Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Kevin Xu April 7, 2020

Context and Complexity in Mouse Vocalizations

by

Kevin Xu

Gordon J. Berman
Adviser

Department of Quantitative Theory and Methods

Gordon J. Berman

Adviser

Robert Liu

Committee Member

Astrid Prinz

Committee Member

Michal Arbilly

Committee Member

2020

Context and Complexity in Mouse Vocalizations

By

Kevin Xu

Gordon J. Berman

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Department of Quantitative Theory and Methods

2020

Abstract

Context and Complexity in Mouse Vocalizations
By Kevin Xu

Many species utilize acoustic communication for different ethological purposes, yet little is known about which features of these vocalizations convey important social information. These acoustic features may be context-dependent, varying across different social situations and having different communicative effects across dissimilar contexts. Mouse ultrasonic vocalizations (USVs) are exhibited across a variety of social contexts and have been shown to provide information about the animals' future behaviors. Previous studies have claimed that mouse vocalizations have characteristics of birdsong-like syntax, including distinct syllable types and repeated temporal sequencing, and that males modify their syntax according to different social stimuli (e.g. males emit more complex syllables and sequences in response to urine compared to female presence). Here, using a large data set of mouse vocalizations and a novel computational method for characterizing the space of vocal features, we reassess the results from these studies to better understand the complexity of vocal behavior in different social contexts. Our analyses show how our previous understanding of mouse vocal behavior may have been a consequence of the choices for quantifying behavior, rather than a general statement about the potential social content. We define vocal behavior in coarser and finer-grained scales, looking at the effects on syntax and temporal structure. Overall, this study provides insight into the complexities of mouse vocalizations in different social contexts. By understanding the different effects of how we describe and measure vocal structures, we can further advance how to analyze vocal repertoires across multiple scales.

Context and Complexity in Mouse Vocalizations

By

Kevin Xu

Gordon J. Berman

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Department of Quantitative Theory and Methods

2020

Acknowledgements

I would like to thank Dr. Gordon Berman for his tremendous support throughout the past few years, for teaching me the necessary skills to complete this project and guiding me during the writing process.

Table of Contents

Context and Complexity in Mouse Vocalizations

Kevin Xu[1], Kelly M. Seagraves[2], S. E. Roian Egnor[2], Gordon J. Berman[1]

[1]Emory University Department of Biology, [2]HHMI Janelia Research Campus

Introduction:

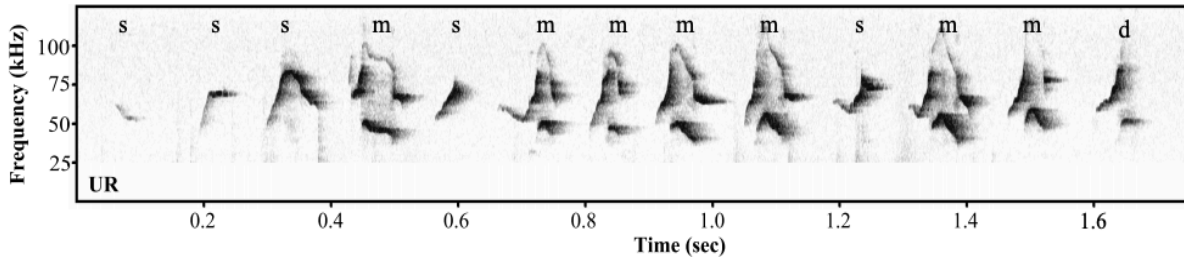
For many species, including humans, vocal communication is fundamental for social interaction and is an important aspect in determining behavior. The structure of communication varies amongst different species, from the fixed utterances of cricket chirps (Doherty and Hoy, 1985) and roars of goitered gazelles (Blank et al., 2014), to the semantic structure of songbirds (Marler, 1990; Berwick et al., 2011), to the complex language of modern humans. Acoustic communication is essential for survival, utilized for a variety of functions including mating (Blank et al., 2014) and courtship (Bradbury, 1998; Behr and von Helversen, 2004). In mice, acoustic communication is emitted in the form of ultrasonic vocalizations (USVs), which are produced in various social contexts — not only for mating and courtship (Neunuebel et al., 2015), but also for social reunion, juvenile play, and territorial defense (D'Amato et al., 2005; Panksepp et al., 2007; Chabout et al., 2012; Hammerschmidt et al., 2012; Petric and Kalcounis-Rueppell, 2013).

Although many studies have focused on the contexts in which animals produce ultrasonic vocalizations, little is known about what features of these vocalizations convey important social information. Holy and Guo (2005) first advanced the idea that mouse USVs have characteristics of birdsong-like syntax, such as discrete, categorized syllables, whose temporal structures include regularly repetitive phrases (Holy and Guo, 2005). It has also been hypothesized that features such as spectral content, typical frequency, duration, and repetition period are informative to the production and perception of USVs (Liu et al., 2004). Liu et al. (2004) found significant differences in these acoustic features between isolated pup calls

and adult encounter calls, thus showing potentially informative cues that a receiver, or an adult female mouse, can use to recognize and distinguish between. Further studies have claimed that mouse USVs are context-dependent and can provide information about the animals' future behaviors (Chabout et al., 2015; von Merten et al., 2014; Hanson and Hurley, 2012; Yang et al., 2013; White et al., 1998). For example, German male mice increased their rates of vocalizations when exposed to female mice within the same population, then decreased their rates when exposed to female mice from a different population and other male mice of the same population (von Merten et al., 2014). Similarly, male mice increased their average number of syllables initially after the removal of a female from the cage (Hanson and Hurley, 2012; Yang et al., 2013). Male mice also produced more calls shortly before and during mating (White et al., 1998) and increased USV production prior to courtship behaviors, such as mounting and sniffing of females (Hanson and Hurley, 2012).

Recently, to better understand these important social features of mouse communication, many studies have focused on investigating syllable types and syntax (Holy and Guo, 2005; Hanson and Hurley, 2012; Portfors, 2007, Perrodin et al., 2020). For instance, Hanson and Hurley (2012) found differences in the percent use of several syllable types for males during and after male-female interaction. Perrodin et al. (2020) looked at how temporal reversal of male mice song and randomization of syllable sequences affected female approach behavior. One study (Chabout et al., 2015) also examined the complexity of syllable types and syntax. They claimed that adult male mice emit more complex syllables and sequences in response to the context of female urine compared to female presence. Utilizing single contours from spectrograms, they classified syllables into four categories based on the presence or absence of instantaneous pitch jumps, or abrupt frequency discontinuities: (1) simple syllables with no pitch jumps ("s"); (2) syllables separated by a single upward ("u") or (3) downward pitch jump ("d"); and

(4) complex syllables with multiple pitch jumps (“m”). Examples of syllables are shown below in the following spectrogram (Chabout et al., 2015).



Analyzing the proportion of syllables in each context, they found that male mice exposed to female urine produced 2.3 times the amount of complex “m” syllables compared to male mice exposed to the female presence context. Additionally, they measured differences in syntax between contexts, finding that male mice produced more transition types to the complex “m” syllable in the urine context than in the female context (Chabout et al., 2015). Through these results, they claimed that complexity was an important feature of acoustic communication that is dependent on social context.

This method of defining complexity from a) the acoustic complexity of the recorded syllables, and b) the transition structure between them, however, could potentially be a result of the particular methodology chosen, rather than a fundamental observation about the animals’ behavior. There’s a difference between spectral complexity, such as how complex a spectrogram looks, versus the complexity of language. In other words, mice in one context may emit more spectrally complex syllables, but this doesn’t necessarily mean they are conveying more complex information. Additionally, the methodology of analyzing complexity from manual definitions of syllable types could lead to different findings depending on the particular choice. There is still considerable debate regarding what syllable categories exist within the mouse vocal repertoire — or whether such “syllables” even exist in the first place (Goffinet et al., 2019). Many studies disagree on the identification of discrete syllable types, with some proposing three (Hammerschmidt et al., 2012), four (Chabout et al., 2015), seven (Holy and Guo, 2005),

ten (Chabout et al., 2012), and even eleven (Grimsley et al., 2011) types. This spectrum of categories could all lead to different results about mice vocal complexity, as complexity could be defined as any syllable type different than the multi-jump ("m") type from Chabout et al. (2015). Perhaps utilizing ten syllable types instead of four will result in a lack of differences in complexity between different social contexts. Even in von Merten et al. (2014), complexity was not defined based on the categorization of syllable types, but rather from the slopes of the frequency contours from the spectrograms. Due to this variability in defining vocal behavior, we looked to develop a new method for analyzing complexity that does not depend on prior assumptions of syllable types or prior categorizations of behavior.

Here, using a large data set of mouse vocalizations and a novel computational method for characterizing the space of vocal features, we reassessed the results from previous studies to better understand the complexity of vocal behavior in different social contexts. By developing a more detailed, consistent method of analyzing the complexity of mouse USVs in different social contexts, we account for discrepancies in describing complexity, and we are able to better understand contextual acoustic communication. Using vocalizations from adult male mice that are exposed to female presence and female chemosensory cues, we first defined vocal behavior in coarse and fine-grained scales. To quantify the structure of the mouse vocal repertoire without making any assumptions about the existence or number of syllable types, we placed each vocalization in a high dimensional space based on frequency contour similarity. We then projected that space down to two dimensions using t-distributed Stochastic Neighbor Embedding (t-SNE), which has been used successfully to describe behaviors such as fly locomotion (Berman et al., 2014). A computational watershed segmentation defined the vocal behavioral space into fine-grained regions, and manual labelling was used to define the behavior into coarse-grained scales. We analyzed the vocal composition between the two social contexts and explored the distinguishability of the distribution of vocalizations between and within contexts. Then, using both

coarse and fine-grained scales, we reassessed the repertoire and temporal complexity of the mouse vocalizations, comparing that with previous literature.

This approach allowed us to analyze the complexity of mouse USVs in different social contexts without depending on the precise methodology of defining our syllable types beforehand. Our novel methods allowed us to understand complexity more quantitatively, creating more consistent analyses. Here, we found congruent results with previous studies (Chabout et al., 2015) related to the composition of syllables between the two contexts; however, we found different results for complexity, when viewing the notion of complexity in a more holistic manner. By understanding the different effects of how we describe and measure vocal structures, we can further advance our understanding of rodent vocal repertoires across multiple scales and between different contexts.

Results:

We recorded a total of 21032 USVs from twelve male mice in each of the following two social conditions: exposure to female urine odor (Odor condition: 10099 total vocalizations, Figure 1a) and exposure to female presence (Female condition: 10933 total vocalizations, Figure 1b) (for setup, see Methods).

