Physics of Living Systems

Modelling collective behavior in groups of mice housed under semi-naturalistic conditions

Xiaowen Chen, Maciej Winiarski, Alicja Puścian, Ewelina Knapska, Thierry Mora, Aleksandra M Walczak

Laboratoire de physique de l'École normale supérieure, CNRS, PSL University, Sorbonne Université, and Université Paris Cité, Paris, France • Center of Excellence for Neural Plasticity and Brain Disorders: BRAINCITY, a Nencki-EMBL Partnership, Nencki Institute of Experimental Biology of Polish Academy of Sciences, Warsaw, Poland

Reviewed Preprint
v3 • April 7, 2025

Reviewed Preprint
v2 • March 24, 2025
Revised by authors

Reviewed Preprint
v1 • April 26, 2024

eLife Assessment

This **valuable** work investigates the social interactions of mice living together in a system of multiple connected cages. It provides **solid** evidence for a statistical approach capturing changes in social interactions after manipulating prefrontal cortical plasticity. This research will be of broad interest to researchers studying animal social behaviour.

https://doi.org/10.7554/eLife.94999.2.sa3

Abstract

In social behavior research, the focus often remains on animal dyads, limiting the understanding of complex interactions. Recent trends favor naturalistic setups, offering unique insights into intricate social behaviors. Social behavior stems from chance, individual preferences, and group dynamics, necessitating high-resolution quantitative measurements and statistical modeling. This study leverages the Eco-HAB system, an automated experimental setup which employs RFID tracking to observe naturally formed mouse cohorts in a controlled yet naturalistic setting, and uses statistical inference models to decipher rules governing the collective dynamics of groups of 10-15 individuals. Applying maximum entropy models unveils social rules in mouse hordes, quantifying sociability through pairwise interactions within groups, exploring how social structure evolves, the impact of individual versus social preferences, and the effects of considering interaction structures among three animals instead of two. Reproducing co-localization patterns of individual mice reveals stability over time, with the inferred interaction strength capturing social structure. By separating interactions from individual preferences, the study demonstrates that altering neuronal plasticity in the prelimbic cortex – the brain structure crucial for sociability – does not eliminate social interactions, but makes the transmission of social information between mice more challenging. The study demonstrates how the joint probability distribution of the mice positions can be used to quantify sociability.

I. Introduction

Social behavior is fundamental for numerous animal species, encompassing human societies. From the dynamic spectacle of Mexican waves in a football stadium to the intricate waggle dance of bees, the diverse manifestations of social interaction raise a pivotal question: How do these social behaviors come to fruition, and what roles do individuals play in their emergence?

In recent decades, the exploration of social behavior has predominantly centered around studying animal dyads in controlled laboratory conditions. However, these experimental paradigms inherently impose limitations on investigating intricate social behaviors that often involve more than two interacting individuals. Studies on social interactions frequently employ tests with brief observation periods, during which animals are evaluated in novel environments, accompanied by the presence of an experimenter that induces stress, influencing social behavior [1–5]. Recently, there has been a notable shift towards conducting experiments in natural settings, involving animal groups such as flocks of birds [6] or swarms of midges [7]. Additionally, semi-naturalistic environments, exemplified by fish in tanks [8–10], marching locusts in arenas [11], flocks of sheep [12], and hordes of rodents [13, 14], are increasingly being utilized. These approaches present unique opportunities for the comprehensive quantification of complex social interactions and sociability.

Mice stand out as a valuable model system for delving into the complexities of social behavior, given their intricate manifestation of various social behaviors. They tend to form cohesive groups, showcasing both amicable and agonistic behaviors. Depending on the environmental context, mice demonstrate territoriality and dynamic social hierarchies [15]. Communication among mice is extensive, primarily mediated through odors, allowing them to convey emotional states such as stress, fear, and preferences in food [16, 17]. Additionally, mice exhibit prosocial behaviors, actively assisting distressed fellow mice in need [18]. Decades of research have extensively explored social interactions between pairs of mice, while the study of mouse groups has only recently become feasible with advancements in high-throughput technologies, particularly radiofrequency identification (RFID) [14, 19]. The Eco-HAB system, utilized in this study, leverages RFID tracking to observe naturally-formed cohorts of mice in a controlled yet naturalistic environment, enabling longitudinal experiments on sociability with minimal human interference [14].

Social behavior arises from chance, individual preferences, group structure, and the transmission of preferences and interactions among group members. To unravel these elements and understand the establishment of social networks and hierarchies, we need high-resolution quantitative measurements of behavior over extended periods and statistical modeling to capture natural variability. In this study, we integrate Eco-HAB recordings with statistical inference to construct models of collective behavior, focusing on the statistics of system states to identify interaction structures within the group.

The prefrontal cortex (PFC) plays a crucial role in processing social information, understanding others' emotions, maintaining social hierarchy, and transmitting information about food safety in both rodents and humans [20, 21]. Neuronal activity of the PFC is correlated with proximity to conspecifics, and studies in mice reveal distinct PFC responses to social and non-social olfactory stimuli [22, 23]. The PFC integrates existing knowledge with new information about self and others, demonstrating dynamic neuronal plasticity [24]. In cognitive tasks involving the PFC and subcortical areas, neuronal connectivity refines more rapidly in the former, highlighting its adaptability to changing environments [25, 26]. Tissue inhibitors of metalloproteinases (TIMPs), particularly TIMP-1, influence synaptic plasticity by inhibiting matrix metalloproteinases (MMPs), especially MMP-9 [27–31]. TIMP-1 is involved in long-term potentiation (LTP), a

crucial process for cellular-level memory formation [32, 33]. This sustained release impedes the updating of neuronal connectivity in the prelimbic part of the PFC (PL), crucial for maintaining social structure [34–36]. Our study employs nanoparticles for gradual TIMP-1 release over several days [37] to impact plasticity in the PL on the changes in group behavior.

Maximum entropy models successfully explain social rules governing collective behavior in bird flocks and mouse hordes [38, 39]. These models help distinguish observed correlations, like the clustering of mice in a specific location, from direct interactions or individual preferences. [39] pioneered the use of these models in studying mouse group behavior, revealing the significance of higher-order interactions in co-localization patterns. While [39] utilized video tracking of groups of four mice, our Eco-HAB setup employs RFID technology for tracking groups of 10-15 mice, providing more compact data with longer recording times but lower spatial resolution. Our focus is on quantifying sociability in mouse hordes through pairwise interactions within groups, ensuring statistical power. We explore whether interactions between pairs can explain collective behavior and examine how social structure evolves over time. We analyze the effects of individual versus social preferences and investigate the impact of considering three animals instead of two. Using a data analysis approach based on wild type C57BL6/J male and female mice, we discuss social structure and sociability changes in mice with temporary prefrontal cortex plasticity modification.

II. Results

A. Recording of mice location in naturalistic environment

Eco-HAB is an automated, ethologically-relevant experimental apparatus that tracts voluntary behavior in group-housed mice [14]. Constructed to simulate notable characteristics of natural murine environment, it consists of four connected large compartments, two of which contain food and water (**Fig. 1A**). Cohorts of 10 to 15 mice are introduced into the Eco-HAB, where they behave freely while their locations are tracked over time. The details of used mouse strains and cohorts' compositions can be found in **Material and Methods**. The compartments are connected with tube-shaped corridors resembling underground tunnels, on whose ends there are 125kHz antennas recording every time a mouse crosses with an accuracy of over 20Hz. Each mouse is tagged with a unique RFID tag. The mice are recorded for 10 days with alternating 12-hour-long light-dark phases that simulate the day-night cycle.

The location of each mouse at each time is reconstructed using the time stamps, reducing the data to a discrete time series, σ_t at time $t = 1, 2, \ldots, T$, with possible values of the locations $\sigma_t = 1, 2, 3, 4$ corresponding to the four compartments. The time resolution for the discretization is set to 2 seconds. As shown by the color-coded location traces in **Fig. 1B**, the majority of mice are often found in the same compartment, especially in the non-active light phases: this corresponds to the ethological behavior – mice tend to sleep in a pile to keep each other warm (see [40]). This suggests that the behavior depends on latent variables, i.a. the circadian clock.

To ensure the relative consistency of the analyzed data, we used the observed rate of transitions to measure the activity of each mouse, and choose an analysis window of 6 hours covering the first half of the dark phase (13:00 - 19:00), which corresponds to the most active time on each day for the entire duration of the experiment. The variability of activity across individuals is larger than the day-to-day variability for a single mouse, suggesting that the level of locomotor activity is a well-defined individual characteristic (**Fig. 1C**).

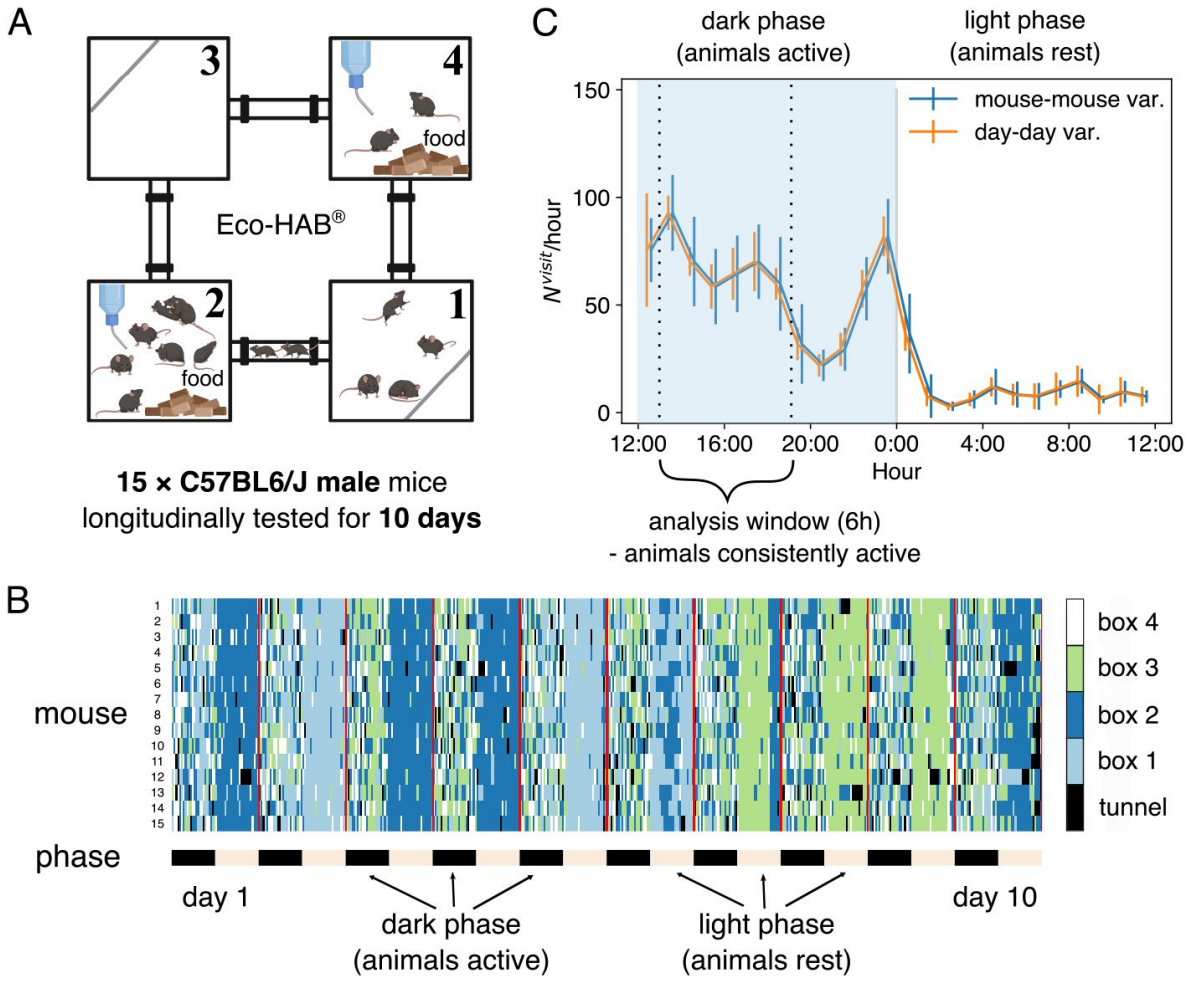


Fig. 1

Mice were tested in Eco-HAB, a system for automated, ecologically-relevant assessment of voluntary behavior in groups of mice. Animals were tested for 10 days.

(A) Schematic of the Eco-HAB system, where four compartments are connected with tunnels. Food and water are available *ad libitum* in compartments 2 and 4. (B) Time series of the location of 15 mice over 10 days, as aligned to the daylight cycle. (C) The activity of the mice is affected by the circadian clock. Error bars represent standard deviation across all mice (mouse-mouse variability, in *blue*) or across all days for the mean activity level for all mice (day-day variability, in *orange*). The two curves are slightly shifted horizontally for clearer visualization. We focus the following analysis on the data collected during the first half of the dark phase, between 13:00 and 19:00 (shaded region).

Pairwise interaction model explains the statistics of social behavior

We first establish a quantification of sociability by building probabilistic interaction models for groups of mice. Following previous work [14], we use the in-cohort sociability, which measures the excess probability of two mice being found in the same compartment compared to the case where they are independent. Mathematically, in-cohort sociability is defined as:

$$C_{ij}^{\text{data}} = \sum_{r=1}^{k} \left(f_{ij}(r,r) - f_i(r) f_j(r) \right) \qquad (1)$$

where $f_i(r) \equiv m_{ir}^{\text{data}}$ and $f_{ij}(r, r')$ are respectively the empirical frequencies of finding a mouse i in compartment r, and a pair of mice (i, j) in compartments r, r'.

As schematically explained in **Fig. 2A**, in-cohort sociability is due to pairwise interactions between each pair of mice, and modifies how likely they are to be found in each compartment with respect to the mice's innate preference for that compartment. However, considering the presence of more than two animals, in-cohort sociability is not an effective measure of social structure of the group: two animals with zero attraction to one another can still be found to have a high in-cohort sociability, if a third animal has a strong social bond with both of them, since they all will be spending time with one another.

Since measurements of location preference, and the incohort sociability, together with the dynamical observables such as the rate of activity, are stable over time (**SI Fig. S1**) it invites a quantitative modeling of the joint-probability distribution of the co-localization of mice.

To distinguish social structure interactions from the effective correlations that define in-cohort sociability, we build a maximum entropy model with pairwise interactions. This approach constrains the joint probability distribution of all the possible co-localization patterns of all mice to reproduce the empirical occupation frequencies m_{ir}^{data} and the in-cohort sociability C_{ij}^{data}, while otherwise remaining as random as possible [39, 41, 42]. With these assumptions, the joint probability distribution of the mice co-localization patterns can be written as

$$P^{(2)}(\sigma) \propto e^{\sum_{i,r} h_{ir} \delta_{\sigma_i, r} + \sum_{i<j} J_{ij} \delta_{\sigma_i, \sigma_j}}, \qquad (2)$$

where h_{ir} is the individual preference of mouse i to be in compartment r, and J_{ij} is the interaction between mouse i and mouse j. The set of parameters $(\{h_{ir}, J_{ij}\})$ is learned through gradient descent (see **SI Fig. S2** and **Materials and Methods** for details). The interactions J_{ij} may be positive or negative. We see that although the structure of the interactions J_{ij} follows that of incohort sociability C_{ij}, they are not identical (**SI Fig. S3**). Likewise, individual mice preferences h_{ir} are not equal to the occupation probability m_{ir} (**Fig. 2C**). Thus, this approach allows us to distinguish direct interactions from indirect ones.

To validate the model, we tested that it is able to predict higher order features of the data, such as the probability of a specific combination of triplets of mice being in the same compartment (**Fig. 2BD**), and the probability of observing K mice in the same compartment (**Fig. 2E**), with the overestimation at large K possibly due to the limit of finite data. Although the model assumes the strength of interaction does not depend on which compartment the mice are in, our minimal model can predict probability of K mice in certain compartments (**SI Fig. S4**, compartments 1 and 3). We call in-state probability the distribution of box occupancy of each mouse given the position of all other mice (see **Material and Methods**). The model prediction for in-state probabilities match the observed one, showing that the model gives an unbiased estimate of individual mouse positions **SI Fig. S5**). Moreover, models with triplet interactions show signs of overfitting under cross-validation, which is mitigated when the triplet interactions are suppressed

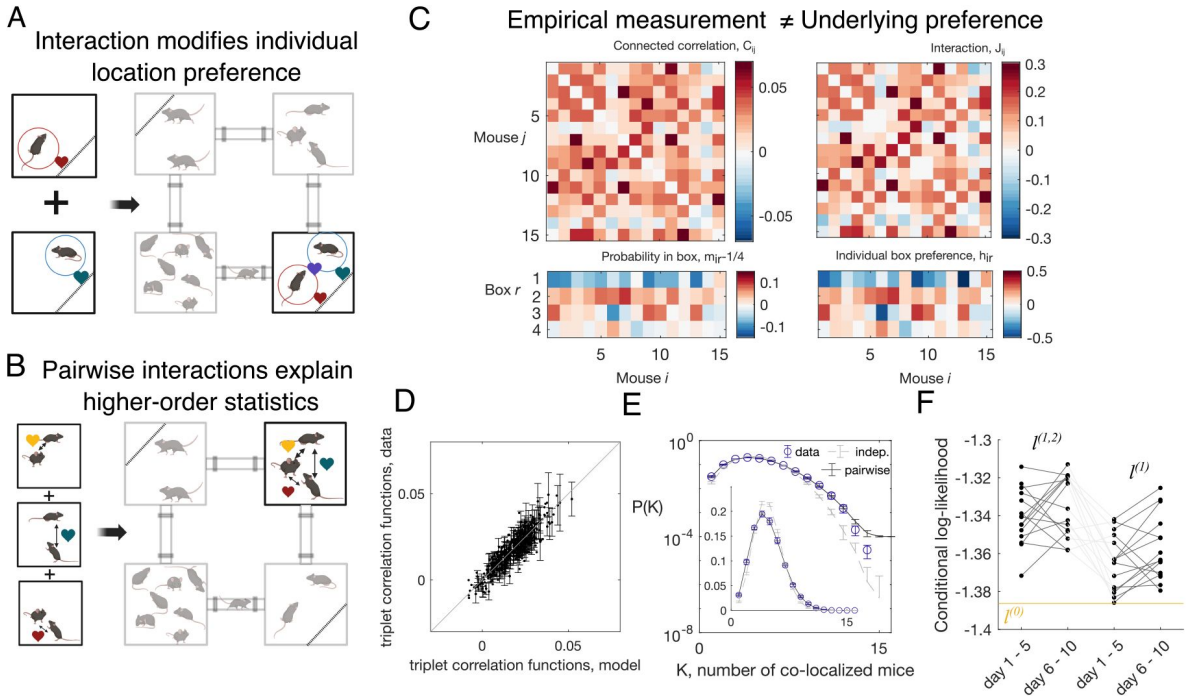


Fig. 2

Mice in Eco-HAB interact pairwisely.

