

Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning

Weizhe Hong^{a,1}, Ann Kennedy^a, Xavier P. Burgos-Artizzu^b, Moriel Zelikowsky^a, Santiago G. Navonne^b, Pietro Perona^{b,1}, and David J. Anderson^{a,1}

^aDivision of Biology and Biological Engineering 156-29, Howard Hughes Medical Institute, California Institute of Technology, Pasadena, CA 91125; and ^bDivision of Engineering and Applied Sciences 136-93, California Institute of Technology, Pasadena, CA 91125

Contributed by David J. Anderson, August 16, 2015 (sent for review May 20, 2015)

A lack of automated, quantitative, and accurate assessment of social behaviors in mammalian animal models has limited progress toward understanding mechanisms underlying social interactions and their disorders such as autism. Here we present a new integrated hardware and software system that combines video tracking, depth sensing, and machine learning for automatic detection and quantification of social behaviors involving close and dynamic interactions between two mice of different coat colors in their home cage. We designed a hardware setup that integrates traditional video cameras with a depth camera, developed computer vision tools to extract the body “pose” of individual animals in a social context, and used a supervised learning algorithm to classify several well-described social behaviors. We validated the robustness of the automated classifiers in various experimental settings and used them to examine how genetic background, such as that of Black and Tan Brachyury (BTBR) mice (a previously reported autism model), influences social behavior. Our integrated approach allows for rapid, automated measurement of social behaviors across diverse experimental designs and also affords the ability to develop new, objective behavioral metrics.

social behavior | behavioral tracking | machine vision | depth sensing | supervised machine learning

Social behaviors are critical for animals to survive and reproduce. Although many social behaviors are innate, they must also be dynamic and flexible to allow adaptation to a rapidly changing environment. The study of social behaviors in model organisms requires accurate detection and quantification of such behaviors (1–3). Although automated systems for behavioral scoring in rodents are available (4–8), they are generally limited to single-animal assays, and their capabilities are restricted either to simple tracking or to specific behaviors that are measured using a dedicated apparatus (6–11) (e.g., elevated plus maze, light-dark box, etc.). By contrast, rodent social behaviors are typically scored manually. This is slow, highly labor-intensive, and subjective, resulting in analysis bottlenecks as well as inconsistencies between different human observers. These issues limit progress toward understanding the function of neural circuits and genes controlling social behaviors and their dysfunction in disorders such as autism (1, 12). In principle, these obstacles could be overcome through the development of automated systems for detecting and measuring social behaviors.

Automating tracking and behavioral measurements during social interactions pose a number of challenges not encountered in single-animal assays, however, especially in the home cage environment (2). During many social behaviors, such as aggression or mating, two animals are in close proximity and often cross or touch each other, resulting in partial occlusion. This makes tracking body positions, distinguishing each mouse, and detecting behaviors particularly difficult. This is compounded by the fact that such social interactions are typically measured in the animals’ home cage, where bedding, food pellets, and other moveable items can make tracking difficult. Nevertheless a home-cage environment

is important for studying social behaviors, because it avoids the stress imposed by an unfamiliar testing environment.

Recently several techniques have been developed to track social behaviors in animals with rigid exoskeletons, such as the fruit fly *Drosophila*, which have relatively few degrees of freedom in their movements (13–23). These techniques have had a transformative impact on the study of social behaviors in that species (2). Accordingly, the development of similar methods for mammalian animal models, such as the mouse, could have a similar impact as well. However, endoskeletal animals exhibit diverse and flexible postures, and their actions during any one social behavior, such as aggression, are much less stereotyped than in flies. This presents a dual challenge to automated behavior classification: first, to accurately extract a representation of an animal’s posture from observed data, and second, to map that representation to the correct behavior (24–27). Current machine vision algorithms that track social interactions in mice mainly use the relative positions of two animals (25, 28–30); this approach generally cannot discriminate social interactions that involve close proximity and vigorous physical activity, or identify specific behaviors such as aggression and mounting. In addition, existing algorithms that measure social interactions use a set of hardcoded, “hand-crafted” (i.e., pre-defined) parameters that make them difficult to adapt to new experimental setups and conditions (25, 31).

In this study, we combined 3D tracking and machine learning in an integrated system that can automatically detect, classify, and quantify distinct social behaviors, including those involving

Significance

Accurate, quantitative measurement of animal social behaviors is critical, not only for researchers in academic institutions studying social behavior and related mental disorders, but also for pharmaceutical companies developing drugs to treat disorders affecting social interactions, such as autism and schizophrenia. Here we describe an integrated hardware and software system that combines video tracking, depth-sensing technology, machine vision, and machine learning to automatically detect and score innate social behaviors, such as aggression, mating, and social investigation, between mice in a home-cage environment. This technology has the potential to have a transformative impact on the study of the neural mechanisms underlying social behavior and the development of new drug therapies for psychiatric disorders in humans.

Author contributions: W.H., P.P., and D.J.A. designed research; W.H. performed research; W.H., X.P.B.-A., and S.G.N. contributed new reagents/analytic tools; W.H., A.K., M.Z., P.P., and D.J.A. analyzed data; and W.H., A.K., M.Z., P.P., and D.J.A. wrote the paper.

The authors declare no conflict of interest.

¹To whom correspondence may be addressed. Email: whong@caltech.edu, perona@caltech.edu, or wuwe@caltech.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1515982112/-DCSupplemental.

close and dynamic contacts between two mice in their home cage. To do this, we designed a hardware setup that synchronizes acquisition of video and depth camera recordings and developed software that registers data between the cameras and depth sensor to produce an accurate representation and segmentation of individual animals. We then developed a computer vision tool that extracts a representation of the location and body pose (orientation, posture, etc.) of individual animals and used this representation to train a supervised machine learning algorithm to detect specific social behaviors. We found that our learning algorithm was able to accurately classify several social behaviors between two animals with distinct coat colors, including aggression, mating, and close social investigation. We then evaluated the robustness of our social behavior classifier in different experimental conditions and examined how genetic backgrounds influence social behavior. The highly flexible, multistep approach presented here allows different users to train new customized behavior classifiers according to their needs and to analyze a variety of behaviors in diverse experimental setups.

Results

Three-Dimensional Tracking Hardware Setup. Most current mouse tracking software is designed for use with 2D videos recorded from a top- or side-view camera (24–28). Two-dimensional video analysis has several limitations, such as difficulty resolving occlusion between animals, difficulty detecting vertical movement, and poor animal tracking performance against backgrounds of similar color. To overcome these problems, we developed an integrated hardware setup with synchronized image acquisition and software to record behavior using synchronized video cameras and a depth sensor. Depth sensors detect depth values of an object in the z-plane by measuring the time-of-flight of an infrared light signal between the camera and the object for each point of the image (32), in a manner analogous to sonar.

We compared two commercially available depth sensors, the Kinect Sensor from Microsoft Corporation and the Senz3D depth and gesture sensor from Creative Technology Ltd. (Fig. S1). We developed customized software to acquire raw depth images from both sensors. Although the Kinect sensor has been recently used for behavioral tracking in rats (31), pigeons (33), pigs (34), and human (35), we found that its spatial resolution was not sufficient for resolving pairs of mice, which are considerably smaller; in contrast, the Senz3D sensor's higher 3D resolution made it better suited for this application (Fig. S1). This higher resolution was partly because the Senz3D sensor was designed for a closer working range (15–100 cm) than the Kinect (80–400 cm). In addition, the Senz3D sensor's smaller form factor allowed us to build a compact customized behavior chamber with 3D video acquisition capability and space for insertion of a standard mouse cage (Fig. 1A).

We installed a side-view conventional video camera in front of the cage as well as a top-view video camera and the Senz3D sensor on top of the cage (Fig. 1A and B). Videos taken from the side-view and top-view cameras provided additional and complementary data, such as luminosity, for the postacquisition image analysis and behavior analysis and allowed users to manually inspect and score behaviors from different angles. Data were acquired synchronously by all three devices to produce simultaneous depth information and top- and side-view grayscale videos (*Methods*). Representative video frames from each of the three devices during three social behaviors (aggression, mounting, and close investigation) are shown in Fig. 1C.

Mice are nocturnal animals, and exposure to white light disrupts their circadian cycle. Therefore, we recorded animal behaviors under red light illumination, which is considered “dark” for mice, because mice cannot perceive light within the red-to-infrared spectrum. Both video cameras and the depth sensor in our system work under red light and do not rely on white-light illumination.

