Report

Current Biology

Mouse spontaneous behavior reflects individual variation rather than estrous state

Highlights

- Spontaneous behavior of female mice is only negligibly affected by estrous state
- Females and males exhibit strongly individualized patterns of exploration
- Female spontaneous behavior is less variable than male behavior

Authors

Dana Rubi Levy, Nigel Hunter, Sherry Lin, ..., Rockwell Anyoha, Rebecca M. Shansky, Sandeep Robert Datta

Correspondence

r.shansky@northeastern.edu (R.M.S.), srdatta@hms.harvard.edu (S.R.D.)

In brief

Levy et al. track open-field behavior of female mice over weeks and find that behavior reflects individual identity far more than estrous state. Open-field exploration patterns in males are much more variable than in females, arguing for the inclusion of both sexes in studies of spontaneous behaviors.

Levy et al., 2023, Current Biology *33*, 1–7
April 10, 2023
https://doi.org/10.1016/j.cub.2023.02.035

CellPress

Please cite this article in press as: Levy et al., Mouse spontaneous behavior reflects individual variation rather than estrous state, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.02.035

Current Biology

CellPress
OPEN ACCESS

Report

Mouse spontaneous behavior reflects individual variation rather than estrous state

Dana Rubi Levy,[1] Nigel Hunter,[1] Sherry Lin,[1] Emma Marie Robinson,[1] Winthrop Gillis,[1] Eli Benjamin Conlin,[1] Rockwell Anyoha,[1] Rebecca M. Shansky,[2,*] and Sandeep Robert Datta[1,3,*]
[1]Department of Neurobiology, Harvard Medical School, Boston, MA, USA
[2]Department of Psychology, Northeastern University, Boston, MA, USA
[3]Lead contact
*Correspondence: r.shansky@northeastern.edu (R.M.S.), srdatta@hms.harvard.edu (S.R.D.)
https://doi.org/10.1016/j.cub.2023.02.035

SUMMARY

Behavior is shaped by both the internal state of an animal and its individual behavioral biases. Rhythmic variation in gonadal hormones during the estrous cycle is a defining feature of the female internal state, one that regulates many aspects of sociosexual behavior. However, it remains unclear whether estrous state influences spontaneous behavior and, if so, how these effects might relate to individual behavioral variation. Here, we address this question by longitudinally characterizing the open-field behavior of female mice across different phases of the estrous cycle, using unsupervised machine learning to decompose spontaneous behavior into its constituent elements.[1,2,3,4] We find that each female mouse exhibits a characteristic pattern of exploration that uniquely identifies it as an individual across many experimental sessions; by contrast, estrous state only negligibly impacts behavior, despite its known effects on neural circuits that regulate action selection and movement. Like female mice, male mice exhibit individual-specific patterns of behavior in the open field; however, the exploratory behavior of males is significantly more variable than that expressed by females both within and across individuals. These findings suggest underlying functional stability to the circuits that support exploration in female mice, reveal a surprising degree of specificity in individual behavior, and provide empirical support for the inclusion of both sexes in experiments querying spontaneous behaviors.

RESULTS

Mice exhibit a stereotyped pattern of behavior when exploring an open field.[1,5] This pattern is altered when the shape of the arena is changed, when sensory cues are introduced, when mice are hungry or in pain, or when brain dynamics are reconfigured through optogenetic perturbations or treatment with neuro- or psychoactive drugs.[1,2,3,6–8] Importantly, each of these different experimental manipulations evokes a characteristic change in behavior that is similar across mice, such that, e.g., the identity and dose of an administered drug can be accurately predicted from behavior in the open field alone.[3] These observations support the broad proposal that the state of a mouse at any given moment determines how it explores the world.[9]

However, mouse behavior is also individualized. For example, different (and yet genetically identical and similarly housed) mice explore the same arena by expressing distinct patterns of behavior that are characteristic of each mouse[3,10,11]; these inter-mouse differences likely reflect stochastic aspects of gene expression and development as well as differences in experience, which collectively impinge upon neural circuit structure and function.[12] Thus, exploratory behavior—and likely other types of naturalistic, spontaneous behavior—reflects a balance between the tendency of mice to generate state-dependent behaviors appropriate for a given context and their tendency to behave as unique individuals.

The day-to-day internal state of female rodents rhythmically varies due to the estrous cycle, in which levels of circulating gonadal hormones systematically rise and fall.[13–15] These hormones alter gene expression, connectivity, and synaptic function across the brain, serving as powerful neuromodulators that can ultimately affect behavior.[16–19] Although estrous state clearly influences sex-specific behaviors, it remains uncertain whether female gonadal hormones more pervasively influence other forms of behavior expressed by both sexes, like exploration.[16,17,20–26] On the one hand, several studies have reported significant estrous-related changes in rodent open-field behavior (e.g., time in center,[27] center entries,[28] and distance traveled[21]); furthermore, the striatal and dopaminergic circuits that control the structure and dynamics of spontaneous exploration have been shown to be sensitive to gonadal hormones.[5,29,30] On the other hand, researchers have also reported the absence of cycle-related differences in exploratory behaviors, or behavioral changes that are linked to interactions between the estrous phase and strain, age, or context.[22,23,25,31,32]

Inconsistencies across these studies may derive, at least in part, from the widespread use of low-dimensional summary statistics (e.g., average velocity) to measure complex and dynamic patterns of exploratory behavior. In addition, studies querying the behavioral effects of the estrous cycle often include only a small number of behavioral sessions, grouping the data by estrous phase and averaging across individual mice. Because

Please cite this article in press as: Levy et al., Mouse spontaneous behavior reflects individual variation rather than estrous state, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.02.035

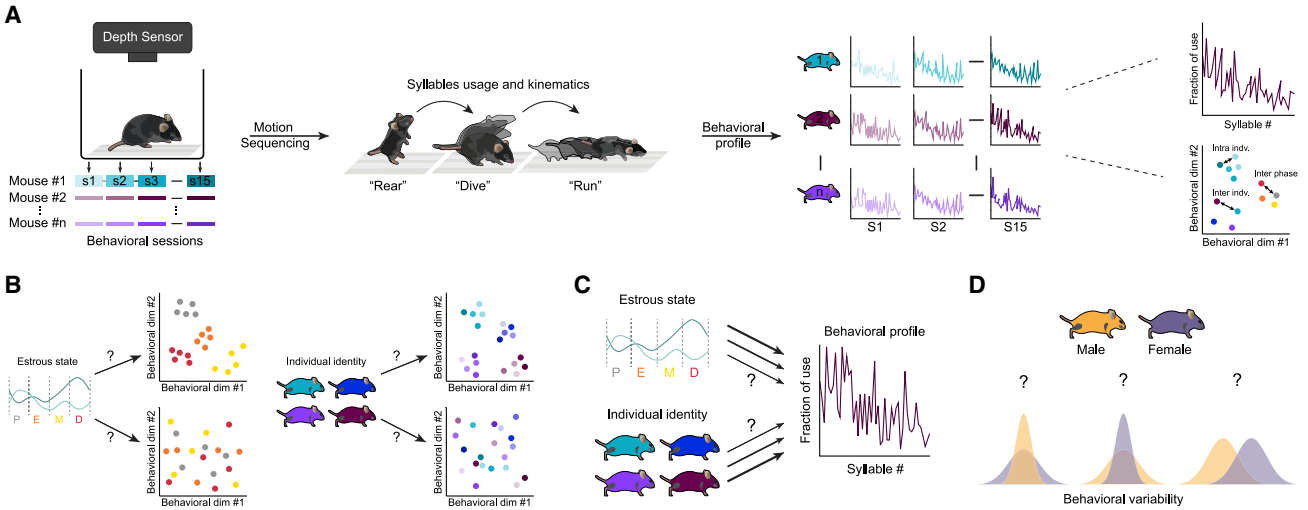

Figure 1. Weighing the relative influence of estrous state and individual variation on spontaneous behavior

(A) Individual mice (n = 16 females, n = 16 males) were recorded in an open-field arena using a depth camera for 15 consecutive sessions, and the estrous phase of female mice was examined each day. Depth video recordings were analyzed using the motion sequencing algorithm (see method details), which segments behavior into sub-second behavioral modules (termed "syllables"; see examples). The fraction of use of each syllable is then calculated for each session, and differences in behavior assessed between individuals (inter-indv.), between sessions for each individual (intra-indv.), and between estrous phases within each individual (inter-phase). Here and throughout, D = diestrus, E = estrus, M = metestrus, and P = proestrus.

(B and C) The effect of the estrous phase (adapted from McLean et al.[14]) and individual mouse identity on female mouse behavior (as captured by how frequently each behavioral syllable was used) was explored (B), as was the relative contribution of each to the use of behavioral syllable and the variability in their use (C).

(D) Differences in the magnitude of intra-individual and inter-individual behavioral variability between males and females were also assessed.

See also Figure S1.

this approach focuses on the estrous cycle as the only source of behavioral variability, it by design neglects other sources of inter-mouse behavioral variability—like individual behavioral biases—that are only measurable with repeated sampling of behavior over time.

To directly quantify the relative contribution of estrous state and individual identity to behavior, here we perform a longitudinal assessment of mouse spontaneous behavior using motion sequencing (MoSeq). MoSeq is a well-validated unsupervised machine learning algorithm that decomposes behavior into sub-second component modules (e.g., rears, runs, and grooms), which are referred to as "syllables."[1,4] This approach yields a high-dimensional description of behavior that captures the use of behavioral modules and sequences while allowing simultaneous measurement of conventional behavioral parameters such as velocity and position.[1,2,3]

We first studied the naturalistic behavior of adult C57BL/6J female mice exploring an open-field arena over multiple consecutive days while tracking their estrous cycle (Figures 1 and S1). Commonly used summary statistics failed to identify a significant effect of estrous phase on open-field exploration (Figures S2A–S2C). In addition, overall syllable distribution was not different between estrous phases (Figures 2A, S2A, S2D, and S2F), and phase-dependent patterns of behavior were not clustered in a low-dimensional embedding of syllable space (Figure 2B). Decoding analysis revealed that MoSeq-identified behavioral syllables and sequences could not predict the specific phase of the estrous cycle associated with any given behavioral session (Figures 2C, 2G, and S2H). We noted, however, that averaging all instances of a particular estrous phase expressed by an individual mouse (rather than considering each session separately) enabled decoders to predict estrous phase based upon syllable use at a level modestly above chance (Figures 2C, right panel, and 2G). The handful of syllables that conveyed the most information about estrous phase were associated with active exploration (e.g., running and rearing, Table S1; Figure S2G). Together, these findings suggest that subtle changes in exploratory behavior are associated with each estrous phase, but that these changes are masked by substantial session-to-session variability.