(A, B) The schematics showing pairwise interactions: two mice are more likely to be found in the same compartment than the sum of their individual preference implies (A); also, the probability for three mice being in the same compartment can be predicted from the pairwise interactions (B). (C) From pairwise correlation C_{ij}, defined as the probability for mouse i and mouse j being in the same compartment (subtracted by the prediction of the independent model), and the probability for mouse i to be found in compartment r (subtracted by the model where each mouse spends equal amount of time in each of the four compartments), m_{ir} − 1⁄4, pairwise maximum entropy model learns the interaction strength between a pair of mice, J_{ij}, and the local field h_i^r, which gives the tendency for each mouse in each compartment. The data shown is the aggregated 5-day data from day 1 to day 5 of the C57 males (cohort M1). The pairwise maximum entropy model can predict higher order statistical structures of the data (schematics in panel B), such as the probability for triplets of mice being in the same compartment (subtracted by the prediction of the independent model, mathematically $f_{ijk}(r, r, r) - f_i(r)f_j(r)f_j(r)$) (D), and the probability of K mice being found in the same compartment (E). (F) Conditional log-likelihood of mouse locations, predicted by the pairwise model ($I^{(1,2)}$), the independent model ($I^{(1)}$), and the null model assuming no compartment preference or interactions ($I^{(0)}$, the yellow line), for each mice (cohort M1 before TIMP-1 treatment, N = 15).

close to zero using L2 regularization (see **Material and Methods** and **SI Fig. S6**). These results show that pairwise interaction among mice are sufficient to assess the observed collective behavior.

Choosing timescales for analysis

To construct an interaction model based on the steady state distribution, we first need to consider the proper timescales for which we average over the observables, i.e. mean and correlation of the mice co-localization patterns. If the timescale is too short, then the error of estimation may be large. More severely, the system may not have enough time to equilibrate, and the time average will not result in the steady state distribution. On the other hand, if the timescale is too long, we loose biologically meaningful information about the temporal evolution of the system, such as the adaptation of the mice in a new environment and evolution of the social interaction strength.

To identify the proper timescale, we systematically conduct cross-validation for pairwise maximum entropy models constructed using K days of data, where K = 1, 2, 3, 4, 5, 10, and each day of the data contains the 6 hours when the mice are most active. The data is then separated into training sets that consist of 5-hours of the data from each day and test sets that consist of 1-hour of the data each day. Pairwise maximum entropy models with L2 regularization with strength β_J imposed on the pairwise interactions are learned from the training set, and the training- and the test-set likelihoods are computed. As shown in **SI Fig. S7**, the test-set likelihood decreases as the regularization strength increases for cumulate data with number of days $K \geq 4$, indicating that the pairwise model generalizes well. We choose K = 5 for subsequent analysis, as it does not overfit, and it gives us temporal information about how the interaction structures may change over the 10-day experiment.

Stability of sociability over time

The data-driven model and its inferred parameters allow us to explore various aspects of social behavior. As the models are built using accumulated data from 5 days in the 10-day experiment, we first assess the temporal consistency of the chamber preference h_{ir} and the inferred interaction parameters J_{ij} of the four cohorts of C57BL6/J male mice (see **Table 1**), M1 (N = 15), M2 (N = 13), M3 (N = 10), M4 (N = 12, before BSA injection).

As shown in **SI Fig. S8**, for all four cohorts, the box preference tendency is consistent, with a Pearson's correlation coefficient between models learned from the first and the last five days being 0.6 ± 0.1. For cohort M1, M3 and M4, the distribution of the individual chamber preferences is consistent over time, with the two-sample F - test for equal variance being non-significant, and the two-sample t-test for equal mean being either non-significant or with a p-value of 0.04 for cohort M1. For cohort M2, both the mean and the variance of the chamber preferences significantly changed between the first and the last 5 days of the experiment. Nonetheless, the Pearson's correlation between the two data segments remain large at p = 0.57, indicating a consistency over time for the same cohort. Different cohorts exhibit different distributions of chamber preferences.

For the inferred interaction parameters J_{ij}, the distribution is consistent between the first 5 days and the last 5 days of the data. Specifically, for all four cohorts, the standard deviations of the interactions do not change between the first 5 days and the last 5 days of the data, as shown by two-sample F -test for equal variance. The mean of the inferred interactions does not change for cohort M1 and M3, however for cohort M2 and M4, two-sample t-test for equal mean returns a p = 0.0034 and p = 0.0073 respectively for the interactions. Notably, across from the four different cohorts, two-sample F -test with Bonferroni correction shows that the variance of all eight 5-day modeling are not significantly different. In contrast, cross-cohort comparison between male (cohort M1 before drug injection) and female (cohort F1 before drug injection) shows significantly different variance ($p < 0.01$ between the last 5 days of F1 and the first 5 days or the last 5 days of

Cohort ID	Gender	N_{mice}	T_{days}	Day 1, WT	NP	Day 1, NP
M1	male	15	10×2	180511	TIMP-1	180526
M2	male	13	10	181012	N/A	N/A
M3	male	10	10	200701	N/A	N/A
M4	male	9 (out of 12)	10×2	200713	BSA	200727
F1	female	13 (out of 14)	10×2	180413	TIMP-1	180430

TABLE 1

Summary of experiments used in this study.

The column N_{mice} gives the number of mice in the cohort used for the analysis, with cohort M4 and F1 containing dead or inactive mice after injection. The original number of mice is included in the parenthesis, and the exclusion procedure is described in **Materials and Methods: Exclude inactive and dead mice from analysis**. The column "NP" indicates the load of the injected nanoparticles. The column "Day 1" indicates the first day of observation in each of the 10-day experiment.

M1) and mean ($p < 0.001$) of the inferred interaction strength, which shows that these measures of sociability can be used to distinguish strains or genders. Nonetheless, the strengths of the individual interactions in the specific pairs of mice i and j, J_{ij}, vary more notably, as given by Pearson's correlation coefficient 0.015 ± 0.165 (see **SI Fig. S8C**). This implies that the maximum entropy model does not infer a social structure that is stable over time.

Quantifying the influence of social versus individual preferences

Further, we ask how important social interactions are for determining mice behavior, by measuring how much the data can be explained by the individual preferences for specific spaces within the territory vs. the interactions with other mice. Mice are social animals, yet they perform many behaviors based on their individual moment-to-moment needs, and it is unclear *a priori* how much the social interactions influence mice behavior in comparison to their individual preference.

For each mouse i, we consider three nested models with increasing descriptive power: first, the null model assuming each mouse has the same probability of being found in each compartment, $P^{(0)}(\sigma_i) = 1/4$; second, the independent model that assumes no interactions among mice, and the probability of finding each mouse in each compartment is solely determined by their individual preferences, $P^{(1)}(\sigma_i) = f_i(\sigma_i)$; third, the inferred pairwise interaction model based on voluntarily spending time with other mice considered, $P^{(2)}(\sigma_i \mid \{\sigma_{j\neq i}\})$, using **Eq. 2**.

We then quantified how well each model explains the data by comparing the mean log-likelihoods of finding a mouse in a given compartment, conditioned on the location of all other mice. As shown in **Fig. 2F**, including information on pairwise interactions increases the log-likelihood of the data by as much as including information on individual compartment preferences, as shown by the similar values of the probability ratios $P^{(2)}/P^{(1)}$ and $P^{(1)}/P^{(0)}$. The likelihood is consistent between the first five days and the last five days of the experiment, but exhibits variability across different cohorts of animals within the same strain (**SI Fig. S9**).

Another possible measure of sociability is the mutual information between a single mouse's location within the territory and the location of the rest of the cohort, which tells us how accurately the position of a single mouse can be predicted if the positions of all other mice are known (see details in **Material and Methods**). The possible values of the mutual information is between 0 and 2 bits, where 0 bits means no predictability, and 2 bits means perfect predictability. In our Eco-HAB data, the average mutual information for each mouse is 0.0323± 0.0151(SD) bits for cohort M1, with the largest value being 0.06 bits, indicating that despite non-zero sociability, the precise mouse position at any single moment is difficult to predict from the network.

Effect of impairing neuronal plasticity in the PL on subterritory preferences and sociability

As a next step we investigate the effects of impairing neuronal plasticity in the prelimbic cortex (PL), the brain structure containing neural circuits indispensable for both maintaining proper social interactions and encoding individual preferences [22]. To that end, we inject animals with a Tissue Inhibitor of MetalloProteinases (TIMP-1), an enzyme regulating the activity of synaptic plasticity proteins. Changing its physiological levels was previously shown to disrupt the neuronal plasticity in various brain structures [43, 44], including the prefrontal cortex (PFC) [33, 45]. Now, we ask if we can identify changes in behavioral patterns after the mice have been injected with nanoparticles (NP) gradually releasing TIMP-1 (NP-TIMP-1) into the PL [37]. It has been demonstrated in the Eco-HAB that it can reduce the mice's interest in chasing other animals (a proxy for their social ranks), and diminish persistence in seeking reward related to social olfactory cues [45]. Here, we measure the behavior of a cohort of $N = 15$ C57BL6/J mice, before and after the injection of NP-TIMP-1. A cohort of mice is introduced into the Eco-HAB and their free behavior is measured for 10 days (see schematics in **Fig. 3A**). Then, neuronal activity in the

PL of the subjects is impaired by injecting nanoparticles releasing TIMP-1 into the PL. After recovery animals are re-introduced into the Eco-HAB, and their behavior is measured for another 10 days. As a control we also have a cohort (male cohort M4, $N = 9$) that is injected with nanoparticles loaded with bovine serum albumin, a physiologically neutral substance having no impact on neuronal plasticity (BSA, vehicle). The detailed experimental procedure can be found in [45]. To provide a perspective on both sexes, a female cohort is also included in this study (female cohort F1, $N = 13$); it was processed identically to the experimental group of males described above. For each five days of the experiment, we infer a pairwise model (eq. 2) and study the changes of the inferred interactions, as well as individual preferences for specific spaces within the territory. The choice of five-day aggregated data for analysis is in line both with the proper timescales needed for the pairwise maximum entropy model to not overfit, and with the literature that TIMP-1 release from the TIMP-1-loaded nanoparticles is stable for 7-10 days after injection [37] (i.e. 2-5 days after the mice are reintroduced to Eco-HAB).

We can assess the change in both the interaction strength J_{ij} and the individual preferences for the compartments containing food Δh_i following the prolonged release of TIMP-1. The individual preferences for compartment containing food shows an increase in both its mean and its variance across all mice following treatment (**Fig. 3C**), with a return to pre-treatment levels after five days, consistent with the time course of TIMP-1 release [37]. In comparison, the control cohort M4 shows the increase in preference for the compartments containing food after injection of the BSA-loaded nanoparticles, which does not return to base level after 5 days (**SI Fig. S10C**). For the interaction strength, as shown by **Fig. 3B**, the variability of interactions J_{ij} is not significantly different before and after the TIMP-1 injection for the male cohort M1 (two-samples F -test for equal variance with Bonferroni correction; see **Materials and Methods**), although there exists a few outliers with strong interactions, which again returns to base level after 5 days of TIMP-1 injection. The increase of interaction variability is significant in the female cohort, when comparing the first 5 days after treatment with the first 5 days before treatment (**SI Fig. S11**), and is not observed for the male BSA cohort M4 (**SI Fig. S10**). We can also ask how TIMP-1 induced modification of PL plasticity affects individual mice, by comparing the pairwise specific interactions J_{ij} before and after drug treatment. However, we cannot conclude much as Pearson's correlation coefficient between J_{ij} shows almost no significant correlation across the four time periods in the above datasets for both the TIMP-1 treated cohort and the BSA-treated control cohort (**SI Fig. S12**).

To quantify the sociability of the entire cohort, we compute conditional likelihoods as introduced in the previous paragraphs, as it measures how much the pairwise model explains the observed data compared to a model where mice behave independently. **Fig. 3D** shows that for cohort M1, the model's likelihoods sharply increase following treatment, meaning that the behavior is more predictable. Represented by the independent model, the individual compartment preferences explain most of this increase, suggesting that TIMP-1 treatment reorganizes preferences for specific subterritories. These differences decay back to pre-treatment levels after 5 days, following the time course of drug release. A slightly smaller increase in model's likelihood is observed in the control cohort M4 after BSA injection (**SI Fig. S10D**), suggesting that at least part of the change in compartment preferences can be due to the injection procedure rather than change in the neuronal plasticity itself. In contrast, the increasing model likelihood is not found in the female cohort F1, where the conditional likelihood remains constant after TIMP-1 treatment. However, the contribution of the pairwise interaction is increased (**SI Fig. S11E**), which points to a sex specificity of observed effects.

This observation is further confirmed by the sociability measure of mutual information between single mouse location and the positions of the rest of the cohort, which was introduced in previous paragraphs. The mutual information either does not change (for the male cohort M1, **Fig. 3E**), or increases (for the female cohort F1, **SI Fig. S11**) after the injection of TIMP-1.

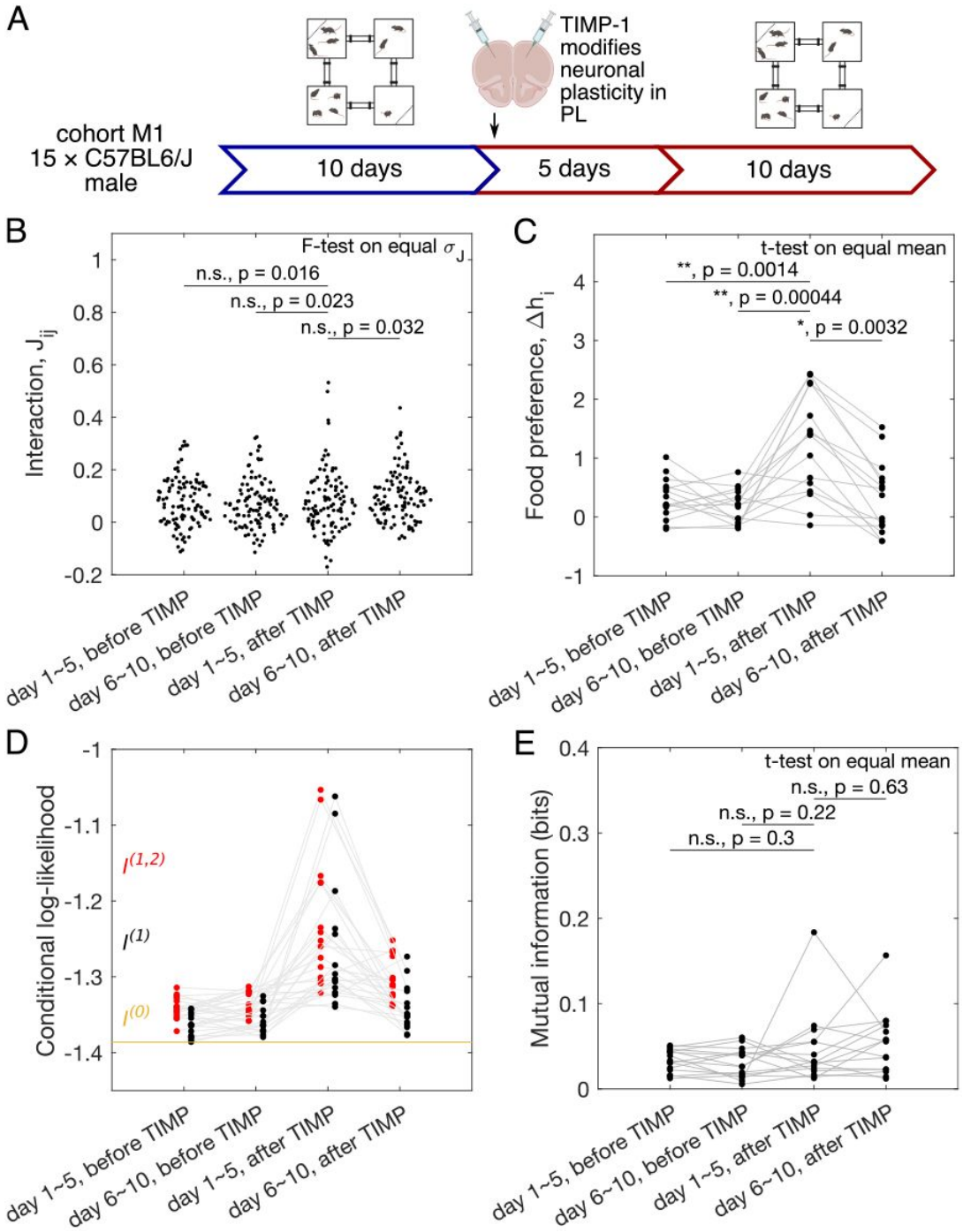


Fig. 3

Quantification of sociability, and the impact of the impaired neuronal plasticity in the prelimbic cortex (PL).

(A) The schematic of the experiment, in which neuronal plasticity in the PL of the tested subjects was impaired with TIMP-1 treatment. A cohort of C57BL6/J male mice (N = 15) was tested in Eco-HAB for 10 days, and then removed from the cages for neuronal plasticity manipulation procedures. After a recovery period, they were placed back in Eco-HAB for another 10 days. For each of the five-day aggregate of the experiment, both before and after TIMP-1 treatment, we plot (B) the model-inferred interactions J_{ij}, (C) preference for food compartments Δh_i, (D) conditional log-likelihood for the pairwise model, $I^{(1,2}$, the independent model, $I^{(1)}$, and the baseline null model, $I^{(0)}$, (E) mutual information between single mouse position and the rest of the network given by the inferred pairwise model.

