High-throughput ethomics in large groups of *Drosophila*

Kristin Branson[1], Alice A Robie[1], John Bender[2], Pietro Perona[1] & Michael H Dickinson[1]

We present a camera-based method for automatically quantifying the individual and social behaviors of fruit flies, *Drosophila melanogaster*, interacting in a planar arena. Our system includes machine-vision algorithms that accurately track many individuals without swapping identities and classification algorithms that detect behaviors. The data may be represented as an ethogram that plots the time course of behaviors exhibited by each fly or as a vector that concisely captures the statistical properties of all behaviors displayed in a given period. We found that behavioral differences between individuals were consistent over time and were sufficient to accurately predict gender and genotype. In addition, we found that the relative positions of flies during social interactions vary according to gender, genotype and social environment. We expect that our software, which permits high-throughput screening, will complement existing molecular methods available in *Drosophila*, facilitating new investigations into the genetic and cellular basis of behavior.

Drosophila has emerged as an important genetic model organism for the study of neurobiology and behavior. Research on fruit flies has led to insight into many behaviors of medical interest including drug abuse[1,2], aggression[3,4], sleep deprivation[5], aging[6] and memory loss[7]. The multitude of genetic manipulations possible in *Drosophila* make it an ideal model system to study general principles of behavioral neuroscience. For example, toolkits have recently been developed for altering the physiology of specific populations of neurons in intact flies[8–11]. However, analysis of the behavioral effects of these manipulations is hampered by the absence of thorough and quantitative methods for measuring behavior[12].

Machine vision has shown promise for automating tracking and behavior analysis of *Drosophila* and other animals. Several algorithms have been developed that can successfully track the trajectories of single, isolated flies[13–16]. Although useful, tracking only a single fly limits the types of behaviors that can be analyzed as well as the throughput of the system. Several tracking systems can be used to follow multiple, unmarked, interacting animals but fail when the animals are in close proximity to one another, and thus these systems cannot keep individual identities distinct[17–22]. The commercially available Ethovision system (Noldus) can track the identities of multiple interacting animals but requires tagging the animals with colored markers. The problem of tracking individuals in groups has been researched for studies of eusocial insects (ants and bees)[23,24], but robust implementations are not publicly available. Recently, systems have been developed to automatically detect components of aggression and courtship behavior in flies[4,25], in addition to tracking their positions. However, these systems cannot be used with large populations or unmarked flies, and detectors for new behaviors cannot be created without additional programming.

We propose a general-purpose, automated, quantitative and high-throughput system for measuring the behavior of interacting fruit flies. Our system uses machine vision techniques to automatically track large groups of unmarked flies while maintaining their distinct identities. We thus obtained trajectories: the position and orientation of each fly in each frame of a recorded video. Our system also includes automatic behavior detectors based on machine learning, which condense these trajectories into ethograms: meaningful, quantitative statistics of social and individual behavior. Because our system can be used to quickly measure many detailed statistics of fly behavior, it can be used to discover and quantify subtle behavioral differences between different populations of flies and between individuals within a population. We have designed our tracker to be adaptable to other laboratory setups, and our machine learning software can be used to specify new, automatic behavior detectors without programming. We therefore envision it will foster a more effective exploitation of genetic tools in behavioral neuroscience.

RESULTS

System overview

The behavioral arena used initially to test and develop our system consisted of a 24.5 cm diameter platform with an overhead FireWire camera and infrared lighting (**Fig. 1**). The software component consisted of a tracker for computing fly trajectories from captured digital video (**Fig. 2**) and a behavior detector, which may be trained from examples (**Fig. 3**). The system was accurate: the *x*-*y* position of a fly was estimated with a median error of 0.03 mm (2% of body length) and orientation with a median error of 4° (**Fig. 2e** and **Supplementary Figs. 1** and **2** online). Identity errors were

[1]California Institute of Technology, Pasadena, California, USA. [2]Case Western Reserve University, Cleveland, Ohio, USA. Correspondence should be addressed to P.P. (perona@caltech.edu) or M.D. (flyman@caltech.edu).

RECEIVED 1 DECEMBER 2008; ACCEPTED 8 APRIL 2009; PUBLISHED ONLINE 3 MAY 2009; DOI:10.1038/NMETH.1328

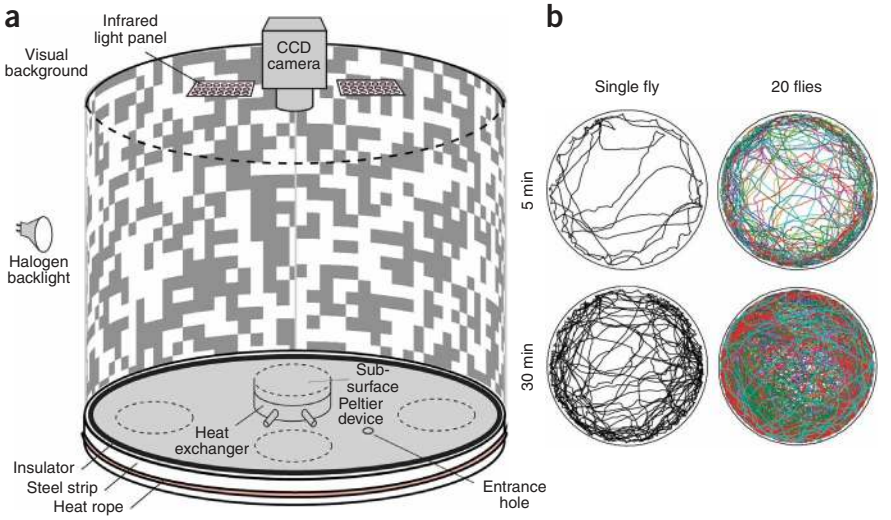

Figure 1 | Walking arena with sample trajectories. (**a**) Schematic diagram of the walking arena. A 24.5 cm tall printed paper cylinder is backlit by an array of 8 halogen lights (only one is shown). At the top is a 1,280 × 1,024 pixel charge-coupled device (CCD) camera with an 8 mm lens and infrared pass filter, and two arrays of 850 nm light–emitting diodes. The circular, 24.5 cm-diameter, 6 mm-thick aluminum base is thermally controlled by four Peltier devices and heat exchangers mounted on the underside (only one is shown) and is surrounded by a heat barrier composed of an insulating strip and a galvanized steel ring heated by thermal tape. Flies are loaded into the chamber through a hole in the floor with replaceable stopper. (**b**) The *x*-*y* position of a single fly or of 20 flies within the arena for 5 and 30 min of a trial.

absent with minimal user supervision and occurred every 1.5 h per fly in fully automatic mode (Online Methods and **Supplementary Table 1** online).

