

Reviewed Preprint

v1 • February 9, 2026

Not revised

Reviewed Preprint

v2 • August 11, 2026

Revised by authors

✉ For correspondence:

wokelsch@uni-mainz.de

¶ These authors shared first authorship.

Competing interests: No competing interests declared

Funding: See [page 25](#)

Reviewing editor: Gordon J Berman, Emory University, United States

© 2026, Reinwald et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Individual differences drive social hierarchies in male mouse societies

Jonathan R Reinwald^{1,2,3,¶}, Sarah Ghanayem^{1,2,¶}, David Wolf^{1,2}, Max F Scheller¹, Julia Lebedeva^{1,2}, Philipp Lebhardt⁴, Oliver Götz⁴, Corentin Nelias^{1,2}, Wolfgang Kelsch^{1,2} ✉

¹Dept. of Psychiatry and Psychotherapy, University Medical Center Mainz, Johannes-Gutenberg University, Mainz, Germany • ²Dept. of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany • ³Dept. of Neuroimaging, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany • ⁴Dept. of Clinical Health Technologies, Institute for Manufacturing Engineering and Automation, Fraunhofer Society, Mannheim, Germany

eLife Assessment

This **important** study introduces NoSeMaze, a semi-naturalistic platform for continuous, high-dimensional tracking of social and cognitive behaviors in group-housed mice, and uses it to show that individual social rank is stable across changing social contexts. By integrating automated dominance measures, proactive social behaviors, and reinforcement-learning-based profiles, the authors demonstrate a novel framework for examining how stable individual differences shape social structure. The findings provide **compelling** evidence that dominance traits are stable across changing social contexts and largely independent of non-social cognitive performance, supporting the view that social rank reflects an intrinsic dimension of individuality however, the broader functional significance of dominance in this paradigm remains somewhat ambiguous, including the extent to which the measured behaviors capture how dominance operates in more naturalistic social settings. This work will be of broad relevance for behavioral neuroscience and social behavior research.

<https://doi.org/10.7554/eLife.109354.2.sa2>

Abstract

Social hierarchies structure groups and confer advantages on high-ranking individuals. In mice, individual position in hierarchies may emerge situationally from current group compositions, or, alternatively, may remain largely stable across groups as an internalized feature. Dominance and subordination are expressed in behaviors like tube competitions or agonistic chasing. The interaction of these behaviors in the shaping of social position in larger male mouse groups remains largely unknown. To address these questions, we developed the NoSeMaze, a semi-naturalistic, open-source, modular platform that enables automated long-term tracking of unperturbed groups. Across more than 4,000 mouse-days, hierarchies derived from incidental competitions in the integrated tube tests were non-despotic, transitive, and stable even when group compositions changed. This stability supports an internalized component of competition-based social rank. Chasing was also stable across contexts. Notably, chasing was concentrated among high-ranking individuals, consistent with ongoing negotiation of social rank among individuals at the upper end of the hierarchy. The link between chasing and social rank strengthened in groups with less well-defined rank structure, where mice rely more on aggressive signaling to assert their position. Chasing and social rank were associated with certain dimensions of simultaneously measured physical and cognitive features. In summary, high-dimensional tracking with the NoSeMaze reveals that social position in mice is multifaceted and shaped by

stable dimensions of individual behavior that persist across changing social contexts. The approach thus enables longitudinal modeling of individuality and social position as key resilience factors.

Introduction

In many social species, individuals organize into hierarchical structures characterized by dominance and subordination relative to others. Social rank shapes access to resources and reduces conflict ^{1–3}, and is the most prominent yet identified resilience factor ^{4–6}. In mice, social position is expressed across different behavioral domains, including competitive encounters and agonistic behaviors such as chasing ^{7–9}. However, the precise nature of the relationship between these behavioral domains remains unclear, particularly in relatively stable social groups and under semi-naturalistic longitudinal conditions ^{9–13}. Moreover, prior studies have yielded inconsistent findings on whether specific facets of social position are associated with individual differences in reward-related behaviors, such as impulsivity, cognitive flexibility, and inhibitory control ^{14–16}.

Social structures in animals are often assumed to emerge from group dynamics that unfold over time ^{2,3}. These processes are difficult to capture with reductionist paradigms ^{17–20} that are not designed for tracking moment-to-moment interactions over long periods and under ecologically valid conditions. Classic assays typically assess social and cognitive behaviors separately, often in different contexts or at different time points, limiting our ability to test how these domains interact and how stable they are within individuals ¹⁷. Moreover, assays such as the tube test ²¹ rely on forced encounters and manual handling, which may introduce confounds ^{22,23} and thereby distort natural behavior.

In mice, small groups tend to form highly despotic hierarchies ^{24–26}, whereas larger groups exhibit more complex structures ²⁷. These patterns suggest a strong influence of emergent group-level dynamics on social structure ^{12,27}. Yet, animals do not enter social groups as blank slates ^{28–30}; stable latent factors in the individual may also contribute to hierarchy formation ¹³. Thus, it remains unclear to what extent an individual's social position is internalized and persists across different social contexts ³¹, or instead is primarily an emergent property of group-level dynamics. Disentangling these possibilities requires experimental conditions that allow unperturbed, continuous tracking of all individuals in sufficiently large groups, together with systematic changes of group composition to modulate social context.

While stable hierarchies and behavioral identity domains have been described previously in semi-naturalistic settings ^{9–13}, many studies quantify these features either within fixed group compositions or in separate assays. Here, we therefore use systematic group reshuffling in 10-member societies to directly test whether individual differences in social rank and chasing persist across distinct social groups. In this study, we use social hierarchy to denote the group-level structure inferred from incidental competitions in the integrated tube tests, and social rank for an individual's level within that hierarchy. Social position serves as an umbrella term for social rank and chasing behaviors. By continuously measuring social rank, chasing, and reinforcement-learning behavior in the same individuals, we further test how chasing and social rank are related to each other and how they relate to individual styles in non-social reinforcement learning.

To enable these tests, we developed the Non-invasive Sensor-rich Maze (NoSeMaze). The NoSeMaze provides a semi-naturalistic environment for fully automated, continuous, long-term behavioral assessment of individual mice living in groups of ten. Importantly, the system continuously monitors social and cognitive domains in parallel, without human intervention. Specifically, the system tracks social rank as real-time outcomes of incidental competitions in integrated tube tests, chasing, and self-initiated reinforcement learning via a go/no-go olfactory task. By combining these measurements across multiple group compositions, the NoSeMaze enables testing of the relative contributions of stable individual characteristics to social rank and chasing, and of how social and cognitive domains interact in shaping an individual's social position.

The resulting longitudinal dataset, encompassing more than 4,000 mouse-days, was used to address three specific goals: (1) establish a high-resolution, longitudinal framework that captures individualized profiles of competition-based social rank, chasing, and reinforcement learning in group-living mice; (2) disentangle how chasing relates to hierarchy structure derived from incidental competitions in the integrated tube tests, and whether either behavior associates with individual differences in non-social reward-seeking; and (3) determine whether social rank, chasing, and non-social reward-seeking behaviors represent stable individual characteristics or dynamic features across time and changing group composition. As a secondary analysis, motivated by oxytocin's established role in social recognition memory ^{32,33}, we also tested whether OXTR deletion in the anterior olfactory nucleus produces detectable shifts in rank dynamics.

Results

Ecological longitudinal assessment in the NoSeMaze

The NoSeMaze provides a complex, semi-naturalistic environment designed for long-term automated assessment of reinforcement learning and social-position-related behaviors in mouse societies. It contains an open-field arena and a housing arena with nesting material and free access to food (Fig. 1A [↗](#), Supplementary Fig. S1 [↗](#)). The arenas are connected by two tubes equipped with RFID readers, which enable automated tracking of social interactions by detecting the identity and movement of individual RFID-tagged mice. Access to water is provided via the olfactory stimulus-outcome learning module, in which mice perform trials ad libitum to meet their water needs. The module is connected to the open-field arena, requiring mice to cross the tubes regularly to access food and water.

We investigated a large population of 79 male mice on the C57BL/6J background that were homozygous for floxed oxytocin receptor alleles (OXTR^{fl/fl}), enabling conditional deletion of the receptor. In 26 mice, OXTR was selectively deleted in the AON pars centralis (OXTR^{AAON}) via viral recombination, induced only after mice had reached full adulthood ³³. OXTR^{AAON} mice show selective impairments in de-novo social olfactory recognition learning ^{32,33}. The remaining 53 mice had normal OXTR expression. This manipulation was included as a secondary biological perturbation. The primary analyses and conclusions focus on the platform and cross-context stability, and genotype is treated as a covariate unless stated otherwise.

The NoSeMaze was originally optimized for male mice in which social hierarchies had been mostly characterized ⁹. Each mouse carried a subcutaneous RFID chip for identification. The study population comprised two age cohorts: younger (16-30 weeks) and older adults (55-97 weeks) (Fig. 1B [↗](#)). The two age cohorts were run as separate experimental series, and group reshuffling was performed within each cohort (Fig. 1C [↗](#), see Supplementary Table S2 [↗](#)). Mice lived in groups of 9-10 for multiple rounds in the NoSeMaze, with different group members in each round (Fig. 1D [↗](#), see Supplementary Table S1 [↗](#)). This allowed us to test which individual behaviors were stable across different group compositions. Before the start of the experiment, mice had been housed in groups of three to five in their home cages, with regular reshuffling to ensure familiarity among potential group members (Fig. 1C-D [↗](#)). To prevent bias in group assignment and behavioral interpretation, both the composition of experimental NoSeMaze groups and the home-cage groupings between rounds were assigned randomly and blinded to prior behavioral outcomes, individual performance, or genotype. Two to four NoSeMazes operated in parallel, with each round spanning three weeks. Between rounds, mice were reassigned to different home cage groups over several weeks (mean between round time: 3.8 weeks; SD: 2.8 weeks, see also Supplementary Table S1 [↗](#)). In total, more than 4,000 mouse-days from 21 NoSeMaze rounds were analyzed (see Supplementary Table S2 [↗](#)).

Individual styles of reinforcement learning in the NoSeMaze

Before the first round in the NoSeMaze, mice were acclimatized to the reward port (see Methods). To assess individual differences in reinforcement learning, they then performed a go/no-go task with two olfactory conditioned stimuli (CS) in the stimulus-outcome learning module. Mice

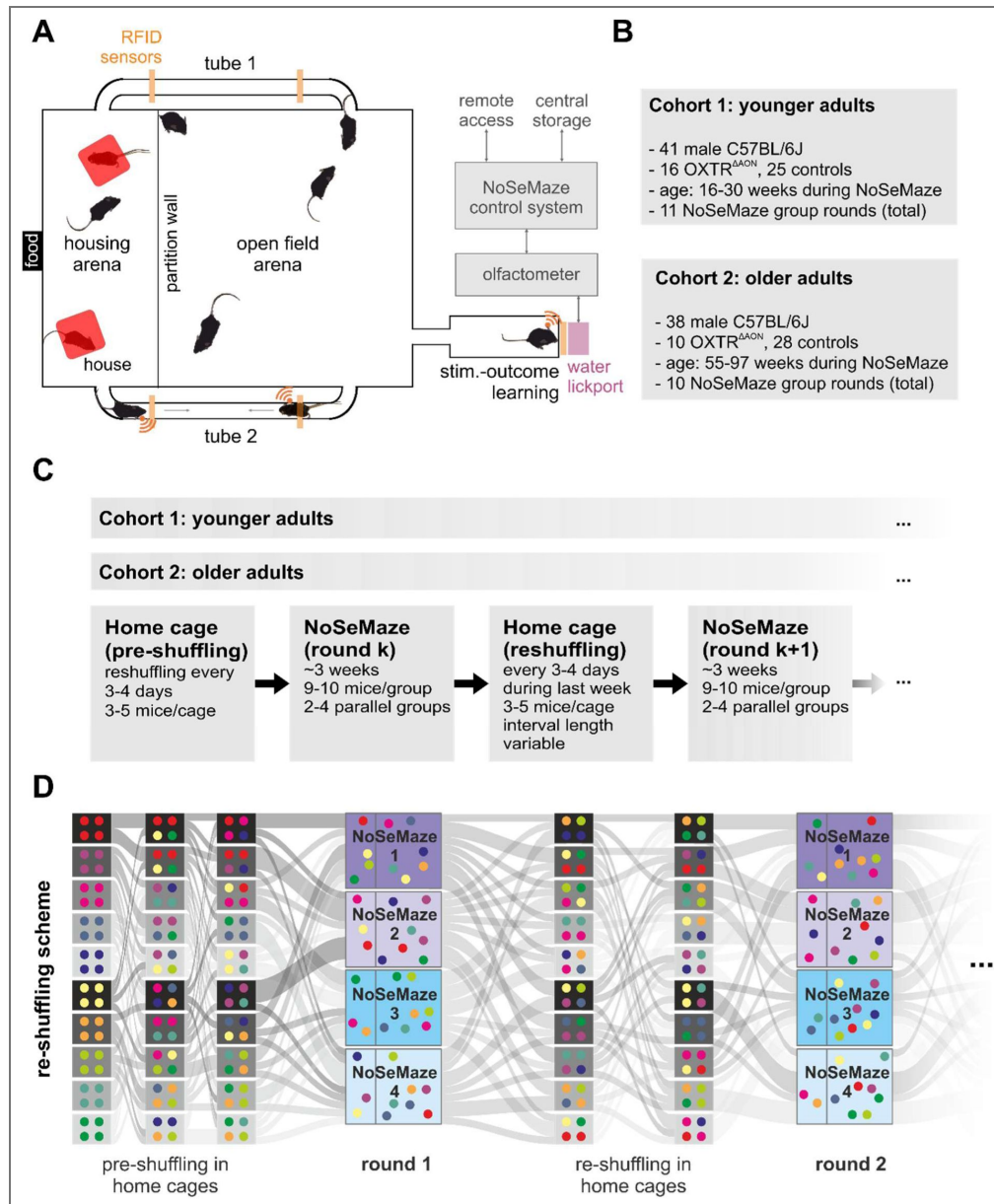


Fig. 1. NoSeMaze setup, cohort structure, and accelerated longitudinal reshuffling design.

A, The NoSeMaze (see Supplementary Fig. S1) tracks social rank, chasing, and reinforcement learning in mouse societies ($n = 9-10$ per group) over multiple weeks without human intervention. RFID-tagged animals are automatically identified at the tube entrances and during olfactory stimulus-outcome learning tasks at the water lickport. **B**, Overview of the two adult cohorts studied in separate experimental series. Cohort 1 (“younger adults”) comprised 41 male C57BL/6J-background mice (16 OXTR^{ΔAON}, 25 controls), aged 16–30 weeks during NoSeMaze testing, and contributed 11 NoSeMaze group rounds in total. Cohort 2 (“older adults”) comprised 38 male mice (10 OXTR^{ΔAON}, 28 controls), aged 55–97 weeks during NoSeMaze testing, and contributed 10 NoSeMaze group rounds in total. **C**, Experimental timeline. Before entry into the NoSeMaze, home-cage compositions were repeatedly reshuffled every 3–4 days (3–5 mice per cage) to promote familiarity among cohort members. Mice were then housed in the NoSeMaze for approximately 3 weeks per round in groups of 9–10, with 2–4 groups run in parallel. Between consecutive NoSeMaze rounds, mice returned to reshuffled home cages, where they were again reshuffled in the last week before the next NoSeMaze round. Reshuffling occurred within cohorts only, not across cohorts. **D**, Example of the reshuffling design across two consecutive NoSeMaze rounds. Colored dots represent individual mice and gray ribbons indicate reassignment from pre-shuffling home cages to NoSeMaze round 1, then to reshuffled home cages, and finally to NoSeMaze round 2. This accelerated longitudinal design systematically altered social group composition across rounds while maintaining animals within their age-defined cohort. *NoSeMaze*, *Non-invasive Sensor-rich Maze*; *RFID*, *radio frequency identification*

initiated trials by inserting their head into the odor port, where they were identified via RFID. In go trials, licking twice during the CS predicting reward (CS+) led to water delivery, while licking twice during the CS predicting no reward (CS-) in no-go trials triggered a timeout of six seconds (Fig. 2A). Reward contingencies alternated every three days between the two odors (i.e., ‘reversal’; Fig. 2B). As expected, activity levels peaked during the dark cycle (Fig. 2C). On average, mice performed 463 trials per day (Supplementary Fig. S2A). The number of trials required to reach 80% correct hits and correct rejections increased after the first reversal, suggesting that the initial reversal was challenging. With subsequent reversals, performance gradually improved, indicating that mice generally learned the task structure (Fig. 2D).

Licking behavior varied widely across individuals, as shown for stable phases of the last 150 trials before reversals (Fig. 2E-G). Inter-individual differences were already seen in the pre-CS licking rates (Fig. 2H) that may serve as a proxy of impulsivity³⁴. Mice also differed in their licking patterns during reward anticipation. While all mice exhibited increased anticipatory licking during CS+ trials (CS+ modulation peak, cf. Fig. 2E-G), this measure of reward sensitivity was independent of pre-CS licking rates, suggesting distinct underlying mechanisms (Fig. 2I, $\rho = -0.02$, $p = 0.79$, Supplementary Fig. S2B). Correct hit rates were relatively high across animals (Fig. 2J), indicating effective CS+ recognition. The high variability in pre-CS licking and correct rejection rates (Fig. 2J) among individuals may again reflect individual differences in impulsivity and inhibitory control³⁴. Pre-CS licking was strongly negatively correlated with correct rejection rates (Fig. 2K, $\rho = -0.88$, $p < 0.001$, Supplementary Fig. S2C), suggesting impaired response suppression in these mice, leading to more timeouts in CS- trials (cf. Fig. 2B).

To quantify behavioral flexibility among mice, we measured CS+ and CS- switch latencies, defined as the trials required to restore (70% correct hits for CS+) or suppress (50% correct rejections for CS-) licking relative to each animal’s pre-reversal CS+ rate. As individualized measures, these latencies capture flexible adaptation rather than absolute performance. Higher pre-CS licking was associated with shorter CS+ switch latencies but slower adaptation to CS- reversals (Fig. 2L, see Supplementary Material – Systematic Statistical Reporting and Supplementary Fig. S3), suggesting that impulsivity facilitates detection of the rewarded cue but hinders economic response suppression.

We therefore examined whether multiple reward-seeking features cluster individuals into different learning strategies. K-means clustering based on CS+ and CS- median switch latencies, correct hit and rejection rates, CS+ modulation peak, and pre-CS licking identified three distinct learning strategies (Fig. 2M). *Cautious no-go learners* (Fig. 2N, green) exhibited the lowest impulsivity and adapted quickly to CS- reversals but updated CS+ associations more slowly, resulting in the lowest correct hit rates. In contrast, *impulsive go learners* (Fig. 2N, purple), characterized by the highest pre-CS licking, quickly adjusted to CS+ reversals but exhibited poor inhibitory control during CS- reversals, leading to frequent timeouts and a low correct rejection rate. Lastly, *flexible learners* (Fig. 2N, orange), who showed relatively low impulsivity, efficiently adapted to both CS+ and CS- reversals, achieving high performance in correct hits and rejections. This last cluster thus displayed an optimal balance between retrieving rewards with minimal timeouts while adapting effectively when cognitive flexibility was required. Flexible learners formed the largest group and also exhibited the strongest modulation of anticipatory CS+ licking.

Detailed statistical reporting for all analyses, including n , degrees of freedom, and 95% confidence intervals, is provided in the Supplementary Material – Systematic Statistical Reporting. **Extended Data** includes additional linear mixed-effects models controlling for potential confounding variables.

In summary, the NoSeMaze provides quantifiable measures of reinforcement learning, including flexibility, response adaptation, and impulsivity. The three behavioral clusters demonstrate how the NoSeMaze captures individual differences in learning flexibility and inhibitory control.