Impaired neuronal plasticity in the PL affects the structure of social interactions

The increasing variability of pairwise interactions and the non-decreasing mutual information between single mouse location and the location of the rest of the group upon TIMP-1 requires further investigation in the face of previous results showing that injecting TIMP-1 to the PL of wide type animals reduces their sociability. Thus, we examined the detailed group structure of pairwise interactions. We define the dissatisfaction triplet index (DTI) among a triplet of mice as $F_{ijk} \equiv -J_{ij}J_{jk}J_{ki}$ if and only if among the three pairwise interactions among mice i, j and k, exactly one of them is negative (see **Fig. 4A** for schematics), and otherwise zero. Notice that DTI is analogous to the concept of "frustration" in physics of disordered systems. A positive DTI means a triplet of pairwise interactions where all of them cannot be satisfied simultaneously – for instance, if mouse i likes to be with j and k, but j and k do not like to be together. We define the global DTI by averaging the local DTI's across all possible triplets of mice. The larger the global DTI is for a cohort, the more difficult it is for the cohort to form cliques with multiple mice where the interactions among each possible pairs are positive, which may suggest possible difficulty in transmitting information between mice.

As shown in **Fig. 4B**, PL-targeted plasticity disruption with TIMP-1 significantly increases the global DTI for the male mice cohort M1 and the female cohort F1 (two-sample Welch's t-test, variability from random halves of the data; see **Materials and Methods** for details of the significant test). In contrast, in the control cohort M4, injecting male mice with BSA either does not significantly change the global DTI or decreases it. Notably, the difference of the DTI is not due to the control group M4 has less mice, as subsampling both on the level of the inferred interactions (**SI Fig. S13**) and on the level of the mice locations (**SI Fig. S14**) give the same DTI for cohorts M1 and F1. This increase of the global DTI is due to the increasing variance of the interaction J_{ij}, which is related to more of the negative interactions. Randomly shuffling J_{ij} does not change the global DTI, indicating that no network structure was found that contributes to this global DTI (**SI Fig. S15**).

III. Discussion

We demonstrate how the joint probability distribution of the mice positions in the Eco-HAB can be used to quantify sociability. By building a pairwise interaction model whose parameters are learned directly from the data, we quantify how much the combined activities of all mice in the cohort are influenced by their individual preferences and how much by the social context. This approach shows that pairwise interactions between mice are sufficient to describe the statistics of collective behavior in larger groups. Additionally, the pairwise interaction model can capture changes in the social interactions of the network induced by alterations in the neuronal plasticity of the prelimbic cortex (PL) in the tested subjects. The Eco-HAB, combined with this analytical approach, provides a toolbox to quantify sociability in mice, which can be applied generally to different mouse strains to study various behavioral phenotypes, including characteristics associated with neurodevelopmental disorders such as autism. Compared to traditional experimental methods like the three-chamber test, our study combines the advantages of an ecologically relevant and automatic experimental apparatus with the powerful tools of statistical inference. This disentangles the effects of individual preferences versus pairwise social interactions in generating the patterns of mouse positions within their territory.

The challenge in studying social behavior lies in finding a balance between being specific enough to capture the properties of sociability while avoiding the loss of generalizability. Including excessive details, such as the classification of precise social behavior among mice, may lead to a more accurate description of the specific mice cohort studied, such as the construction of a precise

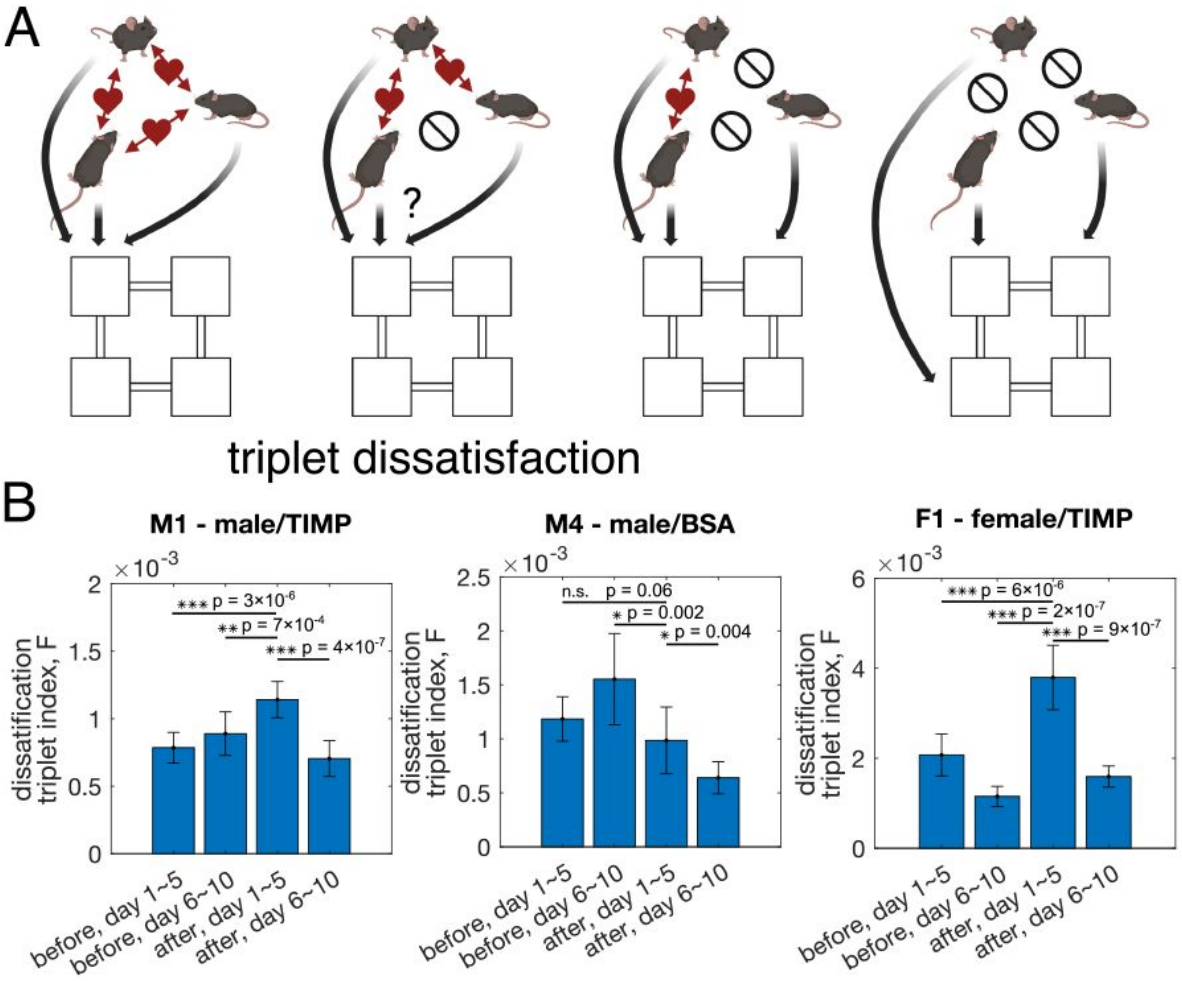


Fig. 4

Effect of TIMP-1 on the structure of the interaction network.

(A) Schematics of how triplets of mice may enter a state of "dissatisfaction" due to competitive pairwise interactions. Dissatisfaction reduces the space of preferable states due to competitive interactions. (B) The global dissatisfaction triplet index (DTI), *F*, computed using inferred interaction from 5-day segments of the data shows that for both male and female mice treated with TIMP-1, the global DTI is significantly increased after drug treatment. Two-sided Welch's *t* -test is performed to test the significance for the difference of the global DTI between the first 5 days after drug injection against the other 5-day segments of the data. Error bars estimate the data variability, which is generated by taking random halves of the data (see **Materials and Methods** for details).

social network. However, it is difficult to assess comparability across different cohorts of mice. Alternatively, as used in this paper, one can construct minimal models and use the ensemble statistics of the models to quantify social properties of a mouse strain without explicitly constructing social networks for each cohort. For example, our study found that the inferred interaction has similar ensemble statistics across four different male cohorts of the same strain but differs across different sexes. This provides evidence to support our argument for a coarse-grained description of mouse social behavior.

Another challenge in studying social behavior lies in the interplay of timescales. We found in this paper that in order to gather enough statistics and to avoid over-fitting, we need to construct probabilistic models using aggregated five-day data, which poses a challenge to balance model construction with enough data versus studying the temporal evolution of sociability. Is the variability of the social network a true property of the social interaction of the mice cohort, or is it due to variabilities of the inferred model caused by the finite data? To address this question, one needs to consider various timescales. For example, mice-mice interactions occur at a much shorter timescale compared to the timescale of changes in the social network, while in between, there are the timescales of adaptation to the new environment and the circadian cycle. These issues need to be addressed using a combination of theoretical tools and experimental validation methods in future works.

Additionally, we have simplified our analysis by focusing on a 6-hour time window each day, during which the rate of locomotor activity is most stable. This approach allowed us to circumvent issues related to individual or strain differences in the circadian cycle, such as the observed “lunch hour” in C57 male mice. One avenue for future research involves reintroducing the circadian cycle as a latent variable to better explain the system. However, caution must be exercised to differentiate between group behavior influenced by the circadian cycle of individual mice and emergent behavior resulting from interactions.

We will now discuss the relationship between our study and that of Shemesh et al., wherein the authors applied a similar approach, investigating the social behavior of groups of 4 mice in a complex experimental environment using statistical modeling of the joint probability distribution of mice locations [39]. They found that triplet interactions are necessary to describe collective behavior, while we found that triplet interactions can be predicted by the pairwise model. We suspect the difference in our results could arise from three factors. First, the experimental arena is quite different between the Eco-HAB and what was used in Shemesh et al. – while the Eco-HAB mimics a naturalistic environment with tubes and connected chambers, the experimental arena in Shemesh et al. contains only one chamber, which could cause a difference in how mice interact. Second, even if the interaction patterns do not change, the nature of the data is different: our data is more coarse-grained spatially, as the state of each mouse is determined by the large compartment it is in, whereas in Shemesh et al., the location is more precise (e.g., a door, a pillar, etc.). As suggested by a comparison to renormalization theory in physics, at coarser spatial scales, the importance of higher-order interactions is likely to decrease. Finally, our studies include more mice (10 to 15) compared to Shemesh et al. (4 mice), which may also influence the importance of higher-order interactions. Larger group size also means that including triplet interactions in our model causes overfitting, which restricts the models we have access to to pairwise interactions. To further investigate these effects, future experiments in Eco-HAB could include mice cohorts of smaller sizes.

How do we move forward, and what is the ideal experiment to study social behaviors? We believe that Eco-HAB offers a balance between a semi-natural environment and controllability, which works well in studying social behavior. One direction for future experimental studies is to focus on the biological function of social interactions. For example, how do mouse cohorts respond to novel odors and transmit information among the cohort? What is the speed of information transmission related to sociability? The current configuration of the Eco-HAB already allows for

the introduction of novel odors accessible to all mice, while the next generation of experiments will localize the introduction of information to individuals. From the analysis perspective, as presented in this manuscript, our current model is purely static. Our model describes the joint probability distribution of mice positions within the territory at concurrent time points and does not model the dynamics of the cohort. To take into account the dynamic aspect of social behaviors, such as dominant mice actively chasing others, one will need to build dynamical models of interaction. For example, this can be done by modeling the probability of transitioning to another compartment of each mouse as a function of the history of its previous location and the locations of all other mice [46].

IV. Materials and methods

Animals

Animals were treated in accordance with the ethical standards of the European Union (directive no. 2010/63/UE) and Polish regulations. All the experiments were pre-approved by the Local Ethics Committee no. 1 in Warsaw, Poland. C57BL/6 male mice were bred in the Animal House of Nencki Institute of Experimental Biology, Polish Academy of Sciences or Mossakowski Medical Research Centre, Polish Academy of Sciences.

The animals entered the experiments when 2-3 month old. They were littermates derived from several breeding pairs. The mice were transferred to the animal room at least 2 weeks before the experiments started and put in the groups of 12-15 in one cage (56 cm× 34 cm × 20cm) with enriched environment (tubes, shelters, nesting materials). They were kept under 12h/12h light-dark cycle. The cages were cleaned once per week. Five cohorts of mice were used in the analysis of this manuscript (see details in **Table 1**).

Exclude inactive and dead mice from analysis

Mouse whose trajectory does not cover all four compartments within the 6-hour period for at least one day of the experiment is defined as inactive, and excluded from the analysis. Including inactive mice in the maximum entropy model will results in unstable learned parameters, as shown by bootstrapped results. For the same mouse cohort before and after injection of drug (M1, M4, and F1), if a mouse is dead or inactive in either phase of the experiment, its trajectory is masked out from the data for consistency of comparison before and after. Specifically, for cohort F1, mouse number 13 (in the original ordering of the 14 mice) is inactive after the drug application. For cohort M4, mouse number 3 and 11 (in the original ordering of the 12 mice) died after surgery, mouse number 9 (in the original ordering) is inactive in the 10th day after drug injection. The total number of mice used in the analysis is given in **Table 1**.

Longitudinal observation of social structure in the Eco-HAB

Cohorts of mice with the same gender and same strain were placed in the Eco-HAB systems and observed for 10 days, removed from the system to undergo stereotaxic injections with TIMP-1 loaded nanoparticles. After 4 to 6 days of recovery, the mice were placed back to a cleaned Eco-HAB, and observed for 10 days.

Activity level

The activity level for a given mouse i during a given time period (t_i, t_f) on day d, is computed by counting the number of times the mouse passes by any antenna, and denoted by $\phi_i^{(d)}(t_i, t_f)$.

Averaging this quantity over all N mice, one obtain the mean activity level for all mice during a given time period. Mathematically, $\overline{\phi}^{(d)}(t_i, t_f) \equiv \sum_i \phi_i^{(d)}(t_i, t_f)/N$.

The standard deviation across all days is the day-to-day variability of mean activity level.

Averaging this quantity over all T days, one obtain the mean activity level for each mouse. Mathematically, $\widetilde{\phi}_i(t_i,t_f) = \sum_d \phi_i^{(d)}(t_i,t_f)/T$. The standard deviation across all mice is the mouse-to-mouse variability of the mean activity level.

Mice location

The raw data consists of time points when mice cross an antenna, as well as the identity of the specific antenna, which are placed at the ends of the four tunnels. The location of a mouse at any given time point is deduced from the most recent time stamps before and after the current time point. For simplicity, for the time points when a mouse is in the tunnel, the location of the mouse is set to be the compartment it will enter. The time resolution is set to 2 seconds, as two adjacent time stamps with separation less than 2 seconds are likely an artifact of mice sniffing the tunnel and returning to the previous compartment.

Pseudocounts

Observing the mice for a finite amount of time means sometimes we have the situation where the mice is stuck in the same compartment for the entire 6 hours of observation. This is not common, but this messes up our statistical inference or model building procedure. To avoid this situation, we use pseudocounts that smoothen the observed statistics. We define

$$m_{ir} = \frac{1}{T+\lambda}\left(\frac{\lambda}{q} + \sum_{t=1}^{T} \delta_{\sigma_i(t)r}\right) \tag{3}$$

and

$$\Gamma_{ij} = \frac{1}{T+\lambda}\left(\frac{\lambda}{q} + \sum_{t=1}^{T} \delta_{\sigma_i(t)\sigma_j(t)}\right), \tag{4}$$

where $q = 4$ is the number of possible states, T is the total number of time points in the data, and λ is the parameter for the pseudocount. In our analysis, after scanning through a range of values for λ, we set $\lambda = 8$ around which value the outcome remained largely unchanged.

The probability model

Gauge fixing for the local field

the probability model is equivalent for the local fields upon a constant, i.e. $P(h)$ and $P(h_{ir} + \delta h_i)$ are equivalent. We overcome this redundancy by enforcing the sum of all local fields for each mouse to be zero, i.e. $\sum r\, h_{ir} = 0$.

Learning the probability model

We train the model using gradient descent, at each learning step k updating the parameters by $J_{ij}^{(k+1)} = J_{ij}^{(k)} - \alpha(c_{ij}^{\text{model}} - c_{ij}^{\text{data}})$ and $h_{ir}^{(k+1)} = h_{ir}^{(k)} - \alpha(m_{ir}^{\text{model}} - m_{ir}^{\text{data}})$, where $\alpha = 0.25 \sim 0.8$ is the step size of learning. The stopping condition is set such that when the difference between the model predicted correlation and magnetization is less than the data variability, estimated by extrapolation from random halves of the data. In addition, because we are interested in quantifying social properties of mice cohort using the statistics of learned parameters, we add to the stopping condition that the mean and the variation of inferred interaction reach a stable value, with change less than 0.005 over 100 learning steps.

A. Computing higher-order correlations

The connected three-point correlation function gives the frequency of finding three mice in the same compartment, subtracting the contributions from the mean and the pair-wise correlation. Mathematically,

$$C_{ijk} = \sum_{r=1}^{p} \left\langle (\delta_{\sigma_i r} - \langle \delta_{\sigma_i r} \rangle)(\delta_{\sigma_j r} - \langle \delta_{\sigma_j r} \rangle)(\delta_{\sigma_k r} - \langle \delta_{\sigma_k r} \rangle) \right\rangle \tag{5}$$

If we only subtract the individual preference, then we define

$$C^*_{ijk} = \sum_{r=1}^{p} \langle \delta_{\sigma_i r} \delta_{\sigma_j r} \delta_{\sigma_k r} \rangle - \langle \delta_{\sigma_i r} \rangle \langle \delta_{\sigma_j r} \rangle \langle \delta_{\sigma_k r} \rangle \tag{6}$$

Comparing in-state probability between model prediction and data

Given time-series data and the inferred joint probability distribution of mice location, we can compare the in-state probability of single mouse, as given by model prediction versus data observation.

More precisely, for each time point t, given the inferred compartment preference h_{ir} for mouse i, the inferred pairwise interaction J_{ij}, and the position of all other mice $j \neq i$, we use the pairwise maximum entropy model (**eq. 2**) to compute the marginal probability of mouse i being in each of the four compartments as the *model-predicted* "in-state probability", $P_{ir}^{\text{in-state}}(t)$. These model-predicted in-state probabilities are then binned according to their percentiles across all observed time points for each mouse and each box, such that the number of time points in each bin is equal. Then, for all time points that belong to each bin, the frequency count of whether mouse i is actually observed in compartment r is computed as the *observed* in-state probability. Agreement between the model-predicted and the observed in state probability across all the bins shows the model is an unbiased estimator, which is what we see after averaging - the in-state probability across the four compartments for fixed percentiles in **SI Fig. S5**.