To test the potential of using multiple fly trajectories for automated behavior analysis, we carried out three proof-of-concept experiments. First, we defined automatic detectors for several individual and social behaviors exhibited by flies walking in a circular arena. Then, we used these detectors to produce ethograms for flies in different gender groupings. To determine whether these ethograms are useful descriptions of the flies' behavior, we used them to classify flies according to gender (male versus female) and genotype (wild type versus *fruitless*). The Fruitless protein is a transcription factor that plays a role in the sex determination pathway in flies. Male *fruitless* mutants exhibit several behavioral abnormalities, including inter-male courtship chains[26]. Second, we quantified differences in the behavior of individuals within a population and found that those differences were stable throughout each trial. Third, we examined the spatial distributions of the relative positions of flies during social interactions. We compared the distributions for pairs of flies of the same and different sex as well as for male *fruitless* mutants. All analyses described below were derived from seventeen 30-min trials, each comprising 20 flies, a total of 170 fly-hours. In four trials we used only females; in six, only males; in five, half male and half female; and in two trials, we used male flies homozygous for the fru^1 allele of *fruitless* (fru^1/fru^1). Examples of each of the four trial types are available in **Supplementary Videos 1–4** online.

Automatic ethograms

We created automatic detectors for eight behaviors with a wide range of sequence durations, velocities and accelerations (**Fig. 3a**, **Supplementary Video 5** and **Supplementary Table 2** online). These behaviors represented the majority of the flies' actions in our circular arena. We trained most detectors from a few manually segmented trajectories (Online Methods). The software is user-friendly, and detectors for new behaviors can be created without additional programming. Six of the behaviors involve basic locomotor actions, and two of the behaviors relate to social interactions between flies. Most of the time the flies either walked at a relatively constant velocity ('walk') or stopped in place ('stop'). The next-most common behavior was 'sharp turn', in which a fly made a large, rapid change in orientation. Other locomotor classifications included 'crabwalks', in which the fly walked with a substantial sideways component, and 'backups', in which the fly's translational velocity was negative. 'Jumps' consisted of rapid translations within the arena. A 'touch' occurred when the head of one fly came in contact with another fly. 'Chases' were cases in which one fly (always a male) followed another across the arena. An automatic detector for a given behavior (for example, the walk detector) input the trajectory for an individual fly (**Fig. 3b**) (or for a pair of flies, for social behaviors), derived per-frame statistics such as the translational speed, angular speed or distance to the second fly (for social behaviors), then segmented the trajectory into bouts in which the fly was and was not performing the given behavior (**Fig. 3c**).

By collecting the statistics of these eight behaviors into a vector, we created ethograms: rich, quantitative descriptions of each individual fly's behavior. For each fly, we computed one such description, consisting of the frequency with which each individual fly performed each behavior (we discuss other descriptions, the fraction of time a fly performed a behavior and mean behavior duration, in **Supplementary Fig. 3** online). To visualize differences among female, male and male fru^1/fru^1 flies, we grouped the flies by type and displayed the behavior frequency in pseudocolor (**Fig. 3d**). Inspection of this 'behavioral microarray' suggested that the behavioral vectors of female, male and fru^1/fru^1 male flies differed consistently. We quantified these differences by computing the mean and standard error behavior vectors for each type of fly (**Supplementary Fig. 4** online).

To determine whether these ethograms are powerful descriptors of behavior, we tested whether we could predict the sex of a fly (male versus female) and its genotype (wild-type males versus fru^1/fru^1 male) based solely on components of the automatically generated behavioral vector (**Fig. 3e**). We found that predictors based on the statistics of each of the eight behaviors independently distinguished sex with accuracies all better than chance, with touch frequency performing best (96.8% accuracy) and sharp turn frequency performing best of the locomotor behaviors (83.9% accuracy). A predictor based on the combination of all behaviors had an accuracy of 96.9%. Even a predictor based solely on locomotor behaviors (excluding touches and chases) predicted sex with an accuracy of 95.5%. We are not advocating using

