Ultrasonic signals associated with different types of social behavior of mice.

Daniel T. Sangiamo[1,2], **Megan R. Warren**[1], **Joshua P. Neunuebel**[1,*]

[1]Department of Psychological and Brain Sciences, University of Delaware, Newark, Delaware, USA

[2]Neuroscience Program, University of Illinois, Urbana-Champaign, Illinois, USA

Abstract

Communication plays an integral role in human social dynamics and is impaired in several neurodevelopmental disorders. Mice are used to study the neurobiology of social behavior; however, the extent to which mouse vocalizations influence social dynamics has remained elusive because it is difficult to identify the vocalizing animal among mice involved in a group interaction. By tracking the ultrasonic vocal behavior of individual mice and using an algorithm developed to group phonically similar signals, we show that distinct patterns of vocalization emerge as male mice perform specific social actions. Mice dominating other mice were more likely to emit different vocal signals than mice avoiding social interactions. Furthermore, we show that the patterns of vocal expression influence the behavior of the socially-engaged partner but no other animals in the cage. These findings clarify the function of mouse communication by revealing a communicative ultrasonic signaling repertoire.

Summary Sentence:

Sound source localization system unmasks function of social vocalizations by revealing behavior-dependent vocal emission.

Introduction:

Throughout the animal kingdom, social behavior is fundamental for survival and reproduction[1, 2]. Social behavior consists of dynamic, complex interactions between animals and can promote cooperation or competition within a group[3]. Moreover, a diverse set of

signals influences the interactions between individuals within a group[3, 4]. Vocalizations, in particular, play a substantial role in group dynamics, by warning group members of specific predators[5] or indicating that reproductive opportunities exist[6]. When vocalizations are detected by a conspecific, the auditory system encodes the signals[7], and the information then helps guide the behavioral response of the animal[8]. The animal's behavioral response creates a new set of social communication signals that can influence the dynamics of an interacting partner[9]. This continuous exchange of social information produces a feedback loop that depends on the perception of sensory cues.

Significant advances in understanding the neurobiology of social behavior as well as complex social network dynamics have been made using mice as a model system[10–12]. Mice (*Mus musculus*) are social animals that engage in diverse behaviors[13] accompanied by ultrasonic vocalizations (USVs)[14]. USVs are auditory signals spanning a range of 35 to 110 kHz[15]. Mice readily approach and investigate the source of USVs[16], and female mice prefer to spend time near a vocalizing male than a mute male[17]. Together, the measures used to quantify mouse behavior imply that USVs are relevant to the dynamics underlying social interactions.

USVs show a diverse acoustic profile, suggesting that multiple categories or types of vocalizations exist[18, 19]. The types of USVs that mice produce vary with recording conditions or context[20, 21], genetics[22], and an animal's developmental stage[23]. Although several variables appear to influence the types of vocalizations produced, the extent to which mouse vocalizations influence social dynamics is unknown. This uncertainty exists because of the challenges involved in identifying the source of a vocalization when animals socially interact. If vocalizations from individual mice could be identified while the animals engage in distinct social behaviors, then a relationship between vocalizations and behavior may emerge; this could help elucidate the role communication plays in social dynamics, thus paving the way to a mechanistic understanding of the neural basis of social behavior.

To determine how mouse vocalizations affect social dynamics, we have to overcome several technological barriers. First, it is essential to accurately identify which mouse was vocalizing during group interactions. Because mouse ultrasonic vocalizations are imperceptible by the human auditory system and lack distinct visual cues that might facilitate our ability to identify the vocalizer[20], tracking the vocal repertoires of specific, socially engaged mice is a formidable task. Second, categorizing different types of vocalizations requires a systematic, unbiased approach. Third, a high-throughput method for automatically detecting the behaviors of multiple animals is needed to provide critical information about an animal's actions. Once we can identify the vocalization types emitted by mice engaged in specific acts, then we can examine patterns in social communication and begin to determine the role these signals play in social dynamics.

Here, by combining a sound source localization system to track the vocal activity of individual adult mice with a machine-learning algorithm to automatically detect specific social behaviors and with a vocalization-clustering program to group similarly shaped vocal signals, we were able to associate different types of ultrasonic vocalizations with distinct behaviors of the vocalizing mouse. We show that specific patterns of vocalization influence

the behavior of only the socially-engaged partner. Our results demonstrate the utility of mouse communication by unmasking a functional correlate of ultrasonic signaling and provide a basis for deciphering the neural underpinnings of social behavior.

Results:

Quantification of Behavioral Repertoires

We examined the relationship between vocal expression and behavior in 11 groups of adult mice, each consisting of two females in estrus and two males. Mice in each group freely interacted for five hours in a competitive environment known to promote complex vocal and social interactions[24]. A custom-built anechoic chamber containing a camera and 8-channel microphone array was used to record movements and vocalizations simultaneously. Freely moving animals were automatically tracked[25], and then a machine learning-based system[26] was applied to create classifiers for extracting behaviors of male mice (Fig. 1A; Supplementary: Videos 1–6, Table 1), specifically, walking, circling another male [mutual circle], fighting, male chasing female, fleeing, and male chasing male.

Every male performed each behavior; however, the characteristics of each behavior differed among males (Fig. 1B, Supplementary Fig. 1). In general, the number and duration of each of the behaviors differed significantly between male mice (Fig. 1C,D). In some social interactions, the two males each have the same role (i.e. during fight and during mutual circle), whereas in other interactions they have distinct roles (chaser and chased, fled from and flee). When accounting for the role a mouse played in a behavior, the velocity of the males was significantly different across all behaviors, except for fleeing and being chased, which had similar velocities (Fig. 1E).

The frequencies with which individuals performed behaviors varied as a function of time (Supplementary Fig. 2). Males circled each other, chased females, or walked more frequently early in the experiment, but this decreased over time. In contrast, the frequency in which males fought with, fled from, and chased other males was constant over time. As the experiment progressed, the duration of each walk and female-chase decreased, whereas the durations of each fight, circle, male-chase, and flee remained constant (Supplementary Fig. 3). Over time, the velocity of the males increased when fighting and chasing other males, but not when walking, circling, fleeing, chasing females (Supplementary Fig. 4).

Localization of Sound

The location where each vocal signal originated was triangulated using a microphone array and then probabilistically assigned to individual animals[24, 27]. Fig. 2 shows examples. Of the 211,090 detected signals, 135,078 were assigned to individual mice. Other than unassigned signals occurring more frequently when two mice were in close proximity, no apparent temporal or spatial biases were introduced by the system (Supplementary Fig. 5). Of the assigned signals, males produced 82.5%. Although prior studies have shown that mice vocalize in different contexts[20, 21], it is unknown how mice vocalize while performing specific behaviors. To answer this question, we first examined the vocal rate (i.e., the number of vocal signals emitted during a behavior) of each male while engaged in particular

acts. Notably, we showed that vocal rates differed significantly between behaviors (Supplementary Fig. 6, Table 2), except between walking and fleeing from another male, which had similar vocal rates. When examining only the behaviors with vocal emission, we found that the vocal rate of fighting males decreased over time (Supplementary Fig. 7), whereas the vocal rate during the other behaviors remained constant over time. These results demonstrate the variability in vocal rate during specific behaviors as animals freely interact.

Determining Vocal Repertoires

The acoustic structure of mouse vocalizations is quite diverse[18, 19]. To capture this diversity, we developed a clustering approach that automatically groups vocal signals based on their acoustic structure (Supplementary Fig. 8). Grouping phonetically similar signals resulted in 22 categories in total (Fig. 3; Table 3). Two permutation tests independently showed that the algorithm successfully clustered acoustically similar signals: first, we calculated the dissimilarity of signals within a group and between groups (Supplementary Fig. 9), which showed that signals within a group were always more similar to each other than signals in different groups (permutation test, all p<0.001); second, we compared the variation in the acoustic structure within each group to the same number of randomly selected signals. The acoustic structure of signals within a group was significantly less variable than expected by chance (permutation test, all p<0.001), indicating that the clustering technique successfully categorized signal types. The acoustic structure of each signal type for each male mouse was similar (Supplementary Fig. 10). Categorizing vocal signals enabled us to examine patterns in vocal emission across male mice, which revealed marked diversity in the signal types used by each mouse (Fig. 4A). The duration, interval (frequency range), and amplitude (intensity) of signals differed between categories (Fig. 4B–D). Interestingly, when comparing how frequently each male produced each signal type, we found that there were significant differences in expression patterns (Supplementary Fig. 11). These findings highlight the variability in the shape, duration, interval, and amplitude of vocal signals of freely-interacting mice.

Behaviorally-Dependent Vocal Emission

Because vocal emission and behavior vary across individual males, we reasoned that distinct signals may correlate with specific behaviors. Indeed, a pattern emerged when we assessed examples from individual males (Fig. 5) and pooled across males (Supplementary: Fig. 12,13). Vocal signals that decreased in pitch (types 12–22) were typically affiliated with dominant behaviors (i.e., courting and aggressive interactions), whereas vocal signals that increased in pitch (types 1–7) were frequently emitted by males engaging in non-dominant behaviors (i.e., avoiding other animals). Other signal types (8–11) occurred more uniformly. An unbiased search for structure independent of clustering revealed that the changes in pitch for signals emitted during dominant and non-dominant behaviors were significantly different (dominant: slope = −0.152±0.007 kHz/ms, average±SEM; non-dominant: slope = 0.278±0.002 kHz/ms; 2-sample t-test, T=71.7, p<10^{-10}). These patterns suggest that mice are more likely to emit specific types of vocal signals during particular behaviors.

To quantify the dependence of vocal expression on social context, we calculated the proportion of signal types produced during each behavior (Fig. 6A). There was a significant interaction between vocal signal type and behavior (2-way ANOVA, F=67.6, p<10^{-8}), indicating that vocal expression varies depending on the behavior. We determined chance levels of context-dependent signaling using a permutation test, where each signal was randomly assigned a signal type while maintaining the number of signals in each category, the identity of the vocalizer, and the social context. Comparing the proportions of signals emitted during distinct actions to the randomly generated distributions showed that emission of specific signals occurred above chance when mice were engaged in dominant behaviors. In contrast, other signals were emitted above chance by mice performing acts associated with social avoidance. An inverse relationship existed such that signals emitted above chance in a context classified as either dominant or non-dominant often were emitted below chance during the other context. Furthermore, when we calculated proportion as the percentage of behaviors in which each type of vocal signal was emitted — thus removing any contribution of vocal repetition within a behavior — we observed similar results (Supplementary Fig. 14), consistent with the notion that vocal expression depends on the behavior of the individual. Importantly, the proportions of assigned and unassigned signals in each behavior were similar (Supplementary Fig. 15), suggesting that assigned signals accurately captured the relationship between vocal signals and behavior.

Because many auditory features describe an animal's vocal repertoire, other attributes of vocalizations may contribute to their social meaning. Therefore, we additionally grouped signals by duration, interval, and amplitude; this yielded 22 groups, with group boundaries determined using multiple approaches (Supplementary Fig. 16; see online methods). Here we observed no consistent link between groups and behaviors. (Fig. 6B–D, Supplemental Fig. 17). These results suggest that the shapes of vocal signals emitted depend on the behavior of the male mice, rather than the duration, interval, or amplitude of a signal.

The preceding analyses combined data from all 22 males; therefore, the results might be an artifact of pooling the data. While the examples in Supplementary Fig. 18 suggests that that individual mice are more likely to emit specific types of vocal signals during particular behaviors, a direct quantification is needed. We therefore calculated the proportion of each vocal signal type emitted during each behavioral event for each mouse, which puts equal weight on each animal's contribution. Here, we observed the same pattern of behavior-dependent vocal expression at the individual mouse level as that seen in the pooled data (Supplementary Fig. 19A, Supplementary Video 7).

Over the course of 5 hours, there were times when a mouse aggressively pursued the other male. At other times, the same mouse was pursued by the other male. Interestingly, when a mouse vocalized while in pursuit, the signals increased in pitch, whereas when the same mouse vocalized while being chased, the signals decreased in pitch (Supplementary Fig. 19B–C). To quantify differences in vocal expression as an animal was acting dominantly and non-dominantly, we computed vocal chase indices for each animal and all vocal signal types. For each signal type, we calculated indices by taking the difference between the proportion produced during dominant and non-dominant behaviors and then dividing by the sum. Positive and negative values indicate a higher proportion of signals emitted during dominant

and non-dominant behaviors, respectively. Interestingly, signals that increased in pitch were associated with a negative index, whereas signals that decreased in pitch were associated with a positive index (Supplementary Fig. 19D). The majority of the post-hoc comparisons revealed that the indices for vocal signals with positive and negative slopes were significantly different (Supplementary Fig. 19E). These results verify that the dependence of vocal expression on behavior was observed in each of the male mice.

While our unbiased clustering method revealed a novel relationship between vocal expression and behavior, we had to rule out the possibility that over-clustering may have undermined the results. That is, distributing more categories of signal types across a finite number of behaviors may have reduced the proportion of signals emitted during specific behaviors and, consequently, affect the calculation of chance levels. To test this possibility, we reduced the number of signal types by 50 percent (i.e., the progressive k-means clustering algorithm was terminated after partitioning vocal signals into 11 categories). As expected, similarly shaped signals were merged (Supplementary Fig. 20A,B). As seen previously, vocal signals increasing in pitch occurred above chance during non-dominant behaviors and below chance in dominant behaviors and the opposite was the case for negatively sloped signals (Supplementary Fig. 20C). We also employed a deep-learning method to categorize vocal signals ("DeepSqueak"[19]), which identified only six types of signals (Supplementary Fig. 20D,E) and yielded the same the pattern of behavior-dependent vocal expression (Supplementary Fig. 20F). This evidence confirms that specific vocal signals are more likely to be produced by males during distinct behaviors, regardless of the clustering approach and number of categories of vocalizations.