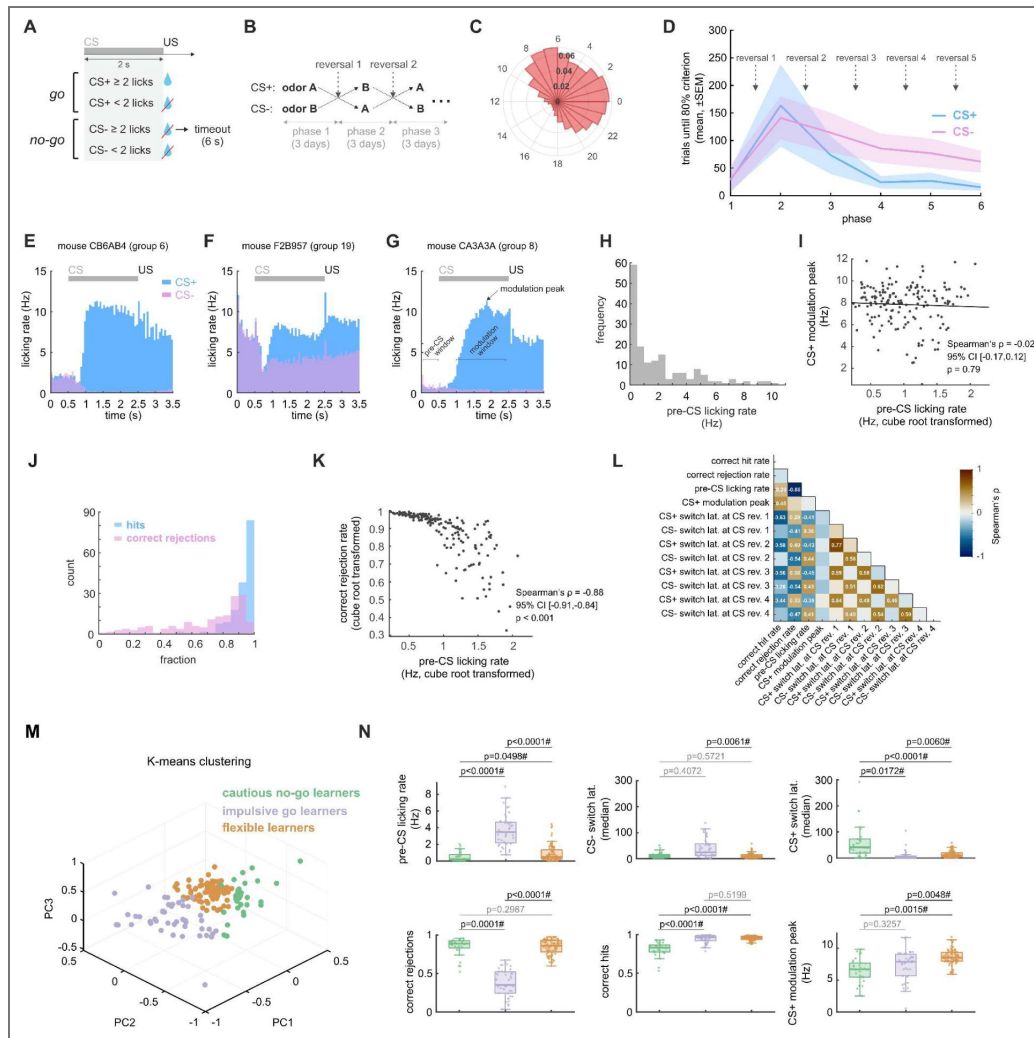


Fig. 2. Inter-individual differences in reinforcement learning in the NoSeMaze.

A, Mice performed stimulus-outcome learning trials for daily water intake, initiated by head insertion into the odor port. In CS+ trials, licking twice during odor presentation released water, while in CS- trials, it triggered a 6-second timeout. **B**, Reward contingencies switched every three days. **C**, Trial activity peaked during the dark phase. **D**, Number of trials required to reach 80% performance (hits or correct rejections, mean $\pm$ SEM) in each phase (cf. Fig. 2B). Performance decreased after the first reversal and improved in later phases, consistent with learning the task structure. Only data from the first round in the NoSeMaze was considered. **E-G**, Example lick patterns from three mice show variability in pre-CS licking rates and modulations during CS presentation in the stable phases of the last 150 trials before reversals. **H**, Pre-CS licking rates varied widely across animals. **I**, No correlation was observed between pre-CS licking rates and CS+ modulation peaks (see Supplementary Fig. S2B). **J**, The fraction of correct rejections was more broadly distributed across mice than hits. **K**, Pre-CS licking rates negatively correlated with correct rejection rates (see Supplementary Fig. S2C). **L**, Spearman's correlation matrix (data from 17 groups) revealed significant relationships between reinforcement-learning metrics, including pre-CS licking rate, correct hit and rejection rates, CS+ modulation peak, and latency to switch after contingency reversals. Only statistically significant correlation coefficients are shown (Bonferroni-corrected for multiple comparisons). **M**, K-means clustering based on reward-seeking features identified three distinct learning strategies: impulsive go learners (purple), cautious no-go learners (green), and flexible learners (orange). **N**, The three learning strategies (color labeling as in **M**) differed across reward-seeking features including hit and rejection rates, pre-CS licking rates, CS switch latencies, and CS+ modulation peak. Horizontal bars denote statistical comparisons between clusters (p-values from permutation tests with $n = 100,000$ permutations using the median). P-values were corrected for multiple comparisons using the Benjamini-Hochberg false discovery rate (FDR) procedure ($\alpha = 0.05$). Significant comparisons after FDR correction are marked with #. CI, confidence interval; CS+, rewarded conditioned stimulus; CS-, unrewarded conditioned stimulus; SEM, standard error of the mean. Note: The regression line is illustrative only; statistical inference is based on Spearman's correlation, which does not assume linearity.

Social interactions in the tubes

The NoSeMaze enables automated, real-time tracking of two distinct social interactions in the tubes: incidental dyadic competitions and voluntary tube chasing. Here, social hierarchy refers to the group-level structure reconstructed from incidental competitions in the integrated tube tests, and social rank to the individual's level herein. Social position is used as an umbrella term for both social rank and chasing.

Tube competitions occur incidentally when two mice enter the tube from opposite ends at the same time, revealing dominance as the other retreats (Fig. 3A [↗](#); see Supplementary Videos S1 [↗](#)–3 [↗](#) for insightful examples). The dominant mouse is detected sequentially at both ends, whereas the subordinate is detected twice at its entry site – once upon entry and again when retreating. Each dyadic tube competition results in symmetrical win/loss outcomes, enabling the construction of interaction networks and social ranks using David's scores (Fig. 3B [↗](#)), which also consider the opponents' social ranks. Competitions occurred primarily during the dark cycle (Supplementary Fig. S4A–B [↗](#)).

Social hierarchies derived from tube competitions were well-structured and stable, showing transitivity values exceeding 0.7 and moderate steepness across groups (Fig. 3C [↗](#), see 'Source Data'). Stability scores above 0.8 and high uncertainty-by-repeatability values [↗](#) confirmed robust social hierarchy sampling over the three-week period (Fig. 3C [↗](#)). David's scores also showed stable correlations across weeks (Fig. 3D [↗](#), $\rho = 0.45 - 0.54$, all $p < 0.001$), indicating consistent social ranks over time. Moreover, David's scores correlated strongly with dynamic social hierarchy measures such as the Elo rating (Fig. 3E [↗](#), $\rho = 0.87$, $p < 0.001$), demonstrating their convergent validity in capturing social rank. Notably, social ranks with and without adjustment for entry time differences were highly correlated (Fig. 3F [↗](#), $\rho = 0.84$, $p < 0.001$), supporting that social rank assessments were largely independent of entry time. Finally, David's scores from the NoSeMaze significantly correlated with those from a manual tube test conducted after the experiment, when animals were again individually housed (Supplementary Fig. S4C [↗](#)).

Body weight showed a moderate positive association with social rank (Supplementary Fig. S5A [↗](#)). Interestingly, this association was driven by a significantly lower fraction of losses in heavier individuals (Fig. 3G [↗](#), $\rho = -0.44$, $p < 0.001$, Supplementary Fig. S5B [↗](#)), while body weight did not correlate to the fraction of wins (Fig. 3G [↗](#), $\rho = 0.06$, $p = 0.395$, Supplementary Fig. S5C [↗](#)). Thus, larger mice were less likely to be defeated in tube competitions, but did not win more frequently.

Unlike incidental tube competitions, tube chasing is a voluntarily initiated interaction in which one mouse actively chases another through the tube (Fig. 4A [↗](#)). To quantify these behaviors, we calculated the fraction of actively initiated chases and the fraction of times being chased for each mouse, relative to the total number of chasing events in its group (Fig. 4A [↗](#), Supplementary Fig. S6A [↗](#), left panel). Chasing occurred predominantly at night (Fig. 4B [↗](#)). The expression of chasing varied widely across individuals (cf. Fig. 4A [↗](#), Supplementary Fig. S6A [↗](#), left panel) and was notably skewed.

More specifically, a small subset of mice initiated most chases (Fig. 4C [↗](#)), with the two most active chasers accounting for more than 50% of all chases in a group. Across weeks, individual chasing behavior remained stable within the same NoSeMaze round (Fig. 4D [↗](#), $\rho = 0.82$ to 0.92 , all $p < 0.001$, Supplementary Fig. S7 [↗](#)). Body weight was unrelated to active chases (Fig. 4E [↗](#), $\rho = 0.09$, $p = 0.235$), but heavier animals were less likely to be chased ($\rho = -0.34$, $p < 0.001$).

In summary, social interactions in the NoSeMaze revealed distinct, quantifiable differences in chasing and competition-based social rank, which remained stable within a NoSeMaze round. Incidental tube competition outcomes established reliable social hierarchies, while chasing reflected individual differences in proactive social behavior.

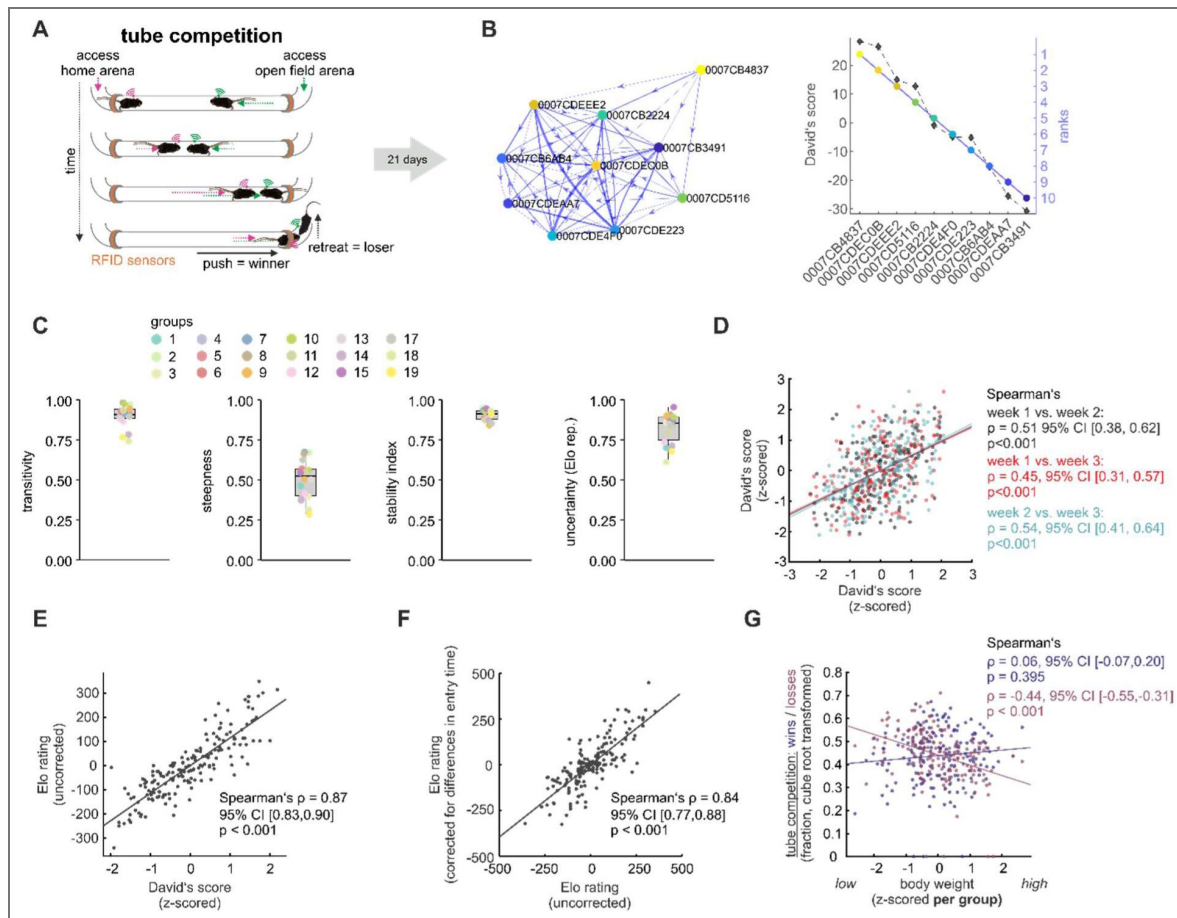


Fig. 3. Social rank derived from incidental competitions in the integrated tube tests of the NoSeMaze.

A, Incidental encounters in the tubes triggered dyadic tube competitions. **B**, Network graph derived from the cumulative tube competitions in an example group over three weeks. Outgoing arrows indicate wins, incoming arrows indicate losses. Individual hierarchical positions were calculated using the David's score and can be converted into linear social ranks from 1 (highest) to 10 (lowest). **C**, Box plots of metrics characterizing social hierarchy for 18 groups, including transitivity, steepness, stability, and uncertainty-by-repeatability (for details, see 'Source Data'). Group labels correspond to original experimental group IDs. Only groups with available tube-competition data are shown. Therefore, numbering is non-consecutive (e.g., groups 16, 20, and 21 are absent; see Supplementary Table S2). **D**, Scatterplots show that David's scores (z-scored) were significantly correlated across the three different weeks, indicating temporal stability of social rank within one NoSeMaze round (data from 18 groups). **E**, David's scores were strongly correlated with Elo ratings, demonstrating convergent validity of social rank measures. **F**, Elo ratings corrected for differences in entry time highly correlated to uncorrected Elo ratings. **G**, Relative body weight (z-scored per group) was significantly negatively correlated to the fraction of losses in tube competitions (cube root transformed, red), with heavier animals being less likely to lose. No association was found between body weight and the fraction of wins in tube competitions (cube root transformed, dark blue). Body weight was measured prior to the animals' introduction to the NoSeMaze, respectively. CI, confidence interval; SD, standard deviation; Note: Regression lines are illustrative only; statistical inference is based on Spearman's correlation, which does not assume linearity.

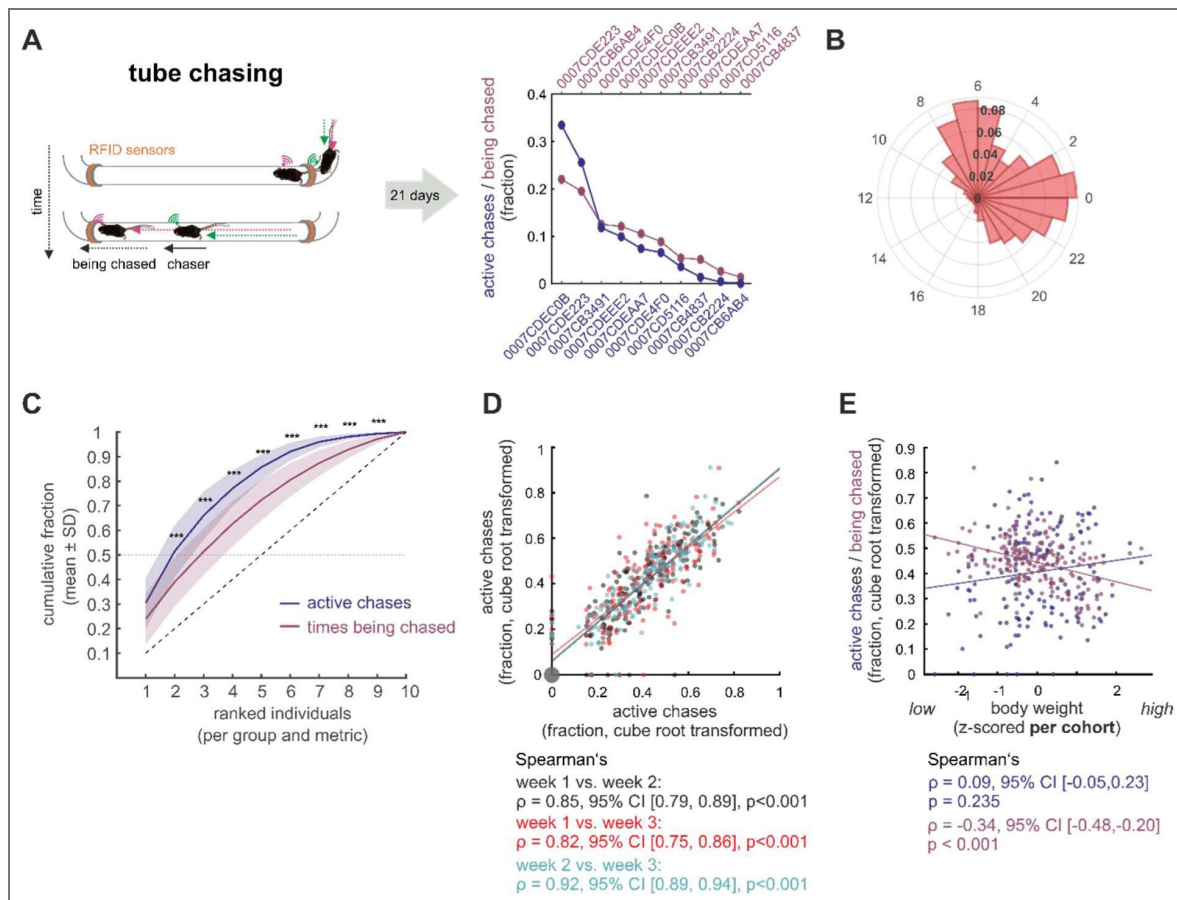


Fig. 4. Chasing in the NoSeMaze.

A, Schematic of the NoSeMaze setup for automated tracking of tube chasing. Chasing was quantified as the fraction of chases initiated by each individual relative to the total number of chases in their group ('active chases'), as well as by the fraction of times being chased (see Supplementary Fig. S6A). **B**, Chasing occurred predominantly during the dark phase. **C**, Cumulative fraction of chases initiated (dark blue) and received (dark red) by individuals within each group, ordered by the fraction of chases they initiated. Chasing was concentrated in a small subset of individuals, with the two most active chasers accounting for more than 50% of all chases. Significant stars mark differences between the two curves (unpaired permutation test, $n=10,000$ permutations, ***, $p < 0.001$). **D**, Active chases were highly consistent across weeks. **E**, Active chases were not significantly correlated to relative weight before entering the NoSeMaze. However, the fraction of times being chased was negatively correlated with weight, indicating that heavier mice were less likely to be chased. CI, confidence interval; SD, standard deviation; Note: Regression lines are illustrative only; statistical inference is based on Spearman's correlation, which does not assume linearity.

Stability of individual behaviors across social contexts

To assess whether behavioral features persist beyond a single NoSeMaze round, we examined their stability across rounds with different group members. More specifically, we quantified across-round stability with Spearman's correlations between rounds 1 and 2 and with intraclass correlation coefficients (ICCs) computed across all rounds. High between-round correlations and high ICCs would indicate that these behaviors are dominated by stable between-mouse differences rather than transient adjustments to changing social contexts.

We found strong correlations between rounds for social rank and chasing. Specifically, stability was high for the tube competition-based David's score (Fig. 5A [↗](#), $\rho = 0.57$, $p < 0.001$), as well as for the fraction of active chases (Fig. 5B [↗](#), $\rho = 0.75$, $p < 0.001$) and of times being chased (Fig. 5C [↗](#), $\rho = 0.54$, $p < 0.001$). Notably, the fraction of active chases showed slightly higher across-round stability than tube-derived David's score and the fraction of being chased. This is in line with active chases capturing an individual propensity to initiate this behavior, whereas competition-based social rank is a relational measure that is also influenced by the set of competitors and interaction opportunities in each reshuffled group. Nonetheless also social rank and being chased were overall stable. Across the full series, ICCs for these social measures were also in the good–excellent range, indicating high across-round stability (Fig. 5D [↗](#), ICC = 0.548 to 0.890).

For reinforcement learning, some features were stable across rounds while others were not (Fig. 5E [↗](#)). Correct hit rates correlated only moderately between rounds ($\rho = 0.34$, $p = 0.008$), with even lower ICCs (ICC = 0.21). In contrast, pre-CS licking ($\rho = 0.69$, $p < 0.001$, Supplementary Fig. S8A [↗](#)) and correct rejection rates ($\rho = 0.78$, $p < 0.001$, Supplementary Fig. S8B [↗](#)) showed higher between-round correlations but only moderate ICCs. Despite shifts in absolute levels, the pattern is consistent with their interpretation as stable individual styles related to impulsivity and inhibitory control (cf. Fig. 5E [↗](#)). Reward sensitivity, measured by peak CS+ licking modulation during anticipation, remained also stable across rounds ($\rho = 0.56$, $p < 0.001$, cf. Fig. 5E [↗](#), Supplementary Fig. S8C [↗](#)) with a similar ICC. In contrast, the stability of CS+ and CS− switch latencies varied substantially. CS+ switch latencies exhibited low-to-moderate between-round correlations ($\rho = 0.16$ to 0.45 across reversals; median: $\rho = 0.50$) and only fair ICCs, whereas CS− switch latencies showed no reliable between-round correlation and poor ICCs (Fig. 5E [↗](#)). The generally shorter switch latencies in the second round (Supplementary Table S3 [↗](#)) are consistent with meta-learning of the reversal rule, with a stronger effect on suppressing responses to CS− than on re-learning CS+. All statistical details are provided in the Supplementary Material – Systematic Statistical Reporting [↗](#).