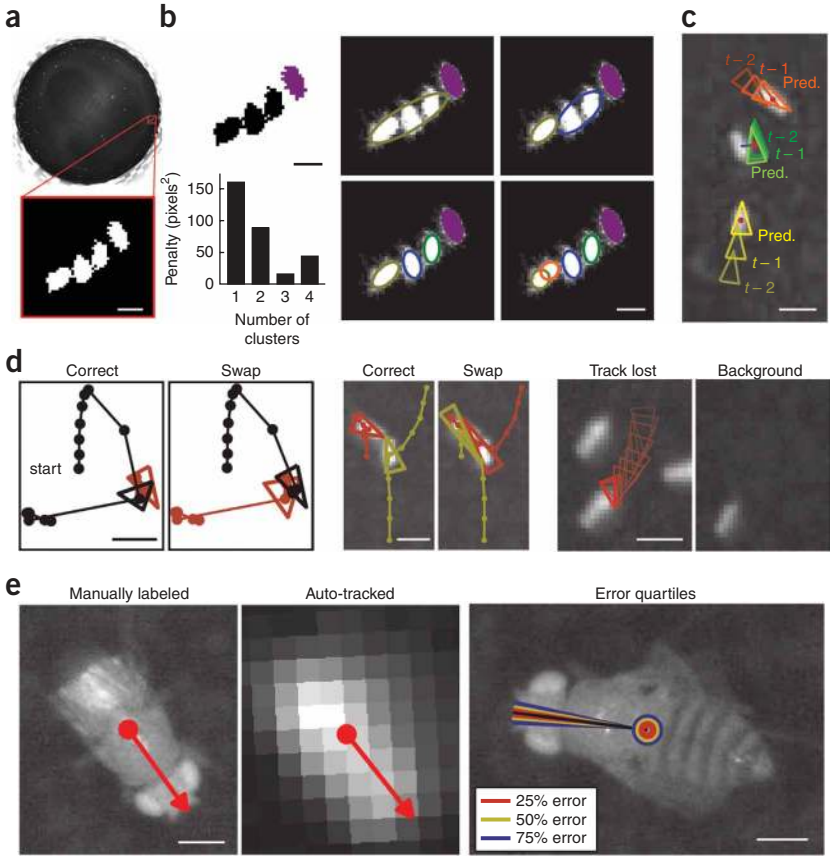

Figure 2 | Tracking algorithm and evaluation. (**a**) Example frame of entire arena with the foreground/background classification for pixels in the inset. (**b**) Detection of individual flies involves grouping foreground pixels. The purple component corresponds to one fly; the large black component corresponds to three. The tracker splits this large component into 1–4 clusters. The penalty based on cluster size is shown for each choice. (**c**) Identity matching involves pairing predicted and detected positions. Red dots indicate the detected fly positions in frame t; triangles indicate the tracked positions at frames $t-2$ and $t-1$ and the predicted position (pred.) at frame t. Blue lines indicate the lowest-cost match between predicted and detected positions. (**d**) Identity errors consist of swaps and lost identities. In the first example, the fly (black) jumps near a stationary fly (red), and identities are swapped. Plotted are the correct and automatically computed trajectories (left). Triangles indicate the positions of the flies at the frame of the swap; circles indicate their trajectories. In the second example, a large connected component is split incorrectly (middle); the trajectories are superimposed on the frame in which the swap occurred. In the third example, the lower left fly is still during the majority of the trial, becoming part of the background model (right); shown is the frame in which the fly's trajectory is lost as well as the background model at that instant. (**e**) A comparison of the center and orientation of a fly manually labeled on a high-resolution image (60 pixels mm^{-1}) to those automatically computed from a low-resolution image (4 pixels mm^{-1}). Quartiles of the sampled center position and orientation errors plotted on an example high-resolution image. The median error was 0.0292 mm (0.117 pixels) for the center and 3.14° for the orientation. Scale bars, 2.5 mm (**a**–**d**) and 0.5 mm (**e**).

behavioral statistics for sexing flies. Our mixed-sex trials (**Figs. 4** and **5**) used a fly's median image area for determining sex, a technique that achieves 96.2% accuracy. Instead, these behavior prediction accuracies are evidence that the ethograms were strongly correlated with gender.

Predictors of genotype (wild-type versus fru^1/fru^1 males) were even more robust (**Fig. 3f**). Frequency of backups achieved the best performance (99.3% accuracy). Using all behaviors or all locomotor behaviors, *fruitless* males could be classified with 100% accuracy. This technique of behavioral profiling could easily be extended to include more behaviors or more features of each behavior (**Supplementary Note** online).

Behavioral variation between and within individuals

We observed that the trajectories of individual flies looked qualitatively different (**Fig. 4a**). For example, some flies traveled more than others and some spent a larger fraction of time near the arena wall. Because our algorithm kept track of each fly's trajectory, we could easily gather data on a large number of flies and explore statistical differences in behavior across individuals. To this end, we computed behavioral statistics separately for the first 15 min and the second 15 min of each 30-min trial and calculated the correlation between the two halves. We considered three statistics of locomotor behavior: the mean speed during walking episodes, the fraction of frames the fly was classified as walking and the mean duration of walking episodes (**Fig. 4b**). The correlation between the first- and second-half statistics was significant ($P < 2.2 \times 10^{-16}$) and positive for all three walking metrics, indicating that individuals maintained behavioral tendencies throughout the 30-min trials. Thus, although within the tested strain of wild-type flies we found substantial differences in walking behavior, each individual fly walked consistently over time.

We also investigated whether there were consistent differences in chasing behavior across individual flies during a 30-min trial. For the first and second half of each trial, we computed the frequency with which a fly begins chasing another fly, the frequency with which other flies begin chasing a given fly and the mean time duration of chase sequences initiated by a given fly (**Fig. 4c**). As with the walking experiments, we computed the correlation between behavioral statistics gathered during the first and second half of each trial. We found small, but significant, positive correlations for frequency of chasing ($P = 3.89 \times 10^{-16}$) and frequency of being chased ($P = 1.54 \times 10^{-3}$) but no significant correlation for duration of chase sequences ($P = 0.261$).

Gender differences and fly-fly interactions

Because our data consisted of the location and orientation of all individuals at all times, we could examine the spatial distributions of the relative positions of flies during social interactions. We compared the distributions of inter-fly distances for different gender pairings in single-sex and mixed-sex trials (for example, male-to-male distance in mixed-sex trial) (**Fig. 5a**). As a control, we