Decoding Behavior Based on Vocal Emission

To identify the extent that vocalization conveys information about an animal's behavior, we trained a multi-class support vector machine (mcSVM) to discriminate behaviors based on the proportion of signal types emitted during randomly selected behavioral events (Fig. 7A,B). After training, we assessed how accurately the classifier could predict behaviors from data excluded from training. The process of training and testing was repeated 1000 times to generate a 95% confidence interval of classifier accuracies (range: 14.7–39.0). With a chance performance of 12.5% (for eight behaviors), the accuracy of each classifier was substantially above chance (Fig. 7C). By contrast, classifiers trained to discriminate behaviors based on duration, interval, or amplitude did not perform above chance (Fig. 7D–F). On average, classifiers trained with the types of signals emitted could accurately predict each category of behavior above chance, with the highest accuracy for predicting walking and being chased (Fig. 7G). As shown in the confusion matrix (Fig. 7K), some specific classification errors were more likely to occur than others. In particular, misclassified non-dominant behaviors were more often classified as other non-dominant behaviors, and incorrectly categorized dominant behaviors were more often classified as other dominant behaviors. Even when the number of vocal signal types was reduced, mcSVMs were able to predict behavior accurately (Supplementary Fig. 21). Classifiers trained using the duration, interval, or amplitude of vocal signals had variable performance, and were mostly unable to predict behavior category better than chance (Fig. 7H–J). Moreover, no clear patterns in misclassification emerged (Fig. 7L–N). While mcSVMs are useful tools for extracting

predictive information, calculating a mutual information (MI) score is a more direct approach for quantifying the behavioral information conveyed by vocal emission[6]. MI scores revealed that significant behavioral information was transmitted by vocal emission (MI: 0.07 bits; permutation test, $z=348.5$, $p<10^{-4}$). These results strongly suggest that the types of vocal signals emitted provide information about an animal's behavior.

Behavioral Responses to Vocal Emission

Communication modulates the decision making of animals that receive signals[28]. To test the impact of behavior-dependent vocal signals on the behavior of a conspecific, we dissected components of behaviors in which each male played a distinct role (chases and flees) on a sub-second time scale. For every frame of a behavior, we calculated the velocity of each of the four mice. The locomotion speeds of the sender (vocalizing animal) and receivers (other three animals) were compared before and after vocal emission. Receiver 1 was always the engaged social partner, and the other two receivers were determined by the distance from the sender at the time of vocal emission. Sender-emitted signals that decreased in pitch and occurred above chance were classified as dominant signals, whereas sender-emitted signals that increased in pitch and occurred below chance were classified as non-dominant signals. During male–male chases (Fig. 8A), when the chasing male emitted dominant signals, the average speed of the male being chased peaked (Fig. 8B). This was not observed for females (i.e. receivers 2 and 3) (Fig. 8C,D) nor for the sender (Fig. 8E), which showed a delayed peak in speed. When the chasing male emitted non-dominant signals, no such patterns were observed. The changes in behavior were immediate, as no consistent patterns emerged during the five seconds after vocal emission (Fig. 8F–I). Note that the negative results may be confounded by the fact that animals could be engaged in any number of behaviors during this time.

To quantify changes in velocity, we computed acceleration before and after emission of dominant and non-dominant vocal signals (Fig. 8J–Q). Chased males accelerated before the emission of dominant signals and decelerated immediately after emission; the changes in speed were significantly different. Neither of the females nor the sender showed a similar pattern, although in some conditions these mice significantly decelerated (Fig. 8K) or accelerated (Fig. 8M). There were no significant differences in acceleration before and after emission of non-dominant signals produce by the chasing males (Fig. 8J–M) or males being fled from (Supplementary Fig. 22A–Q). When males chased females, there were significant differences in acceleration before and after vocalization only when dominant signals were emitted by the chasing males (Supplementary Fig. 23A–Q). Together, these results suggest that dominant signals were directed towards specific individuals and that the auditory information in the signal modulates the social partner's behavior.

To determine if dominant signal emission were associated with changes in behavior, we conducted a permutation test (Fig. 8R–U). Vocal signals emitted by the chasing male were randomly shifted to different times within chases, pre- and post-accelerations were computed for the four mice, and t-scores were calculated to assess pre- and post-acceleration differences. This procedure was repeated 1,000 times, and the distributions of t-scores were compared to the t-scores calculated when dominant or non-dominant vocal signals were

emitted. The change in pre- and post-acceleration for the vocalizing male and both female bystanders at the time of signal emission was not different than expected by chance. In contrast, the change in pre- and post-acceleration of the chased male at the time of signal emission was greater than chance. Moreover, changes in the pre- and post-acceleration of fleeing males (Supplementary Fig. 22R–U) and pursued females (Supplementary Fig. 23R–U) were greater than chance when dominant signals were emitted. These results suggest that specific behavior-dependent dominant vocal signals emitted by a male mouse shape behavioral responses of a social partner.

We next examined the role of vocal emission by the non-dominant male (chased or fleeing male). The speeds of the sender (non-dominant vocalizing animal) and receivers (other three animals) were compared before and after vocal emission. Sender emitted signals that increased in pitch and occurred above chance were classified as non-dominant signals, whereas sender emitted signals that decreased in pitch and occurred below chance were classified as dominant signals. During male–male chases (Supplementary Fig. 24) and flees (Supplementary Fig. 25), there were no noticeable behavioral changes associated with any vocal signals. The only subtle detectable change following vocal emission was a decrease in the acceleration of the females furthest from a chase. These results suggest that non-dominant vocal signals play a less discernible role in the dynamics of social behavior.

Discussion:

Communication is the transfer of information between two or more animals, often prompting changes in behavior[6]. In humans, deficits in communication characterize several neurodevelopmental disorders. Such disorders have a strong genetic basis, and mouse models with mutations in disorder-associated genes have been developed[29]. However, little is understood about vocal communication between mice. While USVs are assumed to function as communication signals, there has been no direct evidence linking adult USV emissions to distinct and immediate changes in behavior. This gap in knowledge was due to technological limitations in tracking vocalizations of individual mice as they interact with other animals. However, a tool for assigning USVs to individual animals was previously described[27, 30] and used in this study. Additionally, novel computational tools enabled us to automatically classify both vocal signals and social behavior of freely moving mice. We analyzed 111,498 male-assigned vocal signals and 32,352 examples of different behaviors, and showed the variability in vocal activity and behaviors of individual mice, similar to studies in other animals[31–35]. Further analysis revealed two prominent findings. First, the repertoire of vocal signals emitted by male mice is biased by specific behaviors. This discovery suggests that mouse vocal expression transmits information about an individual's actions. This provides an entry point for examining the relationship between social dynamics and an animal's emotional state, including in mouse models of neurodevelopmental disorders. Second, by showing that dominant vocal signals are associated with behavioral changes in only the engaged social partner, we provide evidence that in adult mice, USVs function as communication signals.

Our correlational and descriptive results support the notion that adult USVs are communication signals mediating behavioral changes; however, we made the assumption

that mice are capable of perceiving and discriminating vocal signal types. Multiple lines of evidence suggest that mice can indeed do this. First, the strain of mice used in the current study is capable of perceiving high-frequency sounds until ~18 months of age[36], which exceeds our oldest recorded animal's age by 12 months. Second, neural representations of specific USVs are present in the inferior colliculus[37] and auditory cortex[38, 39], suggesting a sensorineural basis for perceptibility and discriminability. Third, mice tested using operant conditioning procedures reliably discriminate between different types of vocalizations[40] that were similar in shape to the signal categories in the current study. Finally, female mice prefer vocalizing males over genetically engineered mute mice[41]. This evidence suggests that mice perceive and discriminate different types of USVs. The approach used in this study provides an avenue for identifying the neural underpinnings of decision making during innate social behavior.

Because social behavior in freely behaving and interacting animals is complex and can be difficult to interpret, many studies use standard tests to assess a specific aspect of social behavior. For instance, tube tests assess social dominance[42], three-chamber tests evaluate sociability[43], and discrimination tasks probe decision making[44]. In these approaches the range of measurable behaviors is limited, thus increasing reproducibility and throughput, but reducing the dimensionality and dynamics of the behavior. By contrast, studies of unrestrained natural behaviors yield larger, more precise, and high-dimensional observations that provide the level of detail necessary to generate a meaningful theoretical framework for elucidating the neural basis of behavior[45–47]. New technologies enabling the acquisition of datasets that are more naturalistic, together with sophisticated computational approaches for data analysis can lead to a more holistic, thorough understanding of animal behavior[45–47]. For instance, elegant work in which subtle differences in an animal's movement were extracted from large, diverse datasets has revealed the microstructure of locomotor repertoires[48, 49]. In the present study, the ability to track USVs and behavior of freely interacting mice, together with our newly developed computational tools, has yielded insight into the relationship between USVs and behavior in individual male mice. Nevertheless, our approach also has limitations. For example, to ensure a vocal emission was from a single animal, vocal signals were extracted as continuous sounds; however, this reduces the complexity of vocal signals in comparison to previously reported vocal repertoires[18]. Future studies that capture the full complexity of vocalizations of individual mice may reveal additional aspects of vocal behavior in different social contexts.

One current, long-standing hypothesis about the conserved properties of communication signals between species suggests that an animal's motivational state governs the types of sounds an animal emits while in close contact with conspecifics[50]. In particular, sounds that decrease in frequency reflect an increasing hostile motivation and sounds that increase in frequency reflect lower hostility or fear. Our results provide quantitative evidence that supports the notion of conserved rules linking an animal's motivation and the structure of vocal signaling. We showed that male mice that were fighting, being fled from, or pursuing other males were more likely to emit signals descending in pitch, whereas male mice in a non-aggressive role were more likely to emit signals ascending in pitch. However, our findings additionally suggest that the motivation–structural rules may need expanding: we showed that male mice are also more likely to produce negatively sloped vocal signals when

they are pursuing females, which is typically considered an affiliative behavior. This suggests that vocalizations of decreasing pitch may reflect an animal's current social status rather than only hostile or aggressive intentions.

In conclusion, within the complex, adaptable nature of innate behavior and vocal expression, our study uncovered a clear pattern linking particular social behaviors and vocal communication in male mice. These findings will provide a foundation for examining the neural circuits linking sensory cues - specifically USVs - to social behavior.

Materials and Methods:

Animals

Adult (13–21 weeks) males (n=22) and females (n = 22) of the B6.CAST-Cdh23Ahl+/Kjn strain (Jackson Laboratory; Bar Harbor, ME; stock number = 002756) were used because mice expressing the *Cdh23* gene (formally *Ahl*) are less likely to suffer from high-frequency hearing loss as they age[36]. No animals or data points were excluded from the analyses. Mice were kept on a 12-hour light/dark cycle (lights off at 7 p.m. EST or ECT) and had *ad libitum* access to food and water. Mice were weaned and genotyped at three weeks of age. Tail samples were sent to TransnetYX, Inc. (TransnetYX, Inc.; Cordova, TN) for genotyping. Only mice expressing *Cdh23* were used in experiments (homozygous and heterozygous animals). After weaning, mice were tagged with a light-activated microtransponder (PharmaSeq, Inc.; Monmouth Jct, NJ; p-Chip injector) that was implanted into their tail and then group-housed with same-sex siblings (3–5 mice per cage) until they were randomly selected to participate in experiments. All cages contained environmental enrichment. The University of Delaware Animal Care and Use Committee approved all experimental protocols, which complied with standards from the National Institutes of Health.

Experiment preparation

Mice were singly housed at least two weeks before the behavioral experiment to minimize group housing effects on the social behavior of the mice[51]. At least two days before the start of the experiment, the mice were marked with a unique back pattern to facilitate our ability to identify each mouse[25]. Back-patterns, created using harmless hair dye (Clairol Nice 'N Easy, Born Blonde Maxi), consisted of two vertical lines, two horizontal lines, one diagonal slash, or five dots. Each pattern was randomly assigned to one of the four mice. The day after painting, mice were exposed to a mouse of the opposite sex for 10 minutes. If animals attempted to copulate, the opposite sex exposure session was prematurely ended. Estrous stages of the female mice were determined two hours before the anticipated start of the experiment using non-invasive lavage and cytological assessment of vaginal cells[24]. Cells were collected by washing with 30 μl of saline, placed on a slide, stained with crystal violet, and examined under a light microscope. Estrous stage was determined based on the proportion of cell types observed. Proestrus consisted of mostly nucleated basal epithelial cells. In estrus, the majority of cells were cornified squamous epithelial cells that lacked a nucleus. During metestrus, cells were a mixture of neutrophils and cornified squamous epithelial cells. Diestrus consisted of mostly neutrophils. If both females were in estrus, the experiment was conducted at the start of the animals' dark cycle (7 pm). If the females were

not in estrus, they were tested every afternoon until both females were in estrus or until they exceeded 21 weeks of age.

