



Published in final edited form as:

Nat Neurosci. 2021 September ; 24(9): 1243–1255. doi:10.1038/s41593-021-00888-4.

Reduced sociability and social agency encoding in adult *Shank3* mutant mice are restored through gene re-expression in real-time

Daniel K Lee^{#1,2}, S William Li^{#2,3}, Firas Bounni^{#2}, Gabriel Friedman^{#2}, Mohsen Jamali², Leah Strahs⁴, Omer Zelinger⁴, Pauline Gabrieli⁵, Michael A Stankovich², Jack Demaree⁴, Ziv M Williams^{1,2,6,**}

¹Harvard-MIT Division of Health Sciences and Technology, Boston MA

²Department of Neurosurgery, Massachusetts General Hospital, Harvard Medical School, Boston MA

³Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston MA

⁴Northeastern University, Boston MA

⁵Harvard University, Cambridge MA

⁶Harvard Medical School, Program in Neuroscience, Boston MA

These authors contributed equally to this work.

Abstract

Despite a growing understanding of the molecular and developmental basis of autism spectrum disorder (ASD), how the neuronal encoding of social information is precisely disrupted in ASD and whether it contributes to abnormal social behavior remains unclear. Here, we disrupted and then restored expression of the ASD-associated gene *Shank3* in adult male mice while tracking the encoding dynamics of neurons in the medial prefrontal cortex (mPFC) over weeks. We find that *Shank3* disruption led to reduction of neurons encoding the experience of other mice and increase in neurons encoding the animal's own. This shift was associated with loss of ability by neurons to distinguish other-from-self and, therefore, the inability to encode social agency. Restoration of *Shank3* expression in the mPFC reversed this encoding imbalance and increased sociability over

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

**Correspondence should be made to zwilliams@mgh.harvard.edu.

Author contributions

S.W.L., F.B., L.S., O.Z., P.G., M.A.S., and J.D. performed the experiments. D.K.L., S.W.L., G.F., M.J., and P.G. analyzed the data. S.W.L. developed the conditional gene knockout studies, D.K.L., G.F., and S.W.L. edited the manuscript and Z.M.W. conceived and designed the study, wrote the manuscript, and supervised all aspects of the research.

Competing interests

The authors declare no competing interests.

Reporting Summary

Please see additional details in the “Life Sciences Reporting Summary” attached to this article

Code availability

All software used in this study are listed in the Reporting Summary along with their versions. The primary MATLAB codes used to perform the statistical and data analyses in this study are available from the corresponding author upon reasonable request.

5-8 weeks. These findings reveal a neuronal encoding process that is necessary for social behavior and that may be disrupted in ASD.

Introduction

Animal models offer the opportunity to study some of the basic neuronal encoding processes that underlie social behavior and to evaluate how these processes may be disrupted by conditions such as autism spectrum disorder (ASD). Prior investigations in rodents and primates, for example, have provided evidence for the presence of neurons that encode information about other individuals, demonstrating changes in activity when the animals observe another's appetitive or aversive experiences¹⁻¹¹. They have also provided evidence for neurons that encode another animals' familiarity or dominance^{12,13}, further suggesting that social context plays a prominent role in such computations. While these observations have provided insight into the mechanisms by which social information is likely encoded by neurons in the mammalian brain, they do not reveal whether or what computations may be disrupted in ASD or how their disruption may relate to abnormal social behavior.

Individuals with ASD often show difficulty in evaluating social cues and in appropriately interpreting another person's experiences or emotions, and may display a preoccupation with self or limited interest in others¹⁴⁻¹⁹. Heterozygous mutations of the *Shank3* gene are associated with up to one percent of ASD cases^{20,21} and are a major contributor in Phelan-McDermid syndrome. In mice, both congenital knockout and haploinsufficiency of *Shank3* leads to diminished social interest in others and is associated with the disruption of cortical and subcortical circuits thought to be involved in ASD²²⁻²⁶ and its ontogenesis^{22,25,27,28}. These mutations are also often related to excitatory synaptic dysfunction and neurotransmission that may be potentially restorable^{24,29,30}. Yet, whether and how *Shank3* mutations or haploinsufficiency disrupt the processes by which social information is precisely encoded by individual neurons in behaving animals remains largely unknown.

Here, we approached these questions by using a FLEEx-switch strategy that allowed us to simultaneously track the encoding properties of neurons in freely interacting mice as we disrupted and then gradually restored *Shank3* expression in real-time. We show how changes in social behavior due to *Shank3* disruption temporally relate to the ability of neurons to encode information about other agents as well as to distinguish one's own experience from another's. We also demonstrate how targeted recovery of *Shank3* expression in the medial prefrontal cortex (mPFC) gradually restores this imbalance and increases sociability over time.

Results

Studying the relation between *Shank3* expression, neuronal encoding and social behavior

To study the relation between *Shank3* expression, neuronal encoding and social behavior, we used a Cre-dependent FLEEx switch approach that allowed us to disrupt and then gradually restore *Shank3* expression *in vivo* as neuronal recordings and behavioral testing were concurrently made (Supplementary Table 1). To create Cre-dependent switch constructs

capable of controlling *Shank3* expression in adult animals actively performing social behavioral tasks, we first crossed *CAGGCre-ER*^{+/-} with *Shank3*^{fx/fx} mice in order to create animals that could serve as functional *Shank3*^{+/-} knockouts³⁰⁻³². Further, because the *Shank3* gene was under receptor-mediated CreER control, we could manipulate *Shank3* expression in real-time as neuronal recordings were made. Tamoxifen (TMX) or endoxifen was used for either brain-wide or region-specific targeting, respectively, to activate CreER function and thus drive increased *Shank3* expression in the *Shank3*^{fx/+};*CreER*^{+/-} mice. *Shank3*^{fx/+};*CreER*^{+/-} littermates that were not given TMX and *Shank3*^{fx/+};*CreER*^{-/-} littermates that were given TMX were used for control comparison (Methods).

In our study, we used heterozygous constructs to mimic the physiological expression of *Shank3* and because, unlike animals with homozygous mutations, they generally do not exhibit motoric dysfunction, anxiety or self-injurious behavior that could potentially confound social testing^{22,33}. All animals were adult male mice, ages four months or older to allow for consistency, and included either (i) *Shank3*^{fx/+};*creER*⁻, (ii) *Shank3*^{fx/+};*creER*⁺, (iii) *Shank3*^{+/+};*creER*⁻ or (iv) *Shank3*^{+/+};*creER*⁺ constructs. In total, we used 77 animals of which 28 were wild-type (WT) and 49 were heterozygous (HET) *Shank3* mice (Extended Data Fig. 1a; Supplementary Table 1).

Here, we used microelectrode arrays that were implanted in the animal's mPFC (Extended Data Fig. 1b), an area suggested to be involved in social behavior^{1,7,12,26,34} and disorders such as ASD^{14,21,35,36}. The responses of recorded neurons were evaluated across different task conditions and their encoding properties were compared both within and between animals. The animals' behaviors were tracked using a ceiling mounted digital camera; with the task condition and animal position being determined in an automated, genotype-blinded manner (Methods).

To first study the encoding properties of individual neurons and to further differentiate information related to other social agents from the individual's own, we used a task design that allowed the recorded animals to undergo paired interactions across varying task conditions. These conditions were varied across three main orthogonal axes that defined the animal's interactions – *social agency* (self vs. other), *experience valence* (positive vs. negative) and *identity* (familiar vs. non-familiar). All trials were given in pseudo-random fashion and were separated by one-minute neutral inter-trial periods. To provide experiences that were salient to the animals and to identify neurons which responded selectively to another's specific experiences (i.e., rather than simply to any experience), we used a confined tube enclosure for the aversive experience³⁷⁻³⁹ and a food-bated enclosure for the appetitive experience⁴⁰. For example, an appetitive experience may be given to the other social agent in one trial but may be given to the recorded animal themselves on another. On other trials, by comparison, either the recorded animal or other agent may be given an aversive experience (Fig. 1a–b).

To further dissociate neuronal responses that may reflect the other animal's experiences from those that more simply reflect the enclosures themselves, we also pseudo-randomly replaced the other animals with inanimate totems as well as alternated the enclosures in which the recorded animal's themselves were placed in. Thus, for instance, the recorded

animal may observe another animal in an aversive enclosure on one trial but, alternatively, may observe the same enclosure with an inanimate totem in another. Finally, to evaluate the influence that social context itself had on neuronal response, we alternated the identity of the other animal as familiar and non-familiar.

All enclosures were well-habituated and did not require conditioning (e.g. tone-shock pairing)²⁸, therefore, allowing us to consistently study neuronal response across multiple interleaved trials. All task conditions were counterbalanced so that an equal number of self- and other- trials as well as an equal number of appetitive- and aversive- conditions were performed.

Single-neuron representations of other and self during social interactions

Together, we recorded from 188 mPFC neurons in the WT mice, 180 neurons in the HET mice prior to TMX and 837 neurons from the HET mice after TMX or endoxifen administration (Fig. 1c and Extended Data Fig. 1c). The process by which neurons encode social information is described by changes in their spiking activity⁴¹. For example, if a neuron demonstrated a difference in activity when another animal received an aversive vs. an appetitive experience but not to the animal's own, then this individual neuron would be considered to selectively encode information related to the other social agent's experience. Here, using a three-way analysis of variance that considered social agency, experience valence and identity for the main effects across the 188 WT neurons recorded, we found 112 neurons that displayed task-related modulation (three-way ANOVA, $p < 0.0125$ with *post-hoc* comparisons). Of these neurons, 30.0% ($n = 33$ of 112) responded to variations in the other social agent's experience (Fig. 1d, *bottom*) meaning that these cells exhibited a relative change in their activity when the other mouse underwent a positive compared to negative experience. A similar proportion of neurons was also observed when using a modeling approach that decoded the experiences on a trial-by-trial basis (36%, $n = 37$; Supplementary Table 2), further suggesting that these encoding properties were robust. The overall probability of observing these numbers of cells by chance was unlikely given the number of neurons recorded (bootstrap analysis, $p < 1.0 \times 10^{-9}$).

Neurons that encoded the other animal's experience were largely distinct from those that encoded the animal's own experience. Of neurons that were modulated by the task, 25.9% ($n = 29$ of 112) encoded the animal's own experiences, meaning that they displayed a relative change in their activity when the recorded animal themselves were undergoing a positive compared negative experience (three-way ANOVA, $p < 0.0125$). Of the neurons that encoded the animal's own experience, however, only 9 also encoded the other animal's experience; a proportion that was significantly lower than expected from chance (bootstrap analysis, $p < 0.01$) (Fig. 2a). Figure 1d illustrates two such representative cells from the same WT animal; one that differentially responded to the other social agent's experiences and the second that differentially responded to the animal's own. This relative lack of overlap in neuronal response to self- vs. other-experience was also apparent at the level of the population when considering differences in each neuron's firing rates (2D KS test, $K\text{-stat} = 0.52$, $p = 9.2 \times 10^{-3}$; Fig. 2 and Supplementary Fig. 1). Therefore, in WT animals, most

neurons in the mPFC that encoded information about the others experience encoded little information about the animal's own experience, and *vice versa*.

Sensitivity of other-encoding neurons to social information

We next assessed whether other-encoding neurons in the mPFC respond to the other social agent's experience rather than to the aversive or appetitive stimuli themselves (Fig. 3). It is possible, for example, that similar neuronal responses would have been observed if the recorded animals were simply presented with an appetitive or aversive enclosure. This was unlikely since most neurons that responded to the other's experience displayed little response to the animal's own. Nonetheless, to evaluate for this possibility more directly, we also examined the trials in which the other animal was replaced with an inanimate totem (Fig. 1b). Under this setting, only 3.6% of neurons ($n = 4$ of 112 cells) responded to the appetitive and aversive stimuli, of which 3 were among the neurons found to encode the other's experience (Chi-square test, $\chi^2(1) = 10.9$, $p = 9.6 \times 10^{-4}$). These neurons therefore did not simply respond to the stimuli themselves.

We also considered other factors indirectly related to the stimuli – for instance, smell emitted when the other animal was eating or vocalizations associated with stress. To this end, we performed an additional control in which the other animal was given the same appetitive and aversive stimuli, but now visual access to the other was blocked by an odor- and sound-permeable barrier (i.e., the recorded mouse was not allowed to view whether or which enclosure their partner was placed in). Now, though, only 7.1% ($n = 8$ of 112) of neurons responded to the aversive and appetitive conditions, of which only one neuron overlapped with those encoding the other's experience (Chi-square test, $\chi^2(1) = 62.1$, $p < 1.0 \times 10^{-7}$; Fig. 3a).

Neuronal responses to the other animal also did not reflect the recorded animal's own enclosure or experience. Only 3.7% ($n = 7$) of neurons displayed a difference in response to the other's experience based on which specific enclosure the recorded animal was simultaneously placed in (three-way ANOVA, $p < 0.0125$ with *post-hoc* comparisons). Moreover, only 1 of the neurons that displayed a significant response to the other's experience also demonstrated a difference in response based on which enclosure the recorded animal was in (Chi-square test, $\chi^2(1) = 62.1$, $p < 1.0 \times 10^{-7}$; Extended Data Fig. 2a), suggesting that neurons which responded to the other's experience were largely insensitive to recorded animal's own specific enclosure.

Finally, to confirm that neuronal responses to the other's experience indeed reflected the social context of the animal's interaction, we evaluated whether the other animal's identity influenced neural activity. Familiarity with a conspecific plays a crucial role in how animals and humans perceive and respond to the experiences of others, and is often used to evaluate the effect that social context has on such interactions¹². Of neurons that were modulated by the task, 18% ($n = 20$ of 112) responded distinctly to the other animal when the other animal was familiar to the recorded mouse compared to when the other animal was unfamiliar (three-way ANOVA, $p < 0.0125$; Fig. 3c). Familiarity with the other animal, however, also markedly influenced neuronal response to the other's experience: neurons that encoded the other's experience displaying a significantly greater difference in activity between the

positive and negative experiences when their partner was familiar to them (two-sample t-test, $t_{s(64)} = -3.2$, $p = 2.1 \times 10^{-3}$) than when the partner was not familiar (Fig. 3d and Extended Data Fig. 2b). In other words, neuronal responses to ‘what’ the other was experiencing also reflected to ‘whom’ the experience belonged. Extended Data Fig. 2 further illustrate the most common feature combinations represented by neurons in the mPFC across the different social agency, experience valence and identity condition combinations. Taken together, neuronal responses to the other’s experience reflected the social context of their interaction rather than simply information about the sensory stimuli or enclosures themselves.

Loss of *Shank3* disrupts social encoding and self-other distinction in individual neurons

We next turned to the HET mice in which *Shank3* expression was disrupted. Here, we recorded from 180 neurons in the HET mice prior to TMX administration. Overall, the firing rates of these neurons were similar as those in WT (3.37 ± 0.31 vs. 3.54 ± 0.37 spikes/sec for the WT and HET mice, respectively, two-sample t-test, $t_s = -0.52$, $p = 0.61$), and had similar waveform morphologies (Fig. 1c and Extended Data Fig. 1b). Further, the HET mice displayed a slightly higher proportion of task-modulated neurons, indicating that the quality of recordings across the WT and HET mice was comparable (112 of 188 vs. 131 of 180 for the WT and HET mice, respectively; Chi-square test, $\chi^2(1) = 7.1$, $p = 0.0075$). Finally, we confirmed that all recording locations in the WT and HET mice were confined to the same mPFC area (Extended Data Fig. 1b).

While many neurons in the HET mice were modulated by the task conditions, markedly fewer neurons encoded information about the other social agent’s experience. Of task-modulated neurons, only 9.2% ($n = 12$ of 131; Fig. 1d) encoded the other agent’s experience, a proportion that was significantly smaller than that observed in WT (Chi-square test, $\chi^2(1) = 10.5$, $p = 1.2 \times 10^{-3}$; Fig. 2a). The degree to which neurons in the HET mice responded to the other’s experiences was also significantly diminished – displaying a smaller overall difference in activity when the other animal underwent a positive compared to negative experience (two-sample t-test, $t_{s(43)} = 2.6$, $p = 0.013$; Fig. 2b). This difference in activity between the WT and HET animals was consistent across the course of the trials (Extended Data Fig. 3).

Decrement in the proportion of neurons that encoded the other’s experience was associated with a proportional increase in neurons that responded to the animal’s own experiences. Overall, 38.2% ($n = 50$ of 131) of neurons in the HET mice encoding experiences related to self. Further, when considered across the population, the ratio of neurons that responded to other- vs. self-experience decreased from 1.1:1 in the WT mice to 1:4 in the HET mice (Chi-square test, $\chi^2(1) = 14.0$, $p = 1.8 \times 10^{-4}$). Similar results were found by decoding analysis (Extended Data Table 3) as well as when performing within- vs. between-group comparisons, indicating that these differences in neuronal encoding were robust (Extended Data Fig. 4). These differences in encoding between the WT and HET mice was also largely independent of the statistical thresholding used ($p < 0.05$ to 0.0025; Extended Data Fig. 5a) and was consistent across animals (Extended Data Fig. 5b), indicating that reduced *Shank3* expression was associated with a loss of neuronal response to the other’s experience.

We also observed that the reduced neuronal response to the other's experience was associated with a loss of distinction between other and self. As noted above, in WT mice, only 9 neurons that encoded the other's experience also encoded the animals own. In the HET mice, by comparison, all neurons (100%) that responded to the other's experience also responded to the animal's own experience (9 of 33 vs. 12 of 12, Chi-square test, $\chi^2(1) = 18.7$, $p = 1.5 \times 10^{-5}$; Fig. 2a). This lack of self-other distinction was also notable at the level of the population when examining the raw firing rates across cells (2D KS test, K-stat = 0.37, $p = 0.70$; Fig. 2a and Supplementary Fig. 1). Cells in the HET mice therefore responded similarly irrespective of whether the experience was given to the other animal or self, together suggesting that reduced expression of *Shank3* was associated with a diminished ability of neurons to represent social agency (i.e., the ability to distinguish other from self).

Controls for potential differences in sensorimotor and anxiety-related behaviors

One potential explanation for these findings could be that the HET animals simply displayed a generalized sensory perceptual deficit and may therefore not be able to differentiate between the aversive or appetitive stimuli themselves. Here, we found that although the HET mice did not show a preference for other animals undergoing aversive compared to appetitive conditions when using a control place-preference assay¹² (ANOVA, $p > 0.5$; Extended Data Fig. 6a), they did display a difference in respiratory rate (ANOVA, $p = 0.0031$; Extended Data Fig. 6b) to suggest that they were able to discriminate between these conditions. To further validate these observations, we also measured the animals' corticosterone levels and found that they were higher in both the WT and HET mice when observing another animal having an aversive compared to appetitive experience (ANOVA, $p < 0.05$). Although this increase in corticosterone levels was smaller in HET mice than in WT mice, the genotype of the animals had no independent effect on this change (two-way ANOVA, $p = 0.23$; Extended Data Fig. 6c). Lastly, we tested the responses of the mPFC neurons themselves to the other's identity. Here, we theorized that if difference in encoding in the HET mice was due to a nonselective sensory deficit, then we should also observe a diminished response when comparing familiar and non-familiar animals. However, 29.8% ($n = 39$ of 131) of neurons in the HET mice displayed a significant difference in response when paired with a familiar compared to non-familiar animal (i.e., independently of their experience) (bootstrap test, $p < 1.0 \times 10^{-9}$; Fig. 3c and Supplementary Fig. 4); together suggesting that the HET animals did not simply display a sensory perceptual deficit.