By contrast, decoders accurately predicted individual mouse identity more than 85% of the time across all behavioral sessions (Figures 2F and 2G); these data demonstrate that each of the 16 tested female mice exhibits a characteristic pattern of behavior that is sufficiently stable over time to enable individual identification based upon behavior expressed within a single experimental session alone. Consistent with this observation, individual patterns of behavior were well clustered in a low-dimensional embedding of syllable space (Figures 2D, 2E, S2A, and S2E). Importantly, most behavioral syllables identified by MoSeq conveyed at least some information about individual identity, suggesting that inter-individual differences in behavior reflect broad changes in syllable use, rather than the modulation of a small set of individualized syllables (Table S1; Figure S2G). Taken together, these data reveal that female mice express an individual-specific pattern of behavior when exploring the open field and that this individualized behavioral profile is relatively invariant across multiple experimental sessions encompassing different estrous phases.

To further explore the relative effect of estrous state and individual identity on behavior, we quantified the degree to which

Please cite this article in press as: Levy et al., Mouse spontaneous behavior reflects individual variation rather than estrous state, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.02.035

CellPress
OPEN ACCESS

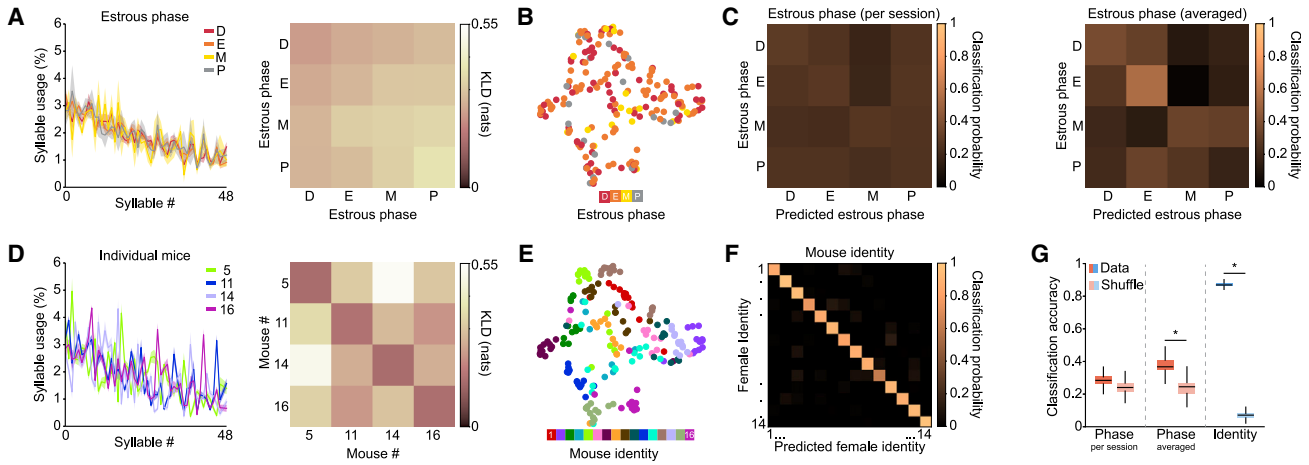

Figure 2. Female spontaneous behavior reflects individual variation rather than the estrous phase

(A) Left: syllable usage distribution for each of the estrous phases averaged across four representative mice (presented also in D). Mean ± SEM (shaded area) is presented. MoSeq assigns each syllable a unique identifier (Syllable #) based upon how often that syllable was used across all the data subject to modeling (in this case, across all female mice and sessions in our dataset), such that Syllable "0" is the most often used and subsequent syllables are less used, and we maintain this syllable ordering across all syllable usage plots herein but vary how the data are aggregated (for example, by estrous phase or individual identity). See Figure S2D for a similar analysis across all mice. Right: Kullback-Leibler divergence (KLD) values of pairwise comparisons between phase-related data presented in the left panel. Here, we compute a phase-dependent KLD by pooling all the sessions corresponding to a particular phase across mice, calculating the pairwise KLD between all sessions corresponding to the same or a different phase (as labeled) and then plotting the average value of those pairwise comparisons in each cell. These values quantify how similar syllable usage distributions are across the different estrous phases, with lower values indicating greater similarity. It is important to note that because phase-based comparisons incorporate data from different animals, if different individual mice exhibit distinct patterns of behavior, those differences will be incorporated into this metric.

(B) Uniform manifold approximation and projection (UMAP) plot depicting syllable usage in females for each session color coded by estrous phase. To assess cluster quality, K-means clustering analysis was performed on high-dimensional data, and clustering quality was quantified using the adjusted rand index (ARI), where higher values indicate a greater match between clustering and data labels; for phase (number of clusters = 4) ARI = 0.03.

(C) Confusion matrix for classification accuracy of a decoder trained to predict estrous phase based on syllable usages. Decoder was trained on data from individual sessions (left) or averaged data per phase per mouse (right) (see method details).

(D) Left: same data as in (A), but syllable use is now averaged across all sessions corresponding to the four representative mice. See Figure S2E for a similar analysis across all mice. Right: same as the right panel in (A), but KLD values are now computed among the four representative mice depicted in the left panel. Note that values on the diagonal indicate intra-individual variability.

(E) UMAP plot depicting syllable usage in females for each session color coded by mouse identity. Clustering analysis was performed as described in (B). For individuals (number of clusters = 16) ARI = 0.43.

(F) Same as (C) but for the prediction of individual mouse identity.

(G) Quantification of overall decoder performance (n = 1,000 restarts for data and shuffled data). Asterisk (*) denotes statistical significance, here indicating that the mean decoding distribution exceeds the 95th percentile shuffle threshold. For all relevant panels, box plots depict median, interquartile range, and upper/lower adjacent values (black lines).

See also Figure S2 and Table S1.

each contributed to the overall variability observed in our data. This analysis demonstrates that inter-individual differences in syllable use are far greater than phase-related differences in behavior expressed by each mouse (Figure 3A). Furthermore, modeling reveals that estrous state contributes little to the overall behavioral variation observed across the dataset, whereas the identity of each mouse accounted for almost all the explained variability (Figure 3B).

Given these findings, we wondered whether the amount of variation observed in female open-field behavior (either within individual mice or between mice) was greater or less than that observed in male behavior. To address this question, we longitudinally characterized open-field behavior in age-matched male mice and similarly found that male behavior largely reflected mouse identity (Figures S3A and S3B; Table S2). However, both intra- and inter-individual variability in male behavior was substantially greater than that observed in females (Figures 4A, 4B, S3, and S4); furthermore, variability in syllable use significantly increased over time in male but not female mice (Figures 4A and 4B, right panels). These findings demonstrate that despite hormonal fluctuations associated with the estrous cycle, female open-field behavior—within individual mice, between different mice, and over time—is significantly more stable than that expressed by males.

DISCUSSION

The potential for estrous-driven variation to alter experimental outcomes has been a primary motivation for the widespread exclusion of female rodents in behavioral neuroscience research.[33–35] This concern has been previously considered in several meta-analyses, which have found that the overall distribution of behavioral variation (assessed by aggregating hundreds of behavioral metrics) is no greater in female rodents than in male rodents.[36–39] Similarly, one recent study found that bulk movement in the home cage is less variable across

Please cite this article in press as: Levy et al., Mouse spontaneous behavior reflects individual variation rather than estrous state, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.02.035

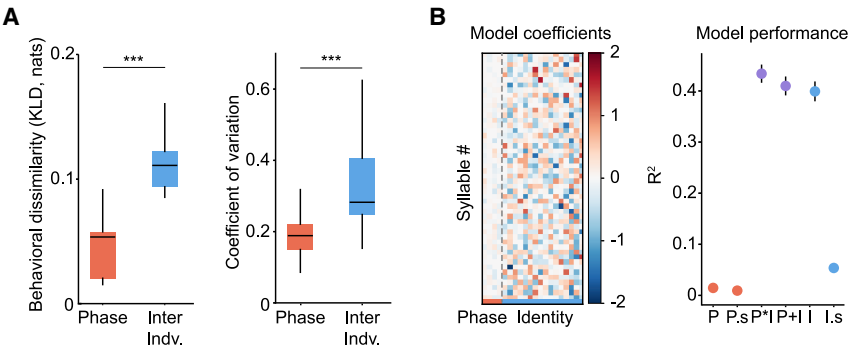

Figure 3. Individual identity, not the estrous phase, accounts for variability in female behavior

(A) Behavioral dissimilarity among phases (within each mouse) and individuals (inter indv) as measured by KLD of syllable usage distributions (left) and by the CV of the usage of each syllable (right). Wilcoxon signed-rank test was performed. For KLD: n = 16 mice, T = 1.0, p = $6.10*10^{-5}$. For CV: n = 49 syllables, T = 32, p = $7.72*10^{-9}$.

(B) Left: weights from linear models fit to model syllable usage as a function of estrous phase and mouse identity. Right: model performance as quantified by R^2 (mean ± SEM) for models trained to predict syllable usage as a function of phase (P); phase with phase labels shuffled within a mouse (P.s, n = 100); phase, identity, and their interaction (P*I); phase and identity without interaction (P+I, weights are shown in the left panel); mouse identity (I); mouse identity, with shuffled identity labels (I.s, n = 100). No significant differences in model performance were identified between P*I, P+I, and I models (assessed via ANOVA). For all relevant panels, box plots depict median, interquartile range, and upper/lower adjacent values (black lines). ***p < 0.001.

females than males.[40] However, these papers do not make clear whether the observed behavioral variation in females is primarily the consequence of estrous or individual variation.

Our granular, high-dimensional, and longitudinal analysis reveals that the behavioral effects of estrous state are modest (at best) compared with that of individual variation during spontaneous open-field behavior. For example, stable individual differences in behavior are apparent in single experimental sessions, whereas stereotyped estrous phase-dependent behavioral differences are not. To give an intuitive sense of the scale of the effects of the estrous cycle on exploratory behavior, in previous experiments we have demonstrated that mice treated with even small doses of neuro- or psychoactive drugs exhibit far greater changes in behavior than those observed when we measure inter-individual variability.[3] Thus, from a practical perspective, the impact of estrous state on exploratory behavior is likely to be negligible.

Given the well-documented synaptic, cellular, and network-level effects of gonadal hormones on neural circuits implicated in exploration, the fact that estrous minimally influences spontaneous behavior suggests the existence of neural mechanisms that stabilize individual behavior in female mice across different estrous phases.[29,30] Furthermore, our data clearly demonstrate that female exploratory behavior is less variable (regardless of estrous state) than that of males. Taken together, our findings argue for the inclusion of both sexes in experiments querying behavior and support the perspective that females—rather than males—should be the default sex used in studies of exploratory behavior in circumstances in which both sexes cannot be tested.