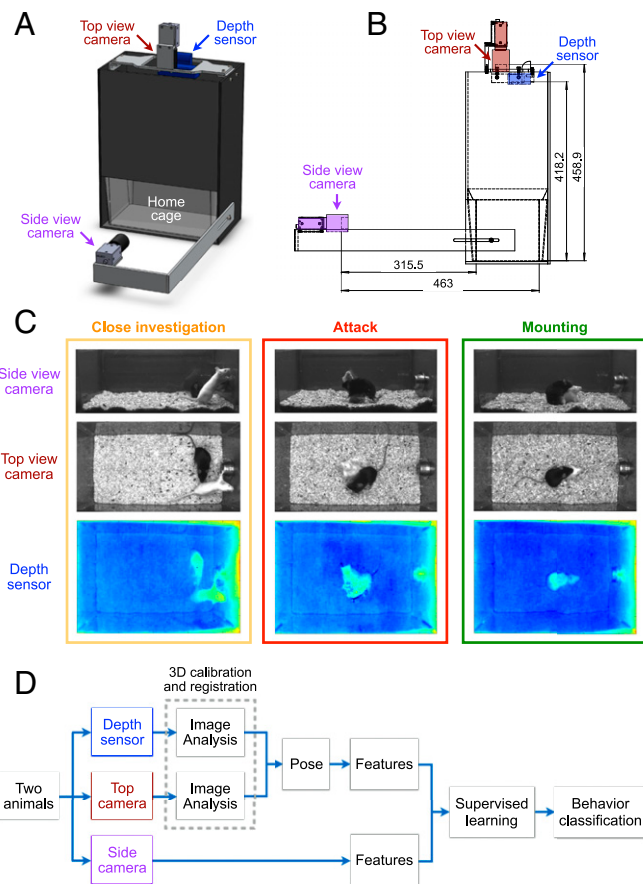


Fig. 1. Equipment setup and workflow. (A and B) Schematic illustrating the customized behavior chamber. A standardized mouse cage can be placed inside the chamber. The front-view video camera is located in front of the cage, and the top-view video camera and the Senz3D sensor are located on top of the cage. Unit: millimeters. (C) Representative synchronized video frames taken from the two video cameras and the depth sensor. (D) A workflow illustrating the major steps of the postacquisition image analysis and behavior analysis.

To separate animal identities, our system is currently limited to tracking and classifying two mice of different coat colors.

The major steps of the postacquisition image analysis and behavior analysis (Fig. 1D and Fig. S2) are described in the following sections.

Image Processing, Pose Estimation, and Feature Extraction. To integrate the monochrome video recordings from the top-view camera with the data from the depth sensor, we registered them into a common coordinate framework using the stereo calibration procedure from MATLAB's Computer Vision System Toolbox, in which a planar checkerboard pattern is used to fit a parameterized model of each camera (Fig. 2A and B and Fig. S3). The top-view camera and depth sensor were placed as close as possible to each other, to minimize parallax (Fig. 2A). We then projected the top-view video frames into the coordinates of the depth sensor (Fig. S3) to obtain simultaneous depth and intensity values for each pixel.

We performed background subtraction and image segmentation using reconstructed data from the top-view camera and depth sensor to determine the location and identity of the two animals (Fig. S4 and *Methods*). To obtain a low-dimensional representation of animal posture (“pose”), we fit an ellipse to each animal detected in the segmented video frames (Fig. 2C). The body orientation of each animal was determined from its

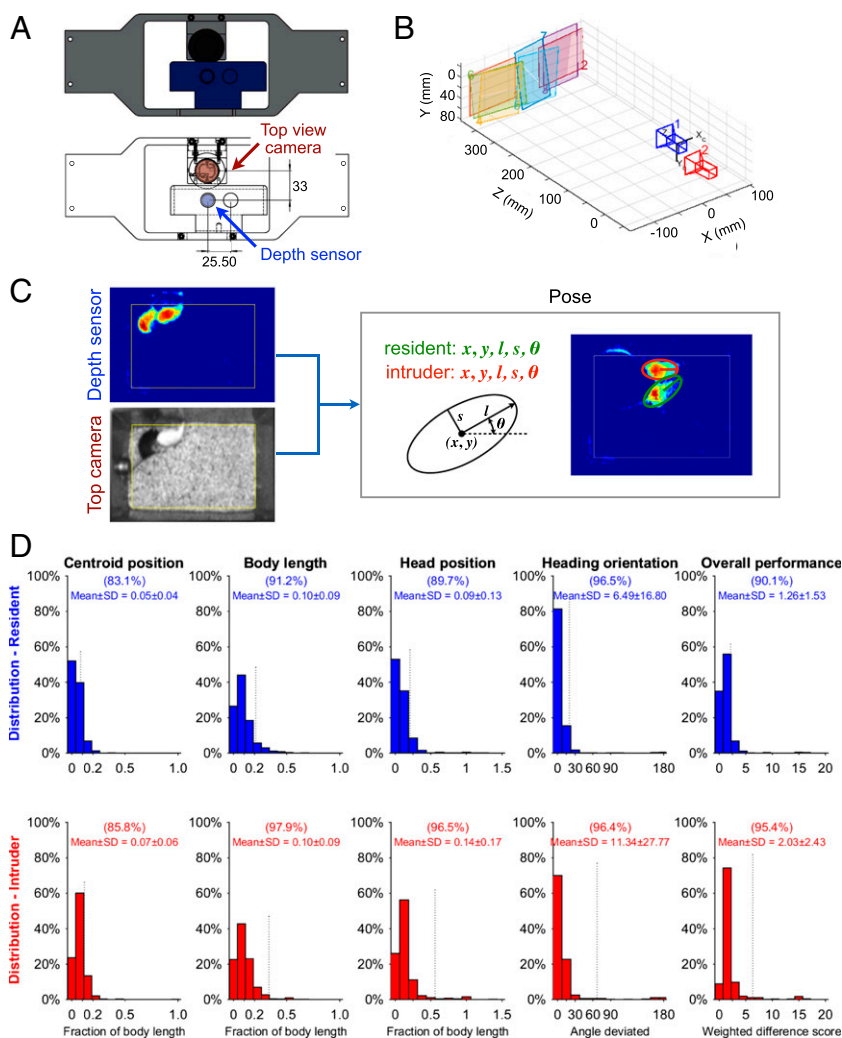


Fig. 2. Image processing, animal tracking, and pose estimation. (A) Schematic illustrating the setup of the top-view video camera and the depth sensor on top of the cage. The top-view camera and depth sensor were placed as close as possible to minimize the parallax effect. Unit: millimeters. (B) MATLAB-generated schematic showing 3D registration of the top-view video camera and the depth sensor into a common coordinate system. Locations of checkerboard patterns (*Methods* and Fig. S3) used for calibration are shown on the left, and the calculated positions of the two cameras are shown on the right. (C) Pose estimation using information from both top view camera and depth sensor. An ellipse that best fits an animal detected in the segmented 3D video frames is used to describe the position, orientation, shape, and scale of the animal. Head orientation is determined by the standing position, moving direction, and a set of features extracted using a previously developed machine learning algorithm (*Methods*). The pose of an animal is thus described by an ellipse using a set of five parameters: centroid position (x, y), length of the long axis (l), length of the short axis (s), and head orientation (θ). (D) Validation of pose estimation against ground truth (manually annotated ellipses in individual video frames). Each histogram represents the distribution of differences of individual pose parameters and overall performance between pose estimation and ground truth (see *Methods* for the definition of differences of individual pose parameters and overall performance). Numbers in the parenthesis at the top of each plot represent the percentage of frames to the left of the dashed lines, which represent the 98% percentiles of the differences between two independent human observers (Fig. S5). $n = 634$ frames.

position and movement direction, as well as from features detected by a previously developed machine learning algorithm (24, 27) (*Methods*). Thus, the pose of each animal is described by a set of five parameters from the fit ellipse: centroid position (x, y), length of the major axis (l), length of the minor axis (s), and body orientation (θ). To evaluate the performance of the automated pose estimation, we constructed a ground truth dataset of manually annotated ellipses and calculated the differences between automatic and manual estimation for the parameters including centroid position, body length, head orientation, and head position (Fig. 2D). We also evaluated the overall performance by computing the weighted differences between the machine annotated and manually annotated ellipses using a previously developed metric (27) (Fig. 2D and *Methods*). We found that the algorithm was able to track the position and the body orientation

of the animals in a robust manner (Fig. 2D and *Movie S1*), compared with the performance of two independent human observers (Fig. S5).

