CRCNS; RRID:SCR_005608) data sharing website
311 (https://crcns.org, hc-3 dataset, ec013 rat) (Mizuseki et al., 2014). The rat foraged for randomly
312 dispersed water or foods on an elevated open field (180 cm × 180 cm) for 17 - 53 min. The rat's
313 position was tracked by monitoring two light-emitting diodes mounted above the head with an
314 overhead video camera at 30 Hz. The 30-Hz tracking data were resampled to 39.0625 Hz for
315 offline processing. This study was carried out in accordance with the recommendations of the
316 Regulations on Animal Experiments in Osaka City University. The protocol was approved by
317 the Animal Care and Ethics Committee of Osaka City University. Sample size was not
318 predetermined. The original data used in this study had already been analyzed and published

319 previously (Diba and Buzsáki, 2008; Mizuseki et al., 2009; 2014), and re-analyzed with the
320 STEFTR method.
321
322 **Flies**
323 Fruit flies *D. melanogaster* were raised on standard yeast-based media at 25ºC and 40 - 60%
324 relative humidity under a 12-hour light/dark cycle. *Canton-S* flies aged between 6-8 days after
325 eclosion were used as a wild-type. After eclosion, the males were housed singly, while females
326 were housed in groups until the experiment.
327
328 The locomotion measurement was performed as described previously with minor modifications
329 (Kohatsu and Yamamoto, 2015; Kohatsu et al., 2011). Briefly, a male fly was tethered with a
330 metal wire on its dorsal thorax and positioned over an air-supported Styrofoam ball (diameter, c.
331 a. 6 mm). The locomotion trajectory of the fly was recorded by monitoring the rotations of the
332 Styrofoam ball using an optical computer mouse sensor (BSMRU21BK; BUFFALO INC.,
333 Nagoya, Japan). The sensor detected the movements of the ball in the horizontal (Δx) and
334 vertical (Δy) directions, which correspond to lateral and forward movements of the male fly,
335 respectively. The Δx and Δy values, together with timestamps, were sent to a computer at 60 Hz
336 via an Arduino Due microcontroller (Switch Science, Japan) with a custom sketch program. The
337 60-Hz data were down-sampled to 4-Hz data for the information gain analysis. The
338 measurements were obtained at 25±1°C and 50±10% relative humidity and within 4 hours after
339 light onset.
340
341 Female pheromones were applied to the male fly by placing the female's abdomen in contact
342 with the male's foreleg at the onset of the measurement. A manipulator (M-3333, Narishige,
343 Tokyo, Japan) actuated a pipette with a volume of 200 μL (FUKAEKASEI Co., Ltd., China), in
344 which a live female with her abdomen exposed toward a male fly was captured. We manually

345 controlled the position of the manipulator to contact the female's abdomen to the male's foreleg.
346 This contact procedure was omitted in the control experiments.
347
348 Visual stimulus was applied directly after pheromone application by starting horizontal
349 movements of the female fly in front of the male fly as described (Kohatsu et al., 2011). The
350 visual stimulus consisted of ten left-right horizontal movements of the female that lasted for 40
351 s. Each movement started with the female fly in the front of the male fly (i.e., center) and
352 continued as the female fly moved left until it reached the left end of the rail (i.e., 5 mm away
353 from the center), then moved right until it reached the right end of the rail (i.e., 5 mm away from
354 the center), and ended when it came back to the center with a constant velocity of 5 mm/s. This
355 movement was driven by a stepper motor (42BYG Stepper Motor, Makeblock Co., Ltd.,
356 Shenzhen, China) controlled by a custom sketch program (processing software version 3.3.7).
357 We defined one round (4 s in total) as the movement related to the female starting to moving
358 away from the center, reaching the left end of the rail, passing the center, moving away to reach
359 the right end of the rail, and coming back to the center again. Each of the "moving away" and
360 "coming back" periods lasted for 1 s.
361
362 The Δx and Δy values with timestamps obtained in the final eight rounds were used for the
363 analysis. To detect the characteristic parameters for the chasing behavior, we used Δx and Δy
364 values during the period when the female was moving away from the male (2 s/round). We set
365 the angle of the chasing behavior as 0 degree when the male moved forward. Angles between 0
366 and 90 degrees indicate that the male fly is moving towards the female that was moving away
367 from the male.
368
369 As the parameters (velocity, bearing, and their time-differential values) were not normally
370 distributed (Shapiro-Wilk test; see Supplementary Table 13), their values were compared

371 between conditions (with/without pheromone) using the Mann-Whitney U test followed by
372 Bonferroni correction for multiple comparison. We used the Steel-Dwass test to compare values
373 of the parameters between rounds. Statistical analyses were conducted using R software version
374 3.4.4. No statistical methods were used to pre-determine sample sizes, but our sample sizes are
375 similar to those in previous studies (Kohatsu and Yamamoto, 2015; Kohatsu et al., 2011).
376
377 **Bats**
378 Three adult Japanese horseshoe bats (*Rhinolophus ferrumequinum nippon*, body length: 6–8 cm,
379 body mass: 20–30 g) were captured from natural caves in the Hyogo and Osaka prefectures in
380 Japan as previously described (Yamada et al., 2016). The bats were housed in a temperature-
381 and humidity-controlled colony room [4 m (L) × 3 m (W) × 2 m (H)] with a 12-h light/dark
382 cycle at Doshisha University in Kyoto, Japan, and were allowed to fly freely and given access to
383 mealworms and water. Captures were conducted under license and in compliance with current
384 Japanese law. This study was carried out in accordance with the recommendations of Principles
385 of Animal Care (publication no. 86-23 [revised 1985)] of the National Institutes of Health) and
386 all Japanese laws. The protocol was approved by the Animal Experiment Committee of
387 Doshisha University.
388
389 Methods for acoustic navigation measurement in bats have been described elsewhere (Yamada
390 et al., in revision). Briefly, the experiments were conducted in a flight chamber constructed
391 using steel plates [9 (length) × 4.5 (width) × 2.5 m (height)] under lighting with red filters (>650
392 nm) to avoid visual effects on the bats. An obstacle environment was constructed using plastic
393 chains (diameter: 4 cm) suspended from the ceiling of the chamber. The chains were arranged at
394 15-cm intervals along the *x*-axis and at 22-cm intervals along the *y*-axis so that the bat was
395 forced to fly in an S-shaped pattern without passing between chains. Three naive bats were
396 observed for 12 continuous repeated flights so that their echolocation behavior in unfamiliar and

397 familiar spaces could be compared. In this study, the first three flights were defined as
398 unfamiliar flights, while the last three flights were defined as familiar flights.
399
400 The flight behavior of the bats was recorded at 125 frames/s using two digital high-speed video
401 cameras (MotionPro X3; IDT Japan, Inc., Japan) placed in the left and right corners of the flight
402 chamber. Based on a direct linear transformation technique, the successive 3D positions of the
403 flying bats, as well as the locations of other objects, were reconstructed using motion analysis
404 software (DIPPMotionPro ver. 2.2.1.0; Ditect Corp., Japan). The statistical calculations were
405 performed with SPSS version 23 (IBM Corp.).
406
407 **Calcium imaging of worm's neurons**
408 Calcium imaging of the worms' ASH neurons was performed according to the previous method
409 with some modifications (Tanimoto et al., 2017). Briefly, transgenic strains expressing GCaMP3
410 (Tian et al., 2009) and mCherry (Shaner et al., 2004) in ASH sensory neurons under the *sra-6*
411 promoter (KDK70034 and KDK70072; 20 ng/µl of *sra-6p::GCaMP3*, 20 ng/µl of *sra-*
412 *6p::mCherry*, 10 ng/µl of *lin-44p::GFP*, 50 ng/µl of PvuII-cut N2 genomic DNA as a carrier in
413 N2 background) were placed on an NGM agar plate on a robotic microscope system, OSB2
414 (Tanimoto et al., 2017). Although these transgenic worms were immobilized with the
415 acetylcholine receptor agonist levamisole (Lewis et al., 1980) for high-throughput data
416 acquisition through simultaneous imaging of multiple worms, the previous study revealed that
417 the ASH activity is essentially unaffected by levamisole-treatment (Tanimoto et al., 2017). A
418 constant gas flow of 8 mm/min was delivered, in which the mixture rate of 2-nonanone gas with
419 air was changed to create a temporal gradient of odor concentration. The temporal change in
420 odor concentration was measured by a custom-made semiconductor sensor before and after the
421 series of calcium imaging experiments on each day. The fluorescence signals of GCaMP3 and
422 mCherry in ASH neurons were divided into two channels using W-View (Hamamatsu, Japan),

423 an image splitting optic, and captured by an electron multiplying charge-coupled detector (EM-
424 CCD) camera (ImagEM; Hamamatsu, Japan) at 1 Hz. The intensities of fluorescence signals
425 from cell bodies were extracted and quantified by ImageJ (NIH) after background subtraction.
426 The average ratio over 30 s prior to the odor increase was used as a baseline (F_0), and the
427 difference from F_0 (ΔF) was used to calculate the fluorescence intensities of GCaMP3 and
428 mCherry ($F = \Delta F/F_0$). The ratio between florescence intensities of GCaMP and mCherry
429 (GCaMP/mCherry) was used in the figure.
430
431 **RESULTS**
432 **Estimation of behavioral states**
433 As the first part of the analysis, we classified the trajectory into several behavioral states based
434 on the distribution of a basic behavioral feature. Behavior of animals consists of several states
435 (Jonsen et al., 2013; Patterson et al., 2008), where basic behavioral features such as speed and
436 direction change are likely distributed probabilistically with a center value that is optimal for
437 each state. Thus, behavior can be more easily characterized when the behavioral features are
438 analyzed for each state rather than for the entire behavior as a whole. In fact, classifying the
439 trajectory into several states is one of the essential preprocessing steps in trajectory mining of
440 people and vehicles in data science (Zheng, 2015).
441
442 For the state classification, we calculated the averages (*Ave*) and variances (*Var*) of four basic
443 behavioral features: velocity (*V*), temporal changes in velocity (i.e. acceleration, *dV*), bearing
444 (*B*), and temporal changes in bearing (*dB*). These 8 features were represented in the form of
445 histograms. Based on the hypothesis that basic behavioral features are likely to be distributed
446 probabilistically for each state, we chose one of the features whose histogram likely consists of
447 multiple probabilistic distributions, and classified the feature values into several clusters with
448 the EM algorithm (Dempster et al., 1977). The EM algorithm, which is an iterative method to

449 estimate model parameters that maximize the likelihood of the model for data, was used to
450 estimate parameters such as the average and variance of each cluster. Each cluster was then
451 regarded as a behavioral state. As proof-of-principle experiments, we analyzed the trajectories
452 of worms and penguins, whose behavioral states have been studied previously using other
453 methods (Pierce-Shimomura et al., 1999; Yoda et al., 2001).
454
455 The roundworm *Caenorhabditis elegans* has been used as a model animal for quantitative
456 behavioral analysis owing to the ease of tracking behavior (movement for a few cm on agar
457 surface can be easily recorded with an inexpensive high-resolution camera), optical monitoring
458 neural activities, and genetic analyses and manipulations (De Bono and Maricq, 2005). Further,
459 the neuronal wiring in *C. elegans* has been described in complete detail (White et al., 1986). In
460 this study, we focused on the avoidance behavior to the repulsive odor of 2-nonanone (Fig. 2A,
461 left) (Bargmann et al., 1993; Kimura et al., 2010). We chose this behavioral paradigm for the
462 proof-of-principle experiment because the odor avoidance behavior has been quantitatively,
463 although not fully, analyzed previously (Kimura et al., 2010; Yamazoe-Umemoto et al., 2015).
464 The behavior of the worms was recorded with an USB camera for 12 minutes, and the position
465 of each worm's centroid was extracted every second.
466
467 We observed multiple peaks in the histogram corresponding to the variance values for bearing
468 change (*dB_Var*, Fig. 2B). These values were automatically classified into 5 clusters by the EM
469 algorithm (Fig. 2C). Upon mapping the clusters on to the trajectories, we found that cluster 0
470 corresponds to relatively straight part, while the other clusters correspond to more complex
471 parts (Fig. 2D, upper panel).
472
473 Cluster 0 and clusters 1-4 mainly corresponded to “run” and “pirouette,” respectively, which are
474 the two classical behavioral states of worms in various types of sensory behavior (Fig. 2D,