Cluster assignments across rounds showed moderate-to-strong contingencies ($\chi^2 = 25.51$, $p < 0.0001$, Cramer's $V = 0.4573$, Fig. 5F [↗](#)). Flexible learners, and especially impulsive go-learners, remained mostly stable across rounds (contingencies: 62.5% and 72.2%, respectively, Supplementary Fig. S8D [↗](#)). In contrast, cautious no-go learners were more likely to shift (contingency: 36.4%).

Together, convergent evidence from Spearman's correlations between the first two rounds and ICCs across all rounds indicates that social rank, chasing, and reward-seeking features including impulsivity, inhibitory control, and reward sensitivity measured in the NoSeMaze represent stable individual styles.

Because the number of rounds in the NoSeMaze was unbalanced between animals (Supplementary Table S3 [↗](#)) and age varied across groups (Supplementary Table S1–2 [↗](#)), we also performed robustness checks. Stability estimates changed only minimally when recomputed in age-adjusted ICC models (between-group mean age and within-group age deviation; see Methods) and when restricting analyses to a balanced first-two-session subset (sessions 1–2 only; mice with both sessions) (Supplementary Table S4 [↗](#)). Mixed models indicated that some chasing and reinforcement-learning measures showed modest age- and/or session-related shifts in absolute levels (Supplementary Table S5 [↗](#)), but importantly, these did not affect the observed stability patterns.

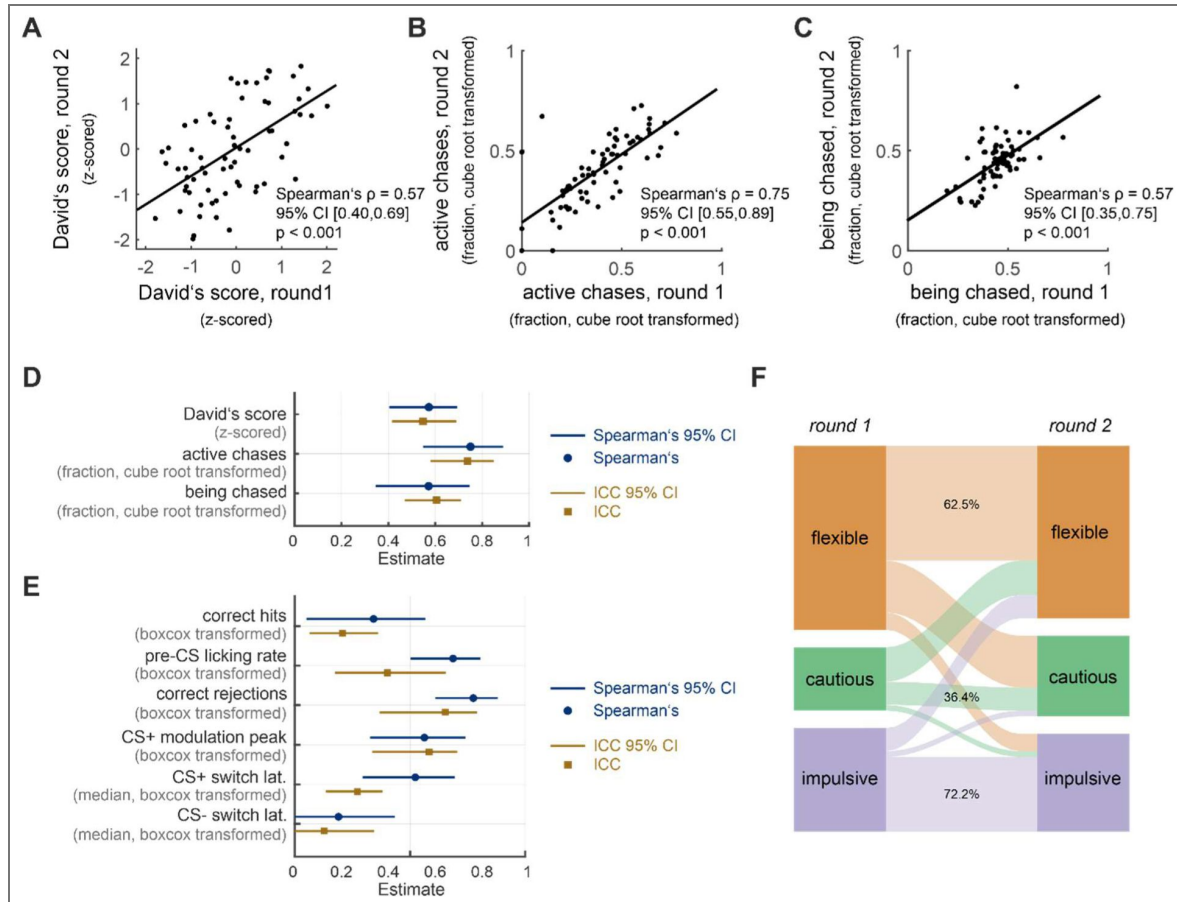


Fig. 5. Stability of social rank, chasing, and reward-seeking features across repeated NoSeMaze rounds.

A–D, Social rank and chasing were stable across NoSeMaze rounds. Individual David's scores (**A**), fraction of active chases (**B**), and fraction of times being chased (**C**) were all significantly correlated between round 1 and 2 (Spearman's $\rho = 0.57$, 0.75 , and 0.54 , respectively; all $p < 0.001$). Lines in **A–C** are illustrative only; inference is based on Spearman's correlation. (**D**) Intraclass correlation coefficients (ICC) confirm similar stability for the three metrics across all rounds. Blue circles show Spearman's ρ between rounds 1 and 2 with 95% CIs; golden squares show ICC with 95% CIs estimated across all rounds to quantify across-round stability of each metric. **E**, Stability of key reward-seeking features. As in **D**, blue circles depict Spearman's ρ between rounds 1 and 2 (95% CIs), whereas orange squares give ICCs computed across all rounds (95% CIs) to index stability over the full longitudinal series. Features include correct hits and correct rejections, pre-CS licking rate, CS+ modulation peak, and switch latencies for CS+ and CS– (see also Supplementary Fig. S8A–C). **F**, Sankey diagram illustrates consistency in individual reward-seeking strategies across rounds. Most mice retained their behavioral strategy classification (flexible, cautious, or impulsive) between rounds, with the majority of flexible and impulsive animals remaining in the same category (see Supplementary Fig. S8D). CI, confidence interval; CS+, rewarded conditioned stimulus; CS–, unrewarded conditioned stimulus. Note: Regression lines are illustrative only; statistical inference is based on Spearman's correlation, which does not assume linearity.

As a secondary analysis, we tested whether $\text{OXTR}^{\Delta\text{AON}}$, which impairs de novo social recognition memory required for social clique formation in this cohort ³⁶, also affects the measures reported here. Consistent with largely internalized features, $\text{OXTR}^{\Delta\text{AON}}$ produced only transient effects. Social rank was significantly lower during the first week in $\text{OXTR}^{\Delta\text{AON}}$ mice, but normalized over the following two weeks (Supplementary Fig. S9A [↗](#)), and no effects of $\text{OXTR}^{\Delta\text{AON}}$ were observed on chasing (Supplementary Fig. S9B-C [↗](#)) or reinforcement learning (Supplementary Fig. S10 [↗](#)). In line with this, all reported stability relationships remained robust and statistically significant when controlling for $\text{OXTR}^{\Delta\text{AON}}$ (see Extended Data Main [↗](#)).

Relationship between social rank, chasing, and reward-seeking features

Finally, we examined whether and how competition-based social rank, chasing, and cognitive styles were related, both at the individual and group level. We first compared social rank and chasing. Notably, David's scores from tube competitions showed a moderately positive association with the fraction of active chases (Fig. 6A [↗](#), $\rho = 0.34$, $p < 0.001$), suggesting that mice of high social rank were more likely to initiate chases. However, this association varied substantially across groups (range: $\rho = -0.03$ to $\rho = 0.62$, cf. Fig. 6B [↗](#)). We thus asked whether the strength of this relationship depended on how clearly the group's social hierarchy was structured. Indeed, at the group level, the strength of the association between David's scores and active chases was negatively related to transitivity (Fig. 6B [↗](#), $\rho = -0.64$, $p < 0.001$), indicating that in groups with less clearly defined social hierarchies, active chases aligned more strongly with social rank from tube competitions. This suggests that aggressive signaling to assert social position becomes more important when dominance-subordination relationships are ambiguous. Conversely, in highly transitive groups, where social hierarchies were well established, active chases were less closely aligned with social rank, implying that social structure reduces the need for proactive reinforcement of dominance. Supporting this interpretation, the fraction of active chases performed by the mouse with the highest social rank also decreased with higher group transitivity (Supplementary Fig. S11 [↗](#), $\rho = -0.61$, $p < 0.001$).

To gain further insight into the interplay between social rank and chasing, we examined the structural properties of both behaviors in more detail. When considering only social rank, the David's score correlated strongly with both the fraction of wins ($\rho = 0.76$, $p < 0.001$) and the fraction of losses ($\rho = -0.67$, $p < 0.001$) in tube competitions (Fig. 6C [↗](#), Supplementary Fig. S12A-B [↗](#)). The integrated tube tests thus meet the assumption of symmetry, where both dominant pushing and subordinate retreat equally contribute to social rank ^{1,37}.

Tube crossings in the NoSeMaze are motivated by the intent to eat, drink, sleep, or socialize. Accordingly, competitions within the tube arise by chance when two animals enter from opposite sides at the same time. Social rank therefore captures the consistent outcomes of repeated incidental competitions (push versus retreat). This measure is not confounded by differences in tube engagement, as social rank was neither associated with participation in tube competitions ($\rho = 0.11$, $p = 0.139$; Fig. 6C [↗](#), Supplementary Fig. S12C [↗](#)) nor with the overall number of tube detections when controlling for chasing (Spearman's partial correlation between detection count and z-scored David's score, corrected for the fraction of active chases: $\rho = 0.022$, $p = 0.758$).

In contrast, these assumptions were not met for chasing (Fig. 6C [↗](#)). This is best illustrated by constructing a 'chasing David's score'. This score was positively associated with the total number of chasing events ($\rho = 0.45$, $p < 0.001$, Fig. 6C [↗](#), Supplementary Fig. S12D [↗](#)), reinforcing that chasing reflects volitional, actively initiated behavior. Importantly, the constructed 'chasing David's score' also correlated strongly with the fraction of active chases ($\rho = 0.85$, $p < 0.001$, Fig. 6C [↗](#), Supplementary Fig. S12E [↗](#)) but was entirely unrelated to the fraction of times being chased ($\rho = -0.01$, $p = 0.897$, Fig. 6C [↗](#), Supplementary Fig. S12F [↗](#)). This fundamental difference suggests that, within a group, chasing did not follow a symmetric dominance-subordination gradient as seen in social rank. Instead, it reflected an unbalanced distribution of initiators and recipients. Consistently, the relationship between the fractions of active chases and being chased differed significantly from the relationship between the fraction of wins and losses in the tube competition

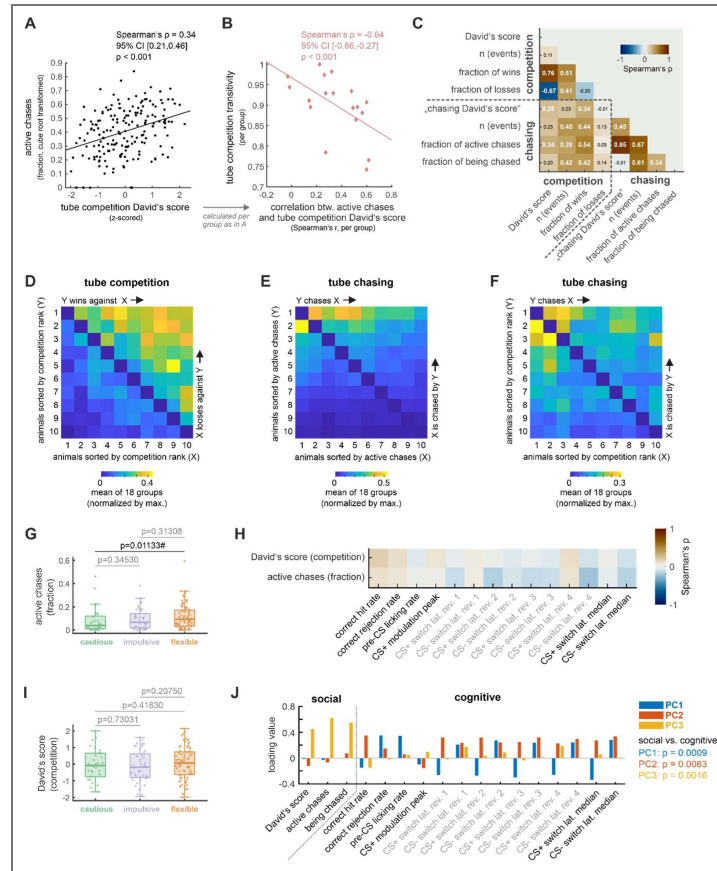


Fig. 6. Relationship between social rank, chasing, and reward-seeking features.

A, Social rank (David's score) was positively correlated with the fraction of active chases. **B**, Across groups, the correlation coefficient calculated between David's scores from tube competitions and fraction of active chases (x-axis, calculated separately within each group, as in **A**) was negatively associated with the group-level transitivity of the social hierarchy derived from tube competitions. Groups with lower transitivity showed a stronger association between social rank and chasing, suggesting that aggressive status signaling via chasing becomes more relevant when social hierarchies are less well-defined. **C**, Correlation matrix showing relationships between social rank and chasing metrics. Color scale indicates Spearman's correlation coefficients; statistically significant values (Bonferroni-corrected) are shown in white. Note that the David's score from tube competitions correlated to both, fraction of wins and losses in tube competitions, while the 'chasing David's score' was strongly correlated only to the fraction of active chases, but not to the fraction of times being chased. **D-F**, Wins and losses in tube competitions (**D**) were concentrated in the top-right, occurring most consistently between animals of opposing social ranks. In contrast, chasing interactions (**E, F**) clustered in the top-left, and thus mostly within a subset of mice with high social rank (**F**), in which they may help clarifying dominance. **G**, Comparison of the fraction of active chases between the three different clusters defined from the reward-seeking task. Cautious mice initiated less chases than flexible individuals (p-values from permutation tests with $n = 100,000$ permutations using the median). Box plots show the median (line), interquartile range (IQR, box), and whiskers extending to $1.5 \times$ the IQR. P-values were corrected for multiple comparisons using the Benjamini-Hochberg false discovery rate (FDR) procedure ($\alpha = 0.05$). Significant comparisons after FDR correction are marked with #. **H**, Correlation heatmap between David's score (social rank from tube competitions), active chases, and reward-seeking features showed no significant associations after FDR correction for multiple comparison. **I**, Comparison of the David's score (social rank from tube competitions) between the three different clusters defined from the reward-seeking task showed no significant difference (p-values from permutation tests with $n = 100,000$ permutations using the median). Box plots show the median (line), interquartile range (IQR, box), and whiskers extending to $1.5 \times$ the IQR. **J**, Loading values from a principal component analysis of social rank (David's score), chasing (active chases and being chased), and reward-seeking (cognitive) features for the top three components (PC1–PC3). Bars indicate each feature's contribution to the respective component. PC1 and PC2 were dominated by cognitive features ($p = 0.0009$ and $p = 0.0063$), PC3 by social rank and chasing ($p = 0.0016$; permutation tests, $n = 10,000$). The largely non-overlapping loadings suggest orthogonality between social and cognitive domains. CI, confidence interval; CS+, rewarded conditioned stimulus; CS–, unrewarded conditioned stimulus. Note: Regression lines are illustrative only; statistical inference is based on Spearman's correlation, which does not assume linearity.

(cf. Supplementary Fig. S6A-B, Supplementary Fig. S13). As a result, the fraction of active chases provided a more direct measure of volitional engagement in chasing than its David's score, as it directly quantified proactive initiation.

With this distinction in mind, we further explored the specific relationship between incidental tube competitions and active chases (Fig. 6D-F, Supplementary Fig. S14). The distribution of wins and losses in incidental tube competitions (Fig. 6D, Supplementary Fig. S14A) differed from those of active chases and being chased (Fig. 6E, Supplementary Fig. S14B). Specifically, a subset of mice expressed high levels of both active chases and being chased (Fig. 6F), and this subset preferentially mapped onto mice with high social ranks in tube competitions (top left cluster, Fig. 6F, Supplementary Fig. S14C). These findings suggest that chasing may serve to negotiate relative social rank among animals with high social rank.

Finally, we examined how social rank and chasing mapped onto reward-seeking strategies. Cautious no-go learners were less likely to engage in active chases compared to flexible learners (Fig. 6G, permutation test with 10,000 permutations, $p = 0.01133$). Apart from this, no significant group differences were found for competition-based David's score (cf. Fig. 6I, see Supplementary Material – Systematic Statistical Reporting for details) or fraction of times being chased (Supplementary Fig. S15). Moreover, neither social rank nor active chases were associated with individual reward-seeking features, including impulsive pre-CS licking, reward sensitivity, inhibitory control, or cognitive flexibility (Fig. 6H, for Spearman's ρ , raw and corrected p -values, see Supplementary Material – Systematic Statistical Reporting). Consistent with this dissociation, principal component analysis showed that social and cognitive domains loaded onto distinct components, further supporting their independence (Fig. 6J, permutation test (10,000 permutations) on the mean loadings values; PC1: $p = 0.0009$, PC2: $p = 0.0063$, PC3: $p = 0.0016$). Notably, all reported between-metric relationships remained robust and statistically significant when controlling for $\text{OXTR}^{\Delta\text{AON}}$ (see Extended Data Main).

Together, these findings suggest that the here examined non-social cognitive features are largely orthogonal to the individuals' social rank and propensity to chase.

Discussion

Ecologically enriched, yet experimentally controlled assessments allow us to study behavioral individuality and social structure in group-living animals over extended timescales. Here, we show that individual mice carry stable, individual-specific, and multi-faceted profiles of social position and cognitive styles across changing group contexts. Our approach goes beyond prior work by testing cross-context stability under repeated, systematic group reshuffling in larger mouse societies, while measuring competition outcomes, chasing, and reinforcement-learning behavior in parallel. To capture this, we had developed the NoSeMaze, an open-source modular platform for continuous tracking of spontaneous behavior in socially housed mice. Although ecological approaches are increasingly valued in neuroscience¹⁷, few existing systems support long-term, automated monitoring of individuals across multiple behavioral domains in larger mouse groups. Most current paradigms are tailored to specific behaviors, such as aggression^{24,25}, conflict³⁸, or locomotor activity³⁰, and offer only limited integrative capacities^{12,19,39–41}. The NoSeMaze addresses this gap by the simultaneous, longitudinal assessment of social rank, chasing, and olfactory reinforcement learning. Its modular, open design supports future expansion to other behavioral domains and integration with neurobiological tools, providing a flexible framework for studying social structure and behavioral individuality. This study focused primarily on the relation of social rank and chasing. We however also considered their relation to additional variables including the loss of oxytocin receptors in the olfactory cortex in the adult ($\text{OXTR}^{\Delta\text{AON}}$), involved in de novo social recognition learning. The propensity to chase was largely unaffected by $\text{OXTR}^{\Delta\text{AON}}$. Mice carrying $\text{OXTR}^{\Delta\text{AON}}$ displayed a transient reduction in social rank during the first week that normalized thereafter. This transient effect contrast to the persistent impairment by $\text{OXTR}^{\Delta\text{AON}}$ in forming higher-order social bonds that enable membership in stable cliques, as identified by video tracking of self-paced interactions in the same cohort³⁶. Together, these

findings suggest that OXT-dependent olfactory learning is critical for the formation of social context-dependent higher-order bonds, but plays a limited role in shaping hierarchy-related behaviors.

Social hierarchies^{1,42} were assessed from incidental dyadic interactions in the novel integrated tube-test module. This method enabled social rank estimation without experimenter intervention and was validated against manual tube testing. Importantly, participation frequency and differences in tube entry time did not confound the resulting social ranks, underscoring the robustness of the automated incidental rank assessment. Social rank was equally shaped by dominance and subordination in tube competitions. This supports Bernstein's assertion³⁷ that social ranks are maintained not only by dominance but also by voluntary subordination. While this study focused on male mice, in which social hierarchies are best established⁹, future work is needed to explore sex-specific expressions of social structure and their neurobiological underpinnings in female groups. We therefore restrict our interpretation to males. Critically, male social ranks were robustly maintained within the same group over time. Even more interestingly, the longitudinal accelerated design, in which animals lived in different groups composed of different individuals, allowed us to test whether social rank is predominantly driven by current group composition or reflects an internalized component that generalizes across social contexts. Indeed, individuals occupied similar relative social ranks across time and groups with different members. This temporal and contextual stability supports interpreting social rank as a stable, internalized characteristic of individuals. In this sense, 'internalized' refers to stability across repeated rounds and reshuffled groups in this paradigm and to relatively stable tube-competition outcomes. The tube-derived social rank describes here a dimension of individual social behavior. The capacity to quantify stable individual differences across changing social contexts highlights the value of the NoSeMaze for lifespan-oriented studies of behavioral individuality. A future direction is to extend this framework to earlier developmental stages to understand which early experiences shape later trajectories of social position.