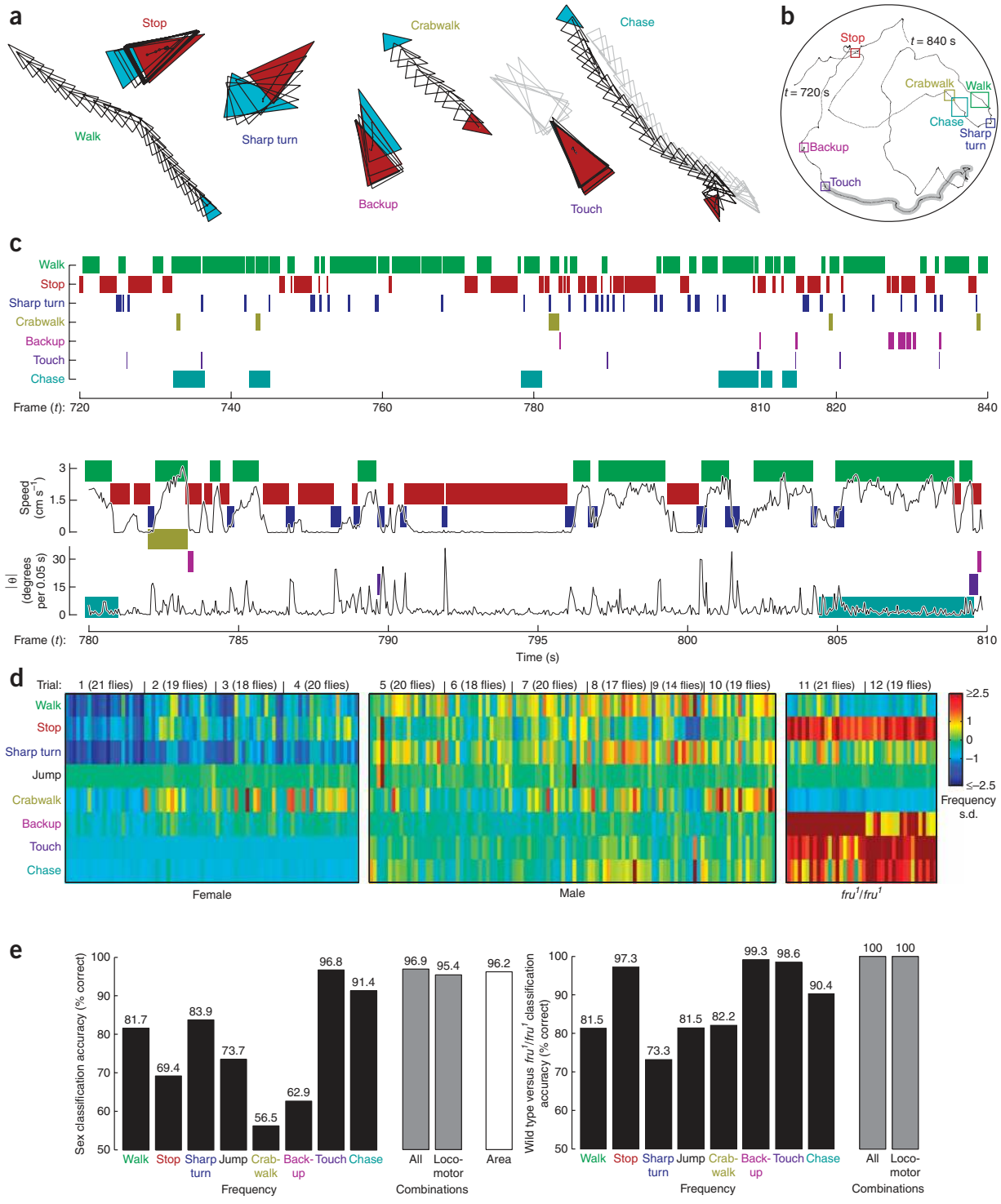

Figure 3 | Ethograms of eight automatically-detected behaviors. (**a**) Examples of behaviors detected (from the trajectory shown in **b**). Triangles indicate the fly's positions in every frame. Cyan and red triangles are plotted at the start and end of each behavior example, respectively; only the start of the walk example is shown. For touching and chasing, we plotted in gray the position of the other fly. (**b**) Sample 2-min trajectory for a male fly in a mixed-sex arena. The colored boxes indicate trajectory segments in **a**. (**c**) Behavior classifications for the 2 min trajectory (top). A mark at $t = 780$ for the 'chase' row indicates that the fly was chasing at that time. Plots of translational and angular speed for a 30-s span of the trajectory ($t = 780$–810 s), superimposed over the behavior classifications (bottom). (**d**) Example behavioral vectors for female, male and male fru^1/fru^1 flies in single-sex trials. Each column corresponds to a fly and each row to a behavior ($n = 78$ (female), 108 (male) and 40 (fru^1/fru^1). Color indicates the s.d. from the mean frequencies (onsets per minute) for each behavior. (**e**) Accuracy of sex prediction from automatically detected behaviors. Black bars, cross-validation error of single-threshold classifiers based on frequency. Gray bars, logistic regression classifiers from all eight and the six locomotor behaviors. White bar, accuracy of classifying sex based on the image area of the fly (Online Methods). (**f**) Accuracy of genotype prediction (wild type versus fru^1/fru^1), as in **e**.

created a semisynthetic dataset by artificially staggering in time all 20 trajectories relative to one another. We left the first fly's trajectory unchanged but shifted the second fly's trajectory in time so that it started at $t = 1.5$ min, with the last 1.5 min of its original trajectory wrapped around to fill the time from $t = 0$ to $t = 1.5$ min; we shifted the third fly's trajectory by 3 min, the fourth by 4.5 min and so forth. These data approximated trajectories in which the flies do not interact.

The peaks in the male-to-male and male-to-female distributions compared to the synthetic data indicated that males actively approach other flies to a distance of 2.5–3.5 mm. In addition, the relatively low frequency of close interactions (<4 mm) between females suggested that they maintained a larger buffer between themselves. These findings were robust across trial type (for example, males approached other males as closely in mixed-sex arenas as in single-sex arenas). We also observed that the flies' centroids never moved within 1.5 mm of each other, which is expected given this distance roughly corresponds to a fly's body width.

To explore spatial differences during social interactions, we created a new behavioral classification termed 'encounter'