475 lower panel) (Lockery, 2011; Pierce-Shimomura et al., 1999). “Run” constitutes relatively
476 straight movement, while “pirouette” is characterized by short, straight movement divided by
477 frequent large changes in angle (turns and reversals). Runs and pirouettes are usually classified
478 based on a threshold value for the duration between consecutive large angle changes (Pierce-
479 Shimomura et al., 1999), unlike in this method. We collectively regarded clusters 1-4 as “cluster
480 1” because the positional information of worms does not appropriately reflect their actual
481 locations during pirouettes due to insufficient spatiotemporal resolution of the recording system
482 for relatively long-distance navigation, such as odor avoidance behavior (Yamazoe-Umemoto et
483 al., 2015). We found that more than 90% of the cluster 0 and 1 corresponded with the run and
484 pirouette, respectively (Fig. 2E and F). Therefore, we concluded that the STEFTR method
485 properly classified the odor avoidance behavior into distinct behavioral states.
486
487 Next, we applied the same process to the trajectories of penguins obtained using GPS devices.
488 Penguins are good model wild animals for studying long-distance navigation given their
489 relatively large body size for attaching GPS devices and their habit of returning to a colony,
490 which makes the attachment and the recovery of GPS data easy (Yoda, 2018; Yoda et al., 2001).
491 In this study, GPS and depth sensors were attached to 11 penguins from a colony in the
492 Antarctic Continent; the depth sensors were used to evaluate the accuracy of state estimation
493 (see below). The penguins moved by walking and swimming about 10 km for feeding, and each
494 dataset contained up to 2 days of data recordings (Table 1 and Fig. 3A). Like in the case of
495 worms, 8 basic behavioral features were extracted from the penguin trajectory data and
496 represented as histograms. We chose the average velocity (*V_Ave*) as it showed multiple peaks
497 in the histogram. The EM algorithm classified it into 5 clusters (Fig. 3B lower panel, C and D
498 upper panel).
499

Interestingly, the clusters exhibited significantly different distributions in multiple behavioral features. For example, the values for the duration of each bout were much longer in cluster 0 than in clusters 1, 2, and 3 (Fig. 3D, middle panel). In addition, although the clusters were classified only based on the horizontal velocities, the depths for cluster 0 and 1 were significantly closer to zero than those for clusters 2, 3, and 4 (Fig. 3D, lower panel). These results are consistent with the idea that each cluster reflects a behavioral state that is a complex function of multiple behavioral features.

To evaluate whether the clusters actually reflect different behavioral states, we manually classified the penguin behavior into four states (resting, transit by walking, transit by swimming, and diving) based on diving depth (from depth sensor), movement velocity and distance from the colony (both calculated from the GPS positional information) (Yoda et al., 2001). Penguins stayed and rested at the colony (location does not change much; depth is zero), moved on land and ice mainly by walking (location changes relatively slowly; depth is near-zero), swim in the sea to go to the foraging area (location changes quickly; depth is relatively shallow in general and sometimes increases when they are moving towards the feeding area), dive deeply at the foraging area (location does not change much; dives occur continuously in bouts), and then come back to the colony by swimming and walking. The resting at the colony and swimming correlated with clusters 0 and 4, respectively (Fig. 3E and F). In addition, most of clusters 1 and 2 correlated with walking, while about 50% of cluster 3 corresponded to diving (Fig. 3F). Thus, when a behavioral state is classified to a cluster other than cluster 3, the penguin is likely to be resting, walking, or swimming. If a behavioral state is classified to cluster 3, which is ~10% of all the behavior recorded, the penguin is either diving or swimming. Remarkably, although the clustering is only based on the trajectories of 11 penguins for a few days, the false positive rates were less than 10% and the sensitivity of the analysis was greater

525 than 90% in all the cases (Fig. 3G). Thus, we concluded that the STEFTR method can
526 reasonably estimate different behavioral states only based on trajectory data.
527
528 The STEFTR method was also applied to the trajectories of flying seabirds in the wild and rats
529 in the laboratory. The seabirds, *Calonectris leucomelas,* traveled ~100 times longer distances
530 (up to 1,000 km) with ~10 times the speeds compared to penguins (Matsumoto et al., 2017;
531 Yoda et al., 2014). Like in the case of penguins, the average velocity (*V_Ave*) showed multiple
532 peaks and was classified into 5 clusters (Fig. 4A and B) with significantly different profiles for
533 duration and directedness (Fig. 4C). In the case of rats, the variance of velocity change
534 (*dV_Var*) was chosen, and the EM algorithm classified the behavior into 4 clusters (Fig. 4D and
535 E). Interestingly, significant differences were observed in several features such as duration and
536 speed (Fig. 4F). Such information can help ecologists estimate the candidates for feeding areas
537 where fishes may be more densely distributed and discover biologically important marine areas.
538 It can also help neuroscientists in estimating candidate conditions to further explore specific
539 neural activities.
540
541 **Comprehensive extraction of behavioral features modulated by learning**
542 As a second part of the STEFTR method, comprehensive feature extraction was performed by
543 comparing a specific behavioral state in two different conditions, such as cluster 0 of worms
544 before and after learning. Comprehensive semi-automated analysis can be very helpful to
545 compare behavioral features in two conditions. This is because even when the overall result of
546 behavioral responses are different in two conditions, it is still difficult to quantitatively
547 determine which part of the trajectories are different (Fig. 2A left and right, for example).
548 Furthermore, even if several behavioral features are found to be different, it is possible that
549 other more prominent feature differences may exist. We considered that learning-dependent
550 changes in behavior should be one of the best models for comprehensive feature extraction

551 because the differences in behavioral features should reflect learning-dependent changes in
552 neural/brain activities.
553
554 As an useful index for feature extraction, we chose information gain, the index for decision tree
555 analysis (Quinlan, 1986). Binary decision tree analysis is for splitting a dataset into two sub-
556 groups by automatically selecting the best feature and its parameter showing the largest
557 information gain (i.e., difference of uncertainty, or "information entropy", between before and
558 after division). Each data point is then classified into one of the sub-classes based on whether it
559 has a larger or smaller value than an automatically determined threshold. When applied for
560 binary classification, decision tree analysis automatically evaluates the classification
561 performance of a large number of features as designed by the researchers. This analysis results
562 in the extraction of certain features, allowing us to easily understand the utility of particular
563 features in the classification. This approach is substantially different from those that employ
564 support vector machines and/or deep neural networks, wherein the relationships between the
565 classification and the features of the data cannot be easily discerned.
566
567 We first analyzed learning-dependent changes in worm odor avoidance behavior. Worm odor
568 avoidance behavior is enhanced by pre-exposure to the odor as a type of non-associative
569 learning, and pre-exposed worms migrate significantly longer distances from the odor source
570 than control worms do during the same period (Fig. 2A) (Kimura et al., 2010). This
571 phenomenon is interesting because prior exposure to a stimulus generally causes reduction,
572 instead of enhancement, of the response to the stimulus through adaptation or habituation.
573 Although this is a simple form of learning, this odor learning is modulated by multiple
574 neuromodulators, including dopamine, octopamine (the worm counterpart of mammalian
575 noradrenaline), and neuropeptides (Kimura et al., 2010; Yamazoe-Umemoto et al., 2015).
576 Previous quantitative analyses have shown that the enhanced odor avoidance behavior is not

577 caused by changes in speed, but mostly by increases in run duration (Yamazoe-Umemoto et al.,
578 2015). However, this did not rule out the possibility that other behavioral features play more
579 profound effects.
580
581 As an example of comprehensive feature extraction from a behavioral state, we focused on
582 learning-dependent changes in cluster 0 (run) because the values of their centroid migration are
583 quantitatively more reliable than cluster 1 (pirouette) as mentioned above. In addition to the
584 basic behavioral features used for the estimation of behavioral states (*V*, *dV*, *B*, and *dB*), we
585 also calculated directedness (*Dir*) (Gorelik and Gautreau, 2014), and the odor concentration (*C*)
586 and temporal change in odor concentration (*dC*) that each worm experienced during the odor
587 avoidance behavior; *C* and *dC* were calculated based on actual measurements of the dynamic
588 odor gradient (Tanimoto et al., 2017; Yamazoe-Umemoto et al., 2018). For these, we calculated
589 the initiation (*Ini*), middle (*Mid*), termination (*Ter*), and all (*All*) values of a cluster 0 segment
590 (Fig. 5A). In addition, different time windows (1-6 s in this case) were used to calculate these
591 values because a behavioral feature could be apparent only within a specific temporal window
592 (for example, velocity of run (i.e. cluster 0) starts decreasing 2 s before the end of a run (Pierce-
593 Shimomura et al., 1999)). We also calculated durations (*Dur*) of cluster 0 and 1, and the
594 weathervane index (*WV*) (Iino and Yoshida, 2009). Information gain for each of these features
595 were compared between naive/mock and pre-exposed conditions (Fig. 5B, for example). The
596 information gain values for each of the features have been summarized in Table 2, and the
597 details are described in Supplementary Tables 2-11.
598
599 Through this analysis, we were able to find new as well as previously known behavioral
600 features that exhibited learning-dependent changes. First, we found that the duration of each
601 cluster 0 (*Clst0Dur*) exhibited higher information gain (Table 2), which corresponded to
602 significantly increased cluster 0 duration (Fig. 6B). This result is consistent with the findings

603 from previous reports (Kimura et al., 2010; Yamazoe-Umemoto et al., 2015), highlighting the
604 reliability of this method. We also found that the velocity at the beginning of each cluster 0
605 (*V_Ini*) consistently exhibited higher information gain in the average and median values in
606 multiple time windows (Table 2); these values were also significantly different in the pre-
607 exposed worms as compared to the control worms (Fig. 6C). The previous study has not
608 identified this difference as only average values per run (*i.e.* cluster 0) have been calculated in
609 the study (Yamazoe-Umemoto et al., 2015). Although the contribution of this behavioral feature
610 to enhanced odor avoidance is unclear at present, our results indicate that the STEFTR method
611 can reveal characteristic feature(s) under specific conditions, which is difficult for human
612 analyses to accomplish.
613
614 Odor stimuli during runs, which likely drive the worms' odor avoidance behavior, were also
615 found to be consistently modulated (Table 2). In fact, odor concentration (*C*) was significantly
616 lower, and the temporal change in odor concentration (*dC*) was significantly closer to zero (i.e.
617 shallower) in a learning-dependent manner (Fig. 6D and E). Because the previous study
618 demonstrated that worm odor avoidance behavior depends on *dC* rather than *C* at least in the
619 naive condition (Tanimoto et al., 2017), one possibility is that the changes in the responsiveness
620 of worms to *dC* is the underlying reason for the enhanced odor avoidance. However, it is also
621 possible that the odor-experienced worms were somehow located farther away from the odor
622 source than the unexperienced worms, and hence, sensed lower odor concentrations and
623 shallower odor concentration change than the latter.
624
625 **Responsiveness of sensory neurons to odor increase was modulated by the odor learning**
626 If the change in sensitivity to *dC/dt* is the reason underlying enhanced odor avoidance behavior,
627 it should be associated with changes in neural activity. Thus, we analyzed the responsiveness of
628 a likely candidate, ASH nociceptive neurons (Bargmann, 2006; Kaplan, 1996). Previously, we

629 have established the OSB2 microscope system that allows for *in vivo* calcium imaging of *C.*
630 *elegans* neurons in the presence of odor stimuli resembling those that the worms experience
631 during the odor avoidance behavior in the plates (Fig. 7A) (Tanimoto et al., 2017). Using the
632 OSB2 system, we found that ASH neurons are the major sensory neurons to cause pirouettes
633 upon increases in 2-nonanone concentration (Tanimoto et al., 2017). However, whether the
634 ASH response is modulated by 2-nonanone experience has not yet been studied.
635
636 We found that ASH responses were indeed modulated by prior odor experience. When the
637 worms were stimulated with a 5 nM/s odor increase rate, which is the lowest rate of change to
638 cause the threshold-level behavioral response in the previous study (Tanimoto et al., 2017),
639 ASH neurons in naive as well as mock-treated worms exhibited robust responses (Fig. 7B).
640 However, the ASH responses were significantly reduced in the pre-exposed worms (Fig. 7B and
641 C). This suggests that prior odor experience causes a reduction in the neuronal response to a
642 slight increase in odor concentration, subsequently causing longer run durations and enhanced
643 odor avoidance behavior (Fig. 7D).
644
645 **Extracted behavioral features of mutant strains correspond to gene function**
646 Next, we comprehensively analyzed learning-dependent behavioral changes in the mutant *C.*
647 *elegans* strains. Many mutant strains of *C. elegans* showing impaired learning have already
648 been isolated and characterized (Bargmann, 2006; Sasakura and Mori, 2013), and the behavioral
649 abnormalities observed in these mutants should reflect the role of the causal genes in neural
650 function. In fact, we have previously shown that two different genes involved in the enhanced
651 odor avoidance behavior cause different abnormalities in behavioral features when mutated
652 (Yamazoe-Umemoto et al., 2015). However, as the behavioral features exhibited by a mutant