In more constrained or despotic conditions, chasing can serve as a unidirectional, dominance-related behavior directed at subordinates^{9,43,44}. The larger groups observed in the NoSeMaze reveal a more nuanced role for chasing behavior. Chasing levels are individually stable, but the expression and meaning of chasing are context-sensitive. Chasing was neither broadly distributed nor consistently directed down the social hierarchy. Instead, it was initiated by a small subset of individuals – primarily those occupying high competition-based social ranks – and frequently occurred reciprocally within this group, suggesting intra-elite social dynamics rather than broad dominance enforcement. Rather than solely serving to impose social hierarchy, chasing appeared to function as a means through which individuals with high social rank monitor, negotiate, and maintain their relative standing within the top tier. Notably, the identity of frequent chasers remained stable over time and persisted across changing group compositions, indicating that the propensity to initiate chases reflects a consistent individual-level tendency rather than a purely situational response.

However, its coupling to social rank depended on the group's hierarchy structure. In groups with less clearly defined social hierarchies (i.e., lower transitivity), active chases aligned more strongly with social rank, suggesting that mice in less structured groups rely more on proactive signaling to clarify social rank. Indeed, the top-ranked mice in these groups exhibited relatively high levels of active chases. This context-sensitivity highlights chasing's dual role: it serves as a tool for negotiating social rank among mice at the upper end of the hierarchy, and additionally functions to establish or reinforce hierarchical clarity when social structures are ambiguous. These dynamic aspects of chasing, including its asymmetric initiator–recipient structure and proactive engagement, differ from the nature of tube competitions, which are incidental encounters. Together, tube-derived social rank and chasing describe complementary dimensions of social position, and, alongside other features such as clique formation³⁶, contribute to describe facets of a broader multidimensional social behavior. Within this complex environment, chasing emerges as a flexible behavioral propensity that is dissociable from formal social rank. Specifically, in the NoSeMaze, chasing contributes dynamically to the maintenance, negotiation, or clarification of

social hierarchy structure. These findings reveal a novel aspect of social dynamics: chasing is not merely a dominance display but a flexible context-dependent mechanism shaped by both individual disposition and group-level social structure.

In addition to behavioral strategies, physical features can also influence patterns of social interaction. Body weight is a factor implicated in social rank and aggression⁴⁵, yet studies in mice have yielded mixed results^{2,46}. In the NoSeMaze, heavier animals were less likely to lose tube competitions or to be targeted in chases, but they were not more likely to win tube competitions or initiate chases. This suggests that weight may buffer against subordination rather than confer dominance, acting as a passive protective factor that reduces the likelihood of being challenged by others.

Beyond social and physical characteristics, we also assessed reward-guided cognitive features, including impulsivity, reward sensitivity, and inhibitory control. These features were relatively stable considering the dynamics of reinforcement learning. Mice adopted distinct cognitive styles: impulsive ‘go’, cautious ‘no-go’, and flexible learners. The styles that were most effective in retrieving rewards also showed the highest stability over months. The cognitive profiles also map onto known axes of behavioral variation observed across species^{47–50}. Although social and cognitive domains revealed internally consistent profiles, they were in many cases orthogonal to each other. This is in line with meta-analyses showing weak associations between social position and cognitive characteristics in mice¹⁴. However, when viewed as composite styles, some links emerged: cautious ‘no-go’ learners were least likely to chase others, while flexible learners showed the highest levels of proactive chasing. Contrary to expectation, impulsivity did not reliably predict active chases, suggesting that social engagement in dominance negotiation may require flexible rather than impulsive behavior.

Together, these findings highlight the value of the NoSeMaze as an integrative ecological platform for dissecting social group organization and behavioral individuality. Different dimensions of social position map onto well-established factors in resilience. The approach may thus provide a suitable testbed for modeling and causally dissecting resilience mechanisms. The NoSeMaze aims to increase environmental complexity and group dynamics while retaining experimental control, enabling longitudinal high-dimensional phenotyping of individuals. At the same time, it is a controlled laboratory group-housing habitat optimized to capture a subset of the determinants of social complexity present in natural communities. The strength of this design lies in the continuous, observer-independent observation of complex behaviors in defined environmental and social contexts. Accordingly, we interpret our findings as applying to the social contexts and environmental conditions tested here. Building on this, its modular design allows for introducing additional environmental and social factors like stressors, mating behavior, or resource competition in the future to understand their respective impact in modifying social behaviors. Further, the NoSeMaze is compatible with longitudinal, individualized neuroscience, including chemogenetic⁵¹ or pharmacological circuit manipulations that can be applied in a time-locked, subject-specific manner via the lickports. The NoSeMaze can also be combined with intermittent task-based fMRI protocols, mirroring study designs currently emerging in human neuroscience^{52,53}. Together, this makes the NoSeMaze suited for unraveling the neural, genetic, and environmental contributions to individuality over time and across contexts.

Conclusion

The NoSeMaze provides a novel and powerful platform for studying the emergence of social structure and behavioral individuality in mouse societies. By enabling continuous, automated tracking of social hierarchy, chasing, and cognitive strategies within the same ecological context, the NoSeMaze revealed that individual behavioral features are remarkably stable over time, even across changing group compositions. This suggests that social position is not merely a product of immediate social environments, but rather reflects a stable individual characteristic that persists across different social contexts.

Crucially, social position is not fully described by a single behavioral dimension. We focused here on two separable dimensions: competition-based social rank and proactive chasing. Future work should integrate additional dimensions of social organization, such as affiliative bonding and higher-order network measures ³⁶, to capture additional aspects of social organization. Chasing behavior played a dual role, reflecting a stable individual propensity while also adapting to group-level structure. It was most pronounced among high-ranking individuals and aligned more closely with social rank when hierarchies were less well clarified, suggesting that chasing serves both stable and context-sensitive functions. At the same time, social and cognitive domains were largely independent, challenging the idea of unified behavioral types and supporting a multidimensional view of individuality.

Finally, the NoSeMaze exemplifies how the complexity of group behavior in semi-naturalistic conditions can be captured without sacrificing experimental control. Its modular design and compatibility with longitudinal neural interventions provides a tool for linking behavioral characteristics to neural circuit dynamics, genetic factors, and environmental influences. By combining high-dimensional phenotyping with ecological validity, the NoSeMaze opens new avenues for understanding how stable individual differences in social and cognitive function arise and how they shape, and are shaped by, the brain across time and context.

Methods

Animals

Animal strain and housing conditions

A total of 79 adult male homozygous OXTR^{fl/fl} mice (B6.129(SJL)-Oxtr^{tm1.1Wsy/J}, RRID:IMSR_JAX:008471 [DOI](#), Jackson Laboratory) backcrossed > F10 to C57BL/6J background (Charles River, Sulzfeld) were used for the experiments. Animals entered the NoSeMaze as adults (see Supplementary Table S2 [DOI](#) for ages at NoSeMaze entry across rounds) and had been continuously group-housed after weaning (3–5 mice/cage), i.e., they were not developmentally or socially naïve at study onset. Of the 79 mice, 26 were injected six weeks before the start of the experiment with an AAV expressing Cre recombinase (*rAAV1/2-CBA-Cre*) into the AON pars centralis to induce bilateral OXTR deletion (OXTR^{ΔAON}). The remaining 53 animals received an AAV expressing only dTomato (*rAAV1/2-CBA-dTomato*). Details on viral constructs, stereotactic surgery, and injection procedures are provided in Wolf et al. ³². Genotype-dependent effects are mainly addressed in a separate manuscript ³⁶. Each mouse was implanted in the neck with a subcutaneous RFID chip (Euro I.D. ID 100 trovan®) for individual identification using a microchip syringe injector under brief anesthesia with 1.5% isoflurane. During home-cage housing, mice were kept in groups of 3–5 under a 12h light/dark cycle (room temperature 24 °C, air humidity 55%). Food and water were provided ad libitum. Before first entry into the NoSeMaze, home-cage group composition was reshuffled every 3–4 days over several weeks to ensure familiarity among potential future group members (cf. Fig. 1C–D [DOI](#)). During the experiment, mice were tested across multiple NoSeMaze rounds and were systematically reassigned to different NoSeMaze groups between rounds (cf. Fig. 1D [DOI](#)), while remaining within their respective age cohort. This repeated reshuffling of NoSeMaze group composition was a central feature of the study and enabled us to test whether individual differences in social rank, chasing, and reinforcement-learning behavior generalized across changing social contexts. During the intervals between NoSeMaze rounds, mice were also reassigned to different home-cage groups. Importantly, both home-cage group assignments and NoSeMaze group compositions were randomized and blinded to prior behavioral outcomes or individual performance. In the NoSeMaze, animals had free access to food. Water was obtained from the self-paced stimulus-outcome learning task that was accessible 24/7. The number of trials determined the water intake, which was continuously monitored by the number of licks in go trials.

Ethics statement

All procedures were in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and the EU 2010/63 directive, and approved by the local animal welfare authority (Referat 35, Regierungspräsidium Karlsruhe, Karlsruhe, Germany).

Sex

In this study, we used male mice, for which social hierarchy is best studied⁹ and for which the NoSeMaze was originally designed and optimized. The cohabitation of male and female mice in the same NoSeMaze leads to mating and thus different dynamics in the social behavior. Future studies will leverage current efforts to determine optimal NoSeMaze conditions for female mouse societies to generalize the observations across sexes.

Non-invasive sensor-rich maze (NoSeMaze) and behavioral data acquisition

NoSeMaze infrastructure and habituation

The NoSeMaze contains an olfactory learning module (Fig. 1A⁴⁰ and Supplementary Fig. S1⁴⁰), which is controlled by an open-source Python software that we modified based on Erskine et al.⁴⁰. Full documentation is available at <https://github.com/KelschLAB/NoSeMaze>⁴⁰. Four identical NoSeMazes, each housing $n=9-10$ male mice, were used in parallel (for group sizes, see Supplementary Table S2⁴⁰). All mice were habituated to the NoSeMaze twice in subgroups of 6-7 animals. Each acclimatization period lasted 16-20 hours including one dark cycle, giving the animals enough time to explore the novel environment. The age ranges at which mice entered the NoSeMaze for the different rounds are shown in Supplementary Table S2⁴⁰.

Stimulus-outcome learning task at the water lick port

The task at the olfactory stimulus-outcome learning module, where the animals earned their daily water, was gradually introduced during acclimatization. Decanal (Sigma Aldrich W236217) and octanal (Sigma Aldrich W279706), both diluted in mineral oil, were used as odors and applied at a final concentration of 0.1 % with a custom-built olfactometer. Initially, a drop of water (10 μ l) was given each time the lick port was visited. Then, only one of the two odors was associated with a water drop. The odor presentation started 500 ms after the animal was recognized by its RFID chip and lasted for 2 s (cf. Fig. 2A⁴⁰). When the mice continuously lived in the NoSeMaze, in the go trials, one odor was followed by a water drop in 100% of the trials if the animal licked twice during the odor presentation. In the no-go trials, the other odor (CS-) was presented and not rewarded. If the animals licked more than once during odor presentation, they received a timeout of 6 s to initiate the next trial. The reward contingencies switched every three days (cf. Fig. 2B⁴⁰). The number of trials and licks were registered with a Darlington sensor and assigned to the RFID tag of each mouse recognized by the antenna at the lick port. Together with the licking behavior, the water intake per animal was recorded and controlled daily. All animals consumed at least 2 ml (up to 6 ml) water per day.

Tube competitions and chasing

Two tubes connected the housing area and the social arena, where the access to the water port was located. Thus, mice had to cross the tubes regularly to consume both food and water. Custom written MATLAB scripts detected events of tube competitions and tube chasing. A competition event was defined as a sequence of RFID-tag detections indicating that two mice entered the tube from different sides at the same time and one mouse pushed the other back out of the tube. While the 'dominant' mouse was detected on both sides of the tube, the 'subordinate' mouse was detected twice at its entry detector (when entering and when exiting backwards, see Fig. 3A⁴⁰). Detection patterns of putative competition events were visually inspected and included in the analysis if they met the criteria outlined above. A 'chasing event' was defined as a sequence of detections indicating the simultaneous presence of two mice in the tube, moving in the same direction and in close succession (Fig. 4A⁴⁰). Events were classified as chases if (1) the time

difference between the two animals at the detector was less than 1.5 seconds (Δ time at detector < 1.5 s), and (2) both animals crossed the tube within 2 seconds (Δ time through tube < 2 s). Animals were visually inspected daily for their well-being. Notably, we did not observe any intensive fighting or relevant injuries during the entire duration of the experiment in any of the groups.

Quantification of individual reward-seeking behavioral metrics

Metrics derived from the stimulus-outcome learning task

A detailed summary of the data across NoSeMaze rounds is provided in [Supplementary Table S1](#) [\[1\]](#). Data from 17 NoSeMaze groups were used for the analysis of the reward-seeking domains (see [Supplementary Table S2](#) [\[1\]](#)). In four NoSeMaze groups, the paradigm at the reinforcement learning port did not work properly, so that they had to be excluded from the analyses. Six animals, distributed across different groups, were additionally excluded due to insufficient RFID recognition at the lickport. The variability in licking responses was quantified using the following metrics: (1) correct hit rate, (2) correct rejection rate, (3) pre-CS licking rate before odor presentation (from +0.05 to +0.5 s relative to trial start; see [Fig. 2G](#) [\[1\]](#)), (4) peak licking during the CS+ window (from +1 s to +2.4 s relative to trial start; ‘CS+ modulation peak’, [Fig. 2G](#) [\[1\]](#)), (5) switch latencies for CS+ and CS−, computed separately for each reversal and summarized as the median across reversals, and (6) the number of trials required for animals to achieve an 80% correct hit or correct rejection rate (‘time to criterion’). All trials were used to calculate pre-CS licking rate, correct hit rate, and correct rejection rate. The pre-CS licking rate served as a measure of impulsivity, while licking inhibition during CS− trials reflected inhibitory control. Reward sensitivity was assessed using CS+ modulation peak, defined as the maximum licking rate during CS+ presentation, by concatenating the last 150 CS+ trials before each reversal to ensure that animals had reached a stable state. CS+ switch latency quantified how quickly an animal increased licking behavior in response to the newly rewarded conditioned stimulus (CS+) after a reversal. It was defined as the number of CS+ trials needed after a reversal for the animal to return to at least 70% of its pre-reversal CS+ licking rate. A shorter CS+ latency reflected faster acquisition of the new reward contingency. CS− switch latency captured how quickly an animal suppressed licking in response to the newly unrewarded conditioned stimulus (CS−) after a reversal. It was calculated as the number of CS− trials after a reversal needed for licking to drop below 50% of the pre-reversal CS+ licking rate. A shorter CS− latency indicated faster response inhibition or suppression of previously reinforced behavior. For both metrics, the pre-reversal licking rates were computed from the last 150 CS+ trials preceding each reversal. All rates were calculated from the odor presentation window (0.5–2.5 s). If no switch was detected, the switch latency was set to the total number of available post-reversal CS+ or CS− trials, which varied by animal and phase. Note that both metrics, CS+ and CS− switch latencies, reflect individualized measures, as they referred to each mouse’s own pre-reversal licking rates. Lastly, time to criterion was assessed as the number of trials following a reversal required for the animal to achieve at least 80% correct responses – hits for CS+ and correct rejections for CS− – sustained over a 10-trial moving average. If the animal did not reach this performance level during the post-reversal phase, the time to criterion was set to the total number of available post-reversal trials for the respective stimulus, indicating that the performance criterion was not met within the phase. In contrast to the switch latencies, time to criterion referred to a uniform performance threshold for all animals. For illustrative purposes, some of the data were cube root or Box-Cox transformed in cases of non-normal distribution.

Clustering of reward-seeking profiles

Data from six features were used for the clustering approach: CS+ and CS− median switch latencies, correct hit rates, correct rejection rates, CS+ modulation peak, and pre-CS licking rates. All features were standardized using min-max normalization to preserve individual ranges despite potential skew. Then, K-means clustering was applied. The optimal number of clusters ($k = 3$) was determined using the elbow method. A silhouette score of 0.50 indicated a moderately good separation between clusters. Cluster assignments were visualized in PCA-reduced space using the first three principal components, and statistically compared across features using permutation tests. To assess differences of the six reward-seeking features between clusters, we performed

unpaired permutation tests using 100,000 iterations and a median-based test statistic to account for data skewness. To address repeated measures and within-animal dependencies, a realistic swap ratio was estimated by computing the ratio of between-animal to total variability (i.e., between-animal / [within-animal + between-animal] variability). This swap ratio defined the proportion of data globally shuffled across clusters during each permutation. It was calculated per feature within each cluster and was applied globally during permutation tests to maintain realistic within-subject structure. Specifically, we used the mean of the two swap ratios from the two clusters that were compared. By incorporating both within- and between-subject variability, this approach generated a more robust null distribution. Final p-values were computed using two-tailed tests comparing observed and permuted differences. To control for multiple testing, p-values were adjusted using the Benjamini–Hochberg false discovery rate (FDR) procedure ⁵⁴.

Quantification of individual social behavioral metrics

Social rank and chasing metrics

Individual-level social rank was quantified using the David's score from tube competitions. It was calculated based on the counts of wins and losses between all pairs of animals. Specifically, for each pair of animals i and j , we calculated the number of times that i defeated j (a_{ij}) and the total number of encounters between them ($n_{ij} = a_{ij} + a_{ji}$). The directed proportion of wins was then defined as $P_{ij} = a_{ij}/n_{ij}$ whenever $n_{ij} > 0$. For each individual i , the first-order scores were calculated as

$$w_i = \sum_j P_{ij}, \quad l_i = \sum_j P_{ji}$$

representing the summed proportions of wins and losses, respectively. Pairs with no interactions ($n_{ij} = 0$) produce undefined P_{ij} values, which were excluded from the sums. To capture the strength of opponents, we additionally computed second-order terms, where

$$w_{2,i} = \sum_j P_{ij}w_j, \quad l_{2,i} = \sum_j P_{ji}l_j$$

These weighted proportions assign greater value to a win against a high-ranking opponent than against a low-ranking one, and analogously for losses ⁵⁵. The David's score for each individual was then obtained as

$$DS_i = w_i + w_{2,i} - l_i - l_{2,i}$$

Linear ranking of David's scores within a group yielded each animal's social rank, with 1 denoting the highest social rank. In this study, we determined the social rank for every animal based on the first three weeks in the NoSeMaze. Similarly, a 'chasing David's score' was calculated using the counts of active chases and times being chased between all pairs of animals in place of wins and losses. Moreover, we computed the fraction of active chases and the fraction of times being chased for each animal, defined as the number of actively initiated chases or times being chased divided by the total number of chasing events in its group. Similarly, the fractions of wins and losses in tube competitions were calculated for each individual based on the outcomes of these events.

Group-level characterization of social hierarchy structure

Social hierarchies were evaluated using metrics recommended in recent literature ³⁵, including steepness, transitivity, stability, and uncertainty-by-repeatability (Fig. 3C ³⁵, see also 'Source Data'). Triangle transitivity was used to quantify transitivity of the social hierarchy by calculating the proportion of transitive triangles (i.e., triads of animals) from all triangles ⁵⁶. A triangle is transitive when, if A dominates B and B dominates C, then A also dominates C. The triangle transitivity index therefore reflects the fraction of all triads in the group that meet this criterion. Significance of triangle transitivity was tested with a permutation test by computing 1,000 random networks (while preserving the distributions of observed null, tied and asymmetric edges in the hierarchy network). These random networks' triangle transitivity values were compared against the observed values ⁵⁶. Steepness describes the degree of dominance success over subordinate animals and is defined as the slope of a linear fit of normalized David's scores ⁵⁷. The significance

of a hierarchy's steepness was also assessed with a permutation test, in which 1,000 hierarchies with wins and losses assigned by chance are calculated and compared to the observed steepness. The stability index ranges between 0 and 1, with 0 indicating unstable hierarchies, in which the ordering reverses every other day, and 1 describing stable hierarchies without rank changes³⁵. The uncertainty-by-repeatability is calculated as the intraclass correlation coefficients for each individual after randomizing the order of the interactions. A repeatability score above 0.8 suggests a reasonably robust hierarchy³⁵. The metrics were calculated in R using the packages 'aniDom' (<https://cran.r-project.org/package=aniDom>)³⁵ and 'EloRating' (<https://cran.r-project.org/web/packages/EloRating/index.html>)⁵⁸.