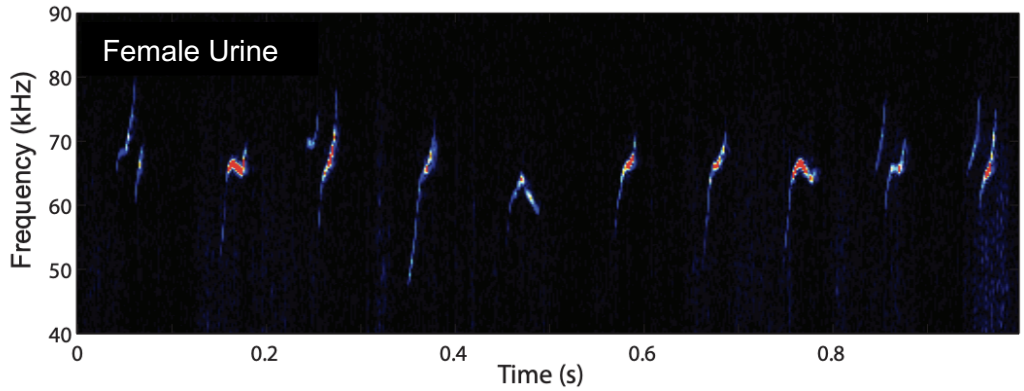


Figure 1a. *Spectrogram of male mouse USVs upon exposure to female mouse urine (figure courtesy of S.E.R Egnor and K. Seagraves)*

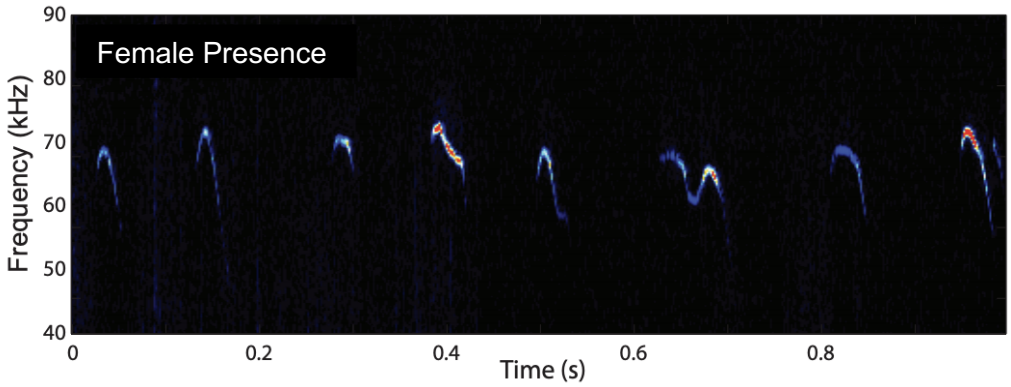


Figure 1b. *Spectrogram of male mouse USVs upon exposure to female mouse (figure courtesy of S.E.R Egnor and K. Seagraves)*

For each vocalization, we automatically extracted the frequency contour, or the dominant frequency as a function of time. To represent every vocalization emitted in both social contexts in the same space, frequency contours were mean-frequency-subtracted, and all pairs of frequency contours were

compared using dynamic time warping to create an all-to-all distance matrix ((21032 x 21031)/2 comparisons). To visualize this high dimensional space, we projected down onto two dimensions using t-distributed Stochastic Neighbor Embedding (t-SNE), forming a map of the acoustic structure of the entire mouse vocal repertoire (for details, see Methods).

Behavioral Map:

Figure 2. *Two-dimensional probability-density function (PDF), generated from embedding all data points via t-SNE and convolving with a Gaussian watershed transformation. Fine-grained peaks correspond to distinct stereotyped vocal behaviors. Frequency contours show local distribution of vocal behaviors around the map.*

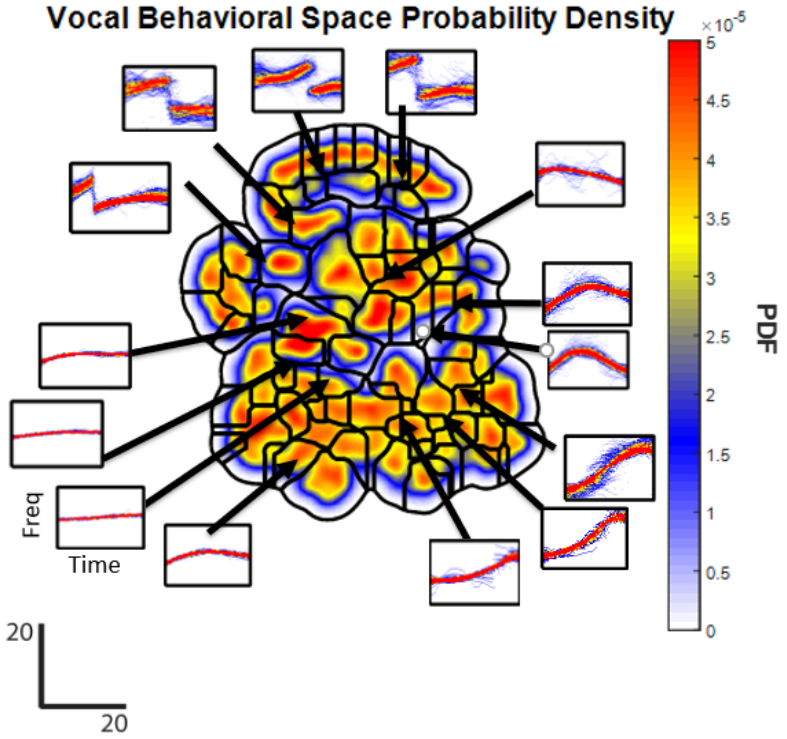


Figure 2 displays the output of the t-SNE embedding, a two-dimensional behavioral map of the entire mouse vocal repertoire across all experimental conditions. A Gaussian watershed algorithm was used to separate local maximum peaks to segment the vocal repertoire into discrete vocalizations, thus creating fine-grained separations in the vocal space. As shown above, 86 unique watershed regions were produced in the behavioral map. Each region contains a different vocal behavior, defined from features of the frequency contours. Regions of red indicate vocal behaviors of higher probability densities, while regions of blue indicate vocal behaviors of lower probability densities. Figure 2 also shows several frequency contours corresponding to specific watershed regions around the map. The distribution of frequency contours

shows localized organization of the vocalizations in the fine-grained two-dimensional space, thus allowing us to make observations on the structure of the mouse vocal repertoire.

Manually Labeled Behavioral Map:

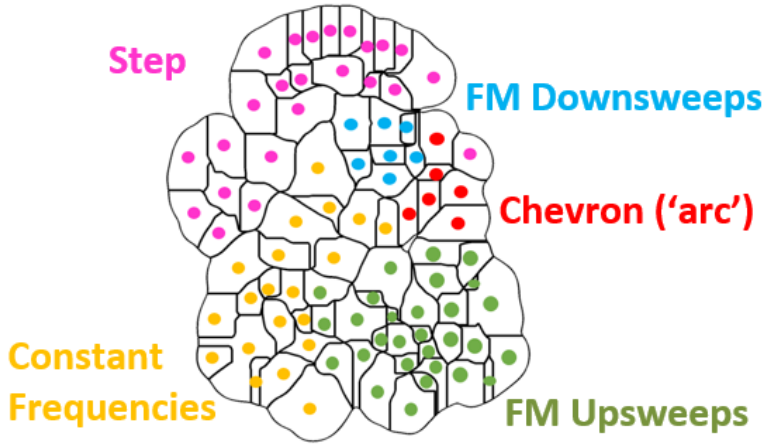


Figure 3. *Manually labeled behavioral map of stereotypical syllable types. Labels correspond to regional frequency contours shown in the previous figure (pink=step, cyan=FM downsweep, red=chevron, yellow=flat (constant frequency), green=FM upsweep). Manually labeled regions correspond to coarse-grained separations of vocal space.*

To examine the vocal space utilizing coarse-grained separations, manual labels were used to define the map (Figure 3). All 86 frequency contour types from the previous figure were structurally examined and labeled accordingly to five stereotypical syllable types: frequency modulated upsweeps (FMU), frequency modulated downsweeps (FMD), constant frequency (CF), chevron (CH, or arc), and steps (ST) (see Methods). Although different studies disagree on the exact number of discrete syllable types, these five align most commonly across many labs. They are important in defining stereotypical features in mice vocal communication; thus, upon labeling each frequency contour, all watershed regions were manually transformed according to their syllable types, generating coarse-grained behavioral separations with 5 unique regions, contrary to the 86 fine-grained regions in Figure 2. In Figure 3, we see that syllable types are distributed adjacently on the map. Regions in the bottom right corner are primarily frequency modulated upsweeps, regions in the bottom left corner are constant frequencies, regions in the top left are step, and regions in the top right appear to

be a combination of frequency modulated downsweeps, chevron, and step. Generating both coarse-grained (Figure 3) and fine-grained (Figure 2) vocal behavioral maps ensure the use of multiple scales, eliminating any nuanced effects that primarily arise due to the consequences of defining behavior.

Information about social context in relation to mouse USVs:

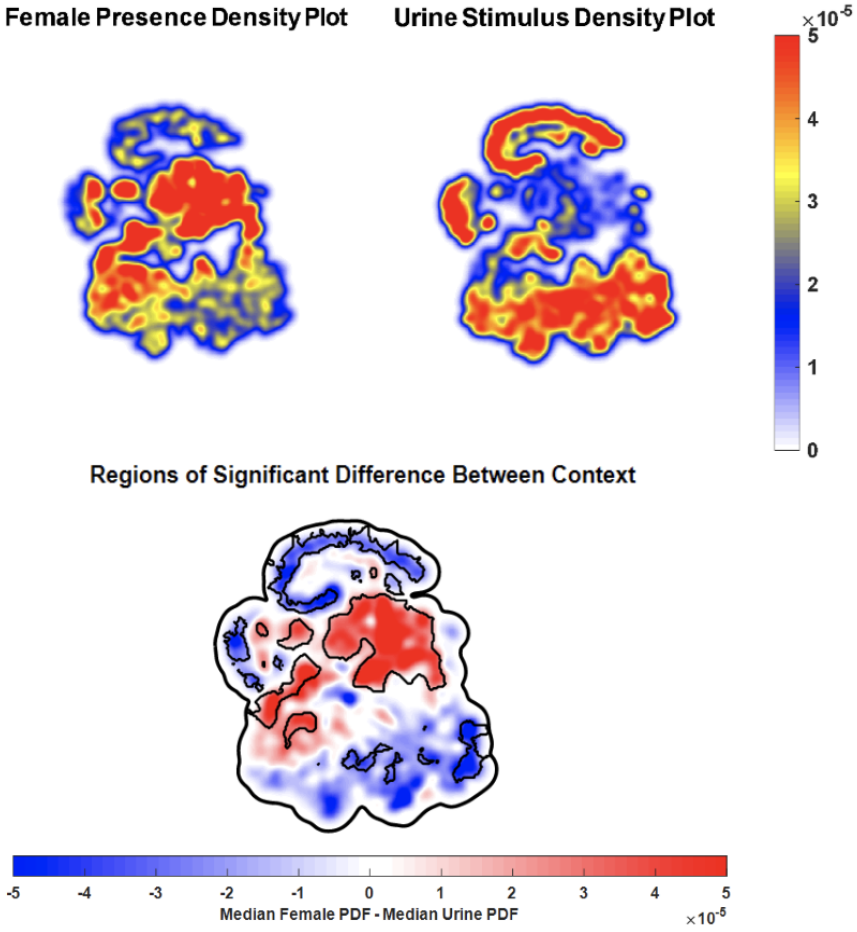


Figure 4a*. Distinct male ultrasonic vocal repertoires elicited by female presence and female chemosensory cues. Probability density of separated vocal repertoire shows regional differences based on social context.*