Behavioral experiment

For each experiment, mice were group-housed (two males and two females) for five hours in a mesh-walled (McMaster-Carr; Robbinsville, NJ; 9318T25) cage. Two different cage shapes were used. One cage was a cuboid with a frame (width = 76.2 cm, length = 76.2 cm, height = 61.0 cm) made of extruded aluminum (8020, Inc.; Columbia City, IN). The other cage was a cylinder (height of 91.4 cm and a diameter of 68.6 cm). The walls of the cylinder were attached to a metal frame by securing the mesh between two pairs of plastic rings surrounding the arena, one pair at both the top and the bottom. Both the cuboid and cylindrical cages were surrounded by Sonex foam (Pinta Acoustic, Inc.; Minneapolis, MN; VLW-35). There were eight and three recordings conducted in the cuboid and cylindrical cages, respectively. Recordings were conducted in a custom-built anechoic chamber. Audio and video data were concurrently recorded for five hours using an 8-channel microphone array and camera. Audio data recorded by each microphone (Avisoft-Bioacoustics; Glienicke, Germany; CM16/CMPA40–5V) was sampled at 250,000 Hz (National Instruments; Austin, TX; PXIe-1073, PXIe-6356, BNC-2110), and then low-pass filtered (200 kHz; Krohn-Hite, Brockton, MA; Model 3384). Video data was captured with a camera (FLIR; Richmond, BC, Canada; GS3-U3-41C6M-C) externally triggered at 30 Hz using a pulse sent from the PXIe-6356 card. The trigger pulse was simultaneously sent to the National Instruments equipment and the camera through a BNC splitter and time-stamped at the sampling rate of the audio recordings. BIAS software was used to record each video (developed by Kristin Branson, Alice Robie, Michael Reiser, and Will Dickson; https:// bitbucket.org/iorodeo/bias/downloads/). Custom-written software controlled and synchronized all recording devices (Mathworks; Natick, MA; version 2014b). Audio and video data were stored on a PC (Hewlett-Packard; Palo Alto, CA; Z620). Infrared lights (GANZ; Cary, NC; IR-LT30) were positioned above the cage and used to illuminate the arena. To determine the microphone positions, a ring of LEDs surrounded each microphone. The LEDs were only turned on during a 15-second recording session before the start of the experiment. During this pre-experimental recording session, a ruler was placed in the cage to determine a pixel-to-meter conversion ratio. Cages were filled with ALPHA-dri bedding (approximate depth of 0.5 inches) to increase the color contrast between the floor of the cage and the mice. Each mouse was recorded individually for 10 minutes after completing the 5-hour recording session.

Data processing

A data analysis pipeline was set up on the University of Delaware's computer cluster and used to determine the trajectory of each mouse, as well as information about the vocal signal assigned to each mouse.

Tracking: The position of each mouse was automatically tracked using a program called Motr[25]. This program fit an ellipse around each mouse for every frame of video, and the following was determined: 1) x and y position of the ellipse's center, 2) direction the mouse was facing, 3) size of the ellipse's semi-major and semi-minor axes. With this information,

the position of each mouse's nose was calculated in every frame. Moreover, the instantaneous speed of each mouse was computed based on the distance the mouse traveled between frames. After trajectories were determined, tracks were visually inspected.

Audio segmentation: Automatic extraction of vocal signals from the eight channels of the microphone array was conducted via multi-taper spectral analysis[52]. Audio data were bandpass filtered between 30 and 110 kHz, and then segments overlapping in time were Fourier transformed using multiple discrete prolate spheroidal sequences as windowing functions (K = 5, NW = 3). An F-test determined when each time-frequency point exceeded the noise threshold at α=0.05[53]. Multiple segment lengths were used to capture a range of spectral and temporal scales (NFFTs = 64, 128, 256). Data were combined into a single spectrogram and convolved with a square box (11 pixels in frequency by 15 in time) to fill in small gaps before the location of continuous regions, with a minimum of 1,500 pixels, were assessed. Because male and female mice vocalize together during courtship chases[24], each continuous sound was extracted as a separate vocal signal as long as there were no harmonics present. Harmonics were considered overlapping signals in which the duration of overlap exceeded 90 percent of the shortest signal's duration and the frequencies of the signals that overlapped in time were integer multiples of each other[24]. When harmonics were present, only information pertaining to the lowest fundamental frequency was included in analyses. For each extracted vocal signal, frequency contours (a series of data points in time and frequency) were calculated[52]. Audio data was subsequently viewed with a custom-written Matlab program to confirm that the system was extracting vocal signals.

Sound source localization: Vocal signals were localized and assigned to specific mice using a method modified from Neunuebel et al.[24] and described in detail by Warren et al.[27]. For each extracted signal, eight sound source estimates were determined. Each of the estimates was generated by localizing the signal using eight different combinations of microphones, where each combination eliminated one microphone from the localization process. For example, the first estimate used data from microphones 2–8. The second estimate used data from microphones 1 and 3–8. This process was repeated until each microphone was omitted once. The average x- and y-coordinates of the eight point estimates were used as the sound source location. The sound source location and the eight point estimates were then used to compute a probability density function over the cage. This function describes the probability that the signal originated from any point of the recording environment. Locations with higher densities indicate a greater likelihood that the signal originated from that position. For each mouse, a density was determined from the probability density function based on the position of the mouse's nose (calculated from the output of Motr). Densities (D) were used to calculate a mouse probability index (MPI) for each mouse using the following formula:

$$MPI_n = \frac{D_n}{\sum_{i=1}^{M} D_i}$$

where n = mouse index and M = the total number of mice. The MPI was used to assign vocal signals to individual mice, with signals only assigned to a mouse when the MPI exceeded

0.95. In short, the system used the time it took for vocal signals to reach each of the eight microphones to generate points of localization. Vocal signals were assigned to a specific mouse when these points of localization were close enough to a mouse's nose that the system could make assignments with 95% certainty.

As previously described[24], potential temporal and spatial biases were assessed by plotting the time and location of all assigned and unassigned vocal signals for each experimental recording (Supplementary Fig. 5A,B). To determine why signals were unassigned (Supplementary Fig. 5C), we calculated the distances between the point of localization and the two closest mice. For the assigned vocal signals, the two distances were plotted relative to each other and then normalized by peak value. This step was repeated for the unassigned vocal signals. Then a difference distribution was calculated (unassigned-assigned), with values closer to 1 representing distances with more unassigned vocal signals and values closer to −1 representing distances with more assigned vocal signals.

Behavioral classifiers

To automatically identify when specific behaviors were occurring, a machine-based learning program was used (Janelia Automatic Animal Behavior Annotator; JAABA)[26]. The program was manually trained to recognize specific behaviors with high accuracy. The following behaviors were classified as defined by previous work[54–57]:

Chase Male - A male follows another male while the two mice are within two body lengths of each other. The classifier was trained on 8,303 frames of video (2,717 positive examples and 5,586 negative examples), using the mouse position data extracted by Motr.

Chased - The male in a chase that is being followed.

Fight - Both of the males engage in physical contact. This behavior involves biting, wrestling, and rolling over each other. The classifier was trained on 20,731 frames of video (11,990 positive examples and 8,741 negative examples).

Flee - One male runs away in response to an action from another male. The classifier was trained on 7,408 frames of video (3,349 positive examples and 4,059 negative examples).

Fled from - The male that the fleeing male is escaping.

Mutual circle - Two males put their heads near the anogenital region of each other and spin in a circle. The classifier was trained on 9,051 frames of video (1,906 positive examples and 7,145 negative examples).

Chase female - A male follows a female while the two mice are within two body lengths of each other. The classifier was trained on 8,982 frames of video (4,054 positive examples and 4,928 negative examples).

Walk - A mouse actively moves around the cage. When walking, no other mice could be within 35 cm. The classifier was trained on 3,512 frames of video (2,756 positive examples and 756 negative examples).

When identifying behaviors, JAABA assigns a confidence score for each frame, which represents the likelihood of a given behavior occurring. For the classifiers used to extract behaviors, confidence scores ranged from −18.3 to 12.8, with positive scores indicating an increased likelihood that the behavior was occurring. The confidence scores, as well as other behavioral characteristics, were used as post hoc refinements to reduce incidents of false positives (detailed below). The necessary refinements were determined by manually inspecting all examples extracted by JAABA for one dataset and then determining what refinements would increase accuracy. All behavioral classifiers had a false positive rate under 5%. The false-positive rate was determined by manually inspecting all examples from a single dataset and 20 random examples from each of the other datasets. For walk and male chase female, 200 examples were inspected from a single dataset as well as 20 random examples from each of the other datasets. Multiple people independently evaluated classifiers. For each extracted behavior, duration was calculated based on the difference between the start and stop times. We calculated the average speed of each behavioral event by taking the mean of the instantaneous speeds for every frame of the behavior.

Male chase male - When extracting chases, examples that were within ten frames of one another were merged, and all examples lasting less than six frames were eliminated. Four additional criteria were used to decrease the likelihood of false positives. First, the trajectories of the two males needed to overlap by at least 20%. Second, the average of the confidence scores needed to exceed 0.5. Third, the average distance between the two males during a potential chase needed to be less than 30 cm. Finally, the average distance between the animal being chased and the closest female needed to be greater than 15 cm. There were 2,136 examples of a male chasing another male detected across all recordings.

Fight - When extracting fights, examples that were within ten frames of one another were merged, and examples lasting fewer than 30 frames were eliminated. Three additional criteria were used to decrease the likelihood of false positives. First, the average of the confidence scores needed to exceed 1.5. Second, the distance between the males and the nearest female at the start of the fight needed to be greater than five cm. Finally, the average speed of the male JABBA determined as having initiated the fight needed to be greater than 7.5 cm/s. There were 1,177 examples of two males fighting detected across all recordings.

Flee and flee from - Three criteria were used to decrease the likelihood of false positives. First, the duration of the event needed to exceed ten frames. Second, the average of the confidence scores needed to exceed 0.7. Finally, the nearest male needed to be within eight cm at the start of the event. There were 851 examples of a male fleeing from another male detected across all recordings.

Mutual circle - When extracting mutual circles, all examples that were within five frames were merged, and all examples lasting fewer than ten frames were eliminated. Two additional criteria were used to decrease the likelihood of false positives. First, the average of the confidence scores needed to exceed 1.5. Second, the distance

between the males at the start of the event needed to be less than five cm. There were 163 examples of two males circling each other detected across all recordings.

Male chase female - When extracting instances of a male chasing a female, all examples within ten frames of one another were merged. Events with fewer than ten frames were eliminated. Three additional criteria were used to decrease the likelihood of false positives. First, the average of the confidence scores needed to exceed 0.5. Second, the trajectories of the two animals involved in the chase needed to overlap more than 20%. Third, the average distance between the male and female needed to be between 7 cm and 30 cm. There were 3,085 examples of a male chasing a female detected across all recordings.

Walk - When extracting walks, examples shorter than 20 frames were eliminated, and the average of the confidence scores that JAABA assigned to each frame of the walk needed to exceed 1.0. There were 21,953 examples of a male walking in isolation detected across all recordings.

Quantification of behaviors

Temporal analyses: Temporal plots were created by dividing each experiment into five 1-hour time bins. Subsequently, we calculated the number of times each male mouse performed every behavior, the durations of the behaviors, and speeds of the behaviors, and then partitioned the results into hourly time bins. A 1-way ANOVA was used to calculate differences over time ($\alpha = 0.05$).

Individual behavioral differences: The number of times each mouse performed every behavior was measured, and proportions were calculated by dividing the number of times that a particular male performed a specific behavior by the total number of times the behavior was performed by all the males (Supplementary Fig. 1). For each behavior, significant differences were quantified using a chi-square test ($\alpha = 0.05$). The same analyses were applied to vocal signal types.

Vocal rate during behaviors: Vocal rate was calculated by dividing the number of vocal signals assigned to the mouse performing the behavior by the duration of the behavior (Supplementary Fig. 6). For behaviors where two mice were performing the same behavior (fight and mutual circle), the vocal rate was calculated for each mouse individually; thus, the number of data points for each behavior was doubled. A 1-way ANOVA ($\alpha = 0.05$) was used to determine if the vocal rate differed between behaviors. If an animal performing a specific behavior emitted vocal signals, then the behavior was considered to coincide with vocal signaling. In cases when no vocal signals were assigned to the animal performing the behavior, and there were no unassigned vocal signals between the beginning and end of the behavior, then the behavior was classified as silent. For the temporal vocal rate analysis (Supplementary Fig. 7), only behaviors with vocalizations were included and partitioned into hourly time bins. 1-way ANOVAs were used to quantify changes in vocal rate over time ($\alpha = 0.05$).

Vocal signal clustering

Each signal's frequency contour was normalized by subtracting the mean frequency and partitioning the signal into 100 time bins[58]. Signals were then progressively clustered based on the shape of their normalized frequency contours using K-means clustering, which is an unsupervised learning algorithm that groups items with similar features. Initially, signals were divided into two groups. Means and standard deviations were calculated for each time bin in each group. If more than 25% of the time bins for more than 3 percent of the vocal signals within a group were outside of 2.5 standard deviations from the mean, all of the signals were re-clustered using one more cluster than was previously used (Supplementary Fig. 8). This process was repeated until 97% of all vocal signals in each cluster had at least 75% of their time bins within 2.5 standard deviations from the mean. Because vocal signals were extracted as continuous sounds to ensure emission was from a single animal, no categories were comprised of signals containing jumps in frequency or harmonics.