Using the five fit ellipse parameters and additional data from the depth sensor, we developed a set of 16 second-order features describing the state of each animal in each video frame (Fig. 3A and B) and 11 “window” features computed over multiple frames, giving 27 total features (*Methods*). Principal component analysis of features extracted from a sample set of movies indicated that the 16 second-order features are largely independent of each other, such that the set of recorded behaviors spanned nearly the entire feature space (Fig. 3C and D).

Supervised Classification of Social Behaviors. We next explored the use of supervised machine learning approaches for automated

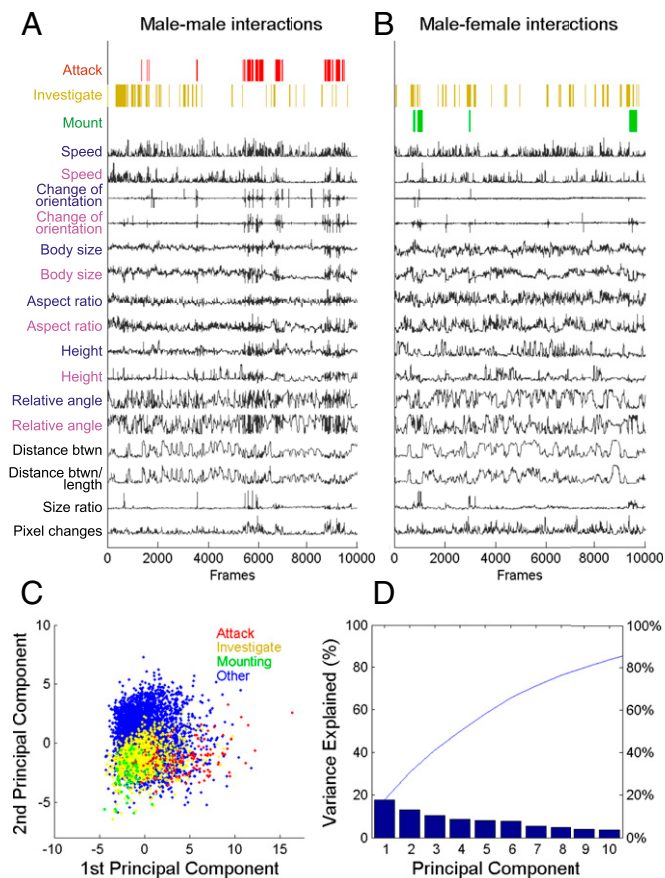


Fig. 3. Feature extraction. (A and B) In each video frame, a set of measurements (features) is computed from the pose and height of animals, describing the state of individual animals (blue: animal 1 or the resident; magenta: animal 2 or the intruder) and their relative positions (black). See [Supporting Information](#) for a complete list and descriptions of features. Two representative video episodes, one during male–male interaction and the other during male–female interaction, are shown. The human annotations of three social behaviors are shown in the raster plot on the top. (C and D) Principle component analysis of high-dimensional framewise features. (C) The first two principal components are plotted. “Other” represents frames that were not annotated as any of the three social behaviors. (D) Variance accounted for by the first 10 principal components; bars show the fraction of variance accounted for by each component, and the line shows the cumulative variance accounted for.

annotation of social behavior. In supervised learning, classifiers are trained using datasets that have been manually annotated with the desired classifier output, to find a function that best reproduces these manual annotations. The performance of the classifier is evaluated using a testing set of ground-truth videos not used for training. The training set and the test set have no overlap and were obtained from separate videos. We used our 27 extracted features to test several supervised learning algorithms, including support vector machine (SVM), adaptive boosting (adaBoost), and random decision forest (TreeBagger). The random decision forest gave us the best performance in prediction accuracy and training speed and was thus selected for further investigation. We trained three social behavior classifiers (attack, mounting, and close investigation; see *Methods* for the criteria used by human annotators) using a set of six videos that contained ~150,000 frames that were manually annotated on a frame-by-frame basis. We chose to generate 200 random decision trees, which was beyond where the error rate plateaued (Fig. 4G); because individual decision trees were built independently,

the process of training the decision forest is parallelizable and can be greatly sped up on a multicore computer. The output of our three behavior detectors for three representative videos is shown in Fig. 4A–D (male–male interactions) and Fig. 4E and F (male–female interactions). As seen in the expanded raster plots (Fig. 4B, D, and F), there is a qualitatively close correspondence between ground truth and prediction bouts for attack, close investigation, and mounting. The contribution of individual features to classifier performance is shown in Fig. 4H.

To measure the accuracy of these behavior classifiers in replicating human annotations, we manually labeled a set of 14 videos (not including the videos used to train the classifier) that contained ~350,000 frames from a variety of experimental conditions and measured classifier error on a frame-by-frame basis. We plotted classifier performance using the detection error tradeoff (DET) curve representing the framewise false negative rate vs. the false-positive rate (Fig. 4I) and the precision-recall curve representing the framewise true positive rate vs. the positive predictive rate (Fig. 4J), using the human annotations as ground truth. These measurements illustrated the tradeoff between the true positive rate vs. the positive predictive value at different classification thresholds from 0 to 1. Here we chose a classification threshold that optimized the framewise precision and recall; the framewise precision, recall, fallout, and accuracy rates at the classification threshold are shown in Fig. 4K. All of the classifiers showed an overall prediction accuracy of 99% for attack, 99% for mounting, and 92% for close investigation. Finally, we measured the precision and recall rates at the level of individual behavioral episodes (“bouts”), periods in which all frames were labeled for a given behavior. We observed a high level of boutwise precision and recall across a range of minimum bout durations (Fig. 4K and [Movies S2–S4](#)).

Use-Case 1: Genetic Influences on Social Behaviors. To explore the utility of our behavior classifiers, we used them to track several biologically relevant behaviors under several experimental conditions. We first used the classifier to annotate resident male behavior during interactions with either a male or a female intruder (Fig. 5; I_m vs. I_f , respectively). We examined the percentage of time resident males spent engaging in attack, mounting, and close investigation of conspecifics (Fig. 5A–C); note that this parameter is not directly comparable across behaviors, because the average bout length for each behavior may be different. We therefore also measured the total numbers of bouts during recording (Fig. 5D–F), the latency to the first bout of behavior for each resident male (Fig. 5G–I), and the distribution of bout lengths for each behavior (Fig. 5J–R). We observed that for our standard strain C57BL/6N, male residents (R_{C57N}) exhibited more close investigation bouts with longer duration toward male (Fig. 5N; I_m) than that toward female (Fig. 5K; I_f) intruders ($P < 0.001$), although the total numbers of bouts were comparable between the two conditions (Fig. 5E). The classifier predictions showed no significant differences from the ground truth in the measured percentage of time spent engaging in each behavior, nor in the bout length distribution of each behavior (Fig. 5K, N, and Q, yellow vs. gray bars) (~350,000 frames total), suggesting that the same classifiers work robustly in both male–male and male–female interactions.