Figure 4b. *Probability density difference between vocalizations from the female-presence and urine odor contexts. Regions circled are regions with bootstrapped p-values less than .05. Differences in the map reveals significant differences in the composition of acoustic communication between the two social contexts.*

To examine how vocalizations vary as a function of social context, we isolated vocal repertoire produced in response to the female presence and in response to the female urine odor, then plotted them on the vocalization map (Figure 4a). Although vocalizations from all regions were produced in both contexts, there were significant differences in the composition of vocalizations produced (Figure 4b) (see Methods). Based on the manually labeled map, step vocalizations and frequency modulated upsweeps

had higher probability densities in the chemosensory urine context, while chevron, frequency modulated downsweeps, and constant frequencies were more common in the female presence context. The composition of step vocalizations being significantly greater in the urine odor context aligns with the results of Chabout et al. (2015), in which they also measured significantly higher proportions of multi-jump "m" syllable types in the urine context compared to the female presence context. Additionally, our results aligned with previous results from other labs, indicating that male mice do indeed communicate differently depending on the social context.

Between and Within Context Findings:

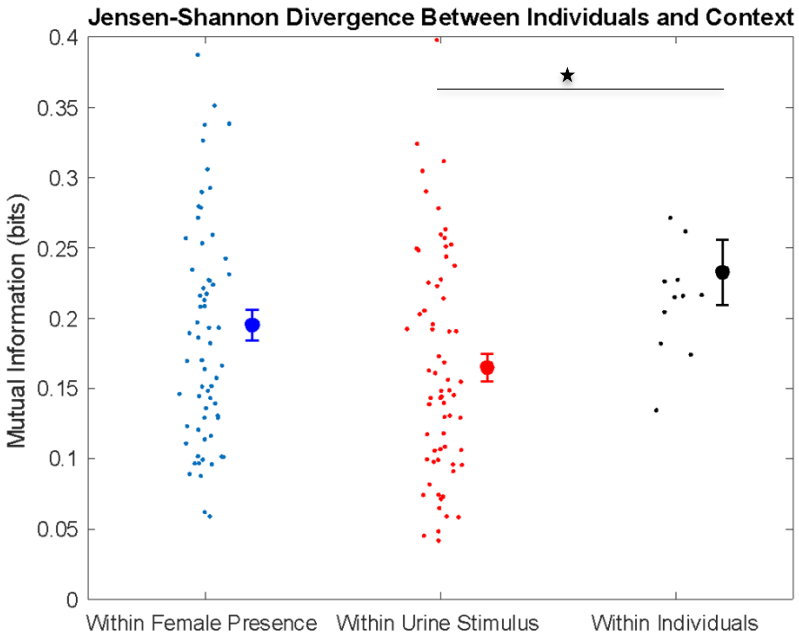


***Figure 5**. Using Jensen-Shannon Divergence, vocalizations between contexts and within individuals (black) are more distinguishable compared to vocalizations between individuals within a context. Asterisks denote significant differences (rank sum test; p<0.05).*

We then investigated vocalizations between and within contexts using the Jensen-Shannon Divergence (see Methods). The Jensen-Shannon Divergence (JS-D) calculates the mutual information between samples from two distributions and the identity of which distribution each sample was drawn from. Thus, it can be interpreted as the distinguishability between two distributions and is measured in units of bits. A Jensen-Shannon Divergence of 0 indicates no distinguishability, while a J-SD of 1 indicates that the chosen distribution can be uniquely identified from a single sample. In the case of the mouse vocal

repertoire, vocalizations within the same context and between individuals were first analyzed. In other words, within the female presence and female urine contexts, the distribution of vocalizations from mouse 1 was compared to those from mouse 2, then the distribution between mouse 1 and mouse 3, mouse 1 and 4, and so on for all combinations of individuals. Afterwards, vocalizations within the same individual and between contexts were analyzed. For instance, for just mouse 1, the distribution of vocalizations between the two social contexts was compared, then the same for mouse 2, and so on for all individual mice. Figure 5 displays the distinguishability in distributions for all comparisons. On the left side in blue, mutual information within the female presence context and between individuals is shown, with its corresponding average and standard error. Likewise, mutual information within the urine odor context, between individuals, is displayed in the red points in the middle of Figure 5. Finally, mutual information within individuals and between contexts is shown on the right in black. The greatest average of mutual information is vocal repertoire within individuals and between contexts (black). This evidence suggests that we can distinguish vocalizations between contexts more easily than we can distinguish between individuals.

Analysis of Vocal Complexity:

To reiterate, past observations on vocal complexity may not have been a fundamental observation of mice behavior, but rather a result of the particular method chosen. For example, in Chabout et al. (2015), complex vocalizations were defined as those with multi-jump syllable types (“m”), while simple vocalizations were labelled as those without jumps (“s”). This methodology of manually defining complexity from spectral features may not be indicative of complex information. Thus, we developed a more quantitative, consistent method of analyzing complexity utilizing two new definitions: repertoire complexity and temporal complexity.

Repertoire Complexity: (See Methods)

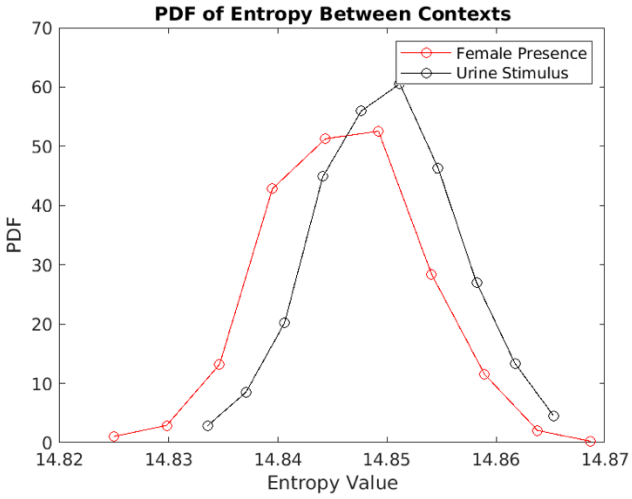


Figure 6*. Probability density function (PDF) of the entropy values between contexts using bootstrap analysis. Large overlap of distributions indicates similar complexity.*

To investigate repertoire complexity in the two contexts, entropy was calculated (see Methods for calculations). Entropy refers to the uncertainty of data and can be utilized to determine the number of binary distinctions in a dataset (i.e. the number of partitions in which you could break the data into), thus directly corresponding to its complexity. If the entropy value is equal to h, then there are 2^h number of distinctions in a data set (Cover and Thomas, 2006). When entropy has a value of zero, only one uniform outcome will occur, and when entropy has a greater value, this will lead to a distribution of results with varied probabilities of appearance. Above, Figure 6 shows a PDF of the entropy values between vocalizations from the female presence and urine odor contexts. To obtain this plot, the vocal repertoire for each context was bootstrapped 1000 times, with random samples of 10,000 vocalizations. Overall entropy was then calculated for each bootstrap. In Figure 6, we see that the PDF displays a large overlap of the entropy distributions, indicating that vocal complexity is actually similar in each context.

Furthermore, entropy values were then investigated on an individual level (see Methods). In Figure 7a, the plot on the left shows a paired plot with average entropy values for each individual's bootstrapped vocal repertoire, separated by social context. Bars around each average entropy value indicate standard error and the blue line connects the individual mouse. A Wilcoxon Signed Rank Test for paired samples was performed on the individual entropy values between contexts, resulting in a

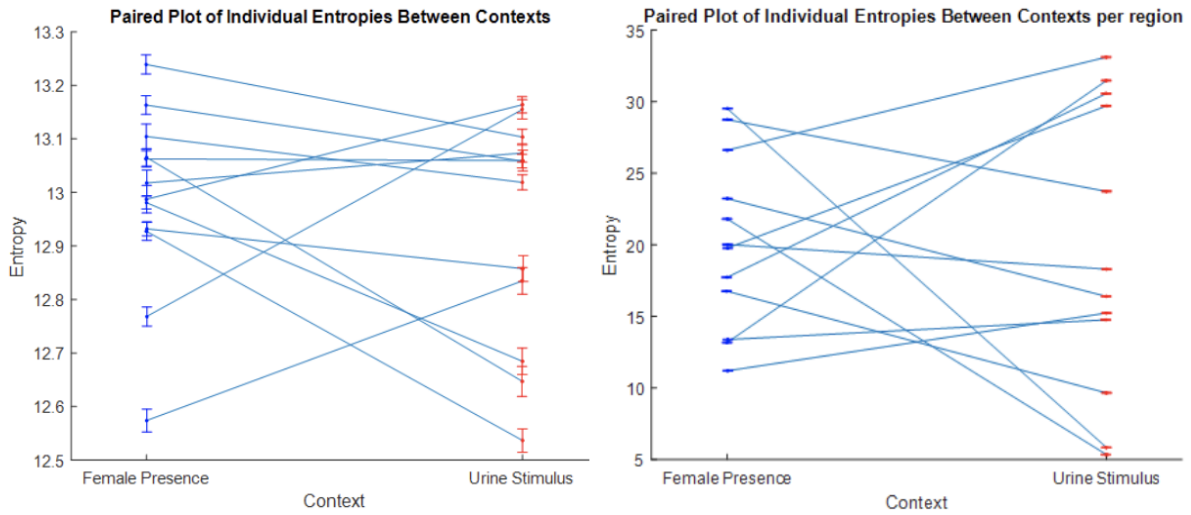


***Figure 7a**. Left paired plot shows entropy values per individual mouse between contexts for the overall repertoire (n=12 mice; Wilcoxon Signed Rank test for paired samples; p>>0.05). **Figure 7b**. Right paired plot shows entropy values per individual mouse between contexts per watershed region (n =12 mice; Wilcoxon Signed Rank test for paired samples; p>>0.05; 86 watershed regions). No significant differences were found between the individual entropy values in each context.*

p-value greater than .05. Thus, no significant differences in individual entropy were measured between contexts. Similarly, in the right plot in Figure 7b, entropy values were investigated individually, yet keeping account of all the watershed regions. For each individual within each watershed region, a bootstrap sample of its vocal repertoire was taken, and its entropy calculated. The average entropy value for each individual within all of the watershed regions was then taken and plotted, as shown in Figure 7b. Again, using the Wilcoxon Signed Rank Test for paired samples to test differences in individual entropy values between contexts, the p-value was significantly greater than .05.