Validating clustering: To quantify the similarity between two vocal signals, a dissimilarity index was calculated (Supplementary Fig. 9). The index was derived by comparing the normalized frequency contours of two signals. The following formula was used to compare frequencies at each of the 100 time bins:

$$\frac{|freq \ \ signal \ \ 2_{binN} - freq \ \ signal \ \ 1_{binN}|}{freq \ \ signal \ \ 2_{binN} + freq \ \ signal \ \ 1_{binN}}$$

After calculating index values for all 100 time bins, the indices were then averaged to determine the dissimilarity index for the two signals. Values closer to zero indicated that the signals shared similar acoustic features, whereas values closer to 1 indicated that signals were less similar. The dissimilarity metric was applied to multiple analyses.

To compare the acoustic structure of vocal signals between the different signal types, a permutation test was conducted. First, vocal signals classified as type 1 were randomly split into two equal groups (total number of signals in the group/2). Dissimilarity indices were calculated between each pair of signals. Next, dissimilarity indices were calculated between type 1 and type 2. Comparison groups were generated by randomly selecting an equal number of signals from type 1 and 2. The number of signals selected was determined by dividing the total number of signals in the group with the fewest signals by 2. The dissimilarity scores for the within group (Supplementary Fig. 9A) and between group (Supplementary Fig. 9B) comparisons were statistically evaluated using a Wilcoxon rank-sum test. This was repeated for each of the possible signal type comparisons (total of 210). Within signal type dissimilarity scores were always significantly lower than between signal type dissimilarity scores when alpha levels were conservatively set using a Bonferroni approach to correct for multiple comparisons (0.05/210).

Random assignment: The within group average root mean square error (RMS) was calculated for the signals assigned to each vocal category by calculating the RMS at each of the 100 time bins and then averaging. A permutation test was then run. Vocal signals were randomly assigned to a new category. The number of categories and the total number of

signals in each category type were maintained. The category label was the only parameter that changed. The average RMS was then calculated for the signals randomly assigned to a category. This was repeated 1,000 times. The real RMS value within each category was then compared to the RMS value from the randomized data, with all real values falling significantly below the randomized values.

Calculating dissimilarity of each vocal signal type between and within mice

To ensure that signals emitted by a particular mouse were not different from signals of the same type emitted by another mouse, a within-mouse dissimilarity index was compared to a between-mice dissimilarity index for each type of vocal signal. To calculate the within-mouse dissimilarity index, all the vocal signals of a particular type emitted by a single mouse were randomly split into two groups (Supplementary Fig. 9C). Using the previously described dissimilarity index formula, signal 1 in group 1 was compared to signal 1 in group 2, then signal 2 in group 1 was compared to signal 2 in group 2. This process was repeated for all signals and each group, and then repeated for each mouse. The final within mouse dissimilarity index value for a particular type was the average of the within mouse dissimilarity index values for each mouse. Next, a dissimilarity index was calculated to compare signals of the same type between mice (Supplementary Fig. 9D). To do this, vocal signals of a particular type, emitted by a particular mouse, were compared to vocal signals of the same type emitted by the other mice. For example, type 1 signals emitted by mouse 1 would be compared to type 1 signals emitted by mouse 2, followed by type 1 signals emitted by mouse 3. To control for differences in sample size, we determined which mouse in each pair emitted the fewest signals, and randomly selected that many signals from the other mouse in the pair. Between mice similarity indices were then calculated in the same manner as within-mouse indices. A 95% confidence interval was calculated for the indexes comparing vocal signal type dissimilarities emitted by the different mice. A 2-sample t-test was used to compare the between mouse dissimilarity to the within mouse dissimilarity ($\alpha = 0.05$). No comparisons reached significance (all p values>0.05).

Quantifying change in pitch of signals emitted during behavior

To evaluate changes in the pitch of vocal signals emitted during dominant and non-dominant behaviors, the slope was calculated for each vocal signal occurring during dominant (mutual circle, fight, chase female, fled from, chase male) or non-dominant (flee, chased, walk) behaviors. All signals used in the analysis were assigned to the animal engaged in the specific behavior. After normalizing the frequency contours of each vocal signal into 100 time bins, the frequency in the first time bin was subtracted from the frequency in the final time bin, and this difference was divided by 100 (i.e., the number of time bins in each normalized frequency contour). There were 26,592 vocal signals emitted during aggressive behaviors and 5,673 vocal signals produced during avoidance behaviors. A 2-sample t-test was used to compare the slopes of vocal signals emitted in aggressive and avoidance behaviors ($\alpha = 0.05$).

Calculating signal type usage

To quantify how frequently each type of vocal signal was emitted, the number of signals of each type that were assigned to each male mouse was measured (Fig. 4A). A 1-way ANOVA was used to quantify statistical differences ($\alpha = 0.05$).

Categorizing vocal signals based on spectral features

As previously described[30], the duration, interval, and amplitude were calculated for each vocal signal. The duration was the temporal length of the vocal signal. Interval was defined as the frequency range or bandwidth of the signal. Amplitude was the maximum height of the voltage trace of the vocal signal. For each of the vocal features, signals were sorted in ascending order.

Equal number: When categorizing based on equal number, the ascendingly sorted signals were split into 22 categories such that each category had the same number of vocal signals (Supplementary Fig. 16A–C).

Equal space: When categorizing based on equal spacing, ascendingly sorted signals were partitioned into 22 categories such that each bin was the same size (Supplementary Fig. 16D–F). The size of each bin was 13.4 ms, 1.93 kHz, and 57.9 mV for the duration, interval, and amplitude, respectively.

Equal space 10 90: When categorizing based on the equal spacing 10 90 method, the bottom and top 10% of all vocal signals were put into their own respective categories. The remaining vocal signals were then split into 20 categories such that each bin was the same size (Supplementary Fig. 16G–I). The size of each bin was 1.74 ms, 0.6 kHz, and 5.1 mV for the duration, interval, and amplitude, respectively.

Individual differences in vocal signal emission

This analysis was identical to the previously described individual differences in behavior analysis, except vocal signals were quantified.

Calculating the proportion of signals emitted during each instance of a behavior

After clustering vocal signals based on the shape and categorizing them based on their spectral features, social context-dependent expression was determined by calculating the proportion of signals emitted during behavior (Fig. 6A–D; Supplementary Fig. 17). The proportion was calculated by dividing the number of times a specific type of vocal signal was emitted during an instance of the behavior by the total number of vocal signals in that behavior. Because we were able to localize the source of the signals and pinpoint which mouse was vocalizing[24, 27], only vocal signals assigned to the mouse performing the behavior were included in the calculation. Proportions were calculated for each instance of a behavior and the average proportion was the mean of all of the individual proportions. This was repeated for each behavior and every vocal signal type. To show that animals emitted different types of vocal signals depending on the behavior they were performing, a 2-way ANOVA was conducted to look for an interaction between behavior and signal type. For each signal type in each behavior the number of samples was equal to the number of

instances of the behavior that contained vocal signals (male chase male = 651, male being chased by a male = 444, fight = 648, fled from = 151, flee = 65, mutual circle = 132, male chasing a female = 1,932, walk = 815).

Determining chance levels of vocal expression

To determine if the proportion of signals emitted during a behavior was significantly above, below, or at chance, a permutation test was performed. This process was carried out for the automated clustering based on shape (Fig. 6A), as well as the three methods of categorization based on spectral features (Fig. 6B–D; Supplemental Fig. 17). In this analysis, each vocal signal was randomly reassigned a signal type, and the number of signals of each type was kept constant. For example, if there were 900 type 1 signals observed in the real data, then type 1 would be randomly assigned to 900 of the vocal signals. For this analysis, the identity of the vocalizer and the time of the vocal signal were held constant; therefore, the proportion of randomly assigned signal types emitted during each instance of a behavior could be calculated. The proportion was calculated by dividing the number of times a randomly-assigned vocal signal type was emitted during an instance of the behavior by the total number of vocal signals in that event. This process was repeated 1,000 times. The distribution of randomly-assigned signal type proportions was then compared to the actual proportions. Finally, a z-score was calculated for each proportion to determine if it was emitted above, below, or at chance.

The proportion of events in which specific types of vocal signals were emitted was calculated by finding the number of instances of a specific behavior that contained a vocal signal type and dividing by the number of times the behavior occurred. Similar to the previous proportion calculation, only vocal signals assigned to the mouse performing the behavior were included. Chance levels were then determined for the proportion of events with specific types of vocal signals (Supplementary Fig. 14). Identical to calculating chance levels based on the proportion of signals emitted during a behavior, each vocal signal was randomly assigned a signal type. The number of signals of each type, the identity of the vocalizer, and the time when the vocal signal was emitted remained constant. We then calculated the proportion of events in which randomly-assigned signal types were emitted by finding the number of instances of a specific behavior that contained a randomly-assigned signal type and dividing by the number of times the behavior occurred. This was repeated 1,000 times, the distributions were compared to the actual proportions, and z-scores were calculated.

Addressing potential bias in vocal signal assignment during behavior

The proportion of assigned and unassigned signal types occurring in each behavior were compared. For this analysis, unassigned signals were temporarily assigned to animal with the highest probability of emitting the signal (see sound source localization methods) and included in the unassigned category for the behavior the animal was performing. First, the number of unassigned and assigned signals of each type within each behavior were calculated. To calculate a proportion, the counts were divided by the total number of unassigned or assigned vocal signals that occurred within the given behavior, respectively

(Supplementary Fig. 15). Distributions were compared using a 2-sampled Kolmogorov-Smirnov test ($\alpha = 0.05$).

Consistency of behaviorally-dependent vocal emission across individuals

Calculating proportion of each vocal signal type in each behavior: For each mouse, the proportion was calculated by dividing the number of times a specific type of vocal signal was emitted during an instance of the behavior by the total number of vocal signals in that event. Only vocal signals assigned to the mouse performing the behavior were included in the calculation. These values were then averaged to get the total for each mouse. The average proportion was the mean of all of the proportions for each mouse. This was repeated for each behavior and every vocal signal type (Supplementary Fig. 19A). For each signal type in each behavior the number of samples was equal to the number mice that emitted vocal signals during the behavior. A 2-way ANOVA was conducted to statically quantify the relationship between behavior and signal type at the individual level.

Calculating vocal chase indices: For all male mice, the average proportion of each vocal signal type emitted during every male-male chase was calculated for times when the mice were acting as both the dominant and non-dominant animal. This produced two average proportions per mouse and vocal signal type (i.e., average proportion when an animal was chasing and being chased). Vocal indices were computed by subtracting the proportion of the signal types produced when the mouse was acting as the non-dominant animal from the proportion of the signal types emitted when the mouse was acting as the dominant animal (Supplementary Fig. 19D). Next, this value was divided by the sum of the two proportions. Animals that did not perform one of the behaviors or emit a specific vocal type while both chasing and being chased were excluded from the calculation. A 1-way ANOVA was used to statistically analyze differences in vocal chase indices ($\alpha = 0.05$). Post-hoc pairwise comparisons were conducted using a Fisher's least significant difference procedure and then a Benjamini-Hochberg procedure was used to correct for multiple comparisons.

Alternate clustering approaches

Clustering based on 11 signal types: Each signal's frequency contour was normalized by subtracting the mean frequency and partitioning the signal into 100 time bins[58]. Signals were then clustered into 11 types based on the shape of their normalized frequency contours using K-means clustering (Supplementary Fig 20).

Clustering using DeepSqueak: As a secondary approach for classifying signal types, the same frequency contours processed using our vocal clustering algorithm were processed using "DeepSqueak" version 2.3.0. This approach used unsupervised learning to classify the shape of each vocal signal's frequency contour. Using "DeepSqueak", signals were partitioned into 6 categories (Supplementary Fig 20).

Decoding behavior based on vocal signal emission

Preprocessing: Decoding was performed on signals automatically clustered based on shape as well as signals categorized by duration, bandwidth, and amplitude using the "Equal

Space 10 90" method of categorization. First, every instance of each behavior was represented as a single 22-dimensional data point. Each of the 22 dimensions represented the proportion of each vocal signal type emitted during that particular behavioral example. Behavioral examples with no vocal signals were excluded.

Training: The same number of examples for each of the behaviors was randomly selected to train the classifier. To determine the number of training examples, 75% of the examples of the least frequently occurring behavior were selected. The behavior with the fewest number of examples was flee, which had 71 examples. As a result, 53 examples of each behavior were randomly selected to train a multi-class support vector machine with error-correcting output codes implemented through the MATLAB fitcecoc() function.

Testing: After training the classifier, an equal number of examples of each behavior was randomly selected to test the classifier. This number was equal to 25% of the number of instances of the behavior with the fewest examples, which was 17. Importantly, the test examples could not have been used to train the classifier. Testing was done using the MATLAB predict() function. The accuracy of the classifier was determined by dividing the total number of correct predictions by the total number of predictions and multiplying it by 100. Because there were eight possible behaviors the classifier could predict, chance levels were 12.5% (100% / 8 behaviors = 12.5%).

Quantifying decoding: The process of randomly selecting examples to train and test the classifier was repeated 1000 times. For each iteration, an accuracy was computed (Fig. 7A–D). To evaluate whether or not classifiers were performing above chance ($\alpha = 0.05$), a z-score was calculated for chance levels. The following formula was used:

$$z \; score = \frac{(chance \; level) - \; mean \; of \; the \; 1000 \; classifier \; accuracies}{standard \; deviation \; of \; the \; 1000 \; classifier \; accuracies}$$

For each iteration, the accuracy of predicting individual behaviors was computed by dividing the number of times a behavior was correctly predicted by 17, with seventeen representing the maximum number of times the behavior could have been correctly predicted (Fig. 7E–H). Confusion matrices were created by counting the number of times the tested behavioral examples were assigned a behavioral label for all 1000 iterations. These counts were then divided by 17,000 (17 examples × 1000 iterations) to get the proportion of times that each behavior was classified as one of the eight behaviors (Fig. 7I–L).