Another possible explanation for these findings could be that the HET mice displayed diminished engagement or interaction with the others. We therefore re-analyzed our data, but now only considering those periods in which the recorded animal was directly engaging with the other animal. However, we again found significantly fewer neurons in the HET mice that responded to the other's experience (11.8% and 35.8% for HET and WT mice, respectively; Chi-square test, $\chi^2(1) = 8.3$, $p = 4.1 \times 10^{-3}$; Fig. 3b) and fewer neurons that responded to the other's experience when confining our data to the time periods in which the recorded animal was in immediate proximity to the other (7.4% and 23.1% for HET and WT mice, respectively; Chi-square test, $\chi^2(1) = 6.0$, $p = 0.014$; Supplementary Fig. 2). We also evaluated for neurons that may have responded directly to physical contact and

found that 24.1% ($n = 27$) of the neurons in WT mice change their activity when the mouse specifically interacted with another. However, of these ‘physical-engagement’ neurons, only 5 overlapped with those that encoded the other’s specific experience and at a probability that was significantly lower than expected by chance (Chi-square test, $\chi^2(1) = 24.9$, $p < 1.0 \times 10^{-6}$). Similar results were also observed in the HET animals (Supplementary Tables 3–4), together suggesting that diminished ability of neurons to encode the specific experience of others or their social agency was not due to nonspecific sensorimotor effects.

Finally, we considered the possibility of generalized behavioral states such as anxiety. Using an elevated zero maze assay⁴², we find that the WT mice preferred the closed areas of the maze, spending approximately 77% of their time in these areas and at a probability significantly above chance ($n = 26$, t-test, $p < 0.00001$). The time spend in the open area, however, did not differ from that of the HET mice ($n = 27$; t-test, $p = 0.73$), nor did the number of visits (t-test, $p = 0.77$) or head-dipping events (t-test, $p = 0.92$; Extended Data Fig. 7). Therefore, consistent with prior reports^{22,33}, the HET mice generally do not display anxiety-related behavior.

Real-time restoration of *Shank3* expression leads to increase in sociability

While the above observations suggested an association between *Shank3* expression and neuronal encoding in the mPFC, they did not reveal whether or to what degree differences in neuronal encoding causally related to the animal’s social behavior. To address this question, we used FLEx switch-mediated control of *Shank3* expression while simultaneously tracking neuronal activity and social behavior in the same animals over time (Fig. 4a). TMX was used here to globally increase *Shank3* expression in the haploinsufficient animals. Increase in α , β and γ SHANK3 isoforms in the *Shank3^{fx/+}:CreER^{+/-}* mice was confirmed using brain synaptosome preparation and Western blot (not all three isoforms are affected equally³⁰). We also confirmed that SHANK3 levels after TMX did not increase in the *Shank3^{fx/+}:CreER^{-/-}* mice, which lacked the CreER recombinase gene (t-test, $t(12) = 2.63$, $p = 0.04$; Fig. 4b).

The HET mice displayed diminished sociability compared to the WT mice. Here, the sociability of the animals was assessed using a modified three-chamber task^{22,25,33,43} and was defined as a preference for another non-familiar mouse over a novel inanimate object. Overall, we find that the WT mice significantly preferred the other animal over the inanimate object (paired t-test, $t(94) = 4.1$, $p = 1.0 \times 10^{-4}$; Fig. 4c and Extended Data Fig. 8a–b) whereas HET mice displayed no such preference (prior to TMX treatment; paired t-test, $t(52) = 0.19$, $p = 0.85$). WT and HET mice did not differ otherwise in movement velocities (two-sample t-test, $t(217) = 0.48$, $p = 0.63$) or distribution of distances traveled (two-sample t-test, $t(217) = 0.32$, $p = 0.75$; Fig. 4d). The HET mice therefore displayed diminished sociability^{22,33}.

Next, to evaluate whether restoration of *Shank3* expression led to a change in social behavior, we examined the time-period after TMX administration (week 5) when SHANK3 expression had already increased^{30–32} and find that the HET mice now displayed a significant preference for the non-familiar animals (week 5, paired t-test, $t(56) = 3.6$, $p = 7.0 \times 10^{-4}$; Fig. 4c). This behavioral change was not observed in the *Shank3^{fx/+}:CreER^{-/-}*

mice lacking the CreER recombinase gene treated with TMX (paired t-test, $t(58) = 1.34$, $p = 0.19$; see additional controls below) nor in the *Shank3^{flx/+}:CreER^{+/-}* mice which possessed the CreER recombinase gene but in which a vehicle (corn oil) was delivered (two sample t-test, $t(15) = 1.10$, $p = 0.29$; Extended Data Fig. 8c,d). These findings therefore suggested that increase in sociability was specifically due to activation of *Shank3* expression rather than the passage of time.

Temporal dependency between *Shank3* expression, neuronal encoding and social behavior

Given these findings, we evaluated the dependency between change in social behavior and neuronal encoding by examining their day-by-day progression. Here, each HET animal underwent consecutive behavioral testing and neuronal recordings on alternating days after TMX administration (Supplementary Table 1). The neuronal and behavioral data were then time-aligned and compared over the course of two months.

First, focusing on neuronal responses in the *Shank3^{flx/+}:CreER^{+/-}* (HET) mice, we find that the percentage of neurons which encoded the other's experience increased from 9.2% ($n = 12$ of 131) prior to TMX to 35.5% after TMX ($n = 49$ of 138 recorded *after* 5 weeks; Chi-square test, $\chi^2(1) = 15.2$, $p = 9.4 \times 10^{-5}$; Fig. 5a). Conversely, the percentage of neurons that encoded the animal's own experience decreased from 38.2% ($n = 50$ of 131) before TMX to 19.5% after TMX ($n = 27$ of 138, 5 weeks; Chi-square test, $\chi^2(1) = 11.38$, $p = 7.4 \times 10^{-4}$). The *Shank3^{flx/+}:CreER^{-/-}* mice, by comparison, displayed no such changes in neuronal encoding (Chi-square test, $\chi^2(1) = 0.095$, $p = 0.76$ and $\chi^2(1) = 0.30$, $p = 0.58$ for other- and self-valence neurons, respectively). We also found no change in HET mice after TMX in the proportion of neurons that were simply task-modulated (Chi-square test, $\chi^2(1) = 0.25$, $p = 0.61$) or in the proportion of neurons that responded to familiarity (Chi-square test, $\chi^2(1) = 0.34$, $p = 0.56$).

Next, we examined the day-by-day relation between changes in neuronal encoding and change in behavior. A cross-correlation analysis that quantified the dot-product between social agency encoding (i.e., other-self ratio) and behavioral sociability (i.e., animal-object preference ratio) as a function of time revealed a positive and significant relation between social-agency encoding and sociability (cross-correlation analysis, $r = 0.88$, $p = 0.0041$; Fig. 5b). The peak time-lag between change in neuronal encoding and change in social behavior for these animals was 5-6 days (permutation test; $p < 0.01$; Fig. 5b, *bottom*). That is, an increase in the proportion of neurons that encoded the other's experience consistently preceded increase in the animal's sociability by approximately 0.75 weeks.

To further validate these findings and to evaluate the degree to which this correlation could be expected by chance, we performed a permutation procedure that randomized the days over which neuronal activity and behavior were recorded 1000 times (Methods). In other words, we asked whether, given the number of data points and days tested, what was the likelihood that this correlation could have been observed by chance. Here, we found that the cross-correlation values for shuffled data were significantly smaller than that observed in the actual data (i.e., r-value of -0.000086 ± 0.0000013 vs. 0.88; permutation test, $p < 1.0 \times 10^{-7}$).

Finally, to confirm that these changes were specific to *Shank3* expression and not simply explained by the passage of time or familiarity with the task, we examined the *Shank3^{fx/+}:CreER^{-/-}* animals who received TMX and experienced the same task conditions but whom lacked the CreER recombinase gene and therefore did not experience an increase in SHANK3. These mice did not show a change in agency encoding and no relation between social behavior and neuronal encoding (cross-correlation, $p > 0.25$; Fig. 5b, *right* and Fig. 6). Similarly, the WT mice did not display a change in sociability (permutation test, $p > 0.2$) or neuronal encoding (permutation test, $p > 0.2$) when tracked over time (Extended Data Fig. 8d) to suggest a simple time-related progression. Together, these findings therefore suggested that the increase in the HET animal's sociability after TMX was temporally dependent on changes in the encoding properties of mPFC neurons and their ability to represent the distinction between other and self.

Specificity of self and other encoding in the mPFC and social behavior

While the above findings revealed a relation between neuronal encoding and social behavior, they did not indicate whether or to what degree activity in the mPFC was sufficient to explain it. For example, it is possible that the responses of mPFC neurons reflect an indirect 'read-out' of activity or changes in other brain areas but played no direct role in the behavior. To address this possibility, we used a new endoxifen-driven Cre-lox recombination approach that allowed us to confine restoration of *Shank3* expression to the mPFC (Fig. 7). Here, these HET mice had the same *Shank3^{fx/+}:CreER^{+/-}* genotype (i.e., were age-matched littermates), underwent the same behavioral conditioning, performed the same three-chamber task and underwent recordings from the same cortical area as those that received TMX. But the increase in SHANK3 expression was now confined to the mPFC (Fig. 7b).

HET mice that underwent local endoxifen-driven Cre-lox recombination and restoration of *Shank3* expression in the mPFC also displayed a marked increase in sociability. As seen in Figure 7a, they showed a significant increase in preference for the non-familiar animals compared to inanimate objects over the 8 week course of testing (two sample t-test, $t(15) = -4.14$, $p = 0.0087$; >5 weeks). By contrast, no change in sociability was observed in HET (*Shank3^{fx/+}:CreER^{+/-}*) mice that had received saline injections into the mPFC (two sample t-test, $t(15) = -1.40$, $p = 0.18$; >5 weeks) and in *Shank3^{fx/+}:CreER^{-/-}* mice, which received endoxifen injection in the mPFC but which lacked the CreER recombinase gene and therefore did not demonstrate increased *Shank3* expression in the mPFC (two-sample t-test, $t_s(15) = -0.41$, $p = 0.69$; Fig. 7a, *right*).

Localized restoration of *Shank3* expression in the mPFC of HET mice also resulted in an increase in the proportion of neurons that encoded the others' experience over time ($n = 115$; $\chi^2(1) = 4.11$, $p = 0.043$). Here, the peak temporal dependency between change in behavior and change in neuronal encoding was at a time-lag of 0 days (r -value = 0.82, $p = 0.0073$; Fig. 7a, *bottom*). Thus, local activation of *Shank3* expression in the mPFC was associated with a near simultaneous change in neuronal encoding and social behavior (i.e., unlike the 5-6 day lag in HET mice after global restoration of *Shank3* expression). The difference in time-lag between local and systemic restoration, though, was not significant (permutation

test, $p > 0.4$). Taken together, these findings indicate that the increase in sociability observed after restoration of *Shank3* expression was both spatially-and-temporally (i.e., ‘time-causally’⁴⁴) dependent on these changes in neuronal encoding properties in the mPFC.

Last, to confirm these findings, we inhibited neural activity in the mPFC in WT animals. To this end, we either injected the reversible GABA agonist muscimol or saline into the same mPFC area on randomly alternating days. We found that after muscimol administration, the mice displayed no preference for another animal over an inanimate object (two-sample t-test, $t_{s(38)} = 0.66$, $p = 0.52$; Fig. 7c). The same mice receiving saline injection, by comparison, displayed significantly preference over the mouse over the object (two-sample t-test, $t_{s(58)} = 2.98$, $p = 4.19 \times 10^{-3}$). These observations therefore together suggested that neuronal activity in this area was indeed necessary for changes in the animals’ sociability.

Discussion

Together, our findings suggest a functional division of labor within the mPFC by which neurons not only represent another animal’s specific experience but also differentiate the other’s specific experience from the animal’s own. By simultaneously tracking variations in social agency, experience valence and identity for the same individual cells, we observe that mPFC neurons represent another animal’s positive and negative experiences differentially, and that they represent the other’s specific experiences distinctly from the animal’s own. Moreover, while these neurons were modulated by the social context of the task, they displayed little response when no other mouse was present.

Previous investigations have provided insight into how discrete aspects of another animal’s experience, such as receipt of reward, may be encoded by cells^{1–7,9}. The current study demonstrates how individual neurons concurrently encode ‘what’ another mouse is experiencing, ‘who’ is experiencing it and to ‘whom’ the experience belongs. In particular, these findings suggest a coding mechanism by which neurons in the mPFC represent information about a conspecific’s specific experience (i.e., whether it is aversive or appetitive) as well as to differentiate the conspecific’s experience from one’s own. They also suggest that these neurons are sensitive to the specific individuals involved (i.e., whether the conspecific is familiar or novel) but show little response to the aversive or appetitive stimuli themselves when paired with inanimate totems (Supplementary Fig. 3).

Individuals with ASD are often described as having inability to organize social information or appropriately interpret the state and experience of others^{14–19}. They may also have heightened sensitivity and preoccupation with self^{45,46}. The current study’s findings in a mouse model of ASD provide a prospective cellular substrate for these observations. In particular, we find that heterozygous disruption of *Shank3* expression is associated with a diminished proportion of mPFC neurons that encode the other social agent’s experiences, but that it is also associated with an increased proportion of neurons to encode the animal’s own. Moreover, by restoring *Shank3* expression in the mPFC and by evaluating the relation between neuronal response and social behavior, our findings suggest that diminished social behavior is temporally correlated both with the ability of the neurons to encode information about the experience of others as well as their social agency. The absence of such changes

in *Shank3* mutant mice with *CreER*^{-/-} constructs or in WT animals further confirmed that these changes were specific to *Shank3* expression.

Another notable finding was that focal restoration of *Shank3* expression in the mPFC was sufficient for increasing the animal's sociability. While loss of *Shank3* expression in the HET animals is largely evident throughout the brain²⁵, findings from the *Shank3* restoration and focal inhibition experiments suggest that the mPFC plays a specific causal role in the animal's social behavior. These findings also suggest that changes in the encoding properties of dmPFC neurons were not simply explained by activity changes elsewhere in the brain, and that the temporal dependency between neuronal encoding in the mPFC and changes in social behavior was likely specific. This interpretation is also supported by the lack of neuronal encoding or behavioral changes in mice lacking the CreER recombinase gene.

While we focused here on the mPFC, it is important to note that ASD is known to involve other cortical-subcortical circuits^{27,36}, and that various neurophysiological features such as cortical volume^{35,47} and connectivity²⁷ are also altered in ASD. Our findings also do not reveal how *Shank3* expression in mouse mPFC relates to other changes observed in *Shank3* mutant mice, such as the excitability and synaptic structure of neurons (or their restoration)^{24,29,30}. Moreover, animal models of social behavior, including the mice used here, do not fully recapitulate the complexity and natural process in which humans interact, or capture the wide spectrum of behavioral phenotypes that are affected by ASD^{14,48,49}.

Together, our study provides a detailed examination of the encoding properties of neurons in the mPFC and how their disruption relates to abnormal social behavior. The ability to track the relationship between neuronal encoding (i.e., how sensory information is represented by neurons) and social behavior (i.e., how animals respond to these sensory stimuli) in real-time and over months-long durations could provide a prospective tool for understanding the behavioral phenomenology of disorders such as ASD. Whereas prior investigations have focused on the molecular and anatomic underpinnings of such disorders (e.g., differences in the synaptic architecture and excitability of neurons)^{14-19,46}, our approach allows one to examine the encoding properties of neurons with prospective implications to the study of a broad variety of genetic disorders^{14,50}.

Methods

Animals

Animal housing.—All procedures were approved by the Massachusetts General Hospital IACUC and kept in strict accordance with the Harvard Medical School Institutional Animal Care and Use Committee guidelines. C57BL/6 male mice were housed on a 12-hour light/dark cycle and food and water was provided *ad libitum* except during experimentation. For the experimentation periods, mice were food restricted to a goal of 85% of initial weight. Littermates of the same sex were housed together in groups for 2-5 months. Experiments were conducted when mice were between 4-12 months old. Animals were randomly assigned to the various experiments; for the main tasks, no subjects were excluded from the analyses.

Animal preparation.—In order to create Cre-dependent FLE_x switch constructs capable of controlling *Shank3* expression, we bred *Shank3^{fx/+}:CreER^{+/-}* and *Shank3^{fx/+}:CreER^{-/-}* male mice by crossing *CAGGs-CreER⁺* males (Stock no.004453, Jackson Labs) with *Shank3^{fx/fx}* females (Stock no.028800, Jackson Labs). Creation of *Shank3^{fx/fx}* animals and *CAGGs-CreER* animals has been described previously^{30–32}. All of the main comparisons in our study were made between littermates. These included all combinations of (i) *Shank3^{fx/+}:creER⁻*, (ii) *Shank3^{fx/+}:creER⁺*, (iii) *Shank3^{+/+}:creER⁻* and (iv) *Shank3^{+/+}:creER⁺* constructs. They were also age- and sex-matched for control. To allow for consistency, females were only used in our study for breeding purposes and were not used for behavioral testing or recordings. Thus, for example, when crossing *CAGGs-CreER⁺* males with *Shank3^{fx/fx}* females, only male pups that were *Shank3^{fx/+}:creER⁺* were used for experiments. All mice were genotyped using samples from ear clippings (Transnetyx, Cordova TN). Additional confirmation of *Shank3* expression is described further below.

***Shank3* genotype and expression confirmation**

Tamoxifen administration: To induce *Shank3* expression in awake-behaving adult mice, tamoxifen was administered per *Mei et al*³⁰. Briefly, mice were given 7 mg/day of tamoxifen (TMX) when between a weight of 26-29g and 8mg/day when between the weight of 30-35g. TMX was dissolved in corn oil (Sigma Life Science) to a volume of 0.35-0.40 ml, depending on the dose of TMX. TMX was delivered with disposable oral gavage needles (Instech, 20ga x 38mm). TMX administration began at the age of 4 months and was given for five days. For control comparison, TMX was given to both the *Shank3^{fx/+}:CreER^{+/-}* and *Shank3^{fx/+}:CreER^{-/-}* mice.

Brain tissue preparation.—Mice were euthanized with pentobarbital (90mg/kg) followed by cervical dislocation and decapitation. The brains were subsequently extracted and sectioned. For histology and electrode localization, the brain was post-fixed in 4% periodate-lysine-paraformaldehyde (PLP) for 48 hours then transferred into a 30% sucrose in PBS solution for an additional 48 hours. For Western blot analysis, the brain was immediately placed in liquid nitrogen and stored at -80°C until use.

Synaptosome Preparation and Western Blot.—Synaptosomes were prepared using Syn-PER synaptic protein extraction reagent (87793, Thermo Fisher Scientific). Briefly, 1 ml of Syn-PER reagent was added to ~100 mg brain tissue sample then centrifuged at $1200 \times g$ for 1 min. The supernatant was collected and further centrifuged at $15000 \times g$ for 20 min. Synaptosomes were recovered from the pellet and re-suspended in the Syn-PER buffer. All buffers contained complete protease inhibitor cocktail (Roche Diagnostics) and protein phosphatase inhibitor cocktail I (Sigma Aldrich). Proteins were separated on 4–12% NuPage MES-SDS gels (Life Technologies) using 5 μg synaptosome loads, transferred to nitrocellulose, and incubated with a rabbit anti-Shank3 primary antibody⁵¹ (1:500; APZ-013, Alomone Labs) for 48 hours at 4°C . Blots were then washed with .05% Tween-containing tris-buffered saline (TBS-T) 3 times and incubated with goat anti-rabbit horseradish peroxidase secondary antibody (1:8,000; 32460, Invitrogen) for 1 hour at room temperature. Blots were washed again with TBS-T ($\times 3$) then incubated

with the WesternBright ECL HRP substrate (K-12045-D20, Advansta) for 5 minutes. Chemiluminescence detection was done using ImageQuant LAS4000 (GE Healthcare Biosciences) and quantification was performed using ImageQuant TL image analysis software (GE Healthcare Biosciences) using β -actin as a loading control. To further confirm antibody selectivity, in a separate control, we combined the rabbit anti-Shank3 primary antibody with SH3 protein prior to incubation. Novex sharp pre-stained protein standard (LC5800, Invitrogen) was used as molecular weight markers.