Validation of social rank metrics

To validate the integrated social rank metrics and demonstrate their convergent validity in assessing social rank, David's scores derived from the integrated tube test were compared to two established dynamic social hierarchy estimation methods: the traditional Elo-rating and the randomized Elo-rating³⁵. Further, to cross-validate the integrated tube test readout, a conventional tube test was performed the day after mice were removed from the NoSeMaze and had returned to standard cages. The standard tube test was performed on ten NoSeMaze groups. In this standard setup, mice were habituated to a classical tube test apparatus consisting of two compartments connected by a tube with a central barrier. Random dyads were drawn from each group and a tube test was conducted by placing the two mice in the tube facing each other. If one mouse pushed the other out of the tube within one minute it was scored a win or loss respectively, otherwise a tie. Each animal performed five such trials to avoid exhaustion²¹. The resulting David's score from the manual tube tests were correlated significantly to those from the integrated tube test (Supplementary Fig. S4C [↗](#)). To assess the influence of entry-time differences in tube competitions, we computed two versions of dynamic Elo ratings for each animal based on its observed win/loss outcomes. The 'uncorrected' Elo rating was calculated using standard update rules, where each win/loss updated an individual's Elo score based on the pre-match Elo ratings and a group-specific K-factor⁵⁸. To evaluate and correct for potential biases due to asymmetries in tube entry time, we first quantified the influence of entry-time differences on tube competition outcomes using a linear mixed-effects model (LME). For this, we compiled a dataset combining tube competition events across all groups. Each event was included twice – once from the perspective of the winner and once from the loser. Each row contained the individual's and the opponent's identity, the pre-match Elo ratings of both animals, and their entry times. From these, the differences in both Elo rating and entry time were computed and included as predictors. Extreme values for entry-time difference were capped at a threshold of 10 seconds to minimize the influence of outliers. The LME was implemented in R (v4.4.1) using the lme4 package⁵⁹ (glmer() function, binomial family). The binary outcome variable indicated wins or losses, with scaled Elo difference and scaled entry-time difference as fixed effects, and mouse identity as a random intercept. The estimated fixed-effect coefficient for entry-time difference was then used to adjust the expected win probability in the Elo update formula. Specifically, for each match, the scaled entry-time difference was multiplied by the LME-derived coefficient and added to the log-odds of the standard Elo expectation. This resulted in a 'corrected' Elo rating, which was updated step-wise throughout the match sequences in each group. Finally, corrected scores were rescaled to fit the distributional properties (mean and SD) of the 'uncorrected' Elo scores to ensure comparability across methods.

Statistics analysis

Note that full details on the statistical tests, including *n*, degrees of freedom, exact p-values and 95% confidence intervals, are provided in the [Supplementary Material – Systematic Statistical Reporting](#) [↗](#), as well as in the **Extended Data** for the LMEs accounting for different covariates.

Basic data assessments

We quantified the number of trials at the lickport, the number of tube competitions, and the number of chasing events per group and day, and illustrated those using boxplots. The time of day at which licking and social interaction events occurred was visualized with polar plots.

Correlation and stability analyses of behavioral metrics

We used Spearman's rank correlation coefficients (ρ) to examine relationships among behavioral metrics, as well as between behavioral metrics and body weight. To account for repeated measures, we also fitted LMEs with mouse identity as a random intercept (see [Extended Data Main](#) [↗](#) and [Extended Data Supplement](#) [↗](#)). Note that the LMEs confirmed all results from the correlation analyses shown in the manuscript. To assess the relationship between social rank, chasing, and body weight, the z-scored body weight (within groups) before entering the NoSeMaze was correlated with the David's score from tube competitions, the fraction of active chases, and the fraction of times being chased. To evaluate the temporal stability of the social metrics within a NoSeMaze round, social metrics were calculated separately for the first, second, and third week, and cross-correlated across weeks. Most mice ($n = 68$) participated in at least two NoSeMaze rounds with reshuffled group members. [Supplementary Table S2](#) [↗](#) summarizes the number of rounds per mouse and missing data due to technical problems. We examined the stability of social and reward-seeking metrics by correlating values from the first and second round (Spearman's correlation, cf. [Fig. 5](#) [↗](#)). These round-1-to-round-2 correlations use one paired observation per mouse and are therefore not inflated by mice contributing >2 rounds. Additionally, we assessed the consistency of cluster assignments based on reinforcement learning metrics between rounds using a chi-square test and visualized transitions with a Sankey diagram. Importantly, to quantify stability across all available rounds, we also estimated intraclass correlation coefficients (ICCs) from variance-component LMEs, interpreting ICCs as the proportion of total variance attributable to stable between-mouse differences. The ICC treated mouse identity as a random intercept and NoSeMaze group (i.e., round-specific social group) as a random effect, with repetition included as fixed effect (i.e., stability across groups while holding repetition means constant). Formally, for a given metric

$$ICC = \frac{\sigma_{\text{mouse ID}}^2}{\sigma_{\text{mouse ID}}^2 + \sigma_{\text{context}}^2 + \sigma_{\varepsilon}^2}$$

with $\sigma_{\text{context}}^2$ containing the group as the relevant context term. The time between rounds extended over several weeks (mean: 6.8 weeks, SD: 2.8 weeks; see [Supplementary Table S2](#) [↗](#) for details). For correlation analyses, we obtained 95% confidence intervals by subject-level bootstrap (resampling mice with replacement; $B = 2,000$ draws; percentile intervals), preserving the within-mouse pairing of rounds. For ICCs, we computed 95% confidence intervals by subject-level bootstrap using the same resampling scheme and refitting the LME for each bootstrap sample.

Variance components for ICC estimation were obtained from LMEs fit by restricted maximum likelihood (REML), which yields less biased variance-component estimates and is well suited for unbalanced repeated-measures designs (i.e., different numbers of rounds per mouse). To assess potential confounding by age structure, age was decomposed into a between-group component (group-mean age) and a within-group component (each animal's deviation from its group mean) and included as covariates. ICCs were recomputed in age-adjusted models (see [Supplementary Table S4](#) [↗](#)). Finally, we performed a conservative sensitivity analysis restricted to the first two sessions per mouse and to mice with observations in both sessions ("balanced first2"), to additionally account for unbalanced participation structure.

Comparative analysis of chasing and social rank

Distributions of the fraction of active chases and the fraction of times being chased, pooled across all groups, were assessed using histograms and empirical cumulative distribution plots ([Supplementary Fig. S6A](#) [↗](#)). Distribution shapes were compared via Kolmogorov-Smirnov tests. A similar analysis was applied to the distributions of wins and losses in tube competitions ([Supplementary Fig. S6B](#) [↗](#)). To further examine differences between the fractions of active chases and of times being chased, we computed cumulative sums over the sorted data within each group, visualized them using shaded error bar plots, and tested group-level differences between the two metrics using unpaired permutation tests (cf. [Fig. 4C](#) [↗](#), 10,000 permutations). Lastly, to explore whether win-loss relationships differed between chasing and tube competition events, we

combined data from both modalities and modeled the fraction of ‘wins’ – defined as successful tube competitions or active chases – using a LME (Supplementary Fig. S13 [↗](#)). Active chases were treated as ‘wins’ in the context of chasing, being chased as ‘losses’. The model included the fraction of ‘losses’, the event type (chasing or competition), and their interaction as fixed effects, with mouse identity as a random intercept. Model assumptions were tested using residual diagnostics and normality tests (Lilliefors, Anderson–Darling, and Shapiro–Wilk). Due to assumption violations, we performed cluster-resampled bootstrapping (10,000 iterations, resampled by subject) to estimate robust confidence intervals for fixed effects. The interaction term captured differences in the win–loss relationship between tube competitions and chasing events.

Relationship between social rank and chasing

First, to assess how tube competition metrics and chasing relate at the individual level, we conducted correlation analyses using Spearman’s rank correlation coefficients (cf. Fig. 6A [↗](#) and Fig. 6C [↗](#)). To account for repeated measures across mice, we also fitted LMEs with mouse identity as a random intercept that confirmed all results (see Extended Data Main [↗](#)). To further examine and illustrate structural patterns within social interactions, we processed win–loss match matrices from tube competition and chasing events for each NoSeMaze group. These match matrices were sorted by the David’s score (competition-based) or by the fraction of active chases (chasing-based) and normalized by dividing all entries by the maximum value within that matrix, thus setting the strongest observed interaction to 1. This procedure was applied at the group level to preserve the relative structure of interactions while allowing comparison across groups with differing activity levels. Normalized matrices were then averaged across all groups and visualized as heatmaps, with rows and columns representing individuals ordered by their social rank or active chases (cf. Fig. 6D–F [↗](#)). Color intensity indicated the relative frequency of interactions between each pair of individuals. To quantify asymmetries in interaction patterns, we modeled quadrant-level differences in the frequencies of normalized dyadic interactions using an LME with a beta distribution and logit link function. Proportions of dyadic events were bounded away from 0 and 1 by a small constant ($\epsilon = 0.0001$) to meet the assumptions of the beta distribution. The model included quadrant as a fixed effect and group identity as a random intercept. Model fitting was performed using the glmmTMB package in R (<https://cran.r-project.org/web/packages/glmmTMB/index.html> [↗](#))⁶⁰. Estimated marginal means were computed for each quadrant using the emmeans package (<https://cran.r-project.org/web/packages/emmeans/index.html> [↗](#))⁶¹, and back-transformed to the original scale (0–1) for interpretability. Pairwise post hoc comparisons between quadrants were conducted on the response scale, with Tukey-adjusted p-values to control for multiple comparisons.

Group-level analysis of social rank–chasing relationships

To evaluate the relationship between social rank and chasing at the group level, we computed Spearman’s correlation coefficients between the David’s scores from tube competitions and the fraction of active chases within each NoSeMaze group. This yielded one correlation coefficient per group ($n = 18$). We then tested whether these group-level coefficients were associated with group-level transitivity scores, using a Spearman’s rank correlation across groups (cf. Fig. 6B [↗](#)). This analysis assessed whether the strength of the within-group association between social rank and active chases varied systematically with the clarity of the social hierarchy. To further characterize this relationship, we also correlated group transitivity with the fraction of active chases performed by the mouse with the highest social rank (cf. Supplementary Fig. S11 [↗](#)), reflecting the extent to which dominant individuals engaged in proactive aggression as a function of group hierarchy structure.

Relationships between social and reward-seeking features

Differences in the David’s score from tube competitions, the fraction of active chases and the fraction of times being chased between the three learning clusters were similarly assessed as for the six stimulus–outcome learning features (cf. **Clustering of reward-seeking profiles**), using the same median-based permutation method accounting for between- and within-animal variability. Again, to control for multiple comparisons, p-values were adjusted using the Benjamini–Hochberg

FDR procedure⁵⁴. Then, association between the social metrics (David's score and active chases) and the reward-seeking metrics (correct hit and rejection rate, pre-CS licking rate, CS+ modulation peak, switch latencies for CS+ and CS- from reversal 1 to 4 and for the median) were assessed with similar Spearman's correlation analyses and LMEs, as described above (cf. **Relationship between social hierarchy and chasing**). Lastly, we performed principal component analysis (PCA) on z-scored data to examine the underlying structure of social and cognitive features more profoundly. Skewed variables were first log- or square root-transformed based on skewness direction and magnitude. PCA was conducted combining social (David's score, active chases and being chased) and reward-seeking features. Loadings were interpreted to assess the contribution of each metric to the top three components. To evaluate domain specificity, permutation tests (n = 10,000) were used to test whether social and cognitive metrics differed significantly in their contributions to each principal component.

Genotype effects

Group differences in social and reinforcement-learning metrics (cf. [Supplementary Fig. S9](#) and [Supplementary Fig. S10](#)) between OXTR^{ΔAON} and control animals were assessed using permutation tests (10,000 iterations) comparing group medians. To account for repeated measures, group labels were shuffled at the level of the animal. Week-wise comparisons for the social metrics were FDR-corrected for multiple comparisons. To test whether the observed behavioral associations were influenced by the OXTR^{ΔAON} genotype, genotype (OXTR^{ΔAON} vs. control) was included as a covariate in LMEs and correlation analyses (see [Extended Data Main](#) and [Extended Data Supplement](#)). Specifically, LMEs were extended with the covariate as additional predictor, and Spearman's partial correlation analyses were performed including the covariate as factor of no interest. In each analysis, we tested whether the primary relationships of interest remained significant when accounting for genotype. All analyses remained statistically significant after inclusion of the genotype covariate, indicating that the observed associations were robust across genotypes. For instance, in the relationship between the David's score from tube competitions and the fraction of active chases, the inclusion of genotype in the LME did not alter the significance of the main association.

Data availability

The full documentation of the NoSeMaze, including hardware information and the open-source Python-based software package, is available on GitHub:

<https://github.com/KelschLAB/NoSeMaze>. Source data files for each main figure and each supplemental figures as well as the extended data are provided with the manuscript. All analysis code and data matrices used to generate the results are available on GitHub:

<https://github.com/KelschLAB/NoSeMaze-social-status>.

Acknowledgements

We thank Michael Bram and Jan Ringkamp for their help in engineering the NoSeMaze, and Dr. Carla Filosa for her initial help in developing the behavioral analyses. We also thank Michal Adveev and Laura Haenschke for help with preparing the competition videos.

We acknowledge the use of ChatGPT (OpenAI, GPT-4, accessed 2025) for assistance with language editing, improving readability, and annotating code. All outputs were reviewed, validated, and verified by the authors, who take full responsibility for the content of this manuscript. The work was funded by BMBF 3R consortium grants 'NoSeMaze1' (161L0277A) and 'NoSeMaze2' (16LW0333K) to W.K., Leibniz Association program grant 'Learning resilience' (K430/2021) to W.K., Boehringer Ingelheim Foundation grant 'Complex Systems' to W.K., BMBF CRCNS grant 'Oxystate' (01GQ1708) to W.K., DFG CRC 379 Project C03 to W.K., and the DFG Clinician Scientist Program 'Interfaces and Interventions in Complex Chronic Conditions' (EB187/8-1) to J.R.

Additional files

[Supplementary Material](#) 

[Supplementary Material - Systematic Statistical Reporting](#) 

[Supplementary Video 1.](#) 

[Supplementary Video 2.](#) 

[Supplementary Video 3.](#) 

[Extended data main](#) 

[Extended data supplement](#) 

[Source data main](#) 

[Source data supplement](#) 

Additional information

Funding

Funder	Grant reference number	Author
Bundesministerium für Forschung, Technologie und Raumfahrt (BMBF)	161L0277A	Wolfgang Kelsch
Bundesministerium für Forschung, Technologie und Raumfahrt (BMBF)	16LW0333K	Wolfgang Kelsch
IPF Leibniz-Gemeinschaft (LG)	K430/2021	Wolfgang Kelsch
Boehringer Ingelheim Stiftung (Boehringer Ingelheim Foundation)	Complex Systems	Wolfgang Kelsch
Deutsche Forschungsgemeinschaft (DFG)	CRC 379 Project C03	Wolfgang Kelsch
Deutsche Forschungsgemeinschaft (DFG)	EB187/8-1	Jonathan Reinwald

Author ORCID iDs

Jonathan R Reinwald:  <https://orcid.org/0000-0001-9508-3942>

Wolfgang Kelsch:  <https://orcid.org/0000-0002-3470-8125>

References

1. **Chase ID** (1980) Social process and hierarchy formation in small groups: a comparative perspective. *American Sociological Review* 905-924 <https://doi.org/10.2307/2094909>
2. **Williamson CM**, Lee W, Curley JP (2016) Temporal dynamics of social hierarchy formation and maintenance in male mice. *Animal Behaviour* **115**:259-272 <https://doi.org/10.1016/j.anbehav.2016.03.004>
3. **Tibbetts EA**, Pardo-Sanchez J, Weise C (2022) The establishment and maintenance of dominance hierarchies. *Philosophical Transactions of the Royal Society B* **377**:20200450 <https://doi.org/10.1098/rstb.2020.0450> | [PubMed](#)
4. **Beery AK**, Kaufer D (2015) Stress, social behavior, and resilience: insights from rodents. *Neurobiology of stress* **1**:116-127 <https://doi.org/10.1016/j.ynstr.2014.10.004> | [PubMed](#)
5. **LeClair KB**, Chan KL, Kaster MP, Parise LF, Burnett CJ, Russo SJ (2021) Individual history of winning and hierarchy landscape influence stress susceptibility in mice. *eLife* **10**:e71401 <https://doi.org/10.7554/eLife.71401> | [PubMed](#)
6. **Sapolsky RM** (2005) The influence of social hierarchy on primate health. *Science* **308**:648-652 <https://doi.org/10.1126/science.1106477> | [PubMed](#)

7. Cum M, Santiago Pérez JA, Iwata RL, Lopez N, Higgs A, Li A, Ye C, Wangia E, Wright ES, García Restrepo C, *et al.* (2024) A Multiparadigm Approach to Characterize Dominance Behaviors in CD1 and C57BL6 Male Mice. *eNeuro* **11**:0342-0324 <https://doi.org/10.1523/eneuro.0342-24.2024> | PubMed
8. Weber EM, Zidar J, Ewaldsson B, Askevik K, Udén E, Svensk E, Törnqvist E (2022) Aggression in Group-Housed Male Mice: A Systematic Review. *Animals* **13** <https://doi.org/10.3390/ani13010143> | PubMed
9. Fulenwider HD, Caruso MA, Ryabinin AE (2022) Manifestations of domination: Assessments of social dominance in rodents. *Genes Brain Behav* **21**:e12731 <https://doi.org/10.1111/gbb.12731> | PubMed
10. Lee W, Khan A, Curley JP (2017) Major urinary protein levels are associated with social status and context in mouse social hierarchies. *Proceedings of the Royal Society B: Biological Sciences* **284**:20171570 <https://doi.org/10.1098/rspb.2017.1570> | PubMed
11. Winiarski M, Madecka A, Yadav A, Borowska J, Wołyniak MR, Jędrzejewska-Szmek J, Kondrakiewicz L, Mankiewicz L, Chaturvedi M, Wójcik DK, *et al.* (2025) Information sharing within a social network is key to behavioral flexibility—Lessons from mice tested under seminaturalistic conditions. *Science Advances* **11**:eadm7255 <https://doi.org/10.1126/sciadv.adm7255> | PubMed
12. Shemesh Y, Sztainberg Y, Forkosh O, Shlapobersky T, Chen A, Schneidman E (2013) High-order social interactions in groups of mice. *eLife* **2**:e00759 <https://doi.org/10.7554/eLife.00759> | PubMed
13. Forkosh O, Karamihalev S, Roeh S, Alon U, Anpilov S, Touma C, Nussbaumer M, Flachskamm C, Kaplick PM, Shemesh Y, *et al.* (2019) Identity domains capture individual differences from across the behavioral repertoire. *Nature neuroscience* **22**:2023-2028 <https://doi.org/10.1038/s41593-019-0516-y> | PubMed
14. Varholick JA, Bailoo JD, Jenkins A, Voelkl B, Würbel H (2021) A Systematic Review and Meta-Analysis of the Relationship Between Social Dominance Status and Common Behavioral Phenotypes in Male Laboratory Mice. *Frontiers in Behavioral Neuroscience* **14** <https://doi.org/10.3389/fnbeh.2020.624036> | PubMed
15. Jupp B, Murray JE, Jordan ER, Xia J, Fluharty M, Shrestha S, Robbins TW, Dalley JW (2016) Social dominance in rats: effects on cocaine self-administration, novelty reactivity and dopamine receptor binding and content in the striatum. *Psychopharmacology* **233**:579-589 <https://doi.org/10.1007/s00213-015-4122-8> | PubMed
16. Sánchez-Salvador L, Prados-Pardo Á, Martín-González E, Olmedo-Córdoba M, Mora S, Moreno M (2021) The role of social stress in the development of inhibitory control deficit: a systematic review in preclinical models. *International Journal of Environmental Research and Public Health* **18**:4953 <https://doi.org/10.3390/ijerph18094953> | PubMed
17. Kahnau P, Mieske P, Wilzopolski J, Kalliokoski O, Mandillo S, Hölter SM, Voikar V, Amfim A, Badurek S, Bartelik A, *et al.* (2023) A systematic review of the development and application of home cage monitoring in laboratory mice and rats. *BMC biology* **21**:256 <https://doi.org/10.1186/s12915-023-01751-7> | PubMed
18. Crabbe JC, Wahlsten D, Dudek BC (1999) Genetics of mouse behavior: interactions with laboratory environment. *Science* **284**:1670-1672 <https://doi.org/10.1126/science.284.5420.1670> | PubMed
19. Voikar V, Gaburro S (2020) Three Pillars of Automated Home-Cage Phenotyping of Mice: Novel Findings, Refinement, and Reproducibility Based on Literature and Experience. *Front Behav Neurosci* **14** <https://doi.org/10.3389/fnbeh.2020.575434> | PubMed
20. Chesler EJ, Wilson SG, Lariviere WR, Rodriguez-Zas SL, Mogil JS (2002) Influences of laboratory environment on behavior. *Nature neuroscience* **5**:1101-1102 <https://doi.org/10.1038/nn1102-1101> | PubMed
21. Fan Z, Zhu H, Zhou T, Wang S, Wu Y, Hu H (2019) Using the tube test to measure social hierarchy in mice. *Nature Protocols* **14**:819-831 <https://doi.org/10.1038/s41596-018-0116-4> | PubMed
22. Mertens S, Vogt MA, Gass P, Palme R, Hiebl B, Chourbaji S (2019) Effect of three different forms of handling on the variation of aggression-associated parameters in individually and group-housed male C57BL/6NCrI mice. *PloS one* **14**:e0215367 <https://doi.org/10.1371/journal.pone.0215367> | PubMed