Calculating mutual information

As previously described[6], mutual information between call type and behavior was calculated using the following formula:

$$MI = \sum_{v \, . \, b} p(v, b) \log\left(\frac{p(v, b)}{p(v)p(b)}\right)$$

MI is the mutual information between vocalization and behavior in bits, where p(v,b) is the joint probability of a particular vocal signal type occurring with a particular behavior, p(v) is the probability of a particular vocal signal type occurring and p(b) is the probability of a particular behavior occurring. MI was calculated for only behaviors with vocalizations emitted by the animal performing them and only vocal signals emitted by the animal performing the behavior. To calculate the probability of emitting a particular vocal signal type or observing a specific behavior, the number of times a specific vocal signal or behavior occurred was divided by the total number of vocal signals or behaviors, respectively. The joint probability of a particular vocal signal type and behavior co-occurring was determined by calculating the ratio of the number of times the particular vocal signal type was emitted during the behavior and the total number of times any vocal signal type was emitted by mice performing any behavior. To estimate spurious information, we randomly shuffled the vocal signal identity 1000 times; thus, breaking any potential statistical relationship between vocal identity and behavior. The actual MI score was compared to the distribution of MI scores generated from the 1000 shuffles, with significance set at $\alpha = 0.05$.

Quantifying the impact of vocal signaling on behavior

We examined the impact of vocal signal emission on behavior using behaviors in which each animal played a distinct role (male-male chases, flees, and male-female chases). For each of these behaviors, similar analyses were conducted—only the vocal signal types were different. The speed of an animal provides a quantifiable readout of a behavioral change that reflects internal motivation and decision making[59–63]. Moreover, complex animal behavior is composed of multiple simple modules at the sub-second time scale[64]. Therefore, instantaneous velocity was computed every 33 ms (equivalent to 1 frame because video sampling rate occurred at 30 frames per second). The velocity of the animals over a sub-second time window (11 video frames which is equivalent to 367 ms) was examined. The frame in which vocal signal emission occurred was centered at zero (time=0). This allowed us to compare of the speeds of the animals before, during, and after vocal emission. To ensure that the entire 367 ms time window was part of a behavior, vocal signals that occurred within 166 ms of the beginning or end of the behavior were excluded from the analysis. To isolate the effect of specific types of vocal signals, the analysis was restricted to signals that did not have another signal emitted by the same mouse within the 367 ms time window. For each of the behaviors, we categorized vocal signals as either dominant or non-dominant. When the sender of a vocal signal was chasing or being fled from, we considered dominant vocal signals to be those that were emitted above chance within the given behavior according to Fig. 6A, and non-dominant signals to be those that were emitted below chance according to Fig. 6A. In contrast, when the sender of a vocal signal was performing a non-dominant behavior (being chased or fleeing), we considered dominant vocal signals to be those that were emitted below chance according to Fig. 6A, and non-dominant signals to be those that were emitted above chance according to Fig. 6A. The following number and type of vocal signals were included in each analysis:

	Signals	
Condition	**Dominant (# included)**	**Non-dominant (# included)**
Male chasing male	13 (71), 14 (46), 16 (57), 18 (22), 19 (33), 20 (20), 21 (22)	2 (2), 4 (7), 6 (13), 7 (8), 9 (38), 10 (18)
Fled from	14 (5), 16 (10), 19 (6), 22 (0)	7 (4), 9 (4)
Male chasing female	12 (337), 13 (330), 14 (220), 16 (272), 19 (146), 20 (66), 21 (53)	1 (5), 3 (22), 4 (78), 5 (17), 6 (65), 7 (81), 8 (19), 9 (250), 10 (79)
Male being chased	12 (14), 13 (8), 14 (6), 16 (6), 17 (1), 18 (0), 19 (2), 20 (2), 21 (1), 22 (0)	1 (7), 2 (14), 3 (8), 4 (46), 5 (7), 6 (59), 7 (81), 9 (96)
Male fleeing	12 (0), 13 (0)	6 (4), 7 (7)

When the vocalizing male (sender) was chasing a male, being chased, fleeing, or being fled from, the other male mouse was receiver 1 because he was actively participating in the behavior. The two females were considered receiver 2 and 3. Each female was assigned to either receiver 2 or 3 based on her distance from the vocalizing male at the time of vocal emission. The closest female to the sender was receiver 2 and the farthest female from the sender was receiver 3. When the vocalizing male was chasing a female, the female being chased was considered receiver 1, the other male was considered receiver 2, and the other female was considered receiver 3. The plots in Fig. 8B–E and Supplemental Figs. 22–25,B–E show the average speed of the sender and receivers (1, 2, and 3) as well as the standard error for both above and below chance vocal signals. To see if there was a delayed response to the vocal signals being emitted in these contexts, the velocity of the animals was calculated for each frame in the 5 seconds following the emission of the signal (Fig. 8F–I and Supplemental Figs. 22–25F–I). The analysis included the same signals that were examined when looking at the velocity 5 frames before and after the vocal signal. Vocalizations are typically emitted in bouts[65]; therefore, isolating a single vocal signal during a specific behavior followed by 5 seconds of silence is rare. Consequently, the last 4.8s of the 5 second window following vocal emission may include other vocal signals or behaviors.

Acceleration at a given frame (i.e., instantaneous acceleration) was calculated by subtracting the speed in the given frame from the speed in the following frame (Fig. 8J–M and Supplemental Figs. 22–25J–M). For each vocal signal examined, acceleration pre- and post-signal emission was calculated by averaging the accelerations in the 5 frames before and 5 frames after the vocal signal. A paired t-test ($\alpha = 0.05$) was used to determine if there was a significant difference between the acceleration immediately before and after vocal emission. A 1-sample t-test was used to determine if the acceleration pre- or post-signal differed from a mean of 0 ($\alpha = 0.05$).

To establish if any change in acceleration was greater than expected by chance, we conducted a permutation test for dominant and non-dominant vocal signals (Fig. 8N–O and Supplemental Figs. 22–25,N–O). For the permutation tests, each vocal signal included in the previous analysis (i.e., quantifying changes in behavior) was randomly shifted to a time that fulfilled two criteria. First, the randomly selected time had to occur during an instance of the

behavior currently being examined (this could be the example in which the real signal occurred). Second, the randomly selected time needed to be at least 166 ms after the start or before the end of the behavior being examined. The acceleration before and after the randomly shifted vocal signals was then calculated as previously described for the real data. For the randomly shifted times, a 2-sample t-test was conducted on the accelerations before and after vocal signal emission and a t-score was computed. The randomization procedure, acceleration calculations, and t-test were repeated 1,000 times. Z-scores were then calculated to see if the change in acceleration was larger than expected by chance.

Blinding

All mice were from the same strain and no comparisons between different groups of mice were conducted. After completing the 11 experiments, behaviors were automatically extracted. An automated behavioral extraction program was used to remove any unconscious biases that may result from manually extracting behaviors. This ensured that experimenters were initially blind to the types of behaviors each mouse performed. After completing the 11 experiments and using our automated vocalization extraction program, vocalizations were partitioned into different categories. All vocalizations were pooled together when categorizing and the identity of the mouse was omitted; thus, experimenters were initially blind to the types of vocalizations each mouse produced.

Statistics

No statistical methods were used to predetermine sample size. All statistical analyses were performed in Matlab (Mathworks; Natick, MA; version 2016a) using a significance level at an a of 0.05. Data distributions were assumed to be normal but this was not formally tested. All t-tests, Kolmogorov-Smirnov tests, and ANOVAs were two-tailed. Post-hoc pairwise comparisons were conducted using a 2-tailed Fisher's least significant difference procedure and then a Benjamini-Hochberg procedure was used to correct for multiple comparisons unless stated otherwise.

Reporting summary

Additional information on research design is available in the Nature Research Reporting Summary linked to this article.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

References:

1. Owen-Smith N Territoriality in the white rhinoceros (Ceratotherium simum) Burchell. Nature 231, 294–296 (1971). [PubMed: 4930978]

2. Pizzari T & Birkhead TR Female feral fowl eject sperm of subdominant males. Nature 405, 787–789 (2000). [PubMed: 10866198]

3. Chen P & Hong W Neural Circuit Mechanisms of Social Behavior. Neuron 98, 16–30 (2018). [PubMed: 29621486]

4. Bahrami B, et al. Optimally interacting minds. Science 329, 1081–1085 (2010). [PubMed: 20798320]

5. Seyfarth RM, Cheney DL & Marler P Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. Science 210, 801–803 (1980). [PubMed: 7433999]

6. Bradbury JW & Vehrencamp SL Principles of Animal Communication (Sinauer Associates, Inc., Sunderland, Massachusetts, 1998).

7. Gruters KG & Groh JM Sounds and beyond: multisensory and other non-auditory signals in the inferior colliculus. Front Neural Circuits 6, 96 (2012). [PubMed: 23248584]

8. Marlin BJ, Mitre M, D'Amour J A, Chao MV & Froemke RC Oxytocin enables maternal behaviour by balancing cortical inhibition. Nature 520, 499–504 (2015). [PubMed: 25874674]

9. Wolpert DM, Doya K & Kawato M A unifying computational framework for motor control and social interaction. Philos Trans R Soc Lond B Biol Sci 358, 593–602 (2003). [PubMed: 12689384]

10. Li Y, et al. Neuronal Representation of Social Information in the Medial Amygdala of Awake Behaving Mice. Cell 171, 1176–1190 e1117 (2017). [PubMed: 29107332]

11. Remedios R, et al. Social behaviour shapes hypothalamic neural ensemble representations of conspecific sex. Nature 550, 388–392 (2017). [PubMed: 29052632]

12. Kingsbury L, et al. Correlated Neural Activity and Encoding of Behavior across Brains of Socially Interacting Animals. Cell (2019).

13. Hong W, et al. Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning. Proc Natl Acad Sci U S A 112, E5351–5360 (2015). [PubMed: 26354123]

14. Sales GD Ultrasound and mating behaviour in rodents with some observations on other behavioural situations. Journal of Zoology 168, 149–164 (1972).

15. Holy TE & Guo Z Ultrasonic songs of male mice. PLoS Biol 3, e386 (2005). [PubMed: 16248680]

16. Shepard KN & Liu RC Experience restores innate female preference for male ultrasonic vocalizations. Genes Brain Behav 10, 28–34 (2011). [PubMed: 20345895]

17. Pomerantz SM, Nunez AA & Bean NJ Female behavior is affected by male ultrasonic vocalizations in house mice. Physiol Behav 31, 91–96 (1983). [PubMed: 6685321]

18. Mahrt EJ, Perkel DJ, Tong L, Rubel EW & Portfors CV Engineered deafness reveals that mouse courtship vocalizations do not require auditory experience. J Neurosci 33, 5573–5583 (2013). [PubMed: 23536072]

19. Coffey KR, Marx RG & Neumaier JF DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations. Neuropsychopharmacology 44, 859–868 (2019). [PubMed: 30610191]

20. Chabout J, et al. Adult male mice emit context-specific ultrasonic vocalizations that are modulated by prior isolation or group rearing environment. PLoS One 7, e29401 (2012). [PubMed: 22238608]

21. Hanson JL & Hurley LM Female presence and estrous state influence mouse ultrasonic courtship vocalizations. PLoS One 7, e40782 (2012). [PubMed: 22815817]

22. Sugimoto H, et al. A role for strain differences in waveforms of ultrasonic vocalizations during male-female interaction. PLoS One 6, e22093 (2011). [PubMed: 21818297]

23. Grimsley JM, Monaghan JJ & Wenstrup JJ Development of social vocalizations in mice. PLoS One 6, e17460 (2011). [PubMed: 21408007]

24. Neunuebel JP, Taylor AL, Arthur BJ & Egnor SR Female mice ultrasonically interact with males during courtship displays. Elife 4 (2015).

25. Ohayon S, Avni O, Taylor AL, Perona P & Roian Egnor SE Automated multi-day tracking of marked mice for the analysis of social behaviour. J Neurosci Methods 219, 10–19 (2013). [PubMed: 23810825]

26. Kabra M, Robie AA, Rivera-Alba M, Branson S & Branson K JAABA: interactive machine learning for automatic annotation of animal behavior. Nat Methods 10, 64–67 (2013). [PubMed: 23202433]

27. Warren MR, Sangiamo DT & Neunuebel JP High Channel Count Microphone Array Accurately and Precisely Localizes Ultrasonic Signals from Freely-Moving Mice. J Neurosci Methods (2018).