653 strain could be different from one another, identification of abnormal behavioral features is
654 often laborious and time-consuming.
655
656 In addition to studying the previously described mutants with defective enhanced odor
657 avoidance behavior (*egl-3* and *egl-21* for neuropeptide biosynthesis, and *dop-3* for dopamine
658 receptor) (Kass et al., 2001; Suo et al., 2004; Yamazoe-Umemoto et al., 2015), we also analyzed
659 mutant strains found to be involved in the phenomenon in this study (*ocr-2* and *osm-9* for TRP
660 channels; *tax-4* for CNG channel; *eat-4* for vesicular glutamate transporter; *pkc-1* for protein
661 kinase) (Colbert et al., 1997; Komatsu et al., 1996; Land et al., 1994; Lee et al., 1999; Tobin et
662 al., 2002) (Table 3).
663
664 Neuropeptide mutant strains did not exhibit learning-dependent changes in behavioral features,
665 except for the velocity of *egl-3(ok979)*. This result is consistent with the previous finding that
666 neuropeptide signaling is required for the acquisition of odor memory (Yamazoe-Umemoto et
667 al., 2015). *egl-3(ok979)* may have exhibited stronger phenotypes than *egl-3(n589)* because they
668 are nonsense and missense mutants, respectively. Also consistent with the previous report
669 (Yamazoe-Umemoto et al., 2015), the *dop-3* mutants exhibited abnormalities in direction-
670 related behavioral features (*B* and *Dir*) while the changes in cluster 0 durations and velocities
671 are similar to those of wild-type worms (Table 3). Furthermore, with respect to the newly added
672 mutant strain, similar patterns are observed in *ocr-2* and *osm-9* mutants of the TRP channel
673 involved in sensory perception. On the other hand, *tax-4*, which is also involved in sensory
674 perception but expressed in a different set of sensory neurons (Komatsu et al., 1996; Tobin et

675 al., 2002), and *eat-4* and *pkc-1* mutants showed different patterns of abnormalities. Taken
676 together, our results suggest that the patterns of features extracted from mutant strains may
677 reflect functional groupings of the mutated genes. Thus, profiling and classification of extracted
678 mutant features of unknown genes may be useful in the estimation of their physiological
679 functions.
680
681 **Feature extraction of fly sexual behavior**
682 Next, we applied the technique to comprehensive feature extraction of animal behavior under
683 specific situations in two different conditions—heterosexual chasing behavior of *Drosophila*
684 *melanogaster* with or without pheromone sensation. On an experimental tracking system (Fig.
685 8A), male flies chased the target female flies' abdomens after tapping them with their forelegs
686 to sense the cuticular pheromone, although males do not show such chasing behavior before
687 tapping (Kohatsu et al., 2011). While this pheromone-driven behavior has been generally used
688 for the observation of neural activity during courtship behavior in fruit flies, the behavioral
689 features have not yet been elucidated in great detail.
690
691 In this study, we used a tracking system as described in previous studies, where a male was
692 exhibited the moving female abdomen with 8 time left-right round trip after a pheromone
693 sensation (Fig. 8A) (Kohatsu et al., 2011). We detected positive information gains in the
694 velocity, changes in velocity, bearing, and changes in bearing in a pheromone sensation-
695 dependent manner (Fig. 8B). Unexpectedly, the information gains of velocity were higher in the
696 earlier round trips and decreased over the trips (Fig. 8B). This suggests that the pheromonal
697 effect promoting chasing behavior decreases over time. To confirm the result of the STEFTR
698 method, we re-analyzed the speed of the male locomotion along with time-series. Consistent
699 with the STEFTR result, the velocity of males that had tapped the female significantly

700 decreased over the trips (orange group in Fig. 8C; significant differences between rounds 1, 4,
701 and 8), whereas that of control flies (gray group in Fig. 8C) remained mostly unchanged
702 (significant difference only between rounds 1 and 5). Thus, the STEFTR method can even
703 uncover behavioral features that fluctuate over time. The decreased tracking velocity may
704 reflect a decrease in motivation in the fly brain, which can be assessed directly by observing the
705 temporal changes in neuronal activity related to the courtship-motivation circuit in the fly brain
706 (Yamamoto and Koganezawa, 2013; Zhang et al., 2018).
707

708 **Feature extraction of learning-dependent modulation of acoustic navigation in bats**

709 To further demonstrate the general applicability of the method, we examined features of
710 acoustic navigation in bats. We have previously reported that bats improve their flight trajectory
711 in an indoor space with obstacles in a learning-dependent manner (Yamada et al., in revision).
712 Here, we optimized features such as velocity (*V*), distance to the obstacle chain array (*R_obs*
713 and *R_x*), and horizontal bearing of the flight (*B_hori*) for the experimental paradigm (Fig. 9A
714 and B). Interestingly, although the velocity (*V*) itself was modulated by flight experience, the
715 change in velocity (*dV*) was not (Fig. 9C and D), suggesting that bats determine flight speed
716 before initiating navigation, but not during navigation, at least in this experimental conditions.
717 As the vocalizations of bats reflect their attention or decisions (Moss and Surlykke, 2010), our
718 results suggest that the STEFTR method can be used to elucidate such higher brain functions
719 during navigation.
720

721 **DISCUSSION**

722 Measuring and analyzing behavior is one of the most prominent steps in understanding brain
723 function. In order to utilize "behavioral big data", we developed a hybrid
724 supervised/unsupervised technique, the STEFTR method, to estimate behavioral states and to
725 efficiently extract behavioral features solely from the trajectories of animal movement. The

behavioral states of worms and penguins estimated with the STEFTR method were in reasonable agreement with the ones based on previous knowledge, highlighting the validity of our method. In addition, one of the learning-dependent behavioral features extracted from worms corresponded to a change in neural activity. Furthermore, we were able identify temporally dynamic changes through feature extraction from fly courtship behavioral data.

One of the advantages of the STEFTR method is its versatility. Multiple methods have been reported for the behavioral analysis of specific animals under specific conditions (Baek et al., 2002; Branson et al., 2009; Brown et al., 2013; Dankert et al., 2009; Kabra et al., 2013; Mathis et al., 2018; Robie et al., 2017; Stephens et al., 2008; Vogelstein et al., 2014; Wiltschko et al., 2015). However, the animals and experimental conditions for which each of these methods can be applied are rather limited. For example, it is still not easy to effectively and robustly extract an animal's posture from a video image, especially in the wild. Even in laboratories, the parameters generally need to be adjusted again when imaging conditions changed (Dell et al., 2014; Egnor and Branson, 2016). In contrast to these methods, our method allows for behavioral analysis based on positional information that can be extracted from animal video images as well as from different methods such as GPS device.

As a first step of the STEFTR method, we estimated behavioral states from animal trajectories. State estimation is one of the critical processes of movement analysis of animals in the wild as well as of cars and people in ecology and data science, respectively (Egnor and Branson, 2016; Jonsen et al., 2013; Patterson et al., 2008; Zheng, 2015). However, the analytical methods that can be applied to the analysis of various types of animals (and cars and people) are still in debate (Gurarie et al., 2016; Zheng, 2015). In the STEFTR method, we aimed to analyze behavior without previous knowledge of the animal and/or experimental condition and independent of the spatiotemporal scale of the behavior. We analyzed migratory velocity and

752 direction, the most fundamental elements of moving objects, with appropriate moving-averaged
753 data. Interestingly, in case of both worms and penguins, the moving average with a temporal
754 window that is about 1% of median length of the behavioral record provided results that were
755 reasonably consistent with behavioral states categorized by manual labeling (Fig. 2D-F and 3E-
756 G). This result suggests that the ~1% temporal window works as a low-pass filter for properly
757 estimating behavioral states in a scale-free manner. We estimated the number of behavioral
758 states to be equal to or less than 5 (Jonsen et al., 2013), which worked well in the above-
759 mentioned cases.
760
761 State estimation based only on trajectory analysis using the STEFTR method is not perfect, as
762 shown in the case of penguins (Fig. 3E). However, the estimated behavioral states likely
763 provide us with important information for further experiments, such as when and where in the
764 spatiotemporal behavioral profile of the animal behavior should be analyzed in detail, especially
765 in the case where the behavior has not been studied intensively in quantitative manner. For
766 example, relatively small movements at places distant from their nest in the wild may
767 correspond to the feeding area. For neurobiological/physiological analysis, the transition from
768 one state to the other could be triggered by a specific change in the sensory stimulus and
769 associated with specific neural activities. It should also be noted that the STEFTR method
770 allows semi-automatic state estimation and feature extraction, which is suitable for large-scale
771 behavioral analysis of mutant strains of laboratory model animals.
772
773 The Hidden Markov Model (HMM) has been used for the estimation of behavioral states
774 (Gurarie et al., 2016; Jonsen et al., 2013). However, the HMM depends on the idea of the
775 Markov process, wherein the state transition is probabilistic. In general, state transition in
776 animal behavior is triggered by external as well as internal stimuli, which are not always
777 probabilistic. While the STEFTR method assumes a probabilistic normal distribution of a

behavioral feature in a state, it does not assume probabilistic transition between states.

For comprehensive feature extraction, we used information gain, an index used in decision tree analysis. Decision tree analysis is one of the machine learning techniques used for classification. Classification analysis involves classifying new, unlabeled data into appropriate classes using characteristic features and the parameters that have been extracted from the known class-labeled data. In the present study, however, the classification itself was not meaningful because the data were already classified (with/without learning or with/without sex pheromone). Instead, we focused on the procedure in the classification that identifies features useful for distinguishing between the two classes. In other words, behavioral features that are different between two classes (*i.e.,* conditions) should be able to effectively classify the behavioral data of animals in two conditions.

Although the STEFTR method does not directly provide information about brain/neural activity underlying animal behavior, it provides us with clues required to formulate hypotheses related to the experimental investigation of the neural activity, as shown in the case of learning-dependent changes in behavioral features and neural activity (Fig. 7). For example, animals in the wild experience continuously changing visual, auditory, and olfactory stimuli, each of which contain multi-dimensional information (color, shape, tone, different chemical compounds, etc.). Therefore, it is difficult to identify which aspect(s) of the particular stimulus actually triggers a change in animal behavior. Estimation of behavioral states using the STEFTR method will allow us to identify the behavior-triggering stimulus by focusing on the timing and/or place of the behavioral transition. Similarly, large-scale recording of neural activities from moving animals in the laboratory itself is difficult to interpret. However, state estimation and feature extraction of the behavior will greatly facilitate the identification of neural activities that are associated with behavioral transitions and/or specific behavioral features.

804

805 **FIGURE LEGENDS**

806 Figure 1. A workflow of the STEFTR method. Trajectory data of animals are used to calculate 8
807 basic behavioral features, and one of them are analyzed by the EM algorithm to estimate
808 behavioral states (upper panels). From a behavioral state, behavioral features are
809 comprehensively evaluated by using information gain (lower panels).

810

811 Figure 2. State estimation of worms. (A) Examples of the trajectories of 3 worms before (left) or
812 after odor learning (right) in 12 min of 2-nonanone avoidance assay, overlaid on a schematic
813 drawing of a 9 cm plate. One of the trajectories is magnified below. (B) The histograms of 8
814 basic behavioral features. Horizontal and vertical axes indicate the values and the density of
815 each feature. *dB_Var* was chosen (red square) for the EM analysis. (C) Classification of *dB_Var*
816 by the EM algorithm. Each cluster distribution (cluster 0, 1, 2, 3, and 4 are indicated by purple,
817 blue, green, orange, and red dashed lines, respectively, similarly to the following figures) and
818 the sum of clusters (red solid line) are shown. Inset is a magnified view. (D) Comparison of the
819 cluster 0 and 1 (upper panel; purple and green, respectively) with run and pirouette (lower
820 panel; blue and red, respectively) on a trajectory. The initial 2 minutes (gray in both panels)
821 were excluded from the analyses because worms do not avoid the odor during the period
822 (Kimura et al., 2010). (E) Event numbers of cluster 0 (left) and 1 (right) in run and pirouette. (F)
823 Matching matrix of the state estimation.