B. Compute mutual information between single mouse position and the rest of the network

The mutual information between single mouse position and the rest of the network is a measure of collectiveness. For Eco-HAB with 4 compartments, the mutual information is between 0 and 2 bits. If the mutual information is close to 2 bits, knowing where other mice are is a perfect predictor for the position of single mouse. If the mutual information is close to 0 bits, knowing where other mice are do not help predicting the position of the singled-out mouse. The mutual information can be computed as the difference of the entropy of mouse i and the conditional entropy of mouse i with respect to the state of all other mice. Mathematically,

$$I(\sigma_i; \{\sigma_j\}_{j\neq i}) = H(\sigma_i) - \langle H(\sigma_i | \{\sigma_j\}_{j\neq i}) \rangle_{\{\sigma_j\}_{j\neq i}}, \tag{7}$$

where the entropy of mouse i is

$$H(\sigma_i) = -\sum_{r=1}^{4} P(\sigma_i = r) \log P(\sigma_i = r) \tag{8}$$

and the conditional entropy is computed using the conditional probability given $\{\sigma_j\}$ and the inferred pair-wise data, and averaged over all observed data patterns $\{\sigma_j(t)\}$.

$$\begin{aligned} & H\left(\sigma_i|\{\sigma_j\}_{j\neq i}\right) \\ &= -\sum_{\{\sigma_i\}} P\left(\{\sigma_i\}\right) \log P\left(\sigma_i|\{\sigma_j\}_{j\neq i}\right) \\ &= \sum_{\{\sigma_j\}_{j\neq i}} P(\{\sigma_j\}_{j\neq i}) \left[-\sum_{\sigma_i} P(\sigma_i|\{\sigma_j\}_{j\neq i}) \log P(\sigma_i|\{\sigma_j\}_{j\neq i})\right] \\ &\approx \frac{1}{T}\sum_{t=1}^{T}\left[-\sum_{\sigma_i} P^{(2)}(\sigma_i|\{\sigma_j(t)\}_{j\neq i}) \log P^{(2)}(\sigma_i|\{\sigma_j(t)\}_{j\neq i})\right] \end{aligned} \tag{9}$$

To reach the final results, we approximate the ensemble average over all possible mice configurations with a temporal average over all observed mice configuration in the data. We also replace the true underlying conditional probability of $P\left(\sigma_i|\{\sigma_j\}_{j\neq i}\right)$ with the inferred pairwise probability model $P^{(2)}$.

Generate errorbars using random bootstrapped halves of the data

Error bars of the observed statistics $\mathscr{O}$(e.g. pairwise correlation, C_{ij}, and probability in each compartment, m_{ir}), the inferred parameters $\mathscr{P}$(e.g. pairwise interaction J_{ij} and compartment preference h_{ir}), and the sub-sequent results $\mathscr{R}$(e.g. the entropy, $S^{(1,2)}$ and $S^{(1)}$, and the dissatisfaction triplet index F) are bootstrap errors generated by repeatedly taking random halves of the data and computing the deviations in the mean. Specifically, each data set (at least 6 hours in duration) is first divided into time bins of 400 seconds. The length of the time bin is chosen such that it is longer than twice the correlation time for each mouse. Then, random halves of the time bins are chosen to compute the observables, as well as used to train a specific pairwise maximum entropy model, which generates a specific set of learned parameters. The deviation across the random halves, σ_{bs}, can be extrapolated to the full dataset by $\sigma = \sigma_{\mathrm{bs}}/\sqrt{2}$.

Test of significance for comparing observables and inferred parameters

To perform significance test across different days of the experiment, we used the Welch's t -test for the mean of the inferred interaction $\langle J_{ij} \rangle$, the mean of the food preference $\langle \Delta h_i \rangle$, the mean of the mutual information between single mouse position and the rest of the network given by the inferred pairwise model, and for the global dissatisfaction triplet index (DTI), F. We used two-sample F -test for the variability of the inferred interaction. Because we are comparing between pairs of the 5-day aggregate data – the first 5 days before drug injection, the last 5 days before drug injection, the first 5 days after drug injection, and the last 5 days after drug injection – we conduct significance tests using Bonferroni corrections. The number of tests performed for such pairwise comparison is 6. In **Fig. 3**, the asterisks encode the following p-values: *, $p \leq 0.05/6$; * *, $p \leq 0.01/6$; * * *, $p \leq 0.001/6$. For the significance test comparing the global DTI, random halves of the 5-day aggregated data is chosen 10 times, each used to learn the interaction parameters and compute the global DTI. The variance of the global DTI across the 10 random halves is used as variation due to finite amount of data, and is adjusted by $\sigma_\theta = \sigma_{\mathrm{rh}}/\sqrt{2}$. Two-tailed tests are performed, and the Bonferroni correction is applied as the total number of tests performed for the pairwise comparisons for the 5-day aggregated data before and after pharmacological intervention is 6. In **Fig. 4**, the asterisks encode the following p-values: *, $p \leq 0.025/6$; **, $p \leq 0.005/6$; * * *, $p \leq 0.0005/6$.

Maximum entropy model with triplet interactions

To model the joint probability distribution of mice colocalization pattern, one could in principle increase the number of constrains when constructing the maximum entropy model. One example is the triplet correlation, i.e. the probability of any triplets of mice are being found in the same

box, $\Gamma_{ijk} = \langle \delta_{\sigma_i \sigma_j} \delta_{\sigma_i \sigma_k} \rangle$. The corresponding maximum entropy model has the probability distribution

$$P^{(3)}(\sigma) \propto e^{\sum_{i,r} h_{ir} \delta_{\sigma_i r} + \sum_{i<j} J_{ij} \delta_{\sigma_i \sigma_j} + \sum_{i<j<k} G_{ijk} \delta_{\sigma_i \sigma_j} \delta_{\sigma_i \sigma_k}},$$

where h_{ir} is the individual preference of mouse i to be in compartment r and J_{ij} is the pairwise interaction between mouse i and mouse j as before, and G_{ijk} is the triplet interaction among mouse i, mouse j and mouse k. For a cohort with N mice, the numbers of parameters for triplet interactions is $N(N-1)(N-2)/6$. To avoid overfitting, we applied a L2 regularization on the triplet interaction strength, where we now minimize the objective function

$$L^{(3)} = -\frac{1}{T} \sum_{t=1}^{T} \log P^{(3)}(\{\sigma_t\}) + \frac{1}{2} \beta_G \sum_{ijk} G_{ijk}^2.$$

This is naturally translated to learning steps:
$h_{ir}^{(l+1)} = h_{ir}^{(l)} - \alpha_h (m_{ir}^{\text{model}} - m_{ir}^{\text{data}})$, $J_{ij}^{(l+1)} = J_{ij}^{(l)} - \alpha_J (c_{ij}^{\text{model}} - c_{ij}^{\text{data}})$, and
$G_{ijk}^{(l+1)} = G_{ijk}^{(l)} - \alpha_G \left(\Gamma_{ijk}^{\text{model}} - \Gamma_{ijk}^{\text{data}} + \beta_G G_{ijk} \right)$. The stopping condition is the same as before, focusing on the difference between the model predicted correlation and magnetization being less than the data variability, and that the mean and variation of the inferred pairwise interaction reach a stable value.

To test the validity of a triplet interaction model, we varied the regularization strength β_G. For our largest dataset, the cumulative data consist of all 10 days of the experiment on cohort M1 ($N = 15$), we performed cross-validation by splitting the 6-hour data into 6 different training-test set combo (5 hours of data in training set, and 1 hour in the test set). As shown by **SI Fig. S6**, the test set likelihood is found to be maximized when the regularization parameter β_G is large, which corresponds to close to zero triplet interactions. This indicates that even for our largest datasets, including the triplet interactions in the model have increased the model complexity beyond what the data allows, and the best way to avoid overfitting is for us to not consider triplet interactions. Hence, the interaction models we choose to study in this manuscript is restricted to pairwise interaction models.

Data and code availability

All data used in our manuscript and the MATLAB and python code to analyze the data can be found in *https://github.com/statbiophys/social_mice*.

Acknowledgements

This work was partially supported by the European Research Council Consolidator Grant n. 724208, 'BRAINCITY - Centre of Excellence for Neural Plasticity and Brain Disorders' project of the Polish Foundation for Science, and the National Science Center grant 2020/39/D/NZ4/01785.

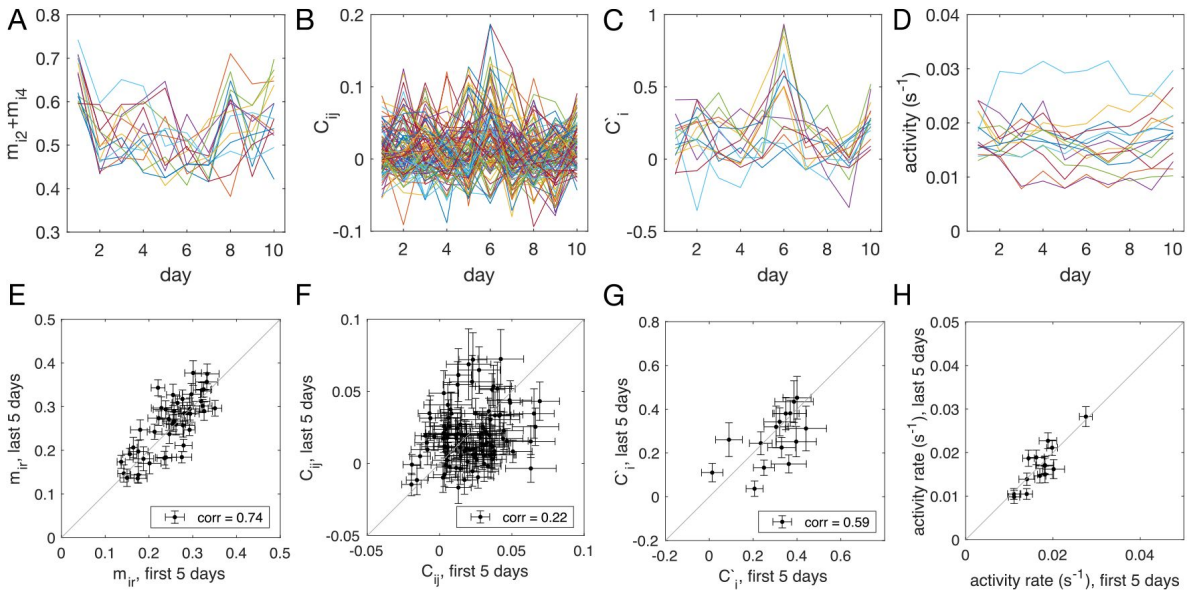


Fig. S1

Stability of the data, given by the time evolution across 10 days of the experiment and the scatter plot between the observables measured using the first 5 days of the data vs. the last 5 days of the data for cohort M1 (*N* = 15).

The observables plotted include (A, E) m_{ir}, probability of mouse *i* being found in compartment *r*, (B, F) C_{ij}, the (connected) pairwise correlation, or the in-cohort sociability, between mouse *i* and mouse *j*, (C, G) the sum of observed in-cohort sociability, $C_i \equiv \sum_{j \neq i} C_{ij}$, which gives a proxy for how mouse *i* is effected by the social interaction, and (D, H) the activity rate, measured by the number of transition event per second. The error bars in panels E-H are extrapolated by bootstrapping random halves of the data.

Fig. S2

The pairwise maximum entropy model is trained such that the model reproduces the probability for each mouse in each compartment, m_{ir}, and the probability for pairs of mice in the same compartment, C_{ij}, as given by the data.

Error bars are generated by bootstrapping random halves of the data.

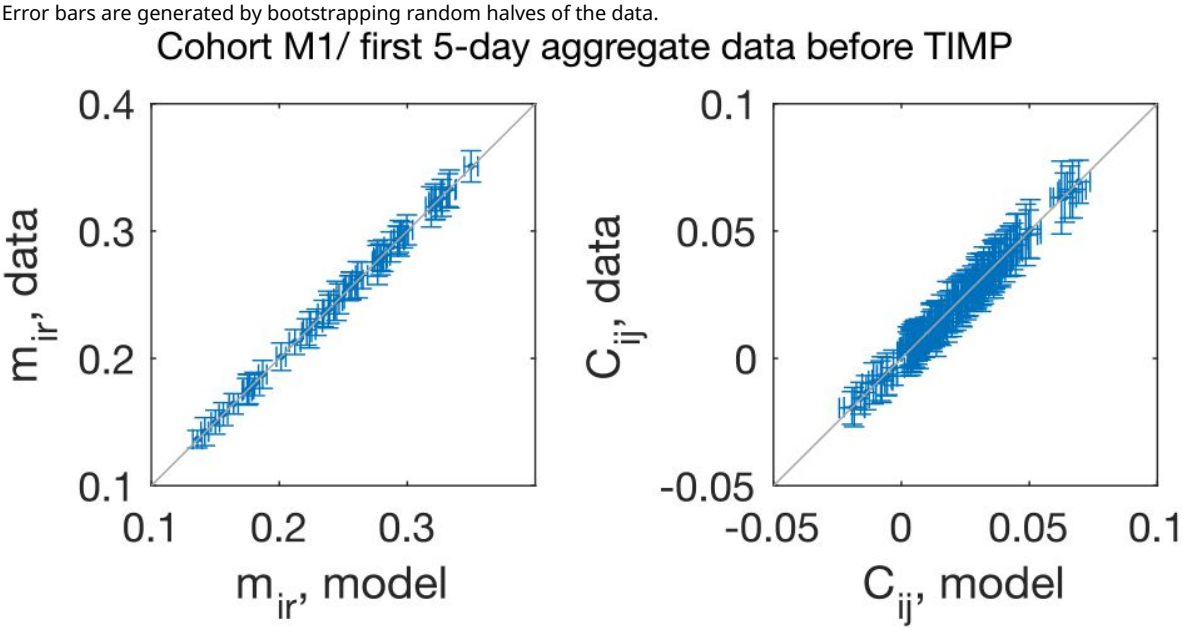
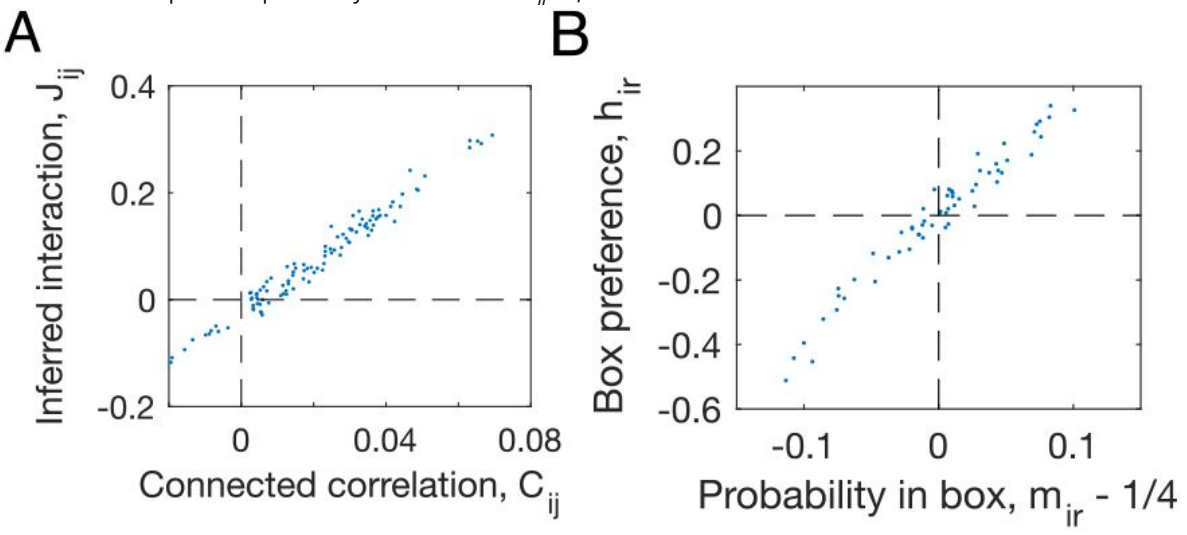


Fig. S3

Learned parameters in the pairwise interaction model versus the observed statistics, plotted for the 5-day aggregate data from the first 5 days of the experiment on male cohort M1 before TIMP-1 treatment.

(A) The inferred interaction J_{ij} versus the connected correlation C_{ij} ; (B) the inferred individual compartment preference h_{ir} versus the in-compartment probability for each mouse m_{ir} - 1/4.



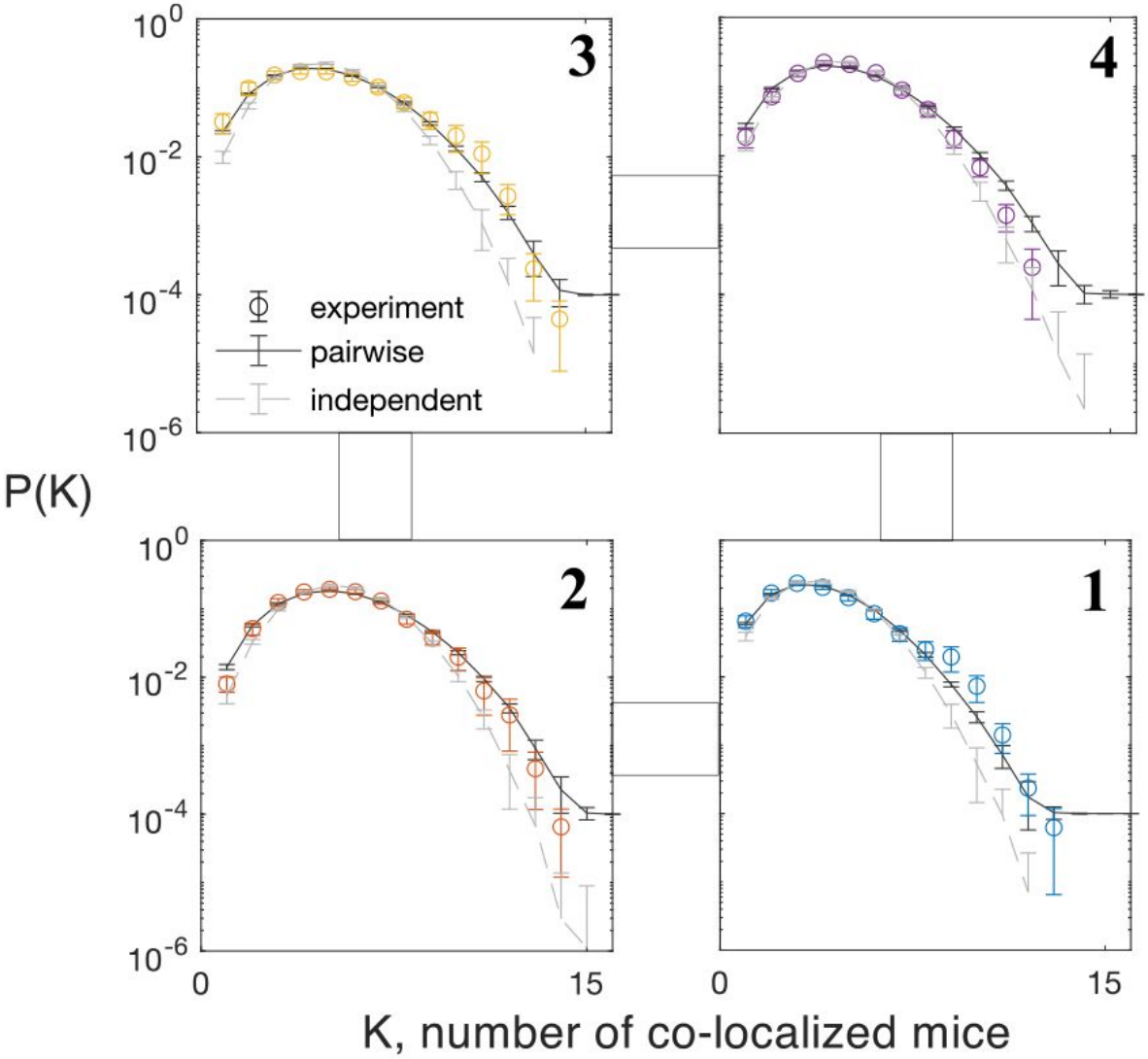


Fig. S4

The probability of *K* mice found in the same compartment, predicted by the pairwise maximum entropy model, the independent model, and computed from the 5-day aggregate data for the first 5 days in male cohort M1 before TIMP-1 treatment (*N* = 15).