28. Coen P, et al. Dynamic sensory cues shape song structure in Drosophila. Nature 507, 233–237 (2014). [PubMed: 24598544]

29. Silverman JL, Yang M, Lord C & Crawley JN Behavioural phenotyping assays for mouse models of autism. Nature reviews. Neuroscience 11, 490–502 (2010). [PubMed: 20559336]

30. Warren MR, Spurrier MS, Roth ED & Neunuebel JP Sex differences in vocal communication of freely interacting adult mice depend upon behavioral context. PLoS One 13, e0204527 (2018). [PubMed: 30240434]

31. Stern S, Kirst C & Bargmann CI Neuromodulatory Control of Long-Term Behavioral Patterns and Individuality across Development. Cell 171, 1649–1662 e1610 (2017). [PubMed: 29198526]

32. Frank MJ, Doll BB, Oas-Terpstra J & Moreno F Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. Nat Neurosci 12, 1062–1068 (2009). [PubMed: 19620978]

33. Favaro L, Gamba M, Gili C & Pessani D Acoustic correlates of body size and individual identity in banded penguins. PLoS One 12, e0170001 (2017). [PubMed: 28199318]

34. Stoeger AS & Baotic A Information content and acoustic structure of male African elephant social rumbles. Sci Rep 6, 27585 (2016). [PubMed: 27273586]

35. Rudorf S, et al. Neural Mechanisms Underlying Individual Differences in Control-Averse Behavior. J Neurosci 38, 5196–5208 (2018). [PubMed: 29760183]

36. Johnson KR, Erway LC, Cook SA, Willott JF & Zheng QY A major gene affecting age-related hearing loss in C57BL/6J mice. Hear Res 114, 83–92 (1997). [PubMed: 9447922]

37. Portfors CV, Roberts PD & Jonson K Over-representation of species-specific vocalizations in the awake mouse inferior colliculus. Neuroscience 162, 486–500 (2009). [PubMed: 19401221]

38. Galindo-Leon EE, Lin FG & Liu RC Inhibitory plasticity in a lateral band improves cortical detection of natural vocalizations. Neuron 62, 705–716 (2009). [PubMed: 19524529]

39. Shepard KN, Lin FG, Zhao CL, Chong KK & Liu RC Behavioral relevance helps untangle natural vocal categories in a specific subset of core auditory cortical pyramidal neurons. J Neurosci 35, 2636–2645 (2015). [PubMed: 25673855]

40. Neilans EG, Holfoth DP, Radziwon KE, Portfors CV & Dent ML Discrimination of ultrasonic vocalizations by CBA/CaJ mice (Mus musculus) is related to spectrotemporal dissimilarity of vocalizations. PLoS One 9, e85405 (2014). [PubMed: 24416405]

41. Tschida K, et al. A Specialized Neural Circuit Gates Social Vocalizations in the Mouse. Neuron (2019).

42. Wang F, et al. Bidirectional control of social hierarchy by synaptic efficacy in medial prefrontal cortex. Science 334, 693–697 (2011). [PubMed: 21960531]

43. Moy SS, et al. Sociability and preference for social novelty in five inbred strains: an approach to assess autistic-like behavior in mice. Genes Brain Behav 3, 287–302 (2004). [PubMed: 15344922]

44. Gold JI & Shadlen MN The neural basis of decision making. Annual review of neuroscience 30, 535–574 (2007).

45. Krakauer JW, Ghazanfar AA, Gomez-Marin A, MacIver MA & Poeppel D Neuroscience Needs Behavior: Correcting a Reductionist Bias. Neuron 93, 480–490 (2017). [PubMed: 28182904]

46. Gomez-Marin A, Paton JJ, Kampff AR, Costa RM & Mainen ZF Big behavioral data: psychology, ethology and the foundations of neuroscience. Nat Neurosci 17, 1455–1462 (2014). [PubMed: 25349912]

47. Berman GJ Measuring behavior across scales. BMC Biol 16, 23 (2018). [PubMed: 29475451]

48. Berman GJ, Choi DM, Bialek W & Shaevitz JW Mapping the stereotyped behaviour of freely moving fruit flies. J R Soc Interface 11 (2014).

49. Marques JC, Lackner S, Felix R & Orger MB Structure of the Zebrafish Locomotor Repertoire Revealed with Unsupervised Behavioral Clustering. Curr Biol 28, 181–195 e185 (2018). [PubMed: 29307558]

50. Morton ES On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. The American naturalist 111, 855–869 (1977).

51. König B Fitness effects of communal rearing in house mice: the role of relatedness versus familiarity. Anim Behav 48, 1449–1457 (1994).

52. Seagraves KM, Arthur BJ & Egnor SE Evidence for an audience effect in mice: male social partners alter the male vocal response to female cues. J Exp Biol 219, 1437–1448 (2016). [PubMed: 27207951]

53. Percival DB & Walden AT Spectral Analysis for Physical Applications (Cambridge University Press, Cambridge, 1993).

54. Grant EC & Mackintosh JH A Comparison of the Social Postures of Some Common Laboratory Rodents. Behaviour 21, 246–259 (1963).

55. Van Oortmerssen GA Biological significance, genetics and evolutionary origin of variability in behaviour within and between inbred strains of mice (Mus musculus). A behaviour genetic study. Behaviour 38, 1–92 (1971). [PubMed: 5100645]

56. Miczek KA, Maxson SC, Fish EW & Faccidomo S Aggressive behavioral phenotypes in mice. Behav Brain Res 125, 167–181 (2001). [PubMed: 11682108]

57. Weissbrod A, et al. Automated long-term tracking and social behavioural phenotyping of animal colonies within a semi-natural environment. Nat Commun 4, 2018 (2013). [PubMed: 23771126]

58. Tabler JM, et al. Cilia-mediated Hedgehog signaling controls form and function in the mammalian larynx. Elife 6 (2017).

59. Schultz W Predictive reward signal of dopamine neurons. J Neurophysiol 80, 1–27 (1998). [PubMed: 9658025]

60. Lemasson BH, Anderson JJ & Goodwin RA Collective motion in animal groups from a neurobiological perspective: the adaptive benefits of dynamic sensory loads and selective attention. J Theor Biol 261, 501–510 (2009). [PubMed: 19699212]

61. Hikosaka O The habenula: from stress evasion to value-based decision-making. Nature reviews. Neuroscience 11, 503–513 (2010). [PubMed: 20559337]

62. Thura D, Cos I, Trung J & Cisek P Context-dependent urgency influences speed-accuracy trade-offs in decision-making and movement execution. J Neurosci 34, 16442–16454 (2014). [PubMed: 25471582]

63. Chang SW, et al. Neural mechanisms of social decision-making in the primate amygdala. Proc Natl Acad Sci U S A 112, 16012–16017 (2015). [PubMed: 26668400]

64. Wiltschko AB, et al. Mapping Sub-Second Structure in Mouse Behavior. Neuron 88, 1121–1135 (2015). [PubMed: 26687221]

65. Chabout J, Sarkar A, Dunson DB & Jarvis ED Male mice song syntax depends on social contexts and influences female preferences. Front Behav Neurosci 9, 76 (2015). [PubMed: 25883559]
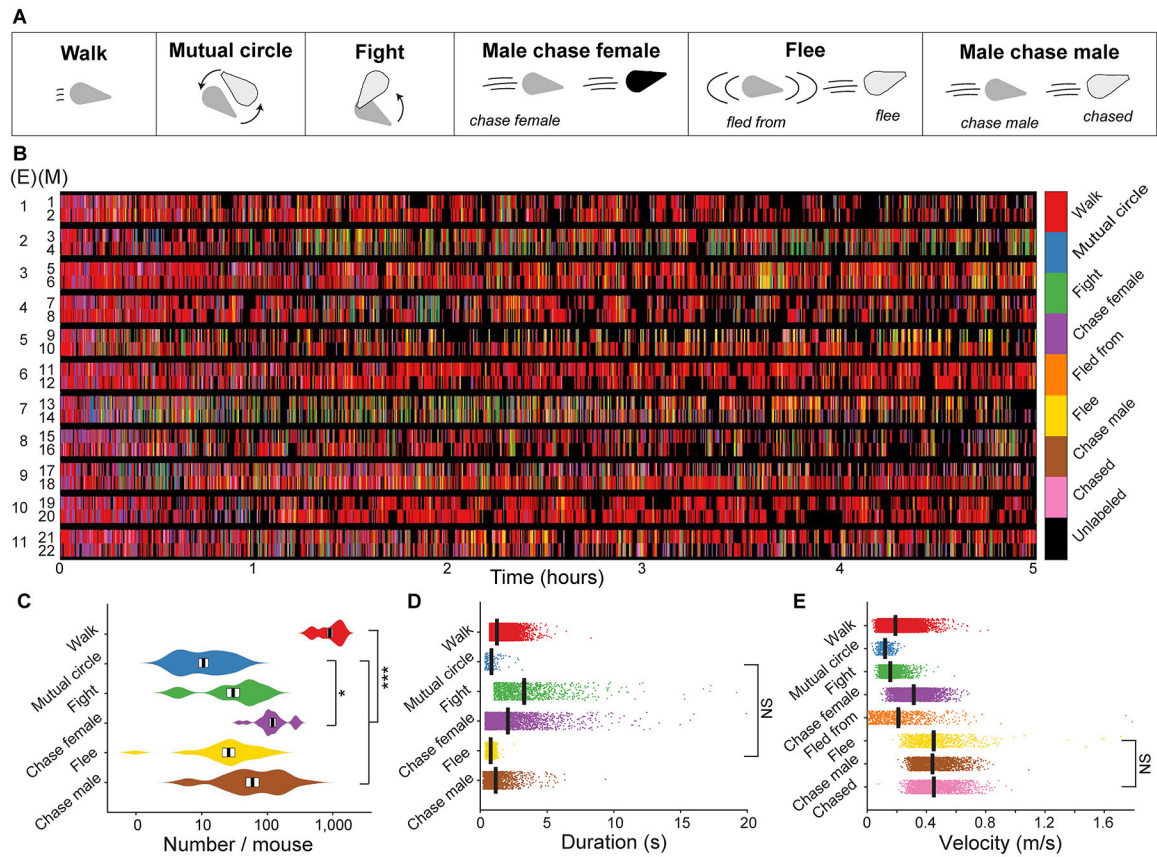
Fig. 1. Dynamic behavioral repertoires.
A, Schematic indicating classified behaviors observed in 11 groups of adult mice, with each group consisting of two females in estrus and two males. For each behavior, the number of mice participating and their actions are portrayed. The sex of the mouse is indicated by the color of the body (male, gray or light gray; female, black). For the behaviors involving two mice, the role each mouse plays in the behavior can be the same (fight, mutual circle) or unique (male chasing female, flee, male chasing male). Each unique role is labeled. **B**, Ethogram depicting the behaviors of each experimentally recorded (E) male mouse (M). **C**, Behavior occurrence distributions (violin plot elements: black vertical line = average, white box = SEM; n=110 biologically independent samples, 1-way ANOVA, F=115.8, p<10^{-44}, 2-sided Fisher's least significant difference and corrected for multiple comparisons using the Benjamini-Hochberg procedure, all significant T −17.20 or T 2.51, all non-significant −2.03 T 1.85, * = p<0.05, *** = p<0.001). **D**, Duration of each behavior (black vertical line = average, width of line = SEM; n=29,365 biologically independent samples, 1-way ANOVA, F=1640.7, p<10^{-100}, 2-sided Fisher's least significant difference and corrected for multiple comparisons using the Benjamini-Hochberg procedure, all significant T −4.34 or T 4.51, non-significant T=−0.639, p=0.52, all significant p<10^{-4}, NS = only non-significant comparison). **E**, Average velocity of the animals performing each behavior (black vertical line = average, width of line = SEM; n=33,692 biologically independent samples, 1-way ANOVA, F=6394.6, p<10^{-100}, Fisher's least significant difference and corrected for multiple comparisons using the Benjamini-Hochberg procedure, all significant T −2.26 or

T 3.39, non-significant, T=0.29, p=0.77, all significant p<0.05, NS = only non-significant comparison).
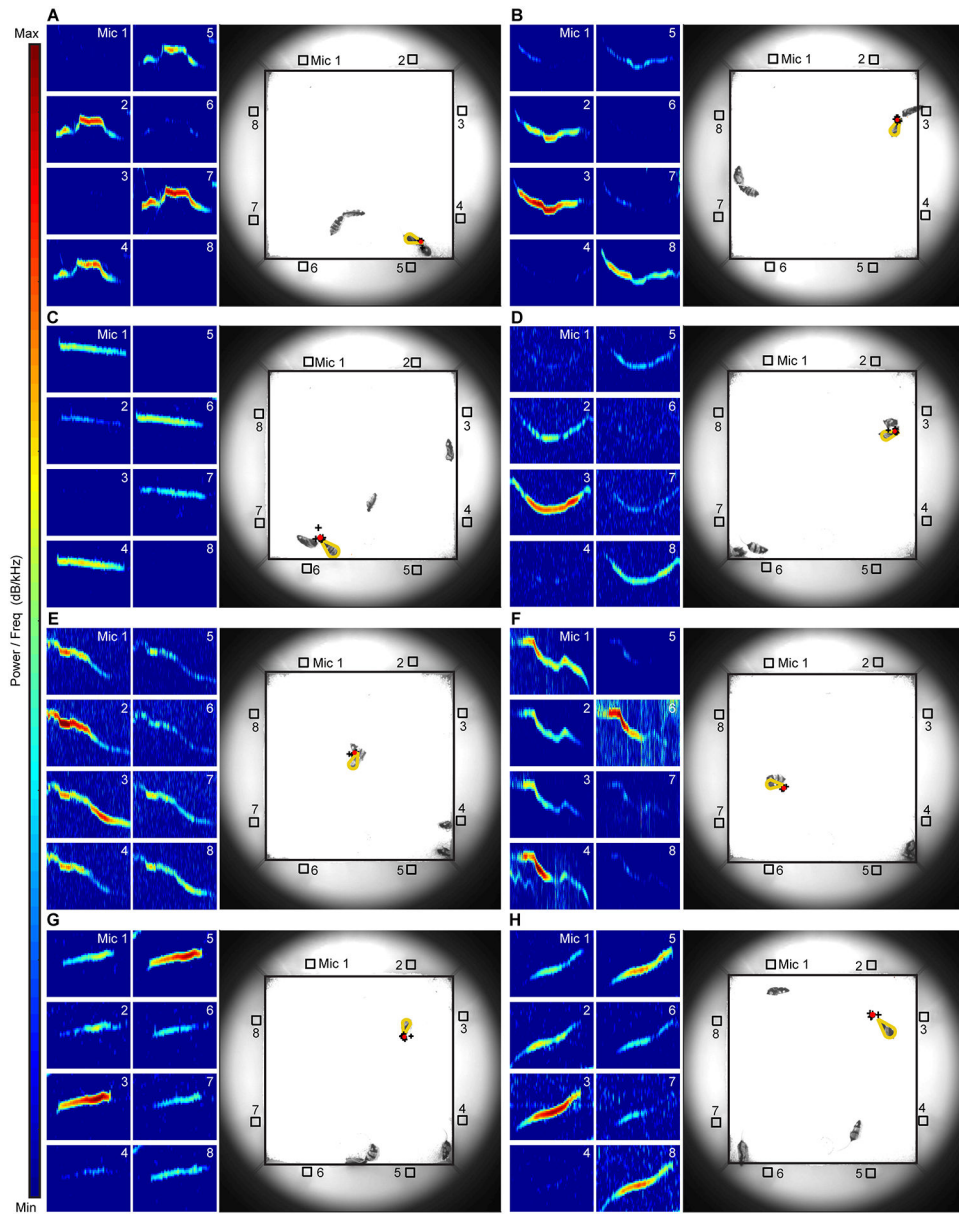
Fig. 2. Localization of vocal signals during social behavior.