In all three analyses of repertoire complexity, no significant differences were observed for the amount of entropy, or disorder within the vocalizations between contexts. This indicates that, contrary to prior findings (Chabout et al., 2015), the complexity of acoustic communication in male mice does not appear to be significantly different in the context of the female presence and urine odor. To further investigate this idea, analysis was performed utilizing temporal complexity.

Temporal Complexity: (See Methods)

To compare temporal complexity between contexts, we investigated the temporal pattern of vocal behaviors by calculating the behavioral transition matrices (see Methods). Shown below in Figure 8 are the one-step transition probability matrices T ($\tau = 1$) for the entire vocal repertoire between each social context (female presence on left, urine stimulus on right). The unit of tau is the number of calls, thus $\tau = 1$ in the transition matrix represents after one call. The matrices describe the probability of a vocal behavior transitioning from one watershed region of the behavioral map (Figure 2) to the next. It represents the sequence of acoustic communication, as each watershed region constitutes a distinct, stereotyped vocal behavior. Areas of white indicate zero probability of that particular transition, while areas in red indicate higher probability. In both matrices, there appears to be a diagonal structure with higher transition probabilities, indicating that transitions may be localized, with higher probabilities of vocal behavior sequencing within the same or nearby regions.

To observe complexity, the transition matrices were then plotted onto their respective fine-grained behavioral maps (Figure 9). The black lines represent the transition probabilities between each vocal region, and the thicker the line the greater the probability. Figure 9 shows that in each context, the overall complexity of transitions appears equal. There are roughly the same number of transitions in each map, and neither social context indicates any pattern or sequence of communication that is any simpler than the other. However, the transition maps do still appear quite different in composition, as the majority of transitions are mainly localized in separate regions of greater probability density within the context-specific vocalization maps.

One-step transition matrices($\tau = 1$) were then created from the coarse-grained behavioral map from Figure 3. The transition matrices consisted of 5 states, based on the manually labeled syllable type regions. Coarse-grained transition matrices were then plotted onto their context-specific behavioral

maps, shown in Figure 10. Likewise, with the transition maps for fine-grained separations (Figure 9), the composition of transitions differs between contexts in the coarse-grained maps. However, the

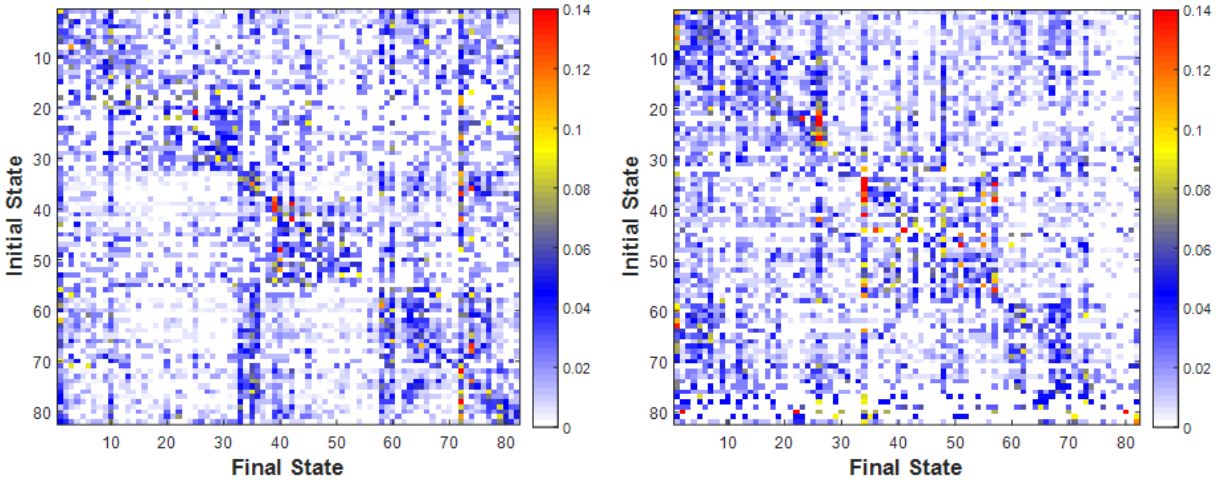


Figure 8. *One-step transition probability matrices T ($\tau = 1$) for all mouse vocalizations within the presence of a female (left) and within the urine odor stimulus (right). Values in the initial and final states represent watershed regions from the fine-grained behavioral map. Areas of white indicate zero probability of a particular vocal transition, and areas of red indicate higher probability.*

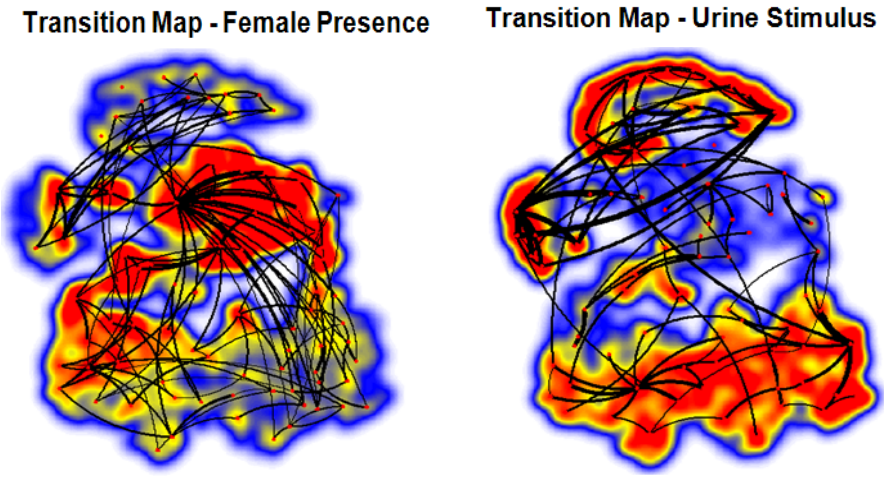


Figure 9*. One-step transition rates($\tau = 1$) plotted on the fine-grained behavioral maps. Each red point represents the maximum of the local PDF, and the black lines represent the transition probabilities between behavioral regions. Line thicknesses are proportional to the corresponding transition values T ($\tau = 1$), and right-handed curvature implies the direction of transition. Composition of transitions differs between contexts, however the complexity of transitions does not. For clarity, all lines representing transition probabilities of less than 0.05 are omitted.*

complexity of transitions looks alike in each map. The number of transitions is about the same, and neither pattern of temporal vocalizations appears significantly more complex than the other.

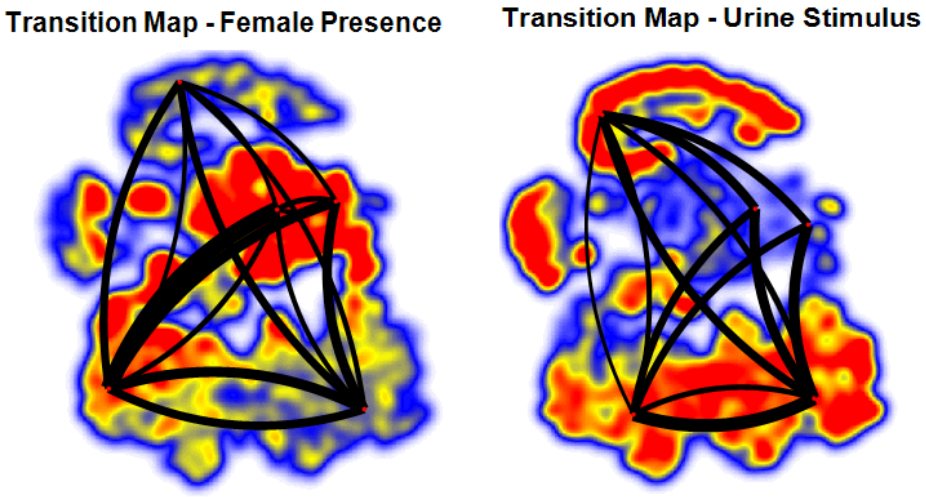


***Figure 10**. One-step transition rates ($\tau = 1$) plotted on the coarse-grained, manually labeled, behavioral maps. Red points represent the maximum of the local PDF, and black lines represent the transition probabilities between behavioral regions. Line thicknesses are proportional to the corresponding transition values T ($\tau = 1$), and right-handed curvature implies the direction of transmission. Similar to Figure 9, the composition of transitions differs between contexts, however the complexity of transitions does not.*

Utilizing multi-scale comparisons with the fine-grained (Figure 9) and coarse-grained (Figure 10) transition maps, we see that in both maps, although the complexity of transitions is not different, the composition of transitions is. These transitions mostly appear in regions of high probability density on the context-specific behavioral maps. In other words, the composition of transitions differs between contexts because of the respective differences in probability densities from the vocalization maps. This makes sense as we would expect vocalizations to transition more often into regions of higher probabilities compared to lower probabilities. To confirm this, we observed the structure of vocal transitions in both contexts without the inherent influence of the differing densities from the context-specific behavioral maps (see Methods). Essentially, the changes in overall densities were removed; then, the vocal transitions for tau=1 were calculated. This allowed us to investigate whether differences in vocal transitions were a direct result of the differing densities, or something beyond that. Thus, in

each context, transition matrices were decomposed to their eigenvalues and eigenvectors, and flux matrices were created by subtracting the first eigenvectors from the transition matrices (for details, see Methods). Then, the number of non-zero elements ('M') in each flux matrix was obtained, and each element of the flux matrices was compared to the value of $\frac{0.05}{M}$. Elements less than this value would indicate that those transitions were influenced beyond just the differing probability densities. Below, Figure 11 shows the flux matrices for both the urine and female contexts. All elements in both matrices were greater than $\frac{0.05}{M}$, indicating that no transitions between contexts differed beyond the influence of different densities. All the changes in vocal transitions between contexts can be explained by the changes in probability densities from the behavioral maps. Thus, what other papers stated as a change in syntactic complexity between contexts was really just a change in the vocalization probability density map. We can conclude then that male mice do not necessarily emit more complex syntax in one context versus the other, but simply utilize different types of vocalizations.