Estrous variation has been proposed to manifest in a species-, strain-, age-, and context-dependent manner.[22,23,31] Thus, a main caveat of our work is that we only explore female and male mouse behavior in a single strain and experimental setting, one that does not impose an explicit task or goal. We have also not made manipulations to explore the causal mechanisms that underlie the striking degree of inter-individual variation we observe. Future work will be required to assess how general our findings are, to identify features of development or experience that are particularly important to defining individualized patterns of open-field behavior, and to explain why behavioral variability is greater in males than females. We speculate that the use of dense, unsupervised behavioral characterization methods, as we have done here, will be useful for unveiling specific relationships between internal states and individual behavioral variation across a variety of contexts.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and Code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animals
- METHOD DETAILS
  - Behavioral procedure
  - Behavioral modeling
  - Cytological evaluation of mouse estrous cycle
  - Phase and identity classifications
  - Identification of putative mouse size-related syllables
  - Assessing the influence of size
  - UMAP visualization
  - Clustering analysis
  - General linear model
  - Kullback-Leibler Divergence (KLD) analysis
  - Coefficient of Variation (CV) analysis
  - Mutual information (MI) analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2023.02.035.

ACKNOWLEDGMENTS

S.R.D. is supported by NIH grants U19NS113201, RF1AG073625, and R01NS114020; the Brain Research Foundation; the Simons Collaboration on

Please cite this article in press as: Levy et al., Mouse spontaneous behavior reflects individual variation rather than estrous state, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.02.035

CellPress
OPEN ACCESS

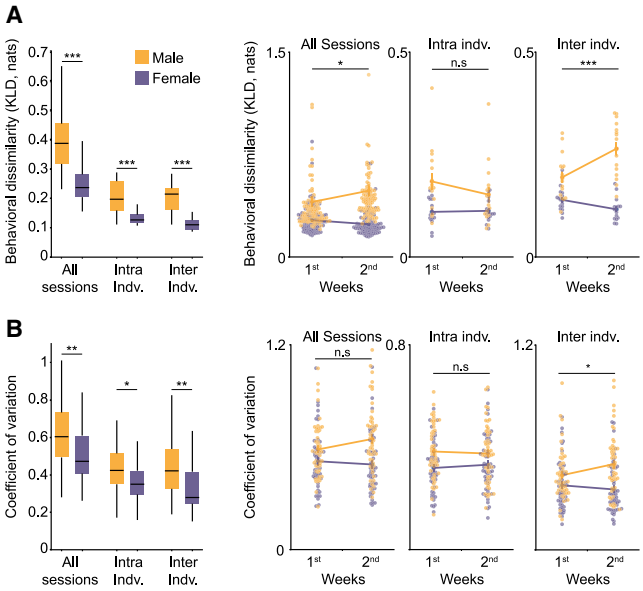

Figure 4. Female behavioral variability is lower than that of males during spontaneous open-field behavior

(A) Left: comparison of male and female behavioral variability, where variability is quantified across all behavioral sessions, across sessions within each individual (intra indv. = intra-individual variability), and across individuals (inter indv. = inter-individual variability) by measuring the KLD of syllable usage distributions. Two-way ANOVA for sex and experimental conditions (exp) as main factors was performed: $Fsex_{(1,438)} = 55.46$, $p = 5.10*10^{-13}$; $Fexp_{(2,438)} = 16.12$, $p = 1.74*10^{-7}$, $Fsex*exp_{(2,438)} = 1.40$, $p = 0.246$. Individual contrasts were performed using Student's t test with Bonferroni correction: $p_{all\ sessions} = 6.42*10^{-11}$, $n_{female} = 188$ sessions, $n_{male} = 192$ sessions; $p_{intra\text{-}indv} = 3.3*10^{-4}$; $p_{inter\text{-}indv} = 1.28*10^{-5}$, n = 16 mice for male and females. Right: variability across the first and second weeks of behavioral assessment (1^{st}: days 1–7, $n_{female} = 93$ sessions, $n_{male} = 94$ sessions; 2^{nd}: days 8–14 $n_{female} = 95$ sessions, $n_{male} = 98$ sessions, excluding the first experimental day in both sexes). Each point represents a single session (for "all sessions" analysis) or a single mouse (for "intra indv." and "inter indv." analysis), with mean ± SEM presented as lines. Two-way ANOVA for sex and time as main factors was performed for each condition, with individual contrasts using Student's t test and Bonferroni correction. Significant interaction effects are marked in the figure. For all sessions: $Fsex_{(1,376)} = 46.2613$, $p = 4.08*10^{-11}$; $Ftime_{(1,376)} = 0.90$, $p = 0.34$; $Fsex*time_{(1,376)} = 4.01$, $p = 0.045$; For intra-individual: $Fsex_{(1,60)} = 14.07$, $p = 3.9*10^{-4}$; $Ftime_{(1,60)} = 0.96$, $p = 0.32$; $Fsex*time_{(1,60)} = 1.33$, $p = 0.25$; For inter-individual: $Fsex_{(1,60)} = 73.03$, $p = 5.77*10^{-12}$; $Ftime_{(1,60)} = 3.8$, $p = 0.054$; $Fsex*time_{(1,60)} = 15.23$, $p = 2.4*10^{-4}$. Male significantly increased their inter-individual variation between the first and second week of behavioral recordings with $p = 0.006$, all other comparisons were not significant when corrected for multiple comparisons. For visualization purposes (but not analytical purposes) 3 of the 784 data points were excluded because they lie outside the range of the graph.

(B) Same as in (A) but depicted is the coefficient of variation of the usage of each syllable. Left panel: two-way ANOVA: $Fsex_{(1,288)} = 27.97$, $p = 2.43^{-7}$; $Fexp_{(2,288)} = 42.56$, $p = 6.38*10^{-17}$, $Fsex*exp_{(2,288)} = 0.52$, $p = 0.59$. Individual contrasts were performed using Student's t test with Bonferroni correction: $p_{all\ sessions} = 0.0063$; $p_{intra\text{-}indv} = 0.015$; $p_{inter\text{-}indv} = 0.0051$. Right panel: two-way ANOVA: for all sessions: $Fsex_{(1,192)} = 17.80$, $p = 3.8*10^{-4}$; $Ftime_{(1,192)} = 0.7$, $p = 0.39$; $Fsex*time_{(1,192)} = 2.52$, $p = 0.11$; For intra-individual: $Fsex_{(1,192)} = 11.47$, $p = 8.5*10^{-4}$; $Ftime_{(1,192)} = 0.038$, $p = 0.85$; $Fsex*time_{(1,\ 192)} = 0.4$, $p = 0.52$; For inter-individual: $Fsex_{(1,192)} = 21.63$, $p = 6*10^{-6}$; $Ftime_{(1,192)} = 0.86$, $p = 0.35$; $Fsex*time_{(1,192)} = 4.04$, $p = 0.045$. $n_{mice} = 16$ and $n_{syllables} = 49$ for both males and females. All individual contrasts were not significant when corrected for multiple comparisons. For all relevant panels, box plots depict median, interquartile range, and upper/lower adjacent values (black lines). *p < 0.05, **p < 0.01, ***p < 0.001; n.s., not significant.
See also Figures S3 and S4 and Table S2.

the Global Brain; and the Simons Collaboration for Plasticity in the Aging Brain. D.R.L. is supported by the Human Frontier Science Program fellowship LT000838/2020 and the Zuckerman STEM Leadership Program and is an awardee of the Women's Postdoctoral Career Development Award. Portions of this research were conducted on the O2 High Performance Compute Cluster at Harvard Medical School. Slide imaging was performed in the HMS/BCH Center for Neuroscience Research #NS072030. We thank members of the Datta lab as well as Dr. Meital Oren-Suissa, Dr. Susana Q. Lima, Dr. Annegret Falkner, and Dr. Vanessa Ruta for useful comments on the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization and methodology, S.R.D. and D.R.L.; supervision and funding acquisition, S.R.D.; investigation, D.R.L., N.H., E.B.C., and E.M.R.; software and formal analysis, D.R.L., S.L., W.G., and R.A.; visualization, D.R.L.; writing, S.R.D., D.R.L., and R.S.

DECLARATION OF INTERESTS

S.R.D. sits on the scientific advisory boards of Neumora, Inc. and Gilgamesh Therapeutics.

INCLUSION AND DIVERSITY

We worked to ensure sex balance in the selection of non-human subjects. One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location. One or more of the authors of this paper self-identifies as a gender minority in their field of research. One or more of the authors of this paper self-identifies as a member of the LGBTQIA+ community. One or more of the authors of this paper received support from a program designed to increase minority representation in their field of research. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list.

Received: October 2, 2022
Revised: December 12, 2022
Accepted: February 10, 2023
Published: March 7, 2023

REFERENCES

1. Wiltschko, A.B., Johnson, M.J., Iurilli, G., Peterson, R.E., Katon, J.M., Pashkovski, S.L., Abraira, V.E., Adams, R.P., and Datta, S.R. (2015). Mapping sub-second structure in mouse behavior. Neuron *88*, 1121–1135.
2. Markowitz, J.E., Gillis, W.F., Beron, C.C., Neufeld, S.Q., Robertson, K., Bhagat, N.D., Peterson, R.E., Peterson, E., Hyun, M., Linderman, S.W., et al. (2018). The striatum organizes 3D behavior via moment-to-moment action selection. Cell *174*, 44–58.e17.
3. Wiltschko, A.B., Tsukahara, T., Zeine, A., Anyoha, R., Gillis, W.F., Markowitz, J.E., Peterson, R.E., Katon, J., Johnson, M.J., and Datta, S.R. (2020). Revealing the structure of pharmacobehavioral space through motion sequencing. Nat. Neurosci. *23*, 1433–1443. https://doi.org/10.1038/s41593-020-00706-3.
4. Lin, S., Gillis, W.F., Weinreb, C., Zeine, A., Jones, S.C., Robinson, E.M., Markowitz, J., and Datta, S.R. (2022). Characterizing the structure of mouse behavior using motion sequencing. Preprint at arXiv. https://doi.org/10.48550/arXiv.2211.08497.
5. Markowitz, J.E., Gillis, W.F., Jay, M., Wood, J., Harris, R.W., Cieszkowski, R., Scott, R., Brann, D., Koveal, D., Kula, T., et al. (2023). Spontaneous

behaviour is structured by reinforcement without explicit reward. Nature *614*, 108–117. https://doi.org/10.1038/s41586-022-05611-2.