23. **Gouveia K, Hurst JL** (2017) Optimising reliability of mouse performance in behavioural testing: the major role of non-aversive handling. *Scientific reports* **7** <https://doi.org/10.1038/srep44999> | [PubMed](#)
24. **Weissbrod A, Shapiro A, Vasserman G, Edry L, Dayan M, Yitzhaky A, Hertzberg L, Feinerman O, Kimchi T** (2013) Automated long-term tracking and social behavioural phenotyping of animal colonies within a semi-natural environment. *Nature communications* **4**:2018 <https://doi.org/10.1038/ncomms3018> | [PubMed](#)
25. **Sofer Y, Zilkha N, Gimpel E, Wagner S, Chuartzman SG, Kimchi T** (2024) Sexually dimorphic oxytocin circuits drive intragroup social conflict and aggression in wild house mice. *Nature neuroscience* **27**:1565-1573 <https://doi.org/10.1038/s41593-024-01685-5> | [PubMed](#)
26. **Barabas AJ, Lucas JR, Erasmus MA, Cheng HW, Gaskill BN** (2021) Who's the Boss? Assessing Convergent Validity of Aggression Based Dominance Measures in Male Laboratory Mice, *Mus Musculus*. *Front Vet Sci* **8** <https://doi.org/10.3389/fvets.2021.695948> | [PubMed](#)
27. **Williamson CM, Lee W, Romeo RD, Curley JP** (2017) Social context-dependent relationships between mouse dominance rank and plasma hormone levels. *Physiology & behavior* **171**:110-119 <https://doi.org/10.1016/j.physbeh.2016.12.038> | [PubMed](#)
28. **Chase ID, Tovey C, Spangler-Martin D, Manfredonia M** (2002) Individual differences versus social dynamics in the formation of animal dominance hierarchies. *Proceedings of the National Academy of Sciences* **99**:5744-5749 <https://doi.org/10.1073/pnas.082104199> | [PubMed](#)
29. **Bengston SE, Jandt JM** (2014) The development of collective personality: the ontogenetic drivers of behavioral variation across groups. *Frontiers in Ecology and Evolution* **2**:81 <https://doi.org/10.3389/fevo.2014.00081>
30. **Freund J, Brandmaier AM, Lewejohann L, Kirste I, Kritzler M, Krüger A, Sachser N, Lindenberger U, Kempermann G** (2013) Emergence of individuality in genetically identical mice. *Science* **340**:756-759 <https://doi.org/10.1126/science.1235294> | [PubMed](#)
31. **Cantor M, Maldonado-Chaparro AA, Beck KB, Brandl HB, Carter GG, He P, Hillemann F, Klarevas-Irby JA, Ogino M, Papageorgiou D, et al.** (2021) The importance of individual-to-society feedbacks in animal ecology and evolution. *Journal of Animal Ecology* **90**:27-44 <https://doi.org/10.1111/1365-2656.13336> | [PubMed](#)
32. **Oettl L.-L, Ravi N, Schneider M, Scheller Max F, Schneider P, Mitre M, da Silva Gouveia M, Froemke Robert C, Chao Moses V, Young WS, et al.** (2016) Oxytocin Enhances Social Recognition by Modulating Cortical Control of Early Olfactory Processing. *Neuron* **90**:609-621 <https://doi.org/10.1016/j.neuron.2016.03.033> | [PubMed](#)
33. **Wolf D, Hartig R, Zhuo Y, Scheller MF, Articus M, Moor M, Grinevich V, Linster C, Russo E, Weber-Fahr W, et al.** (2024) Oxytocin induces the formation of distinctive cortical representations and cognitions biased toward familiar mice. *Nature Communications* **15** <https://doi.org/10.1038/s41467-024-50113-6> | [PubMed](#)
34. **Bari A, Robbins TW** (2013) Inhibition and impulsivity: behavioral and neural basis of response control. *Progress in neurobiology* **108**:44-79 <https://doi.org/10.1016/j.pneurobio.2013.06.005> | [PubMed](#)
35. **Sánchez-Tójar A, Schroeder J, Farine DR** (2018) A practical guide for inferring reliable dominance hierarchies and estimating their uncertainty. *Journal of Animal Ecology* **87**:594-608 <https://doi.org/10.1111/1365-2656.12776> | [PubMed](#)
36. **Nelias C, Ghanayem S, Wolf D, Moor M, Scheller MF, Grinevich V, Reinwald JR, Kelsch W** (2025) Stable clique membership in mouse societies requires oxytocin-enabled social sensory states. *bioRxiv* <https://doi.org/10.1101/2025.08.26.672298>
37. **Bernstein IS** (1980) Dominance: A theoretical perspective for ethologists. In: Omark DR, Strayer FF, Freedman DG (Eds). *Dominance relations: An ethological view of human conflict and social interaction* New York: Garland Press. pp. 71-84

38. **Torquet N**, Marti F, Campart C, Tolu S, Nguyen C, Oberto V, Benallaoua M, Naudé J, Didiennie S, Debray N, *et al.* (2018) Social interactions impact on the dopaminergic system and drive individuality. *Nature communications* **9**:3081 <https://doi.org/10.1038/s41467-018-05526-5> | PubMed
39. **Castelhano-Carlos M**, Costa PS, Russig H, Sousa N (2014) PhenoWorld: a new paradigm to screen rodent behavior. *Translational psychiatry* **4**:e399-e399 <https://doi.org/10.1038/tp.2014.40> | PubMed
40. **Ersikine A**, Bus T, Herb JT, Schaefer AT (2019) AutoMouse: High throughput operant conditioning reveals progressive impairment with graded olfactory bulb lesions. *PloS one* **14**:e0211571 <https://doi.org/10.1371/journal.pone.0211571> | PubMed
41. **Kiryk A**, Janusz A, Zglinicki B, Turkes E, Knapska E, Konopka W, Lipp H.-P, Kaczmarek L (2020) IntelliCage as a tool for measuring mouse behavior—20 years perspective. *Behavioural brain research* **388**:112620 <https://doi.org/10.1016/j.bbr.2020.112620> | PubMed
42. **Lindzey G**, Winston H, Manosevitz M (1961) Social dominance in inbred mouse strains. *Nature* <https://doi.org/10.1038/191474a0> | PubMed
43. **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**:693-697 <https://doi.org/10.1126/science.1209951> | PubMed
44. **Miczek KA**, Maxson SC, Fish EW, Faccidomo S (2001) Aggressive behavioral phenotypes in mice. *Behavioural brain research* **125**:167-181 [https://doi.org/10.1016/s0166-4328\(01\)00298-4](https://doi.org/10.1016/s0166-4328(01)00298-4) | PubMed
45. **Clutton-Brock TH**, Huchard E (2013) Social competition and selection in males and females. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**:20130074 <https://doi.org/10.1098/rstb.2013.0074> | PubMed
46. **Van Loo PL**, Mol JA, Koolhaas JM, Van Zutphen BF, Baumans V (2001) Modulation of aggression in male mice: influence of group size and cage size. *Physiology & behavior* **72**:675-683 [https://doi.org/10.1016/s0031-9384\(01\)00425-5](https://doi.org/10.1016/s0031-9384(01)00425-5) | PubMed
47. **Cervantes MC**, Laughlin RE, Jentsch JD (2013) Cocaine self-administration behavior in inbred mouse lines segregating different capacities for inhibitory control. *Psychopharmacology* **229**:515-525 <https://doi.org/10.1007/s00213-013-3135-4> | PubMed
48. **Dongelmans M**, Durand-de Cuttoli R, Nguyen C, Come M, Duranté EK, Lemoine D, Brito R, Ahmed Yahia T, Mondoloni S, Didiennie S, *et al.* (2021) Chronic nicotine increases midbrain dopamine neuron activity and biases individual strategies towards reduced exploration in mice. *Nature Communications* **12**:6945 <https://doi.org/10.1038/s41467-021-27268-7> | PubMed
49. **Huang Y**, Luan S, Wu B, Li Y, Wu J, Chen W, Hertwig R (2024) Impulsivity is a stable, measurable, and predictive psychological trait. *Proceedings of the National Academy of Sciences* **121**:e2321758121 <https://doi.org/10.1073/pnas.2321758121> | PubMed
50. **Pittaras E**, Callebert J, Chennaoui M, Rabat A, Granon S (2016) Individual behavioral and neurochemical markers of unadapted decision-making processes in healthy inbred mice. *Brain Structure and Function* **221**:4615-4629 <https://doi.org/10.1007/s00429-016-1192-2> | PubMed
51. **Magnus CJ**, Lee PH, Bonaventura J, Zemla R, Gomez JL, Ramirez MH, Hu X, Galvan A, Basu J, Michaelides M, *et al.* (2019) Ultrapotent chemogenetics for research and potential clinical applications. *Science* **364**:eaav5282 <https://doi.org/10.1126/science.aav5282> | PubMed
52. **Tost H**, Reichert M, Braun U, Reinhard I, Peters R, Lautenbach S, Hoell A, Schwarz E, Ebner-Priemer U, Zipf A, *et al.* (2019) Neural correlates of individual differences in affective benefit of real-life urban green space exposure. *Nature neuroscience* **22**:1389-1393 <https://doi.org/10.1038/s41593-019-0451-y> | PubMed
53. **Reichert M**, Gan G, Renz M, Braun U, Brüßler S, Timm I, Ma R, Berhe O, Benedyk A, Moldavski A, *et al.* (2021) Ambulatory assessment for precision psychiatry: Foundations, current developments and future avenues. *Experimental Neurology* **345**:113807 <https://doi.org/10.1016/j.expneurol.2021.113807> | PubMed

54. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**:289-300 <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
55. David HA (1987) Ranking from unbalanced paired-comparison data. *Biometrika* **74**:432-436 <https://doi.org/10.1093/biomet/74.2.432>
56. Shizuka D, McDonald DB (2012) A social network perspective on measurements of dominance hierarchies. *Animal Behaviour* **83**:925-934 <https://doi.org/10.1016/j.anbehav.2012.01.011>
57. de Vries H, Stevens JMG, Vervaecke H (2006) Measuring and testing the steepness of dominance hierarchies. *Animal Behaviour* **71**:585-592 <https://doi.org/10.1016/j.anbehav.2005.05.015>
58. Neumann C, Duboscq J, Dubuc C, Ginting A, Irwan AM, Agil M, Widdig A, Engelhardt A (2011) Assessing dominance hierarchies: validation and advantages of progressive evaluation with Elo-rating. *Animal Behaviour* **82**:911-921 <https://doi.org/10.1016/j.anbehav.2011.07.016>
59. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of statistical software* **67**:1-48 <https://doi.org/10.18637/jss.v067.i01>
60. Brooks ME, Kristensen K, Van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Mächler M, Bolker BM (2017) glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R package.
61. Lenth R (2023) emmeans: Estimated Marginal Means, aka Least-Squares Means_. R package. version: v 1.8.5

Peer reviews

Reviewer #2 (Public review):

Summary:

This manuscript presents the "NoSeMaze", a novel automated platform for studying social behavior and cognitive performance in group-housed male mice. The authors report that mice form robust, transitive dominance hierarchies in this environment and that individual social rank remains largely stable across multiple group compositions. They further demonstrate that social dominance and aggressive behaviors, like chasing, are partially dissociable and that dominance traits are independent of non-social cognitive performance. The study includes a genetic manipulation of oxytocin receptor expression in the anterior olfactory nucleus, which showed only transient effects on social rank.

Strengths:

(1) Innovative Methodology:

The NoSeMaze platform is a technically elegant and conceptually well-integrated system that enables fully automated, long-term monitoring of both social and cognitive behaviors in large groups of group-housed mice. It combines tube-test-like dominance contests, voluntary chase-escape interactions, and an embedded operant olfactory discrimination task within a single, ethologically relevant environment. This modular design allows for high-throughput, minimally invasive behavioral assessment without the need for repeated handling or artificial isolation.

(2) Experimental Scale and Rigor:

The study includes 79 male mice and over 4,000 mouse-days of observation across multiple group reshufflings. The use of RFID-based identification, automated data logging, and longitudinal design enables robust quantification of individual trait stability and group-level social structure.

(3) Multidimensional Behavioral Profiling:

The integration of social (tube dominance, proactive chasing), physical (body weight), and cognitive (olfactory learning task) measures offers a rich, multi-dimensional profile of each individual mouse. The authors' finding that social dominance traits and non-social cognitive performance are largely uncorrelated reinforces emerging models of orthogonal behavioral trait axes or "animal personalities".

(4) Clarity and Data Analysis:

The analytical framework is well-suited to the study's complexity, with appropriate use of dominance metrics, mixed-effects models, and permutation tests. The analyses are clearly explained, statistically rigorous, and supported by transparent supplementary materials.

Weaknesses:

(1) Scope Limitations (Sex):

The study is limited to male mice, which represents a common but problematic bias.

(2) Ambiguity of Dominance as a Construct:

While the study robustly quantifies social rank and hierarchy structure, the broader functional meaning of "dominance" remains unclear.

<https://doi.org/10.7554/eLife.109354.2.sa1>

Author response:

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

Public Reviews:

Reviewer #1 (Public review):

Summary:

The goal of the study was to address the question of the degree to which social position in a group is a stable trait that persists across conditions. Reinwald et al. use a custom-built cage system with automated tracking and continuous testing for social dominance that does not require intervention by the experimenter. Remixing of individuals from different groups revealed that social position was rather stable and not really predictable from other measures that were taken. The authors conclude that social position is multifaceted but dependent on characteristics like personality traits.

Strengths:

(1) Reductionistic, highly controlled setting that allows for the control of many confounding variables.

(2) Very interesting and important question.

(3) Confirms the emergence of inter-individual behavior-driven differences in inbred mice in a shared environment.

(4) Innovative paradigm and experimental setup.

(5) Fresh perspective on an old question that makes the best use of modern technology.

(6) Intelligent use of behavioral and cognitive covariables to generate a non-social context.

(7) Bold and almost provocative conclusion, inviting discussion and further elaboration.

We thank Reviewer #1 for this constructive and balanced evaluation of our work, and for highlighting both the conceptual importance of the question and the strengths of our automated, highly controlled approach.

Weaknesses:

(1) Reductionistic, highly controlled setting that blends out much of the complexity of social behavior in a community.

Our goal in developing the NoSeMaze was to provide an enriched yet standardized environment that allows animals to interact freely in a complex setting where arena geometry and access contingencies are held constant across groups. As the Reviewer also notes as a strength, this highly controlled setting with minimal experimenter interference enables us to minimize confounds and provides reproducibility across rounds and groups. We now clarify this explicitly and frame the design as a trade-off between ecological complexity and experimental controllability.

The NoSeMaze is designed as an open system that can incorporate additional configurations. In this first study with the system, we intentionally used a design that focused on single-sex groups without mating, no intruders or external threats, stable environmental conditions, and adult mice. This allowed us to establish social behavior under one defined condition. From here, future studies can add certain levels of ecological and social complexity to progressively understand their impact on specific behaviors and group dynamics.

Accordingly, we have revised the manuscript to clarify the scope and the role of social and environmental conditions to the here observed phenomena. We also describe how future studies with systematically modified conditions can be used to understand how they change behaviors. We further replaced potentially over-broad terms (e.g., “naturalistic,” “real-world”) with more precise wording throughout.

We modified the following sections:

Abstract

We removed “... within naturalistic mouse groups.” (ll. 52-54) and changed the sentence to “The approach thus enables longitudinal modeling of individuality and social position as key resilience factors.”

We changed “... in naturalistic groups.” (l. 38) to “... in larger male mouse groups.” and “... from naturalistic tube competitions ...” (ll. 41-42) to “... from incidental competitions in the integrated tube tests ...”. We also deleted “naturalistic” in l. 50.

Discussion (ll. 691-701)

“...The NoSeMaze aims to increase environmental complexity and group dynamics while retaining experimental control, enabling longitudinal high-dimensional phenotyping of individuals. At the same time, it is a controlled laboratory group-housing habitat optimized to capture a subset of the determinants of social complexity present in natural communities. The strength of this design lies in the continuous, observer-independent observation of complex behaviors in defined environmental and social contexts. Accordingly, we interpret our findings as applying to the social contexts and environmental conditions tested here. Building on this, its modular design allows for introducing additional environmental and

social factors like stressors, mating behavior, or resource competition in the future to understand their respective impact in modifying social behaviors.”

Conclusion

We changed “... complexity of real-world behavior ...” (ll. 729-730) to “... complexity of group behavior in semi-naturalistic conditions ...”

(2) The motivation to enter the test tube is not "trait" (or at least not solely a trait) but the basic need to reach food and water; chasing behavior would be less dependent on this stimulus.

Tube traversals may reflect different motivations, including routine movement between compartments to access food, the water lickport, the open arena, or the housing area. Nevertheless, we do not interpret tube-entry motivation itself as a ‘trait’. Importantly, our hierarchy readout does not quantify which animals enter the tubes, nor is it confounded by tube-entry frequency itself (cf. ll. 527-532). Rather, it captures the consistent outcomes of incidental dyadic competitions once two animals meet in the tube (push vs. retreat), aggregated across many interactions. Thus, the stable signal we report lies in repeatable competition outcomes, not in traversal propensity. Consistent with this interpretation, the overall number of tube competition events was not associated with social rank, arguing against systematic competition avoidance by low-ranking animals. In contrast, chasing is a voluntarily initiated, asymmetric interaction between an initiator and a recipient and therefore adds a social-action component beyond incidental access-linked encounters, providing a complementary readout of social behavior.

We also noted that the term ‘trait’ may not be the optimal description in this context and replaced it throughout the ms. with more precise wording, including “internalized social rank” and “propensity to chase”.

Results

We added the following paragraph (ll. 524-532):

“Tube crossings in the NoSeMaze are motivated by the intent to eat, drink, sleep, or socialize. Accordingly, competitions within the tube arise by chance when two animals enter from opposite sides at the same time. Social rank therefore captures the consistent outcomes of repeated incidental competitions (push versus retreat). This measure is not confounded by differences in tube engagement, as social rank was neither associated with participation in tube competitions ($p = 0.11$, $p = 0.139$; Fig. 6C, Supplementary Fig. S12C) nor with the overall number of tube detections when controlling for chasing (Spearman’s partial correlation between detection count and z-scored David’s score, corrected for the fraction of active chases: $\rho = 0.022$, $p = 0.758$).”

Discussion

We changed the following paragraphs:

Line 608-611

“Importantly, participation frequency and differences in tube entry time did not confound the resulting social ranks, underscoring the robustness of the automated incidental rank assessment.”

Lines 653-662

“These dynamic aspects of chasing, including its asymmetric initiator–recipient structure and proactive engagement, differ from the nature of tube competitions, which are incidental

encounters. Together, tube-derived social rank and chasing describe complementary dimensions of social position, and, alongside other features such as clique formation [36], contribute to describe facets of a broader multidimensional social behavior. Within this complex environment, chasing emerges as a flexible behavioral propensity that is dissociable from formal social rank. Specifically, in the NoSeMaze, chasing contributes dynamically to the maintenance, negotiation, or clarification of social hierarchy structure.”

(3) Dominance is only one aspect of sociality, social structure is reduced to rank. The information that might lie in the chasing behavior is not optimally used to explain social behavior beyond the rank measure.

In this manuscript, we focused on the relationship between social rank derived from incidental tube competitions and chasing. We agree that social rank derived from tube competitions captures only one aspect of social structure. We have now clarified this point more explicitly in the revised manuscript. Here, our specific aim was to determine how chasing relates to competition-derived social rank and whether it provides information beyond rank in these mouse groups.

We also refer to another manuscript dedicated to additional aspects of social structure captured in the video data, including approach and interaction behavior as well as social clique formation. There, these measures are again examined in relation to chasing and social rank. We also plan future studies leveraging chasing in the NoSeMaze as a readout for neurophysiological investigations.

In the present study, social rank is based on dominance and subordination in incidental competitions in the integrated tube test. We treat chasing as a distinct, volitional social dimension rather than redundant “rank information”. Importantly, our chasing analyses add beyond social rank information in three ways: (1) structural asymmetry (initiator vs recipient roles) not captured by symmetric social rank measures; (2) elite-centric reciprocal dynamics rather than broad top-down enforcement; and (3) context dependence, with social rank–chasing coupling strengthening when group transitivity is lower. We revised the Discussion/Conclusion accordingly and additionally note that other aspects of social organization (e.g., affiliative bonding and higher-order network structure such as clique/rich-club organization; Nelias et al., 2025) require complementary measures (e.g., video-derived interaction networks) and are not the scope of the present study.