To examine how genetic backgrounds influence social behaviors, we compared two strains of resident male mice, C57BL/6N and NZB/B1NJ (Fig. 5). NZB/B1NJ mice were previously shown to be more aggressive than C57BL/6N (36). Consistently, we found that NZB/B1NJ resident males spent more time attacking BALB/c intruder males, and significantly less time engaging in close investigation, than did C57BL/6N resident males (Fig. 5A and B; R_{NZB}) ($P < 0.05$). This likely reflects a more rapid transition from close investigation to attack, because the average latency to attack was much shorter for NZB/B1NJ than for C57BL/6N males (Fig. 5G). Interestingly, NZB/B1NJ animals exhibited both a higher number of attack bouts (Fig. 5D) ($P < 0.05$) and longer average

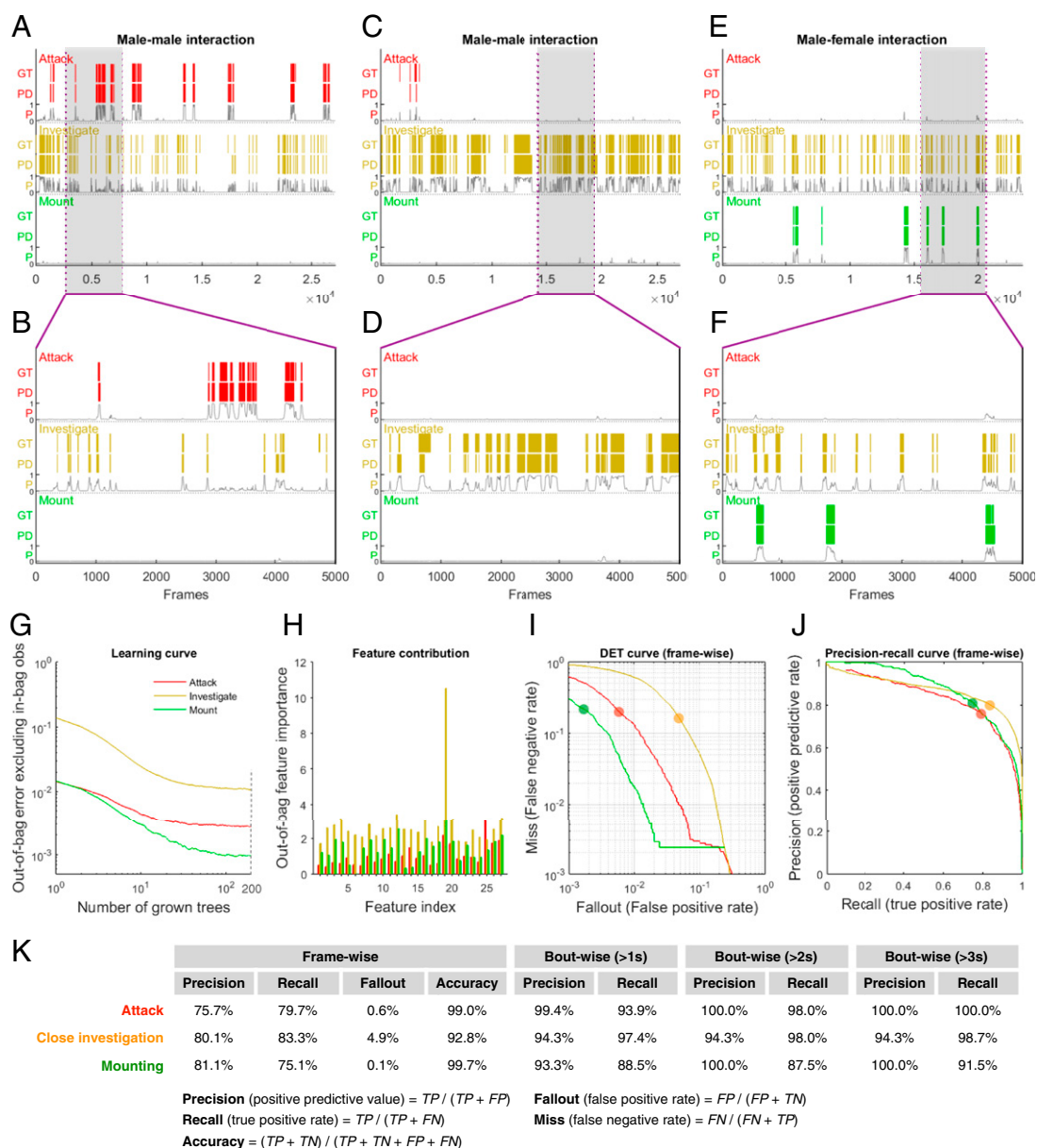


Fig. 4. Supervised classification of social behaviors. (A–K) Classification of attack, mounting, and close investigation using TreeBagger, a random forest classifier. (A–F) Raster plots showing manual annotations of attack, close investigation, and mounting behaviors, as the ground truth, vs. the machine learning classifications of these social behaviors. Three representative videos with different experimental conditions were used as the test set. A and C illustrate two representative examples of male–male interactions. GT, ground truth; P, probability; PD, machine classification/prediction. (G) Learning curve of different behavior classifiers represented by out-of-bag error as a function of the number of grown trees. (H) Contribution of distinct features to individual classifiers. See *Supporting Information* for a complete list and descriptions of features. (I) DET curve representing the false negative rate vs. the false positive rate in a frame-wise manner. (J) Precision–recall curve representing true positive rate vs. positive predictive value in a frame-wise manner. (K) Table of precision, recall, fallout, and accuracy at the level of individual frames, as well as precision and recall at the level of individual behavioral episodes (“bouts”) for a range of minimum bout durations (>1 s, >2 s, and >3 s). Classification thresholds in A–F and K are 0.55 (attack), 0.5 (close investigation), and 0.4 (mounting) and are highlighted in red, orange, and green dots, respectively, in I and J.

attack durations compared with C57BL/6N animals (Fig. 5 *M* and *P*) ($P < 0.05$). These data illustrate the ability of the method to reveal differences between the manner in which NZB/B1NJ and C57BL/6N males socially interacted with intruder animals of a different strain. In all measurements, the classifier prediction showed no significant differences from the ground truth (Fig. 5), suggesting that the same classifiers work robustly with distinct strains of animals that exhibit very different social behaviors.

Use-Case 2: Detection of Social Deficits in an Autism Model. To explore the utility of our behavioral classifiers in detecting social

deficits in mouse models of autism, we examined the behavior in Black and Tan Brachyury (BTBR) T+tf/J (BTBR) mice, an inbred mouse strain that was previously shown to display autism-like behavioral phenotypes, such as reduced social interactions, compared with C57BL/6N animals (1, 37–39). Here we measured parameters of social interactions between BTBR mice (or C57BL/6N control mice) and a “target” animal of the BALB/c strain, in an unfamiliar, neutral cage. By using our behavioral analysis system to track the locations, poses, and behaviors of the interacting animals, we observed significantly less social investigation performed by BTBR animals in comparison with C57BL/6N controls