The subpanels are arranged in the same order as in the Eco-HAB setup. Error bars for the experiment are extrapolated from 50 random halves of the data, for the independent is generated by 50 random cyclic shuffling of the data, and for the pairwise model is from 50 random MCMC samplings (each with 54000 realizations, the same number of data points as the data) for the pairwise model.

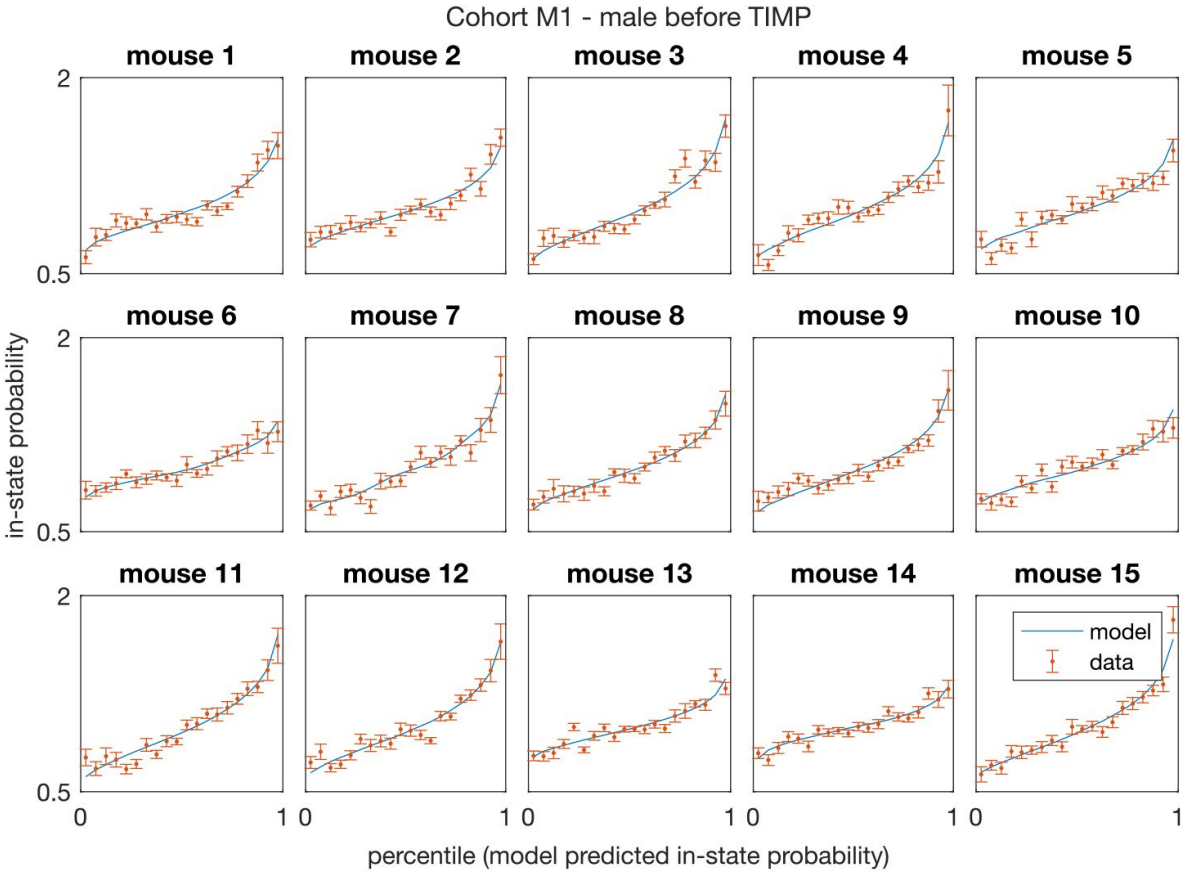


Fig. S5

Model predicted in-state probability matches data observation for the aggregate data of first five days of experiment in mice cohort M1 - C57BL6/J male mice (*N* = 15), which shows the prediction of the inferred pairwise model is unbiased.

Error bars are extrapolated from 20 draws of random halves of the data.

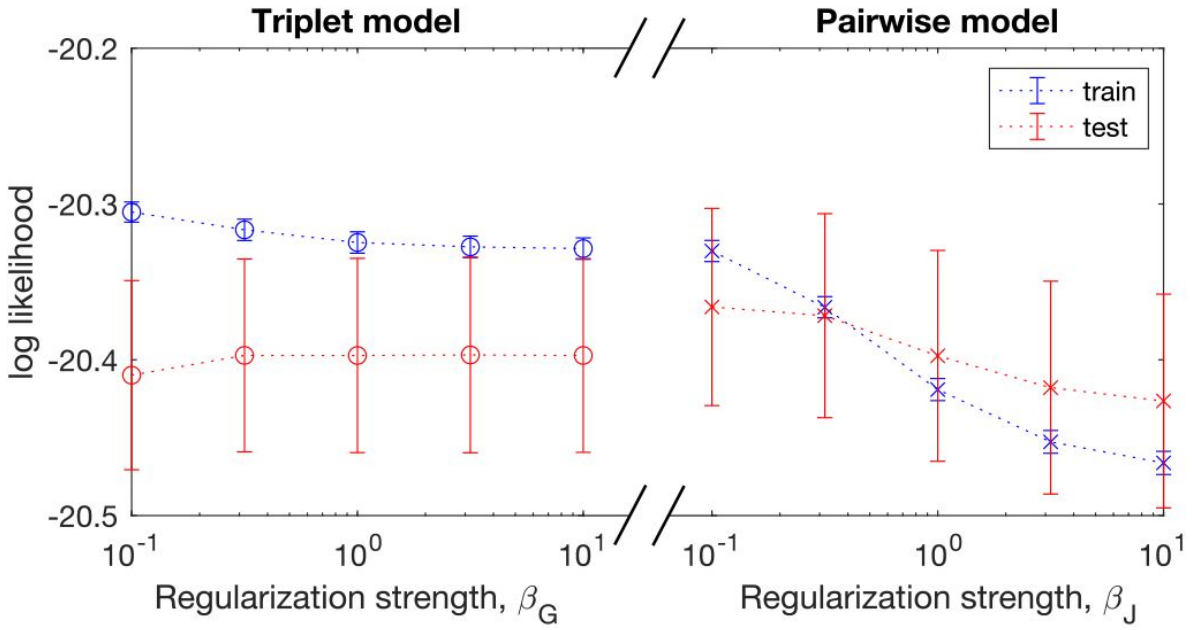


Fig. S6

Cross-validation for maximum entropy models with triplet interactions and models with pairwise interactions on combined 10-day data (cohort M1, *N* = 15).

For the triplet model, L2 regularization β_G is applied only to the triplet interactions. For the pairwise model, L2 regularization β_J is applied only to the pairwise interactions. Error bars are standard error from the mean across 6 different training-test set partitions, each containing 1 hour of data as test set and 5 hours of data as training set. The maximum of test-set log likelihood is achieved when the regularization strength β_G is large in the triplet model, and when the regularization strength β_J is small in the pairwise model, indicating that the triplet model overfits for the data pulled from all 10 days.

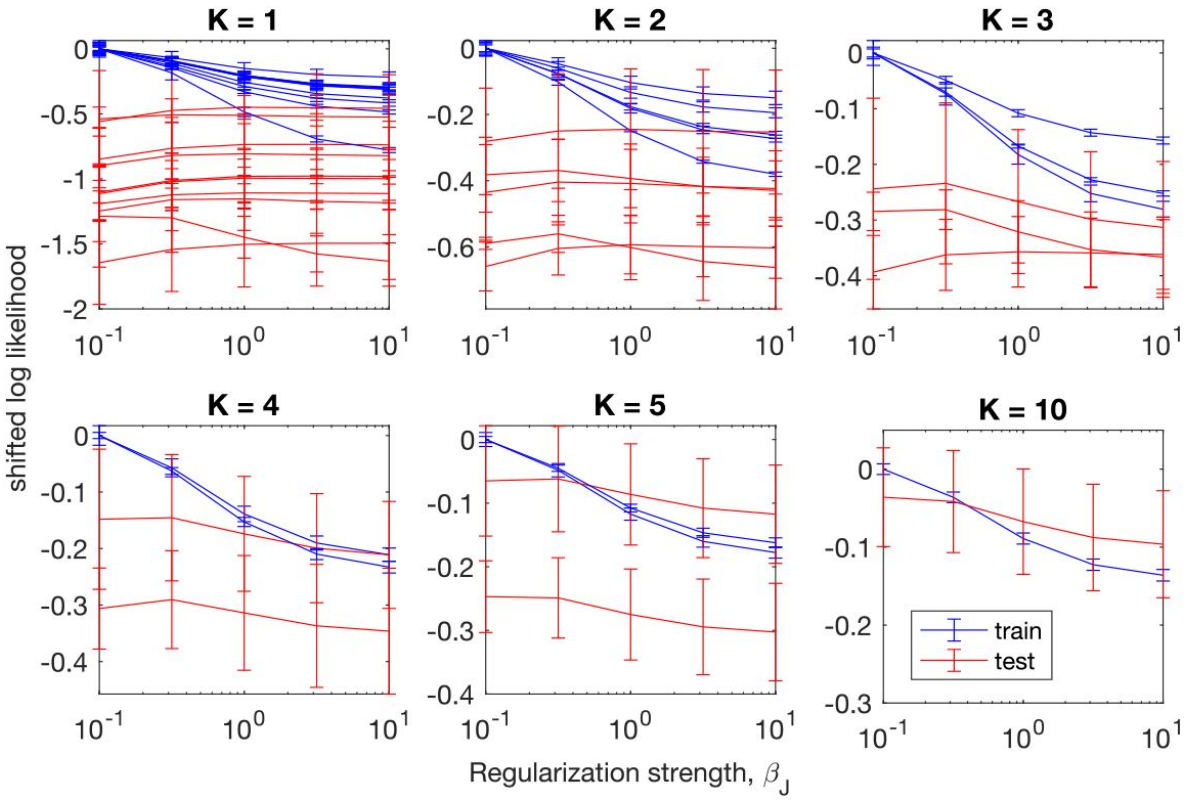


Fig. S7

Cross-validation for maximum entropy models with pairwise interactions on combined data from a total of *K* days (cohort M1, *N* = 15).

The plotted log-likelihoods are shifted by the training-set log likelihood at L2 regularization strength $\beta_J = 10^{-1}$. The L2 regularization strength β_J is applied only to the pairwise interactions. Error bars are standard error from the mean across 6 different training-test set partitions, each containing 1 hour of data as test set and 5 hours of data as training set. For $K = 1$, the test-set likelihood increases as the regularization strength β_J increases, indicating that the pairwise maximum entropy model overfits if we only consider data from each day; while for $K \geq 4$, the test-set likelihood decreases as the regularization parameter increases, indicating no overfitting.

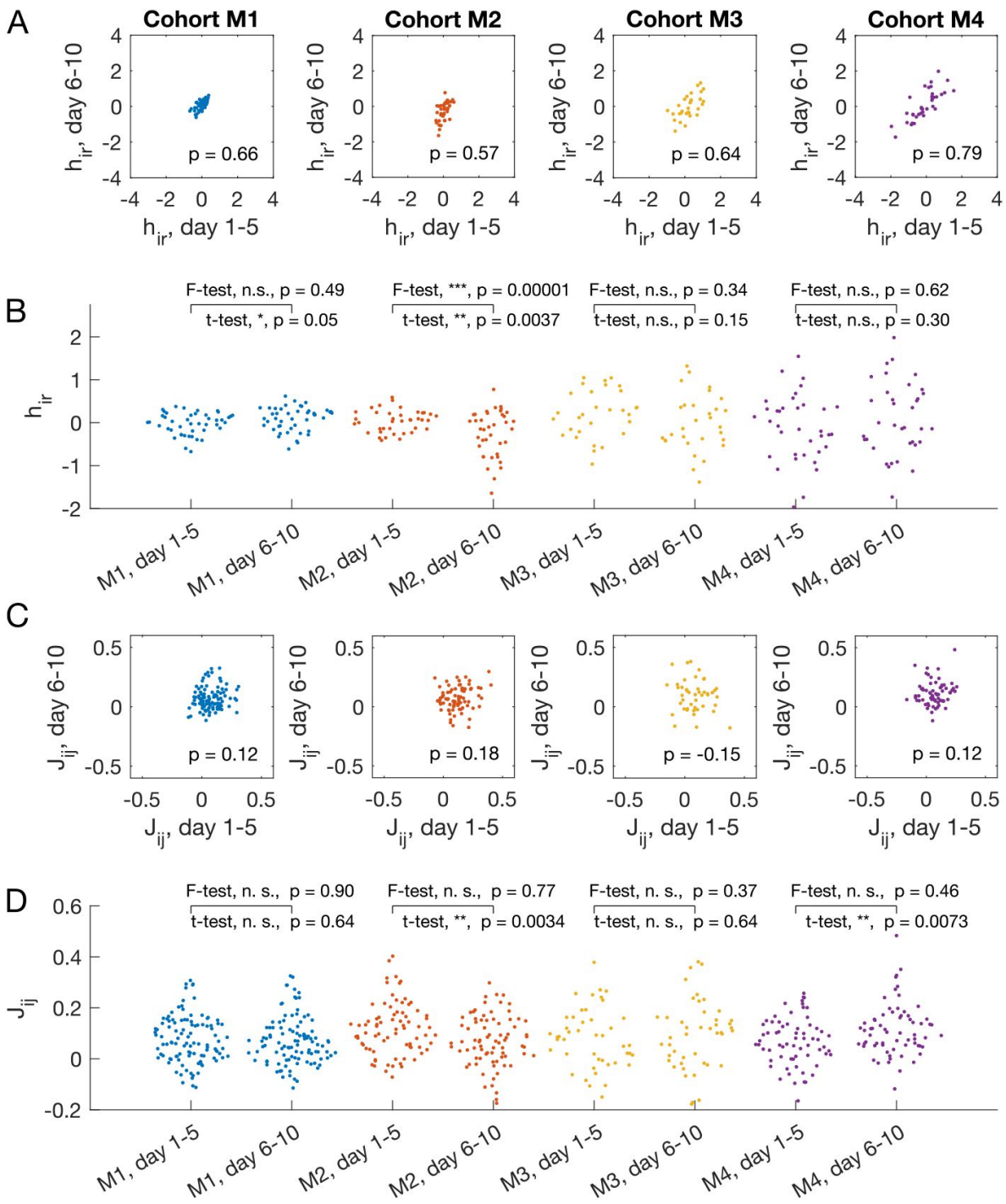


Fig. S8

Temporal consistency of inferred parameters from 5-day accumulated data for mice cohorts M1 (*N* = 15), M2 (*N* = 13), M3 (*N* = 10), M4(*N* = 12)).

(A) Chamber preference for each mouse in each box, h_{ir}, between models learned from the accumulated data from day 1-5, compared to the model learned from accumulated data from day 6-10. The Pearson's correlation coefficient is shown on the plot. (B) Swarm plot for inferred pairwise interactions for each cohort from the first 5 days and the last 5 days. Two-sided *t* -test for equal mean and two-sided *F* -test for equal variance are performed. The asterisks encode the following *p*-values: *, $p \leq 0.05$; **, $p \leq 0.01$; * * *, $p \leq 0.001$. (C) and (D) are the same as (A) and (B) for the pairwise interaction between mouse *i* and mouse *j*, J_{ij}.

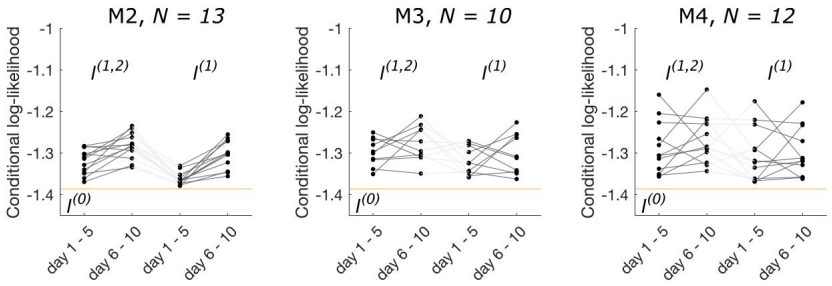


Fig. S9

The conditional log likelihood is different for each cohort of C57BL6/J male mice (N = 13 in cohort M2, N = 10 in cohort M3, and N = 12 in cohort M4 (before BSA injection), exhibiting individuality.