To avoid ambiguity, we also added the following operational definitions in the Introduction and Results:

Introduction

Lines 90-94

“In this study, we use social hierarchy to denote the group-level structure inferred from incidental competitions in the integrated tube tests, and social rank for an individual’s level within that hierarchy. Social position serves as an umbrella term for social rank and chasing behaviors.”

Results

Line 306-308

“Here, social hierarchy refers to the group-level structure reconstructed from incidental competitions in the integrated tube tests, social rank to an individual’s position within that structure, and social position to social rank together with chasing.”

We additionally clarified the distinction between chasing and social rank in the following sections.

Discussion (ll. 632-664)

“In more constrained or despotic conditions, chasing can serve as a unidirectional, dominance-related behavior directed at subordinates [9,43,44]. The larger groups observed in the NoSeMaze reveal a more nuanced role for chasing behavior. Chasing levels are individually stable, but the expression and meaning of chasing are context-sensitive. Chasing was neither broadly distributed nor consistently directed down the social hierarchy. Instead, it was initiated by a small subset of individuals – primarily those occupying high competition-based social ranks – and frequently occurred reciprocally within this group, suggesting intra-elite social dynamics rather than broad dominance enforcement. Rather than solely serving to impose social hierarchy, chasing appeared to function as a means through which individuals with high social rank monitor, negotiate, and maintain their relative standing within the top tier. Notably, the identity of frequent chasers remained stable over time and persisted across changing group compositions, indicating that the propensity to initiate chases reflects a consistent individual-level tendency rather than a purely situational response.

However, its coupling to social rank depended on the group’s hierarchy structure. In groups with less clearly defined social hierarchies (i.e., lower transitivity), active chases aligned more strongly with social rank, suggesting that mice in less structured groups rely more on proactive signaling to clarify social rank. Indeed, the top-ranked mice in these groups exhibited relatively high levels of active chases. This context-sensitivity highlights chasing’s dual role: it serves as a tool for negotiating social rank among mice at the upper end of the hierarchy, and additionally functions to establish or reinforce hierarchical clarity when social structures are ambiguous. These dynamic aspects of chasing, including its asymmetric initiator-recipient structure and proactive engagement, differ from the nature of tube competitions, which are incidental encounters. Together, tube-derived social rank and chasing describe complementary dimensions of social position, and, alongside other features such as clique formation [36], contribute to describe facets of a broader multidimensional social behavior. Within this complex environment, chasing emerges as a flexible behavioral propensity that is dissociable from formal social rank. Specifically, in the NoSeMaze, chasing contributes dynamically to the maintenance, negotiation, or clarification of social hierarchy structure. These findings reveal a novel aspect of social dynamics: chasing is not merely a dominance display but a flexible context-dependent mechanism shaped by both individual disposition and group-level social structure.”

Conclusion (ll. 718-723)

“Crucially, social position is not fully described by a single behavioral dimension. We focused here on two separable dimensions: competition-based social rank and proactive chasing. Future work should integrate additional dimensions of social organization, such as affiliative bonding and higher-order network measures [36], to capture additional aspects of social organization. Chasing behavior played a dual role, reflecting a stable individual propensity while also adapting to group-level structure.”

(4) Focus on rank bears the risk of overgeneralization for readers not familiar with the context.

As already discussed above (see also point 2 and 3), we have sharpened the framing to reduce the risk of overgeneralization. Throughout the manuscript, we now refer to tube-derived social rank explicitly as a dominance-subordination-related axis of social organization rather than a comprehensive measure of “sociality”, and we also avoid the term “traits”, but prefer

the use of “internalized social rank” or “propensity to chase” when discussing stability across rounds and contexts. Also, we removed “personality-like” as the scope of the study is to set specific behaviors in relation and not to enter the field of mouse personality classification.

We clarified the definition of social rank in this manuscript in the Introduction (ll. 90-94) and the beginning of the Results (ll. 306-308, see also point 3).

Abstract

Lines 41-46

“... Across more than 4,000 mouse-days, hierarchies derived from incidental competitions in the integrated tube tests were non-despotic, transitive, and stable even when group compositions changed. This stability supports an internalized component of competition-based social rank. Chasing was also stable across contexts. Notably, chasing was concentrated among high-ranking individuals, consistent with ongoing negotiation of social rank among individuals at the upper end of the hierarchy.”

Lines 50-54

“In summary, high-dimensional tracking with the NoSeMaze reveals that social position in mice is multifaceted and shaped by stable dimensions of individual behavior that persist across changing social contexts. The approach thus enables longitudinal modeling of individuality and social position as key resilience factors.”

Discussion (ll. 622-626)

“This temporal and contextual stability supports interpreting social rank as a stable, internalized characteristic of individuals. In this sense, ‘internalized’ refers to stability across repeated rounds and reshuffled groups in this paradigm and to relatively stable tube-competition outcomes.”

Changes in the Conclusion starting l. 709 as highlighted in answer to the above point 3.

(5) Conclusion only valid for the reductionistic setting, in which environment, social and non-social changes only within narrow limits, and in which the mouse population does not face challenges

Our conclusions are bounded by the conditions tested, namely an enriched but stable environment, controlled group composition and remixing, and the absence of explicit ecological challenges such as resource scarcity or predators. As described above in the answer to point 1, we now state explicitly that our conclusions apply to the social-context variation tested here (controlled group reshuffling) under otherwise stable conditions, and we frame resource-competition and environmental-stressor manipulations as future variation to test of how this manipulation affects the here described behaviors.

We accordingly changed the Discussion as already highlighted in point 1 above and added ll. 693-704.

(6) Animals are not naive at the beginning of the experiment, but are already several weeks old.

Animals entered the study as adults and were continuously group-housed (3-5 mice/cage) before entering the NoSeMaze. To reduce effects of initial apparatus novelty, mice underwent two NoSeMaze habituation sessions prior to data collection (each several hours). We now clarify these points explicitly in the Methods and scope the inference accordingly. One of our future steps is to extend this framework to earlier developmental stages (e.g., adolescence/weaning) to capture full lifespan trajectories. This however first required

establishing the approach in adult mice under controlled conditions as done in the present study.

Methods (ll. 739-747)

“A total of 79 adult male homozygous OXTR^{fl/fl} mice (B6.129(SJL)-Oxtrtm1.1Wsy/J, [RRID:IMSR_JAX:008471](#), Jackson Laboratory) backcrossed > F10 to C57BL/6J background (Charles River, Sulzfeld) were used for the experiments. Animals entered the NoSeMaze as adults (see Supplementary Table S2 for ages at NoSeMaze entry across rounds) and had been continuously group-housed after weaning (3–5 mice/cage), i.e., they were not developmentally or socially naïve at study onset. Of the 79 mice, 26 mice were injected six weeks before the start of the experiment with an AAV expressing Cre recombinase (rAAV1/2-CBA-Cre) into the AON pars centralis to induce bilateral OXTR deletion (OXTR Δ AON). The remaining 53 animals received an AAV expressing only dTomato (rAAV1/2-CBA-dTomato). ...”

Discussion (ll. 629-631)

“... A future direction is to extend this framework to earlier developmental stages to understand which early experiences shape later trajectories of social position.”

In summary, this is a wonderful study, but not one that is easy to interpret. The bold conclusion is valid only within the constraints of the study, but nevertheless points in an important direction. The paradigm is clever and could be used for many interesting follow-ups.

To define social position as a personality trait will elicit strong opposition and much debate; the nuances of the paper might be lost on many readers and call for the (re)-consideration of many concepts that are touched. I find this attitude a strength of the paper, but the approach bears the risk of misunderstanding.

We thank the Reviewer for the helpful comments. As detailed above, we tightened terminology and framing throughout the revised manuscript by defining stability explicitly in terms of repeatability across rounds and reshuffled groups within this paradigm, and described tube-derived social rank as one dimension of social behavior. We hope these edits preserve the conceptual message while reducing the risk of misunderstanding.

Reviewer #2 (Public review):

Summary:

This manuscript presents the "NoSeMaze", a novel automated platform for studying social behavior and cognitive performance in group-housed male mice. The authors report that mice form robust, transitive dominance hierarchies in this environment and that individual social rank remains largely stable across multiple group compositions. They further demonstrate that social dominance and aggressive behaviors, like chasing, are partially dissociable and that dominance traits are independent of non-social cognitive performance. The study includes a genetic manipulation of oxytocin receptor expression in the anterior olfactory nucleus, which showed only transient effects on social rank.

Strengths:

(1) Innovative Methodology:

The NoSeMaze platform is a technically elegant and conceptually well-integrated system that enables fully automated, long-term monitoring of both social and cognitive behaviors in large groups of group-housed mice. It combines tube-test-like dominance contests, voluntary chase-escape interactions, and an embedded operant olfactory

discrimination task within a single, ethologically relevant environment. This modular design allows for high-throughput, minimally invasive behavioral assessment without the need for repeated handling or artificial isolation.

(2) Experimental Scale and Rigor:

The study includes 79 male mice and over 4,000 mouse-days of observation across multiple group reshufflings. The use of RFID-based identification, automated data logging, and longitudinal design enables robust quantification of individual trait stability and group-level social structure.

(3) Multidimensional Behavioral Profiling:

The integration of social (tube dominance, proactive chasing), physical (body weight), and cognitive (olfactory learning task) measures offers a rich, multi-dimensional profile of each individual mouse. The authors' finding that social dominance traits and non-social cognitive performance are largely uncorrelated reinforces emerging models of orthogonal behavioral trait axes or "animal personalities".

(4) Clarity and Data Analysis:

The analytical framework is well-suited to the study's complexity, with appropriate use of dominance metrics, mixed-effects models, and permutation tests. The analyses are clearly explained, statistically rigorous, and supported by transparent supplementary materials.

We thank the Reviewer for their evaluation of our manuscript. We appreciate their recognition of the NoSeMaze as an innovative and conceptually integrated platform, the scale and longitudinal rigor of the dataset, and the clarity and appropriateness of the analytical framework.

Weaknesses:

(1) Conceptual Novelty and Prior Work:

While the study is carefully executed and methodologically innovative, several of its core findings reaffirm concepts already established in the literature. The emergence of stable, transitive social hierarchies, the persistence of individual differences in social behavior, and the presence of non-despotic social structures have all been previously reported in mice, including under semi-naturalistic conditions (e.g., Fan et al., 2019; Forkosh et al., 2019). Although this work extends those findings with greater behavioral resolution and scale, the manuscript would benefit from a clearer articulation of what is genuinely novel at the conceptual level, beyond the technological advance.

We agree with the Reviewer that transitive dominance hierarchies and stable inter-individual differences have been demonstrated previously in mice, including in semi-naturalistic settings. Our intent was therefore to address two more specific conceptual questions that prior work typically has not tested in an integrated way:

(1) Whether an individual's social position generalizes across distinct social contexts created by systematic changes in group composition rather than reflecting stability that is only observable within a fixed group;

(2) How the two frequently employed social dominance metrics tube competition-based social rank and chasing relate to each other in unperturbed larger groups.

Specifically, the conceptual advance is enabled by (A) continuous, handling-free estimation of competition-based social rank from incidental tube contests over weeks, (B) a repeated

group-reshuffling (“accelerated longitudinal”) design that explicitly tests whether an individual’s social rank generalizes across distinct social contexts, and (C) parallel, continuous measurement of chasing and non-social reinforcement-learning behaviors in the same individuals and environment.

This integration also lets us dissociate dominance-related social dimensions (competition-based social rank vs chasing) and test their relation to individual styles in non-social reinforcement-learning. Importantly, these questions are addressed in larger intact societies (9–10 mice), where hierarchy structure is shaped by more complex network-level dynamics than in smaller groups.

We revised the Introduction and Discussion to foreground these conceptual points and to position them more explicitly in the context of prior works.

Introduction

We adapted the following section (ll. 76-98) and integrated the suggested references:

“... In mice, small groups tend to form highly despotic hierarchies [24-26], whereas larger groups exhibit more complex structures [27]. These patterns suggest a strong influence of emergent group-level dynamics on social structure [12,27]. Yet, animals do not enter social groups as blank slates [28-30]; stable latent factors in the individual may also contribute to hierarchy formation [13]. Thus, it remains unclear to what extent an individual’s social position is internalized and persists across different social contexts [31], or instead is primarily an emergent property of group-level dynamics. Disentangling these possibilities requires experimental conditions that allow unperturbed, continuous tracking of all individuals in sufficiently large groups, together with systematic changes of group composition to modulate social context.

While stable hierarchies and behavioral identity domains have been described previously in semi-naturalistic settings [9-13], many studies quantify these features either within fixed group compositions or in separate assays. Here, we therefore use systematic group reshuffling in 10-member societies to directly test whether individual differences in social rank and chasing persist across distinct social groups. In this study, we use social hierarchy to denote the group-level structure inferred from incidental competitions in the integrated tube tests, and social rank for an individual’s level within that hierarchy. Social position serves as an umbrella term for social rank and chasing behaviors. By continuously measuring social rank, chasing, and reinforcement-learning behavior in the same individuals, we further test how chasing and social rank are related to each other and how they relate to individual styles in non-social reinforcement learning.

To enable these tests, we developed the Non-invasive Sensor-rich Maze (NoSeMaze).”

Discussion

We added a brief framing statement at the beginning of the Discussion (ll. 577-583) to clarify the study’s conceptual novelty:

“Ecologically enriched, yet experimentally controlled assessments allow us to study behavioral individuality and social structure in group-living animals over extended timescales. Here, we show that individual mice carry stable, individual-specific, and multi-faceted profiles of social position and cognitive styles across changing group contexts. Our approach goes beyond prior work by testing cross-context stability under repeated, systematic group reshuffling in larger mouse societies, while measuring competition outcomes, chasing, and reinforcement-learning behavior in parallel. ...”

| (2) *Role of OXTR Deletion:*

The inclusion of the OXTR manipulation feels somewhat disconnected from the manuscript's central aims. The effects were minimal and transient, and the authors defer full interpretation to a separate study.

We appreciate the Reviewer's point and agree that the OXTR^{ΔAON} manipulation can appear secondary to the manuscript's central aims. We included OXTR^{ΔAON} because oxytocin-dependent social recognition memory was hypothesized originally to impact potentially also learning an individual's position in social hierarchy networks. Even though the effects were small and transient, we nevertheless believe it is relevant to report them, also in relation to the more profound effects of the genetic manipulation reported in a related manuscript (Nelias et al., bioRxiv 2025, 10.1101/2025.08.26.672298). Therefore, we explicitly account for genotype as a covariate in the analyses such that the main conclusions do not depend on this manipulation.

To improve coherence, we have added one sentence in the Introduction motivating why OXTR^{ΔAON} was included, and finally more clearly signposted that deeper mechanistic interpretation is beyond the scope of the present manuscript.

Introduction (ll. 115-120)

We added one short sentence introducing the rationale behind the perturbation:

"... and (3) determine whether social rank, chasing, and non-social reward-seeking behaviors represent stable individual characteristics or dynamic features across time and changing group composition. As a secondary analysis, motivated by oxytocin's established role in social recognition memory [32,33], we also tested whether OXTR deletion in the anterior olfactory nucleus produces detectable shifts in rank dynamics."

Results

Lines 161-164

"... This manipulation was included as a secondary biological perturbation. The primary analyses and conclusions focus on the platform and cross-context stability, and genotype is treated as a covariate unless stated otherwise. ..."

Lines 451-454

"... As a secondary analysis, we tested whether OXTR^{ΔAON}, which impairs de novo social recognition memory required for social clique formation in this cohort [36], also affects the measures reported here. Consistent with largely internalized features, OXTR^{ΔAON} produced only transient effects. ..."

Discussion (ll. 594-605)

"This study focused primarily on the relation of social rank and chasing. We however also considered their relation to additional variables including the loss of oxytocin receptors in the olfactory cortex in the adult (OXTR^{ΔAON}), involved in de novo social recognition learning. The propensity to chase was largely unaffected by OXTR^{ΔAON}. Mice carrying OXTR^{ΔAON} displayed a transient reduction in social rank during the first week that normalized thereafter. This transient effect contrasts to the persistent impairment by OXTR^{ΔAON} in forming higher-order social bonds that enable membership in stable cliques, as identified by video tracking of self-paced interactions in the same cohort [36]. Together, these findings suggest that OXT-dependent olfactory learning is critical for the formation of social context-dependent higher-order bonds, but plays a limited role in shaping hierarchy-related behaviors."

(3) *Scope Limitations (Sex and Age):*

The study is limited to male mice, and although this is acknowledged, the title and overall framing imply broader generalizability. This sex-specific focus represents a common but problematic bias. Additionally, results from the older mouse cohort are under-discussed; if age had no effect, this should be explicitly stated.

We thank the Reviewer for this relevant point. The study is limited to male mice. We therefore revised the title and the abstract and strengthened the limitations to make the sex-specific scope explicit.

We additionally note ongoing work extending the same framework to female groups, where we find similar hierarchy structure and cross-context stability in the NoSeMaze in preliminary unpublished data.

Regarding age, our design included two separate cohorts of different adult age ranges (young and older adult animals), but age was not the primary experimental factor. As detailed in our response to Reviewer #3 (point 1), we now quantify age structure explicitly and test age effects using a decomposition that separates between-group age differences from within-group age variation (mean age per group and each animal's deviation from that mean). We also recomputed stability estimates with age-adjusted ICC models, and the resulting ICCs are highly similar to the original estimates (cf. new Supplementary Table S4), indicating that the reported metrics' stability is not driven by age differences across groups.

Title

We change the title from "Individual differences drive social hierarchies in mouse societies" to "Individual differences drive social hierarchies in male mouse societies"

Abstract

We also added male in the abstract (ll. 37-38):

"The interaction of these behaviors in the shaping of social position in larger male mouse groups remains largely unknown."

Discussion (ll. 613-617)

"... While this study focused on male mice, in which social hierarchies are best established [9], future work is needed to explore sex-specific expressions of social structure and their neurobiological underpinnings in female groups. We therefore restrict our interpretation to male mice. Critically, male social ranks were robustly maintained within the same group over time. ..."

For a more detailed integration of age in the Methods, Results, and Discussion, we kindly refer to the reply to Reviewer #3, point 1.

(4) *Ambiguity of Dominance as a Construct:*

While the study robustly quantifies social rank and hierarchy structure, the broader functional meaning of "dominance" remains unclear. As in prior work (e.g., Varholick et al., 2019), dominance rank here shows only weak associations with physical attributes (e.g., body weight), cognitive strategy, or neuromodulatory manipulation (OXTR deletion). This recurring pattern, where rank metrics are reliably established yet poorly predictive of other behavioral or biological traits, raises important questions about what such measures actually capture. In particular, it challenges the assumption that outcomes in paradigms like the tube test or chase frequency necessarily reflect dominance per se, rather than other constructs.

We thank the Reviewer for this clarifying point. We agree that stable social rank metrics do not necessarily imply a complete or unitary measure of "dominance." In the revised manuscript, we therefore clarified that, in this study, social rank is operationalized as consistent competitive outcomes in incidental tube-test encounters in the NoSeMaze (see also Reviewer #1, points 2-4).

Our data indicate that this competition-based social rank is related to, but not identical with, other social behaviors such as chasing. Likewise, body weight significantly contributes to social rank, but explains only part of the variance. We therefore do not interpret weak or partial associations with other variables as invalidating the social rank measure. Rather, we interpret them as indicating that social position is multidimensional and only partially captured by any single assay.

In independent subsequent studies that are currently in preparation or revision, we observed that heterogeneity in the neurobiology and response to challenges was best predicted by the competition-based social rank, also compared to the other behaviors assessed here. While these observations are beyond the scope of the present manuscript, they support our view that incidental tube competition and the resulting dominance-subordination structure may provide a biologically informative measure of one important dimension of social position. At the same time, the observation here and in many previous studies that factors such as body weight explain only a limited portion of the variance remains important for our understanding of these constructs.

We have revised the manuscript accordingly to make this distinction more explicit. In particular, we now define more precisely the terms social hierarchy, social rank, and social position in the context of this manuscript (see also reply to Reviewer #1, point 3; ll. 90-94 in the Introduction and ll. 306-308 in the Results), and we use these terms more consistently throughout. We also revised the text to avoid overstating the meaning of "dominance" where the data support a more specific interpretation.

Finally, we now emphasize more clearly both the strength and the limitation of the present approach. The NoSeMaze allows these relationships to be assessed continuously in a minimally perturbed group-housing ecology, laying ground for future incorporation of further variables and dimensions to capture how they shape social organization.