824

Figure 3. State estimation of penguins. (A) The trajectories of 11 penguins (black lines) on the Antarctic Continent (gray area; white area is the sea). Horizontal and vertical axes indicate longitude and latitude, respectively. (B) The histograms of 8 basic behavioral features (upper panels) and the classification by the EM algorithm. *V_Ave* was chosen (red square) and classified into 5 clusters (lower panels). (C) Mapping of the five clusters on the trajectory. (D) Differences in velocity, duration, and depth among the clusters. Each dot represents a cluster bout, and the bars represent the median and the first and third quartiles. Significant differences among clusters suggest that the clusters correspond to different behavioral states. Statistical values were calculated using Kruskal-Wallis test with *post hoc* Dunn's test. $^{**}p < 0.01$, $^{***}p < 0.001$. (E) An example of comparison of the clusters from the STEFTR analysis with the behavioral states by manually classified labels, which is based on diving depth, movement speed recorded from GPS data, and distance from the colony. (F) Event numbers of each clusters. (G) Matching matrix of the state estimation. The statistical details are described in Supplementary Table 1.

Figure 4. State estimation of flying seabirds in the Pacific Ocean (A-C) and rats in the open maze (D-F). (A) From the histograms of 8 basic features of flying seabirds (left panels), the average of velocity (*V_Ave*) was chosen and classified into 5 clusters (right panel). (B) Mapping of the clusters on the trajectory. The gray region is the northern part of Japan (Tohoku and Hokkaido area), while the white region is the sea. (C) Significant differences were observed in duration (left) and directedness (right). (D) From the histograms of 8 basic features of rats (left panels), the variance of velocity change (d*V_Var*) was chosen and classified into 4 clusters

847 (right panel). (E) An example of trajectories of one rat. (F) Significant differences were
848 observed in duration (left) and velocity (right). Each dot represents a cluster bout, and the bars
849 represent the median and the first and third quartiles. Statistical values were calculated using
850 Kruskal-Wallis test with *post hoc* Dunn's test. ***$p < 0.001$. The statistical details are described
851 in Supplementary Table 1.
852
853 Figure 5. Feature extraction of worm behavior. (A) Schematic drawing of the behavioral
854 features. (B) One example (*V_Ini*; average of time window 2) of calculation of information
855 gain.
856
857 Figure 6. Extracted features modulated by the odor learning. (A) Enhanced odor avoidance
858 behavior in worms caused by odor pre-exposure. Left: End-points of 25 worms in each
859 condition plotted on a schematic representation of the assay plate. Right: Avoidance distance
860 (distance between the center line of the plate and the end-point of the behavior) of each worm.
861 Each dot represents a worm. Significant differences were observed between the pre-exposed
862 worms and the naive and mock-treated worms (***$p < 0.001$, Kruskal-Wallis test with *post-hoc*
863 Dunn's test). (B, C, D and E) Distributions of extracted features. Duration (B), the initial value
864 of velocity (C; average of time window 2), the average odor concentration (D), and the average
865 odor concentration change (E) of each run (***$p < 0.001$, Kruskal-Wallis test with *post-hoc*
866 Dunn's test). Each dot represents a cluster bout, and the bars represent the median and the first
867 and third quartiles. The statistical details are described in Supplementary Table 1.
868

869 Figure 7. Sensory responses to slight increases in odor concentration were reduced by pre-
870 exposure to the odor. (A) A schematic drawing of calcium imaging of neural activity of worms
871 under odor stimuli. Several immobilized worms were simultaneously exposed to an odor flow
872 whose concentration was changed by controlling syringe pumps. (B) Responses
873 (GCaMP/mCherry) of ASH neurons in naive (n = 25), mock-treated (n = 29), and pre-exposed
874 (n = 26) worms. Thick lines with gray shadows indicate mean ± standard error of mean, while
875 thin lines indicate individual responses. (C) Distributions of peak values during the odor-
876 increasing phase (t = 40-80 s) shown in panel B. The bars represent the median. (***$p < 0.001$,
877 Kruskal-Wallis test with *post-hoc* Dunn's test). (D) A model relationship between odor
878 concentration change and behavioral response during navigation along the odor gradient. When
879 naive and mock-treated worms sensed a slight increase in the odor concentration, which is a
880 sign of migrating in the wrong direction, they stopped a run and started a pirouette to search for
881 a new direction. In contrast, the pre-exposed worms did not respond to a slight increase in odor
882 concentration, leading to longer run durations (and shorter pirouette durations in total as a
883 consequence), which likely contribute to the enhanced avoidance distance. The statistical details
884 are described in Supplementary Table 1.
885
886 Figure 8. Pheromone-driven responses of male fruit flies decreased over time. (A) A schematic
887 drawing of the experimental setup. A female fly was actuated leftward and rightward in front of
888 the male fly. The locomotion of the male fly was monitored by an optical sensor, which
889 recorded lateral (Δx) and forward (Δy) movements at 4 Hz. (B) Information gain. Darker and
890 lighter colors mean larger and smaller values, respectively. (C) Distribution of velocity in the

chasing behavior of male flies. Control (without female tapping, gray dots) and experimental (with female tapping, orange dots) groups are shown. Solid and dotted lines represent the median and the first and third quartiles, respectively. Asterisks indicate the statistical significance between the control and test groups (Mann-Whitney U test followed by Bonferroni test for multiple comparison correction, $p < 0.05$). Different characters in each group indicate statistical significance among rounds (Steel-Dwass test, $p < 0.05$). The statistical details are described in Supplementary Tables 12 and 13.

Figure 9. Learning-dependent changes in bat acoustic navigation. (A) The experimental setup for monitoring the 3D flight trajectory of a bat during obstacle avoidance flight in a chamber. (B) Representative flight trajectories of a bat in the horizontal plane during repeated flights in the obstacle course. The figure on the top combines the first three (red) and last three (blue) flight trajectories. Each behavioral feature was collected in three segments: earlier, middle, and later terms. The figure on the bottom shows an expanded view of the earlier term in the first flight. Definition of the horizontal bearing of the flight (*B_hori*), distance (*R_obs*), and bearing (*B_obs*) of the bat to the nearest edge point of the obstacle chain array, longitudinal directional distance to the frontal chain array (*R_x*), and lateral directional distance to the inside pitch of the chains array (*R_y*) are indicated. Time windows for the analysis of each behavioral feature were 0.1, 0.2, or 0.3 s before or while ($t = 0$) passing through the chain array. (C) A list of extracted features of bat acoustic navigation modulated by flight experience. (D) Distributions of *V(-0.3)* and *dV(-0.3)* are plotted. The bars represent the median and the first and third quartiles. (*$p <$ 0.05, Kruskal-Wallis test with *post-hoc* Dunn's test). The statistical details are described in Supplementary Table 1.

ACKNOWLEDGEMENTS

We thank Drs. André Brown, Katsuyoshi Matsushita, Ken-ichi Hironaka, Takuma Degawa, Daisuke Yamamoto, Soh Kohatsu and the Kimura laboratory members for suggestions and comments on this work. Nematode strains were provided by the Caenorhabditis Genetics Center (funded by the NIH Office of Research Infrastructure Programs P40 OD010440) and by National Bioresource Project funded by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. This work was supported by Interdisciplinary graduate school program for systematic understanding of health and disease (for S.J.Y.), by the Tohoku Ecosystem-Associated Marine Science (for K. S.), by KAKENHI 25249020 and JP16H06536 (for K.H.), 24681006, JP16H01769, JP16H06541 (for K.Y.), 16H04655, 18H05069, and17K19450 (for A.K.), JP16H06542 (for S.H.), JP16H06539 (for T.M.), JP16H06545 (for K.D.K.) from the MEXT, by PRESTO JPMJPR14D8 (for S.H.) by JST, and by Osaka University Co-Creation Program (for T.M. and K.D.K.).

AUTHOR CONTRIBUTIONS

S.J.Y, T.M. and K.D.K. designed the experiments, S.J.Y, K.O., K.I., N.K., T.K., D. T., Y.Y., Y.I., F.H., K.F., Y.T., A.Y.-U., K.H., K.S., K.Y., A.T., Y.I., A.K., S.H. and K.D.K. performed the experiments, S.J.Y. and T.M. analyzed the data, and S.J.Y. and K.D.K. wrote the manuscript. All authors reviewed the manuscript.

REFERENCES

Alivisatos, A. P., Chun, M., Church, G. M., Greenspan, R. J., Roukes, M. L., and Yuste, R. (2012). The Brain Activity Map Project and the Challenge of Functional Connectomics. *Neuron* 74, 970–974. doi:10.1016/j.neuron.2012.06.006.

Anderson, D. J., and Perona, P. (2014). Toward a Science of Computational Ethology. *Neuron* 84, 18–31. doi:10.1016/j.neuron.2014.09.005.

Baek, J.-H., Cosman, P., Feng, Z., Silver, J., and Schafer, W. R. (2002). Using machine vision

942 to analyze and classify Caenorhabditis elegans behavioral phenotypes quantitatively. *J*
943 *Neurosci Methods* 118, 9–21.

944 Bargmann, C. I. (2006). Chemosensation in C. elegans. *WormBook*, 1–29.
945 doi:10.1895/wormbook.1.123.1.

946 Bargmann, C. I., Hartwieg, E., and Horvitz, H. R. (1993). Odorant-selective genes and neurons
947 mediate olfaction in C. elegans. *Cell* 74, 515–527.

948 Berman, G. J. (2018). Measuring behavior across scales. *BMC Biol* 16, 23. doi:10.1186/s12915-
949 018-0494-7.

950 Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer Verlag.

951 Branson, K., Robie, A. A., Bender, J., Perona, P., and Dickinson, M. H. (2009). High-
952 throughput ethomics in large groups of Drosophila. *Nat Meth* 6, 451–457.
953 doi:10.1038/nmeth.1328.

954 Brenner, S. (1974). The genetics of Caenorhabditis elegans. *Genetics* 77, 71–94.

955 Brown, A. E. X., and de Bivort, B. (2018). Ethology as a physical science. *Nat Phys* 14, 653–
956 657. doi:10.1038/s41567-018-0093-0.

957 Brown, A. E. X., Yemini, E. I., Grundy, L. J., Jucikas, T., and Schafer, W. R. (2013). A
958 dictionary of behavioral motifs reveals clusters of genes affecting Caenorhabditis elegans
959 locomotion. *Proc Natl Acad Sci USA* 110, 791–796. doi:10.1073/pnas.1211447110.

960 Colbert, H. A., Smith, T. L., and Bargmann, C. I. (1997). OSM-9, a novel protein with
961 structural similarity to channels, is required for olfaction, mechanosensation, and olfactory
962 adaptation in Caenorhabditis elegans. *J Neurosci* 17, 8259–8269.

963 Dankert, H., Wang, L., Hoopfer, E. D., Anderson, D. J., and Perona, P. (2009). Automated
964 monitoring and analysis of social behavior in Drosophila. *Nat Meth* 6, 297–303.
965 doi:10.1038/nmeth.1310.

966 De Bono, M., and Maricq, A. V. (2005). Neuronal substrates of complex behaviors in C.
967 elegans. *Annu Rev Neurosci* 28, 451–501. doi:10.1146/annurev.neuro.27.070203.144259.

Dell, A. I., Bender, J. A., Branson, K., Couzin, I. D., de Polavieja, G. G., Noldus, L. P. J. J., et al. (2014). Automated image-based tracking and its application in ecology. *Trends in Ecology & Evolution* 29, 417–428. doi:10.1016/j.tree.2014.05.004.

Dempster, A. P., Laird, N. M., Rubin, D. B., 1977 (1977). Maximum likelihood from incomplete data via the EM algorithm. *J Royal Stat Soc B* 39, 1–38. doi:10.2307/3315718.

Diba, K., and Buzsáki, G. (2008). Hippocampal network dynamics constrain the time lag between pyramidal cells across modified environments. *J Neurosci* 28, 13448–13456. doi:10.1523/JNEUROSCI.3824-08.2008.

Egnor, S. E. R., and Branson, K. (2016). Computational Analysis of Behavior. *Annu Rev Neurosci* 39, 217–236. doi:10.1146/annurev-neuro-070815-013845.