Histological confirmation

Histologic preparation.—For all histological experiments, animals were perfused transcardially with phosphate buffered saline (PBS) then 4% periodate-lysine-paraformaldehyde (PLP). Brains were extracted and post-fixed in 4% PLP for 48 h, washed 3 times in PBS, placed in 30% sucrose in PBS for 48h, and sectioned at 50 μ m thickness using with a vibratome (Leica VT1200S). Electrolytic lesions were used to confirm localization of the areas recorded from whereas fluorescent dye injection was used to confirm cannula localization.

Immunofluorescence confirmation of local increase in Shank3 expression.—For immunofluorescence staining, brain sections were blocked in a 0.2% Triton X solution (PBST) with 5% NGS solution (PBST-NGS) for 1h at room temperature followed by incubation in a rabbit anti-Shank3 primary antibody solution (1:200; Alomone APZ-013) in PBST-NGS for 72h at 4 °C. Sections were washed 4x in PBST for 10min at room temperature, then incubated in a Alexa 488-conjugated goat anti-rabbit secondary solution (1:500; Jackson ImmunoResearch 111-545-144) in PBST-NGS for 2h at room temperature. Sections were again washed 4x in PBST, then mounted onto glass slides, cover-slipped using Vectashield with DAPI (Vector Laboratories H-1200), and sealed with clear nail polish. All fluorescent images were taken using a fluorescence microscope (Keyence BZ-X710) with either CFI Plan Apo λ 2x, CFI Plan Apo λ 4x, or CFI Plan Apo λ 40x objectives.

Neuronal recording

Stereotactic implantation.—Multi-electrode microarrays were used for single-unit recordings (Microprobes for the Life Sciences). Electrodes were implanted under isoflurane anesthesia, followed by antibiotic administration and pain relief. The arrays were implanted under stereotactic guidance in the medial prefrontal cortex (mPFC), and were aligned to correspond to the dorsal anterior cingulate and pre-limbic (ACC/PL) areas using a Kopf microdrive. Each array contained 16 recording electrodes and 2 reference/grounds, with two rows of electrodes being placed in each hemisphere. Neuronal recordings began at least two weeks following surgery to allow recovery. The locations of implantations were confirmed by histology as described previously⁵².

Electrophysiological recordings.—A Plexon multichannel acquisition processor was used to amplify and band-pass filter the neuronal signals (150 Hz – 8 kHz; 1 pole low-cut and 3 pole high-cut with 1000 \times gain; Plexon Inc., TX). Signals were then digitized at 40 kHz and processed to extract action potentials in real-time by a Plexon MEA workstation. Putative neurons were required to separate clearly from any channel noise, to demonstrate

waveform morphology consistent with that of a cortical neuron, and to have at least 99% of spikes separated by a minimum refractory inter-spike interval of 1 ms. No multiunit activity was used.

Mean spiking activity in WT mice was slightly but non-significantly lower than for that of HET mice (3.37 ± 0.37 vs. 3.54 ± 0.31 spikes/sec, two-sample t-test, $t_s = -0.52$, $p = 0.61$) during neutral trials. Overall, we found no direct correlation between neuronal firing rate and neuronal response to self (two-sample t-test (self vs remaining neurons), WT: $t_s = 1.81$, $p = 0.073$, HET: $t_s = 0.55$, $p = 0.58$) or other (two-sample t-test (other vs remaining neurons), WT: $t_s = 0.17$, $p = 0.87$, HET: $t_s = 0.48$, $p = 0.63$) for either WT or HET animals. Using the waveform morphologies of neurons and their inter-spike intervals⁵³, we identified neurons as putative pyramidal neuron or interneuron. For the pyramidal neurons in the WT mice, 46% (n=25) responded to self-experience and 54% (n=29) to other-experience, whereas for the pyramidal neurons in the HET mice 83% (n=39) responded to self-experience and only 17% (n=8) to other-experience. For interneurons in the WT mice, 50% (n=4) responded to self-experience and 50% (n=4) to other-experience whereas, for interneurons in the HET mice 73% (n=11) responded to self-experience and only 27% (n=4) to other-experience. Overall, these numbers suggested that there was no significant difference in the distribution of neurons displaying self- or other-related responses based on neuronal subtype (Chi-square test, $\chi^2(3) = 3.5$, $p = 0.33$).

Localized manipulation of *Shank3* expression and neuronal activity in the mPFC

Endoxifen injection.—To elicit a local increase in *Shank3* expression in the mPFC, we injected either endoxifen (Sigma-Aldrich E8284) or a saline control in *Shank3^{flx/+}:CreER^{-/-}* and *Shank3^{flx/+}:CreER^{+/-}* mice. Here, we dissolved the endoxifen in 20% DMSO at a concentration of 4mM to allow for injection⁵⁴. Isoflurane-anesthetized mice were head-fixed on a Kopf stereotaxic frame, followed by bilateral craniotomies lateral of the sagittal suture and anterior of bregma. Using a syringe pump (Harvard Apparatus PHD Ultra) with 10 μ L syringes (Hamilton 1700) and connected to 34 gauge needles (WPI NanoFil) by PE-10 and fused silica capillary tubing, we injected 350 nL of 4 mM endoxifen or saline bilaterally (three injections per hemisphere, six total). We used the following coordinates relative to bregma: 1) AP +1.78 mm DV -2.00 mm ML +/-0.30mm, 2) AP +1.34 mm DV -2.00 mm ML +/-0.30 mm, 3) AP +0.86 mm DV -2.00 mm ML +/-0.30mm. A small amount of cranioplastic cement (Ortho-Jet, Lang Dental) was placed on the skull to cover the craniotomy. Animals were allowed to recover for 2-3 days before behavioral experiments.

Muscimol injection.—For muscimol injections, 8 WT mice were implanted bilaterally with custom 8 mm 23 gauge guide cannulae fabricated from hypodermic needles (Covidien 8881850310). For the surgeries, isoflurane-anesthetized mice were head-fixed on a Kopf stereotaxic frame, followed by small bilateral burr holes centered at AP +1.34 mm and ML +/-0.35 mm relative to bregma. Two bone screws were then placed posterior to bregma for support. Using a 3D-printed custom bilateral cannulae insertion tool attached to a Kopf cannula holder (1776-P1), the cannulae were placed 1 mm above the injection area of interest. Custom made 8 mm 29 gauge stylets were inserted into the guide cannulae after

the surgery to prevent clogging. Animals were allowed to recover for at least one week before behavioral experiments. To transiently locally inhibit neural activity in the mPFC, we injected 350 nL of muscimol (1 mg/mL; Tocris 0289) or a saline control at a rate of 250 nL/min via a syringe pump (Harvard Apparatus PHD Ultra) with 10 μ L syringes (Hamilton 1700) connected to the internal cannulae with PE-10 and fused silica capillary tubing. We waited three to five minutes after the end of the injection to allow for drug diffusion before removing the internal cannulae. All injections were verified histologically by injecting 100 nL DiI through 9 mm 30 gauge internal cannulae.

Animal tracking

The animals were recorded using a ceiling mounted digital camera (Canon Vixia, HF R500). Each animal's position within their enclosure, movement velocity and head-direction were determined in an automated, genotype-blinded manner using Ethovision XT (Noldus). For additional cross-validation, a blinded tester visually inspected each trial (as recorded by the digital camera), and corrected any discrepancies in the above automated parameters using custom-written scripts in MATLAB (Mathworks). Neuronal and behavioral data were aligned using a NI DAQ and Plexon MEA system.

Social task designs

Paired interaction task.—For neuronal recordings, pairs of mice were placed in rectangular enclosure (24 \times 30 \times 32 cm), separated by a translucent barrier that provided full view of the other, and that was permeable to smell and sound. To evaluate for the selectivity of neuronal response, we focused on three main social features that defined their paired interaction. These included social *agency* (self vs. other), *experience* valence (positive vs. negative) and *identity* (familiar vs. non-familiar). All trials lasted 20 seconds, with trial onset being defined as the time point the recorded mouse was introduced into its enclosure. The trials were separated by a one-minute period, in which the recorded mouse was placed in a separate isolated ante-chamber. The stimuli were removed and the enclosure was cleaned after each trial. The task conditions were therefore counterbalanced so that an equal number of self- and other- trials as well as an equal number of appetitive- and aversive- conditions were performed and the conditions were given in pseudo-random fashion.

To first evaluate for neuronal responses that may be selective for *agency* (i.e., self vs. other), we randomly varied whether the experiences were given to the other animal or to the recorded animal itself. For example, the other animal may be given an appetitive stimulus on one trial whereas the recorded animal may be given an appetitive stimulus on another. Second, to evaluate for responses that may be selective to the specific *experience* (i.e., rather than simply any salient sensory cue), we used food reward (odorless pellets, 8 mg; BioServ) as the appetitive experience and a narrow tube enclosure as the aversive experience. Here, we selected these stimuli because the other's experiences needed to be clearly apparent and salient to the recorded mouse. We also needed to limit potential confounds such as associative learning which changes over the course of training (e.g., as in fear conditioning through tone-shock pairing) and which does not easily allow for the dissociation of self-other agency²⁸. Finally, we varied the other's social *familiarity* (i.e. familiar or non-familiar mice) in order to evaluate for the effect of social context on the animal's interaction. Here,

familiar mice were defined as those that had been pair-housed with the recorded mouse for four weeks or longer. A non-familiar pairing was only considered valid if the other mouse had not been seen by the selected recorded mouse in the last four weeks, and two or fewer such pairings had ever been made in total.

Therefore, taken together, an example trial sequence may be that on trial #1, the recorded mouse is paired with another familiar mouse that is subject to an aversive stimulus whereas, on trial #2, it is paired with a non-familiar mouse that is subject to an appetitive experience and, on trial #3, it is itself subject to an aversive experience, *etc.* All trial combinations (i.e. based on variations in agency, experience and familiarity) were repeated such that recording sessions consisted of 45 trials including sensory and behavioral controls.

Three-chamber task.—For behavioral testing, a standard three-chamber task was used as described previously⁵⁵. Here, we tested the animal's sociability as defined by the animal's preference for a non-familiar (i.e. novel/stranger) mouse over novel inanimate object. Briefly, the mice were placed in a three-chamber apparatus for one hour the day prior to testing for acclimation. The mice were then subjected to three conditions in the following order: an empty vs. empty chamber, a non-familiar mouse vs. a novel inanimate object and then a repeated condition, with switched sides, to limit the possibility of spatial bias. Non-familiar mice were housed separately and were not used more than once. The positions of the animals were automatically tracked as described further below. Trials in which tracked mice did not leave the starting, central chamber or trials in which tracked mice were immobile in any particular chamber for greater than 50% of the recorded time were excluded. The three-chamber enclosure measured 45 × 60 × 30 cm.

Place preference/avoidance task.—To understand whether animals recognized another animal's experience and to allow for comparison with our main results, the animals performed a place preference/avoidance task⁵⁶. Here, the test subjects were presented with two age and sex-matched familiar conspecific partners undergoing either an appetitive or aversive experience (Extended Data Fig. 6). The test subjects were placed into clear arenas (30 x 30 x 30cm, L x W x H) that included two closed-off areas positioned on opposite sides. Each area contained a familiar conspecific animal and had an open mesh on the front to allow for interaction. For each trial, the test subject was first allowed to habituate to the arena for five minutes. The two familiar conspecifics were then introduced into each of the two closed-off areas, with one animal undergoing an appetitive experience (food-bated enclosure) and the other undergoing an aversive experience (confined tube enclosure) for five minutes. In order to control for possible location bias by the test subjects, the location of the animals undergoing the appetitive or aversive experiences were counter-balanced across the opposite areas. Further, to determine whether the animals perceived the other's experience rather than simply the appetitive or aversive stimuli themselves, we also included control trials in which the two conspecifics were replaced with inanimate totems that were “experiencing” the same appetitive (food-bated) or aversive (confined tube) conditions. Here, we calculated the test subjects' preference as the percentage of time in which they were within 3 cm of the appetitive vs. aversive areas over the course of the trial. This allowed us to explicitly determine whether the animals could discriminate between the

appetitive and aversive conditions. In other words, we would expect the HET mice to display diminished investigative or approach behaviors towards the others under any conditions; irrespective of whether the other animals are specifically having an appetitive or aversive experience.

Respiratory rate recordings.—For real-time respiratory rates (RR) recordings, shaved test subjects were fitted with a mouse jacket (Lomir Biomedical) and pulse sensor (World Famous Electronics) connected to an Arduino via a commutator. These animals were habituated to the jackets for one hour per day for at least five days to ensure their behavior were not hindered. Using custom Python code and built in Arduino serial port, raw physiological data were collected and digitized at 500Hz. This data was then band-passed at 1-5Hz. To evaluate RR during enclosure-investigation, we took non-overlapping 1s samples while animals were investigating either enclosure and fit a sinusoidal wave to the data. RR for each trial were calculated as mean frequency of the fitted wave. Results across counter-balanced trials were averaged for each subject. We excluded RR data for trials where animals failed to investigate one of the two enclosures more than 10s.

Additional task controls

Controls for sensory related responses.—Several controls were performed to evaluate for other sensory-related aspects of the task. First, to test whether neuronal responses to the other's experience reflected the aversive and appetitive stimuli themselves (i.e., independently of the social partner), we performed a task control in which we replicated the experiments but now replaced the partner with an inanimate object. Therefore, all aspects of the task, including presentation of the appetitive and aversive stimuli were the same, but in the absence of a social partner. Second, to evaluate whether neuronal responses could be explained by factors indirectly related to the stimuli, such as the smell emitted when the other was eating, we added a second set of controls in which the other animal was given the same aversive and appetitive stimuli. Here, however, direct visual access to the other was blocked with an odor- and sound-permeable barrier (a darkened, non-translucent version of the dividing plastic barrier). Third, all trials were evaluated in relation to a neutral condition in which neither an appetitive nor aversive experience was given.

Controls for motor related responses.—We evaluated for motoric differences in the recorded animals that could potentially contribute to neuronal response. To this end, and as detailed above, we continuously recorded each animal's position and tail-head orientation. We then sub-selected only time periods in which the animals were in direct view or close proximity to the other. Here, '*direct view*' was defined by the recorded animal's head being within a 20 degree range of the other. '*Proximity*' was defined as the recorded animal's center body position being within the half of the recording enclosure closest to partner mouse (i.e., the animals were near vs. far from their partner). Second, in a separate set of controls, we evaluated baseline differences in motor behavior. Here, we evaluated the animal's movement velocity as well as the distribution of distances traveled when placed in a separate empty enclosure.

Elevated zero maze.—To determine any *Shank3*-dependent differences in anxiety phenotype, HET and littermate WT animals were first habituated to a well-lit open field arena for one hour on at least two separate days. A custom elevated zero maze apparatus (lane width 10cm, diameter 60cm, height 60cm) was divided into four equal quadrants, where two opposite quadrants were enclosed by clear acrylic walls (height 30cm). Animals were introduced into one of the two closed areas and allowed to freely explore for 5min, while their position were tracked with EthovisionXT 12 (Noldus). Anxiety-like behaviors were calculated based upon the number of open area visits, percent time spent in the open areas, and percent time spend head dipping in open areas.

Corticosterone.—We measured corticosterone levels as a physiological index of the animals' ability to differentiate between the other's aversive and appetitive experiences. A separate cohort of HET and littermate WT animals performed a modified version of the paired interaction task as described above. Here, pairs of familiar conspecifics performed the task that included two main variables: social agency (self vs. other) and experience valence (positive vs. negative). Unlike the main task, animals performed one 10min trial per day. At least 7 days of rest were given in between trials in order for the mice to recover blood loss. Upon completion of each trial, animals were briefly anesthetized with 5% isoflurane for less than 20s and ~150 μ L blood samples were collected with a facial vein puncture (Animal Lancet; Goldenrod). The blood was centrifuged and plasma was collected and stored at -80°C . Plasma corticosterone levels were measured using an ELISA kit according to the manufacture's protocol (ADI-900-097; Enzo Life Sciences) and a microplate reader at 405nm (Absorbance 96; Enzo Life Sciences)^{57,58}. Contaminated samples were excluded.

Statistical analyses

Analysis of Variance.—To evaluate the responses of individual cells, a 3-way analysis of variance (ANOVA) with factors for agency (self, other), experience valence (aversive, appetitive) and partner identity (familiar, non-familiar) was conducted on a *per* neuron basis across all tested trials. *Post-hoc* analysis was conducted at one through three factor levels to determine whether a neuron significantly differentiated between trial types (i.e. encoded for a particular factor) while controlling for multiple comparisons. For example, a neuron would be considered to encode the valence of a partner's experience if other-aversive and other-appetitive trials differed significantly from one another by *post-hoc* analysis. Significance was set at $p < 0.0125$ for the primary reported findings. Thus, a neuron subtype would be 'other-experience' encoding if it displayed a consistent difference in activity when the other animals were having an aversive vs. appetitive experience (3-way ANOVA, $p < 0.0125$). On the other hand, the neuron would be considered as 'self-experience' encoding if it displayed a consistent difference in activity when the recorded animals themselves were having an aversive vs. appetitive experience. Trial firing rates were calculated using a spike density function with a sliding Gaussian and a standard deviation of 250ms. To evaluate for differences in the proportion of neurons within the different recorded populations, we used a standard Chi-square test ($p < 0.05$). The test was used to determine whether there was a significant difference in the number of neurons that responded to a particular main effect (e.g., self vs. other agency). Task-modulated neurons were considered across the recorded neuronal population when comparing main effects.

Neural population modulation.—To evaluate for differences in population response, we examined the distribution of firing activities across cells *per* condition. For example, to determine the degree to which neurons were modulated by differences in another's experience, we would subtract the mean firing rates recorded during trials in which the other was given an appetitive experience from the mean firing rates recorded during trials in which the other was given an aversive experience on a *per* cell basis. Analogous analysis was completed for self-trials, such that each neuron had two firing rate values (one for other-trials and the second for self-trials). Differences in the distribution of neural responses were then compared using a 2-dimensional 2-sample Kolmogorov-Smirnov test ($p < 0.05$). A one sample t-test was used to analyze whether a neuronal population exhibited a preferential increase or decrease in firing rate. A two-sample t-test was used to compare whether neuronal populations significantly differed from one another along a particular axis (i.e. during self or other trials only). Finally, to evaluate for changes in neural population response over time, analyses were repeated using successive 5s windows advance from trial onset to the end of the trial. Peak difference in neural population response was determined by permutation analysis (5000 random permutations).

Modeling and decoding analysis.—To determine the degree to which the trial conditions could be predicted from neuronal activity on a *per*-trial basis, we used a linear decoder (Fisher discriminant with quadratic boundaries) as described previously^{59–62}. For the tested conditions (e.g., other aversive vs. appetite experience), the ratio of the variance in neuronal activity between variables was compared to the variance within conditions based on:

$$S_W^{-1} S_B v = \lambda v,$$

whereby S_W and S_B are the within condition scatter matrices and between condition scatter matrices, respectively. The prediction vector v corresponds to the largest eigenvalue of the matrix on the left-hand side of the equation. The prediction vector defines a projection of the recorded activity into a scalar unit. For validation, we divided the neuronal data into a training set consisting of 80% of the trials and tested the accuracy of the prediction on the remaining 20% of trials for validation. This operation was repeated 1000 times using a random sampling of the total trials. A bootstrapping procedure was then used to determine whether the neuron could discriminate between conditions at an accuracy that was greater than chance. As before, we used an FDA correction to account for repeated comparison across conditions (i.e. experience valence, social agency and identity; $p < 0.0125$).