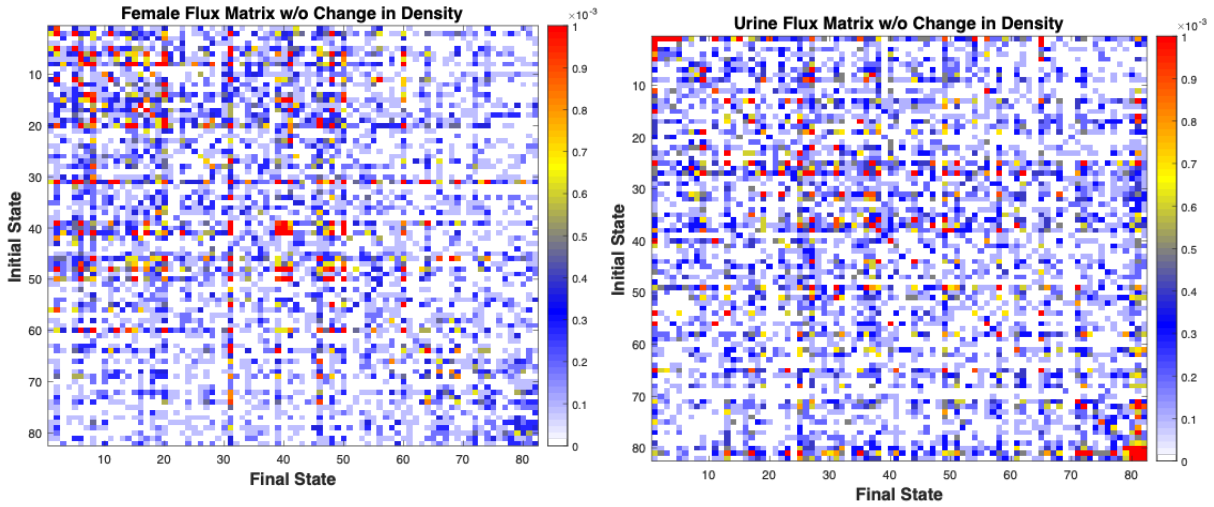


Figure 11. *Flux matrices for each context, created by subtracting the first eigenvector of the transition matrices (Figure 8). Flux matrices show transitions of vocalizations between watershed regions after removal of the probability densities. The structure does not differ between contexts, indicating that differences between transition matrices are a result of differences in behavioral densities.*

To quantify temporal complexity through long-term dynamics, we decomposed transition matrices into their eigenvalues and eigenvectors, and looked at the leading eigenvalue spectra (|λ|) of transition matrices from tau=1 to tau=100 (see Methods). Eigenvalues describe the predictability of a transition matrix, so for each tau, transition matrices were created in both contexts, and leading eigenvalues were calculated for each individual's vocal repertoire. The individuals' eigenvalues were then averaged, and shown in Figure 12 below, for both contexts, $|\lambda_\mu(\tau)|$ was plotted as a function of τ for μ= 2-6 (solid color lines). Predictions from the Markov Model were also plotted (dashed lines). It's important to note that Markov eigenvalues should decay to zero, but do not due to finite sampling effects. Eigenvalues of higher magnitude and slower decay over tau represent more complex temporal

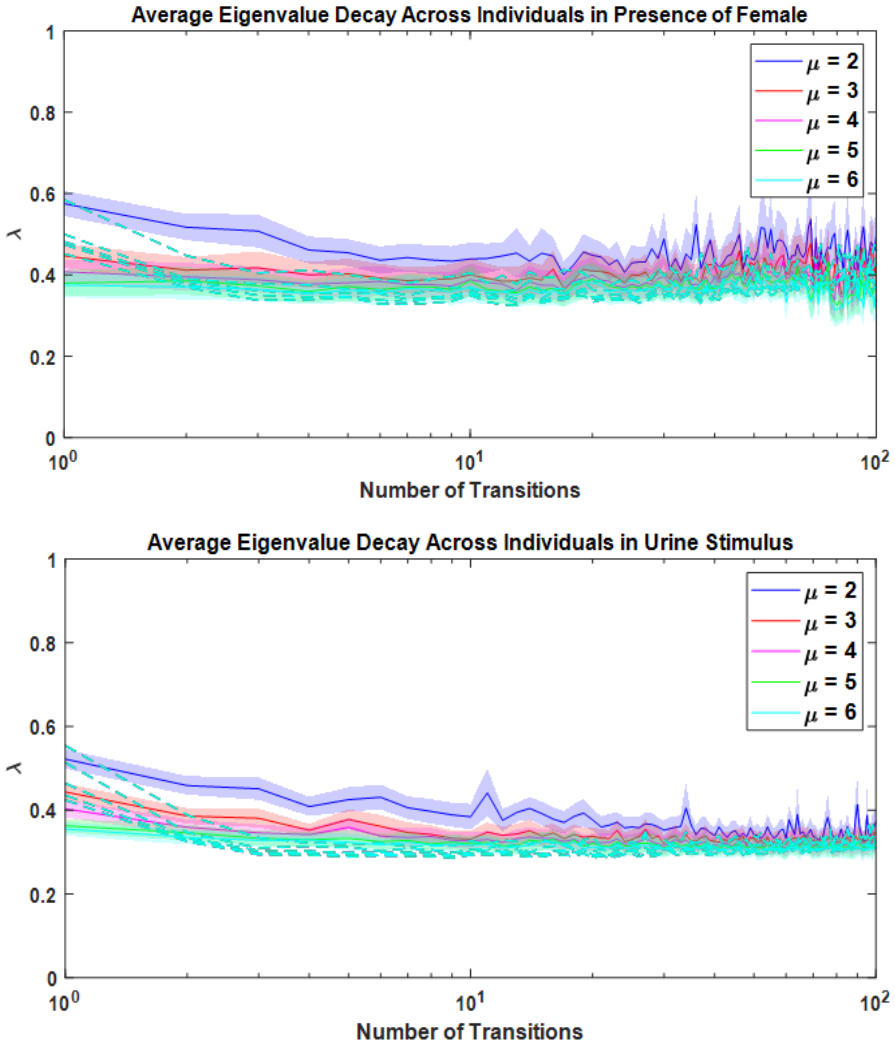


***Figure 12** (**to the left and below**). Absolute value of the leading eigenvalues of the transition matrices T (τ) as a function of τ for both social contexts (female presence above, urine stimulus below). Curves represent the average over all mice, and thicknesses represent the SEM. Dashed lines are the predictions for the Markov Model T_M (τ). There does not appear to be any significant differences in eigenvalue magnitude nor in eigenvalue decay between contexts.*

dynamics. In both contexts, the eigenvalue spectra of the transitions from the data has greater magnitude and slower rate of decay than that of the Markov Model. This indicates that the temporal dynamics from the vocal repertoire in both contexts are more complex than the Markov model. However, when comparing the data between contexts, the eigenvalue magnitudes appear similar for all μ in both contexts, as well as the rates of eigenvalue decay. This points to the notion of equal complexity between contexts through long-term communication.

In order to compare the eigenvalue spectra between the two contexts in more depth, the 2nd and 3rd leading eigenvalues (μ=2,3) of the transition matrices for both contexts were isolated and plotted, shown in Figure 13 below. The left figure shows the trajectory of just the 2nd leading average eigenvalue (μ=2) and the right shows that of just the 3rd leading average eigenvalue (μ=3). Red lines indicate vocalizations with the mice in the presence of a female, and blue lines indicate vocalizations with the mice exposed to female urine odor. Both figures show similar magnitudes of eigenvalues and similar rates of decay as a function of tau. Thus, utilizing insights from both Figures 12 and 13, we can conclude that temporal complexity in each context, through long-term communication, is similar.

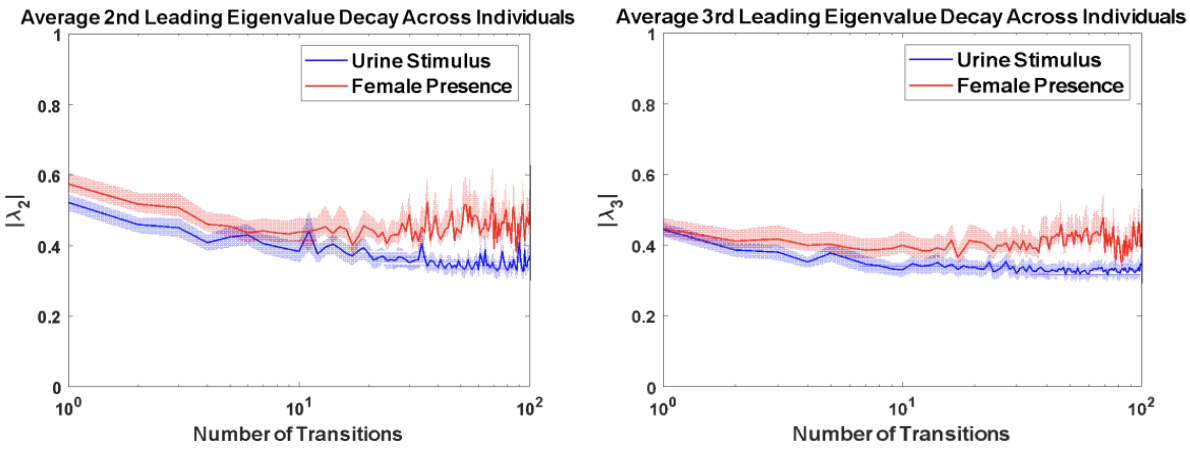


Figure 13. *Average leading eigenvalues (μ=2,3) of the transition matrices as a function of τ, separated by context. Left figure shows decay of the 2nd leading eigenvalue (μ=2), while the right figure shows the decay of the 3rd leading eigenvalue (μ=3). Red line indicates the presence of a female, and blue line indicates urine odor stimulus. Based on magnitude and decay, there are no significant differences in complexity between contexts.*

Overall, in this section, vocal complexity was analyzed through entropy and temporal dynamics, utilizing coarse- and fine- grained separations. We were able to conclude that the complexity of male communication did not differ significantly whether in the presence of a female and the urine odor context. First, we analyzed the entropy of the vocal repertoire, across the entire data as well as through each individual mouse, finding no significant differences between the contexts. Then we analyzed the temporal features of the mice's acoustic behavior across multiple scales. We calculated short-term transition matrices between contexts and compared their complexities. Additionally, we measured the eigenvalue spectra for long-term transitions between contexts, comparing the magnitudes and rates of decay. Through all analyses, no significant differences in complexity were found between contexts. Thus, we conclude that male mice do not communicate any more complex information or complex syntax in one social contexts versus another, but rather they just communicate different types of vocalizations.

Discussion:

In this study, we used a novel computational method to analyze the vocalizations of male mice in the presence of a female and with a female urine odor stimulus. We first created a fine-grained vocalization map of the vocal repertoire (Figure 2). Then, we manually labeled regions of the map according to previously identified syllable types to build a coarse-grained behavioral map (Figure 3). We then analyzed significant differences in the composition of vocalizations between contexts (Figure 4a-b), and also compared the distinguishability of the vocalizations between contexts and between individuals, finding more distinguishability between contexts (Figure 5). We compared the complexities of vocal repertoires between contexts utilizing newly developed definitions for repertoire and temporal complexity (Figure 6-13). Our observations suggested that across multiple scales, although the types of vocalizations change between different social contexts, the complexity of vocalizations does not. We believe that previous insights about vocal complexity were simply a result of the choices made in

defining the vocal behavior, rather than being an intrinsic social feature of mouse acoustic communication.