6. Zhang, Z., Roberson, D.P., Kotoda, M., Boivin, B., Bohnslav, J.P., González-Cano, R., Yarmolinsky, D.A., Turnes, B.L., Wimalasena, N.K., Neufeld, S.Q., et al. (2022). Automated preclinical detection of mechanical pain hypersensitivity and analgesia. Pain *163*, 2326–2336. https://doi.org/10.1097/j.pain.0000000000002680.
7. Li, C., Hou, Y., Zhang, J., Sui, G., Du, X., Licinio, J., Wong, M.L., and Yang, Y. (2019). AGRP neurons modulate fasting-induced anxiolytic effects. Transl. Psychiatry *9*, 111. https://doi.org/10.1038/s41398-019-0438-1.
8. Burnett, C.J., Li, C., Webber, E., Tsaousidou, E., Xue, S.Y., Brüning, J.C., and Krashes, M.J. (2016). Hunger-driven motivational state competition. Neuron *92*, 187–201. https://doi.org/10.1016/j.neuron.2016.08.032.
9. Tinbergen, N. (1951). The Study of Instinct (Clarendon Press).
10. Freund, J., Brandmaier, A.M., Lewejohann, L., Kirste, I., Kritzler, M., Krüger, A., Sachser, N., Lindenberger, U., and Kempermann, G. (2013). Emergence of individuality in genetically identical mice. Science *340*, 756–759. https://doi.org/10.1126/science.1235294.
11. Forkosh, O., Karamihalev, S., Roeh, S., Alon, U., Anpilov, S., Touma, C., Nussbaumer, M., Flachskamm, C., Kaplick, P.M., Shemesh, Y., et al. (2019). Identity domains capture individual differences from across the behavioral repertoire. Nat. Neurosci. *22*, 2023–2028. https://doi.org/10.1038/s41593-019-0516-y.
12. Honegger, K., and de Bivort, B.d. (2018). Stochasticity, individuality and behavior. Curr. Biol. *28*, R8–R12. https://doi.org/10.1016/j.cub.2017.11.058.
13. Ajayi, A.F., and Akhigbe, R.E. (2020). Staging of the estrous cycle and induction of estrus in experimental rodents: an update. Fertil. Res. Pract. *6*, 5. https://doi.org/10.1186/s40738-020-00074-3.
14. McLean, A.C., Valenzuela, N., Fai, S., and Bennett, S.A.L. (2012). Performing vaginal lavage, crystal violet staining, and vaginal cytological evaluation for mouse estrous cycle staging identification. J. Vis. Exp. *67*, e4389.
15. Byers, S.L., Wiles, M.V., Dunn, S.L., and Taft, R.A. (2012). Mouse estrous cycle identification tool and images. PLoS One *7*, e35538. https://doi.org/10.1371/journal.pone.0035538.
16. Knoedler, J.R., Inoue, S., Bayless, D.W., Yang, T., Tantry, A., Davis, C.H., Leung, N.Y., Parthasarathy, S., Wang, G., Alvarado, M., et al. (2022). A functional cellular framework for sex and estrous cycle-dependent gene expression and behavior. Cell *185*, 654–671.e22. https://doi.org/10.1016/j.cell.2021.12.031.
17. Inoue, S. (2022). Neural basis for estrous cycle-dependent control of female behaviors. Neurosci. Res. *176*, 1–8. https://doi.org/10.1016/j.neures.2021.07.001.
18. Jennings, K.J., and de Lecea, L. (2020). Neural and hormonal control of sexual behavior. Endocrinology *161*, bqaa150. https://doi.org/10.1210/endocr/bqaa150.
19. Gutierrez-Castellanos, N., Husain, B.F.A., Dias, I.C., and Lima, S.Q. (2022). Neural and behavioral plasticity across the female reproductive cycle. Trends Endocrinol. Metab. *33*, 769–785.
20. Chari, T., Griswold, S., Andrews, N.A., and Fagiolini, M. (2020). The stage of the estrus cycle is critical for interpretation of female mouse social interaction behavior. Front. Behav. Neurosci. *14*, 113. https://doi.org/10.3389/fnbeh.2020.00113.
21. Burke, A.W., and Broadhurst, P.L. (1966). Behavioural correlates of the oestrous cycle in the rat. Nature *209*, 223–224. https://doi.org/10.1038/209223a0.
22. Lovick, T.A., and Zangrossi, H., Jr. (2021). Effect of estrous cycle on behavior of females in rodent tests of anxiety. Front. Psychiatry *12*, 711065. https://doi.org/10.3389/fpsyt.2021.711065.
23. Meziane, H., Ouagazzal, A.M., Aubert, L., Wietrzych, M., and Krezel, W. (2007). Estrous cycle effects on behavior of C57BL/6J and BALB/cByJ female mice: implications for phenotyping strategies. Genes Brain Behav. *6*, 192–200. https://doi.org/10.1111/j.1601-183X.2006.00249.x.
24. Finger, F.W. (1969). Estrus and general activity in the rat. J. Comp. Physiol. Psychol. *68*, 461–466. https://doi.org/10.1037/h0027490.
25. Ogawa, S., Chan, J., Gustafsson, J.A., Korach, K.S., and Pfaff, D.W. (2003). Estrogen increases locomotor activity in mice through estrogen receptor α: specificity for the type of activity. Endocrinology *144*, 230–239. https://doi.org/10.1210/en.2002-220519.
26. Mullenix, P. (1981). Structure analysis of spontaneous behavior during the estrous cycle of the rat. Physiol. Behav. *27*, 723–726. https://doi.org/10.1016/0031-9384(81)90246-8.
27. Rocks, D., Cham, H., and Kundakovic, M. (2022). Why the estrous cycle matters for neuroscience. Biol. Sex Differ. *13*, 62. https://doi.org/10.1186/s13293-022-00466-8.
28. Koonce, C.J., Walf, A.A., and Frye, C.A. (2012). Type 1 5α-reductase may be required for estrous cycle changes in affective behaviors of female mice. Behav. Brain Res. *226*, 376–380. https://doi.org/10.1016/j.bbr.2011.09.028.
29. Krentzel, A.A., and Meitzen, J. (2018). Biological sex, estradiol and striatal medium spiny neuron physiology: A mini-review. Front. Cell. Neurosci. *12*, 492. https://doi.org/10.3389/fncel.2018.00492.
30. Zachry, J.E., Nolan, S.O., Brady, L.J., Kelly, S.J., Siciliano, C.A., and Calipari, E.S. (2021). Sex differences in dopamine release regulation in the striatum. Neuropsychopharmacology *46*, 491–499. https://doi.org/10.1038/s41386-020-00915-1.
31. Miller, C.K., Halbing, A.A., Patisaul, H.B., and Meitzen, J. (2021). Interactions of the estrous cycle, novelty, and light on female and male rat open field locomotor and anxiety-related behaviors. Physiol. Behav. *228*, 113203. https://doi.org/10.1016/j.physbeh.2020.113203.
32. Scimonelli, T., Marucco, M., and Celis, M.E. (1999). Age-related changes in grooming behavior and motor activity in female rats. Physiol. Behav. *66*, 481–484. https://doi.org/10.1016/s0031-9384(98)00314-x.
33. Shansky, R.M., and Murphy, A.Z. (2021). Considering sex as a biological variable will require a global shift in science culture. Nat. Neurosci. *24*, 457–464. https://doi.org/10.1038/s41593-021-00806-8.
34. Beery, A.K., and Zucker, I. (2011). Sex bias in neuroscience and biomedical research. Neurosci. Biobehav. Rev. *35*, 565–572. https://doi.org/10.1016/j.neubiorev.2010.07.002.
35. McCarthy, M.M., Arnold, A.P., Ball, G.F., Blaustein, J.D., and De Vries, G.J. (2012). Sex differences in the brain: the not so inconvenient truth. J. Neurosci. *32*, 2241–2247.
36. Becker, J.B., Prendergast, B.J., and Liang, J.W. (2016). Female rats are not more variable than male rats: a meta-analysis of neuroscience studies. Biol. Sex Differ. *7*, 34. https://doi.org/10.1186/s13293-016-0087-5.
37. Prendergast, B.J., Onishi, K.G., and Zucker, I. (2014). Female mice liberated for inclusion in neuroscience and biomedical research. Neurosci. Biobehav. Rev. *40*, 1–5. https://doi.org/10.1016/j.neubiorev.2014.01.001.
38. Beery, A.K. (2018). Inclusion of females does not increase variability in rodent research studies. Curr. Opin. Behav. Sci. *23*, 143–149. https://doi.org/10.1016/j.cobeha.2018.06.016.
39. Kaluve, A.M., Le, J.T., and Graham, B.M. (2022). Female rodents are not more variable than male rodents: a meta-analysis of preclinical studies of fear and anxiety. Neurosci. Biobehav. Rev. *143*, 104962. https://doi.org/10.1016/j.neubiorev.2022.104962.
40. Smarr, B., and Kriegsfeld, L.J. (2022). Female mice exhibit less overall variance, with a higher proportion of structured variance, than males at multiple timescales of continuous body temperature and locomotive activity records. Biol. Sex Differ. *13*, 41. https://doi.org/10.1186/s13293-022-00451-1.
41. Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al. (2020). Array programming with NumPy. Nature *585*, 357–362.
42. Waskom, M. (2021). seaborn: statistical data visualization. J. Open Source Softw. *6*, 3021. https://doi.org/10.21105/joss.03021.

43. McKinney, W. (2010). Data structures for statistical computing in python. Proceedings of the Python in Science Conference, 56–61.

44. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). Scikit-learn: machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

45. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272.

46. Seabold, S., and Perktold, J. (2010). Statsmodels: econometric and statistical modeling with python. Proceedings of the Python in Science Conference *61*, 92–96.

47. Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. Comput. Sci. Eng. *9*, 90–95. https://doi.org/10.1109/MCSE.2007.55.

48. Cora, M.C., Kooistra, L., and Travlos, G. (2015). Vaginal cytology of the laboratory rat and mouse: review and criteria for the staging of the estrous cycle using stained vaginal smears. Toxicol. Pathol. *43*, 776–793. https://doi.org/10.1177/0192623315570339.