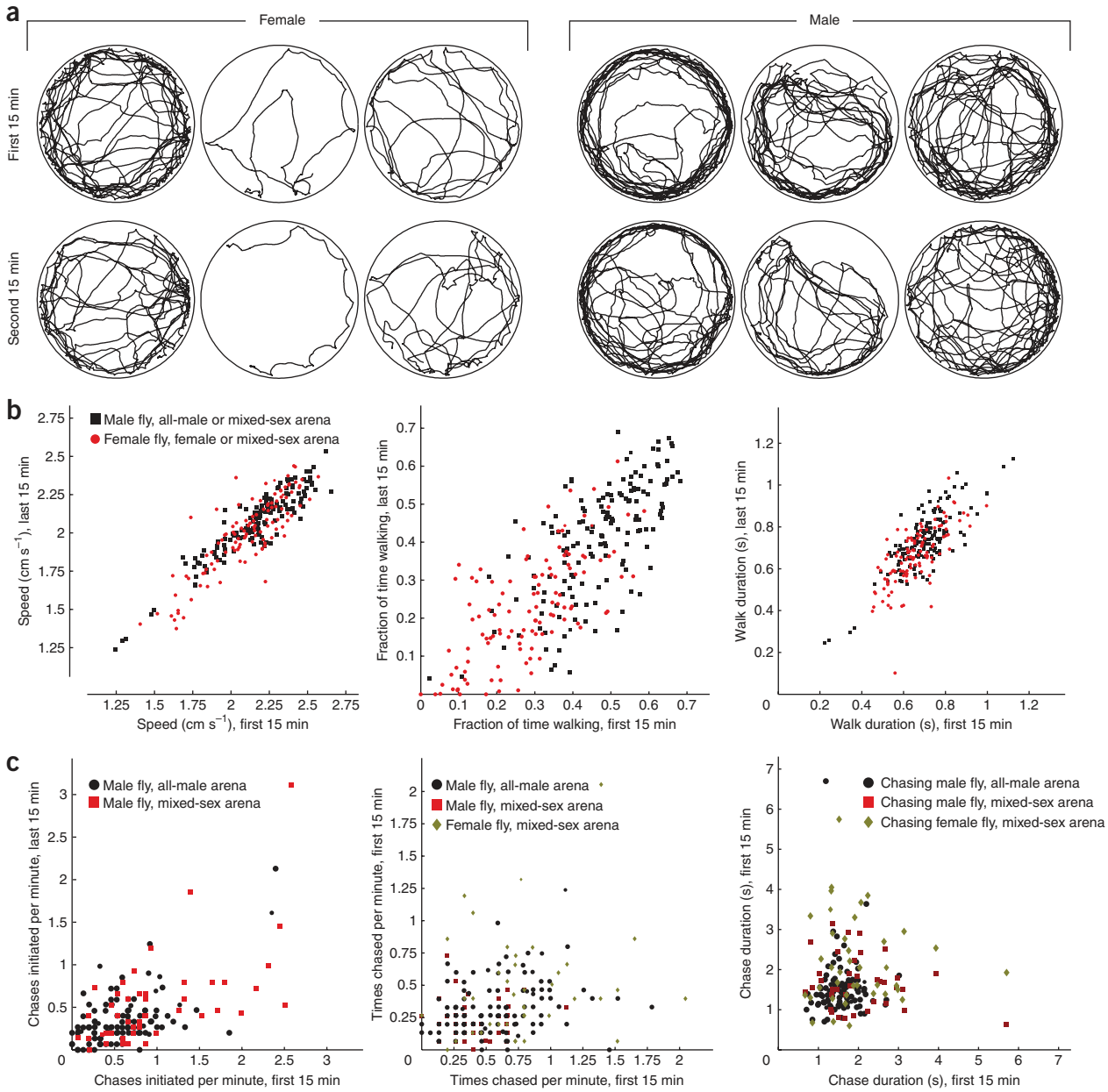

Figure 4 | Differences within and among individual flies. (**a**) The first and second halves of trajectories for three male and three female flies from the same trial. (**b**) Scatter plots of walking statistics from each individual fly in the first 15 min of its trajectory against the same statistics from the last 15 min of its trajectory for flies in all trial types (female, $n = 132$ and male, $n = 159$). Walking statistics examined were: mean speed in frames in which fly was classified as walking: $r = 0.889$, $P < 2.2 \times 10^{-16}$ (P, the probability that the null hypothesis of r non-positive is correct; left); fraction of frames fly is classified as walking: $r = 0.689$, $P < 2.2 \times 10^{-16}$ (middle); and mean duration of sequences of consecutive walking frames: $r = 0.765$, $P < 2.2 \times 10^{-16}$ (right). (**c**) Chasing behavior differences. We repeated the above procedure for chasing behavioral statistics: Frequency with which the fly begins chasing another fly: $r = 0.592$, $P = 3.89 \times 10^{-16}$ (left), frequency with which a fly is chased by another fly: $r = 0.213$, $P = 1.54 \times 10^{-03}$ (middle), and mean duration of chases: $r = 0.054$, $P = 0.261$ (right).

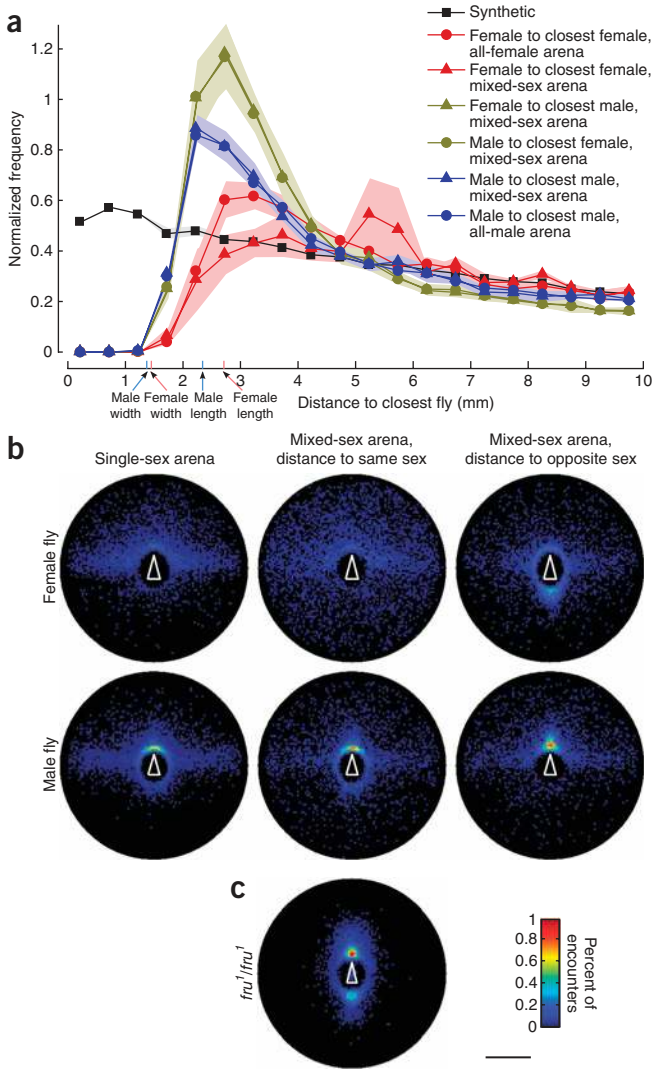

Figure 5 | Spatial analysis of social interactions. (**a**) Normalized histogram of inter-fly distances to the nearest fly for each fly in each frame. The frequency was normalized both by the total number of counts and by the area of the bin. Each encounter was counted only once by ignoring all but the first frame in which both flies were stopped. The 'synthetic' condition shows a control in which we decorrelated fly positions by staggering the trajectories in time and collapsed data from all conditions. The lightly shaded regions indicate 1 s.d. in normalized frequency, approximated by randomly splitting the flies into five groups. For comparison, the pink and blue tick marks indicate the mean fly widths and heights for female and male flies, respectively. (**b**) Histogram of the *x*-*y* relative position of one fly in the coordinate system of another at the closest point of an encounter. Each plot corresponds to a different social condition, as indicated. The white triangle in each plot shows the fixed position of the given fly. The pixel color indicates the frequency with which the closest fly is in the corresponding location bin. (**c**) Histogram of the *x*-*y* mutual position between fru^1/fru^1 males. Scale bar, 0.5 cm.

describing those trajectory intervals in which the distance between a pair of flies was less than 10 mm. For each encounter, we computed the relative location of one fly in the coordinate system of the other at the time when the distance between them was minimal. We computed histograms of these relative locations over all encounters of each gender pairing and trial type (**Fig. 5b**). These histograms were consistent with our qualitative knowledge of courtship behavior. For interactions involving males, the majority of the encounters occurred very near the other fly, when the flies were almost in direct contact. In contrast, the relative locations of the female-female encounters were more diffuse. It is apparent from the hot spot near the head of the flies in **Figure 5b** that males often took a position so that another fly was right in front of them, an orientation that is consistent with their chasing behavior. Conversely, a hot spot is visible directly behind females in mixed-sex trials, indicating that they are being chased by males. Notably, two hot spots are apparent in the encounter histograms of fru^1/fru^1 males (**Fig. 5c**), indicating a social phenotype that is intermediate between that of males and females. The data in this figure represent a quantitative and reproducible measure of the chaining phenotype that is characteristic of many male *fruitless* mutants[26].