Model fitting and cross-correlation analysis.—To evaluate for temporal changes in relation to TMX or endoxifen administration, we used the ratio of other-to-self encoding neurons for the neuronal point-data and the ratio of time spent in the three-chambers (NF-F or NF-T) for the data point-data. Here, we used the *ratio* of two values rather than simply a single value (e.g., the proportion of other-encoding neurons) since the latter could vary based on quality of recordings or simple differences in motoric behavior/motivation from week-to-week. The temporal dependency between neuronal and behavioral data was assessed by cross-correlation analysis (i.e., the cross correlation between two series as a

function of the temporal displacement of one relative to the other). Here, the two time-series y_{1k} and y_{2k} were lagged by $k = 0, \pm 1, \pm 2, \dots, \pm \text{length}_{\min}(y_1 | y_2)$, and cross-covariance, c , was calculated for each possible pairing $(y_{11}, y_{21}), (y_{12}, y_{22}), \dots$ etc, by,

$$c_{y_1 y_2}(k) = \begin{cases} \frac{1}{T} \sum_{t=1}^{T-k} (y_{1t} - \bar{y}_2) \\ \frac{1}{T} \sum_{t=1}^{T-k} (y_{2t} - \bar{y}_1) \end{cases}$$

whereby $T = \text{length}_{\min}(y_1 / y_2)$, and \bar{y} is the sample mean of the series across time. Cross-correlation was in turn calculated for each cross-covariance value $r_{y_1 y_2}(k) = c_{y_1 y_2}(k) \times (e_{y_1} e_{y_2})^{-1}$ where e_{y_1} and e_{y_2} are the square root of variance. Optimal lag time was based on k with the greatest cross-correlation estimate. Because neuronal recording and behavioral testing was performed on alternating days over the two-month course of testing, time-lags were calculated in 2 day intervals. Significance of cross-correlation was determined by permutation test ($n = 1000$; $p < 0.01$).

Neural population decoding.—A supervised learning approach was used to quantify the degree to which the population's activities were informative of each the three primary task features that defined the animals' experiences. The approach was designed in three parts. First, we constructed support vector machines (SVMs) with nonlinear kernels in order to map the population's activity patterns onto high-dimensional feature spaces whereby

$$\min_{w, b, \zeta} \left(\frac{1}{2} S^T S + C \sum_{i=1}^n \zeta_i \right)$$

subject to

$$y_i (S^T \phi(x_i) + b) \geq 1 - \zeta_i$$

Here, $y \in \{1, -1\}^n$ corresponded to the stimulus conditions (e.g., appetitive vs. aversive condition), x corresponded to the neural activity; $\zeta_i = \max(0, 1 - y_i(w x_i - b))$ and C is the regularization factor. As described previously⁶²⁻⁶⁴, these approaches are well-suited for identifying neuronal 'subspaces' in relation to specific task features and account for potential nonlinearities in the response properties of the neurons. Second, to determine the decoding accuracy of the models, we used trials randomly sampled and bootstrapped from the validation data (20% of randomly selected trials). Finally, to determine significance, this process was repeated 1000 times and compared to models trained on neuronal data that was randomly shuffled ($H_0=50\%$ chance; permutation test, $p < 0.001$).

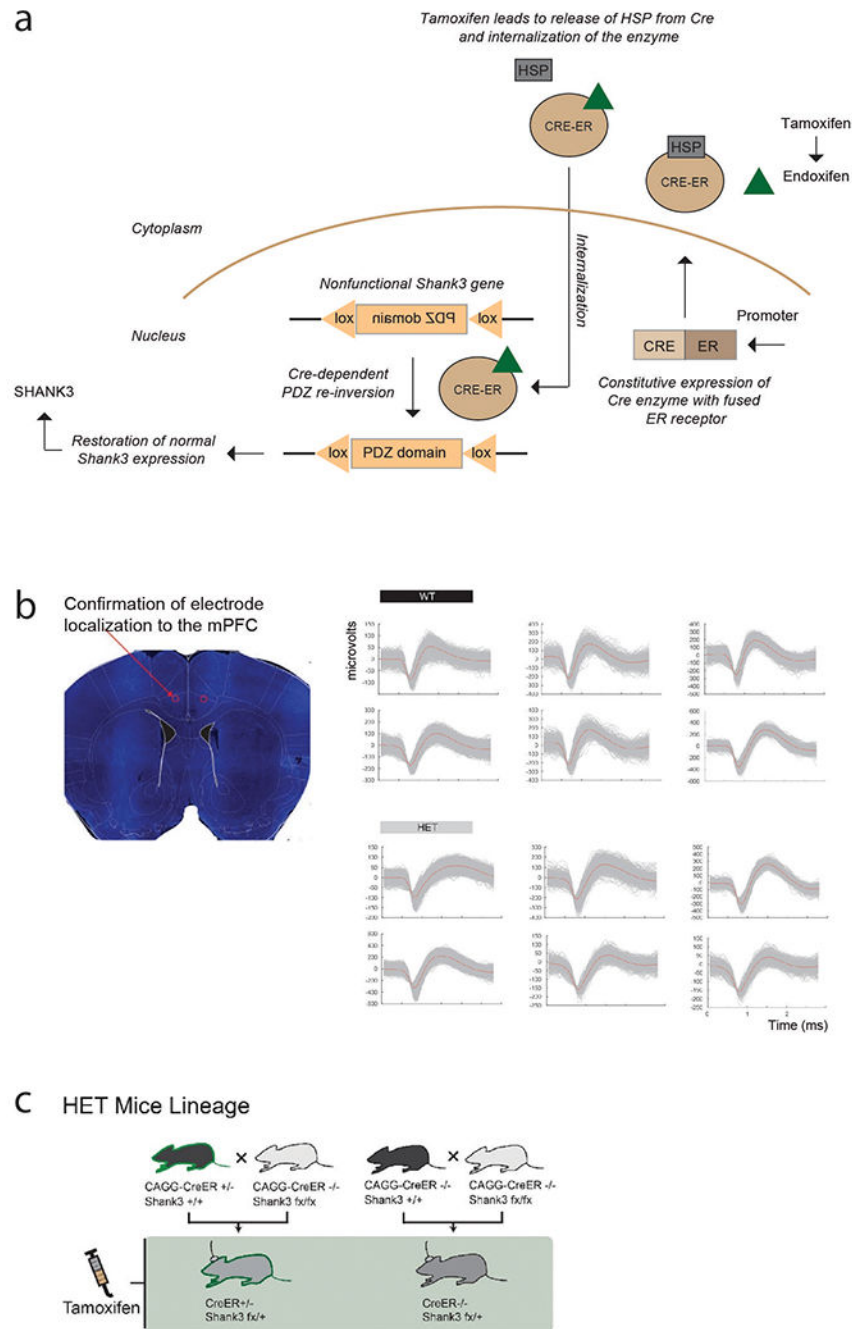
Statistics.—No statistical method was used to predetermine sample size. Statistical analyses were conducted in MATLAB (Mathworks). Significance was set at $\alpha < 0.05$ for all statistical tests unless otherwise indicated. The data distribution was assumed to be normal,

but this was not formally tested. Two-tailed tests were used unless otherwise indicated. Permutation tests were also used to avoid assumptions about the distributions of the data. Data are expressed as mean \pm s.e.m unless otherwise indicated.

Blinding

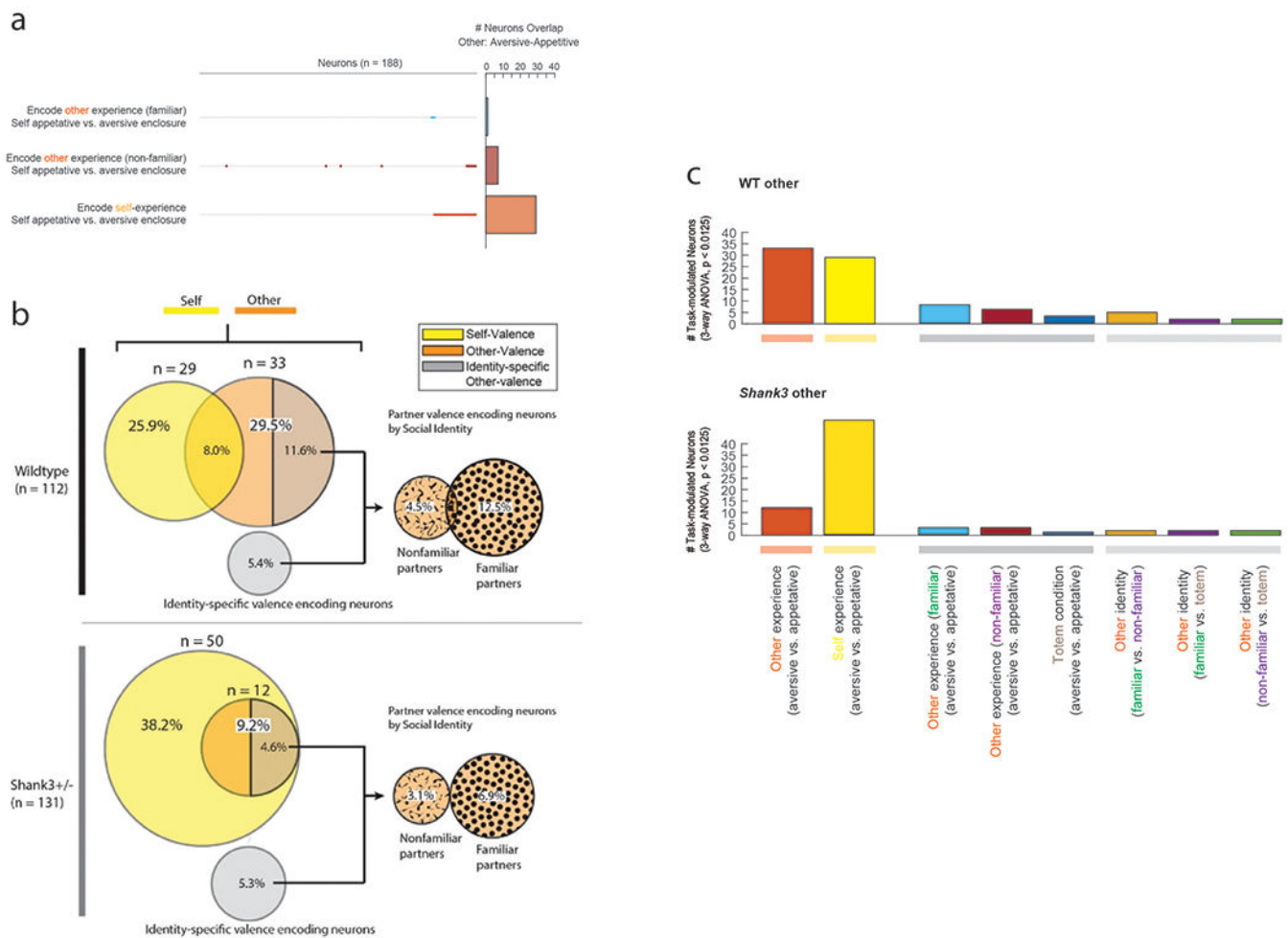
All data collection and analyses were performed blind to animal genotype (WT vs HET), gavage (TMX vs corn oil); injection (saline vs endoxifen; saline vs muscimol). All analyses were performed blind to trial condition (appetitive vs aversive).

Extended Data

**Extended Data Fig. 1. Animal preparation and electrophysiological recordings**

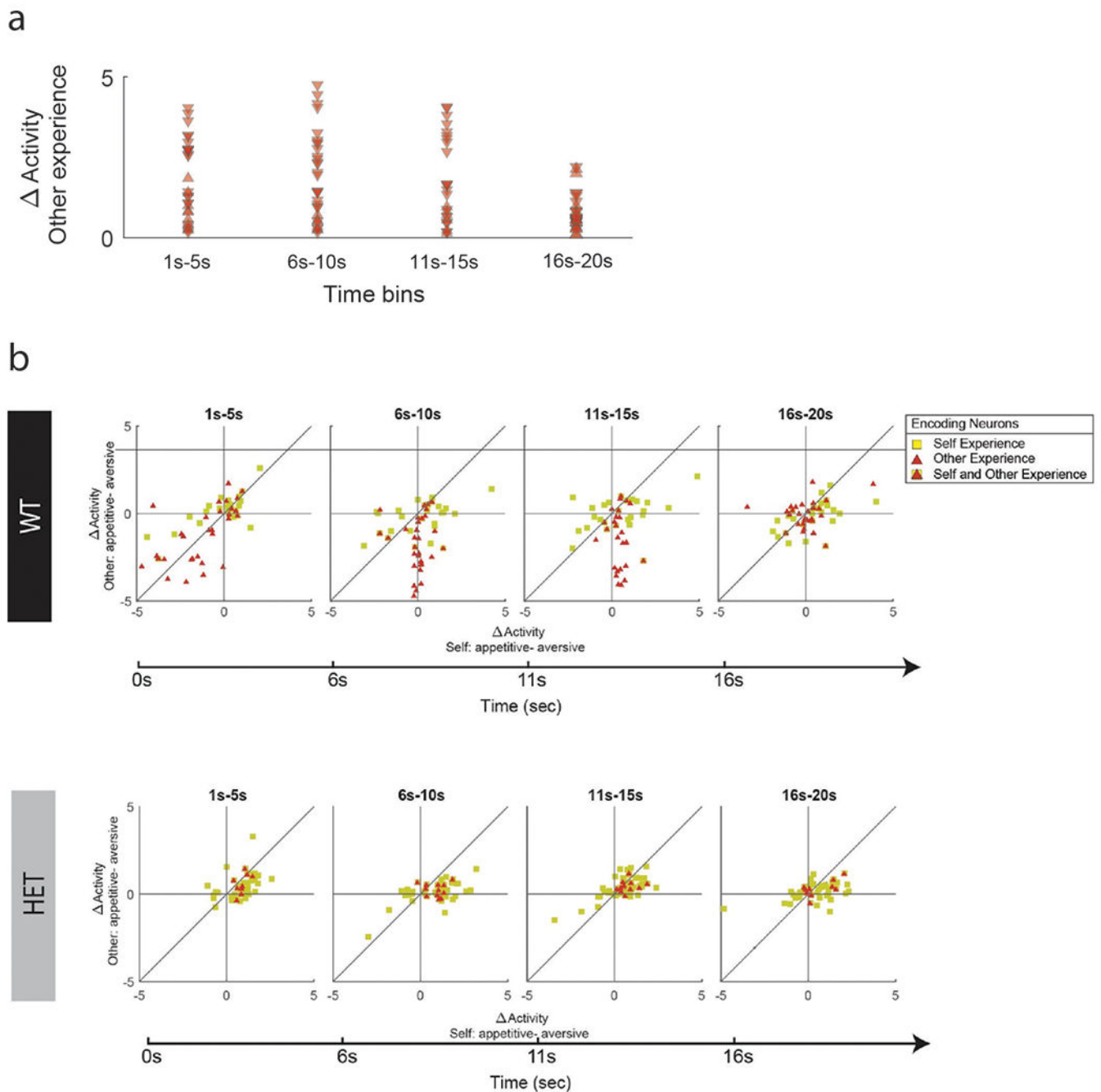
a. Illustration depicting the Cre-lox system and its use in restoring *Shank3* expression in adult mice. Cre-dependent FLEX switch constructs capable of controlling *Shank3* expression are made by creating *Shank3^{flx/+}:CreER^{+/-}* mice in which Cre recombinase (CRE) is fused to an estrogen receptor (ER) protein. Here, an inverted PDZ domain (the membrane anchoring portion of SHANK3) is 'floxed' by two lox sites and, therefore,

renders the mice heterozygous *Shank3*. CRE-ER is expressed constitutively. Normal *Shank3* expression is restored by either delivering systemic TMX or local endoxifen to the mice. Binding of endoxifen to the ER receptor leads to dislodgement of the associated heat shock protein (HSP) from CRE which then allows it to enter the nucleus. Once in the nucleus, CRE leads the floxed PDZ sequence, flanked by two loxP sites, to be inverted through recombinase-mediated cassette exchange. Finally, inversion of the PDZ sequence leads to normal production of SHANK3. **b.** Representative waveform morphologies and recording locations. Displayed *left* is a histological section with bilateral electrolytic lesions indicating the microelectrode recording locations from the medial prefrontal cortex (mPFC). *Right* are representative waveform morphologies from six putative neurons recorded each from the mPFC of WT and HET mice. **c.** *Shank3^{fx/+}:CreER^{+/-}* mice were created by crossing *CAGG-Cre-ER^{+/-}* with *Shank3^{fx/fx}* mice. *Shank3^{fx/+}:CreER^{-/-}* mice were created for control comparison. Only littermates were used for comparisons and all were male mice.



Extended Data Fig. 2. Neuronal responses to experience valence, agency, and identity
a. Neurons that responded to other’s experiences based on variations in the recorded animal’s own enclosure. Overlap in encoding across neurons is displayed on the *left* and their total numbers are shown to the *right*. Overall, only a few neurons (n = 7) displayed

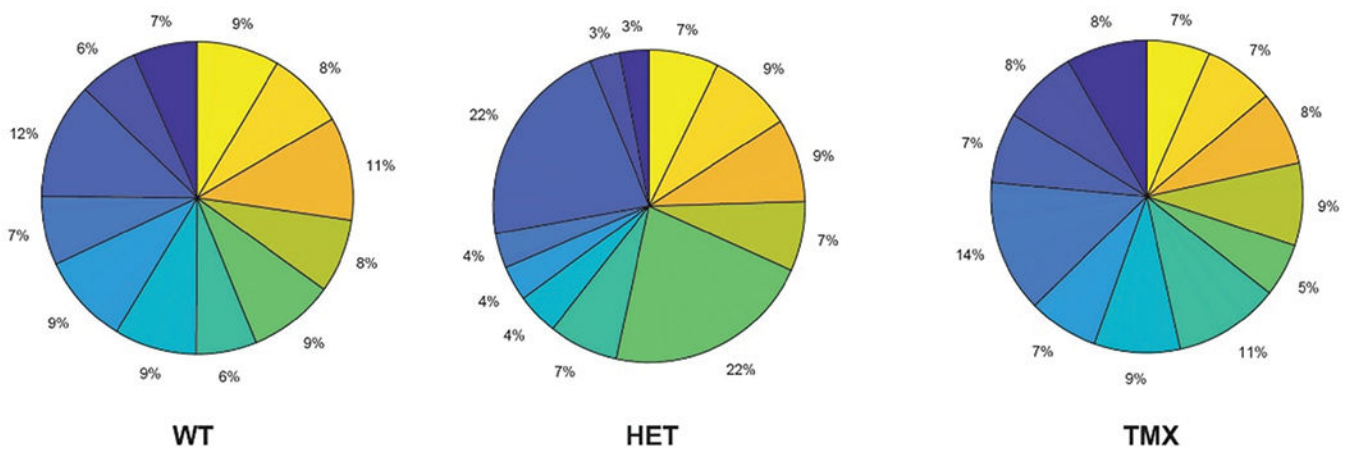
a difference in response to the other's experience based on which specific enclosure the recorded animal was simultaneously placed in (three-way ANOVA, $p < 0.0125$ with *post-hoc* comparisons). **b.** Neuronal responses based on the other's social identity. Breakdown of neurons that responded to the other's identity (familiar and non-familiar) and their relation to neurons that responded to the other's experience. On the *left* are the distribution of neurons that responded to the other's experience. While many of the neurons that responded to the other's experience also responded to the other's identity (11.6%), some neurons responded to the other's identity alone (5.4%; i.e., irrespective of the other's experience). On the *right* are the distribution of neurons that responded to the other's identity based on whether the other animals were specifically familiar or non-familiar to recorded mouse. **c.** Relative proportions of all task modulated neurons using a three-way ANOVA that accounted for all terms describing experience valence (positive vs. negative), social agency (self vs. other) and social context/identity (familiar vs. non-familiar vs. totem) with *post-hoc* comparisons and correction for multiple comparisons at a $p < 0.0125$. To the *left* are the numbers of neurons that responded to self- or other-experience valences. To the *right* are the remaining top six most common feature combinations. For example, while many neurons responded to another's aversive vs. appetitive experience, some neurons only responded to the other's experience when they were familiar to the recorded animal. Other neurons, by comparison, only responded to differences between familiar vs. non-familiar animals while displaying little modulation to their specific experience. In total, 112 neurons and 131 neurons displayed task-related modulation in the WT and *Shank3* mice, respectively.