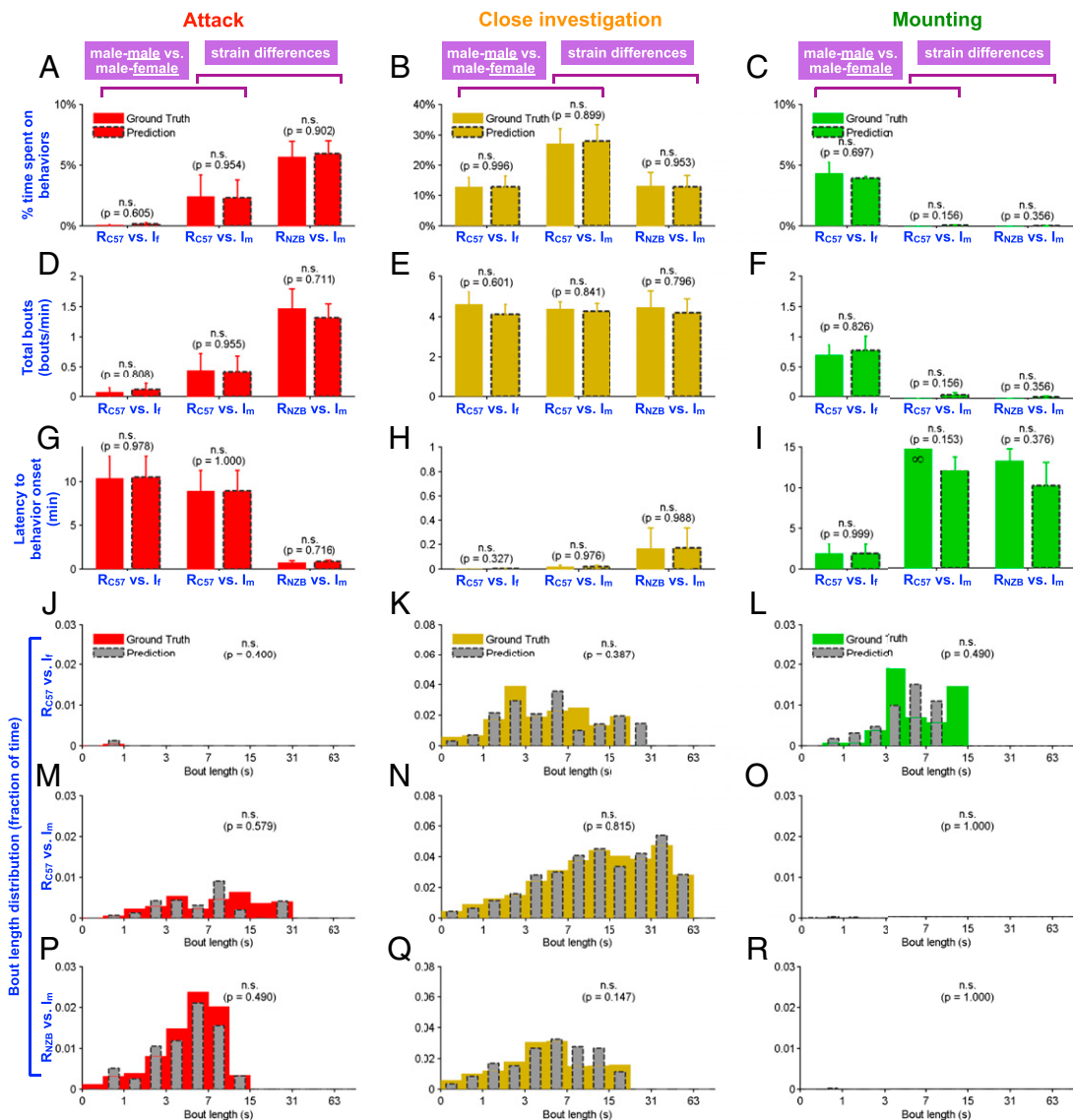


Fig. 5. Genetic influences on social behaviors. (A–R) We examined the effects of the genetic and environmental influences on attack, mounting, and close-investigation behaviors in three different experimental conditions and validated the performance of the social behavior classifiers in these conditions. In each of panels A–I, the left two bars are from trials in which C57BL/6N male residents were tested with female intruders, the middle two bars are from C57BL/6N male residents tested with male intruders, and the right two bars are from NZB/B1N1 male residents tested with male intruders. All intruders are BALB/c. (A–C) Percentage of time spent on attack, close investigation, or mounting behavior during 15-min behavior sessions. (D–F) Total bouts per minute of individual behaviors during the same behavior sessions. (G–I) Latency to the first bout of individual behaviors during the same behavior sessions. (J–R) Histograms of behavioral bout duration (fraction of total time), as measured by the classifier and as measured by hand, for each type of resident–intruder pair and each behavior class. (J, M, and P) Attack. (K, N, and Q) Close investigation. (L, O, and R) Mounting. RC57: C57N male resident; RNZB: NZB male resident; Im: BALB/c male intruder; If: BALB/c female intruder.

(Fig. 6A–C), consistent with previous reports (38, 39). In particular, the BTBR animals displayed shorter bouts of (Fig. 6B), and reduced total time engaged in (Fig. 6C), social investigation.

To determine whether this reduction of social investigation reflects less investigation of the BALB/c mouse by the BTBR mouse (in comparison with the C57BL/6N controls), or vice versa, we measured the social investigation behavior performed by the BALB/c mouse. BALB/c animals did not exhibit reduced social interactions with the BTBR mice in comparison with C57BL/6N controls (Fig. S6A and B). This suggests that the reduction of social investigation observed in BTBR animals is indeed due to less investigation of the BALB/c mouse by the BTBR mouse.

Finally, we asked whether pose estimation and supervised behavioral classifications offered additional information beyond

tracking animal location alone. We first measured “body–body” distance—the distance between centroid locations of two interacting animals (illustrated in the schematic in Fig. 6D)—a measurement that only used the output from tracking animal location alone but not from pose estimation or behavioral classifiers. We observed a trend to decreased time spent at short body–body distances (<6 cm) in BTBR animals (Fig. 6D and E), but this effect was not statistically significant. When we measured “head–body” distance—the distance between the front end of the subject and the centroid of the other animal (illustrated in the schematic in Fig. 6F)—a measurement that used output from both tracking and pose estimation, but not from supervised behavioral classifications, we observed a statistically significant reduction in time spent at short (<4 cm) head–body distances in BTBR animals paired with BALB/c mice (Fig. 6F and G),

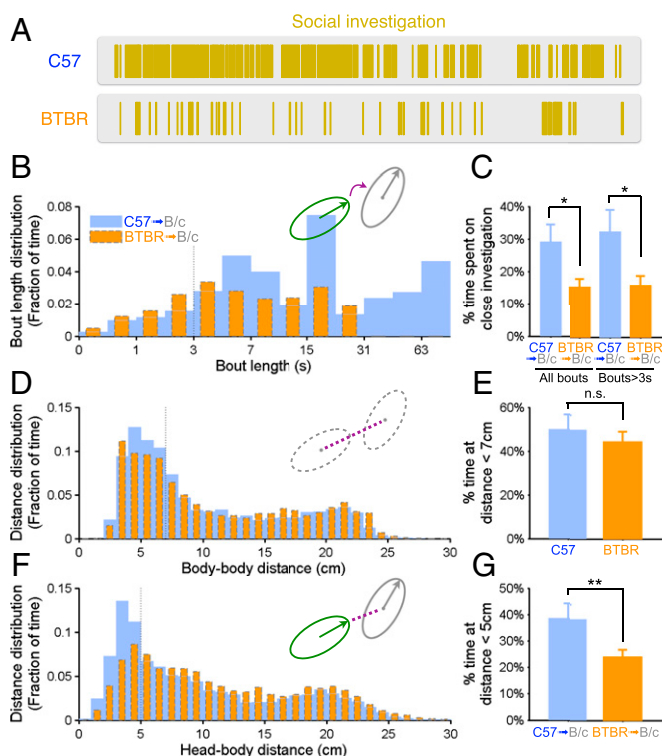


Fig. 6. Detection of social deficits in BTBR animals. C57BL/6N or BTBR animals were tested with a BALB/c male in an unfamiliar, neutral cage. (A) Raster plots showing the supervised classifier-based machine annotations of social investigation behavior exhibited by C57BL/6N or BTBR tester mice in the first 5 min of their interactions with BALB/c target animals. (B) Histograms of behavioral bout duration (fraction of total time) for social investigation exhibited by C57BL/6N or BTBR animals toward BALB/c during the first 5 min of their interactions. (C) Percentage of time spent on close investigation during the same behavior sessions. (D) Distribution of the distance between centroid locations of two interacting animals (fraction of total time) during the same behavior sessions. (E) Percentage of time the centroids of two interacting animals are within 6 cm during the same behavior sessions. (F) Distribution of the distance between the front end of the subject (BTBR or C57BL/6N) and the centroid of the BALB/c animal (fraction of total time) during the same behavior sessions. Note that no significant difference between tester strains is evident using this tracker-based approach to analyze the interactions. (G) Percentage of time the front end of the tester (BTBR or C57BL/6N) mouse is within 4 cm from the centroid of the target BALB/c animals during the same behavior sessions. Metrics in D and E are based solely on output from the tracker, metrics in F and G are based on output from the tracker and pose estimator, and metrics in A–C are derived from the automated behavioral classifier. See Fig. S6 for metrics equivalent to D–G analyzed for the BALB/c target mouse.

compared with that in C57BL/6N animals paired with BALB/c. This difference did not reflect reduced investigation of BTBR animals by BALB/c mice, because the latter did not show a significant difference in time spent at short head–body distances toward BTBR vs. C57BL/6N mice (Fig. S6 C and D). Rather, the difference reflects reduced close investigation of BALB/c mice by BTBR mice in comparison with C57BL/6N controls. These data together suggest that our behavioral tracking system was able to detect social behavioral deficits in BTBR mice, a mouse model of autism, and that compared with animal location tracking alone, pose estimation and supervised behavioral classification provide additional useful information in detecting behavioral and phenotypic differences.

Discussion

Although a great deal of progress has been made in marking, imaging, and manipulating the activity of neural circuits (40–42),

much less has been done to detect and quantify the behaviors those circuits control in freely moving animals, particularly in the case of social behaviors. Social behaviors are especially hard to quantify, because they require separating and maintaining the identities, positions, and orientations of two different animals during close and dynamic interactions. This is made particularly difficult by occlusion when the animals are close together—and most social behaviors in mice occur when the animals are in proximity to each other. Moreover, social behavioral assays are ideally performed in the home cage, where bedding absorbs familiar odors and allows digging, nesting, and other activities. The fact that bedding is textured and may be rearranged by the mice presents additional challenges for tracking and pose estimation. Most mouse trackers were developed for use in a novel arena with a bare solid-color floor, to facilitate the separation of the animal from background (4, 5, 7, 43); this type of arena, however, can be stressful to animals, and may perturb their social behavior.