Please cite this article in press as: Levy et al., Mouse spontaneous behavior reflects individual variation rather than estrous state, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.02.035

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
JorVet Dip Quick Stain	Fischer Scientific	NC9581034
Black Polyethylene Tank (OFA)	US plastics	#14317
Deposited data		
Analyzed data and related scripts	This paper	https://github.com/dattalab/spontaneous-behavior-reflects-individuality-not-estrous
Post-MoSeq behavioral database	This paper	https://doi.org/10.5281/zenodo.7622958
Experimental models: Organisms/strains		
Mouse: C57BL/6J	The Jackson Laboratory	#000664
Software and algorithms		
Motion Sequencing (Moseq)	The Datta lab	http://www.moseq4all.org/
OLYMPUS OlyVIA 2.9	Olympus Software Imaging Solutions	https://www.olympus-sis.com/
Python version 3.6	Python Software Foundation	https://www.python.org
Numpy	Harris et al[41]	https://numpy.org/
Seaborn	Waskom et al[42]	https://seaborn.pydata.org/
Pandas	Mckinney et al[43]	https://pandas.pydata.org/
Scikit-learn	Pedregosa et al[44]	https://scikit-learn.org/stable/
Scipy	Virtanen et al[45]	https://scipy.org/
Statsmodels	Seabold et al[46]	https://www.statsmodels.org/stable/index.html
Matplotlib	Hunter et al[47]	https://matplotlib.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Sandeep Robert Datta (srdatta@hms.harvard.edu)

Materials availability

This study did not generate new unique reagents.

Data and Code availability

- The data that support the findings of the current study are available at Zenodo and are publicly available as of the date of publication. DOIs are listed in the key resources table.
- The behavioral analysis code and Moseq software used to model and analyze the data is freely available to all academic researchers online: http://www.moseq4all.org/.
- All original code has been deposited at Github and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

Animals used for this study were adult (3 months old) male and female C57BL/6J mice obtained from Jackson laboratories (stock number #000664, n=16 for each sex for all experiments). Mice were kept on a reverse 12-h light–dark cycle with food and water

Please cite this article in press as: Levy et al., Mouse spontaneous behavior reflects individual variation rather than estrous state, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.02.035

ad libitum and tested during the dark phase under dim red light. All mice were grouped-housed (four mice in a cage, randomly assigned). Mouse husbandry and experiments were performed following institutional and federal guidelines and were approved by Harvard Medical School's Institutional Animal Care and Use Committee.

METHOD DETAILS

Behavioral procedure

At the day of the testing, mice were brought to the behavioral room and allowed to habituate in their home cage for a minimum of 30 minutes under dim red light. Following habituation, each individual mouse was placed in the middle of a circular 17″ diameter open-field arena (OFA) enclosed with 14″-high opaque walls (US Plastics 14317), immediately after which video recording began. The arena walls were sanded to eliminate reflective artifacts in the depth video. Depth videos of mouse behavior were acquired at 30 Hz using a Kinect v2 camera and app for Windows (Microsoft). Mice were allowed to freely explore the enclosure for a 20-min period after which they were returned to their home cage. Arenas were thoroughly cleaned using 70% Ethanol solution and allowed to air dry between trials and at the end of each experimental day. All animals were tested for 15 consecutive days. Female and male experiments were performed on consecutive weeks. To focus our conclusions on exploration rather than novelty responses, we excluded the first encounter of each mouse with the behavioral apparatus from analysis. For variability analysis, number of female experimental days was matched to males.

To control for possible handling and habituation related effects when comparing male and female data variability, an additional experiment was performed in which a novel batch of aged-matched (3 months old) males and females (n=16 for either sex) were habituated to the behavioral apparatus for 10 consecutive days, and then recorded for additional 7 days. Here, females were not tested for estrous phase, and thus male and female mice were matched for handling; further, providing an extensive habituation minimized any novelty/stress related behavioral effects in these control cohorts. Female and male experiments were performed on the same day to control for batch effects. Open field arenas were thoroughly cleaned using 70% ethanol and left to vent for 30 minutes between sexes. Behavioral results were analyzed as for the main experiments and reported in Figure S3.

Behavioral modeling

Depth videos of all mice used in this study were preprocessed, extracted, modeled, and analyzed using the Motion Sequencing (MoSeq) algorithm as previously described,[1,4] using the pipeline available here: http://www.moseq4all.org/. In brief, MoSeq is an unsupervised, ethologically inspired machine learning algorithm which automatically segments animal spontaneous behavior into subsecond motifs (termed "syllables", e.g., rear, run, pause). Moseq labels each frame with a syllable label and identifies how often within a given session each syllable is used. Only those syllables used in more than one percent of frames across all sessions were used for analysis (n=49 for females and matched for the males and for the control experiment presented in Figure S3). Syllable number are sorted by fraction of use across all sessions, as seen in Figures 2A, 2D, S2D, and S2E. The behavior depicted by each syllable was later manually annotated and presented in Table S1 (for females) and Table S2 (for males). In addition, MoSeq identifies the transition statistics that governs how often every syllable transitions into every other syllable (referred to in Figure S2H as "bigrams"). The MoSeq pipeline also quantified more traditional scalar metrics used to describe behavior (e.g., velocity, size, location, distance to center). Total distance traveled during each session was computed by adding the between-frame Euclidean distance between the mouse centroid coordinates. Time in center was calculated as total time mouse center point was present within 11cm circular radius around the center of open field arena. Sessions which could not be extracted due to technical errors or insufficient recording quality (e.g., reflections on OFA walls, sessions where mouse size was at 2.5% top/bottom of mouse size area distribution) were excluded from analysis prior to modeling; overall 9/240 of female sessions and 32/224 sessions in male data were excluded. For the handling/stress control experiment, 18/112 of female sessions and 12/112 sessions in male data were excluded.

Cytological evaluation of mouse estrous cycle

At the end of each experimental day a vaginal swab was collected from each female mouse and used to determine its estrous state as previously described.[13–15] Briefly, swabs were gently collected form the vaginal opening using a saline-dipped cotton swab at a similar time each day, spread on a glass slide and left to dry in room temperature, after which they were immediately stained (JorVet Dip Quick Stain). Slides were then imaged at 10X magnification (Olympus VS120 Virtual Slide Microscope). Relative presence and proportion of leukocytes, cornified epithelial cells and nucleated epithelial cells were used to determine the estrous phase: Diestrus (D), Proestrus (P), Estrus (E) and Metestrus (M) (see Figure S1 for representative examples. Image contrast was adjusted for visualization purposes). Contaminated slides (i.e., significant urine residue or cellular debris) were removed from the analysis and corresponding behavioral sessions were excluded as well. Labeling was done by two independent observers and compared; behavioral trials corresponding to images in which labels were in disagreement were removed from analysis. Overall, 31/231 sessions were excluded. In accordance with the established distribution of phases across the cycle (in which P and M phases are brief and more difficult to detect[13,15,48]), the final number of included sessions per phase was D=62, E=99, M=20, P=19; the final number of included phases per mouse (after applying technical exclusion criteria described above for both behavioral recording and cytological assessment): 7/16 mice - four phases, 8/16 - three phases, one mouse was included only in the E and D phases and was therefore excluded from all decoder and associated analysis.

Please cite this article in press as: Levy et al., Mouse spontaneous behavior reflects individual variation rather than estrous state, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.02.035

Phase and identity classifications

To predict estrous phase (out of the four estrous phases) and individual mouse identity (out of all same-sex mice) a random forest classifier was trained on session-based behavioral syllable usages (the usage of each syllable was used as an input feature, n_{trees} = 250, class_weights=balanced, so that weights are inversely proportional to class frequencies (i.e phase/mouse identity) in the input data). To maintain a balanced training set, mice that had less than ten total behavioral sessions or fewer than three of the four estrous phases in the data were excluded from all classification analysis in the manuscript (2/16 of females and 2/16 males).

All classifications were performed on all behavioral sessions, other than the classifier presented in right panel of Figure 2C and corresponding middle panel of Figure 2G, in which syllable usages were averaged over repetitions of the same phase within a mouse (e.g., two estrus days of a single mouse were averaged to give the mean behavioral "estrus" signature in that mouse, and similarly for all other estrous phases in that mouse, for all mice). To maintain a balanced training set, training data was bootstrapped by subsampling 9 sessions per phase (in the case of the phase decoding) or 9 sessions per mouse (in the case of the individual decoders) for each decoder restart. In the case of both average and per-session phase decoders, tests were performed on a single held-out mouse; in the case of the individual identity decoders, tests were performed on a single held-out session. The distribution of classification accuracy across 1000 bootstrapped restarts is presented and compared to classification accuracy of shuffled data.

Decoders presented in Figure S2B were trained on scalar quantiles (0.2, 0.4, 0.6, 0.8) calculated in each sessions: 2D velocity, 3D velocity, and distance to center, as well as total distance traveled and total time in center (each scalar was considered separately in Figure S2B and all measurements were aggregated together as one input vector in Figure S2C). Decoder presented in Figure S2H used syllable bigrams (the probability of a two syllables sequence) observed above chance in behavioral data as input. Here, syllable sequence was extracted for each session, and bigram probabilities were defined as the probability of two syllables to appear one after the other in a specific order, for each pair of syllables in the dataset (so that number of possible bigrams = $49^2 - 49$, since self-transition were excluded). The probability for each bigram was calculated across all sessions. To identify bigrams that are observed above chance levels, syllable sequence was then permuted (n=2,000) and bigrams probabilities were calculated for each permutation resulting in a probability distribution for each bigram across the permutated data set. Decoder analysis included only bigrams which probability was greater than |±2SD| from the mean of the permutated bigram probability distribution (n=2185).

Identification of putative mouse size-related syllables

To identify candidate syllables whose usage might be correlated with mouse size, a linear model (ElasticNet with α=0.01) was fit to model mouse size-related features as a function of log-transformed, z-scored syllable usage data calculated per mouse. Size features included height, length, width and body area of the mouse as measured by the MoSeq algorithm from the recorded 3D videos. Each measure was calculated as the robust range per mouse. i.e. the differences between 95th and 5th percentile values of that measure per animal. Model weights were compared against shuffled data (n=1000 shuffled per measurement). To provide the most rigorous analysis, syllables with weights surpassing two-sided top 2.5 percentile of the shuffle weight distribution for any of the measurements were identified as correlating with mouse size. These syllables (n=8 for males, n=15 for females) were excluded from the decoder presented in Figures S4A and S4B, and from the variability analysis presented in Figures S4C and S4D.