A-H, Examples showing signals assigned to male mice. Each signal is detected on microphones in the array and visualized in the spectrograms. Numbers correspond to the microphone recording the signal (Mic 1–8). Intensity is indicated by the color bar. Ordinate shows frequency (range of 40 kHz) and abscissa indicates time. Duration of the examples shown ranges from 11.7–85.2 ms. Photographs show the location of the mice in the recording arena at the time of signal emission. Microphones are represented by open-faced squares and numbered 1–8. Vocal signals were probabilistically assigned to individual mice (outlined in gold) based on time delays observed between all possible microphone pairs. A total of 8 point estimates (each shown with a black +) were generated by localizing the signal using 8 different combinations of microphones, where each combination eliminated one microphone from the localization process. Because of the precision of the system, full

visibility of each of the 8 plus signs was often precluded. The point estimates were used to determine the location of the sound source (red dot). Localization of each vocal signal was repeated independently with similar results in all 11 experiments.
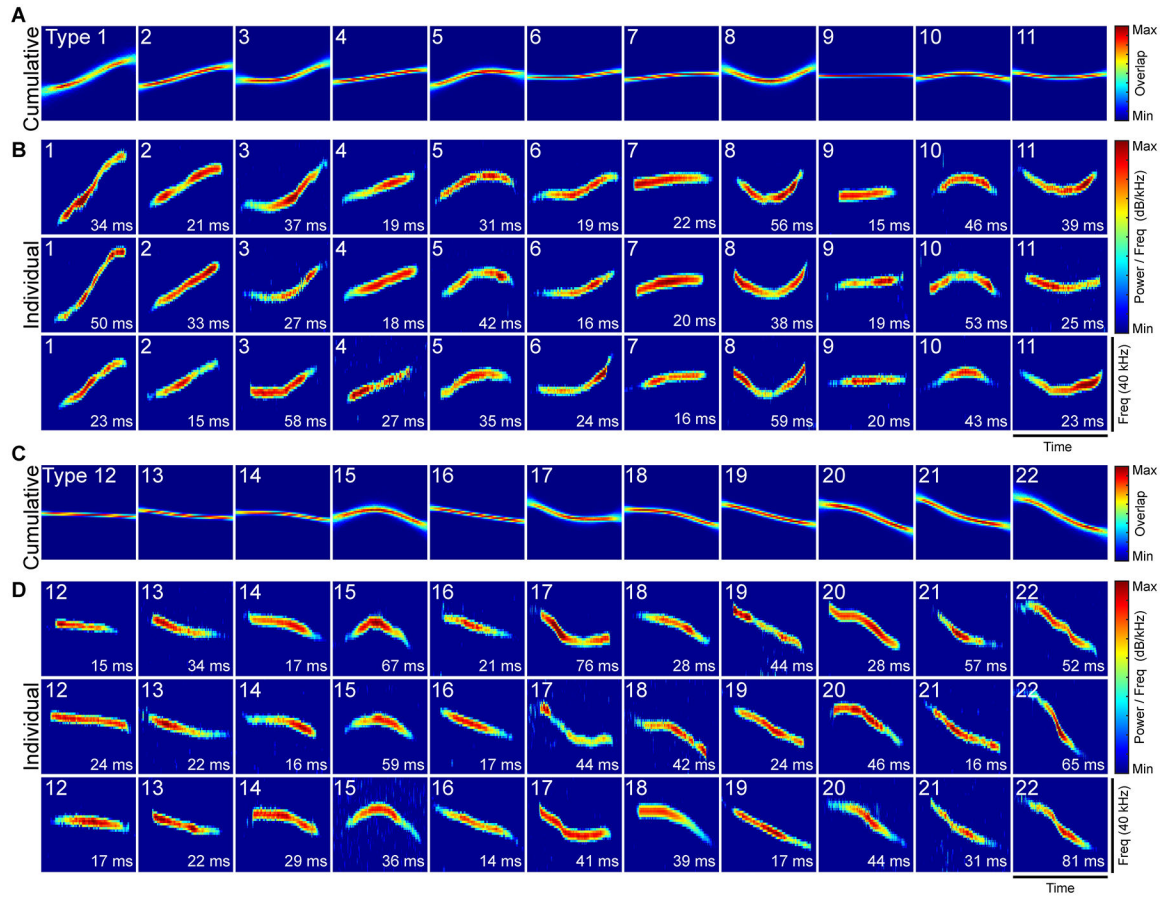
Fig. 3. Categories of vocal signals.

The clustering algorithm grouped all vocal signals that were recorded during the 11 recordings into 22 different types. Vocal signal types sorted by changes in pitch in descending order. General spectral features of each signal type are shown in the cumulative plots (**A,C**) and spectrograms of three representative vocal signals (**B,D**). Cumulative plots were created by overlaying the normalized frequency contours of each of the vocal signals that were classified as a particular type. Maximum and minimum signal overlap is shown in red and blue, respectively. The duration of each individual example is shown in lower right corner. Intensity is indicated by color (red = maximum, blue = minimum).
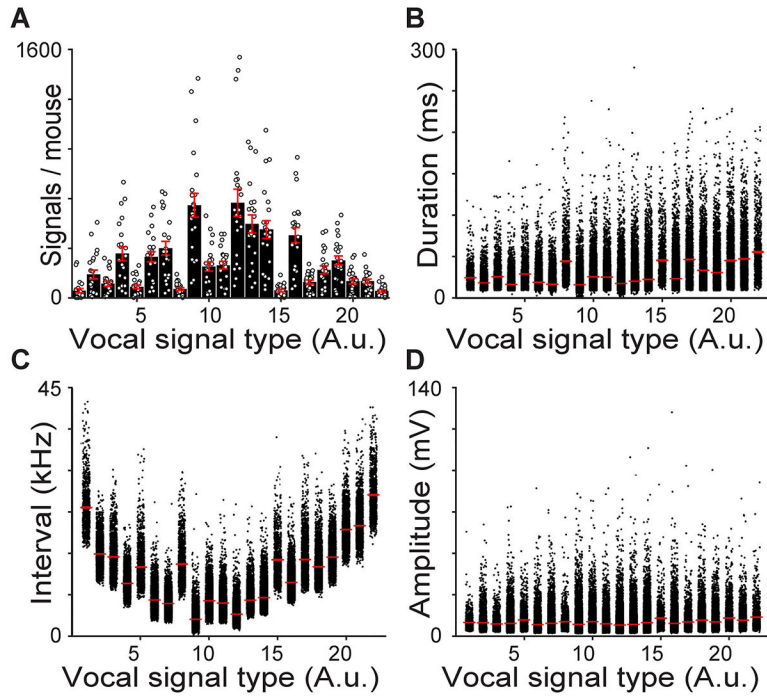
Fig. 4. Quantification of vocal repertoires.
A, Distributions of vocal signals clustered by shape (Average±SEM; n=484 biologically independent samples, 1-way ANOVA, F=21.0, p<10^{-54}). Average±SEM of the duration (**B**), interval (**C**), and amplitude (**D**) for each vocal signal type. **B-D**, n=111,498 biologically independent samples, 1-way ANOVAs, all F>14.2, all p<10^{-38}, red line = average, width of line = ±SEM
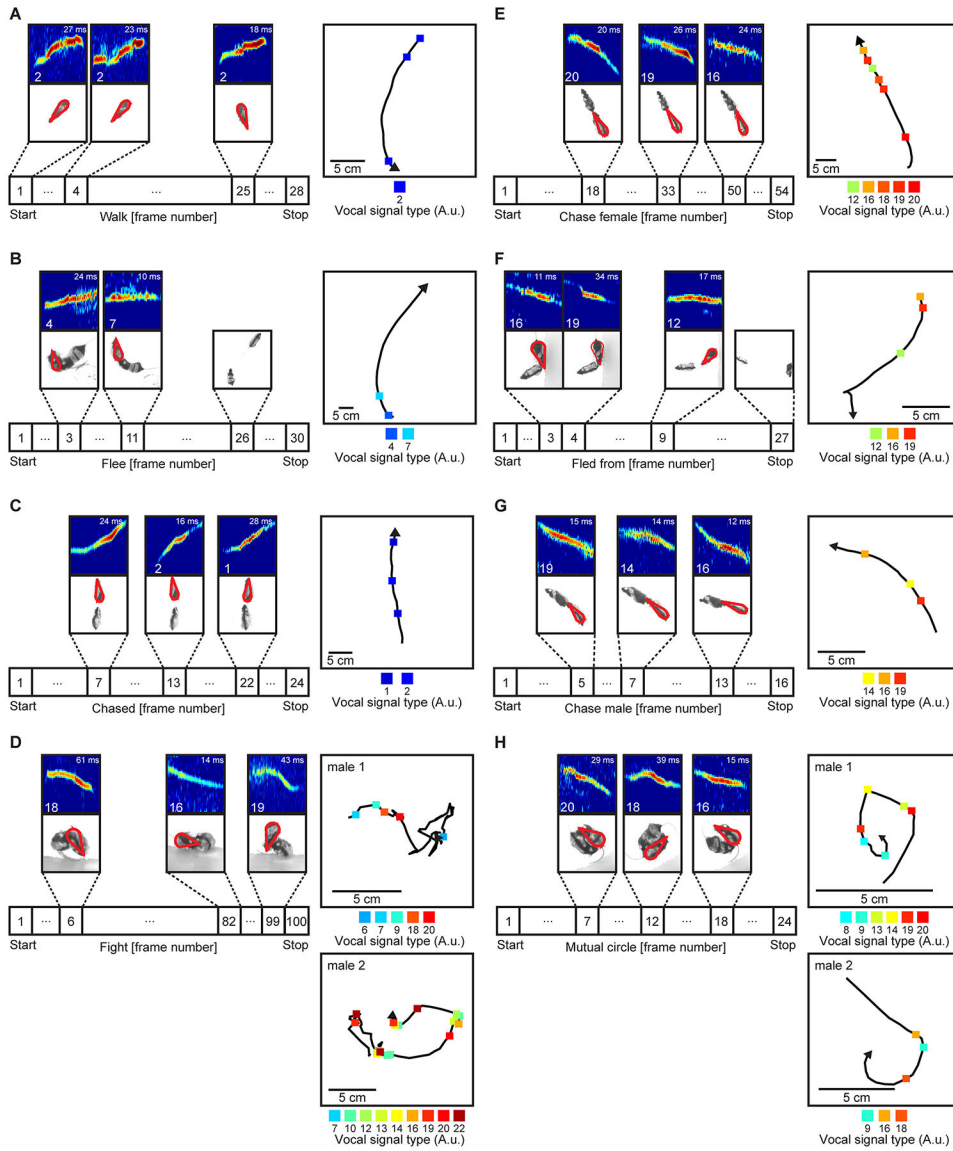
Fig. 5. Examples of vocal emission during distinct behaviors.