Here we describe and test a hardware and software platform that integrates 3D tracking and machine learning for the automated detection and quantification of social behavior in mice. We used this tool to track animal trajectories and orientations in the context of an animal's home cage and detect specific social behaviors, including attack, mounting, and close investigation. Our approach significantly extends existing methods for behavior tracking and classification, which typically do not work well when pairs of mice are in close contact or partially overlapping and/or do not provide specific behavior classification such as attack (25, 28). The automated behavior scoring method we have introduced here should greatly facilitate study of the neural circuits and genes that regulate social behavior.

Our system annotates behavioral data at three levels: (i) simple video tracking, which locates the centroid of an ellipse fit to each mouse in each frame; (ii) pose estimation, which combines information from the video and depth camera recordings to determine the orientation (head vs. tail), height, and other postural features of each mouse relative to the other; and (iii) automated behavioral classification and scoring using the supervised machine learning-based classifiers. We show that tracking analysis alone was incapable of detecting differences in the frequency of social interactions between control C57BL/6N mice and BTBR mice, a previously reported autism model (1, 37–39). Application of the pose estimator, by contrast, detected a significant difference between strains, as did the automated behavior classifier. The classifier also provided additional metrics, such as investigation bout-length distribution, that were not available from the pose estimator. These data suggest that our system may be useful for detecting and quantifying subtle differences in social behavior phenotypes caused by genetic or circuit-level perturbations.

A major advantage of the technology described here is increased throughput and decreased labor intensiveness. Typically, it takes about 6 h of manual labor to score each hour of video, on a frame-by-frame basis at 30 Hz, particularly if multiple behaviors are being analyzed. A typical study using social behavior as a quantitative readout may require analyzing dozens or scores of hours of video recordings (44). Our system reduces that time requirements of analysis to an initial commitment of several hours to manually generate a training set of annotations and a few minutes to train the classifier, after which large numbers of additional videos can be scored in a matter of minutes. This not only eliminates major bottlenecks in throughput but should improve the statistical power of behavioral studies by enabling larger sample sizes; this is often a problem for behavioral assays, which typically exhibit high variance (45). Our method also opens up the possibility of using behavioral assays as a primary, high-throughput screen for drugs or gene variants affecting mouse models of social interaction disorders, such as autism (1). In addition to this time- and labor-saving advantage, whereas human observers may fail to detect behavioral events due to fatigue or

flagging attention, miss events that are too quick or too slow, or exhibit inconsistencies between different observers in manually scoring the same videos, supervised behavior classifiers apply consistent, objective criteria to the entire set of videos, avoiding potential subjective or irreproducible annotations.

Use of the depth sensor offers several unique advantages over the traditional 2D video analysis. Depth values improve detection of the animal's body direction and provide better detection of vertical movements that are relevant to some behaviors. Because the depth sensor is able to detect mice by their height alone, the system works under red light illumination, is insensitive to background colors, and is particularly useful in more natural environments such as home cages. This is helpful in studying social behavior, because removing an animal from its home cage for recording or exposing animals to white light illumination heightens stress and affects behavior.

Although a previous study reported the use of depth cameras (Kinect) to track rats during social interactions (31), that approach differs from ours, in that it used the camera output to construct a 3D pose of each animal that was then used to manually classify different behaviors. In contrast, in our method the output of the position and pose tracker was passed through a set of feature extractors, producing a low-dimensional representation with which machine learning algorithms were used to train classifiers to detect specific social behaviors recognizable by human observers. Using the feature extractors removed uninformative sources of variability from the raw video data and reduced susceptibility of the classifier to overtraining, producing automated behavioral annotations that were accurate and robust.

Our system used coat color to separate and keep track of the identities of both animals and is limited to tracking and phenotyping two mice. Future improvements could enable the tracking of more than two animals, and/or animals with identical coat colors. Our system is also limited to detecting the main body trunk of the animals and is unable to track finer body parts, such as limbs, tails, whiskers, nose, eyes, ears, and mouth. Although detecting the main body trunk is sufficient to build robust classifiers for several social behaviors, constructing a more complete skeleton model with finer body-part resolution and tracking should provide additional information that may allow the classification of more subtle behaviors, such as self-grooming, tail-rattling, or different subtypes of attack. This is currently technically challenging given the resolution of available depth-sensing cameras and could be potentially solved through future improvements in depth-sensing technology.

In summary, we describe the first application, to our knowledge, of synchronized video and depth camera recordings, in combination with machine vision and supervised machine learning methods, to perform automated tracking and quantification of specific social behaviors between pairs of interacting mice in a home-cage environment, with a time resolution (30 Hz) commensurate with that of functional imaging using fluorescent calcium or voltage sensors (46). Integration of this methodology with hardware and technology for manipulation or measurement of neuronal activity should greatly facilitate correlative and causal analysis of the brain mechanisms that underlie complex social behaviors and may improve our understanding of their dysfunction in animal models of human psychiatric disorders (1, 12).

Methods

Hardware Setup. The customized behavioral chamber was designed in SolidWorks 3D computer-aided design software and manufactured in a machine shop at California Institute of Technology. To obtain depth images of animals, a depth sensor (Creative Sens3D depth and gesture camera) was installed on top of the cage. The Creative Sens3D depth and gesture camera is a time-of-flight camera that resolves distance based on the speed of light, by emitting infrared light pulses and measuring the time-of-flight of the light signal between the camera and the subject for each point of image. Because mice are nocturnal animals but are insensitive to red light, animal behaviors are recorded under red light illumination (627 nm). Although the depth sensor itself

contains an RGB color camera, it does not perform well under red light illumination; therefore, we installed a monochrome video camera (Point Gray Grasshopper3 2.3 MP Mono USB 3.0 IMX174 CMOS camera) on top of the cage, next to the depth sensor. The top-view camera and depth sensor were placed as close as possible to each other, to minimize parallax, and were aligned perpendicular to the bottom of the cage. To obtain side-view information, an additional monochrome video camera of the same model was installed in front of the cage.

Video Acquisition and Camera Registration. Frames from the two monochrome video cameras were recorded by StreamPix 6.0 from NorPix Inc. Similar software did not exist for the Sens3D camera, and therefore we developed customized C# software to stream and record raw depth data from the depth sensor. All three devices were synchronized by customized MATLAB scripts. Frames of each stream and their corresponding timestamps were recorded at 30 frames per second and saved into an open source image sequence format (SEQ) that can be accessed later in MATLAB using a MATLAB Computer Vision toolbox (vision.ucsd.edu/~pdollar/toolbox/doc).

To integrate the monochrome information from the top-view camera with the depth information from the depth sensor, we registered the top-view camera and the depth sensor into a common coordinate frame using the Stereo Calibration and Scene Reconstruction tools from MATLAB's Computer Vision System Toolbox (included in version R2014b or later), in which a calibration pattern (a planar checkerboard) was used to fit a parameterized model of each camera (Fig. S3 A–D). We then projected the top-view video frames into the coordinates of the depth sensor (Fig. S3 E–H) to obtain simultaneous depth and monochrome intensity values for each pixel.

Animal Detection and Tracking. We performed background subtraction and image segmentation using data from top-view and depth sensor to determine the location and identity of the two animals (e.g., resident and intruder in the resident-intruder assay or two individuals in the reciprocal social interaction assay). Specifically, we first determined the rough locations of animals in each frame using image segmentation based on the depth information. The 3D depth of the unoccupied regions of the mouse cage was stitched together from multiple frames to form the depth background of the entire cage. This background was then subtracted from the movie to remove objects in the mouse's home environment, such as the water dispenser. We then performed a second round of finer-scale location tracking, in which the background-subtracted depth images were segmented to determine the potential boundary of the animals, and the identities of the animals were determined by their fur colors (black vs. white) using data from the monochrome camera. Each segmented animal in each frame was fit with an ellipse, parameterized by the centroid, orientation, and major- and minor-axis length. Body orientation was determined using an automated algorithm that incorporated data from the animal's movement velocity and rotation velocity and from feature outputs produced by a previously developed algorithm (24, 27).