Assessing the influence of size

To explore the relationship between mouse size and individual behavioral patterns, we compared the degree of confusion between pairs of mice (as assessed via a classifier) with their size differences. To perform this analysis, we designed "held-out" decoders that excluded all sessions from a single mouse from the training data, and then used those sessions as the test data (decoder input features were behavioral syllables, and the overall design was as described above). This approach identified the individual mouse or mice that the classifier most confused with the held-out query mouse (Figures S4E and S4F). We also measured the size differences between individual mice, using the z-scored height, length, width and body area of the mouse as measured by the MoSeq algorithm from the recorded 3D videos. Each measure was calculated as the robust range per mouse. i.e. the differences between 95th and 5th percentile values of that measure per animal, and the overall size difference was determined by the Euclidian distance between measurement vectors for each pair of mice. In order to facilitate direct comparisons between classifier confusions and size, both datasets were normalized to a 0-to-1 scale. To allow for positive relationships, size distances were converted to size similarity such that similarity = 1-distance. Classification confusion and size similarity were then directly compared and fitted using a linear model as described in figure legend. Self-distances (both in size analysis and in the classification confusion analysis) were excluded from the final correlation analysis. As described above, and to allow evaluation of these results versus all other decoder results described in the manuscript, mice that had less than ten total behavioral sessions or fewer than three of the four estrous phases in the data were excluded from classification analysis and from all related analysis described here (2/16 of females and 2/16 males).

UMAP visualization

UMAP (Uniform Manifold Approximation and Projection) based dimensionality reduction was performed for visualization purposes only, with n_neighbors=8, min_dist=0.1, n_components=2 and metric = Euclidean. Syllable usages in all sessions were used and z-scored before UMAP analysis was applied.

Please cite this article in press as: Levy et al., Mouse spontaneous behavior reflects individual variation rather than estrous state, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.02.035

Clustering analysis

Clustering was performed using a k-means algorithm applied on z-scored sessions-based syllable usage with 10 random seed initiations and 300 iterations. Clustering analysis was performed twice: i) with $n_{clusters}$=4, the results of which were compared to labeling of sessions by phase, and ii) with $n_{clusters}$=16, the results of which were compared to labeling of session by individual identity (using adjusted Rand index. 40 initial random seeds values were compared and maximum ARI is presented).

General linear model

A linear regression model (ElasticNet with α=0.01) was fit to model log-transformed, z-scored, session based, syllable usage for each syllable as a function of either): i) phase; ii) phase, with phase labels shuffled within a mouse (n=100); iii) phase, identity and their interaction; iv) phase and identity without interaction; v) mouse identity; vi) mouse identity, with shuffled identity labels (n=100). Coefficient of determination (R^2) for each model are presented in Figure 3B as well as parameters for model iv. Performance of models iii,iv and v was compared using ANOVA for each syllable in order to explore the influence of adding phase information on explained variability.

Kullback-Leibler Divergence (KLD) analysis

Kullback-Leibler Divergence (KLD) analysis (relative entropy) was performed to calculate the pairwise dissimilarity of syllable usage distributions (P,Q) using the scipy.stats.entropy module such that $D_{KL}(P\|Q) = \sum_{x \in X} P(x) log\left(\frac{P(x)}{Q(x)}\right)$.

For phase, syllable usage distributions were averaged per phase, per mouse. KLD between phases was then calculated for all possible pairwise combinations of phases within each mouse, and then averaged to give a single value indicating the overall dissimilarity of estrous phases within a single mouse. Dissimilarity of sessions within each individual (referred to as ''intra-individual'' across figures) was calculated as KLD for all possible pairwise combinations of sessions recorded for that individual mouse, and then averaged to give a single value indicating the overall dissimilarity of behavior across sessions for each single mouse. Dissimilarity of individuals (referred to as ''inter-individual'' across figures) was measured by calculating the pairwise KLDs of average syllable usage distribution of each mouse against all other mice, and then averaged to produce a single value for each mouse indicating its overall dissimilarity from all other mice. For ''all sessions'' condition presented across figures, KLD was calculated for each behavioral session against all other sessions (in all pairwise combinations) and then averaged per session, to give a single number indicating the dissimilarity of behavior observed in that session in comparison to all other sessions. For analysis of dynamics of variability presented in Figures 4A and 4B, the above was performed separately for each half of the experimental days (first 7 days – 1st week, versus second 7 days – 2nd week). For variability analysis presented in Figures S3C and S3D, measurements from the first week of female data was compared to those from the second week of male data. For analysis in Figures S3E and S3F, an independent experimental data was used controlling for mouse handling as described above. Analysis in Figures S4C and S4D excludes putative size-correlated syllables as described above. KLD values of self-distances (KLD=0) were removed from all analysis.

Coefficient of Variation (CV) analysis

Coefficient of variation analysis was performed to estimate the variation in usage of each syllable across phases and individuals such that $CV = \frac{SD}{\bar{X}}$. For phases, CV for each syllable was calculated across phases within each mouse (using the mean syllable usage across repetitions per phase), and then averaged across mice to give a single CV value per syllable. For intra-individual analysis, CV was calculated per syllable across repetition for each mouse, and then averaged across mice to give a single measurement per syllables. For inter-individual condition, CV for each syllable was calculated across mice, using the mean syllable usage across sessions per mouse. For ''all sessions'' analysis, CV for each syllable was calculated across all behavioral sessions. Details of specific dataset used for CV analysis in each figure is as detailed for KLD analysis above, and mentioned in the appropriate figure legends.

CV analysis shown in Figure S2F was calculated to test for the dynamic of syllable use throughout a single session. First, each 20 minute-long behavioral session was divided into five minute non-overlapping bins, and the syllable usage distribution was calculated for each bin. The CV across these bins was calculated for each syllable per session, and then averaged per estrous phase to give a single number indicating the variation in usage of a single syllable throughout the 20 minutes session for each phase.

Mutual information (MI) analysis

Mutual information was calculated to evaluate the information each syllable holds regarding the differences between phases or individuals using the sklearn.feature_selection.mutual_info_classif module, to evaluate MI between continuous data set (syllable usage) and discrete target (either phase or identity).

QUANTIFICATION AND STATISTICAL ANALYSIS

All quantification and statistical analysis were done in Python 3.6, using the following modules: Numpy,[41] Pandas,[43] Scikit-learn,[44] Scipy,[45] Statsmodels,[46] Seaborn[42] and Matplotlib.[47] Specific details of statistical test, statistics, number of samples and p-values are described in the figure legends. Null hypothesis was rejected with α>0.05. When appropriate, p-values were adjusted for multiple comparisons using the Bonferroni method.

Current Biology, Volume *33*

Supplemental Information

Mouse spontaneous behavior reflects individual variation rather than estrous state

Dana Rubi Levy, Nigel Hunter, Sherry Lin, Emma Marie Robinson, Winthrop Gillis, Eli Benjamin Conlin, Rockwell Anyoha, Rebecca M. Shansky, and Sandeep Robert Datta

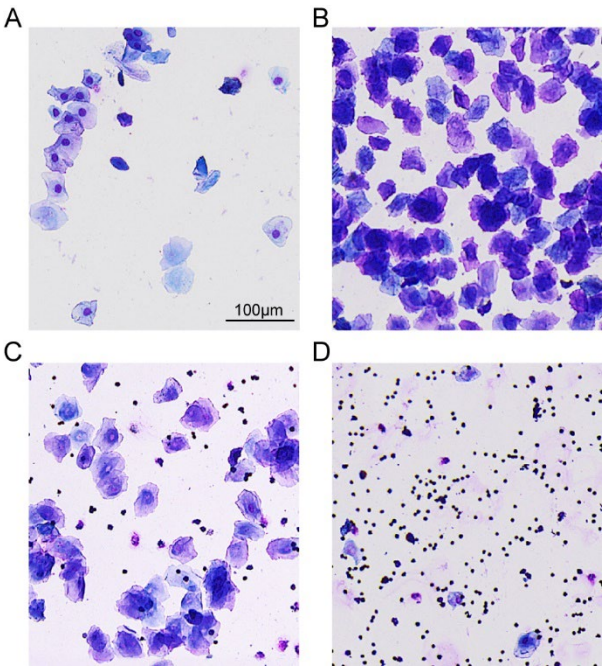

Figure S1: Identification of the four estrous phases, related to Figure 1.

Representative images of vaginal smears used to define the four estrous stages in female mice: **A)** Proestrus; **B)** Estrus; **C)** Metestrus; **D)** Diestrus

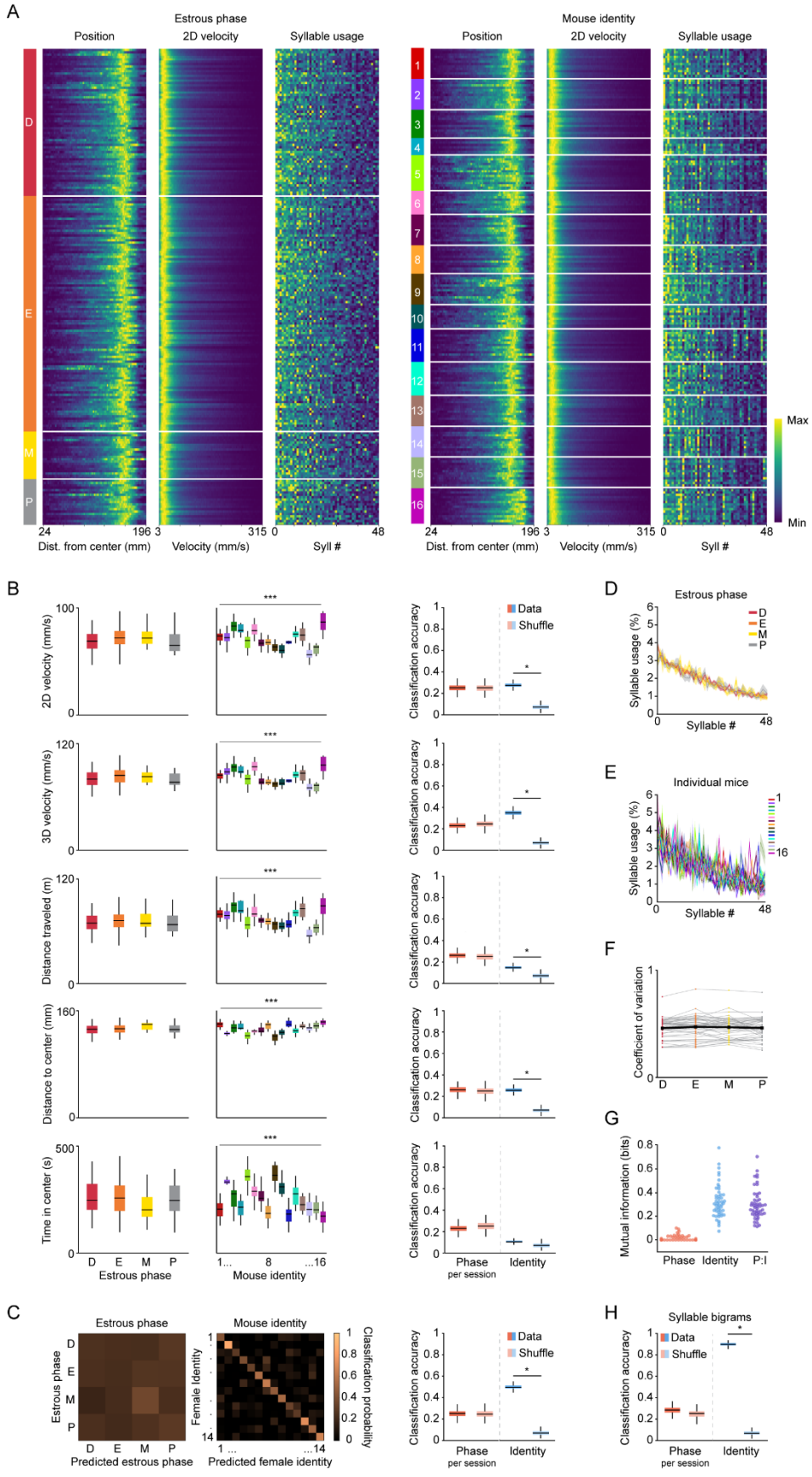
A
Estrous phase
Position
2D velocity
Syllable usage
Mouse identity
D
E
M
P
Max
Min
Dist. from center (mm)
Velocity (mm/s)
Syll #
B
2D velocity (mm/s)
3D velocity (mm/s)
Distance traveled (m)
Distance to center (mm)
Time in center (s)
Estrous phase
Mouse identity
Classification accuracy
Data
Shuffle
Phase
Identity
per session
C
Predicted estrous phase
Predicted female identity
Female identity
Classification probability
D
Syllable usage (%)
Syllable #
E
Individual mice
F
Coefficient of variation
G
Mutual information (bits)
Phase
Identity
P:I
H
Syllable bigrams