Comparisons with other studies:

In the first portion of our study, we produced a fine and coarse-grained behavioral map of the vocal repertoire. This method was similar to what was done in Berman et al. (2014), where they worked with mapping stereotyped behaviors of freely moving fruit flies (Berman et al., 2014). Using our behavioral maps, we first compared the composition of vocalizations between the two contexts, finding significant differences in the use of syllable types. Step vocalizations and frequency modulated upsweeps had higher usage in the urine context, while chevron, frequency modulated downsweeps, and constant frequencies were more common in the female presence context. This context-dependent change in the vocal structure aligns similarly with other studies. For example, in Chabout et al. (2015), male mice produced a greater proportion of syllable types with multiple pitch jumps ("m") in the urine condition compared to the female presence condition. Hanson and Hurley (2012) also found that syllables with at least one segment with at least one harmonic showed the greatest change in male vocal repertoire with respect to mounting behaviors, significantly increasing 10 seconds prior to mounting. And recently, Sangiamo et al. (2020) quantified the dependence of different vocal expressions on certain social contexts, finding that male mice vocalizations transmit information about their social actions (Chabout et al., 2015; Hanson and Hurley, 2012; Sangiamo et al., 2020).

We also analyzed the distinguishability of vocalizations between contexts and between individuals. Evidence using the Jensen-Shannon Divergence showed that vocalizations between contexts were more distinguishable than between individuals. This type of direct comparison had not been done before. Hanson and Hurley (2012) found significant differences in syllable types between different individual

males as well as different social contexts and social behaviors, but did not compare the magnitude of differences between individuals versus between contexts. Although many other studies have alluded to the idea that vocal repertoire differs between contexts, not many also offer evidence of significant differences between individuals.

Finally, our study found that the complexity of the vocal repertoire did not differ between contexts; thus, suggesting that complexity may not be an intrinsic feature of mouse acoustic communication. This result differs with the result of Chabout et al. (2015), in which they concluded that males emitted more complex syllables and sequences in response to female urine compared to female presence. We believe that their method for analyzing complexity from manual definitions of spectral features may not have been a fundamental observation about mice vocal behavior. In this study, all syllable types without pitch jumps were labeled simple “s”, while syllables containing a single downward pitch jump were denoted “d” and a single upward pitch jump were denoted “u”. Syllables with multiple pitch jumps were labeled “m”, and these were considered the complex vocalizations. Using these definitions, they made the claim that more complex vocalizations were emitted in the female urine context due to the higher proportion of “m” syllables, as well as the higher probability of transitioning to the “m” syllables. These results differed from our study, in which we developed a computational method for better understanding complexity that depended on few assumptions, and without having to make prior definitions of syllable types. Through coarse and fine-grained scales, complexity was analyzed through entropy calculations and short-term and long-term temporal dynamics. Our analyses align more with the results of Perrodin et al. (2020), where they found that both randomization and temporal reversal of male syllable sequences had no effect on female approach behavior in comparison to original songs. They claimed that females listen for and perceive only the presence/absence of specific syllable types, independent of

the syllable order. This then agrees with our conclusion that male social communication does not depend on the temporal dynamics of the repertoire, but rather just the composition of syllables types.

Why might the composition of vocalizations differ between two contexts, but not their complexities?

The two social contexts we compared in this study differed in a variety of ways. In the presence of a female there were many other cues: tactile, visual, and auditory as well as additional olfactory cues. These differences in cues may have led to different states of arousal, hence producing different types of vocalizations. In addition, in the female case, the possibility exists that some of the vocalizations may have come from the female. However, we believe that there are few, if any, female vocalizations in this data set for two reasons: 1) previous experimenters have suggested that when a mouse is added to a cage, as in these recordings, that mouse is less likely to vocalize (Chabout et al., 2012) and 2) we observed very few overlapping vocalizations (1 possible overlap out of 10933 vocalizations).

On the other hand, although the composition of vocalizations varied due to different states of arousal from the social contexts, this difference had no effect on vocal complexity. That's because there may not be any feature of differing complexity in mouse communication. For example, in 2005, Holy and Guo (2005) first found that mice USVs contained distinct syllable types with repeated temporal sequencing, similar to the songs of songbirds. This correlation of mice vocalizations and bird songs alludes to the idea that mice acoustic features resemble simple calls without any complicated linguistic structure. Our studies agree with this claim, in which we don't believe that complexity is even a feature of social communication. Thus, in any social context, we wouldn't expect to see any differences in the complexity of vocal communication.

How does complexity allow us to better understand vocal behavior and acoustic communication?

Ever since the initial claim that mouse USVs have characteristics of birdsong-like syntax, many studies thereafter have analyzed mouse vocalizations. Although it is known what types of social contexts produce USVs, little is still known about what features of these vocalizations convey important social information. As a result, different studies have attempted to investigate possible features such as duration, syllable type, and temporal structure. Here, we developed a computational method to understand complexity as a possible acoustic feature related to social context. Through this, we were able to gain better insight about important components of social communication and the mechanisms behind it.

The reason why understanding complexity is quite important is because in some species, complexity significantly enhances the ability for survival and reproduction. For example, with male peacocks, complex structures in the tails and feathers are crucial for successful courtship behaviors (Burgess, 2001). Feathers that display patterns of higher complexity will result in greater attraction of females and greater likeliness of reproduction. Similarly, human communication is quite complex, and this gives humans a survival advantage to think, adapt, and innovate. Thus, in the case of mice, vocal complexity is important to study as it may have an evolutionary advantage resulting in a higher chance of survival and reproduction.

Advantages and Disadvantages to Our Methods:

One major advantage of our method is the ability to capture variability using multiple scales. By analyzing the vocal repertoire in both fine and coarse-grained scales, we are able to take into account any differences based on the choices we make. In other words, if previous studies analyze complexity using only one set of manual definitions of vocal behavior, then utilizing only fine-grained separations of our behavioral map provides no progressive insights. Thus, incorporating both fine- and coarse-grained

separations of behavior is advantageous for providing multi-dimensional analyses not dependent on a single definition of behavior. Furthermore, our method of analysis is advantageous because it is map-based, with few assumptions, and with the use of watershed analysis and manual labeling, large numbers of vocalizations can be analyzed. On the other hand, one disadvantage of our current method is that it emphasizes vocalizations that appear frequently and are similar to each other. Rare vocalizations, or vocalizations that are extremely variable, will be distributed across the map, and won't contribute significantly to any one region. In addition, the other disadvantage is that the number of watershed regions that are found depends on the level of smoothing.

Future steps and questions:

It may be useful to utilize our newly developed computational methods to investigate any changes in vocal complexity in relation to different social behaviors. Sangiamo et al. (2020) tracked different patterns of vocalizations in mice while they performed specific social behaviors, such as dominance over other mice or avoidance of social interactions. Similarly, analyzing the complexity of vocalizations in mice while they performed these different complex social behaviors may be insightful. Additionally, our computational methods can be applied to understanding vocal complexity across species. For example, prairie voles are monogamous while mice are not, therefore major differences in innate social behavior could correlate with differences in vocal complexity. Furthermore, not only does our method have to be analyzed for acoustic communication, it can also be used with understanding motion, such as with Drosophila. By capturing variability and using multiple scales, insights about the complexity of physical behaviors could also be made.

Methods:

All animal experiments were conducted by Kelly Seagraves, Ph.D. and Roian Egnor, Ph.D. from the Janelia Research Campus.

Subjects:

Twelve male SWR/J mice (age range: 14-45 weeks) were maintained on a reversed dark-light cycle (12h on: 12h off), with ad lib access to food and water. Males were isolate-housed for at least 7 days prior to experiments. Eight SWR/J females (age range: 20-30 weeks) were used for urine stimuli and for male-female interaction stimuli.

Urine collection:

Urine for eliciting vocalizations was collected by placing individual female mice in a sterile empty cage until a sufficient quantity was produced. Urine was stored at -15 ± 1°C until a few minutes before use.

Determining estrus state:

Male bedding was added to the home cages of female stimulus mice to induce estrus cycling (Dalal, 2001; Whitten, 1956). The estrus state of all regularly cycling females was checked daily between 0800-1200, using vaginal cytology. Vaginal cell samples were collected via lavage (Mclean, 2012; Caligioni, 2009), stained with Wright Stain (Thermo Fisher Scientific Inc.) on a glass microscope slide, and viewed at 10x magnification on a light microscope. Proestrus was characterized as a mixture of basal and squamous cells, with few to no neutrophils, and estrus was characterized as a large proportion of squamous cells, with few to no basal cells or neutrophils. Metestrus and Diestrus were combined together due to few observations of Diestrus and were characterized as a mixture of basal cells, squamous cells, and neutrophils.

Recording vocalizations:

Vocalizations were recorded using custom-written MATLAB software on two ultrasonic microphones (CM16/CMPA40-5V; Avisoft Bioacoustics). Signals were amplified by 20 dB (40dB adjustable preamplifier; Avisoft Bioacoustics), lowpass filtered at 200 kHz (Krohn-Hite Model 3384 Four Channel Filter; Krohn-Hite Corp.) and digitized at 450450 Hz (NI PXIe-6356 DAQ device; National Instruments Corp., Texas, USA).

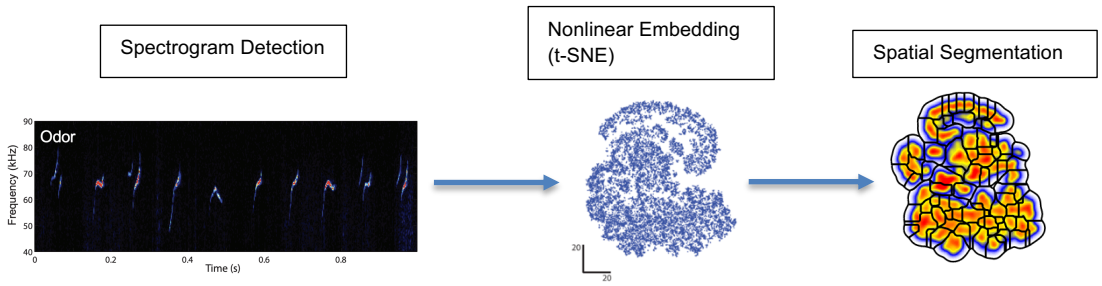
Automatic detection of vocalizations (Figure 1):

Ultrasonic frequency contours, continuous in time and frequency, were automatically detected using Ax (custom-written MATLAB-based software, available for download at https://github.com/JaneliaSciComp/Ax). Acoustic segmentation was manually verified and corrected for. Signals exceeding the amplitude limit of the recording system were excluded from analysis (Tabler, 2017). To determine whether the female mouse produced many of the vocalizations in the female condition each file was manually inspected for overlapping vocalizations. No overlapping vocalizations were identified in the female condition, consistent with other laboratories (Hanson and Hurley, 2012).