To evaluate the performance of the automated estimation of animal poses, we constructed a ground truth data set of 634 manually annotated ellipses and calculated the frame-wise differences in fit and measured centroid position, length of the major axis (estimated body length), heading orientation, and head position (Fig. 2D):

$$d(\theta_1, \theta_2) = \frac{|\theta_1 - \theta_2|}{l},$$

where θ_n denotes the value of the pose parameter (centroid position, estimated body length, or estimated head position) with $n \in \{1$ for measured value, 2 for fit value), and l is the measured length of the major axis.

We also calculated framewise differences in fit and measured heading orientation:

$$d(\theta_1, \theta_2) = \lfloor \text{mod}(\theta_1 - \theta_2, 360) \rfloor,$$

where $\lfloor x \rfloor = \min(x, 360 - x)$ and θ_n denotes the value of the heading orientation with $n \in \{1$ for measured value, 2 for fit value}.

We evaluated the overall performance of individual frames by computing the framewise weighted differences between the machine-annotated and manually annotated ellipses using a previously developed metric (27),

$$d(\theta_1, \theta_2) = \sqrt{\frac{1}{D} \sum_{i=1}^D \frac{1}{\sigma_i^2} (\theta_1^i - \theta_2^i)^2},$$

where D is the number of pose parameters, σ_i^2 denotes the variance of the differences between human annotations of the i th pose parameter, and θ_n^i denotes the value of the i th pose parameter with $n \in \{1$ for measured value, 2 for fit value}.

We also compared the differences between two independent human observers (Fig. S5) and determined the 98% percentile, which was used as the threshold for evaluating the performance of the prediction.

Using the five fit ellipse parameters and additional data from the depth sensor, we developed a set of 16 second-order features describing the state of each animal in each video frame, and 11 “window” features computed over multiple frames, giving 27 total features. These features are described in [Supporting Information](#) and [Fig. S7](#).

Supervised Learning. In supervised learning, classifiers are trained using datasets that have been manually annotated with the desired classifier output, to construct a function that best reproduces these manual annotations. We used the 27 extracted features to test several supervised learning algorithms, including SVM, adaptive boosting (adaBoost), and random decision forests (TreeBagger). The random decision forest gave us the best performance in prediction accuracy and training speed and was thus selected for further investigation. We trained three social behavior classifiers (attack, mounting, and close investigation) using a set of six videos of male–male and male–female interactions, in which a total of ~150,000 frames were manually annotated on a frame-by-frame basis.

We then trained an ensemble of 200 random classification trees using the TreeBagger algorithm in MATLAB; output of the classifier was taken as the mode of the bagged trees. We chose to use 200 trees for classification because this number was well beyond where error rate plateaued; because trees in the ensemble can be trained in parallel, increasing the size of the population was not computationally expensive.

To measure the accuracy of the three decision forest classifiers in replicating human annotations, we manually labeled a different set of 14 videos from a variety of experimental conditions that contained ~350,000 frames total, and used them as our test set. We evaluated the performance using the DET curve representing false negative rate $FNR = FN / FN + TP$ and the false positive rate $FPR = FP / FP + TN$, as well as the precision-recall curve representing the true positive rate $TPR = TP / TP + FN$ vs. the positive predictive value $PPV = TP / TP + FP$. We also measured the accuracy by computing the fraction of true positive and true negative in all classes $ACC = TP + TN / TP + TN + FN + FP$, where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

Boutwise precision was defined as

$$precision_{bout} = \frac{\sum_{i \in TP} T_i}{\sum_{i \in TP} T_i + \sum_{i \in FP} T_i}$$

where $\sum_{i \in TP} T_i$ is the total time of the true-positive bouts and $\sum_{i \in FP} T_i$ is the total time of the false-positive bouts. Here a true-positive bout was the classified bout in which >30% of its frames were present in the ground truth; a false-positive bout was the classified bout in which $\leq 30\%$ of its frames were present in the ground truth.

Boutwise recall was defined as

$$recall_{bout} = \frac{\sum_{i \in TP} T_i}{\sum_{i \in TP} T_i + \sum_{i \in FN} T_i}$$

where $\sum_{i \in TP} T_i$ is the total time of the true-positive bouts and $\sum_{i \in FN} T_i$ is the total time of the false-negative bouts. Here a true-positive bout was the ground truth bout in which >30% of its frames were present in the classification; a false-negative bout was the ground truth bout in which $\leq 30\%$ of its frames were present in the classification.

Animal Rearing. Experimental subjects were 10-wk-old wild-type C57BL/6N (Charles River Laboratory), NZB/B1NJ (Jackson Laboratory), and BTBR T+tf/J (Jackson Laboratory). In the resident-intruder assay to examine attack, social

investigation, and mounting behaviors (Figs. 4 and 5), intruder mice were BALB/c males and females, purchased at 10 wk old (Charles River Laboratory). In the reciprocal social interaction assay to examine social investigation (Fig. 6), interaction partners were BALB/c males, purchased at 10 wk old (Charles River Laboratory). The intruder male was gonadally intact and the intruder females were randomly selected. Animals were housed and maintained on a reversed 12-h light-dark cycle for at least 1 wk before behavioral testing. Care and experimental manipulations of animals were in accordance with the NIH *Guide for the Care and Use of Laboratory Animals* and approved by the Caltech Institutional Animal Care and Use Committee.

Behavioral Assays. The resident-intruder assay was used to examine aggression, mounting, and close investigation of a resident mouse in its home cage (Figs. 4 and 5). Resident males in their home cages were transferred to a behavioral testing room containing a customized behavioral chamber equipped with video acquisition capabilities (described in *Video Acquisition and Camera Registration*). An unfamiliar male or female (“intruder”) mouse was then introduced into the home cage of the tested resident. The resident and intruder were allowed to interact with each other freely for 15–30 min before the intruder was removed. If excessive tissue damage was observed due to fighting, the interaction was terminated prematurely.

A reciprocal social interaction assay was used to examine social investigation between two interacting animals in an unfamiliar, neutral cage (Fig. 6). The procedure followed was similar to the resident-intruder assay, except that both interacting individuals were introduced to an unfamiliar, neutral cage and were allowed to interact with each other for 15 min before they were removed.

Manual Annotation of Behaviors. Two synchronized videos were scored manually on a frame-by-frame basis using a Computer Vision MATLAB toolbox (vision.ucsd.edu/~pdollar/toolbox/doc/). The human observer performing annotation was blind to experimental conditions. In manual scoring, each frame was annotated as corresponding to aggression, mounting, or close-investigation behavior. Aggression was defined as one animal engaging in biting and tussling toward another animal; mounting was defined as one animal grabbing on to the back of another animal and moving its arms or its lower body; close investigation was defined as the head of one animal closely investigating any body parts of another animal within $1/4 \sim 1/3$ body length.

Statistical Analysis. Statistical analysis was performed using MATLAB (MathWorks). The data were analyzed using two-sample *t* test, two-sample Kolmogorov–Smirnov test, and Mann–Whitney *U* test.

Note Added in Proof. The depth-sensing camera used in this paper, Sens3D, was recently discontinued by Intel Corp. and Creative, Inc. However, an alternative device, DepthSense 325, is being sold by SoftKinetics, Inc. Sens3D and DepthSense 325 are identical products, except for their product packaging and branding. They offer identical functionalities and are supported by an identical software development kit (SDK).

ACKNOWLEDGMENTS. We thank Xiao Wang and Xiaolin Da for manual video annotation; Michele Damian, Louise Naud, and Robert Robertson for assistance with coding; Allan Wong for helpful suggestions; Prof. Sandeep R. Datta (Harvard University) for sharing unpublished data; Celine Chiu for laboratory management; and Gina Mancuso for administrative assistance. This work was supported by grants from the Moore and Simons Foundations and postdoctoral support from the Helen Hay Whitney Foundation (W.H.), the National Science Foundation (M.Z.) and Sloan-Swartz Foundation (A.K.). D.J.A. is an Investigator of the Howard Hughes Medical Institute.