Frank, E., Hall, M. A., and Witten, I. H. (2016). "The WEKA Workbench (Online Appendix)," in *Data Mining* (Morgan Kaufmann).

Gomez-Marin, A., Paton, J. J., Kampff, A. R., Costa, R. M., and Mainen, Z. F. (2014). Big behavioral data: psychology, ethology and the foundations of neuroscience. *Nat Neurosci* 17, 1455–1462. doi:10.1038/nn.3812.

Gorelik, R., and Gautreau, A. (2014). Quantitative and unbiased analysis of directional persistence in cell migration. *Nat Protoc* 9, 1931–1943. doi:10.1038/nprot.2014.131.

Gurarie, E., Bracis, C., Delgado, M., Meckley, T. D., Kojola, I., and Wagner, C. M. (2016). What is the animal doing? Tools for exploring behavioural structure in animal movements. *J Anim Ecol* 85, 69–84. doi:10.1111/1365-2656.12379.

Iino, Y., and Yoshida, K. (2009). Parallel use of two behavioral mechanisms for chemotaxis in Caenorhabditis elegans. *J Neurosci* 29, 5370–5380. doi:10.1523/JNEUROSCI.3633-08.2009.

Jonsen, I. D., Basson, M., Bestley, S., Bravington, M. V., Patterson, T. A., Pedersen, M. W., et al. (2013). State-space models for bio-loggers: A methodological road map. *Deep Sea Research Part II: Topical Studies in Oceanography* 88–89, 34–46. doi:10.1016/j.dsr2.2012.07.008.

995 Kabra, M., Robie, A. A., Rivera-Alba, M., Branson, S., and Branson, K. (2013). JAABA:
996 Interactive machine learning for automatic annotation of animal behavior. *Nat Meth* 10,
997 64–67. doi:10.1038/nmeth.2281.

998 Kaplan, J. M. (1996). Sensory signaling in Caenorhabditis elegans. *Curr Opin Neurobiol* 6,
999 494–499.

1000 Kass, J., Jacob, T. C., Kim, P., and Kaplan, J. M. (2001). The EGL-3 proprotein convertase
1001 regulates mechanosensory responses of Caenorhabditis elegans. *J Neurosci* 21, 9265–9272.

1002 Kimura, K. D., Fujita, K., and Katsura, I. (2010). Enhancement of odor avoidance regulated by
1003 dopamine signaling in Caenorhabditis elegans. *J Neurosci* 30, 16365–16375.
1004 doi:10.1523/JNEUROSCI.6023-09.2010.

1005 Kohatsu, S., and Yamamoto, D. (2015). Visually induced initiation of Drosophila innate
1006 courtship-like following pursuit is mediated by central excitatory state. *Nature*
1007 *Communications* 6, 6457. doi:10.1038/ncomms7457.

1008 Kohatsu, S., Koganezawa, M., and Yamamoto, D. (2011). Female contact activates male-
1009 specific interneurons that trigger stereotypic courtship behavior in Drosophila. *Neuron* 69,
1010 498–508. doi:10.1016/j.neuron.2010.12.017.

1011 Kokubun, N., Lee, W. Y., Kim, J.-H., and Takahashi, A. (2015). Chinstrap penguin foraging
1012 area associated with a seamount in Bransfield Strait, Antarctica. *Polar Science* 9, 393–400.
1013 doi:10.1016/j.polar.2015.10.001.

1014 Komatsu, H., Mori, I., Rhee, J. S., Akaike, N., and Ohshima, Y. (1996). Mutations in a cyclic
1015 nucleotide-gated channel lead to abnormal thermosensation and chemosensation in C.
1016 elegans. *Neuron* 17, 707–718.

1017 Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., and Poeppel, D. (2017).
1018 Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron* 93, 480–490.
1019 doi:10.1016/j.neuron.2016.12.041.

1020 Land, M., Islas-Trejo, A., and Rubin, C. S. (1994). Origin, properties, and regulated expression
1021 of multiple mRNAs encoded by the protein kinase C1 gene of Caenorhabditis elegans. *J.*

Biol. Chem. 269, 14820–14827.

Landhuis, E. (2017). Neuroscience: Big brain, big data. *Nature* 541, 559–561. doi:10.1038/541559a.

Lee, R. Y., Sawin, E. R., Chalfie, M., Horvitz, H. R., and Avery, L. (1999). EAT-4, a homolog of a mammalian sodium-dependent inorganic phosphate cotransporter, is necessary for glutamatergic neurotransmission in caenorhabditis elegans. *J Neurosci* 19, 159–167.

Lewis, J. A., Wu, C. H., Berg, H., and Levine, J. H. (1980). The genetics of levamisole resistance in the nematode Caenorhabditis elegans. *Genetics* 95, 905–928.

Lockery, S. R. (2011). The computational worm: spatial orientation and its neuronal basis in C. elegans. *Curr Opin Neurobiol* 21, 782–790. doi:10.1016/j.conb.2011.06.009.

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., et al. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci*, 1–12. doi:10.1038/s41593-018-0209-y.

Matsumoto, S., Yamamoto, T., Yamamoto, M., Zavalaga, C. B., and Yoda, K. (2017). Sex-related differences in the foraging movement of streaked shearwaters calonectris leucomelas breeding on Awashima Island in the sea of Japan. *Ornithological Science* 16, 23–32. doi:10.2326/osj.16.23.

Mizuseki, K., Diba, K., Pastalkova, E., Teeters, J., Sirota, A., and Buzsáki, G. (2014). Neurosharing: large-scale data sets (spike, LFP) recorded from the hippocampal-entorhinal system in behaving rats. *F1000Res* 3, 98. doi:10.12688/f1000research.3895.1.

Mizuseki, K., Sirota, A., Pastalkova, E., and Buzsáki, G. (2009). Theta Oscillations Provide Temporal Windows for Local Circuit Computation in the Entorhinal-Hippocampal Loop. *Neuron* 64, 267–280. doi:10.1016/j.neuron.2009.08.037.

Moss, C. F., and Surlykke, A. (2010). Probing the natural scene by echolocation in bats. *Front Behav Neurosci* 4. doi:10.3389/fnbeh.2010.00033.

Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., and Matthiopoulos, J. (2008). State-space models of individual animal movement. *Trends in Ecology & Evolution* 23, 87–94.

doi:10.1016/j.tree.2007.10.009.

Pierce-Shimomura, J. T., Morse, T. M., and Lockery, S. R. (1999). The fundamental role of pirouettes in Caenorhabditis elegans chemotaxis. *J Neurosci* 19, 9557–9569.

Quinlan, J. R. (1986). Induction of Decision Trees. *Mach Learn* 1, 81–106. doi:10.1023/A:1022643204877.

Quinlan, J. R. (1992). *C4.5*. San Mateo: Morgan Kaufmann.

Robie, A. A., Hirokawa, J., Edwards, A. W., Umayam, L. A., Lee, A., Phillips, M. L., et al. (2017). Mapping the Neural Substrates of Behavior. *Cell* 170, 393–406.e28. doi:10.1016/j.cell.2017.06.032.

Sasakura, H., and Mori, I. (2013). Behavioral plasticity, learning, and memory in C. elegans. *Curr Opin Neurobiol* 23, 92–99. doi:10.1016/j.conb.2012.09.005.

Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N. G., Palmer, A. E., and Tsien, R. Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotech* 22, 1567–1572. doi:10.1038/nbt1037.

Stephens, G. J., Johnson-Kerner, B., Bialek, W., and Ryu, W. S. (2008). Dimensionality and Dynamics in the Behavior of C. elegans. *PLoS Comput Biol* 4, e1000028. doi:10.1371/journal.pcbi.1000028.

Suo, S., Ishiura, S., and Van Tol, H. H. M. (2004). Dopamine receptors in C. elegans. *Eur. J. Pharmacol.* 500, 159–166. doi:10.1016/j.ejphar.2004.07.021.

Tanimoto, Y., Yamazoe-Umemoto, A., Fujita, K., Kawazoe, Y., Miyanishi, Y., Yamazaki, S. J., et al. (2017). Calcium dynamics regulating the timing of decision-making in C. elegans. *eLife* 6, e21629. doi:10.7554/eLife.21629.

Tian, L., Hires, S. A., Mao, T., Huber, D., Chiappe, M. E., Chalasani, S. H., et al. (2009). Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nat Meth* 6, 875–881. doi:10.1038/nmeth.1398.

Tobin, D., Madsen, D., Kahn-Kirby, A., Peckol, E., Moulder, G., Barstead, R., et al. (2002). Combinatorial expression of TRPV channel proteins defines their sensory functions and subcellular localization in C. elegans neurons. *Neuron* 35, 307–318.

Vogelstein, J. T., Park, Y., Ohyama, T., Kerr, R. A., Truman, J. W., Priebe, C. E., et al. (2014). Discovery of Brainwide Neural-Behavioral Maps via Multiscale Unsupervised Structure Learning. *Science* 344, 386–392. doi:10.1126/science.1250298.

White, J. G., Southgate, E., Thomson, J. N., and Brenner, S. (1986). The Structure of the Nervous System of the Nematode Caenorhabditis elegans. *Philosophical Transactions of the Royal Society B: Biological Sciences* 314, 1–340. doi:10.1098/rstb.1986.0056.

Wiltschko, A. B., Johnson, M. J., Iurilli, G., Peterson, R. E., Katon, J. M., Pashkovski, S. L., et al. (2015). Mapping Sub-Second Structure in Mouse Behavior. *Neuron* 88, 1121–1135. doi:10.1016/j.neuron.2015.11.031.

Yamada, Y., Hiryu, S., and Watanabe, Y. (2016). Species-specific control of acoustic gaze by echolocating bats, Rhinolophus ferrumequinum nippon and Pipistrellus abramus, during flight. *J Comp Physiol A* 202, 791–801. doi:10.1007/s00359-016-1121-0.

Yamamoto, D., and Koganezawa, M. (2013). Genes and circuits of courtship behaviour in drosophila males. *Nat Rev Neurosci* 14, 681–692. doi:10.1038/nrn3567.

Yamazoe-Umemoto, A., Fujita, K., Iino, Y., Iwasaki, Y., and Kimura, K. D. (2015). Modulation of different behavioral components by neuropeptide and dopamine signalings in non-associative odor learning of Caenorhabditis elegans. *Neurosci Res* 99, 22–33. doi:10.1016/j.neures.2015.05.009.

Yamazoe-Umemoto, A., Iwasaki, Y., and Kimura, K. (2018). Measuring Spatiotemporal Dynamics of Odor Gradient for Small Animals by Gas Chromatography. *BIO-PROTOCOL* 8. doi:10.21769/bioprotoc.2797.

Yoda, K. (2018). Advances in bio-logging techniques and their application to study navigation in wild seabirds. *Advanced Robotics* 58, 1–10. doi:10.1080/01691864.2018.1553686.

Yoda, K., Naito, Y., Sato, K., Takahashi, A., Nishikawa, J., Ropert-Coudert, Y., et al. (2001). A

new technique for monitoring the behaviour of free-ranging Adélie penguins. *Journal of Experimental Biology* 204, 685–690.

Yoda, K., Shiomi, K., and Sato, K. (2014). Foraging spots of streaked shearwaters in relation to ocean surface currents as identified using their drift movements. *Progress in Oceanography* 122, 54–64. doi:10.1016/j.pocean.2013.12.002.

Zhang, S. X., Miner, L. E., Boutros, C. L., Rogulja, D., and Crickmore, M. A. (2018). Motivation, Perception, and Chance Converge to Make a Binary Decision. *Neuron* 99, 376–388.e6. doi:10.1016/j.neuron.2018.06.014.

Zheng, Y. (2015). Trajectory Data Mining. *ACM Trans. Intell. Syst. Technol.* 6, 1–41. doi:10.1145/2743025.

Table 1. Summary of recording and analysis conditions of animal behavior.