Generation of Fine-Grained Behavioral Map (Figure 2):

In order to visualize the vocal repertoire, methods similar to Tabler (2017) were applied. First, frequency contours were mean frequency subtracted in order to eliminate effects of individuality. This did not affect individuality of the frequency contour shape. Then all pairs of frequency contours were compared using dynamic time warping (Sakoe and Chiba, 1978) to create an all-to-all distance matrix ((21032 x 21032)/2 comparisons). Dynamic time warping is a method that adjusts sequences over time so that each vocal repertoire can be measured across similar time frames. All the acoustic structures from the frequency contours were fully captured by the high dimensional distance matrix. The data in the high dimensional space was then projected down into two dimensions using t-Distributed Stochastic

Neighbor embedding (t-SNE) (van der Maaten and hinton, 2008) (transition entropy = 5, relative convergence of the cost function to .0001). T-SNE is a nonlinear, dimensionality reduction method that aims to preserve local structure within a data set. Points that are nearby in a higher-dimensional representation remain nearby in the low-dimensional space, at the expense of large length scale distortions. This technique differs from other non-linear embedding methods such as a PCA, multidimensional scaling, and Isomap (Tenenbaum et al., 2000), which preserve global structures but distort local structures. Since t-SNE preserves these local neighbor relationships of frequency contours, regions of the behavioral maps can be interpreted as rough categories of syllable types based on similarities of frequency contour structure (Tabler, 2017). Thus, from our data of 21,000+ mouse calls from the two social contexts, t-SNE was utilized to visualize the vocalizations on a behavioral map, where vocalizations with similar frequency contours occupied local regions in the map. The figure below displays a general framework for generating the behavioral map. From the spectrograms containing the frequency contours, t-SNE was applied, producing a two-dimensional behavioral space. Lastly, a watershed segmentation was applied to a Gaussian-smoothed density, isolating individual regions of high probability density (Berman et al., 2014).



Manual labeling of syllables and Generation of Coarse-Grained Behavioral Map (Figure 3):

After watershed-based segmentation was applied over the behavioral map, regions of local maximum probability density were segregated. The vocalizations in each watershed region resembled frequency contours of similar structure embedded locally from the t-SNE. From each region, we extracted the

frequency contours and manually identified syllables that were defined relatively consistently across laboratories and strains. These were the FM upsweep (FMU), FM downsweep (FMD), constant frequency (CF), single step (ST), and chevron (CH) vocalizations (Holy, 2005; Panksepp, 2007; Mahrt, 2013; Scattoni, 2008; Hanson and Hurley, 2012; Grimsley, 2011). We used the following definitions: FMU—bandwidth greater than 6 kHz and upward frequency modulation; FMD—bandwidth greater than 6 kHz and downward frequency modulation; CF—bandwidth less than 6 kHz; CH—upward and then downward FM with at least 6 kHz between peak frequency and start and stop frequencies, ST—only one abrupt frequency discontinuity (greater than 2 kHz in less than 5 ms). Each watershed region was manually assigned a syllable type to generate the coarsely separated behavioral map.

Analysis of Vocal Composition Between Contexts (Figure 4a-b):

Behavioral maps were next separately generated by isolating vocalizations from the female presence context and urine-odor stimulus. Regions of significant differences between the two maps were determined using bootstrap analysis, done similarly in Tabler (2017). This was achieved through separately resampling the 2-D embeddings of the vocalizations for each context with replacement 10,000 times and convolving each of these resampled data sets with a Gaussian of width 4 to create distributions, $q_{fem}(\rho|x,y)$ and $q_{urine}(\rho|x,y)$ for each of the PDFs at every point in space $q_{fem}(\rho|x,y) = Prob\ (\rho_{fem}(x,y) = \rho)$ and $q_{urine}(\rho|x,y)$ = Prob ($\rho_{urine}(x,y) = \rho$). These spatially varying PDFs were obtained by fitting a Gaussian mixture model to the sampled PDFs (up to three peaks, chosen at each point by maximizing the Akaike Information Criterion). As we assumed that the two populations were sampled independently, the probability that the probability distribution of the female context was greater than the urine-odor context was given by numerically integrating:

$$P_{fem}\,(x,y) = \int_0^{\infty} \int_0^{\rho\_fem} q_{\ urine}(\rho_{urine}|x,y) q_{\ fem}(\rho_{fem}|x,y) d\rho_{urine} d\rho_{fem}$$

Regions of significant difference were those in which the difference in probability from the integration was less than .05 or greater than 0.95.

Between and Within Context Findings (Figure 5):

Vocalizations between the female presence and urine-odor contexts were compared to vocalizations between individuals within their respective contexts using the Jensen-Shannon Divergence, a method for quantifying the distinguishability between two probability distributions. The Jensen-Shannon Divergence utilizes the equation below, based on the Kullback-Leibler divergence (Lin, 1991), to measure the mutual information between samples from two distributions and the identity of which distribution each sample was drawn from. In other words, it is interpreted to answer the following question: if a sample is randomly drawn from two different distributions with different probabilities, how much uncertainty about the origin of the sample from the two distributions is reduced? (DeDeo, 2013). The Jensen-Shannon Divergence has been previously utilized to measure overall differences in genomic protein-coding sequences between various phyla groups such as bacteria, fungi, insects, vertebrae, and plants (Itzkovitz, 2010).

In the case of the male mice vocal repertoire, we used the Jensen-Shannon Divergence to compare the vocalizations within the same individuals and between the two contexts, and vocalizations within the same context between the twelve different individuals. In other words, in the first case, when looking at mouse 1, the distribution of its vocal repertoire within the female presence context was compared to the distribution within the urine odor context, and the distinguishability was calculated. This same procedure was done for mouse 2, 3, 4, etc. In the other case of the same context and between individual measurement, when looking at just the female presence context, the distribution of the vocal repertoire from mouse 1 and mouse 2 were compared, then mouse 1 and 3, mouse 1 and 4, continuing so on with

all combinations of individuals. This method quantified the distinguishability of communication, whether there were more differences in vocal features and structure between distributions of different contexts or distributions of different individuals.

$$D_{JS}(p||q) = \frac{1}{2}D_{KL}(p||\frac{p+q}{2}) + \frac{1}{2}D_{KL}(q||\frac{p+q}{2})$$ *Jensen-Shannon Divergence Equation*

Analysis of Vocal Complexity (Figure 6-13):

Repertoire Complexity (Figure 6-7a, b):

Vocal complexity through repertoire complexity was measured using the entropy value derived from information theory, which studies the quantification and communication of information and was originally proposed by Claude Shannon. Entropy, calculated from the equation, $H_i = \sum_j p_{j|i} \log p_{j|i}$, refers to the uncertainty of the data, thus directly quantifying its complexity. Entropy is utilized to determine the number of binary distinctions in a dataset, correlating to the total amount of information in a system (Berman et al., 2014). If the entropy value is equal to h, then there are 2^h number of distinctions in a data set (Cover and Thomas, 2006). When entropy has a value of zero, only one uniform outcome will occur, therefore the greater the entropy value, the more probabilistic the distribution and the more complex the data. In the case of the mouse vocal repertoire, entropy offers a more consistent measurement of complexity compared to studies which define complex vocalizations from their definitions of syllable types. By relying on quantitative information, this method overcomes the variability in defining what a complex vocalization is.

Three different analyses using entropy were conducted. First, the entire repertoire of male vocalizations between female presence and urine-odor contexts was bootstrapped, and entropy was calculated for each bootstrap sample. A PDF of the entropy between contexts was generated. Second, entropy was

calculated for each male individual's overall vocal repertoire between the two contexts. Each individual's repertoire for each context was bootstrapped, with the entropy value being calculated for each sample. The average entropy for each individual was subsequently plotted for each context, thus comparing differences of the individuals' complexity between social contexts. Third, individual entropy values were measured again, but keeping account of all watershed regions. For each individual's repertoire within each watershed region, their vocalizations were bootstrapped, and entropy values were calculated for each bootstrap sample. The average was then taken for each individual across all the watershed regions between contexts. A paired plot of entropy values was plotted again to measure entropy differences between context, accounting for each watershed region.

Temporal Complexity (Figure 8-13):

The complexity of mouse vocalizations was also defined in terms of temporal dynamics. First, behavioral transition matrices were calculated using the equation: $[\mathbf{T}(\tau)]_{i,j} \equiv p(S(n+\tau)=i|S(n)=j)$, which describes the probability that an animal's behavior will go from state j to i after tau transition states. The unit of tau is the number of calls, thus a tau value of 1 in the transition matrix represents one call. This method had been previously used to measure the transitions in behavioral states in Drosophila (Berman, 2016). Transition matrices of tau=1 were created for vocalizations within each context, for both the automatically generated, fine-grained behavioral map, as well as the manually labeled, coarse-grained behavioral map, then plotted onto their respective behavioral maps. This method ensured both coarse- and fine-grained scales were accounted for in this analysis of complexity.

Temporal complexity was also observed in terms of long-term dynamics. To do so, transition matrices were described from their eigendecompositions, shown in this equation here: $[\mathbf{T}(\tau)]_{i,j} = \sum_{\mu} \lambda_{\mu}(\tau) u_i^{\mu}(\tau) v_j^{\mu}(\tau)$ This equation was used to understand vocal complexity with large values of tau. $\lambda_{\mu}(\tau)$ represents the

eigenvalue with the μth largest modulus (Berman, 2016). In general, eigenvalues describe the predictability of a transition matrix, thus corresponding to its complexity. To measure long term temporal dynamics, transition matrices between each context were created from tau=1 to tau=100, and eigenvalues were calculated for each individual's vocal repertoire. For each tau, each individuals' eigenvalues were averaged, and plots of the decaying average eigenvalue spectra were created for each context. In order to understand the temporal complexities of the data, we first used the Markov Model as a reference. If the observed dynamics were purely Markovian, then transitions from one state to the next would not depend on the history of behavior, which, in terms of mouse vocal behavior means future vocalizations would depend only on the current vocalization, not on any past types of vocalizations. Thus, Markovian transition matrices were generated from tau=1 to tau=100, and the Markovian eigenvalues were extracted and plotted alongside the eigenvalue spectra from the data for each context. This compared the loss of predictability between the Markov Model and the data. To interpret the eigenvalue spectra, data with higher eigenvalue magnitudes and slower decaying rates over tau represent more complex temporal dynamics.

In order to directly compare the eigenvalues of the vocal repertoire between the two contexts, the averages of the second leading eigenvalues for both contexts were plotted together against the number of transitions. Additionally, the averages of the third leading eigenvalues between the contexts were plotted thereafter. Unlike past studies, this method of understanding temporal complexity offered another unbiased insight into understanding the complexity of the mice's acoustic communication. By creating transition matrices for small and large values of tau and calculating a spectrum of leading eigenvalues, we created a quantitative approach to study the complexity of both short-term and long-term behaviors, without the need to manually define the behaviors.