Extended Data Fig. 3. Neural modulation over the course and within the trials

a. Absolute difference in activity (Z-score) for other experience *per cell* over the time course of the trial. Here, neuronal responses are broken down into 5 second intervals in order to illustrate the time progression of neural modulation. While there was a slightly lower degree of modulation at the very end of the trial, this difference was not significant (repeated measures ANOVA, $p > 0.2$). Similarly, we find no difference in the total number of neurons that displayed significant modulation over the course of the trial (Chi-square, $p > 0.5$). The directions of the arrows indicate whether responses were stronger for the aversive

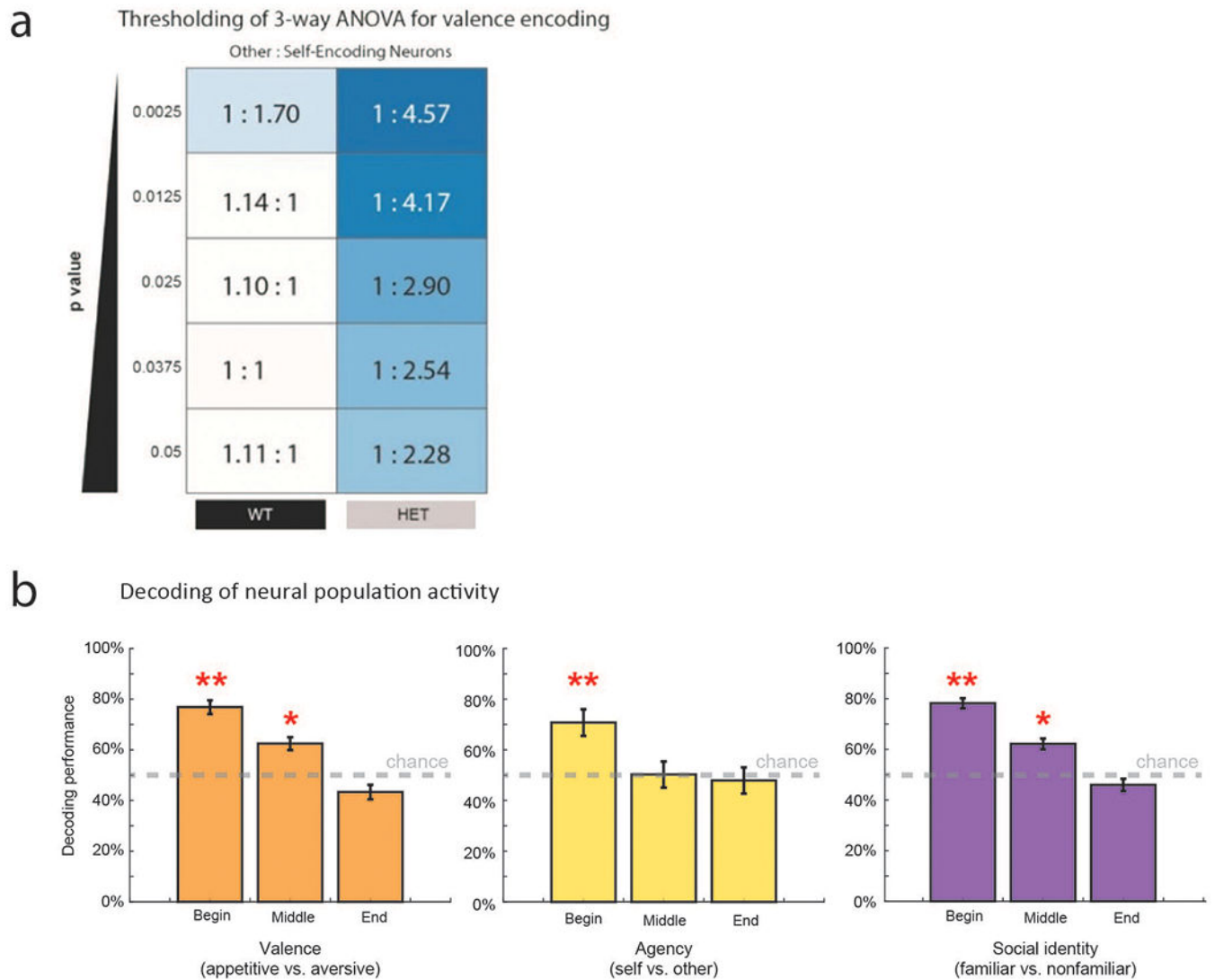
(point up) compared to appetitive (point down) other experiences. To evaluate the effect that the animal's own prior experience may have had on neuronal encoding of the other's experience, we also compared all possible transitions combinations between self- followed by other-experience valence (e.g., self-appetitive followed by other-aversive, self-aversive followed by other-aversive, etc.). Here, we find that only 2 neurons in the WT mice that responded to other-experience were also affected by specific past self-experience (three-way ANOVA, $p < 0.0125$). In other words, the animal's own experience in one trial influenced neuronal responses the following trial in only 1.8% of the neurons. Similarly, we find only 1 neuron in the WT mice whose response to self-experience was also affected by past other-experience; together suggesting that the past trial condition did not influence neuronal response under this task. **b.** This Figure follows the same convention as in Fig. 3d. Here, however, neural activity (Z-score) are divided into successive, non-overlapping 5 second windows over the 20 second trial-span.



Extended Data Fig. 4. Proportional contribution of neurons across animals

To allow for comparison across animals, the relative contribution of each mouse to the overall group ratio for self vs. other encoding are shown in percentages. Thus, a larger percentage means that they contributed relatively more to the proportion of other-encoding neurons whereas a smaller percentage means that they contributed relatively more to the proportion of self-encoding neurons. Additionally, we performed a within-group vs. between-group comparison to evaluate more directly whether differences in the proportions of neurons could be explained by potential dissimilarities in recording quality or variations in anatomical localization between animals. For example, if recordings were indeed made from slightly different areas or subpopulations of cells, then we should observe a similar variance between individual animals that belonged to the same genotype (i.e., HET) compared to between individual animals that belonged to the different genotypes (i.e., HET vs. WT). In other words, variability in the proportions of neurons found between one WT animal and another WT animal should be similar to that found between one WT animal and another HET animal. Examining the proportion of neurons that responded to other-aversive vs. other-appetitive experience, however, we find that the difference between animals within the WT group was significantly smaller than that between the WT and HET group (two-sample F-test for equality of variance; $f\text{-stat} = 0.20$, $p = 0.0076$). These

observations therefore support the main findings and controls further below that those differences in encoding properties between animals were due to differences in *Shank3* expression rather than a systematic variation in subpopulations sampled.

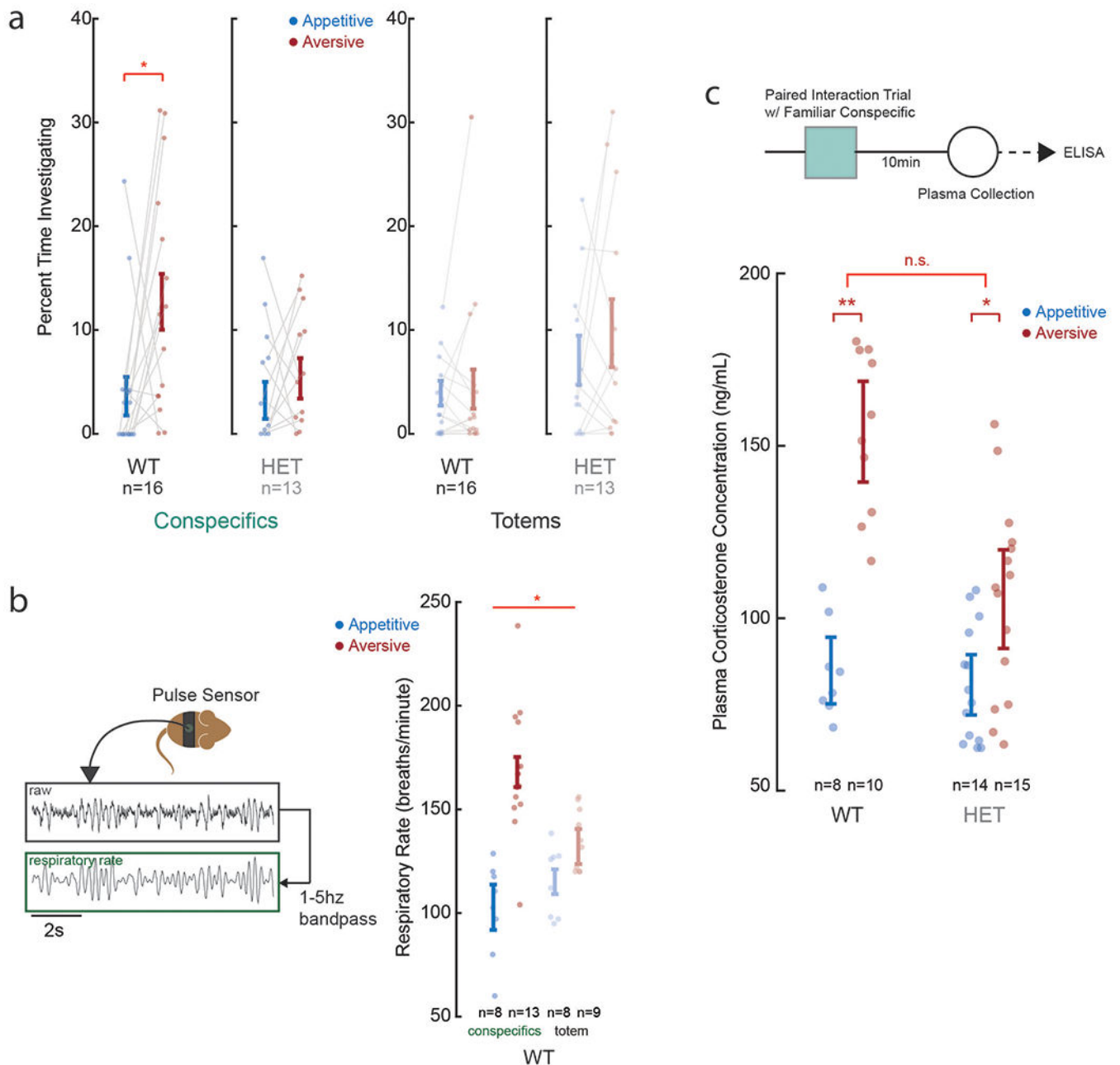


Extended Data Fig. 5. Robustness of neuronal encoding within the population across different statistical methodologies

a. To validate our results and further delineate the response characteristics of the neurons, we repeated the ANOVA analyses at different statistical thresholds from $p = 0.05$ to 0.0025 . Across conditions, other-to-self ratio was relatively stable, with HET and WT results ranging from 1:2.3 to 1:4.6, and 1.1:1 to 1:1.7, respectively. At all significance thresholds, the other-to-self ratio of HET was significantly smaller than that of WT (chi-square, $p < 0.05$).

b. Support vector machines (SVMs) with nonlinear kernels were used to decode neural population activity on validation trials not used for model training. The temporal dynamic of neural population predictions is presented in relation to the beginning, middle and end of the trials and are broken down into the three primary features that described the animals'

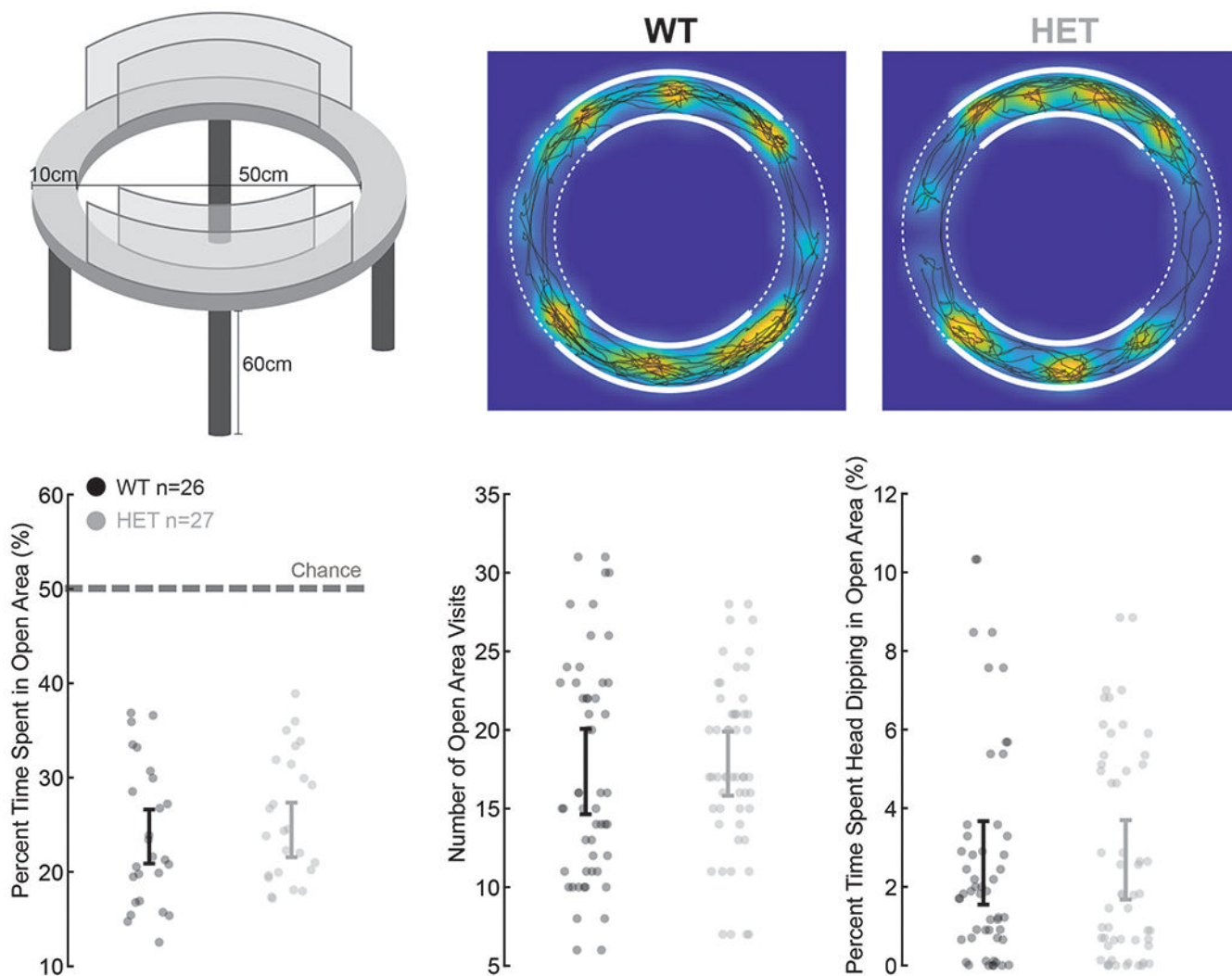
social interactions during the task; their experience valence, social agency and identity. Using neural population activity from validation trials not used for model fitting, we find that activity from the neural population could accurately predict the social agency of the animals' experience with an accuracy of $70.9 \pm 5\%$ at trial onset (permutation test, $p < 0.001$). In other words, the neural population could be used to reliably distinguish one's own experience from that of another. We also found that the neural population could predict the specific valence of the animals' experience with an accuracy of $76.8 \pm 3\%$ (permutation test, $p < 0.001$), suggesting that they reliably distinguished whether the experience was appetitive or aversive. Last, the neural population predicted the social context with an accuracy of $78.2 \pm 2\%$ (permutation test, $p < 0.001$). Overall, all three primary task features (agency, experience valence and social context/identity) were most accurately decoded from the population's response at the beginning of the trials (i.e., within the first 1-6 seconds; permutation test, $** p < 0.001$). Prediction accuracy then gradually dropped over the middle to end of the trial. While decoding accuracy was significant during the middle of the trial for valence and social identity (permutation test, $* p < 0.01$), decoding performance for all three features was at chance by the end of the trial. Representation of these task features was therefore most prominent at the start of the trials (i.e., at which time these conditions were most salient), and then gradually diminished by the end of the trial. To determine significance and error, this process was repeated 1000 times and compared to models trained on neuronal data that was randomly shuffled. Error bars indicate 95% confidence interval.