Animal	worms	penguins	flying seabirds	rats	flies	bats
Condition	naive, mock, preexposed	1	1	1	with or without pheromone	with or without learning
Animal # / condition	50 each	11	54	1	27, 30	3 each
Trajectory # / condition	50 each	11	54	45	27, 30	9 each
Recording time (min-max)	600 s	1626-2787 min	484-8293 min	1031-3194 s	16 s	1.67-2.53 s
Recording time (median)	600 s	2279 min	972 min	1822 min	16 s	1.95 s
Time unit	1 s	1 min	1 min	1 s	0.25 s	0.008 s
Time window for moving average	12 s	20 min	10 min	20 s	-	-
Feature used for EM	*dB_Var*	*V_Ave*	*V_Ave*	*dV_Var*	-	-
Length unit	mm	m	km	cm	a.u.	m
Approx. distance (min-max)	5-30 mm	1-10 km	100-2000 km	180 × 180 cm	1 - 500 mm	6 m

1113

Table 2. Learning-dependent features extracted from cluster 0 in odor avoidance behavior of wild-type worms.

time	value	*Clst0 Dur*	*Clst1 Dur*	*V*				*dV*				*B*				*dB*				*Dir*				*C*				*dC*				*WV*
		All	All	All	Ini	Mid	Ter	All	Ini	Mid	Ter	All	Ini	Mid	Ter	All	Ini	Mid	Ter	All	Ini	Mid	Ter	All	Ini	Mid	Ter	All	Ini	Mid	Ter	All
-	Ave	0.098	0.057		-	-	-		-	-	-		-	-	-	0.072	-	-	-		-	-	-	0.163	-	-	-	0.088	-	-	-	
-	Med	-	-	0.040	-	-	-		-	-	-		-	-	-	0.102	-	-	-	-	-	-	-	0.160	-	-	-	0.088	-	-	-	
-	Q1st	-	-		-	-	-		-	-	-		-	-	-	0.065	-	-	-	-	-	-	-	0.164	-	-	-	0.107	-	-	-	
-	Q3rd	-	-		-	-	-		-	-	-		-	-	-	0.083	-	-	-	-	-	-	-	0.161	-	-	-	0.080	-	-	-	
-	Var	-	-		-	-	-		-	-	-		-	-	-		-	-	-	-	-	-	-		-	-	-	0.080	-	-	-	
-	1 %ile	-	-		-	-	-		-	-	-		-	-	-		-	-	-	-	-	-	-	0.163	-	-	-	0.090	-	-	-	
-	5 %ile	-	-		-	-	-		-	-	-		-	-	-		-	-	-	-	-	-	-	0.165	-	-	-	0.100	-	-	-	
-	95 %ile	-	-		-	-	-		-	-	-		-	-	-		-	-	-	-	-	-	-	0.153	-	-	-	0.108	-	-	-	
-	99 %ile	-	-		-	-	-		-	-	-		-	-	-		-	-	-	-	-	-	-	0.148	-	-	-	0.060	-	-	-	
1	Ave	-	-	-	0.114			-				-				-				-	-	-	-	-	0.147	0.162	0.162	-	0.116	0.093	0.104	-
2	Ave	-	-	-	0.120		0.049	-				-				-		0.048		-				-	0.148	0.166	0.163	-	0.111	0.105	0.122	-
3	Ave	-	-	-	0.109			-				-				-		0.049		-				-	0.151	0.163	0.162	-	0.110	0.099	0.128	-
3	Med	-	-	-	0.100		0.039	-				-				-				-	-	-	-	-	0.149	0.162	0.162	-	0.114	0.089	0.134	-
3	Var	-	-	-				-				-				-				-	-	-	-	-	0.111	0.087	0.041	-		0.060	0.050	-
4	Ave	-	-	-	0.099			-				-				-		0.050		-				-	0.147	0.166	0.165	-	0.082	0.098	0.122	-
4	Med	-	-	-	0.100			-				-				-				-	-	-	-	-	0.149	0.166	0.165	-	0.085	0.097	0.145	-
4	Var	-	-	-				-				-				-				-	-	-	-	-	0.102	0.083	0.040	-	0.057	0.082	0.082	-
5	Ave	-	-	-				-				-				-		0.048		-				-	0.146	0.163	0.166	-	0.090	0.112	0.148	-
5	Med	-	-	-	0.078			-				-				-				-	-	-	-	-	0.148	0.162	0.165	-	0.093	0.110	0.125	-
5	Var	-	-	-				-				-				-				-	-	-	-	-	0.097	0.092	0.054	-	0.061	0.112	0.078	-
6	Ave	-	-	-	0.072			-				-				-		0.056		-				-	0.148	0.164	0.146	-	0.083	0.112	0.169	-
6	Med	-	-	-	0.075			-				-				-				-	-	-	-	-	0.148	0.166	0.147	-	0.085	0.111	0.149	-
6	Var	-	-	-				-				-				-				-	-	-	-	-	0.089	0.095	0.060	-		0.116	0.081	-

Numbers are the sum of information gains obtained from the comparisons of naive vs. preexp. and mock. vs. preexp. Darker and lighter colors correspond to larger and smaller values, respectively. Blank cells, no value for information gain was obtained. "-", not calculated.

Table 3. Patterns of learning-dependent behavioral features of cluster 0 in odor avoidance behavior of mutant worms.

Strain	*Clst0 Dur*	*Clst1 Dur*	*V*				*dV*				*B*				*dB*				*Dir*				*C*				*dC*			
	All	*All*	*All*	*Ini*	*Mid*	*Ter*	*All*	*Ini*	*Mid*	*Ter*	*All*	*Ini*	*Mid*	*Ter*	*All*	*Ini*	*Mid*	*Ter*	*All*	*Ini*	*Mid*	*Ter*	*All*	*Ini*	*Mid*	*Ter*	*All*	*Ini*	*Mid*	*Ter*
wild-type	Up	D/U	Down	Down		Down									Down		Down						Down	Down	Down	Down	Up	Up	Up	Up
egl-3(n589)																							Up	Up	Up	Up				
egl-3(ok979)			Down	Down	Down	Down																	Up	Up	Up	Up				
egl-21																														
dop-3	Up		Down		Down	Down					Up		Up	Up				Up	Down			Down					Up		Up	Up
ocr-2			Up			Up																								
osm-9	Up		Up	Up	Up	Up																								
tax-4						Down									Down	U/D							D/U	D/U	D/U	D/U				
eat-4			Down	Down		Down																	Down							
pkc-1															Down		Down						Up				Down			

Darker and lighter colors mean larger and smaller information gain values, respectively. “Up” and “Down” indicates the value changes of pre-exposed worms compared to naive and mock-treated worms. “U/D” or “D/U” indicates the value of pre-exposed worms is higher than naive and lower than mock-treated worms, and *vice versa*. Blank cells, no value for information gain was obtained. The details of information gain are described in Supplementary Tables 2-11.

1116

Figure 1

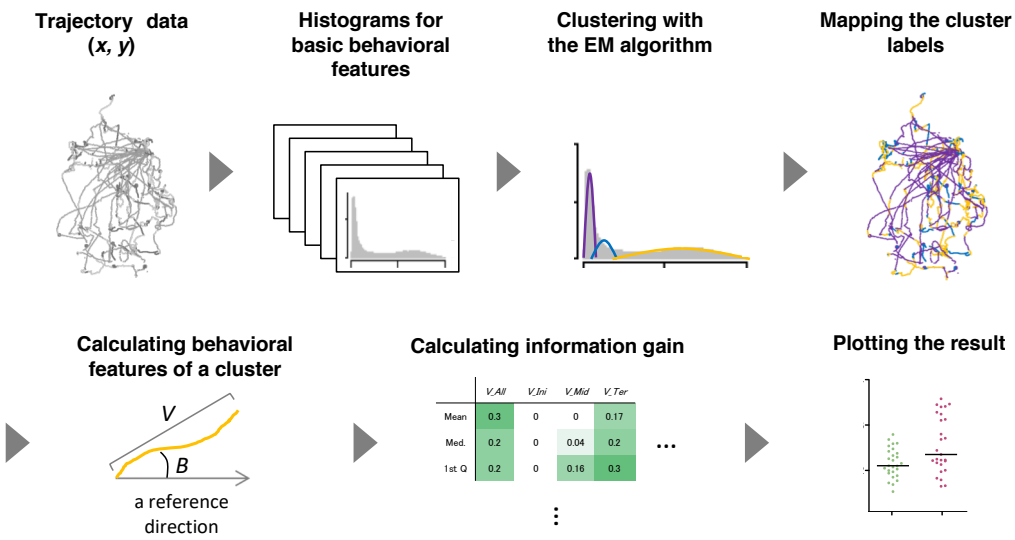
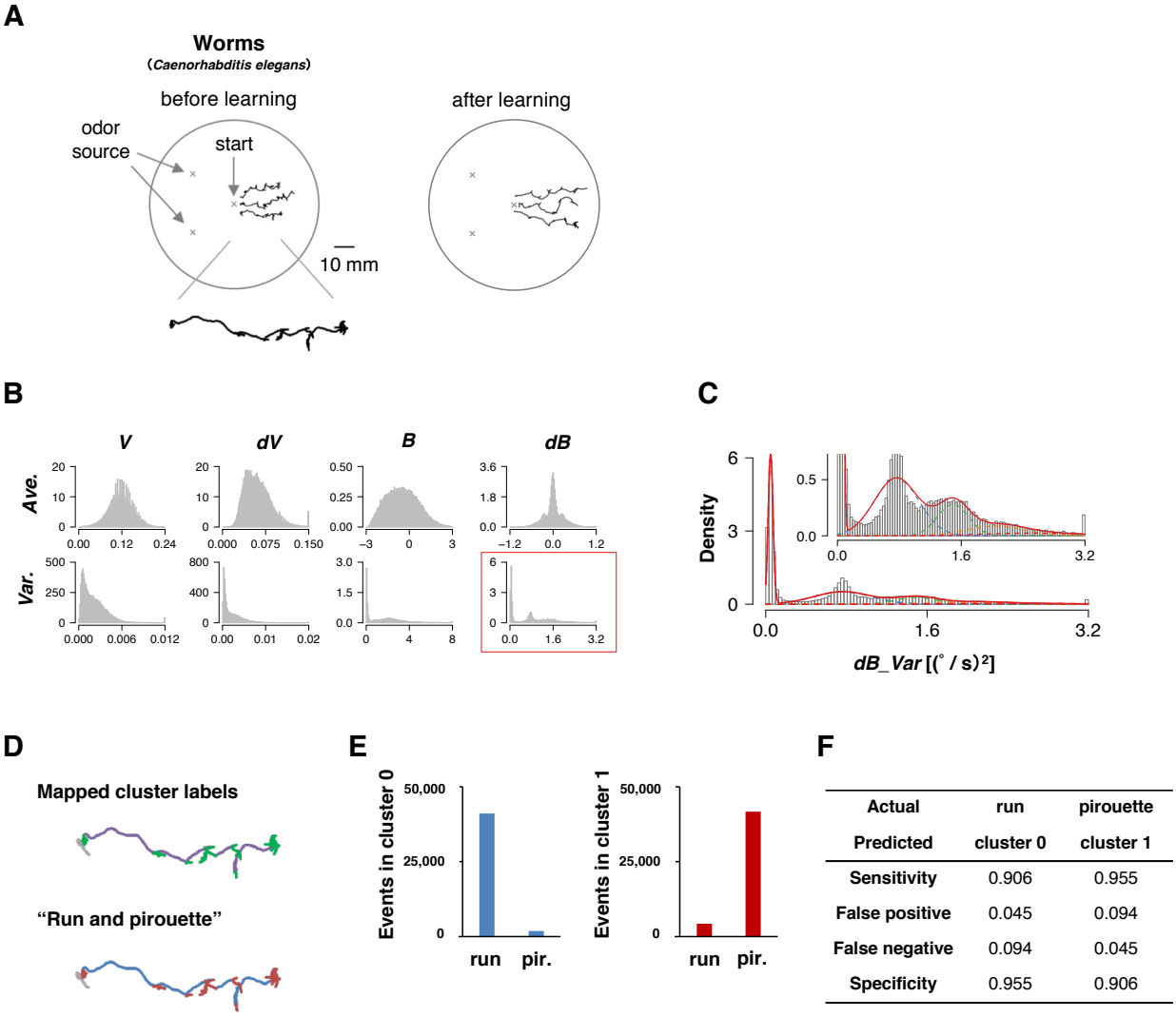

Trajectory data
(x, y)
Histograms for
basic behavioral
features
Clustering with
the EM algorithm
Mapping the cluster
labels
Calculating behavioral
features of a cluster
V
B
a reference
direction
Calculating information gain
V_All
V_Ini
V_Mid
V_Ter
Mean
0.3
0
0
0.17
Med.
0.2
0
0.04
0.2
1st Q
0.2
0
0.16
0.3
...
Plotting the result