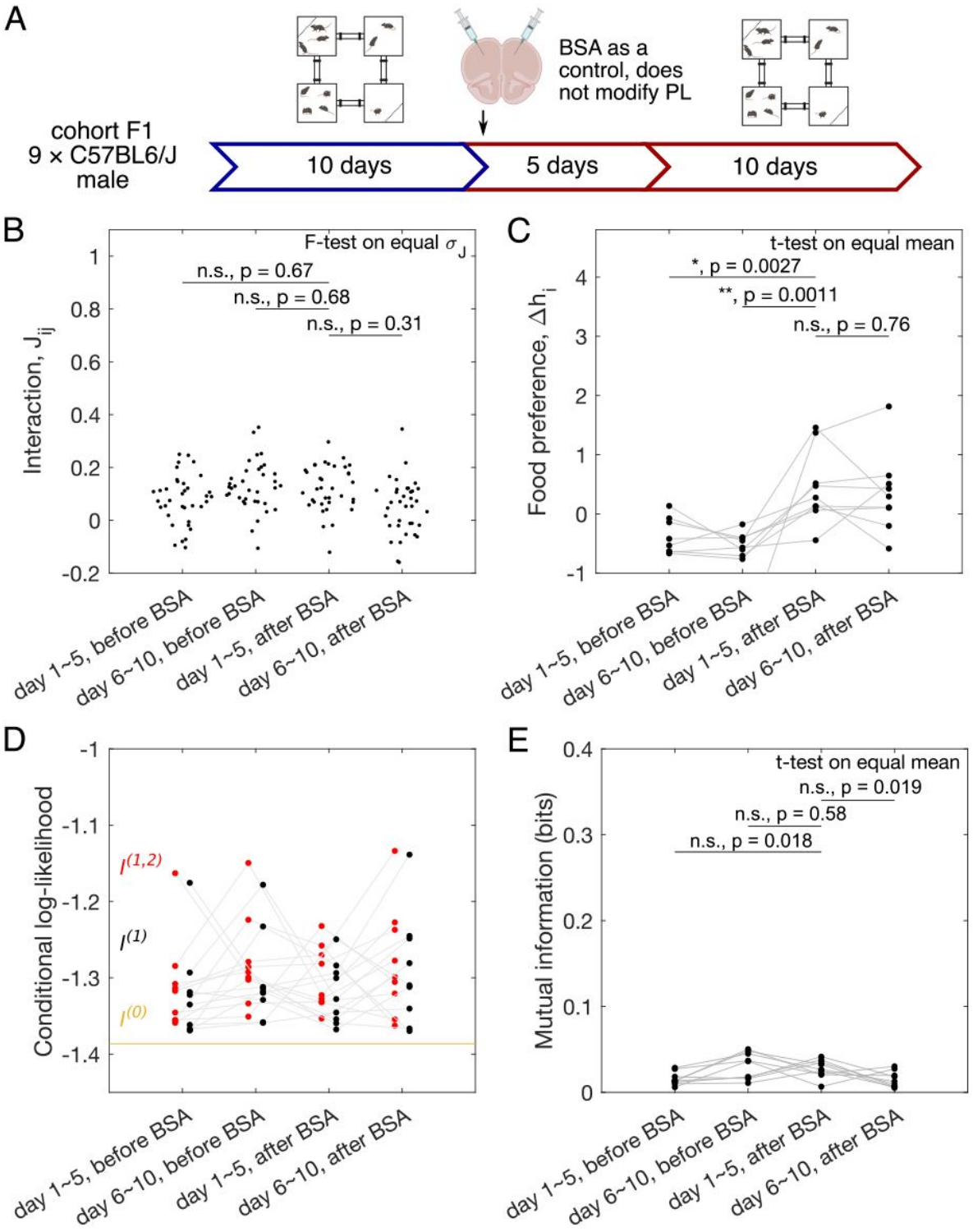


Fig. S10

Quantification of sociability, and the impact of the injection of BSA-infused nanoparticles, a control which does not impair neuronal plasticity in the prelimbic cortex (PL), in male C57BL6/J mice.

This figure follows **Figure 3**, now for cohort M4 (*N* = 9).

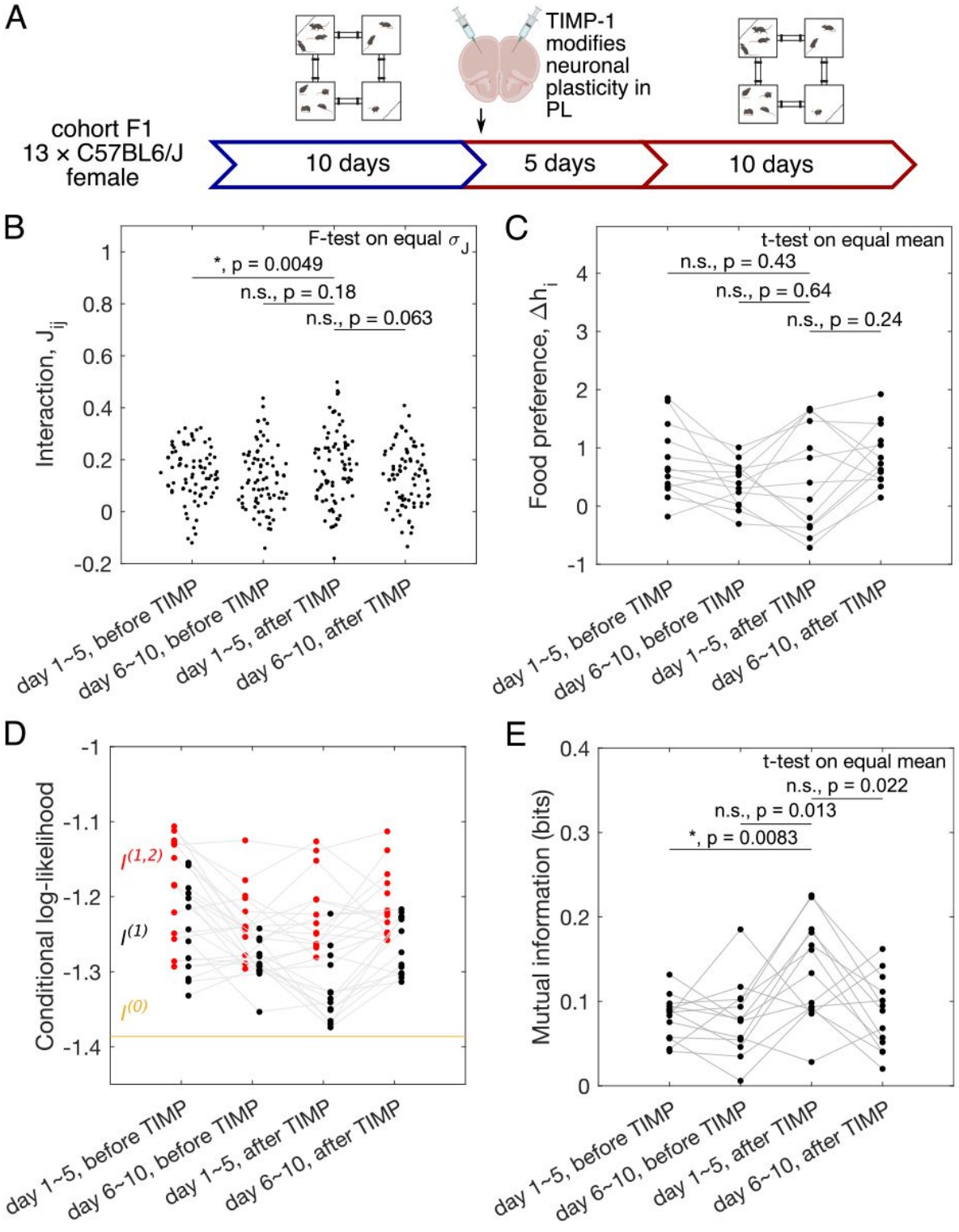


Fig. S11

Quantification of sociability, and the impact of the impaired neuronal plasticity in the prelimbic cortex (PL) in female mice.

This figure follows **Figure 3**, now for cohort F1 (*N* = 13).

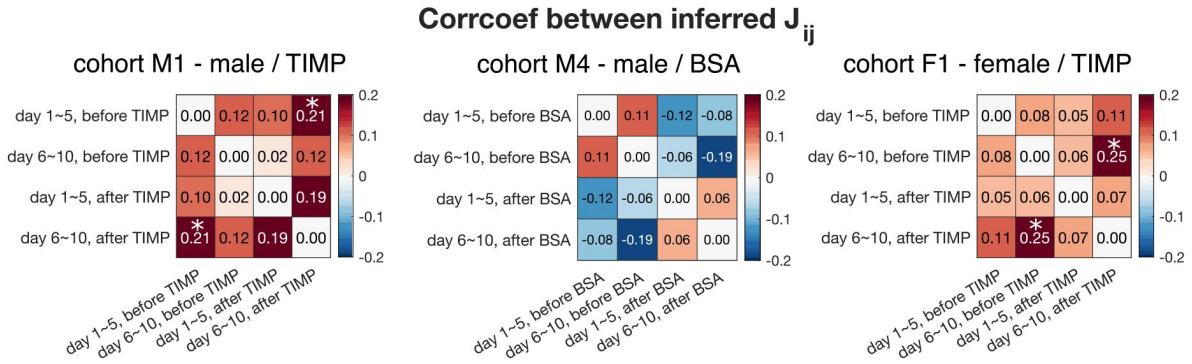


Fig. S12

Pearson's correlation coefficient between inferred interaction J_{ij} from different five-day aggregated data, before and after drug injection, for cohort M1, M4 and F1. Asterisks indicate statistical significance. Almost no correlation is detected between the inferred J_{ij}. The only two comparisons that exhibit statistical significance are between first and last 5 days after TIMP injection for cohort M1 (p-value = 0.033), and between last 5 days before TIMP and last 5 days after TIMP injection for cohort F1 (p-value = 0.028).

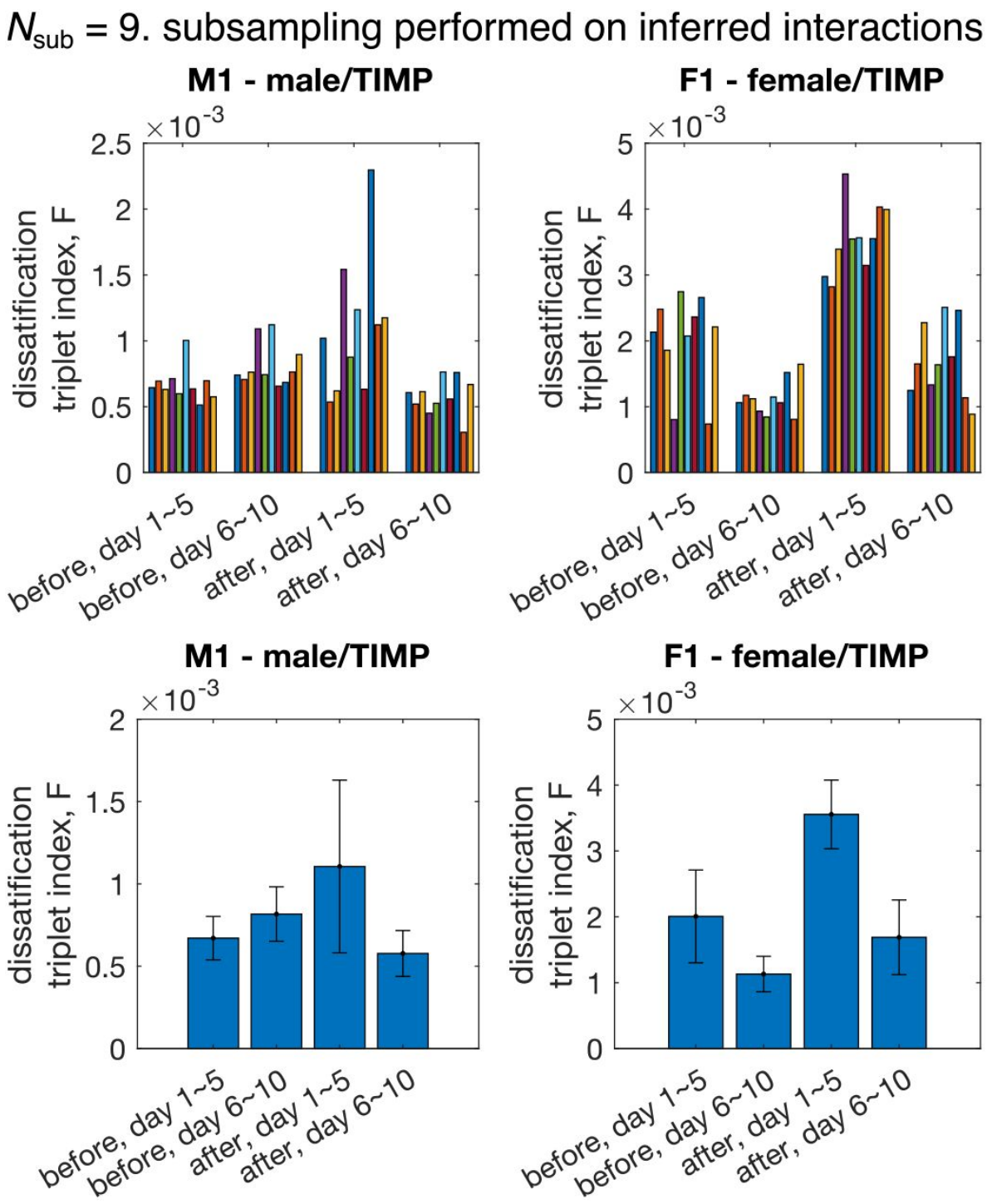


Fig. S13

Dissatisfaction triplet index *F* computed using subgroups of mice for cohorts M1 (N = 15, N_{sub} = 9) and F1 (N = 13, N_{sub} = 9). Subsampling is performed at the level of inferred interactions, i.e. interactions J_{ij} among all N mice are inferred using the pairwise maximum entropy model, then, 10 random subgroups of N_{sub} mice are drawn. *Top:* each colored bar represent 1 realization of the subsampling. Bootstraps of random halves of the data were used in analysis to be consistent **Figure 4**. *Bottom:* average over subgroups of the bar plot in the *top* row. Error bar represents standard deviation across different realizations of subgroups.

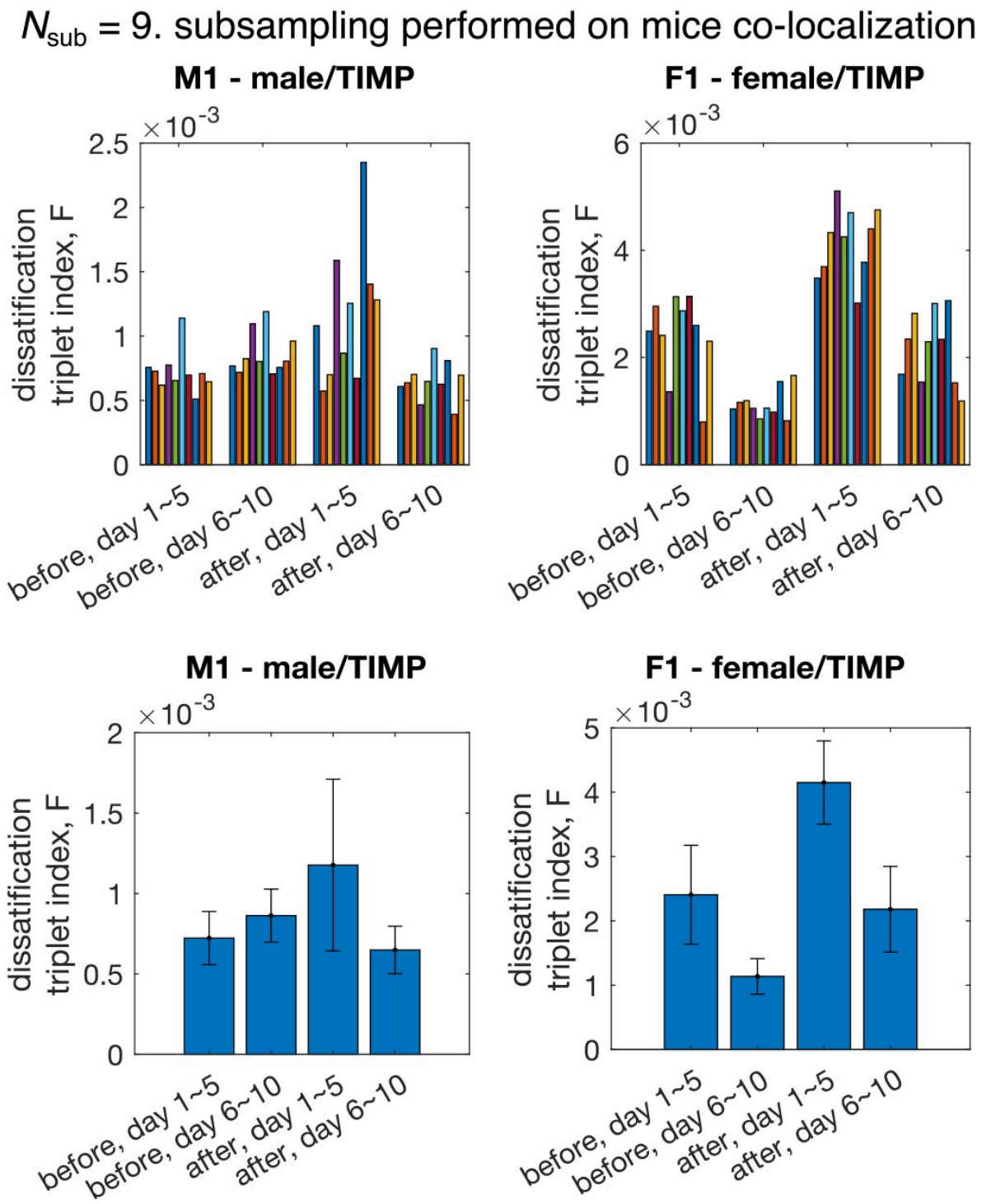


Fig. S14

Dissatisfaction triplet index computed using subgroups of mice from cohorts M1 and F1. Subsampling is performed at the data level, i.e. the co-localization patterns of N_{sub} randomly selected mice are used to infer the interaction strengths, which are then used to compute the dissatisfaction triplet index *F*.