Figure S2: Estrous phase negligibly influences exploratory behavior, related to Figures 2.

A) Heatmaps depicting position, velocity, and syllable usages across all female dataset sorted by phase (left) or individual identity (right). For each measurement, values were binned into 49 bins (to match the number of syllables) and the colormap represents occupancy in each bin, normalized to max/min values for each session, and syllables are sorted by use across all sessions. White lines separate different phases or individuals. **B)** Left: Quantification of female behavior in the open field across estrous phases and in individual mice, as assessed via traditional kinematic scalar metrics. Kruskal-Wallis H-test was performed for all: for 2D velocity: $H_{phase(3)}$=5.56, p=0.13; $H_{individual(15)}$=128.803, p=3.64*10^{-20}. For 3D velocity: $H_{phase(3)}$=5.47, p=0.14; $H_{individual(15)}$=125.09, p=1.93*10^{-19}. For distance traveled: $H_{phase(3)}$=3.58, p=0.30; $H_{individual(15)}$=118.88, p=3.11*10^{-18}. For distance to center: $H_{phase(3)}$=5.27, p=0.15; $H_{individual(15)}$=128.44, p=4.28*10^{-20}. For time in center: $H_{phase(3)}$=4.19, p=0.24; $H_{individual(15)}$=133.66, p=4.06*10^{-21}. n sessions per phase: D=62, E=99, M=20, P=19; n sessions per individuals: 8-15. Right: Quantification of overall decoder performance for each scalar, matched to left panel. **C)** Left: Confusion matrix for classification probability of a decoder trained to predict estrous phase (left) or individual mouse identity (right) based on all scalars described in B concatenated. Right: Classification accuracy. **D-E)** Mean syllable usage distribution for phase **(D)** and individual mice **(E)**. Shaded area represents standard error of the mean. Kruskal–Wallis tests with Bonferroni correction were performed to assess differences in the use of each syllable between estrous phases (none significant). Syllables are sorted by use across all sessions, and the sorting is maintained for both panels. **F)** The overall variability of behavior over time during each session does not vary based upon estrous cycle. To assess this within-session variation, the coefficient of variation of syllable use during each 20 minute behavioral session was computed and graphed, with data binned into 5 minute chunks. Results are presented per syllable per phase, with gray lines representing individual syllables; Thick black line represents the mean across syllables per phase. Friedman test for repeated samples was performed. $X^2_{(3)}$=3.375, p=0.337. **G)** Mutual information analysis for the association of syllable use with estrous phase, individual identity and the interaction of phase and identity. Each datapoint represents a single syllable. Top 10 syllables associated with differences between phases included: runs and darts, walks, running left/right, groom, rears and pause. Syllables associated with differences between individuals include: rears against the arena wall, darts, rears, and turns, although almost every syllable had greater mutual information with individual identity than information about estrous phase. **H)** Classification accuracy of decoders trained to predict phase and mouse identity based on syllable bigrams (sequences of two syllables, see Methods). For all relevant panels, Box plots depict median, interquartile range, and upper/lower adjacent values (black lines). For all decoder panels: asterisk (*)

denotes statistical significance, here indicating that the mean decoding distribution exceeds the 95th percentile shuffle threshold. For all other panels: ***p<0.001

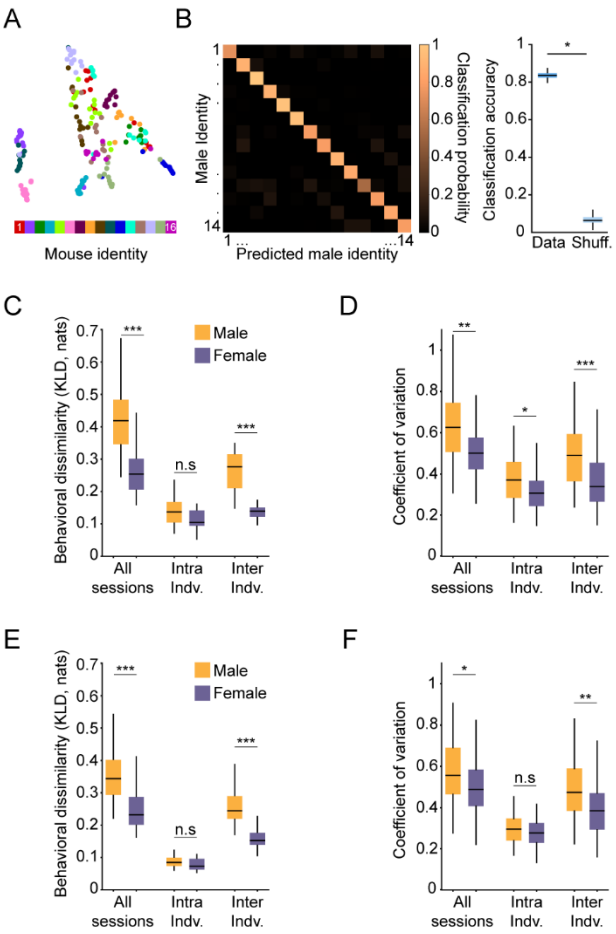

Figure S3: Male behavior is more variable than female behavior in the open field. related to Figure 4.

A) UMAP plot depicting syllable usage in males for each session, colored by mouse identity. To assess cluster quality, K-means clustering analysis was performed on high-dimensional data and clustering quality compared to true labels was quantified using the Adjusted Rand Index (ARI). For individuals (number of clusters = 16) ARI = 0.47. **B)** Classification accuracy for male individual identity based on syllable usages. Left: confusion matrix for decoder performance. Right: overall decoder performance is presented against shuffled data (Shuff.). Asterisk (*) denotes statistical significance, here indicating that the mean decoding distribution exceeds the 95th percentile shuffle threshold. **C)** Comparison of behavioral variability as measured by the KLD of syllable usage distributions during the first week of female behavioral sessions and the last week of male behavioral sessions. Pairwise comparisons were done between syllable usage distributions in each behavioral sessions (“all sessions”), between sessions within each individual (intra indv.), and between individuals (inter indv.). 2-way ANOVA for sex and experimental conditions (exp) as main factors was performed: $Fsex_{(1,249)}=34.8$, $p=1.18*10^{-8}$, $Fexp_{(2,249)}=18.41$, $p=3.47*10^{-8}$, $Fsex*exp_{(2,249)}=1.95$, $p=0.144$. Individual contrasts were performed using student’s t-test with Bonferroni

correction: $p_{all\ sessions}=1.09*10^{-6}$, $n_{female}=93$ sessions, $n_{male}=98$ sessions; $p_{intra\text{-}indv}=1$; $p_{inter\text{-}indv}=2.9*10^{-6}$, n=16 mice for male and females. **D**) Same as in C, but for CV of syllable usage, calculated per syllable. $Fsex_{(1,288)}=31.7$, $p=4.18*10^{-8}$, $Fexp_{(2,288)}=56.46$, $p=2.04*10^{-21}$, $Fsex*exp_{(2,288)}=1.75$, p=0.176. Individual contrasts were performed using student's t-test with Bonferroni correction: $p_{all\ sessions}=0.0027$; $p_{intra\text{-}indv}=0.036$; $p_{inter\text{-}indv}=0.0009$. $n_{mice}=16$ and $n_{syllables}=49$ for both male and females. **E)** Same as C, but for an independent dataset, comparing male and female behavioral variability when matching handling conditions and following prolonged (10 days) habituation to the experimental setup. $Fsex_{(1,252)}=88.57$, $p=3.2*10^{-18}$; $Fexp_{(2,252)}=128.39$, $p=3.56*10^{-39}$, $Fsex*exp_{(2,252)}=4.77$, p=0.009. Individual contrasts were performed using student's t-test with Bonferroni correction: $p_{all\ sessions}=1.14*10^{-14}$, $n_{female}=94$ sessions, $n_{male}=100$ sessions; $p_{intra\text{-}indv}=0.51$; $p_{inter\text{-}indv}=0.0008$, n=16 mice for male and females. **F)** Same as in E, but measured as CV of syllable usage, per syllable. For left panel: $Fsex_{(1,288)}=17.97$, $p=3.02*10^{-5}$; $Fexp_{(2,288)}=102.83$, $p=1.9*10^{-34}$, $Fsex*exp_{(2,288)}=2.37$, p=0.09. Individual contrasts were performed using student's t-test with Bonferroni correction: $p_{all\ sessions}=0.036$; $p_{intra\text{-}indv}=0.63$; $p_{inter\text{-}indv}=0.0051$, $n_{mice}=16$ and $n_{syllables}=49$ for both male and females. For all relevant panels, Box plots depict median, interquartile range, and upper/lower adjacent values (black lines). *p<0.05, **p<0.01, ***p<0.001, n.s = not significant.