Extended Data Fig. 6. WT animals differentiate between others' aversive and appetitive experiences

a. To understand whether animals recognized another animal's experience and to allow for comparison with our main results, the animals performed a place preference/avoidance task where subjects were presented with two age and sex-matched familiar conspecific partners undergoing either an appetitive (food-baited enclosure) or aversive (confined tube enclosure) experience. The percent time spent by the animals investigating the aversive and appetitive enclosures when in the presence of another animal is shown here. While the WT displayed a significant difference the time spent investigating the specific enclosures when compared to HET animals (*two-tailed unpaired test; $t_s(15) = -2.79$; $p = 0.014$), there was

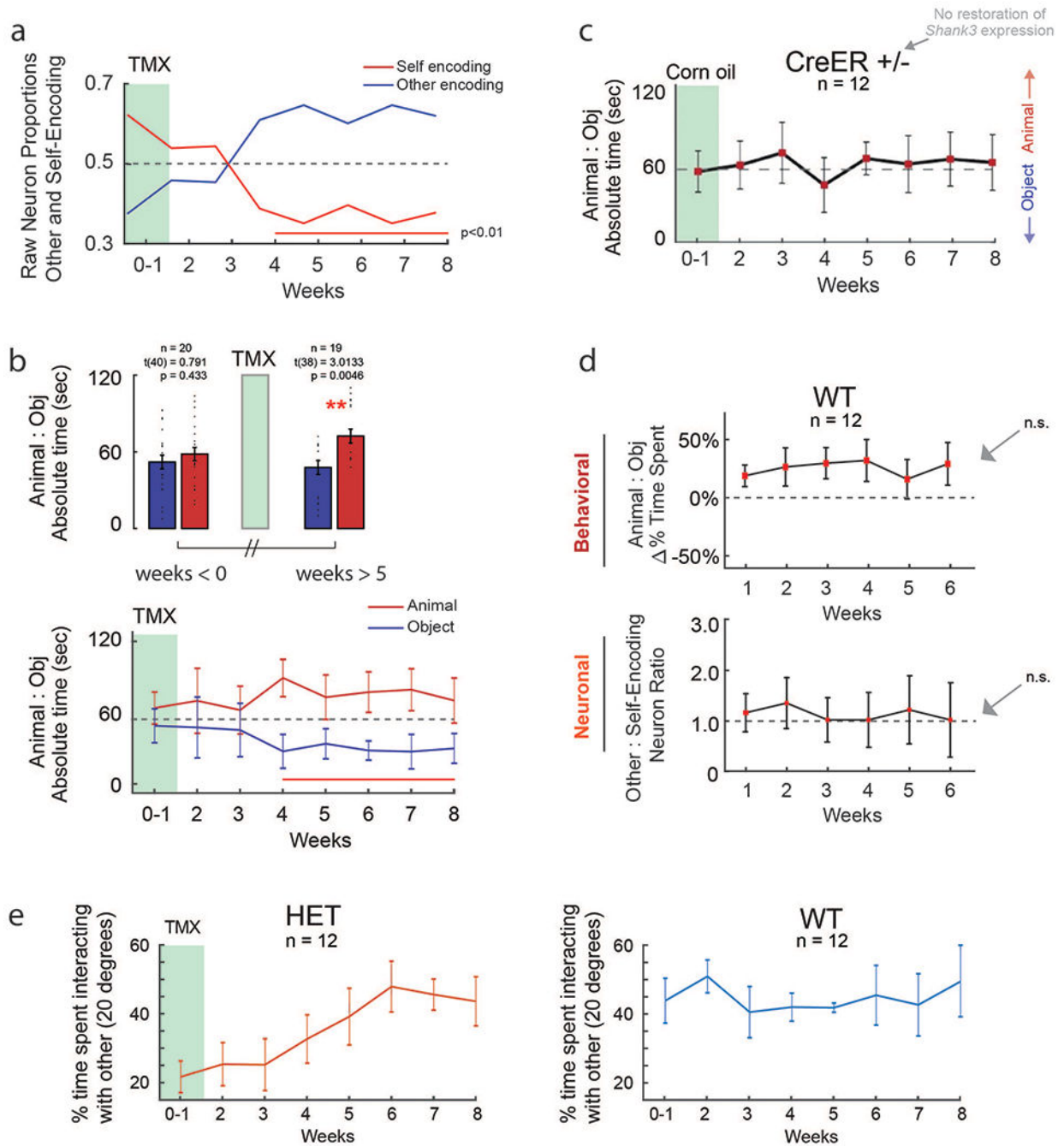
no difference in approach behavior when familiar conspecifics were replaced with inanimate totems in the enclosures (ANOVA, $p > 0.5$ post-hoc comparison). **b.** We confirmed that the WT animals differentiated the other's experiences based on respiratory rate. Using a mouse jacket (Lomir Biomedical) and pulse sensor (World Famous Electronics; digitized at 500 Hz then band-passed at 1-5 Hz), we find a significant difference in the animal's respiratory rate when the other animal was having an aversive vs. appetitive experience (*ANOVA, $p = 0.0031$). **c.** The panel *above* displays the sequence of behavioral testing, plasma collection and corticosterone analysis/quantification by ELISA. The panel *below* displays the plasma levels collected across all WT and HET animals. Here, the levels are given based on whether the animals observed the other having an appetitive vs. aversive experience (two-tailed unpaired t-tests; * $p = 0.021$, ** $p < 0.01$, n.s. $p = 0.23$).



Extended Data Fig. 7. Evaluating for *Shank3* dependent differences in anxiety phenotype between the WT and HET mice

Elevated zero-maze testing was performed on the WT and HET mice. Overall, HET mice displayed no evidence of elevated anxiety compared to WT based on the proportion of

time spent in the open areas, number of visits to the open areas, or percent of time spent head-dipping (two-tailed unpaired t-tests; $p > 0.5$).



Extended Data Fig. 8. *Shank3* dependent changes in proportion of neuronal encoding and sociability

a. Each curve represents the raw proportions of other-encoding and self-encoding neurons at weekly intervals after TMX administration (rather than ratio; with the sum equaling one). Time points in which there was a significant difference are underlined in red (chi-square tests, $p < 0.01$). The proportion of neurons that were task-modulated or which responded to

variations in familiarity did not significantly change in the weeks after TMX administration for *Shank3^{flx/+}:CreER^{+/-}* mice. Specifically, the percentage of task modulating neurons increased only marginally from 73% (n = 131 of 180, pre-tamoxifen) to 74% (n = 84 of 114, post TMX; chi-square, $\chi^2(1) = 0.29$, p = 0.86; Extended Data Fig. 3). Likewise, the percentage of neurons that responded to variations in the other's familiarity remained almost identical (30%, n = 39 of 131, pre-tamoxifen to 29.8%, n = 25 of 84, post-tamoxifen; chi-square, $\chi^2(1) = 2.0 \times 10^{-6}$, p = 1.0). **b.** *Above*, the net amount of time spent with the other animal vs. inanimate object out of the 120 seconds for testing. *Below*, three-chamber testing was obtained from the same individual animals over the consecutive course of 8 weeks after TMX administration. Time points in which there was a significant difference are underlined in red (paired t-tests, * ts(19) = 3.01; p = 0.0046). **c.** To further validate results from the *Shank3^{flx/+}:CreER^{-/-}* mice and to confirm that changes in behavior after TMX was not explained by task familiarity or time-progression, HET mice also received corn oil (i.e., vehicle) instead of TMX. Here, the HET (Cre+/-) mice displayed no change in behavior over time (one-sample t-tests, p > 0.5). **d.** In the *top* panel, the curve represents the proportion of time spent in the chamber of the other animal or inanimate object \pm standard errors of the mean at weekly intervals. In the *bottom* panel, each curve represents the percentage of other-encoding and self-encoding neurons at weekly intervals. The horizontal line represents an equal proportion of neurons. No time point demonstrated a significant difference in either behavior or sociability when compared to all other points (one-sample t-tests, n.s.; p > 0.2). **e.** The proportion of time interacting with the other animal as defined by the amount of time at which the subject mouse was in proximity (within 3 cm) and oriented towards the other animal (within 20 degrees). HET mice displayed a gradual and significant increase in the amount of time spent interacting with the other animal after TMX (paired t-test, p < 1.0×10^{-4}). These findings were not apparent when testing the *Shank3^{flx/+}:CreER^{-/-}* mice which lacked the Cre-lox system after TMX (unpaired t-test, p > 0.5). More remarkably, by 8 weeks after TMX administration, the amount of time spent interacting with the other by the HET mice was essentially indistinguishable from that of the WT mice (unpaired t-test, p > 0.5). All error bars indicate s.e.m.

Extended Data Table 3.
Evaluating the consistency of neuronal encoding using different statistical techniques.

Proportions of neurons based on ANOVA (p < 0.0125) and linear decoding approaches (p < 0.0125). Overall, both statistical approaches were associated with a similar proportion of neurons that responded to self- and other-experience. Both approaches were also associated with a similar difference in encoding when comparing the WT to HET mice as well as when comparing the HET mice before vs. after TMX or endoxifen administration.

	ANOVA		Linear decoder	
	WT	HET	WT	HET
Total neurons	188	180	188	180
Task modulated	112	131	102	120
Other-experience	33 (30%)	12 (9%)	37 (36%)	19 (15%)

	ANOVA		Linear decoder	
	WT	HET	WT	HET
Self-experience	29 (26%)	50 (38%)	31 (30%)	44 (37%)
post-TMX (> 5 wks)				
Task modulated	-	138	-	95
Other-experience	-	49 (36%)	-	29 (31%)
Self-experience	-	27 (20%)	-	18 (19%)
post-Endox. (> 5 wks)				
Task modulated	-	72	-	67
Other-experience	-	30 (42%)	-	26 (39%)
Self-experience	-	14 (19%)	-	14 (21%)

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Anna Aristarkhova, Brigitte Burcescu, Jacob Grondin, Emma Mastrobattista, Carlos Ortega, Jonas St Fleur, and Mackenna Mejdell for assisting in the experimentation and data pre-processing. G.F. is funded by HHMI, S.W.L. is funded by the Autism Foundation Fellowship and Z.M.W. is supported by NIH R01HD059852, NIH R01NS091390, the Presidential Early Career Award for Scientists and Engineers and the Whitehall Foundation.

Data availability

The behavioural and neuronal data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Rudebeck PH, Buckley MJ, Walton ME & Rushworth MF A role for the macaque anterior cingulate gyrus in social valuation. *Science* 313, 1310–1312, doi:10.1126/science.1128197 (2006). [PubMed: 16946075]
2. Chang SW, Garipey JF & Platt ML Neuronal reference frames for social decisions in primate frontal cortex. *Nat Neurosci* 16, 243–250, doi:10.1038/nn.3287 (2013). [PubMed: 23263442]
3. Dolen G, Darvishzadeh A, Huang KW & Malenka RC Social reward requires coordinated activity of nucleus accumbens oxytocin and serotonin. *Nature* 501, 179–184, doi:10.1038/nature12518 (2013). [PubMed: 24025838]
4. Li Yet al. Neuronal Representation of Social Information in the Medial Amygdala of Awake Behaving Mice. *Cell* 171, 1176–1190 e1117, doi:10.1016/j.cell.2017.10.015 (2017). [PubMed: 29107332]
5. Caggiano V, Fogassi L, Rizzolatti G, Thier P & Casile A Mirror neurons differentially encode the peripersonal and extrapersonal space of monkeys. *Science* 324, 403–406, doi:10.1126/science.1166818 (2009). [PubMed: 19372433]
6. Omer DB, Maimon SR, Las L & Ulanovsky N Social place-cells in the bat hippocampus. *Science* 359, 218–224, doi:10.1126/science.aao3474 (2018). [PubMed: 29326274]
7. Haroush K & Williams ZM Neuronal Prediction of Opponent's Behavior during Cooperative Social Interchange in Primates. *Cell*, doi:S0092–8674(15)00123–3 [pii] 10.1016/j.cell.2015.01.045 (2015).

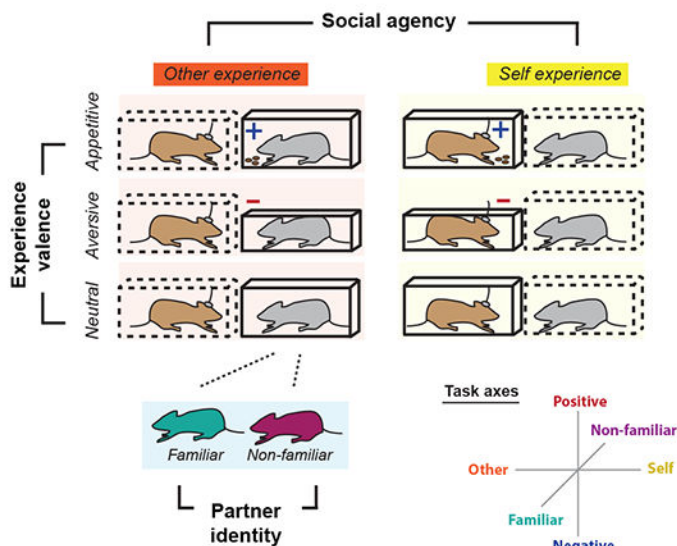
8. Allsop SA et al. Corticoamygdala Transfer of Socially Derived Information Gates Observational Learning. *Cell* 173, 1329–1342 e1318, doi:10.1016/j.cell.2018.04.004 (2018). [PubMed: 29731170]
9. Yoshida K, Saito N, Iriki A & Isoda M Social error monitoring in macaque frontal cortex. *Nat Neurosci* 15, 1307–1312, doi:10.1038/nn.3180 (2012). [PubMed: 22864610]
10. Murugan M et al. Combined Social and Spatial Coding in a Descending Projection from the Prefrontal Cortex. *Cell* 171, 1663–1677 e1616, doi:10.1016/j.cell.2017.11.002 (2017). [PubMed: 29224779]
11. Zhang W & Yartsev MM Correlated Neural Activity across the Brains of Socially Interacting Bats. *Cell* 178, 413–428 e422, doi:10.1016/j.cell.2019.05.023 (2019). [PubMed: 31230710]
12. Burkett J et al. Oxytocin-dependent consolation behavior in rodents. *Science* 351, 375–378, doi:10.1126/science.aac4785 (2016). [PubMed: 26798013]
13. Chang L & Tsao DY The Code for Facial Identity in the Primate Brain. *Cell* 169, 1013–1028 e1014, doi:10.1016/j.cell.2017.05.011 (2017). [PubMed: 28575666]
14. Amaral D, Dawson G & Geschwind DH Autism spectrum disorders. (Oxford University Press, 2011).
15. Lord C & Bishop SL Recent advances in autism research as reflected in DSM-5 criteria for autism spectrum disorder. *Annu Rev Clin Psychol* 11, 53–70, doi:10.1146/annurev-clinpsy-032814-112745 (2015). [PubMed: 25581244]
16. Lamm C, Bukowski H & Silani G From shared to distinct self-other representations in empathy: evidence from neurotypical function and socio-cognitive disorders. *Philos Trans R Soc Lond B Biol Sci* 371, 20150083, doi:10.1098/rstb.2015.0083 (2016). [PubMed: 26644601]
17. Steinbeis N The role of self-other distinction in understanding others' mental and emotional states: neurocognitive mechanisms in children and adults. *Philos Trans R Soc Lond B Biol Sci* 371, 20150074, doi:10.1098/rstb.2015.0074 (2016). [PubMed: 26644593]
18. Uddin LQ et al. Neural basis of self and other representation in autism: an FMRI study of self-face recognition. *PLoS One* 3, e3526, doi:10.1371/journal.pone.0003526 (2008). [PubMed: 18958161]
19. Senju A Spontaneous theory of mind and its absence in autism spectrum disorders. *Neuroscientist* 18, 108–113, doi:10.1177/1073858410397208 (2012). [PubMed: 21609942]
20. Uchino S & Waga C SHANK3 as an autism spectrum disorder-associated gene. *Brain Dev* 35, 106–110, doi:10.1016/j.braindev.2012.05.013 (2013). [PubMed: 22749736]
21. Guilmatre A, Huguet G, Delorme R & Bourgeron T The emerging role of SHANK genes in neuropsychiatric disorders. *Dev Neurobiol* 74, 113–122, doi:10.1002/dneu.22128 (2014). [PubMed: 24124131]
22. Yi F et al. Autism-associated SHANK3 haploinsufficiency causes Ih channelopathy in human neurons. *Science* 352, e2669, doi:10.1126/science.aaf2669 (2016). [PubMed: 26966193]
23. Bariselli S et al. SHANK3 controls maturation of social reward circuits in the VTA. *Nat Neurosci* 19, 926–934, doi:10.1038/nn.4319 (2016). [PubMed: 27273769]
24. Duffney L et al. Shank3 deficiency induces NMDA receptor hypofunction via an actin-dependent mechanism. *J Neurosci* 33, 15767–15778, doi:10.1523/JNEUROSCI.1175-13.2013 (2013). [PubMed: 24089484]
25. Peca J et al. Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. *Nature* 472, 437–442, doi:10.1038/nature09965 (2011). [PubMed: 21423165]
26. Yizhar O et al. Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature* 477, 171–178, doi:10.1038/nature10360 (2011). [PubMed: 21796121]
27. Stoodley C et al. Altered cerebellar connectivity in autism and cerebellar-mediated rescue of autism-related behaviors in mice. *Nat Neurosci* 20, 1744–1751, doi:10.1038/s41593-017-0004-1 (2017). [PubMed: 29184200]
28. Twining RC, Vantrease JE, Love S, Padival M & Rosenkranz JA An intra-amygdala circuit specifically regulates social fear learning. *Nat Neurosci* 20, 459–469, doi:10.1038/nn.4481 (2017). [PubMed: 28114293]
29. Guo B et al. Anterior cingulate cortex dysfunction underlies social deficits in Shank3 mutant mice. *Nat Neurosci* 22, 1223–1234, doi:10.1038/s41593-019-0445-9 (2019). [PubMed: 31332372]

30. Mei Yet al. Adult restoration of Shank3 expression rescues selective autistic-like phenotypes. *Nature* 530, 481–484, doi:10.1038/nature16971 (2016). [PubMed: 26886798]
31. Guo C, Yang W & Lobe CG A Cre recombinase transgene with mosaic, widespread tamoxifen-inducible action. *Genesis* 32, 8–18, doi:10.1002/gene.10021 (2002). [PubMed: 11835669]
32. Schnutgen Fet al. A directional strategy for monitoring Cre-mediated recombination at the cellular level in the mouse. *Nat Biotechnol* 21, 562–565, doi:10.1038/nbt811 (2003). [PubMed: 12665802]
33. Jaramillo TC et al. Novel Shank3 mutant exhibits behaviors with face validity for autism and altered striatal and hippocampal function. *Autism Res* 10, 42–65, doi:10.1002/aur.1664 (2017). [PubMed: 27492494]
34. Singer Tet al. Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 1157–1162, doi:10.1126/science.1093535303/5661/1157 [pii] (2004). [PubMed: 14976305]
35. Zielinski BA et al. Longitudinal changes in cortical thickness in autism and typical development. *Brain* 137, 1799–1812, doi:10.1093/brain/awu083 (2014). [PubMed: 24755274]
36. Stone VE, Baron-Cohen S & Knight RT Frontal lobe contributions to theory of mind. *J Cogn Neurosci* 10, 640–656, doi:10.1162/089892998562942 (1998). [PubMed: 9802997]
37. Ben-Ami Bartal I, Decety J & Mason P Empathy and pro-social behavior in rats. *Science* 334, 1427–1430, doi:10.1126/science.1210789 (2011). [PubMed: 22158823]
38. Abe Cet al. CI neurons mediate a stress-induced anti-inflammatory reflex in mice. *Nat Neurosci* 20, 700–707, doi:10.1038/nn.4526 (2017). [PubMed: 28288124]
39. Land BB et al. Medial prefrontal D1 dopamine neurons control food intake. *Nat Neurosci* 17, 248–253, doi:10.1038/nn.3625 (2014). [PubMed: 24441680]
40. Rowland NE, Giddings AM, Minervini V & Robertson KL Economics of food intake in mice: energy yield of the reinforcer. *Physiol Behav* 136, 104–110, doi:10.1016/j.physbeh.2014.04.036 (2014). [PubMed: 24768647]
41. Trappenberg TFundamentals of computational neuroscience. (Oxford University Press, 2006).
42. Tucker LB & McCabe JT Behavior of Male and Female C57BL/6J Mice Is More Consistent with Repeated Trials in the Elevated Zero Maze than in the Elevated Plus Maze. *Front Behav Neurosci* 11, 13, doi:10.3389/fnbeh.2017.00013 (2017). [PubMed: 28184191]
43. Bozdagi O et al. Haploinsufficiency of the autism-associated Shank3 gene leads to deficits in synaptic function, social interaction, and social communication. *Mol Autism* 1, 15, doi:10.1186/2040-2392-1-15 (2010). [PubMed: 21167025]
44. Pearl J Causality: Models, Reasoning and Inference. 2 edn. (Cambridge University Press, 2009).
45. Orefice L et al. Peripheral Mechanosensory Neuron Dysfunction Underlies Tactile and Behavioral Deficits in Mouse Models of ASDs. *Cell* 166, 299–313, doi:10.1016/j.cell.2016.05.033 (2016). [PubMed: 27293187]
46. Duerden E et al. Self-injurious behaviours are associated with alterations in the somatosensory system in children with autism spectrum disorder. *Brain Struct Funct* 219, 1251–1261, doi:10.1007/s00429-013-0562-2 (2014). [PubMed: 23644587]
47. Libero LE, DeRamus TP, Deshpande HD & Kana RK Surface-based morphometry of the cortical architecture of autism spectrum disorders: volume, thickness, area, and gyrification. *Neuropsychologia* 62, 1–10, doi:10.1016/j.neuropsychologia.2014.07.001 (2014). [PubMed: 25019362]
48. Sukhodolsky DG, Bloch MH, Panza KE & Reichow B Cognitive-behavioral therapy for anxiety in children with high-functioning autism: a meta-analysis. *Pediatrics* 132, e1341–1350, doi:10.1542/peds.2013-1193 (2013). [PubMed: 24167175]
49. Harrington JW & Allen K The clinician’s guide to autism. *Pediatr Rev* 35, 62–78; quiz 78, doi:10.1542/pir.35-2-62 (2014). [PubMed: 24488830]
50. Broke SL The use of the creative therapies with autism spectrum disorders. (Charles C. Thomas, Publisher, 2009).