DISCUSSION

We developed software that allowed us to automatically track and analyze up to 50 individual flies (a density of 0.1 fly cm^{-2} in our arena) simultaneously for long periods of time. We estimate that the behavioral analyses shown in **Figure 3** would have taken a human operator between 3,000 and 5,000 h to produce manually. The observations on individual behavior would have taken much longer. The software (**Supplementary Software 1** and **2** online, updates available at http://www.dickinson.caltech.edu/ctrax) is open-source and was developed to function in a wide array of experimental contexts. Furthermore, it is easy for a biologist to train the system to detect new behaviors by providing a few examples using a graphical user interface designed for this purpose.

The open arena used for most of our analysis required clipping the flies' wings, a manipulation that may affect aspects of their behavior, for example, the production of courtship song. In addition, although the open arena apparatus allowed us to perform the rigorous groundtruthing presented, it is custom-built and would not be instantly available to the research community. However, we analyzed data that were collected in a much simpler and easy-to-replicate chamber, consisting of a backlit plastic chamber with a glass top (Jasper Simon and M.H.D.; unpublished data). This analysis (**Supplementary Videos 6** and **7** online) demonstrated that our software works on data collected from intact flies in an inexpensive and easily reproduced device.

Our method benefits from insight gained from previous approaches to the study of behavior in *Drosophila*. The first, inspired by a classic 'countercurrent' apparatus[27], involves crafting a simple mechanical contraption that isolates behavioral outliers in a large population. This method is easy to perform and thus amenable to high-throughput screens but does not provide detailed measurements on individual flies. In addition, complex behaviors (for example, courtship and aggression) are not easily screened by these techniques. The second, exemplified by tethered flight arenas[28] and 'Buridan's paradigm'[29], involves developing a sophisticated apparatus that provides detailed, time-resolved measurements of individual flies. This approach offers a rich view of behavior but does not allow for high-throughput screens. In addition, behavioral analyses that depend on elaborate, custom-made instruments do not easily proliferate throughout the scientific community. The third approach, exemplified by the use of 'courtship wheels'[30], provides detailed information on the complex

behaviors of individual flies but relies on manual scoring by human observers and is labor-intensive and subjective.

Our system combines the key features of prior behavior analysis methods and is thus a complementary tool to genetic manipulation for the study of the neural bases of behavior. Because each fly is tracked and measured individually, it is possible to quantify the behavior of individual flies as well as fly-fly interactions. The system's flexibility allows many different individual and social behaviors to be defined and automatically detected. The definitions for these behaviors are interpretable and quantitative, allowing researchers to easily reproduce experiments. Finally, the system supports high-throughput screening, facilitating its use with genetic manipulations.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemethods/.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank A. Straw for developing and maintaining the camera interface program, J. Simon for assistance in collecting the data presented in **Supplementary Videos 6** and **7**, W. Korff for help with high-resolution data acquisition and M. Arbietman (University of Southern California) for the gift of the *fruitless* fly lines. Funding for this research was provided by US National Institutes of Health grant R01 DA022777 (to M.H.D. and P.P.).

Published online at http://www.nature.com/naturemethods/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Wolf, F.W. & Heberlein, U. Invertebrate models of drug abuse. *J. Neurobiol.* **54**, 161–178 (2003).
2. Guarnieri, D.J. & Heberlein, U. *Drosophila melanogaster*, a genetic model system for alcohol research. *Int. Rev. Neurobiol.* **54**, 199–228 (2003).
3. Chan, Y.B. & Kravitz, E.A. Specific subgroups of Fru^M neurons control sexually dimorphic patterns of aggression in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **104**, 19577–19582 (2007).
4. Hoyer, S.C. *et al.* Octopamine in male aggression of *Drosophila*. *Curr. Biol.* **18**, 159–167 (2008).
5. Ho, K.S. & Sehgal, A. *Drosophila melanogaster*: an insect model for fundamental studies of sleep. *Methods Enzymol.* **393**, 772–793 (2005).
6. Shaw, P., Ocorr, K., Bodmer, R. & Oldham, S. *Drosophila* aging 2006/2007. *Exp. Gerontol.* **43**, 5–10 (2008).
7. Konsolaki, M., Song, H.J., Dobbs, W. & Garza, D. P2–109 *Drosophila* models of Alzheimer's-related pathways. *Neurobiol. Aging* **25**, S255–S255 (2004).
8. Zhang, F. *et al.* Circuit-breakers: optical technologies for probing neural signals and systems. *Nat. Rev. Neurosci.* **8**, 577–581 (2007).
9. Callaway, E.M. A molecular and genetic arsenal for systems neuroscience. *Trends Neurosci.* **28**, 196–201 (2005).
10. Luo, L., Callaway, E.M. & Svoboda, K. Genetic dissection of neural circuits. *Neuron* **57**, 634–660 (2008).
11. Zhou, C., Rao, Y. & Rao, Y. A subset of octopaminergic neurons are important for *Drosophila* aggression. *Nat. Neurosci.* **11**, 1059–1067 (2008).
12. Anonymous. Geneticist seeks engineer: must like flies and worms. *Nat. Methods* **4**, 463 (2007).
13. Martin, J.R. A portrait of locomotor behaviour in *Drosophila* determined by a video-tracking paradigm. *Behav. Processes* **67**, 207–219 (2004).
14. Ramazani, R.B., Krishnan, H.R., Bergeson, S.E. & Atkinson, N.S. Computer automated movement detection for the analysis of behavior. *J. Neurosci. Methods* **162**, 171–179 (2007).
15. Grover, D., Tower, J. & Tavaré, S. O fly, where art thou? *J. Royal Society Interface* **5**, 1181–1191 (2008).
16. Valente, D., Golani, I. & Mitra, P.P. Analysis of the trajectory of *Drosophila melanogaster* in a circular open field arena. *PLoS ONE* **2**, e1083 (2007).
17. Crocker, J.C. & Grier, D.G. Methods of digital video microscopy for colloidal studies. *J. Colloid Interface Sci.* **179**, 298–310 (1996).
18. Ramot, D., Johnson, B.E., Berry, T.L. Jr., Carnell, L. & Goodman, M.B. The parallel worm tracker: a platform for measuring average speed and drug-induced paralysis in nematodes. *PLoS ONE* **3**, e2208 (2008).
19. Ryu, W.S. & Samuel, A.D.T. Thermotaxis in *Caenorhabditis elegans* analyzed by measuring responses to defined thermal stimuli. *J. Neurosci.* **22**, 5727–5733 (2002).
20. Tsunozaki, M., Chalasani, S.H. & Bargmann, C.I. A behavioral switch: cGMP and PKC signaling in olfactory neurons reverses odor preference in *C. elegans*. *Neuron* **59**, 959–971 (2008).
21. Wolf, F.W., Rodan, A.R., Tsai, L.T.Y. & Heberlein, U. High-resolution analysis of ethanol-induced locomotor stimulation in *Drosophila*. *J. Neurosci.* **22**, 11035–11044 (2002).
22. Soll, D.R. & Voss, E. Two-and three-dimensional computer systems for analyzing how animal cells crawl. In *Motion analysis of living cells* (David R. Soll & Deborah Wessels, eds.) 25–52 (Wiley-Liss, New York, 1997).
23. Khan, Z., Balch, T. & Dellaert, F. MCMC-based particle filtering for tracking a variable number of interacting targets. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**, 1805–1819 (2005).
24. Veeraraghavan, A., Chellappa, R. & Srinivasan, M. Shape-and-behavior-encoded tracking of bee dances. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**, 463–476 (2008).
25. Dankert, H., Wang, L., Hoopfer, E.D., Anderson, D.J. & Perona, P. Automated monitoring and analysis of social behavior in *Drosophila*. *Nat. Methods* **6**, 297–303 (2009).
26. Hall, J.C. Courtship among males due to a male-sterile mutation in *Drosophila melanogaster*. *Behav. Genet.* **8**, 125–141 (1978).
27. Benzer, S. Behavioral mutants isolated by countercurrent distribution. *Proc. Natl. Acad. Sci. USA* **58**, 1112–1119 (1967).
28. Götz, K. Flight control in *Drosophila* by visual perception of motion. *Biol. Cybern.* **4**, 199–208 (1968).
29. Bülthoff, H., Götz, K.G. & Herre, M. Recurrent inversion of visual orientation in the walking fly, *Drosophila melanogaster*. *J. Comp. Physiol. A* **148**, 471–481 (1982).
30. Siegel, R.W. & Hall, J.C. Conditioned responses in courtship behavior of normal and mutant *Drosophila*. *Proc. Natl. Acad. Sci. USA* **76**, 3430–3434 (1979).