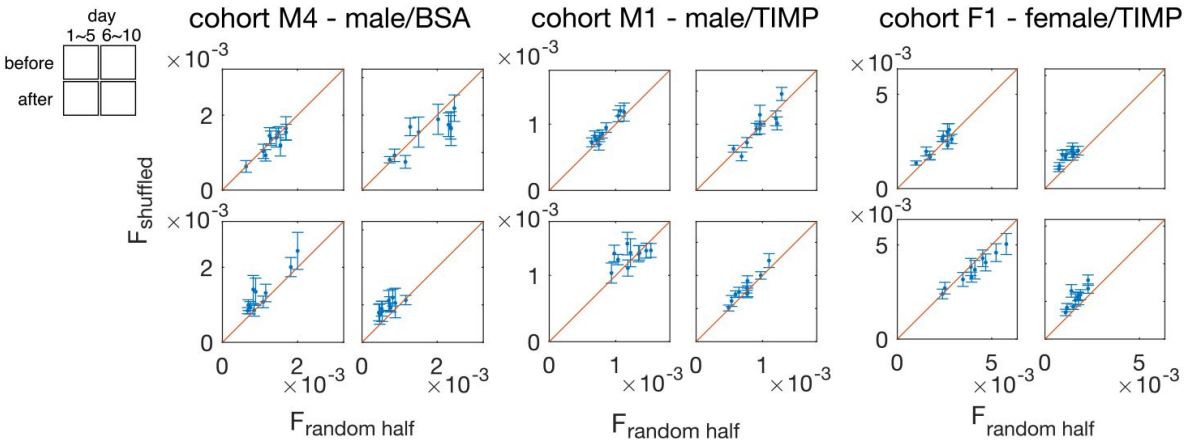


Fig. S15

The global dissatisfaction triplet index (DTI) computed using shuffled interaction, F_{shuffled} vs. the global DTI computed using the inferred interaction, $F_{\text{random half}}$.

Each point corresponds to one random half of the data. The two sets of global DTI's are equal within the range of the error bars, computed by standard deviation across 20 random shuffling of the inferred interaction J_{ij}, which shows the global DTI comes from the value of the inferred interaction, and that there is no additional network structure of the inferred interaction.

References

[1] Allsop S. A., Vander Weele C. M., Wichmann R., Tye K. M. (2014) **Optogenetic insights on the relationship between anxiety-related behaviors and social deficits** *Frontiers in behavioral neuroscience* **8**:241

[2] Hurst J. L., West R. S. (2010) **Taming anxiety in laboratory mice** *Nature methods* **7**:825

[3] Sandi C., Haller J. (2015) **Stress and the social brain: behavioural effects and neurobiological mechanisms** *Nature Reviews Neuroscience* **16**:290

[4] Chesler E. J., Wilson S. G., Lariviere W. R., Rodriguez-Zas S. L., Mogil J. S. (2002) **Identification and ranking of genetic and laboratory environment factors influencing a behavioral trait, thermal nociception, via computational analysis of a large data archive** *Neuroscience & Biobehavioral Reviews* **26**:907

[5] Sorge R. E., Martin L. J., Isbester K. A., Sotocinal S. G., Rosen S., Tuttle A. H., Wieskopf J. S., Acland E. L., Dokova A., Kadoura B., et al. (2014) **Olfactory exposure to males, including men, causes stress and related analgesia in rodents** *Nature methods* **11**:629

[6] Ballerini M., Cabibbo N., Candelier R., Cavagna A., Cisbani E., Giardina I., Orlandi A., Parisi G., Procaccini A., Viale M., et al. (2008) **Empirical investigation of starling flocks: a benchmark study in collective animal behaviour** *Animal behaviour* **76**:201

[7] Attanasi A., Cavagna A., Del Castello L., Giardina I., Melillo S., Parisi L., Pohl O., Rossaro B., Shen E., Silvestri E., et al. (2014) **Collective behaviour without collective order in wild swarms of midges** *PLoS computational biology* **10**:e1003697

[8] Katz Y., Tunstrøm K., Ioannou C. C., Huepe C., Couzin I. D. (2011) **Inferring the structure and dynamics of interactions in schooling fish** *Proceedings of the National Academy of Sciences* **108**:18720

[9] Herbert-Read J. E., Perna A., Mann R. P., Schaerf T. M., Sumpter D. J., Ward A. J. (2011) **Inferring the rules of interaction of shoaling fish** *Proceedings of the National Academy of Sciences* **108**:18726

[10] Gautrais J., Ginelli F., Fournier R., Blanco S., Soria M., Theraulaz G. (2012) **Deciphering interactions in moving animal groups** *PLoS Computational Biology* :e1002678

[11] Buhl J., Sumpter D. J., Couzin I. D., Hale J. J., Despland E., Miller E. R., Simpson S. J. (2006) **From disorder to order in marching locusts** *Science* **312**:1402

[12] Ginelli F., Peruani F., Pillot M.-H., Theraulaz G., Bon R. (2015) **Intermittent collective dynamics emerge from conflicting imperatives in sheep herds** *Proceedings of the National Academy of Sciences* **112**:12729

[13] Kondrakiewicz K., Kostecki M., Szadzińska W., Knapska E. (2019) **Ecological validity of social interaction tests in rats and mice** *Genes, Brain and Behavior* **18**:e12525

[14] Puścian A., Łęski S., Kasprowicz G., Winiarski M., Borowska J., Nikolaev T., Boguszewski P. M., Lipp H.-P., Knapska E. (2016) **Eco-HAB as a fully automated and ecologically relevant assessment of social impairments in mouse models of autism** *Elife* **5**:e19532

[15] Williamson C. M., Franks B., Curley J. P. (2016) **Mouse social network dynamics and community structure are associated with plasticity-related brain gene expression** *Frontiers in Behavioral Neuroscience* **10**:152

[16] Jeon D., Kim S., Chetana M., Jo D., Ruley H. E., Lin S.-Y., Rabah D., Kinet J.-P., Shin H.-S. (2010) **Observational fear learning involves affective pain system and Cav1.2 Ca2+ channels in ACC** *Nature neuroscience* **13**:482

[17] Jr B. G. Galef (2002) **Social learning of food preferences in rodents: rapid appetitive learning** *Current protocols in neuroscience* **21**:8

[18] Gonzalez-Liencres C., Juckel G., Tas C., Friebe A., Brüne M. (2014) **Emotional contagion in mice: the role of familiarity** *Behavioural brain research* **263**:16

[19] Freund J., Brandmaier A. M., Lewejohann L., Kirste I., Kritzler M., Krüger A., Sachser N., Lindenberger U., Kempermann G. (2013) **Emergence of individuality in genetically identical mice** *Science* **340**:756

[20] Bicks L. K., Koike H., Akbarian S., Morishita H. (2015) **Prefrontal cortex and social cognition in mouse and man** *Frontiers in psychology* **6**:1805

[21] Demolliens M., Isbaine F., Takerkart S., Huguet P., Boussaoud D. (2017) **Social and asocial prefrontal cortex neurons: a new look at social facilitation and the social brain** *Social cognitive and affective neuroscience* **12**:1241

[22] Lee E., Rhim I., Lee J. W., Ghim J.-W., Lee S., Kim E., Jung M. W. (2016) **Enhanced neuronal activity in the medial prefrontal cortex during social approach behavior** *Journal of Neuroscience* **36**:6926

[23] Levy D. R., Tamir T., Kaufman M., Parabucki A., Weissbrod A., Schneidman E., Yizhar O. (2019) **Dynamics of social representation in the mouse prefrontal cortex** *Nature Neuroscience* **22**:2013

[24] Denny B. T., Kober H., Wager T. D., Ochsner K. N. (2012) **A meta-analysis of functional neuroimaging studies of self-and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex** *Journal of cognitive Neuroscience* **24**:1742

[25] Yizhar O., Levy D. R. (2021) **The social dilemma: prefrontal control of mammalian sociability** *Current Opinion in Neurobiology* **68**:67

[26] Loureiro M., Achargui R., Flakowski J., Van Zessen R., Stefanelli T., Pascoli V., Lüscher C. (2019) **Social transmission of food safety depends on synaptic plasticity in the prefrontal cortex** *Science* **364**:991

[27] Attwood B. K., Bourgognon J.-M., Patel S., Mucha M., Schiavon E., Skrzypiec A. E., Young K. W., Shiosaka S., Korostynski M., Piechota M., et al. (2011) **Neuropsin cleaves EphB2 in the amygdala to control anxiety** *Nature* **473**:372

[28] Bach D. R., Tzovara A., Vunder J. (2018) **Blocking human fear memory with the matrix metalloproteinase inhibitor doxycycline** *Molecular psychiatry* **23**:1584

[29] Vaillant C., Didier-Bazes M., Hutter A., Belin M.-F., Thomasset N. (1999) **Spatiotemporal Expression Patterns of Metalloproteinases and Their Inhibitors in the Postnatal Developing Rat Cerebellum** *Journal of Neuroscience* **19**:4994

[30] Ould-yahoui A., Tremblay E., Sbai O., Ferhat L., Bernard A., Charrat E., Gueye Y., Lim N. H., Brew K., Risso J.-J., et al. (2009) **A New Role for TIMP-1 in Modulating Neurite Outgrowth and Morphology of Cortical Neurons** *PloS one* **4**:e8289

[31] Dziembowska M., Wlodarczyk J. (2012) **MMP9: a novel function in synaptic plasticity** *The international journal of biochemistry & cell biology* **44**:709

[32] Gorkiewicz T., Balcerzyk M., Kaczmarek L., Knapska E. (2015) **Matrix metalloproteinase 9 (MMP-9) is indispensable for long term potentiation in the central and basal but not in the lateral nucleus of the amygdala** *Frontiers in cellular neuroscience* **9**:73

[33] Okulski P., Jay T. M., Jaworski J., Duniec K., Dzwonek J., Konopacki F. A., Wilczynski G. M., Sánchez-Capelo A., Mallet J., Kaczmarek L. (2007) **TIMP-1 abolishes MMP-9-dependent long-lasting long-term potentiation in the prefrontal cortex** *Biological psychiatry* **62**:359

[34] Wang F., Zhu J., Zhu H., Zhang Q., Lin Z., Hu H. (2011) **Bidirectional control of social hierarchy by synaptic efficacy in medial prefrontal cortex** *Science* **334**:693

[35] Bicks L. K., Peng M., Taub A., Akbarian S., Morishita H. (2021) **Prefrontal cortex and social cognition in mouse and man** *Frontiers in Neural Circuits* **15**:676308

[36] Wang F., Kessels H. W., Hu H. (2014) **The mouse that roared: Neural mechanisms of social hierarchy** *Trends in neurosciences* **37**:674

[37] Chaturvedi M., Molino Y., Sreedhar B., Khrestchatisky M., Kaczmarek L. (2014) **Tissue inhibitor of matrix metalloproteinases-1 loaded poly (lactic-co-glycolic acid) nanoparticles for delivery across the blood–brain barrier** *International journal of nanomedicine* **9**:575

[38] Bialek W., Cavagna A., Giardina I., Mora T., Silvestri E., Viale M., Walczak A. M. (2012) **Statistical mechanics for natural flocks of birds** *Proceedings of the National Academy of Sciences* **109**:4786

[39] Shemesh Y., Sztainberg Y., Forkosh O., Shlapobersky T., Chen A., Schneidman E. (2013) **High-order social interactions in groups of mice** *Elife* **2**:e00759

[40] Rydzanicz M., Kuzniewska B., Magnowska M., Wójtowicz T., Stawikowska A., Hojka A., Borsuk E., Meyza K., Gewartowska O., Gruchota J., et al. (2024) **Mutation in the mitochondrial chaperone TRAP1 leads to autism with more severe symptoms in males** *EMBO Molecular Medicine* **16**:2976

[41] Jaynes E. T. (1957) **Information theory and statistical mechanics** *Physical review* **106**:620

[42] Schneidman E., Berry M. J., Segev R., Bialek W. (2006) **Weak pairwise correlations imply strongly correlated network states in a neural population** *Nature* **440**:1007

[43] Knapska E., Lioudyno V., Kiryk A., Mikosz M., Górkiewicz T., Michaluk P., Gawlak M., Chaturvedi M., Mochol G., Balcerzyk M., et al. (2013) **Reward Learning Requires Activity of Matrix Metalloproteinase-9 in the Central Amygdala** *Journal of Neuroscience* **33**:14591

[44] Puścian A., Winiarski M., Borowska J., Łęski S., Górkiewicz T., Chaturvedi M., Nowicka K., Wołyniak M., Chmielewska J., Nikolaev T., et al. (2022) **Targeted therapy of cognitive deficits**

in fragile X syndrome *Molecular psychiatry* **27**:2766

[45] Winiarski M., Borowska J., Wołyniak R., Jędrzejewska-Szmek J., Kondrakiewicz L., Mankiewicz L., Chaturvedi M., Turzyński K., Wójcik D., Puścian A., et al. (2021) **BioRxiv**

[46] Chen X., Winiarski M., Puscian A., Knapska E., Walczak A. M., Mora T. (2023) **Generalized Glauber Dynamics for Inference in Biology** *Phys. Rev. X* **13**:041053

Author information

Xiaowen Chen

Laboratoire de physique de l'École normale supérieure, CNRS, PSL University, Sorbonne Université, and Université Paris Cité, Paris, France

Maciej Winiarski

Center of Excellence for Neural Plasticity and Brain Disorders: BRAINCITY, a Nencki-EMBL Partnership, Nencki Institute of Experimental Biology of Polish Academy of Sciences, Warsaw, Poland

Alicja Puścian

Center of Excellence for Neural Plasticity and Brain Disorders: BRAINCITY, a Nencki-EMBL Partnership, Nencki Institute of Experimental Biology of Polish Academy of Sciences, Warsaw, Poland

Ewelina Knapska

Center of Excellence for Neural Plasticity and Brain Disorders: BRAINCITY, a Nencki-EMBL Partnership, Nencki Institute of Experimental Biology of Polish Academy of Sciences, Warsaw, Poland

Thierry Mora†

Laboratoire de physique de l'École normale supérieure, CNRS, PSL University, Sorbonne Université, and Université Paris Cité, Paris, France
ORCID iD: 0000-0002-5456-9361

For correspondence: tmora@lps.ens.fr

†These authors contributed equally.

Aleksandra M Walczak†

Laboratoire de physique de l'École normale supérieure, CNRS, PSL University, Sorbonne Université, and Université Paris Cité, Paris, France

†These authors contributed equally.

Editors

Reviewing Editor

Yuuki Watanabe

Graduate University for Advanced Studies, SOKENDAI, Tokyo, Japan

Senior Editor
Christian Rutz
University of St Andrews, St Andrews, United Kingdom

Reviewer #2 (Public review):

The authors have constructively responded to previous referee comments and I believe that the manuscript is a useful addition to the literature. I particularly appreciate the quantitative approach to social behavior, but have two cautionary comments.

(1) Conceptually it is important to further justify why this particular maximum entropy model is appropriate. Maximum entropy models have been applied across a dizzying array of biological systems, including genes, neurons, the immune system, as well as animal behavior, so would seem quite beneficial to explain the particular benefits here, for mouse social behavior as coarse-grained through the eco-hab chamber occupancy. This would be an excellent chance to amplify what the models can offer for biological understanding, particularly in the realm of social behavior

(2) Maximum entropy models of even intermediate size systems involve a large number of parameters. The authors are transparent about that limitation here, but I still worry that the conclusion of the sufficiency of pairwise interactions is simply not general, and this may also relate to the differences from previous work. If, as the authors suggest in the discussion, this difference is one of a choice of variables, then that point could be emphasized. The suggestion of a follow up study with a smaller number of mice is excellent.

https://doi.org/10.7554/eLife.94999.2.sa2

Reviewer #3 (Public review):

Summary:

Chen et al. present a thorough statistical analysis of social interactions, more precisely, co-occupying the same chamber in the Eco-HAB measurement system. They also test the effect of manipulating the prelimbic cortex by using TIMP-1 that inhibits the MMP-9 matrix metalloproteinase. They conclude that altering neural plasticity in the prelimbic cortex does not eliminate social interactions, but it strongly impacts social information transmission.

Strengths:

The quantitative approach to analyzing social interactions is laudable and the study is interesting. It demonstrates that the Eco-HAB can be used for high throughput, standardized and automated tests of the effects of brain manipulations on social structure in large groups of mice.

Weaknesses:

A demonstration of TIMP-1 impairing neural plasticity specifically in the prelimbic cortex of the treated animals would greatly strengthen the biological conclusions. The Eco-HAB provides coarser spatial information compared to some other approaches, which may influence the conclusions.

https://doi.org/10.7554/eLife.94999.2.sa1

Author response:

The following is the authors' response to the original reviews.

We would like to thank the reviewer and the editor for carefully reading our manuscript, and acknowledging the strength of combining quantitative analysis with semi-naturalistic experiments on mice social behavior. Please find below our response to both the public review and the recommendation to the authors. As a summary, we have included additional figures and texts such as

- a new Results subsection "Choosing timescales for analysis " (page 6)

- a new Materials and Methods subsection "Maximum entropy model with triplet interactions" (page 17)

- new supplementary figures, which have current labels of:

- Figure 2 - figure supplement 5

- Figure 2 - figure supplement 6

- Figure 2 - figure supplement 7

- Figure 4 - figure supplement 1

- Figure 4 - figure supplement 2

> ***Public Reviews:***
>
> ***Reviewer #1 (Public Review):***
>
> *Summary:*
>
> *In this manuscript, Chen et al. investigate the statistical structure of social interactions among mice living together in the ECO-Hab. They use maximum entropy models (MEM) from statistical physics that include individual preferences and pair-wise interactions among mice to describe their collective behavior. They also use this model to track the evolution of these preferences and interactions across time and in one group of mice injected with TIMP-1, an enzyme regulating synaptic plasticity. The main result is that they can explain group behavior (the probability of being together in one compartment) by a MEM that only includes pair-wise interactions. Moreover, the impact of TIMP-1 is to increase the variance of the couplings J_ij, the preference for the compartment containing food, as well as the dissatisfaction triplet index (DTI).*
>
> *Strengths:*
>
> *The ECO-Hab is a really nice system to ask questions about the sociability of mice and to tease apart sociability from individual preference. Moreover, combining the ECO-Hab with the use of MEM is a powerful and elegant approach that can help statistically characterize complex interactions between groups of mice -- an important question that requires fine quantitative analysis.*
>
> *Weaknesses:*
>
> *However, there is a risk in interpreting these models. In my view, several of the comparisons established in the current study would require finer and more in-depth analysis to be able to establish firmer conclusions (see below). Also, the current study, which closely resembles previous work by Shemesh et al., finds a different result but does*

> *not provide the same quantitative model comparison included there, nor a conclusive explanation of why their results are different. In total, I felt that some of the results required more solid statistical testing and that some of the conclusions of the paper were not entirely justified. In particular, the results from TIMP-1 require proper interaction tests (group x drug) which I couldn't find. This is particularly important when the control group has a smaller N than the drug groups.*

We would like to thank the reviewer and the editor for carefully reading our manuscript, and acknowledging the strength of combining quantitative analysis with semi-naturalistic experiments on mice social behavior. Thanks to the reviewer's suggestion, we have improved our manuscript by

(1) A proper comparison with Shemesh et al., especially to include maximum entropy models with triplet interactions. We show that triplet models overfit even given the entire 10 day dataset, which limits our study to look at pairwise interactions.