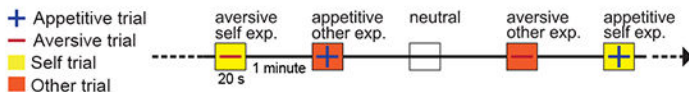
Methods-only References

51. Verpelli C et al. Importance of Shank3 protein in regulating metabotropic glutamate receptor 5 (mGluR5) expression and signaling at synapses. *J Biol Chem* 286, 34839–34850, doi:10.1074/jbc.M111.258384 (2011). [PubMed: 21795692]
52. Shapiro RM, Badalamenti JI & Glick SD A simple and rapid technique for preparing histological sections of brain. *Pharmacol Biochem Behav* 19, 1049–1050, doi:10.1016/0091-3057(83)90415-x (1983). [PubMed: 6197714]
53. Moore GP, Segundo JP, Perkel DH & Levitan H Statistical signs of synaptic interaction in neurons. *Biophys J* 10, 876–900, doi:S0006-3495(70)86341-X [pii] 10.1016/S0006-3495(70)86341-X (1970). [PubMed: 4322240]
54. Benedykcińska A et al. Generation of brain tumours in mice by Cre-mediated recombination of neural progenitors in situ with the tamoxifen metabolite endoxifen. *Dis Model Mech* 9, 211–220, doi:10.1242/dmm.022715 (2016). [PubMed: 26704996]
55. Nadler JJ et al. Automated apparatus for quantitation of social approach behaviors in mice. *Genes Brain Behav* 3, 303–314, doi:10.1111/j.1601-183X.2004.00071.x (2004). [PubMed: 15344923]
56. Rogers-Carter M et al. Insular cortex mediates approach and avoidance responses to social affective stimuli. *Nat Neurosci* 21, 404–414, doi:10.1038/s41593-018-0071-y (2018). [PubMed: 29379116]
57. Hwang D-J et al. Comparative analysis of restraint stress-induced depressive-like phenotypes in C57BL/6N mice derived from three different sources. *Laboratory Animal Research* 36, doi:10.1186/s42826-020-00062-0 (2020).
58. Kinn Rød AM, Harkestad N, Jellestad FK & Murison R Comparison of commercial ELISA assays for quantification of corticosterone in serum. *Scientific Reports* 7, doi:10.1038/s41598-017-06006-4 (2017).
59. Pagan M, Urban LS, Wohl MP & Rust NC Signals in inferotemporal and perirhinal cortex suggest an untangling of visual target information. *Nat Neurosci* 16, 1132–1139, doi:10.1038/nn.3433 (2013). [PubMed: 23792943]
60. Quiñero R, Snyder LH, Batista AP, Cui H & Andersen RA Movement intention is better predicted than attention in the posterior parietal cortex. *J Neurosci* 26, 3615–3620, doi:10.1523/JNEUROSCI.3468-05.2006 (2006). [PubMed: 16571770]
61. Hung CP, Kreiman G, Poggio T & DiCarlo JJ Fast readout of object identity from macaque inferior temporal cortex. *Science* 310, 863–866, doi:10.1126/science.1117593 (2005). [PubMed: 16272124]
62. Wasserman L All of statistics: a concise course in statistical inference. (Springer Texts, 2005).
63. Shanechi M et al. Neural population partitioning and a concurrent brain-machine interface for sequential motor function. *Nat Neurosci* 15, 1715–1722, doi:10.1038/nn.3250 (2012). [PubMed: 23143511]
64. McCullagh PN, J. A. . Generalized linear models. (Chapman & Hall, 1989).

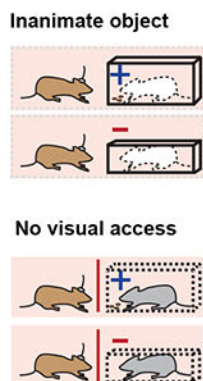
a Paired Interactions



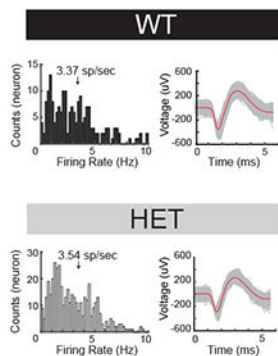
Example of Randomized Trial Sequence



b Sensory Controls



c Single-Neuronal Recordings



d Neuronal Responses

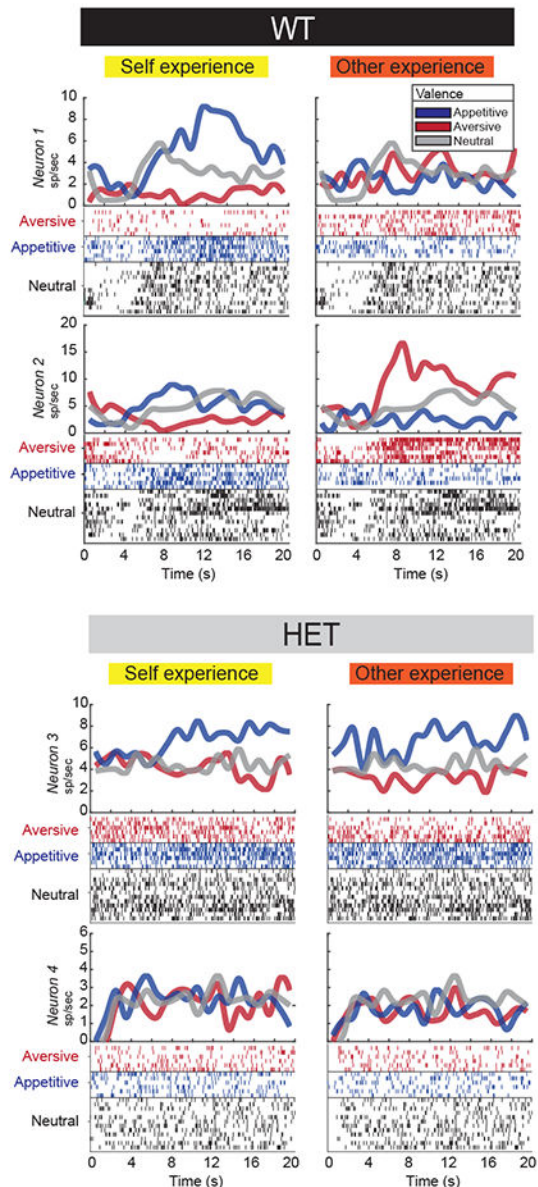


Fig. 1. Neuronal encoding of social information in the mPFC and its relation to *Shank3* expression.

a. Pairs of mice performed a social task in which three main variables were tested. These included social agency (self vs. other), social identity (familiar vs. non-familiar) and experience valence (positive vs. negative). Thus, for example, whereas one trial may involve another mouse undergoing an aversive experience, the subsequent trial may involve the recorded mouse undergoing an appetitive experience. Here, the recorded mouse performed these trials in randomly interleaved fashion, with one-minute inter-trial periods over multiple trials. The figure shows a representative sequence of such trial combinations. The full set of self- and other- enclosure combination is described in the Methods. **b.** To further control for potential variations related to the stimuli themselves, either an inanimate mouse-shaped

object was given or direct visual access was blocked. **c.** Multiple neuronal recordings were taken from the mPFC using penetrating microelectrode arrays. Here, the distribution of firing rates and sample waveform morphologies are displayed for both the WT and HET mice. Additional examples of waveform morphologies as well as histological section confirming the microelectrode locations in the cingulate gyrus of the mPFC can be seen in Extended Data Fig. 1b. **d.** Peri-event histograms and rasters from four representative neurons aligned to trial onset. Here, neuronal responses are divided based on differences in social agency (self vs. other) and experience valence (appetitive vs. aversive). Neutral conditions in which neither the recorded animal or the other agent underwent any experience are displayed for both left and right panels for comparison. Examples are given for both WT mice (out of $n = 112$ task-modulated neurons) and HET mice (out of $n = 131$ task-modulated neurons).

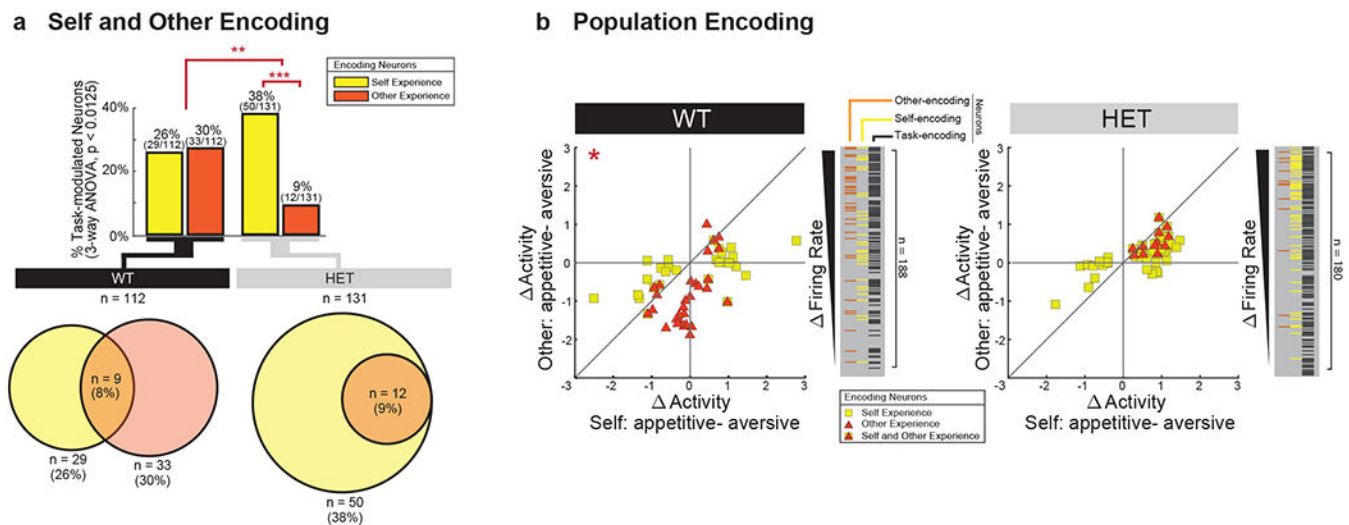


Fig. 2. Diminished *Shank3* expression leads to diminished neuronal encoding of other-experience and diminished distinction between other-and-self.

a. Proportion of task-modulated neurons that encoded for self- and other-experience.

Differences in encoding characteristics within and between the WT and HET mice were made by Chi-square test analysis (** $\chi^2(1) = 15.38, p = 8.78 \times 10^{-5}$, *** $\chi^2(1) = 30.51, p = 3.32 \times 10^{-8}$). Venn diagrams *below* display the degree to which neurons that encoded self- vs. other-experience overlapped. Supplementary Table 2 provides additional results based on decoding. **b.** Scatter plot indicating the differences in neural activity (z-score) on trials in which appetitive (positive) vs. aversive (negative) experiences were given and illustrating the magnitude of response across conditions for individual cells. Here, the y-axis reflects absolute differences in activity based on whether the other animal had an appetitive vs. aversive experience (i.e., other-experience) and the x-axis reflects absolute differences in activity based on whether the recorded animals themselves had an appetitive vs. aversive experience (i.e., self-experience). Therefore, a neuron that responded preferentially to the other's experience would be found midline along the vertical axis of the graph whereas a neuron that responded preferentially to one's own experience would be found midline along the horizontal axis of the graph. Plotted cells are those which displayed significant modulation (three-way ANOVA, $p < 0.0125$) to self (yellow) and other (orange) experience. The overlap between encoding properties *per cell* can be seen within the scatter plots and the *insets* to their right. Differences in the distribution of neuronal responses to self- vs. other-experience were evaluated for by a 2D 2-sample KS test (* $p < 0.001$). $n = 112$ and $n = 131$ task-modulated neurons recorded from $n = 12$ WT and $n = 12$ HET mice, respectively.

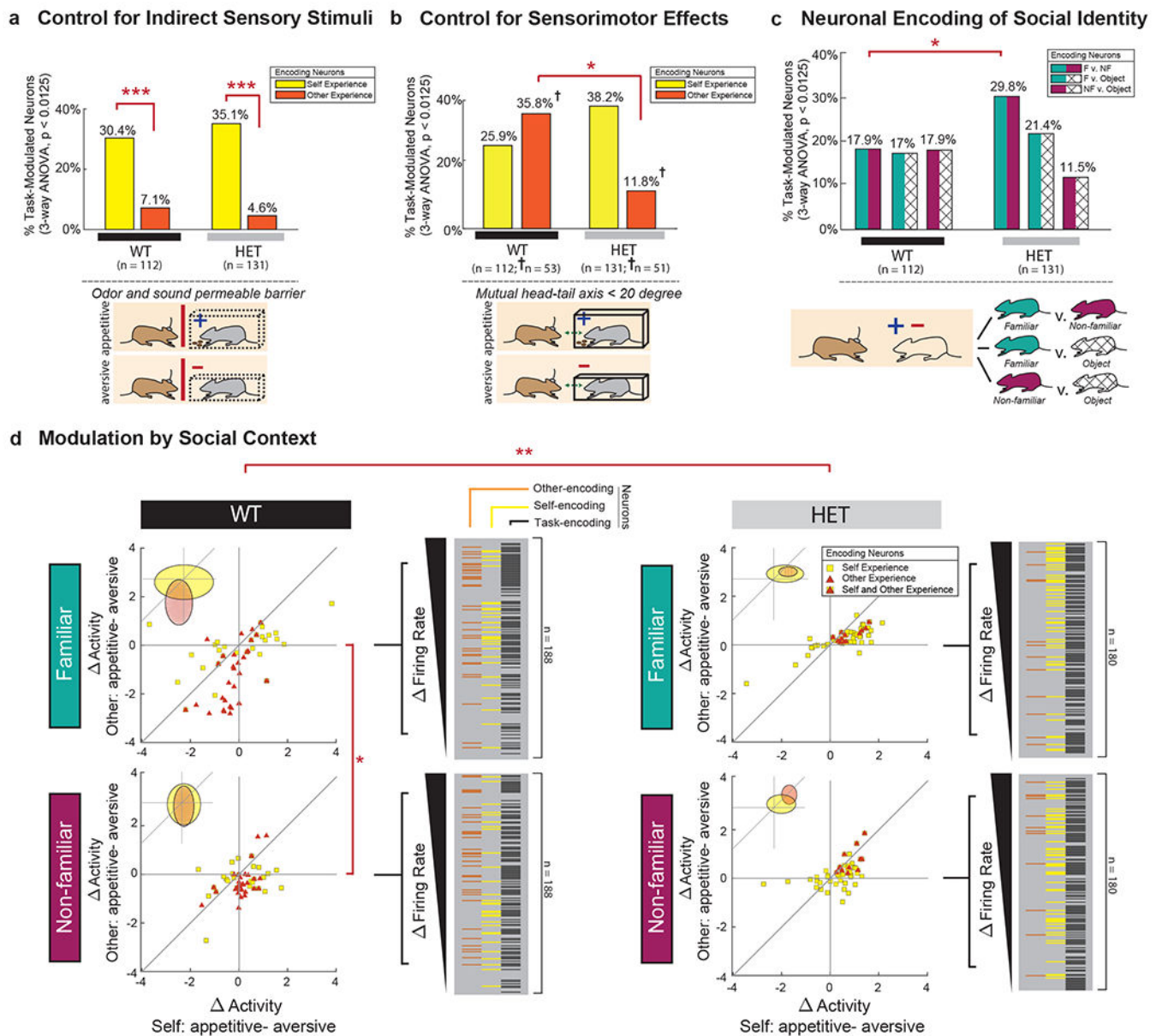
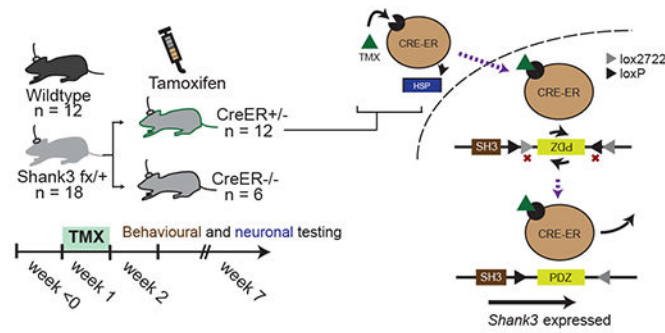


Fig. 3. Neuronal encoding of the other's experience is social context specific.

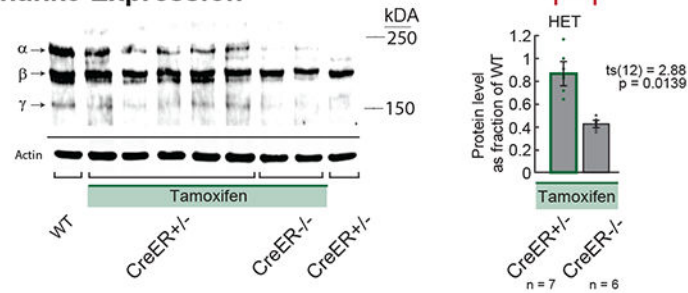
a. Neuronal responses under control conditions in which the view of the other mouse was blocked by an odor- and sound-permeable barrier in both WT (***) $\chi^2(1) = 19.81$, $p = 8.56 \times 10^{-6}$) and HET (***) $\chi^2(1) = 38.39$, $p = 5.8 \times 10^{-10}$) animals. See Fig. 2c for comparison. **b.** Neuronal responses when confining activity to the time periods in which the recorded animal was in direct view of the other (i.e., within 20 degrees). † signifies a sub-selection of neurons used, where $n = 53$ and $n = 51$ task-modulated neurons recorded from $n = 12$ WT and $n = 12$ HET mice, respectively, across trials. Significances are displayed by Chi-square test analysis (* $\chi^2(1) = 8.26$, $p = 0.004$). Supplementary Fig. 2 provides additional results based on alignment to behavioral events. Extended Data Figs. 6–7 provide results on approach- and anxiety-related behaviors. **c.** Proportion of neurons in the WT and HET mice that significantly differentiated between familiar mice (F), non-familiar

mice (NF) and inanimate objects (Object) (* $p < 0.05$, 2D 2-sample KS tests). **d.** Scatter plots displaying differences in neural activity (Z-score) on trials in which a self vs. other, appetitive vs. aversive experience was given (same convention as in Fig. 2b), but additionally divided based on whether the other animal was a familiar or non-familiar partner (* $D = 0.45$, $p = 0.034$, ** $D = 0.727$, $p = 0.0024$, 2D 2-sample KS tests). The elliptical *insets* indicate the first standard deviation of the means for the distribution of firing rates of self-encoding (yellow) and other-encoding (orange) neurons, respectively (three-way ANOVA, $p < 0.0125$). Extended Data Figs. 2 and Supplementary Fig. 4 provides further breakdown of what neurons responded to social identity as well as their specific relation to other's experience. $n = 112$ and $n = 131$ task-modulated neurons recorded from $n = 12$ WT and $n = 12$ HET mice, respectively.