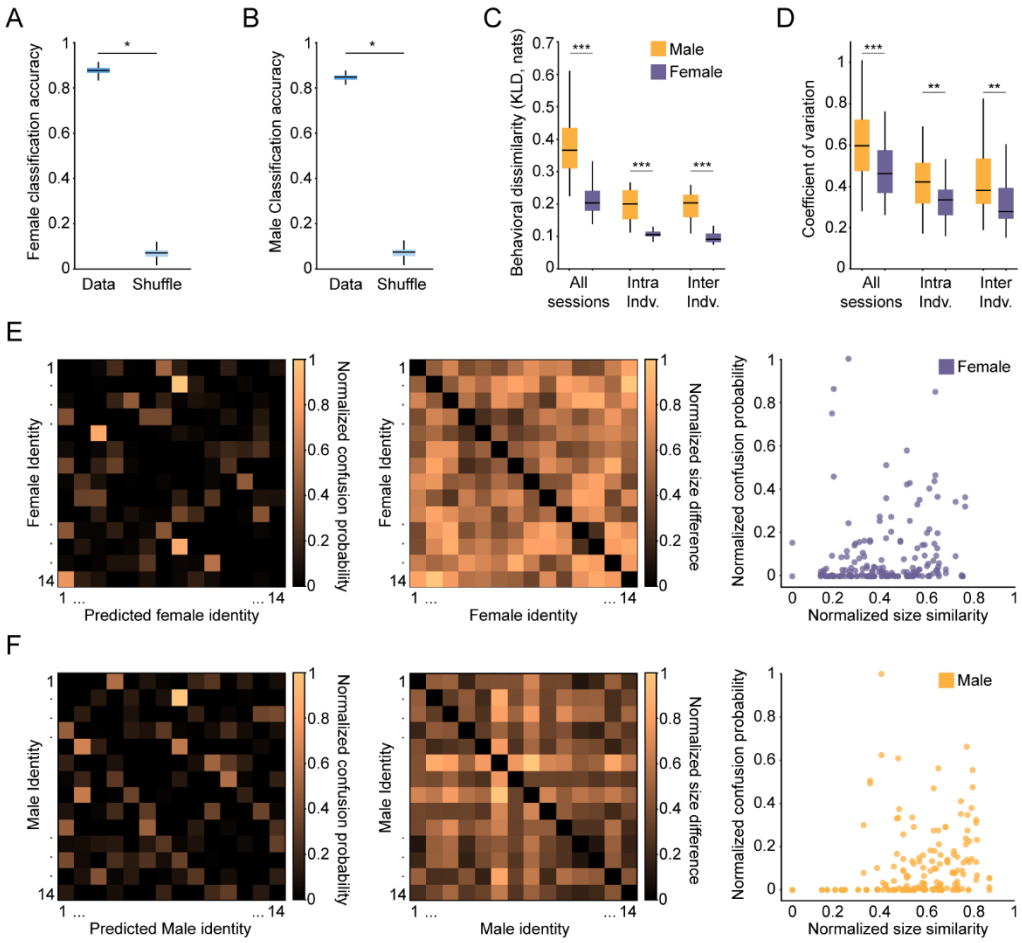

Figure S4: Size differences do not account for inter-individual behavioral variability as measured by MoSeq, related to Figure 4.

A) Classification accuracy for female individual identity based on syllable usages, after exclusion of syllable whose usage correlates with mouse size (see Methods). Asterisk (*) denotes statistical significance, here indicating that the mean decoding distribution exceeds the 95th percentile shuffle threshold. **B)** Same as A but for male individual identity. **C-D)** Behavioral variability analysis as in Fig. 4A and 4B but after exclusion of syllables whose usage correlates with mouse size (see Methods). **C)** Comparison of male and female behavioral variability between all behavioral sessions, between sessions within each individual (intra indv), and between individuals (inter indv.) as measured by KLD. 2-way ANOVA for sex and experimental conditions (exp) as main factors was performed: $Fsex_{(1,438)}=87.1$, $p=5.2*10^{-19}$; $Fexp_{(2,438)}=16.19$, $p=1.64*10^{-7}$, $Fsex*exp_{(2,438)}=2.14$, $p=0.117$. Individual contrasts were performed using student's t-test with Bonferroni correction: $p_{all\ sessions}=1.35*10^{-15}$, $n_{female}=188$ sessions, $n_{male}=192$ sessions; $p_{intra\text{-}indv}=6.87*10^{-6}$; $p_{inter\text{-}indv}=9.93*10^{-7}$, n=16 mice for male and females. **D)** Same as in C, but depicting the distribution of coefficients of variation of the usage of each syllable. For left panel: 2-way ANOVA: $Fsex_{(1,219)}=36.33$, $p=6.96*10^{-9}$;

$F_{exp(2,219)}$=33.42, p=2.15*10^{-13}, $F_{sex*exp(2,219)}$=0.55, p=0.57. Individual contrasts were performed using student's t-test with Bonferroni correction: $p_{all\ sessions}$=0.0009; $p_{intra-indv}$=0.0024; $p_{inter-indv}$=0.0015. $n_{syllables}$=41 for males and $n_{syllables}$=34 for females. **E)** Left: Confusion matrix depicting the output of a "held-out" classifier, which enables classifier-based quantification of the similarity of behavioral patterns expressed by pairs of female mice (see Methods). Middle: Euclidian distances between size measurements of individual females (see Methods). Size differences and classification probabilities, depicted by colormap, were normalized to 0-1 scale. Right: Size similarity (defined as "1-x", x=size difference) plotted against decoder confusion probabilities. For linear fit, R^2 = 0.014. Self-distances were removed from the analysis. **F)** Same as E, but for males. For linear fit, R^2 =0.03. For all relevant panels, Box plots depict median, interquartile range, and upper/lower adjacent values (black lines). *p<0.05, **p<0.01, ***p<0.001

Syllable ID	Associated behavior	MI score phase	MI score identity	CV between individuals
0	Rear	0	0.29	0.24
1	Dart	0	0.57	0.24
2	Low rear	0	0.32	0.36
3	Groom	0.01	0.32	0.19
4	Down from wall rear	0	0.2	0.16
5	Run left	0.04	0.44	0.26
6	Groom	0	0.21	0.15
7	Pause before rear	0	0.35	0.24
8	Turn left	0	0.61	0.27
9	Walk	0.03	0.37	0.34
10	Run	0.03	0.2	0.27
11	Short dart	0.01	0.3	0.24
12	Walk	0.06	0.16	0.25
13	Groom	0.02	0.45	0.22
14	Groom	0.0	0.07	0.19
15	Short dart	0.1	0.38	0.33
16	Dart	0.07	0.35	0.49
17	Down from rear	0.02	0.37	0.34
18	Run right	0.0	0.35	0.27
19	Groom	0.09	0.25	0.38
20	Rear	0.0	0.31	0.24
21	Rear	0.02	0.25	0.40
22	Walk	0.0	0.14	0.26
23	Low rear	0.0	0.2	0.26
24	Down from wall rear	0.03	0.4	0.24
25	Dart	0.0	0.22	0.20
26	Rear	0.0	0.33	0.27
27	Scrunch	0.0	0.38	0.52
28	Wall rear	0.0	0.77	0.52
29	Run	0.04	0.17	0.22
30	Wall rear	0.0	0.64	0.48
31	Wall rear	0.02	0.24	0.25
32	Run right	0.05	0.21	0.26
33	Short pause	0.0	0.31	0.36
34	Scrunch	0.0	0.28	0.24
35	Groom	0.01	0.2	0.31
36	Rear	0.03	0.43	0.56
37	High rear	0.0	0.25	0.28
38	Stretch	0.0	0.25	0.63
39	Rear	0.0	0.28	0.40
40	Wall rear	0.0	0.5	0.45
41	Pause	0.03	0.12	0.16
42	Run left	0.0	0.47	0.60
43	Groom	0.0	0.29	0.41
44	High rear	0.0	0.29	0.35
45	Low rear	0.0	0.23	0.45
46	Rear	0.04	0.55	0.72
47	Low rear	0.0	0.32	0.72
48	High rear	0.02	0.26	0.49

Table S1: Syllable labels for females, related to Figures 2-4 and S2.

This table lists all of the behavioral syllable identified in the MoSeq model for female mice: the human-annotated behaviors associated with each syllable, as well as the degree of mutual information for estrous phase and individual identity (see Methods); also shown is the coefficient of variation of the usage of each syllable across individuals.

Syllable ID	Associated behavior	MI score identity	CV between individuals
0	Pause	0.53	0.22
1	Low rear	0.51	0.38
2	Stretch	0.52	0.36
3	Dart	0.49	0.37
4	Short rear	0.55	0.46
5	Down from rear	0.56	0.37
6	Wall rear	0.27	0.19
7	Walking	0.27	0.30
8	Pause	0.08	0.26
9	Run	0.45	0.42
10	Turn left	0.39	0.28
11	Short pause	0.22	0.32
12	Short dart	0.12	0.18
13	Groom	0.77	0.57
14	Run right	0.53	0.45
15	Low rear	0.54	0.50
16	Run	0.63	0.61
17	High rear	0.25	0.28
18	Down from rear	0.34	0.39
19	Pause	0.19	0.44
20	Dart	0.24	0.36
21	Turn right	0.75	0.42
22	Walk	0.23	0.31
23	Rear	0.37	0.32
24	Short dart	0.44	0.53
25	Rear	0.22	0.27
26	High rear	0.38	0.53
27	High rear	0.42	0.55
28	Down from wall rear	0.31	0.25
29	Run right	0.28	0.41
30	Groom	0.38	0.36
31	High rear	0.47	0.53
32	Low rear	0.48	0.67
33	Rear	0.56	0.55
34	High rear	0.46	0.77
35	Rear	0.36	0.51
36	Run left	0.39	0.37
37	Pause	0.52	0.50
38	Scrunch	0.85	0.82
39	Run right	0.45	0.42
40	Groom	0.24	0.50
41	Run	0.37	0.56
42	Groom	0.21	0.37
43	Groom	0.23	0.35
44	Wall rear	0.33	0.29
45	Stretch	0.59	0.95
46	Stretch	0.29	0.56
47	High rear	0.4	0.71
48	Rear	0.33	0.44

Table S2: Syllable labels for males, related to Figure 4 and S3.

This table lists all of the behavioral syllable identified in the MoSeq model for male mice: the human-annotated behaviors associated with each syllable, as well as the degree of mutual information for individual identity (see Methods); also shown is the coefficient of variation of the usage of each syllable across individuals.