ONLINE METHODS

Flies. Wild-type flies used in these experiments were derived from a laboratory population originating from a collection of 200 wild-caught females. fru^1/fru^1 flies were derived from a fru^1/TM3 stock. In the open-arena experiments, flies were cold-anesthetized 24 h before experiments to clip their wings to one-half their original length so that they could not fly out of the arena. They recovered overnight on food and were wet starved 6 h before experiments. For more information, see the **Supplementary Note**.

Apparatus. The walking arena used in most of our experiments consisted of a temperature-controlled 24.5 cm diameter platform surrounded by a static backlit visual pattern (**Fig. 1**). Flies were maintained in the arena by a thermal barrier around the outside edge of the walking platform and by clipping the wings as described above. The thermal barrier consisted of a rope heater wrapped around a galvanized steel band insulated from the platform by a layer of neoprene. Although some flies would occasionally hop over the arena's edge, most would avoid walking off the platform due to the heat barrier. Above the arena were mounted infrared light–emitting diodes and a 1,280 × 1,024 pixel camera sensitive to near-infrared light. Images were recorded at 20 frames per second by a computer using the Motmot Python camera interface package (Andrew Straw and M.D.; unpublished data). For details, see the **Supplementary Note**. Although our software was developed in conjunction with this set up, it is adaptable to other arrangements with similar characteristics (**Supplementary Videos 6** and **7**).

Tracking algorithm. Our purpose in developing both the algorithm and the apparatus was to create a reliable system for obtaining interesting behavioral statistics for use by behavioral geneticists. Our tracking algorithm combined techniques from the computer vision literature to achieve this goal. The tracking algorithm input a stored video sequence and computed the trajectory of each fly (center position and orientation in each frame). Tracking was achieved by alternating two steps: fly detection and identity assignment. At each new frame, flies were first detected and their positions and orientations were computed. Next, each detected fly in frame t was associated with a fly tracked in the previous frame $t - 1$. Example tracked trajectories are shown in **Figure 1b**. Our tracking algorithm is described below; more details are available in the **Supplementary Note**.

Detection. Detection was based on background subtraction[31]. In our laboratory setting, we ensured that the camera was still and the infrared lighting was constant, thus the only objects moving in the video were flies. The appearance and variability of the arena without flies (the background) was estimated before tracking as the pixelwise median of a set of frames sampled from the entire video sequence. The variability was estimated as the pixelwise median absolute deviation from the background image. Using the median made our algorithm tolerant to flies that did not move for long periods of time. Note that it is good practice to estimate the background model from video taken after the flies have been added because the arena may be inadvertently jostled in the process of introducing flies. Movement of the arena or camera of just one pixel can cause large errors in background subtraction.

In our setup, the flies appeared bright and the background dark (the tracker will also work with dark flies on a light background, as shown in **Supplementary Videos 6** and **7**). Foreground pixels, that is, pixels belonging to flies, were detected when the difference between the pixel and background intensity exceeded a multiple of the background variability (**Fig. 2a**). This step relied on the flies (and only the flies) looking different from the background; poor camera quality and excessive video compression can compromise this step. Next, foreground pixels were grouped together into single fly detections. Ideally, each connected component[32] of foreground pixels would correspond to exactly one fly. We thus initially fit an ellipse to each connected component by fitting a Gaussian to the locations of the corresponding foreground pixels. Owing to flies sometimes coming into contact and inevitable errors in pixel labeling, some connected components may have corresponded to many, to part of one or to no flies. These errors were corrected automatically by detecting connected components that were too large or small and considering multiple splitting or merging hypotheses (**Fig. 2b**).

Identity assignment. Each fly detected in frame t was associated with a trajectory from frame $t - 1$. In the first frame, a unique trajectory label was assigned arbitrarily to each detection. In subsequent frames, assuming that each trajectory had been computed up to frame $t - 1$, it was extended to frame t by assigning each fly detection in t to the trajectory that best predicted its position and orientation (**Fig. 2c**), where predictions were computed by a constant-velocity model. This was a multiple-assignment problem because trajectories and flies had to be in one-to-one correspondence: two flies could not be associated to the same trajectory and vice versa. Thus, the optimal solution needed to be computed simultaneously for all flies. Occasionally, a fly may have escaped or entered the arena or the detection stage may have made an error. For this reason, our software algorithm allowed a trajectory or a detection to be unmatched when the distance was too large and paid a constant penalty. The best overall assignment was computed using the Hungarian method for minimum-weight perfect bipartite matching[33,34]. The assignment step requires that the frame rate be sufficiently high relative to the speed of the flies so that the optimal matching between observations and trajectories is easy for a human observer.