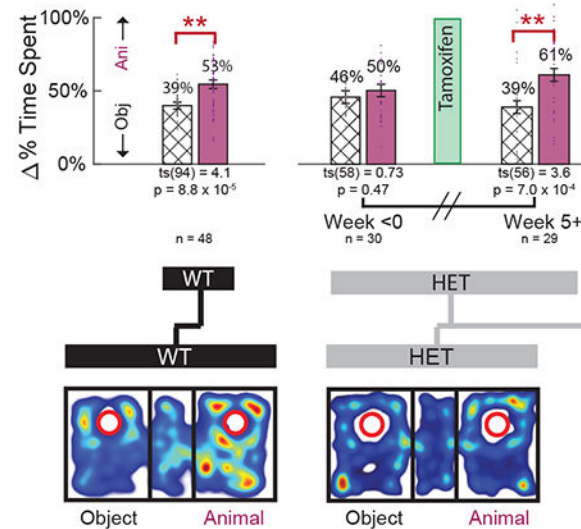
a Conditional In Vivo Restoration of *Shank3* Expression



b *Shank3* Expression



c *Shank3* Related Changes in Social Behavior



d Motoric Testing

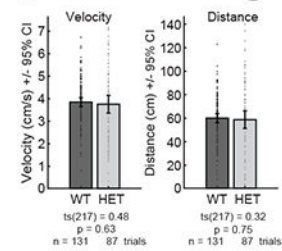


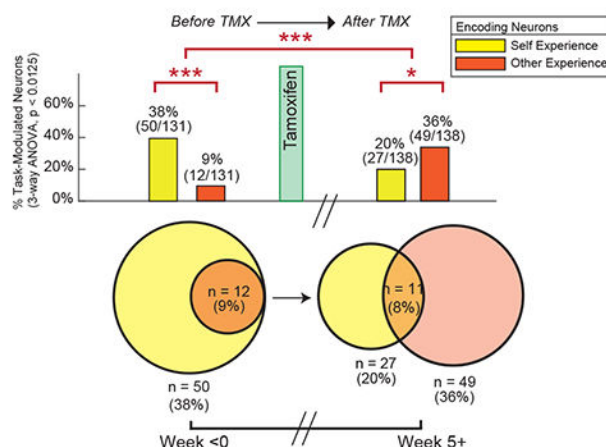
Fig. 4. Increase in *Shank3* expression is associated with increased sociability.

a. Tamoxifen (TMX) was used to activate *CreER* function and therefore provide temporal control of *Shank3* expression in the *Shank3^{fx/+}:CreER^{+/-}* (HET) mice. Neuronal and behavioral evaluations were performed before and after the administration of TMX. See Supplementary Table 1 for additional detail on all the genetic constructs and conditions tested as well as Extended Data Fig. 1a for additional description of the Cre-lox system.

b. *Shank3* expression was quantified *via* synaptosome preparation and Western blot in the *Shank3^{fx/+}:CreER^{+/-}* (i.e., HET) and *Shank3^{fx/+}:CreER^{-/-}* (i.e., control) mice. Heterozygous

knockin of *Shank3* primarily affected the putative SHANK3 α and γ isoforms and, to a lesser extent, the β isoform (see Methods). After administration of TMX, the *Shank3^{fx/+}: CreER^{+/-}* but not *CreER^{-/-}* constructs demonstrated SHANK3 protein levels comparable to those of WT (*left*). Error bars represent standard error of the mean, with a protein level of 1.0 representing WT. See Source Data Fig. 1 for the full unprocessed Western blot. **c.** Behavioral testing for sociability was performed using a standard three-chamber enclosure, wherein mice were presented with a novel animal (purple) vs. novel inanimate object (hash). Error bars represent the standard error of the mean. *Below* are heatmaps of representative trials. Significances are displayed by two-sample t-tests (** $p < 1 \times 10^{-3}$). $n = 48, 30,$ and 29 three-chamber trials from $n = 28$ WT, $n = 18$ HET, and $n = 18$ HET + TMX mice, respectively **d.** HET mice did not exhibit impaired motoric behavior, as assessed by velocity or distribution of distances traveled. Here, $n = 131$ and 87 trials were assessed from $n = 28$ WT and $n = 18$ HET mice, respectively.

a Shank3 Related Change in Neuronal Encoding



b Tracking the Dependency between Neuronal Encoding and Social Behavior in Real-Time

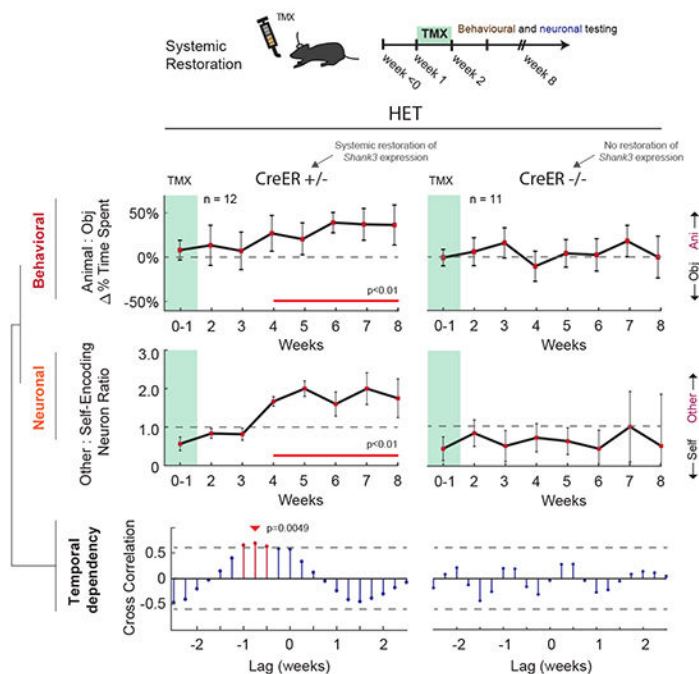


Fig. 5. Restoration of social behavior is dependent on the neuronal encoding of self-and-other.

a. Proportion of task-modulated neurons that encoded for variations in self- and other-experience, before (0 weeks) and after (5 weeks) TMX administration. n = 131 and n = 138 task-modulated neurons recorded from n = 12 HET and n = 12 HET + TMX mice, respectively. The Venn diagram *below* shows the overlap between other- and self-encoding neurons. Significances are displayed by Chi-square test analysis (***) $\chi^2(1) = 30.51$, p = 3.32×10^{-8}). To the *right*, restoration of *Shank3* expression is largely associated with a normalization of neuronal encoding responses (i.e., compared to Fig. 2a; *** $\chi^2(1) =$

28.18, $p = 1.1 \times 10^{-7}$, * $\chi^2(1) = 8.79$, $p = .003$). For additional analyses, see Supplementary Fig. 5. **b.** Behavioral and neuronal recordings were obtained from these $n=12$ animals over the course of 8 weeks after TMX administration, and compared with data from $n=11$ *Shank3^{flx/+}:CreER^{-/-}* mice that lacked the CreER recombinase gene and therefore displayed no increase in *Shank3* expression. In the *top* panel, the curves represent the relative preference of the mice for another animal compared to an object (\pm S.E.M.) at weekly intervals. The horizontal line represents no preference. Whereas the Cre^{+/-} mice displayed a significant increase in sociability over time compared to baseline (weeks 0-1), the Cre^{-/-} mice did not. Time points in which there was a significant difference are underlined in red (one-sample t-tests, $p < 0.01$). In the *middle* panel, the curves represent the proportion of other-encoding compared to self-encoding neurons in the mPFC \pm S.E.M. at weekly intervals. The horizontal line represents an equal proportion of neurons. Here, the Cre^{+/-} mice displayed a significant increase in other-encoding neurons over time compared to baseline (weeks 0-1), whereas the Cre^{-/-} mice demonstrated no change. Time points in which there was a significant difference are underlined in red ($p < 0.01$). The raw proportions of neurons and behavioral metrics are further provided in Extended Data Fig. 8a-b. Changes in the physical interaction times between the mice are shown in Extended Data Fig. 8e. In the *bottom* panel, the lines represent the degree of cross-correlation between changes in social behavior and change neuronal encoding (i.e., their statistical dependency). Red lines indicate time lags that were significant with the arrow indicating the optimal time lag at which change in social behavior proceeded change in neuronal encoding in the mPFC (0.75 weeks for the *Shank3^{flx/+}:CreER^{+/-}* mice).

Controlling for time-progression using *Shank3^{fx/+}; CreER^{-/-}* Mice

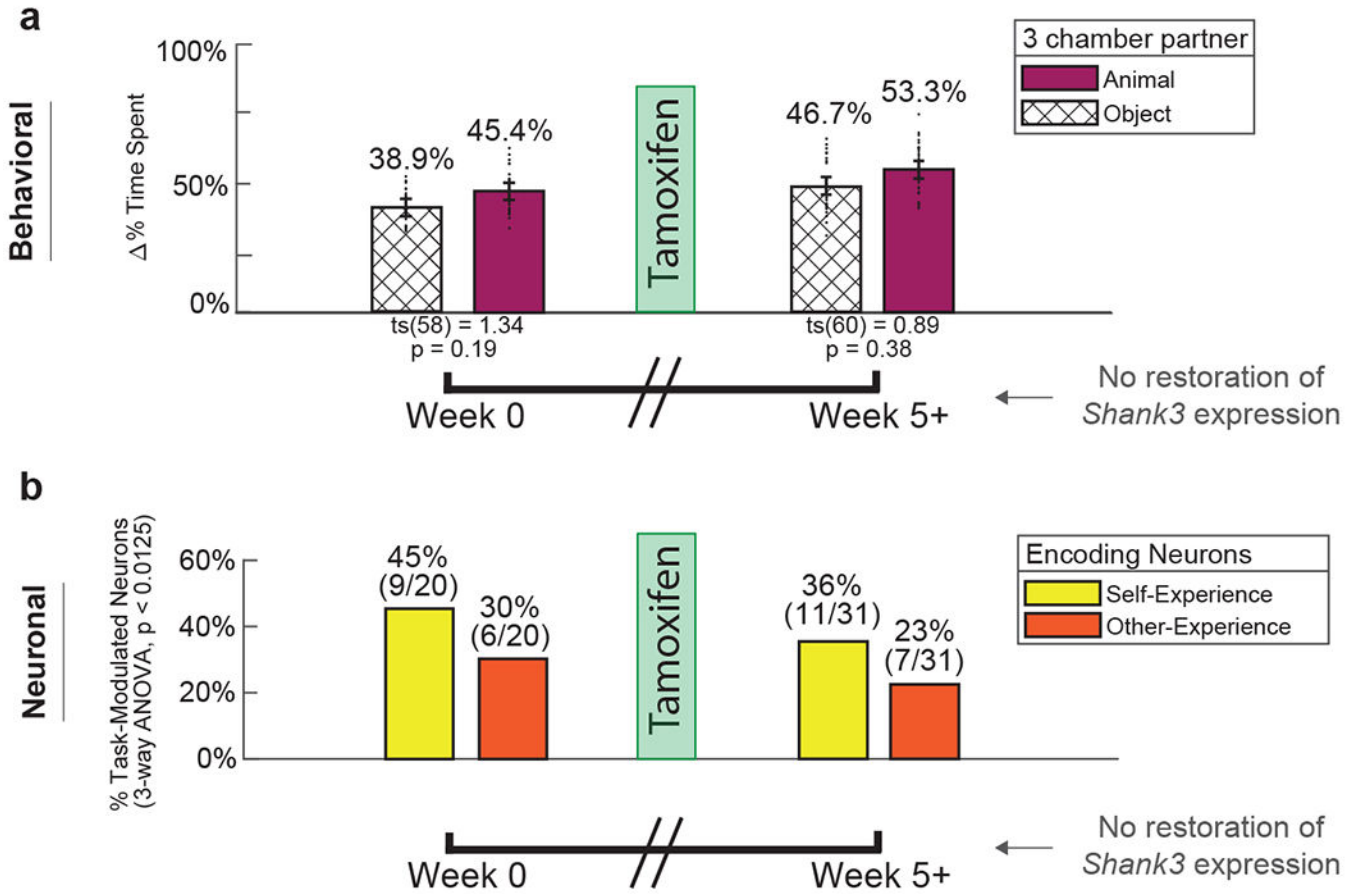
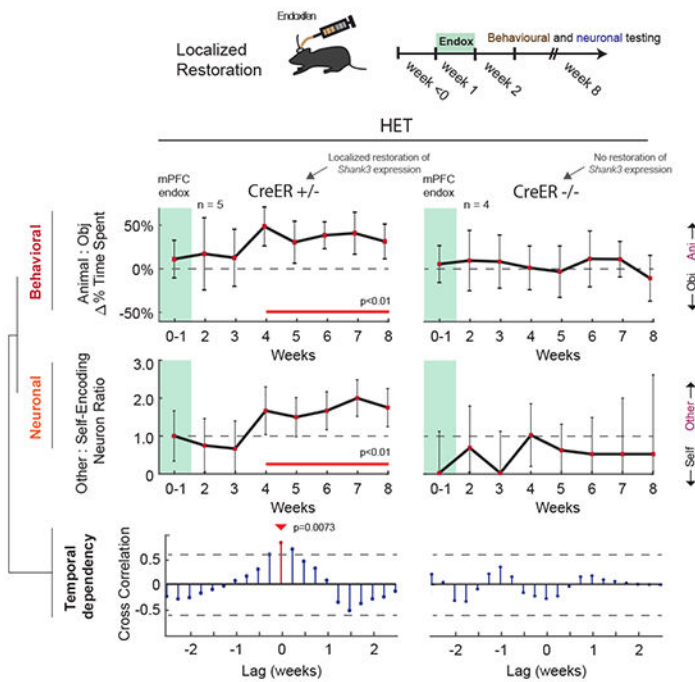


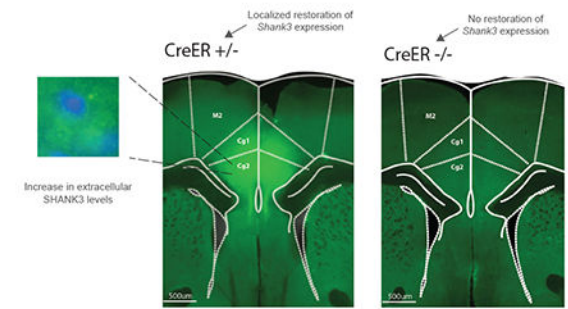
Fig. 6. Controls for time-progression and familiarity using *Shank3^{fx/+}; CreER^{-/-}* mice. Mice lacking the CreER recombinase gene (*Shank3^{fx/+}; CreER^{-/-}*) were given TMX to control for the independent effects of time-progression or familiarity. These mice therefore underwent the same task experiences and TMX administration as the *Shank3^{fx/+}; CreER^{+/-}* mice but did not display an increase in *Shank3* expression. **a.** Behavioral testing for sociability was performed using the same three-chamber enclosure, wherein mice were presented with a novel animal (purple) vs. novel inanimate object (hash). Error bars represent the standard error of the mean. Unlike the *Shank3^{fx/+}; CreER^{+/-}* mice in which *Shank3* expression was restored, these animals displayed no change in sociability before (0 weeks) compared after (5 weeks) TMX administration (paired t-test, $ts(58) = 1.34$, $p = 0.19$) $n = 29$ and $n = 30$ three chamber trials in $n = 11$ Cre^{-/-} mice pre and post TMX administration, respectively. **b.** We recorded from 96 neurons in these *Shank3^{fx/+}; CreER^{-/-}* mice. Proportion of task-modulated neurons that encoded for self- and other-experience. Unlike the *Shank3^{fx/+}; CreER^{+/-}* mice in which *Shank3* expression was restored, these animals displayed no increase in neurons that encoded for other-experience before (0 weeks) and after (5 weeks) TMX administration. Of recorded neurons, 79.2% ($n = 76$ of 96) were task modulated across all recording sessions. Of task modulated neurons,

n = 20 and n = 31 neurons encoded experience valence recorded from n = 11 Cre^{-/-} mice before and after TMX administration, respectively. Pre-tamoxifen baseline (week 0) demonstrated 45% (n = 9 of 20) and 30% (n = 6 of 20) neurons encoding for self- and other-experience, respectively (chi-square, $\chi^2(1) = 0.96$, p = 0.33). Post-tamoxifen recordings (week 5) did not demonstrate relative change in encoding patterns, with 36% (n = 11 of 31) and 23% (n = 7 of 31) of neurons encoding self- and other-experience, respectively (Chi-square, $\chi^2(1) = 1.25$, p = 0.26). Additional detail on the temporal profile of behavioral and neuronal encoding changes can be seen in Figure 4b.

a Specific Relation between Neuronal Encoding in the mPFC and Social Behavior



b Confirming Localized Increase of SHANK3 in the mPFC



c Confirming the Effect of Localized mPFC inhibition

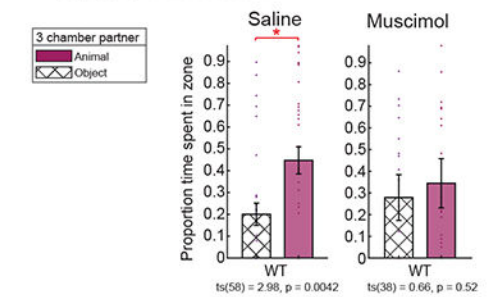


Fig. 7. Restoration of sociability is dependent on social-agency encoding in the mPFC.

a. To further evaluate the spatial selectivity of *Shank3* expression in the mPFC, the Cre-lox system was activated locally using endoxifen injection. Behavioral and neuronal recordings were obtained from each animal over the course of 8 weeks after endoxifen administration, and compared to data from. Data from $n=5$ *Shank3^{flx/+}:CreER^{+/-}* mice in which endoxifen was administered to the mPFC were compared to data from $n=4$ *Shank3^{flx/+}:CreER^{-/-}* mice in which endoxifen was given but which lacked the CreER recombinase gene. In the *top* panel, the curves represent the relative preference (\pm S.E.M.) of the mice for another animal compared to an object at weekly intervals. The horizontal line represents no preference. Cre^{+/-} mice displayed a significant increase in sociability over time compared to baseline (weeks 0-1), whereas Cre^{-/-} mice did not. Time points in which there was a significant difference are underlined in red (one-sample t-tests, $p < 0.01$). In the *middle* panel, the curves represent the proportion (\pm S.E.M.) of other-encoding compared to self-encoding neurons in the mPFC at weekly intervals. The horizontal line represents an equal proportion of neurons. Cre^{+/-} mice displayed a significant increase in other-encoding neurons over time compared to baseline (weeks 0-1), whereas Cre^{-/-} mice did not. Time points in which there was a significant difference are underlined in red ($p < 0.01$ with Bonferonni correction for repeated comparisons). In the *bottom* panel, the lines represent the degree of cross-correlation between changes in social behavior and change neuronal encoding. Red lines indicate time lags that were significant, with the arrow indicating the maximal time lag at which change in social behavior followed change in neuronal encoding (0 weeks for the *Shank3^{flx/+}:CreER^{+/-}* mice). For additional control comparison, no change in social behavior was observed in animals in which no endoxifen was given (Extended Data Fig. 8). **b.** Confirmation of localized increase in SHANK3 levels following endoxifen was

made histologically. On the *left*, increase in SHANK3 levels was confined to the cingulate gyrus of the mPFC in the Cre+/- mice. On the *right*, no increase was observed in the mPFC. **c.** Neural activity was locally inhibited in the mPFC. For control comparison, the same animals were also injected with saline (on randomly interleaved days). The proportions of time spent with another animal vs. object are shown on the *left*. Whereas WT mice receiving saline continued to display a significant preference for another animal compared to object, the same animals receiving muscimol in the mPFC did not (** $t(58) = 2.98$, $p = .0042$, paired t-tests). $n = 29$ and $n = 19$ trials from from $n = 8$ WT mice injected with saline or muscimol in the dmPFC, respectively. Localizations of the injection sites were confirmed histologically (Supplementary Fig. 6).