A-H, Vocal signal expression from individuals during each of the eight innate behaviors. Similar results were repeated independently in all 11 experiments. Timeline shows the position and type of signals emitted by specific mice during the initial, middle, and final stages of the behavior. Vocalizing mouse indicated by red outline. Behaviors in which two mice participated are auto-scaled such that both mice are visible. Signal type and duration indicated in bottom left and top right of spectrogram, respectively. Inset shows trajectory of an individual animal (arrow and squares, direction of movement and type of vocal signals emitted, respectively). During final stages of flees (**B,F**), no vocal activity is shown because the mice were silent. For fight (**D**) and mutual circle (**H**), each inset shows the trajectory and vocal activity of one of the two mice engaged in the behavior.
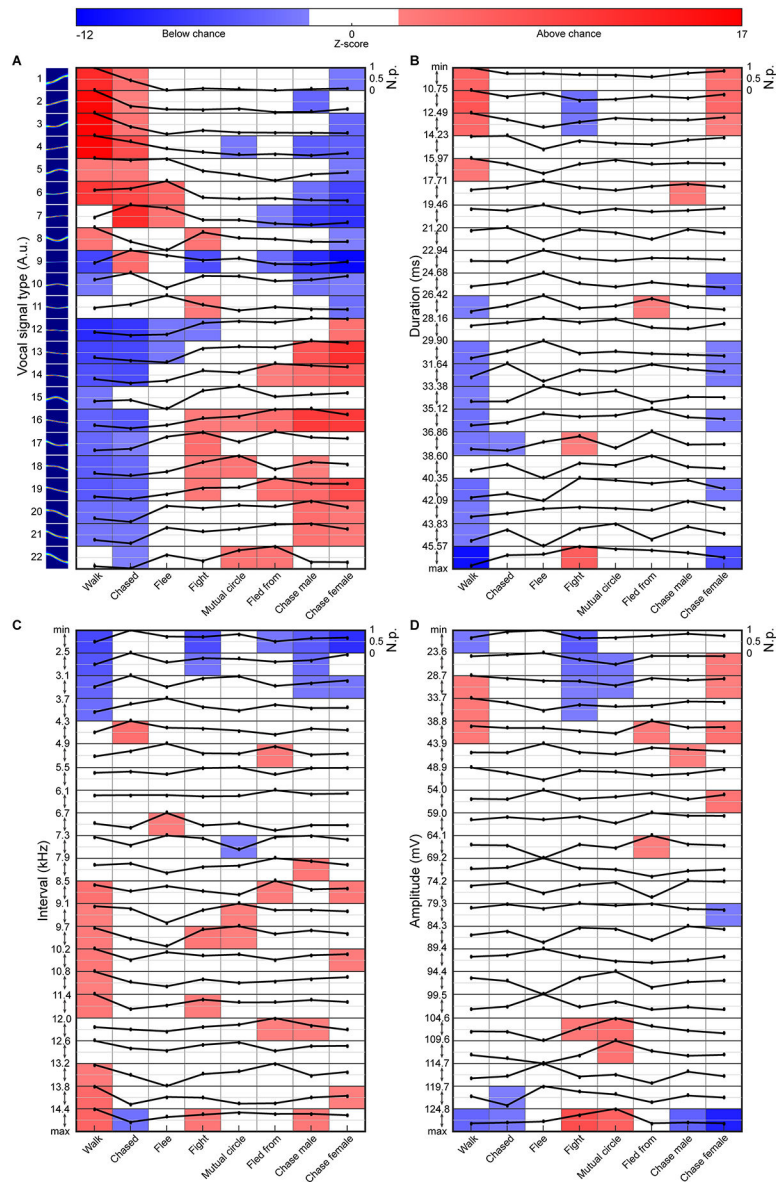
Fig. 6. Behaviorally-dependent vocal emission.
Differential expression of social context-dependent vocal expression indicating when signal types were emitted above (red), below (blue), and at chance (white). Chance level heat maps overlaid with black line indicating normalized probability (N.p.) of signal type emission. Signals categorized based on shape (**A**), duration (**B**), interval (**C**), and amplitude (**D**). **A-D**, 2-sided permutation tests, n=1000 independent permutations for each analysis.
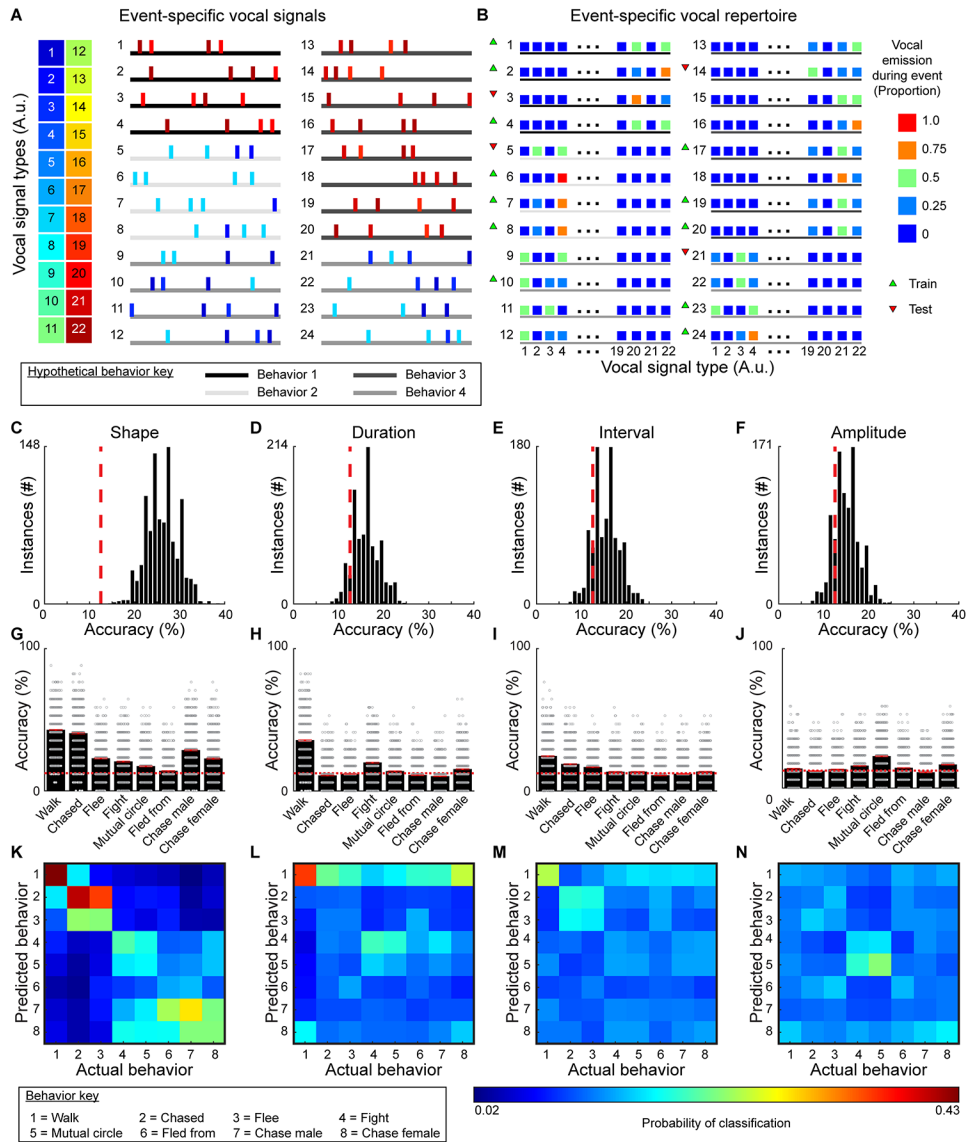
Fig. 7. Decoding behavior based on vocal emission.

A, Schematized vocal emission during 24 hypothetical events for 4 (of 8) different behavior categories (see Supplemental Figs 12, 13 for real events from all categorized behaviors). Horizontal lines denote 4 of the 8 different behaviors, with each behavior indicated by a shade of gray. **B**, Using the proportion of each vocal signal type emitted during each behavioral example, a multi-class support vector machine was independently trained and tested on random subsets of the data. **C-F**, Distributions of classifier accuracies when vocal signals were categorized by shape (2-sided permutation test, n=1000 independent permutations, z=−3.8, p<10^{-4}), duration (z<−1.3, p>0.2), interval (z=−1.0, p>0.3), and amplitude (z=−0.9, p>0.3). Chance levels denoted by red dashed lines. **G-J**, Classifier accuracy for each behavior when signals were grouped based on shape, duration, interval, and amplitude (red line = average, width of line = ±SEM, red dashed lines = chance). For each analysis, the classifier was run 1000 times and every iteration predicted 17 examples of

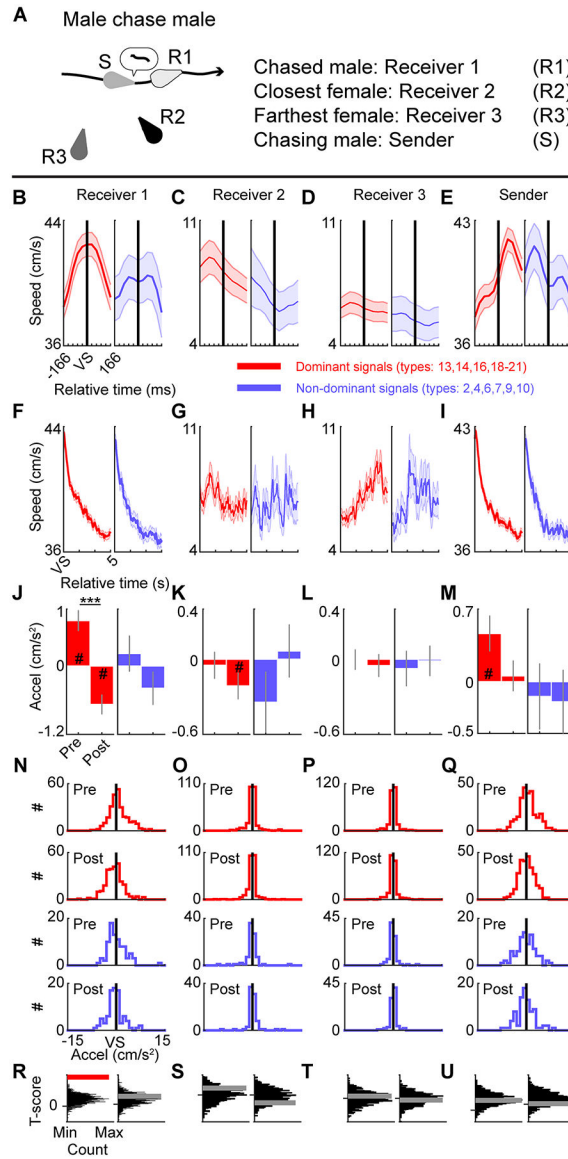each behavior with accuracies shown as a dot. **K-N**, Confusion matrices showing classification errors.

Fig. 8. Vocalizing alters behavior of an engaged social partner.
A, Schematic of a male mouse (Sender) chasing the other male (Receiver 1) as well as the closest (Receiver 2) and farthest (Receiver 3) females. **B-E**, Average speed±SEM for Receiver 1 (**B**), Receiver 2 (**C**), Receiver 3 (**D**), and Sender (**E**) before and after the emission of dominant (red) or non-dominant (blue) vocal signals (VS). There were 271 dominant and 86 non-dominant vocal signals included in the analyses. **F-I**, Average speed±SEM in the 5 seconds after the emission of dominant or non-dominant vocal signals. **J-M**, Acceleration of animals (Average±SEM; # indicates that an animal was accelerating (2-sided 1-sample t-test, all significant T>=2.73, all significant p<0.01) or decelerating (2-sided 1-sample t-test, all significant T<=−2.04, p<0.05); *** denotes significant differences in pre- and post-acceleration, 2-sided paired t-test, T=5.6, p<10^{−8}). **N-Q**, Full data distribution for **J-M**. **R-U**, Randomly generated distributions of T-scores comparing pre- and post-acceleration calculated for signals that were randomly shifted to different times within a chase. T-score

distributions were compared to the T-scores calculated in **J-M**. Red horizontal bar in **R** indicates above chance, 2-sided permutation test, n=1000 independent permutations, p<0.001; gray horizontal bars in **R-U** indicate chance, 2-sided permutation tests, n=1000 independent permutations, all p>0.27. Panels: left and right, dominant and non-dominant signals, respectively.

Table 1.

Behavioral definitions.

Behavior	Definition
Walk	A mouse actively moves around the cage and no other mouse could be within 35 cm.
Mutual circle	Two males put their heads near the anogenital region of each other and spin in a circle.
Fight	Both of the males engage in physical contact. This behavior involves biting, wrestling, and rolling over each other.
Male chase female	A male follows a female while the two mice are within 2 body lengths of each other.
Flee	One male runs away in response to an action from another male.
Fled from	The male that the fleeing male is escaping.
Male chase male	A male follows another male while the two mice are within 2 body lengths of each other.
Male being chased	The male in a chase that is being followed.

Table 2.

Vocal signal emission during behaviors.

Behavior	Detected			Assigned			Unassigned		
	#	Rate	Rank	#	Rate	Rank	#	Rate	Rank
Walk	22466	0.76	6	16619	0.55	6	5847	0.21	6
Mutual circle	1496	10.4	1	1067	7.40	1	429	3.03	1
Fight	12212	3.46	3	7334	2.07	3	4878	1.39	2
Male chasing female	30609	4.50	2	21275	3.14	2	9334	1.36	3
Fleed from/fleeing *	700	1.02	5	484	0.70	5	216	0.32	5
Male chasing/being chased *	6873	2.49	4	4832	1.76	4	2041	0.73	4

*
Behaviors are grouped together because either animal may have emitted an unassigned signal.

Table 3.

Vocal Signal Categories

Figure	Vocal Signal Type
Fig. 2A	15
Fig. 2B	17
Fig. 2C	14
Fig. 2D	8
Fig. 2E	22
Fig. 2F	21
Fig. 2G	7
Fig. 2H	2