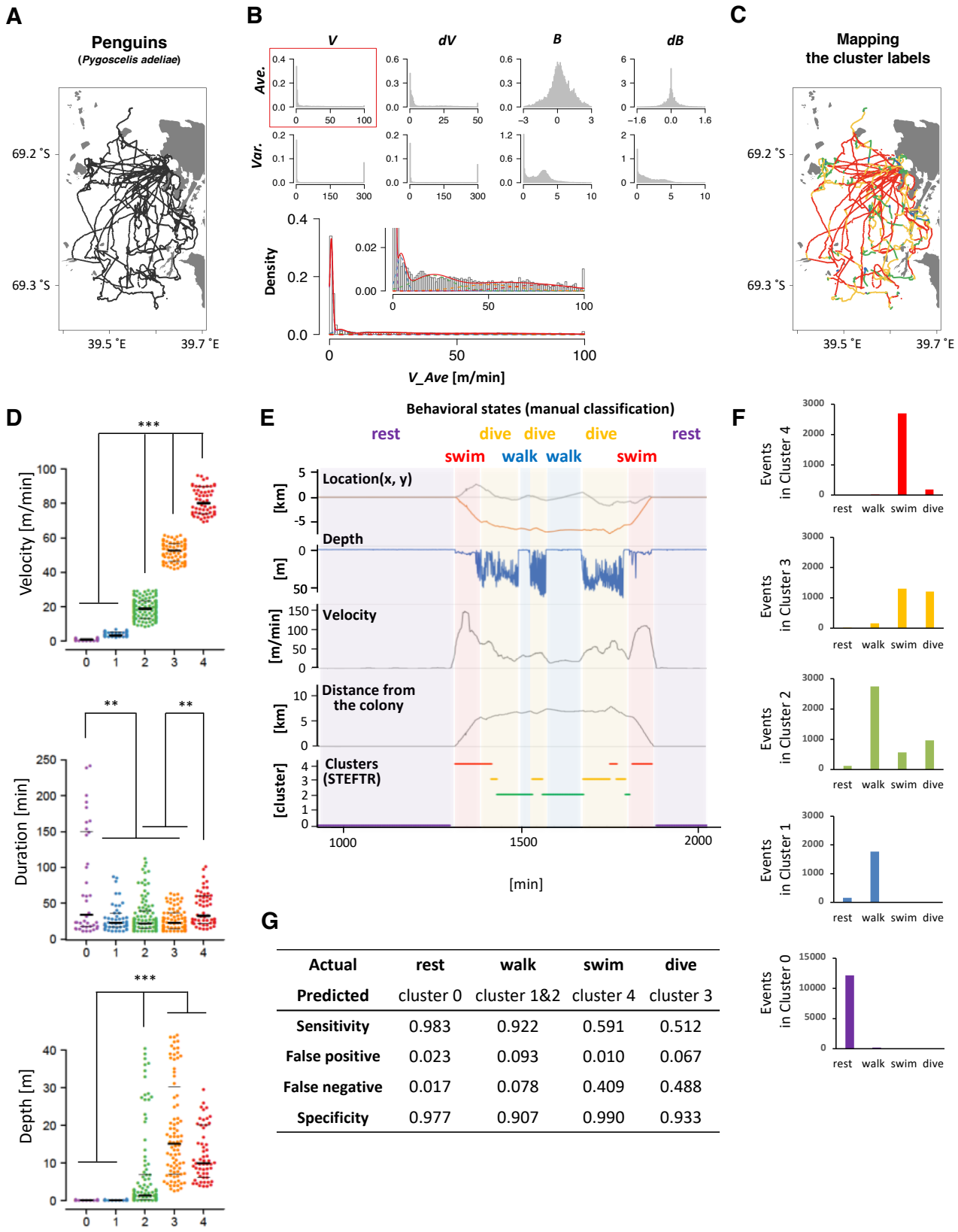
Figure 2



F

Actual	run	pirouette
Predicted	cluster 0	cluster 1
Sensitivity	0.906	0.955
False positive	0.045	0.094
False negative	0.094	0.045
Specificity	0.955	0.906

Figure 3



Actual	rest	walk	swim	dive
Predicted	cluster 0	cluster 1&2	cluster 4	cluster 3
Sensitivity	0.983	0.922	0.591	0.512
False positive	0.023	0.093	0.010	0.067
False negative	0.017	0.078	0.409	0.488
Specificity	0.977	0.907	0.990	0.933

Figure 4

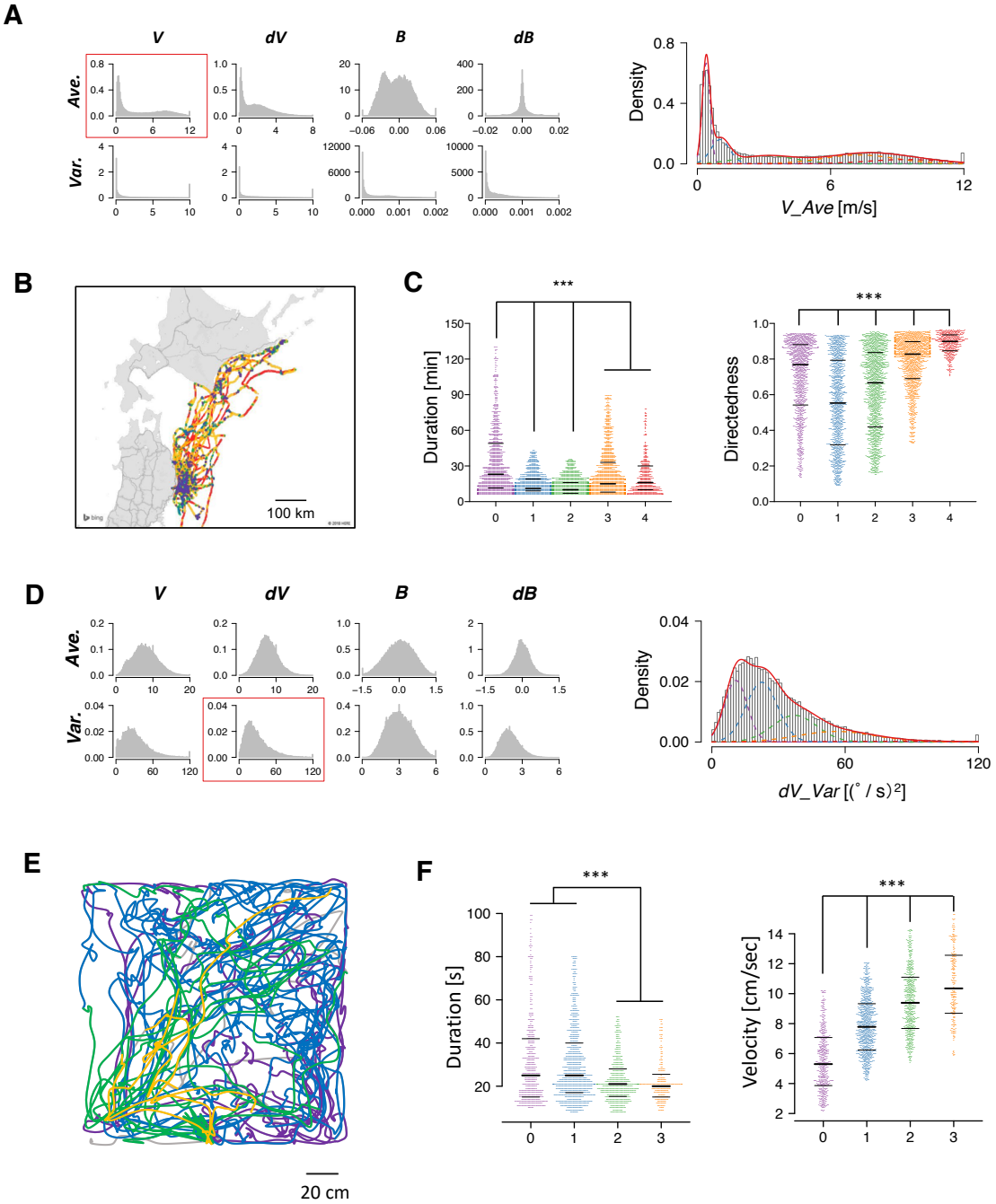
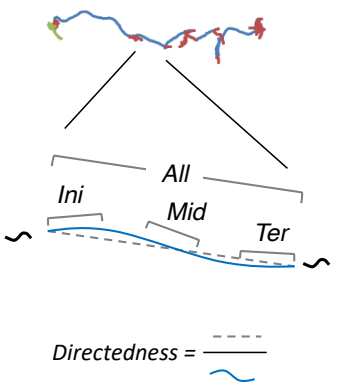
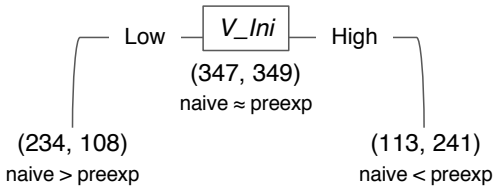

A
V
dV
B
dB
Ave.
Var.
Density
V_Ave [m/s]
B
100 km
C
***
Duration [min]
Directedness
D
dV_Var [(° / s)²]
E
20 cm
F
Duration [s]
Velocity [cm/sec]

Figure 5

A



B



Parameter	naive		preexp		entropy	Info gain
	Low	High	Low	High		
Original	347		349		1[a]	-
V_Ini	234	113	108	241	0.9017[b]	0.0983[c]

($c = a - b$)

Figure 6

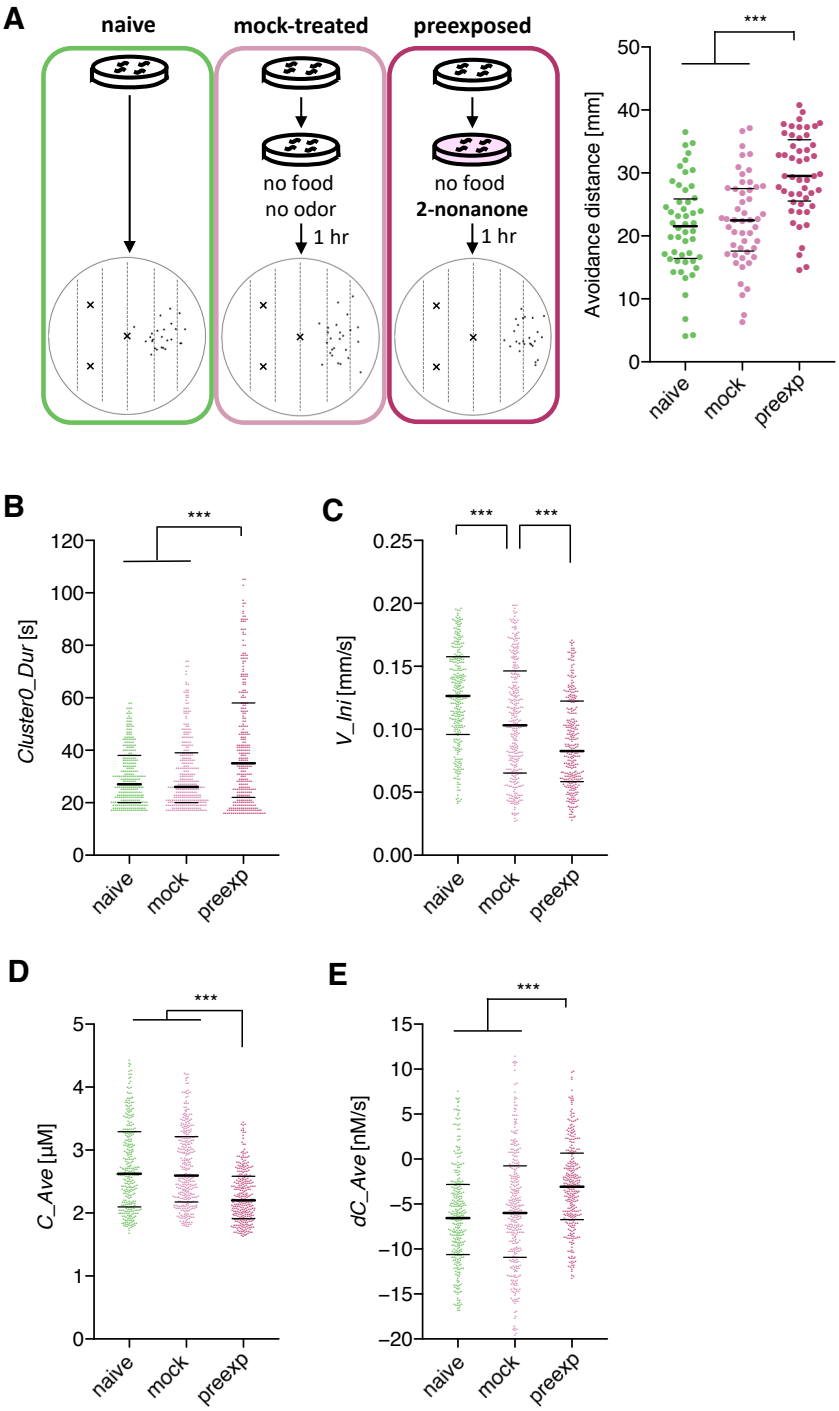

A
naive
mock-treated
preexposed
no food
no odor
1 hr
no food
2-nonanone
1 hr
Avoidance distance [mm]
***
naive
mock
preexp
B
Cluster0_Dur [s]
***
C
V_Ini [mm/s]
***
***
D
C_Ave [µM]
***
E
dC_Ave [nM/s]
***