Measuring Changes in Transitions Upon Removal of Probability Densities (Figure 11):

The structure of vocal transitions was compared between contexts with the removal of overall probability density. This method determined whether compositional differences in vocal transitions of acoustic behaviors between contexts were a direct result of the context-specific differences in probability densities, or something beyond that. Thus, in each context, the first eigenvectors, which are proportional to the steady state probabilities, were subtracted from the transition matrices ('T') to create flux matrices ('S'). The transition matrix ('T') is defined as, $T = \lambda_1 \hat{v}_1 \hat{v}_1^T + \sum_{k=2}^{N} \lambda_k \hat{v}_k \hat{v}_k^T$, and the flux matrix ('S'), is defined as $S = \sum_{k=2}^{N} \lambda_k \hat{v}_k \hat{v}_k^T$, which is the second term in the transition matrix equation. Since the first eigenvalue (λ_1) is always equal to 1, the equation utilized to create the flux matrices was $S = T - \hat{v}_1 \hat{v}_1^T$. In each flux matrix, the number of non-zero elements ('M') was obtained. Each element of the flux matrices was then compared to the value of $\frac{0.05}{M}$. Elements greater than this value indicate that differences in those transitions were indeed explained by the compositional differences in probability densities from the behavioral maps. Elements less than this value indicate that those transitions changed beyond just the differing probability densities.

References:

Behr, O., & von Helversen, O. (2004). Bat serenades—complex courtship songs of the sac-winged bat (Saccopteryx bilineata). Behavioral Ecology and Sociobiology, 56(2), 106-115. doi:10.1007/s00265-004-0768-7

Berman, G. J., Bialek, W., & Shaevitz, J. W. (2016). Predictability and hierarchy in Drosophila behavior. Proceedings of the National Academy of Sciences, 113(42), 11943-11948. doi:10.1073/pnas.1607601113

Berman, G. J., Choi, D. M., Bialek, W., & Shaevitz, J. W. (2014). Mapping the stereotyped behaviour of freely moving fruit flies. J R Soc Interface, 11(99). doi:10.1098/rsif.2014.0672

Berwick, R. C., Okanoya, K., Beckers, G. J. L., & Bolhuis, J. J. (2011). Songs to syntax: the linguistics of birdsong. Trends in Cognitive Sciences, 15(3), 113-121. doi:https://doi.org/10.1016/j.tics.2011.01.002

Blank, D. A., Ruckstuhl, K., & Yang, W. (2014). Roaring function in male goitered gazelles. Behavioural Processes, 106, 152-159. doi:https://doi.org/10.1016/j.beproc.2014.05.009

Bradbury, J. W., & Vehrencamp, S. L. (1998). Principles of Animal Communication (2nd ed.): Sinauer Associates.

Burgess, S. (2001). The Beauty of the Peacock Tail and the Problems with the Theory of Sexual Selection. Journal of Creation, 15(2), 94-102.

Caligioni, C. S. (2009). Assessing reproductive status/stages in mice. Current protocols in neuroscience, Appendix 4, Appendix-4I. doi:10.1002/0471142301.nsa04is48

Chabout, J., Sarkar, A., Dunson, D. B., & Jarvis, E. D. (2015). Male mice song syntax depends on social contexts and influences female preferences. Front Behav Neurosci, 9, 76. doi:10.3389/fnbeh.2015.00076

Chabout, J., Serreau, P., Ey, E., Bellier, L., Aubin, T., Bourgeron, T., & Granon, S. (2012). Adult Male Mice Emit Context-Specific Ultrasonic Vocalizations That Are Modulated by Prior Isolation or Group Rearing Environment. PLOS ONE, 7(1), e29401. doi:10.1371/journal.pone.0029401

Cover, T. M., & Thomas, J. A. (2006). Elements of Information Theory. Hoboken, New Jersey: John Wiley & Sons, Inc.

D'Amato, F. R., Scalera, E., Sarli, C., & Moles, A. (2005). Pups call, mothers rush: does maternal responsiveness affect the amount of ultrasonic vocalizations in mouse pups? Behavior Genetics, 35(1). doi:10.1007/s10519-004-0860-9

Dalal, S., Estep, J., Valentin-Bon, I., & Jerse, A. (2001). Standardization of the Whitten Effect to Induce Susceptibility to Neisseria gonorrhoeae in Female Mice. Contemporary topics in laboratory animal science / American Association for Laboratory Animal Science, 40, 13-17.

DeDeo, S., Hawkins, R., Klingenstein, S., & Hitchcock, T. (2013). Bootstrap Methods for the Empirical Study of Decision-Making and Information Flows in Social Systems. Entropy, 15(12), 2246-2276. doi:10.3390/e15062246

Doherty, J., & Hoy, R. (1985). The Auditory Behavior of Crickets: Some Views of Genetic Coupling, Song Recognition, and Predator Detection. The Quarterly Review of Biology, 60.

Goffinet, J., Mooney, R., & Pearson, J. (2019). Inferring low-dimensional latent descriptions of animal vocalizations. *bioRxiv*. doi:https://doi.org/10.1101/811661

Grimsley, J. M. S., Monaghan, J. J. M., & Wenstrup, J. J. (2011). Development of Social Vocalizations in Mice. PLOS ONE, 6(3), e17460. doi:10.1371/journal.pone.0017460

Hammerschmidt, K., Radyushkin, K., Ehrenreich, H., & Fischer, J. (2012). The Structure and Usage of Female and Male Mouse Ultrasonic Vocalizations Reveal only Minor Differences. PLOS ONE, 7(7), e41133. doi:10.1371/journal.pone.0041133

Hanson, J. L., & Hurley, L. M. (2012). Female Presence and Estrous State Influence Mouse Ultrasonic Courtship Vocalizations. PLOS ONE, 7(7), e40782. doi:10.1371/journal.pone.0040782

Holy, T. E., & Guo, Z. (2005). Ultrasonic Songs of Male Mice. PLOS Biology, 3(12), e386. doi:10.1371/journal.pbio.0030386

Itzkovitz, S., Hodis, E., & Segal, E. (2010). Overlapping codes within protein-coding sequences. Genome Res, 20(11), 1582-1589. doi:10.1101/gr.105072.110

Lin, J. (1991). Divergence measures based on the Shannon entropy. IEEE Transactions on Information Theory, 37(1), 145-151. doi:10.1109/18.61115

Liu, R. C., Miller, K. D., Merzenich, M. M., & Schreiner, C. E. (2004). Acoustic variability and distinguishability among mouse ultrasound vocalizations. The Journal of the Acoustical Society of America, 114, 3412-3422. doi:10.1121/1.1623787

Mahrt, E. J., Perkel, D. J., Tong, L., Rubel, E. W., & Portfors, C. V. (2013). Engineered deafness reveals that mouse courtship vocalizations do not require auditory experience. The Journal of neuroscience : the official journal of the Society for Neuroscience, 33(13), 5573-5583. doi:10.1523/JNEUROSCI.5054-12.2013

Marler, P. R., Krebs, J. R., & Gabriel, H. (1990). Song learning: the interface between behaviour and neuroethology. The Royal Society, 329(1253). doi:https://doi.org/10.1098/rstb.1990.0155

McLean, A., Valenzuela, N., Fai, S., & Bennett, S. (2012). Performing Vaginal Lavage, Crystal Violet Staining, and Vaginal Cytological Evaluation for Mouse Estrous Cycle Staging Identification. Journal of visualized experiments : JoVE, 67. doi:10.3791/4389

Neunuebel, J. P., Taylor, A. L., Arthur, B. J., & Egnor, S. E. (2015). Female mice ultrasonically interact with males during courtship displays. Elife, 4. doi:10.7554/eLife.06203

Panksepp, J. B., Jochman, K. A., Kim, J. U., Koy, J. J., Wilson, E. D., Chen, Q., . . . Lahvis, G. P. (2007). Affiliative Behavior, Ultrasonic Communication and Social Reward Are Influenced by Genetic Variation in Adolescent Mice. PLOS ONE, 2(4), e351. doi:10.1371/journal.pone.0000351

Perrodin, C., Verzat, C., & Bendor, D. (2020). Courtship behaviour reveals temporal regularity is a critical social cue in mouse communication. bioRxiv. doi:https://doi.org/10.1101/2020.01.28.922773

Petric, R., & Kalcounis-Rueppell, M. C. (2013). Female and male adult brush mice (Peromyscus boylii) use ultrasonic vocalizations in the wild. Behaviour, 150(14), 1747-1766. doi:10.1163/1568539X-00003118

Portfors, C. V. (2007). Types and functions of ultrasonic vocalizations in laboratory rats and mice. American Association for Laboratory Animal Science, 46(1).

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing, 26(1), 43-49. doi:10.1109/TASSP.1978.1163055

Sangiamo, D. T., Warren, M. R., & Neunuebel, J. P. (2020). Ultrasonic signals associated with different types of social behavior of mice. Nature Neuroscience. doi:https://doi.org/10.1038/s41593-020-0584-z

Scattoni, M. L., Gandhy, S. U., Ricceri, L., & Crawley, J. N. (2008). Unusual Repertoire of Vocalizations in the BTBR T+tf/J Mouse Model of Autism. PLOS ONE, 3(8), e3067. doi:10.1371/journal.pone.0003067

Tabler, J. M., Rigney, M. M., Berman, G. J., Gopalakrishnan, S., Heude, E., Al-Lami, H. A., . . . Wallingford, J. B. (2017). Cilia-mediated Hedgehog signaling controls form and function in the mammalian larynx. Elife, 6. doi:10.7554/eLife.19153

Tenenbaum, J. B., de Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. Science 290(5500). doi:doi:10.1126/science.290.5500.2319

Van Der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. Journal of Machine Learning Research, 9.

von Merten, S., Hoier, S., Pfeifle, C., & Tautz, D. (2014). A role for ultrasonic vocalisation in social communication and divergence of natural populations of the house mouse (Mus musculus domesticus). PLoS One, 9(5), e97244. doi:10.1371/journal.pone.0097244

White, N. R., Prasad, M., Barfield, R. J., & Nyby, J. G. (1998). 40- and 70-kHz Vocalizations of Mice (Mus musculus) during Copulation. Physiology & Behavior, 63(4), 467-473. doi:https://doi.org/10.1016/S0031-9384(97)00484-8

Whitten, W. K. (1956). Modification of the Oestrous Cycle of the Mouse by External Stimuli Associated with the Male. Journal of Endocrinology, 13(4). doi:https://doi.org/10.1677/joe.0.0130399

Yang, M., Loureiro, D., Kalikhman, D., & Crawley, J. N. (2013). Male mice emit distinct ultrasonic vocalizations when the female leaves the social interaction arena. Front Behav Neurosci, 7, 159. doi:10.3389/fnbeh.2013.00159