Specifically, we added text in the Discussion and Conclusion to clarify that competition-based social rank and proactive chasing represent separable dimensions related to dominance and subordination, but do not exhaust sociality or individuality, and that future work should integrate additional measures such as affiliative behavior and higher-order network structure.

Discussion

Lines 624-6631:

“... In this sense, ‘internalized’ refers to stability across repeated rounds and reshuffled groups in this paradigm and to relatively stable tube-competition outcomes. The tube-derived social rank describes here a dimension of individual social behavior. The capacity to quantify stable individual differences across changing social contexts highlights the value of the NoSeMaze for lifespan-oriented studies of behavioral individuality. A future direction is to extend this framework to earlier developmental stages to understand which early experiences shape later trajectories of social position.”

Conclusion

Lines 718-722:

“Crucially, social position is not fully described by a single behavioral dimension. We focused here on two separable dimensions: competition-based social rank and proactive chasing. Future work should integrate additional dimensions of social organization, such as affiliative bonding and higher-order network measures [36], to capture additional aspects of social organization.”

Reviewer #3 (Public review):

Reinwald et al. present the NoSeMaze, a semi-natural behavioral system designed to track social behaviors alongside reinforcement-learning in large groups of mice. Accumulating more than 4,000 days of behavioral monitoring, the authors demonstrate that social rank (determined by tube competitions) is a stable trait across shuffled cohorts and correlated with active chasing behaviors. The system also provides a solid platform for long-term measurements of reinforcement learning, including flexibility, response adaptation, and impulsiveness. Yet, the authors show that social ranking and chasing are mostly independent of these cognitive traits, and both seem mostly independent of oxytocin signaling in the AON.

Strengths:

- (1) The neuroethological approach for automated tracking of several mice under semi-natural conditions is still rare in social behavioral research and should be encouraged.*
- (2) The assessment of dominance by two independent measures, i.e., spontaneous tube competitions and proactive chasing, is innovative and valuable.*
- (3) The integration of a long-term reinforcement-learning module into the semi-natural system provides novel opportunities to combine cognitive traits into social personality assessments.*
- (4) The open-source system provides a valuable resource for the scientific community.*

Limitations:

- (1) Apparent ambiguity and inconsistency in age structure and cohort participation across rounds, raising concerns about uncontrolled confounds.*
- (2) Chasing behavior appears more stable than tube-test competitions (Figure 4D vs. Figure 3D), which challenges the authors' decision to treat tube competitions as the primary basis for hierarchy determination.*

We thank the Reviewer for the evaluation of our work. We have addressed the limitations raised below with additional analyses, clarifications, and corresponding manuscript revisions.

Major concerns:

(1) Unclear and inconsistent handling of age groups and repeated sampling. The manuscript repeatedly refers to "younger" and "older" adults, but it is unclear whether age was ever controlled for or included in models. Some mice completed only one round, others 2-5 rounds, without explanation of the criteria or balancing.

We thank the Reviewer for this clarifying comment. We now explicitly quantify participation in the different rounds in new Supplementary Table S3. Importantly, our primary stability analyses are implemented using variance-component mixed models (REML) that naturally handle unbalanced repeated-measures data. This is explained in more detail in the Methods section (ll. 1003-1111), where we added a statement that the LMEs are well suited for unbalanced repetitions.

To further address unbalanced participation, we additionally performed a conservative sensitivity analysis restricted to a balanced subset, including only sessions 1 and 2 and only mice with observations in both sessions. For the key social measures, ICC estimates were highly similar in the full dataset and in the balanced subset (e.g., z-scored competition David's score, ICC across cohorts: 0.55 without age adjustment vs. 0.56 in the balanced first-two-session subset; active chasing: 0.74 vs. 0.72; being chased: 0.61 vs. 0.60; Supplementary Table S4), indicating that unbalanced participation did not inflate the stability estimates.

We also addressed age structure explicitly. Although age was not the primary experimental factor, the inclusion of two age cohorts allowed us to assess whether the observed behaviors and their interrelations were robust across most of the adult lifespan (cf. new Figure 1). We therefore decomposed age into a between-group component (mean age per group) and a within-group component (each animal's deviation from its group mean) and included these terms in the relevant LME models. Tube-based dominance rank (David's score, z-scored) showed no age effect ($p_{\text{age, cond.}} = 0.63$), whereas chasing metrics showed modest age associations (cf. new Supplementary Table S5). This however only indicates that chasing was associated with age to some degree. More importantly, recomputing all stability estimates using age-adjusted ICC models yielded nearly identical ICCs for the core social measures (new Supplementary Table S4), indicating that the reported stability was not affected by age differences across groups.

In addition, we revised the study design schematic (new Fig. 1; cf. Reviewer #1, Recommendations for the authors) to depict the separate age cohorts and the reshuffling procedure more clearly.

We adapted the following sections accordingly.

Methods

Lines 980-985

"... Most mice ($n = 68$) participated in at least two NoSeMaze rounds with reshuffled group members. Supplementary Table S2 summarizes the number of rounds per mouse and missing data due to technical problems. We examined the stability of social and reward-seeking metrics by correlating values from the first and second round (Spearman's correlation, cf. Fig. 5). These round-1-to-round-2 correlations use one paired observation per mouse and are therefore not inflated by mice contributing >2 rounds."

Lines 991-993

"The ICC treated mouse identity as a random intercept and NoSeMaze group (i.e., round-specific social group) as a random effect, with repetition included as fixed effect (i.e., stability across groups while holding repetition means constant)."

Lines 1003-1011

“Variance components for ICC estimation were obtained from LMEs fit by restricted maximum likelihood (REML), which yields less biased variance-component estimates and is well suited for unbalanced repeated-measures designs (i.e., different numbers of rounds per mouse). To assess potential confounding by age structure, age was decomposed into a between-group component (group-mean age) and a within-group component (each animal’s deviation from its group mean) and included as covariates. ICCs were recomputed in age-adjusted models (see Supplementary Table S4). Finally, we performed a conservative sensitivity analysis restricted to the first two sessions per mouse and to mice with observations in both sessions (“balanced first2”), to additionally account for unbalanced participation structure.”

Results

Lines 167-173

“... The study population comprised two age cohorts: younger (16-30 weeks) and older adults (55-97 weeks) (Fig. 1B). The two age cohorts were run as separate experimental series, and group reshuffling was performed within each cohort (Fig. 1C, see Supplementary Table S2). Mice lived in groups of 9-10 for multiple rounds in the NoSeMaze, with different group members in each round (Fig. 1D, see Supplementary Table S1). This allowed us to test which individual behaviors were stable across different group compositions. ...”

Lines 441-450

“Because the number of rounds in the NoSeMaze was unbalanced between animals (Supplementary Table S3) and age varied across groups (Supplementary Table S1-2), we also performed robustness checks. Stability estimates changed only minimally when recomputed in age-adjusted ICC models (between-group mean age and within-group age deviation; see Methods) and when restricting analyses to a balanced first-two-session subset (sessions 1–2 only; mice with both sessions) (Supplementary Table S4). Mixed models indicated that some chasing and reinforcement-learning measures showed modest age- and/or session-related shifts in absolute levels (Supplementary Table S5), but importantly, these did not affect the observed stability patterns.”

(2) Stability of chasing appears stronger than the stability of tube competitions. Figure 4D shows highly consistent chasing behavior across weeks, while Figure 3D shows weaker and more variable correlations for tube-based David scores. This is also evident from Figure 5A-B,D. Thus, it appears that chasing, which serves to quantify dominance in similar semi-natural setups, may be a more reliable and behaviorally meaningful measure of dominance than the incidental tube competitions.

Indeed, active chasing showed higher cross-round correlations than tube-derived social rank (e.g., R1–R2 Spearman $\rho = 0.75$ vs 0.57 ; ICC_{across cohort} 0.74 vs 0.55 , see Fig. 5 and new Supplementary Table S4). Importantly, this does not contradict the central finding that tube-derived social rank is stable across time and across remixed groups. Rather, it may highlight that chasing and social rank capture different aspects of dominance-subordination-related behavior with different statistical properties. Chasing reflects an individual’s propensity to actively initiate interactions (a strongly expressed, asymmetric behavior), which can be highly consistent across contexts. By contrast, social rank is a relational measure inferred from symmetric dyadic win–loss outcomes based on incidental competitions in the integrated tube tests and can vary with the specific set of competitors and interaction opportunities in each reshuffled NoSeMaze group, while still remaining substantially stable overall.

Accordingly, we do not interpret the higher repeatability of chasing as evidence that it is the “better” hierarchy measure. Tube competitions yield symmetric dyadic outcomes that directly support formal hierarchy reconstruction (David’s score/Elo, transitivity, steepness) and show convergent validity with traditional tube testing. Chasing, in contrast, is asymmetric and volitional and, in our data, is concentrated in the upper social ranks and modulated by group-level social hierarchy structure, consistent with rank negotiation/signaling rather than a mechanism that assigns a full ordering to all individuals. These different properties lead to very different “win-lose” relationships when comparing tube competition events to chasing events that we specifically illustrated in Supplementary Fig. S13. We therefore revised the Results and Discussion to frame tube-derived social rank and chasing as complementary social dimensions.

Results (ll. 402–413)

“... Specifically, stability was high for the tube competition-based David’s score (Fig. 5A, $\rho = 0.57$, $p < 0.001$), as well as for the fraction of active chases (Fig. 5B, $\rho = 0.75$, $p < 0.001$) and of times being chased (Fig. 5C, $\rho = 0.54$, $p < 0.001$). Notably, the fraction of active chases showed slightly higher across-round stability than tube-derived David’s score and the fraction of being chased. This is in line with active chases capturing an individual propensity to initiate this behavior, whereas competition-based social rank is a relational measure that is also influenced by the set of competitors and interaction opportunities in each reshuffled group. Nonetheless also social rank and being chased were overall stable. Across the full series, ICCs for these social measures were also in the good–excellent range, indicating high across-round stability (Fig. 5D, ICC = 0.548 to 0.890).”

Discussion

Line 653–664 (see also Reviewer #1, point 2)

“These dynamic aspects of chasing, including its asymmetric initiator–recipient structure and proactive engagement, differ from the nature of tube competitions, which are incidental encounters. Together, tube-derived social rank and chasing describe complementary dimensions of social position, and, alongside other features such as clique formation [36], contribute to describe facets of a broader multidimensional social behavior. Within this complex environment, chasing emerges as a flexible behavioral propensity that is dissociable from formal social rank. Specifically, in the NoSeMaze, chasing contributes dynamically to the maintenance, negotiation, or clarification of social hierarchy structure.”

(3) Unbalanced participation across rounds compromises stability analyses. Stability analyses (e.g., ICCs, round-to-round correlations) assume comparable sampling across individuals. However, some mice contribute 1 round, others 2, 3, 4, and even 5 rounds. This imbalance may inflate stability estimates or confound group reshuffling effects, and the rationale for variable participation is not explained.

We thank the Reviewer for raising this point. Indeed, participation was unbalanced across rounds (see the same Reviewer #3, Major concerns 1 and new Supplementary Table S3), mainly due to missing data from occasional technical failures of the RFID detectors or the reinforcement learning water port during acquisition in some groups (for details, see Supplementary Table S2). We added this rationale behind variable participation to our Methods section (for details, see Major concern 1, ll. 981–982, “Supplementary Table S2 summarizes the number of rounds per mouse and missing data due to technical problems.”)

We now quantify round participation (new Supplementary Table S3) and directly address potential bias from unequal sampling in two ways. First, the round-1-to-round-2 (R1–R2) stability correlations use one paired observation per mouse and therefore are not inflated by mice contributing more than two rounds. Second, ICCs were estimated from REML variance-

component mixed models that account for unbalanced repeated-measures by design and use all available observations. To further rule out inflation from unequal sampling or non-random missingness, we additionally report a conservative sensitivity ICC restricted to a balanced subset including only each mouse's first two observed sessions and only mice with both sessions ("first2-balanced"). Full-sample ICCs and first2-balanced ICCs were highly similar (new Supplementary Table S4), indicating that participation imbalance did not affect the stability estimates. For details on the changes made in the manuscript, see Reviewer #3, Major concern 1.

Recommendations for the authors:

Editor's notes:

Should you choose to revise your manuscript, if you have not already done so, please include full statistical reporting including exact p-values wherever possible alongside the summary statistics (test statistic and df) and, where appropriate, 95% confidence intervals. These should be reported for all key questions and not only when the p-value is less than 0.05 in the main manuscript.

Readers would also benefit from noting that the mice were male in the abstract.

We have revised the manuscript accordingly and now report exact p-values values wherever possible for the key results in the main text. Because full statistical reporting for every analysis in the main text would substantially interrupt readability, we provide the complete statistical details in a Supplementary Excel File (Supplementary Material – Systematic Statistical Reporting), including sample sizes, degrees of freedom, the number and type of permutation tests, exact p-values, and 95% confidence intervals. We also provide Extended Data Sheets for all linear-mixed effects models that account for covariates such as age and round of participation in the NoSeMaze (cf. Reviewer #3, Major Concern (1) for the additional analyses). To guide readers to these resources, we now explicitly refer to the Supplementary Material – Systematic Statistical Reporting at several points in the manuscript:

Results

Lines 271-274

“Detailed statistical reporting for all analyses, including n, degrees of freedom, exact p-values, and 95% confidence intervals, is provided in the Supplementary Material - Systematic Statistical Reporting. Extended Data includes additional linear mixed-effects models controlling for potential confounding variables.”

Lines 430-431

“All statistical details are provided in the Supplementary Material – Systematic Statistical Reporting.”

Additional references to the supplementary statistical reporting were inserted at lines 252-254, 562-563, and 567-568.

Methods

Lines 962-965

“Full details on the statistical tests, including n, degrees of freedom, exact p-values, and 95% confidence intervals, are provided in the Supplementary Material – Systematic Statistical Reporting, as well as in the Extended Data for the LMEs accounting for different covariates.”

We also revised the abstract and the title to explicitly state that the mice were male.

Manuscript changes:

Results, figure legends, and supplementary tables expanded to include full statistical reporting; abstract revised to specify male mice.

Reviewer #1 (Recommendations for the authors):

I would recommend being much more explicit about the reductionistic nature of the study and how the limitations are turned here to an advantage, while at the same time acknowledging the challenges of extrapolating beyond these boundaries. Most importantly, dominance should be positioned more clearly and cautiously within a framework of social behavior (and social structure) in general. The authors include cognitive tests, etc., to generate context, but this context is dependent on the same circumstances that possibly contribute to the social structure. The information that lies in the chasing behavior as an additional measured variable might be used better to provide more context.

We thank the Reviewer for these recommendations. We have better clarified the framing to make explicit that the NoSeMaze is a controlled laboratory group-housing system. The specific conditions are now discussed in more detail in relation the observed social behaviors. We describe the tube-derived social rank as one dimension of social organization, and proactive chasing as another one. This study presents a first necessary step to understand their shared and distinguishing features. We also elaborated the Discussion to better clarify the trade-off between ecological complexity and experimental control, and to highlight that the value of the NoSeMaze for continuous, observer-independent phenotyping under standardized conditions. We now clearly state the importance to vary conditions in the system to see how the social behaviors and also their relation to non-social features changes depending on context conditions. For detailed changes, see our responses to Reviewer #1, points (1), (3), (4), and (5), and Reviewer #2, Weakness (4).

Manuscript changes:

Abstract, Introduction, Discussion, and Conclusion revised to clarify scope, construct interpretation, and the complementary roles of competition-based social rank and chasing.

The precision of the description of the experimental design should be improved. When were the cohorts mixed, or did they stay separate? Which animals were old, which were young? This remained a bit confusing.

We revised the presentation of the study design accordingly. Specifically, we clarified that the younger and older adult cohorts were run as separate experimental series and that group reshuffling occurred within, but not across, these cohorts. We also revised the study schematic in Fig. 1 and the corresponding description in the manuscript to more explicitly depict cohort structure, timing of NoSeMaze rounds, and between-round reshuffling (details are provided in the Supplementary Tables 1-3). For new age-related analyses and additional robustness checks, see our response to Reviewer #3, Major Concern (1).

Manuscript changes:

Figure 1 and its legend were revised to depict cohort structure, NoSeMaze rounds, and between-round reshuffling more explicitly. In addition, the Results subsection “Ecological longitudinal assessment in the NoSeMaze” were updated to clarify that the younger and older adult cohorts were run separately, that reshuffling occurred within but not across cohorts, and which animals belonged to each age-defined cohort. We also now cross-reference Supplementary Table S2 for round-specific age information and cohort composition.

Results

Lines 167-170

“The study population comprised two age cohorts: younger (16-30 weeks) and older adults (55-97 weeks) (Fig. 1B). The two age cohorts were run as separate experimental series, and group reshuffling was performed within each cohort (see Supplementary Table S2).”

It is also not fully clear across which groups stability measures were obtained: are these across all groups or within the subgroups with a given characteristic?

We thank the Reviewer for highlighting this point. We now state explicitly that stability measures were computed across all eligible rounds, using mixed-effects models that account for repeated observations of the same mouse and for round-specific group membership. We also added a conservative sensitivity analysis restricted to a balanced first-two-session subset. Full and balanced-subsample stability estimates were highly similar, indicating that the main conclusions are robust to the participation structure. For details, see our response to Reviewer #3, Major Concerns (1) and (3).

Manuscript changes:

Methods and Results revised to clarify the level of analysis and model structure, as well as new supplementary robustness tables (Supplementary Tables 3-5) added.

Minor point: In Figure 3, 18 groups are mentioned, but 19 are shown.

We thank the Reviewer for this clarification. The apparent discrepancy arose because the group labels in Figure 3 follow the numbering of all experimental groups, whereas only groups with available tube-competition data are shown in this panel. Accordingly, group 16 is absent because no tube data were available for that group (see Supplementary Table S2), and groups 20 and 21 are likewise not included for the same reason. We have revised the figure legend and corresponding text to make this explicit and to avoid the impression of a numbering inconsistency.

“Fig. 3: Social rank derived from incidental competitions in the integrated tube tests of the NoSeMaze.”

“C, Box plots of metrics characterizing social hierarchy for 18 groups, including transitivity, steepness, stability, and uncertainty-by-repeatability (for details, see ‘Source Data’). Group labels correspond to original experimental group IDs. Only groups with available tube-competition data are shown. Therefore, numbering is non-consecutive (e.g., groups 16, 20, and 21 are absent; see Supplementary Table S2).”

Reviewer #2 (Recommendations for the authors):

(1) To better distinguish this study from previous literature, we recommend incorporating a more focused discussion (or adding to the intro) of how the findings advance our understanding of social hierarchy beyond prior works.

We have sharpened the conceptual framing in both the Introduction and Discussion. In particular, we now distinguish more explicitly between the well-established observation that hierarchies form and remain stable within fixed semi-naturalistic groups, and the more specific question addressed here: whether an individual’s social position generalizes across changing social contexts created by repeated group reshuffling. We added additional references and also emphasize that the present study integrates continuous measurements of competition-based social rank, chasing, and reinforcement-learning features in the same individuals and environment. For details, see our response to Reviewer #2, Weakness (1).

Manuscript changes:

Introduction and Discussion revised to foreground conceptual novelty relative to prior semi-naturalistic work.

(2) Given the weak associations between dominance rank and other traits such as body weight, cognitive performance, and oxytocin receptor manipulation, we suggest further clarifying what is being captured by these measures.

We agree and have clarified this point throughout the manuscript. We now define tube-derived social rank explicitly as an operational measure based on repeated competitive outcomes from incidental dyadic tube tests, highlighting it as one dimension of social behavior.

We also make clearer that proactive chasing is not redundant with rank, but instead captures a distinct, partly dissociable dominance-related interaction mode. For details, see our response to Reviewer #2, Weakness (4).

We understand the question on the meaning of social rank as the correlations to body weight and cognitive performance are only punctual. We would like to mention here already that the competition-based social rank turns out to be a strong predictor of individual reactivity in a series of challenges. These works are in currently in preparation for publication and will make the relevance of these measures more clear. Related to this, we find in these studies that chasing and competition-based social rank predict different behavioral and neuronal aspects of individual reactivity.

Manuscript changes:

Abstract, Results, Discussion, and Conclusion revised to clarify construct interpretation and multidimensionality.

(3) We recommend modifying the title and abstract to more clearly reflect the male-only design of the study. In addition, please indicate whether any age-related differences were observed. If age had no measurable effect, this should be stated explicitly to justify the combination of age groups.

We revised the title and abstract to make the male-only design explicit. We also now report age-related analyses directly in the manuscript. Briefly, age was modeled by separating between-group age structure from within-group age variation. Some measures showed modest age associations in mean level, but most importantly, age-adjusted stability estimates for the core social metrics were highly similar to the original estimates, indicating that the reported stability is not driven by age differences across groups. For details, see our response to Reviewer #3, Major Concern (1).

Manuscript changes:

Title and abstract revised. Methods, Results, and supplementary robustness analyses expanded to report age effects explicitly.

<https://doi.org/10.7554/eLife.109354.2.sa0>