Figure 7

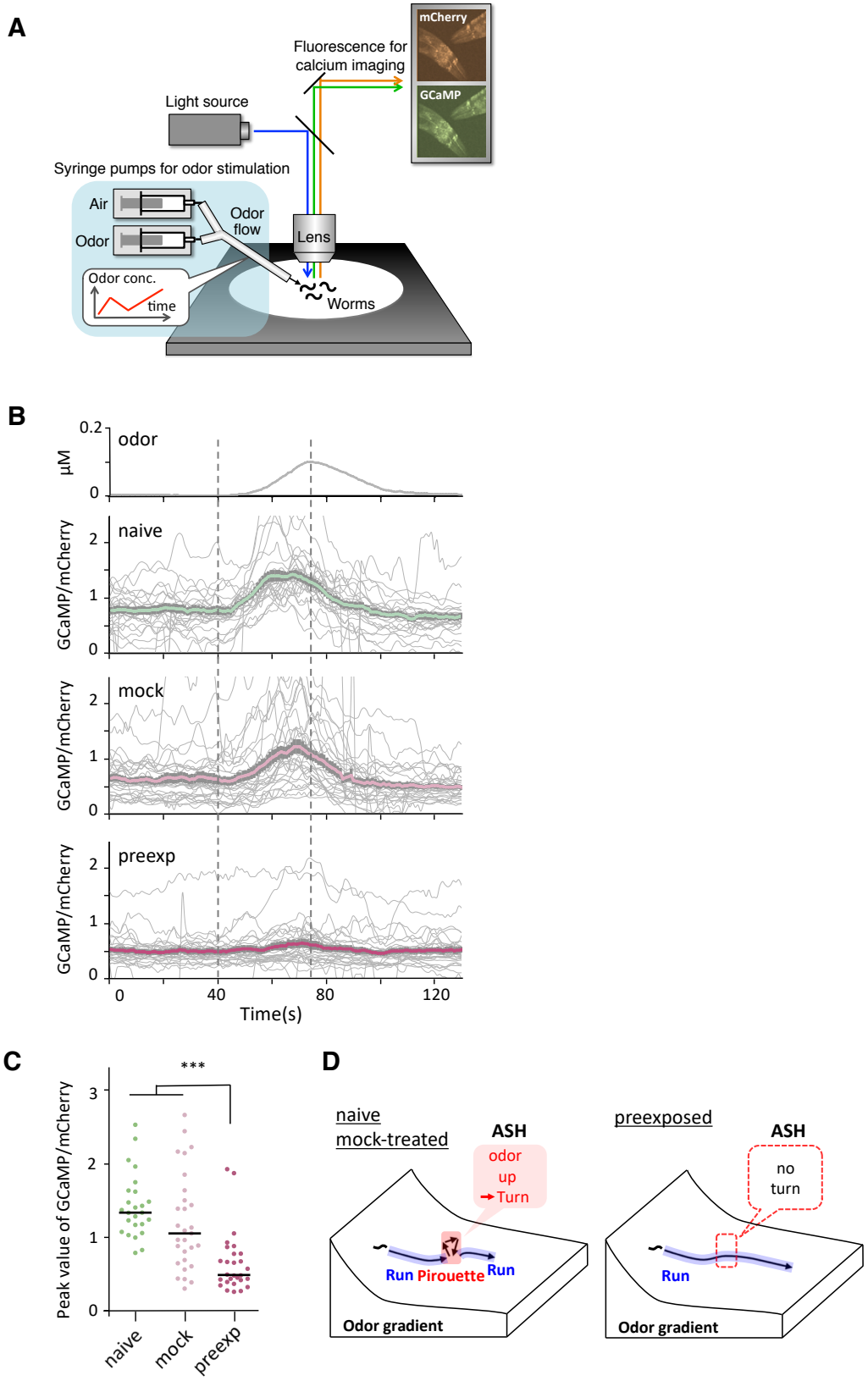
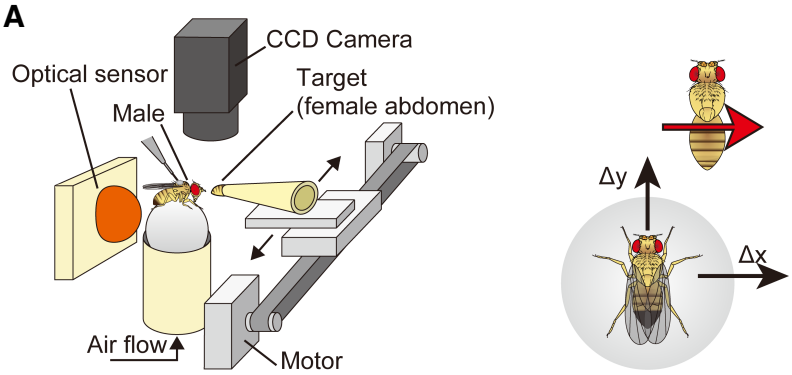

A
Fluorescence for calcium imaging
mCherry
GCaMP
Light source
Syringe pumps for odor stimulation
Air
Odor
Odor flow
Lens
Odor conc.
time
Worms
B
odor
μM
naive
mock
preexp
GCaMP/mCherry
Time(s)
C
***
Peak value of GCaMP/mCherry
naive
mock
preexp
D
naive
mock-treated
ASH
odor up
Turn
Run
Pirouette
Run
Odor gradient
preexposed
ASH
no turn
Run
Odor gradient

Figure 8

A



B

	V	*dV*	*B*	*dB*
Overall	0.279	0.260	0.146	0.176
Round 1	0.254	0.200	0.038	0.201
Round 2	0.268	0.218	0.177	0.277
Round 3	0.222	0.171	0.070	0.214
Round 4	0.175	0.235	0.184	0.204
Round 5	0.206	0.229	0.216	0.224
Round 6	0.168	0.221	0.118	0.240
Round 7	0.165	0.188	0.103	0.158
Round 8	0.027	0.087	0	0.062

C

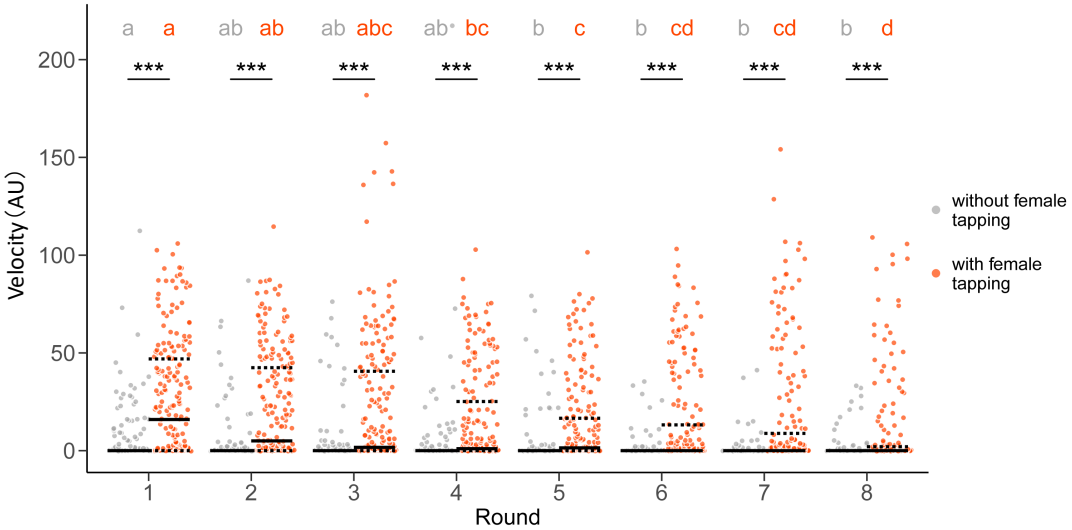
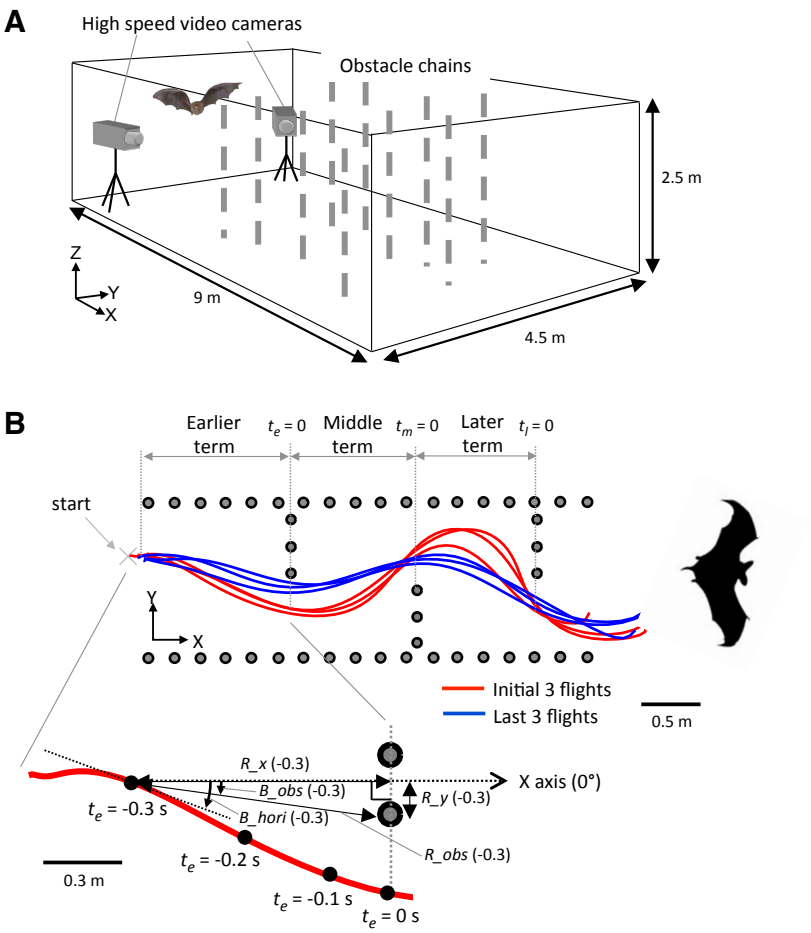
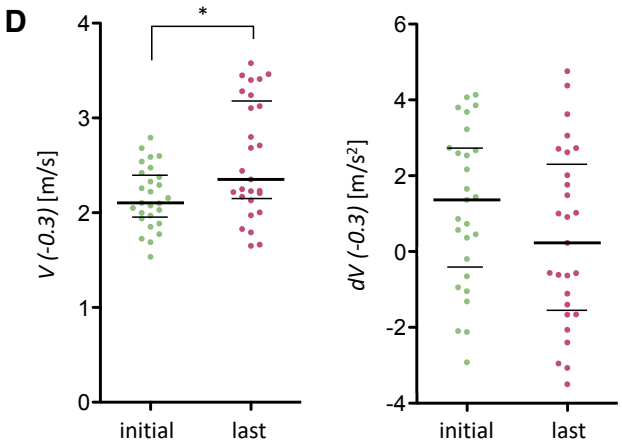


Figure 9



C

Features	Info Gain
R_x (-0.3)	0.297
B_hori (-0.3)	0.285
R_obs (-0.3)	0.269
V (-0.3)	0.216
B_hori (-0.2)	0.213
ddB_hori (-0.3)	0.207
V (-0.2)	0.191
R_x (-0.2)	0.191
R_obs (-0.2)	0.191