Hindsight. The detection step was performed using information from only the current frame, and the matching step assumed that these detections were correct. Errors in the detection step would often result in births or deaths of tracks. After identity assignment, the tracker determined whether each birth and death could be prevented by temporarily splitting, connecting, merging or deleting tracks. This step worked on the assumption that flies rarely entered or left the arena.

Orientation ambiguity. The detection phase could not tell the head from the tail of a fly. To resolve this ambiguity, at each frame, our tracker determined whether to add 180° to the orientation of each fly. Using a variation of the Viterbi algorithm[35], the sequence of orientation offsets was computed that minimized the change in orientation between consecutive frames and the difference between orientation and velocity direction when the fly was moving.

doi:10.1038/nmeth.1328

System evaluation. We measured the quality of our tracker by comparing its measurements with groundtruth on a set of benchmark videos. We distinguished identity, position and sex assignment errors. Identity errors included swapping flies' identities, losing flies' tracks and spurious detections that did not correspond to flies (**Fig. 2d**). Position errors were inaccuracies in the estimated position and orientation of a fly (**Fig. 2e**). Sex assignment errors were mistakes in determining whether a fly was male or female.

Identity errors. We evaluated the frequency of identity errors made by our system on 18 manually annotated video sequences, each containing 10, 20 or 50 wild-type flies, which were either all female, all male or half male and half female. Two 5-min videos were used as benchmarks for each condition. We show example identity errors in **Figure 2d**. To find identity errors, a trained operator examined those video frames in which tracking is hardest: when flies were near each other, there were large differences between predicted and measured positions or at the births and deaths of trajectories. These frames were inspected in slow motion, zoomed in on the difficult-to-follow flies. The operator marked an annotated frame as incorrect if there was an identity error and also classified the type of error. The scoring took approximately 0.5 h for each 10-fly video, 2 h for each 20-fly video and 8 h for each 50-fly video. We observed an identity error on average once every 5 fly-hours in the 10-fly videos, once every 1.5 fly-hours in the 20-fly videos and once every 40 fly-minutes in the 50-fly videos. **Supplementary Table 1** shows the counts per error type per video.

Fixing identity errors manually. Using simple heuristics, a small number of suspicious frames and flies were automatically flagged. An operator could then inspect these frames and manually fix any errors using our graphical user interface. All manually determined identity errors in the benchmark sequences were also flagged automatically, and thus, error detection was 100% accurate with this limited supervision.

Position errors. We simultaneously recorded high-resolution video (15× standard resolution, corresponding to fly lengths of 120 pixels) of a portion of the arena with our standard lower-resolution video of the entire arena (**Fig. 2e**). We labeled the positions manually in the high-resolution video and compared them to those computed by the tracker from the lower-resolution video. The high-resolution labels were transformed into the lower-resolution coordinate system for this comparison (**Supplementary Fig. 1** and **Supplementary Note**). A random sample of 100 flies from 9 5-min video sequences was used. As above, each video contained 10, 20 or 50 flies and each contained either all male, all female or half male and half female flies. We chose frames in the high-resolution video in which flies were fully visible and far from other flies. The manual annotation consisted of a carefully drawn bounding box of the fly and was used to estimate the center position and orientation of the fly. We repeated the above experiment on 50 samples in which the chosen fly was close to another fly. The median error was 0.0292 mm (0.117 pixels) for the center and 3.14° for the orientation (**Fig. 2e** and **Supplementary Fig. 2**). For touching flies, the median errors were slightly larger: 0.0461 mm for the center position and 10.6° for the orientation (**Supplementary Table 3** online).

Gender assignment. As female flies are slightly larger than male flies, a fly's sex could be automatically predicted from its image area. For each trajectory, the median area was computed and sex was assigned by comparing this area to a threshold estimated from single-sex experiments (correcting for biases from lighting variations in different parts of the arena). The hold-one-out error rate was $4/77 = 0.0519$ for females and $3/106 = 0.0283$ for males.

Behavior definitions. All our behavior definitions had the following structure. The fly was performing the defined behavior from frames t_1 to t_2 if all of the following applied. (i) In each frame $t_1,\ldots,t_2$, properties of the fly (for example, speed, distance to another fly) were within given ranges. (ii) In each frame $t_1,\ldots,t_2$, properties of the fly were temporally near (within a given number of frames) frames in which the properties were within tighter ranges. (iii) The summed properties (for example, total distance traveled) of the fly's trajectory in $t_1,\ldots t_2$ were within given ranges. (iv) The mean value of properties of the fly were within given ranges.

Social behaviors operated on properties of pairs of flies rather than individuals. Parameters of each behavior, including the properties and ranges for each of the above rules, are given in **Supplementary Table 2**.

For each behavior, each trajectory was segmented into intervals in which the fly was and was not performing the behavior by maximizing the sum-squared lengths of the positive sequences using a globally optimal, dynamic programming algorithm. Note that this one-versus-all set of behavior detectors resulted in some frames of the trajectory not being labeled at all (our behavior vocabulary is incomplete), and that a fly may have been engaged in multiple behaviors at the same time (for example, chasing and walking).

Our software allowed us to define behavior detectors in two ways. The quickest way was direct manual selection of the ranges of property values defining a behavior. We found this approach intuitive and easy for a couple of behaviors ('back up' and 'touch'). In all other cases, we used example-based training to learn the ranges. Using the latter approach, a user manually segmented sample trajectories to create training data. The parameter ranges were then computed automatically so that the detected segmentations agreed with the manual segmentations (**Supplementary Note**). In both cases, other scientists may inspect the parameter ranges defining specific behaviors and thus reproduce exactly a given experiment.

31. Piccardi, M. Background subtraction techniques: a review. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics* **4**, 3099–3104 (2004).
32. Gonzalez, R.C. & Woods, R.E. *Digital Image Processing* (Prentice Hall, Upper Saddle River, New Jersey, USA, 2007).
33. Papadimitriou, C.H. & Steiglitz, K. *Combinatorial Optimization: Algorithms and Complexity* (Dover Publications, Mineola, New York, USA, 1998).
34. Perera, A., Srinivas, C., Hoogs, A. & Brooksby, G. Multi-object tracking through simultaneous long occlusions and split-merge conditions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* **1**, 666–673 (2006).
35. Cormen, T.H. *Introduction to Algorithms* (MIT Press, Cambridge, Massachusetts, USA, 2001).