(2) Results on cross-validation for both triplet interaction models and pairwise interaction models, completed on aggregates of various length of days. This analysis showed that pairwise models overfit for single-day data, and led us to learn pairwise models only on 5day aggregation of data. We have updated the manuscript (both the text and the figures) to present these results.

(3) New results that subsample the drug groups to the same size as the control group. The conclusions about TIMP-1 treated mice hordes hold when we compare groups of the same size.

> ***Recommendations for the authors:***
>
> ***Reviewer #1 (Recommendations For The Authors):***
>
> *(1) COMPARISON WITH PREVIOUS WORK. The comparison with the cited previous work of Shemesh et al. 2013 rests novelty to the use of ME models in characterizing social interactions between groups of mice as well as sheds doubts on the main claim of the manuscript, namely that second-order correlations are sufficient to describe the joint distribution of occupancies of all mice (in particular triplets; there is no quantification of the variance explained by model in panel Fig. 2D). In my view, to make the claim "These results show that pairwise interaction among mice are sufficient to assess the observed collective behavior", the authors should compare models with 2nd and 3rd order interactions and quantify how much of the total correlation can be explained by pair-wise interactions, triplet interactions, and so on. Without a proper model comparison, it is unclear how the authors can make such a claim. One thing observed by Shemesh et al. is that, on average, J_ij are negative. This does not seem to be the case in the current study and the authors should discuss why.*
>
> *Finally, the explanation provided in the Discussion about this discrepancy (spatial resolution and different group size) are not completely satisfactory. With more animals, one would imagine that the impact of higher order correlations would increase (and not decrease) as the number of terms of 3rd, 4th, ... order will be very big. I would also think that the same could be true for the spatial scale: assessing interactions with a coarser spatial grid (whole cages in the case of the ECO-Hab) would allow for simultaneous interactions among more mice to happen compared with a situation in which the spatial grid is so small that only a few animals can fit in each subdivision.*

We thank the reviewer for the recommendation. In the updated version of the manuscript, we explicitly learn the triplet interaction model. We show that because the number of mice in

our experiment is much larger than Shemesh et al., a triplet model runs into the problem of overfitting.

In particular, we found that the test set likelihood increases monotonically when the L2 regularization strength increases, which corresponds to a suppression of the triplet interaction strength (see additional supplementary figure, now Figure 2 - figure supplement 5). More specifically, for the range of regularization strength (β_G) we tested ($10^{-1} < \beta_G < 10^{1}$),

the maximum test set likelihood is achieved at $\beta_G = 10^{1}$, which corresponds to $\langle G_{ijk}^2 \rangle = 10^{-7}$.

Notice that those learned triplet interactions are very close to zero. This means we should select a model with pairwise interactions over a model with triplet interactions.

We have added the above reasoning in page 5, line 166-169 of the Results section with the sentence “Moreover, models with triplet interactions show signs of overfitting under crossvalidation, which is mitigated when the triplet interactions are suppressed close to zero using L2 regularization”, a new subsection “Maximum entropy model with triplet interactions” in Materials and Methods (page 16-17, line 548 - 563) to describe the protocols of learning and crossvalidation for these triplet interaction models.

Furthermore, we extended the discussion about the difference between Shemesh et al. and our results in the Discussion section. In addition to the difference of spatial scales (chamber vs. location in the chamber), and the difference of group size and its impact on data analysis (N = 15 in our largest cohort and N = 4 in theirs), we added a discussion about the difference of experimental arena, which in Eco-HAB contains connected chambers that mimic the naturalistic environment, and in Shemesh et al. contains a single chamber. The change in the text is on page 12, between line 390 and line 394.

We thank the reviewers for pointing out that the mean 2nd order interaction in Shemesh et al. is negative. One possibility is that the labeled areas in Shemesh et al. are much smaller than in our Eco-HAB setup, which could suggest that mice do have the space to stay in the same area, which will lead to a negative mean 2nd order interaction.

> *(2) ASSESSMENT OF THE TEMPORAL EVOLUTION OF THE INTERACTIONS. The analysis of the stability of the social structure is not conclusive. First, I don't think the authors can conclude that "These results suggest that the structure of social interactions in a cohort as a whole is consistent across all days." If anything is preserved, they would be the statistics of that structure but not the structure itself (i.e., there is no evidence for that). The comparison of the stability of the mean <h_i> and the mean <J_ik> would also require a statistical test to be able to state that "Delta h_i changed more strongly from day to day (Fig. 3D, top panel) relative to the interaction measured as the Jij's." The same is true for the assessment of the TIMP: the differences found in the variability in J_ij and in the mean and variance of the h_i's, look noisy and would require a proper statistical test. The traces look quite variable across days in the control condition, so assessing differences may be difficult. Finally, it would be good to know if the variability in individual J_ij is because they truly vary from day to day or because estimating them within one day is difficult (statistical error). If the reason is the latter, one could decrease the temporal resolution to 2-3 days and see whether the estimated J_ijs are more stable. Perhaps, also for that reason, the summed interaction strength J_i is also more stable, simply because it aggregates more data and has a smaller statistical error.*

We thank the reviewer for pointing out the necessity of assessing the temporal evolution of the interactions. The problem of shorter data duration leads to more noise in the estimation, together with the reviewer’s Comment 4 about the risk of overfitting, led us to add a new Results subsection “Choosing timescales for analysis” (page 6, line 171 to line 189). Specifically, we assess whether the pairwise maximum entropy model overfits using data

from _K-_day aggregates, by computing the log-likelihood of both the training sets and the test sets,which is chosen to be 1 hour from the 6 hour data window of each day. We found that for single day data, the pairwise maximum entropy model overfits. In contrast, for data with aggregates of more or equal to 4 days of data, the pairwise model does not overfit. This new result is supported by an additional supplementary figure, now Figure 2 - figure supplement 6.

To be consistent with later approaches in the manuscript where we consider the effects of TIMP1, we choose the analysis windows to be data aggregates from 5 days. This means for the experiment that collects a total of 10 days of data, there are only two time points, thus a study of the temporal evolution is limited to comparison between the first 5 days and the last 5 days of the experiment. We describe these results in the Results subsection “Stability of sociability over time” (page 6, line 190 - 220). An additional supplementary figure, now Figure 2 - figure supplement 7, shows in details the comparison of the inferred interaction strength J and the chamber preference between the first 5 days and the last 5 days for the 4 cohorts of male C57BL6/J mice, which shows the inferred interactions have a consistent variability across first and last 5 days, and across all cohorts. The small value of Pearsons’ correlation coefficient shows that the exact structure (pairspecific J_{ij}) is not stable. At the end of the Results subsection “Stability of sociability over time”, we explicitly say that “This implies that the maximum entropy model does not infer a social structure that is stable over time.”

> *(3) EFFECT OF TIMP-1. The reported effects of TIMP-1 on the variance of the J_ij seem very small and possibly caused by a few outlier J_ijs (perhaps from one or two animals) which*
>
> *are not present in the control group which seems to have fewer animals (N = 9 minus two mice that died after the surgery vs. N = 14 in the drug group), so the lack of a significant difference in the sigma[J_ij] could simply be due to a smaller N (a test for the interaction group x drug was not done).*
>
> *The clearest effect of TIMP-1 seems to be a change in place preference (h_i) and not the interaction terms (J_ij) (Fig. 3F bottom). But this could be explained by a number of factors that have nothing to do with sociability such as that recovery from surgery makes them eat more/less. The fact that it seems to be present, as recognized by the authors, in the control group with no TIMP-1 and that this effect was not observed in the female group F1, puts into question the specificity and reproducibility of the result.*
>
> *Finally, the effect of TIMP-1 in the DTI would require more statistics (testing the interaction group x drug). The fact that the control group has fewer animals (N = 9 vs. 15 and 13 in the drug groups), and that there is a weaker trend in the DTI of the control group to start high and then decrease, makes this test necessary.*

Now, after we select a proper timescale to learn the pairwise maximum entropy model, we update the manuscript to present results only on 5-day aggregation of data (see updated Figure 3, updated supplementary figures, Figure 3 - figure supplement 1 and 2). For the variance of the J_{ij}, the F-test between different 5-day aggregates before and after TIMP for the male drug group now shows a nonsignificant p-value after applying the Bonferroni correction. For the female drug group, the difference of the J_{ij} variance is still significant.

To test the effect of different group size on DTI, we subsampled the drug groups by 1) subsampling the inferred interactions learned from the original N = 15 or N = 13 data, or 2) subsampling the mice colocalization data and then inferring the pairwise interactions. In both cases, the resulting DTI for the subsampled drug group still exhibits the same global pattern as before, i.e. after TIMP-1 injection, DTI significantly increases, which after 5 days falls back to the baseline level. The results are supported by two additional supplementary figures, Figure 4 - figure supplement 1 and 2. This result is referred to in the text in the Results subsection “Impaired neuronal plasticity in the PL affects the structure of social

interactions" (page 10, line 333 - 336): "Notably, the difference of the DTI is not due to the control group M4 has less mice, as subsampling both on the level of the inferred interactions (Figure 4 - figure supplement 1) and on the level of the mice locations (Figure 4 - figure supplement 2) give the same DTI for cohorts M1 and F1."

> *(4) MODEL COMPARISON. Any quantitative measure of "goodness" of the model , (i.e., comparison of the predictions of the model with triplet frequency as well as the distribution of p(K)) should be cross-validated. In particular, Fig. S2 needs to be cross-validated for the goodness of fit to be properly quantified. Is the analysis shown in Fig. 3F crossvalidated? Because otherwise, there is an expected increase in the likelihood simply explained by an increase in the number of parameters of the model (i.e., adding the J_ij's).*

As discussed in our responses to Comment 1 and 2, we have added results about cross-validation in the new supplementary figures, Figure 2 – figure supplement 5 and 6 , for which we computed the test-set and training-set likelihood for maximum entropy models with pairwise interactions and also for models with triplet interactions. Figure 2 - figure supplement 6 shows the pairwise model does not overfit when we consider the aggregated data from more or equal to 4 days.

> *(5) EFFECT OF SLEEP. The comparison of p(K) between the data and the model requires a bit more investigation: the model underestimates instances in which almost all mice were in the same compartment (i.e., for K >= 13. p(K)_data >> p(K)_MEM; btw where is the pairwise point p(15) in Fig. 2E and Fig. S4?). Could this be because there were still short periods during the dark cycle in which all mice were asleep in one of the cages? As explained by the authors, sleep introduces very strong higher order correlations between animals as they like sleeping altogether. Knowing whether removing light periods was enough to remove this "sleep contamination" or not, would be important in order to interpret discrepancies between the pairwise model and the data.*

Figure 2E shows that the pairwise maximum entropy model (in black) overestimates the data (in blue circles) for P(K) at large K (and not underestimates). In the data, we never observe all 15 mice being in the same box; hence $P_{data}(15) = 0$, and does not show up in the log-scaled figure (same for Figure 2 - figure supplement 3). A possible explanation for the pairwise model overestimating P(K) at large K is that the finite-sized box limits the total number of mice that are comfortably staying in the same box. It can also be due to the fact that the number of time points at which K >= 13 is small and hence causes an underestimation due to finite data. We have added this interpretation of the discrepancy of P(K) to Section "Pairwise interaction model explains the statistics of social behavior" in page 6, line 160.

We thank the Reviewer for raising the point of "sleep contamination". Indeed, Eco-HAB data, as do data from other 24h-testing behavioral systems, demonstrate distinct differences in activity levels during the light and dark phases of the light-dark cycle (Rydzanicz et al., *EMBO Mol. Med.*, 2024). During the light phases, mice primarily sleep and, as noted, they huddle, so many individuals within the cohort tend to remain in close proximity for extended periods. We acknowledge that including such periods in the analysis could potentially introduce confounding effects to the model due to limited movement and interactions, and this is why we decided not to use this data. However, during the dark phases, mice are highly active, with individuals rarely staying in the same compartment for long periods. Specifically, in the dark phases, while there are occasional instances where a few mice may remain in the same compartment for over 1 hour, the majority exhibit considerable mobility, actively exploring and transitioning between compartments. We see no compelling reason to exclude these periods from our analysis, as such activity aligns with the natural behavioral repertoire of the mice and provides robust data for our model. Furthermore, it is well-established that

mammals, including nocturnal species such as mice, are most active shortly after waking, typically at the onset of their active phase (i.e., the beginning of the dark phase). To ensure a conservative approach, we specifically analyzed the first 6 hours of the dark phase when the cumulative number of box visits is at its peak, indicating heightened activity levels. In our view, this period offers an optimal window for studying natural behaviors, including social interactions.

Additionally, prior studies using the Eco-HAB system have consistently demonstrated that mice engage in social interactions both within the compartments and in the connecting tubes during the dark phase (Puścian et al., *eLife*, 2016, Winiarski et al. in press). Given this evidence and the observed behavioral dynamics in our data, the likelihood of mice being asleep during the analyzed periods of the dark phase is very low.

We hope this clarification addresses the reviewer's concerns and highlights the rationale underpinning our analysis choices. Thank you for raising this important point, which allowed us to provide additional context for our approach.

> *(6) COMPARTMENT PREFERENCES. The differences between p(K) across compartments also would require a bit more attention: of a MEM with non-spatially dependent pair-wise interactions shows differences across compartments, it must be because of the terms h_{i,r} terms which contain a compartment index, right? Wouldn't this imply that the independence model, which always underrepresents data events with large K, already contains the difference in goodness of fit between compartments (1, 3) and (2, 4)? In the plots, it does not look like the goodness of the independent model depends on the compartment (the authors could compare directly the models' predictions between compartments). Moreover, when looking at Fig. 2C, it does not look like the value of h_{i,r} in compartments (1,3) is higher than in (2,4) (if anything, it would be the other way around). How can this be explained? It would be good to know if the difference across compartments comes from differences in the empirical p(K) or in the models' prediction? If the difference is in the data p(K), could it be that the compartments 2-4 showing higher p(K=15) (i.e., larger difference with the pairwise MEM prediction) are those chosen by mice to sleep during the light cycle? If not, what could explain these differences across compartments? Could the presence of food and water explain this difference?*

The reviewer is correct, in the pairwise MEM, the difference across compartments enter in the box preference h_{ir}. Greater h_{ir} means compartment r is more attractive to mouse i. Because box 2 and 4 contain food and water, we expect that mice are more attracted to box 2 and 4, and this is what we see in Figure 2C, bottom subpanels. To reduce the number of parameters to look at, we introduce an index $\Delta h_i = h_{i2} + h_{i4} - h_{i1} - h_{i3}$. This index Δh_i is found to be mostly positive (see updated Figure 3C), which makes sense because mice are attracted to food and water.

Next we analyze the difference of *P(K)* across compartments (Figure 2 - figure supplement 3). There is already a difference in the *P(K)* calculated from empirical data. For example, *P(K)* in compartment 2 has a maximum at $K = 5$ while *P(K)* in compartment 1 has a maximum at $K = 3$.

One interesting observation is that it seems from Figure 2 - figure supplement 3 that the pairwise model explains *P(K)* in compartment 1 and compartment 3 better than in compartment 2 and in compartment 4. In compartment 2 and 4, the pairwise MEM overestimates *P(K)* for large K. An alternative MEM could include compartment-specific interaction strength, but it will also introduce 315 new parameters for a mice cohort with size $N = 15$.

> *MINOR*

> *(1) A more quantitative comparison between in-cohort sociability and couplings J_ij as œwell as mean rates and parameters h_i is required. The matrices in Fig. 2C do look similar. So it is not clear how the comparison between these values is contributing to characterizing the correlation structure of the data.*

The comparison between in-cohort sociability and coupling J_{ij} is given by supplementary Figure 2 - figure supplement 2. The key point for the model with the learned J_{ij} reproducing the in-cohort sociability is given by Figure 2 - figure supplement 1.

> *(2) Analysis of "in-state" probability is not explained. To me, it wasn't obvious what Fig. S5 is showing. I was assuming that this analysis was comparing the prediction of the MEM about the position of each animal at each time point, given its preference (h), pairwise interactions (J_ij), and the position of all other animals and the true position of the animal. But it seems like it is comparing the shape of the distribution of this prob across time between the data and the model (I guess the data had to be temporally binned in coarser temporal periods to yield prob values other than 0s and 1s). Also, not clear whether this analysis was done for each compartment separately and then averaged. This needs explanation.*

The in-state probability is comparing the prediction of the MEM about the position of each animal at each time point, given its preference (h), pairwise interactions (J_{ij}), and the position of all other animals and the true position of the animal. To achieve values between 0s and 1s, we bin the data temporally according to the model-predicted in-state probability.

We have added the explanation of in-state probability on page 6, line 163-166. We have also improved the description of in-state probability in Materials and Methods (subsection “Comparing in-state probability between model prediction and data”, line 493 - 503, page 15), and added a pointer from the main text to it.

> *(3) Looks like Fig. S3 is not cited in the text.*

We added a pointer to Fig. S3 (now Figure 2 - figure supplement 2) in line 154.

> *(4) The authors say that "TIMP-1 release from the TIMP-1-loaded nanoparticles diminishes after 5 days." Does that mean from the day of the injection (4-5 days before the "After Day 1") or five days after reintroduced in the ECO-Hab?*

It means five days after the mice were re-introduced in the ECO-Hab. We have updated the text in Results/Effects of impairing neuronal plasticity in the PL on subterritory preferences and sociability (the end of the first paragraph of this subsection) to

“The choice of five-day aggregated data for analysis is in line both with the proper timescales needed for the pairwise maximum entropy model to not overfit, and with the literature that TIMP-1 release from the TIMP-1-loaded nanoparticles is stable for 7-10 days after injection (Chaturvedi et al., 2014) (i.e. 2-5 days after the mice are reintroduced to Eco-HAB).” (line 272 - 276, page 9)

> *(5) In Methods, the authors should report the final N of each of the three groups.*

The number of final N is reported in Table 1 (page 13). In the updated version, we have added a pointer to Table 1 in Materials and Methods/Animals, and in Materials and Methods/Exclude inactive and dead mice from analysis. We have also expanded the caption of Table 1 to clarify the difference between final N and initial N, and added a pointer to Materials and Methods/Exclude inactive and dead mice from analysis.

https://doi.org/10.7554/eLife.94999.2.sa0