- Silverman JL, Yang M, Lord C, Crawley JN (2010) Behavioural phenotyping assays for mouse models of autism. *Nat Rev Neurosci* 11(7):490–502.
- Anderson DJ, Perona P (2014) Toward a science of computational ethology. *Neuron* 84(1):18–31.
- Gomez-Marin A, Paton JJ, Kampff AR, Costa RM, Mainen ZF (2014) Big behavioral data: Psychology, ethology and the foundations of neuroscience. *Nat Neurosci* 17(11):1455–1462.
- Spink AJ, Tegelenbosch RA, Buma MO, Noldus LP (2001) The EthoVision video tracking system—A tool for behavioral phenotyping of transgenic mice. *Physiol Behav* 73(5):731–744.
- Noldus LP, Spink AJ, Tegelenbosch RA (2001) EthoVision: A versatile video tracking system for automation of behavioral experiments. *Behav Res Methods Instrum Comp* 33(3):398–414.
- Ou-Yang T-H, Tsai M-L, Yen C-T, Lin T-T (2011) An infrared range camera-based approach for three-dimensional locomotion tracking and pose reconstruction in a rodent. *J Neurosci Methods* 201(1):116–123.
- Post AM, et al. (2011) Gene-environment interaction influences anxiety-like behavior in ethologically based mouse models. *Behav Brain Res* 218(1):99–105.
- Walf AA, Frye CA (2007) The use of the elevated plus maze as an assay of anxiety-related behavior in rodents. *Nat Protoc* 2(2):322–328.
- Pham J, Cabrera SM, Sanchis-Segura C, Wood MA (2009) Automated scoring of fear-related behavior using EthoVision software. *J Neurosci Methods* 178(2):323–326.
- Cai H, Haubensack W, Anthony TE, Anderson DJ (2014) Central amygdala PKC- δ (+) neurons mediate the influence of multiple anorexigenic signals. *Nat Neurosci* 17(9):1240–1248.
- Tzschentke TM (2007) Measuring reward with the conditioned place preference (CPP) paradigm: update of the last decade. *Addict Biol* 12(3–4):227–462.
- Nestler EJ, Hyman SE (2010) Animal models of neuropsychiatric disorders. *Nat Neurosci* 13(10):1161–1169.
- Dankert H, Wang L, Hoopfer ED, Anderson DJ, Perona P (2009) Automated monitoring and analysis of social behavior in *Drosophila*. *Nat Methods* 6(4):297–303.

14. Eyjolfsson E, Branson S, Burgos-Artizzu XP (2014) Detecting social actions of fruit flies. *Computer Vision—ECCV*. Lecture Notes in Computer Science (Springer, Cham, Switzerland), Vol 8690, pp 772–787.
15. Kabra M, Robie AA, Rivera-Alba M, Branson S, Branson K (2013) JAABA: Interactive machine learning for automatic annotation of animal behavior. *Nat Methods* 10(1): 64–67.
16. Branson K, Robie AA, Bender J, Perona P, Dickinson MH (2009) High-throughput ethomics in large groups of *Drosophila*. *Nat Methods* 6(6):451–457.
17. Tsai H-Y, Huang Y-W (2012) Image tracking study on courtship behavior of *Drosophila*. *PLoS One* 7(4):e34784.
18. Iyengar A, Imoehl J, Ueda A, Nirschl J, Wu C-F (2012) Automated quantification of locomotion, social interaction, and mate preference in *Drosophila* mutants. *J Neurogenet* 26(3-4):306–316.
19. Gomez-Marin A, Partoune N, Stephens GJ, Louis M, Brembs B (2012) Automated tracking of animal posture and movement during exploration and sensory orientation behaviors. *PLoS One* 7(8):e41642.
20. Kohlhoff KJ, et al. (2011) The iFly tracking system for an automated locomotor and behavioural analysis of *Drosophila melanogaster*. *Integr Biol (Camb)* 3(7):755–760.
21. Fontaine El, Zabala F, Dickinson MH, Burdick JW (2009) Wing and body motion during flight initiation in *Drosophila* revealed by automated visual tracking. *J Exp Biol* 212(Pt 9):1307–1323.
22. Card G, Dickinson M (2008) Performance trade-offs in the flight initiation of *Drosophila*. *J Exp Biol* 211(Pt 3):341–353.
23. Wolf FW, Rodan AR, Tsai LT-Y, Heberlein U (2002) High-resolution analysis of ethanol-induced locomotor stimulation in *Drosophila*. *J Neurosci* 22(24):11035–11044.
24. Burgos-Artizzu XP, Perona P, Dollár P (2013) Robust face landmark estimation under occlusion. *Proceedings of the 2013 IEEE International Conference on Computer Vision (IEEE, Washington, DC)*, pp 1513–1520.
25. de Chaumont F, et al. (2012) Computerized video analysis of social interactions in mice. *Nat Methods* 9(4):410–417.
26. Burgos-Artizzu XP, Dollár P, Lin D, Anderson DJ, Perona P (2012) Social behavior recognition in continuous video. *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (IEEE, Washington, DC)*, pp 1322–1329.
27. Dollár P, Welinder P, Perona P (2010) Cascaded pose regression. *Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (IEEE, Washington, DC)*, pp 1078–1085.
28. Ohayon S, Avni O, Taylor AL, Perona P, Roian Egnor SE (2013) Automated multi-day tracking of marked mice for the analysis of social behaviour. *J Neurosci Methods* 219(1):10–19.
29. Weissbrod A, et al. (2013) Automated long-term tracking and social behavioural phenotyping of animal colonies within a semi-natural environment. *Nat Commun* 4:2018.
30. Shemesh Y, et al. (2013) High-order social interactions in groups of mice. *eLife* 2:e00759.
31. Matsumoto J, et al. (2013) A 3D-video-based computerized analysis of social and sexual interactions in rats. *PLoS One* 8(10):e78460.
32. Iddan GJ, Yahav G (2001) Three-dimensional imaging in the studio and elsewhere. *Photonics West 2001: Electronic Imaging*, eds Corner BD, Nurre JH, Pargas RP (SPIE, Bellingham, WA), pp 48–55.
33. Lyons DM, MacDonall JS, Cunningham KM (2014) A Kinect-based system for automatic recording of some pigeon behaviors. *Behav Res Methods*, 10.3758/s13428-014-0531-6.
34. Kulikov VA, et al. (2014) Application of 3-D imaging sensor for tracking minipigs in the open field test. *J Neurosci Methods* 235:219–225.
35. Gonçalves N, Rodrigues JL, Costa S, Soares F (2012) Preliminary study on determining stereotypical motor movements. *Proceedings of the 2012 International Conference of the IEEE Engineering in Medicine and Biology Society (Piscataway, NJ)*, pp1598–1601.
36. Guillot PV, Chapouthier G (1996) Intermale aggression and dark/light preference in ten inbred mouse strains. *Behav Brain Res* 77(1-2):211–213.
37. Ellegood J, Crawley JN (2015) Behavioral and neuroanatomical phenotypes in mouse models of autism. *Neurotherapeutics* 12(3):521–533.
38. McFarlane HG, et al. (2008) Autism-like behavioral phenotypes in BTBR T+tf/J mice. *Genes Brain Behav* 7(2):152–163.
39. Bolivar VJ, Walters SR, Phoenix JL (2007) Assessing autism-like behavior in mice: Variations in social interactions among inbred strains. *Behav Brain Res* 176(1):21–26.
40. Luo L, Callaway EM, Svoboda K (2008) Genetic dissection of neural circuits. *Neuron* 57(5):634–660.
41. Deisseroth K, Schnitzer MJ (2013) Engineering approaches to illuminating brain structure and dynamics. *Neuron* 80(3):568–577.
42. Zhang F, Aravanis AM, Adamantidis A, de Lecea L, Deisseroth K (2007) Circuit-breakers: optical technologies for probing neural signals and systems. *Nat Rev Neurosci* 8(8):577–581.
43. Desland FA, Afzal A, Warraich Z, Mocco J (2014) Manual versus automated rodent behavioral assessment: Comparing efficacy and ease of bederson and garcia neurological deficit scores to an open field video-tracking system. *J Cent Nerv Syst Dis* 6:7–14.
44. Lin D, et al. (2011) Functional identification of an aggression locus in the mouse hypothalamus. *Nature* 470(7333):221–226.
45. Button KS, et al. (2013) Power failure: Why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 14(5):365–376.
46. Chen JL, Andermann ML, Keck T, Xu N-L, Ziv Y (2013) Imaging neuronal populations in behaving rodents: Paradigms for studying neural circuits underlying behavior in the mammalian cortex. *J Neurosci* 33(45):17631–17640.