HHS Public Access

Author manuscript

Neuron. Author manuscript; available in PMC 2019 November 07.

Published in final edited form as:

Neuron. 2018 November 07; 100(3): 700–714.e9. doi:10.1016/j.neuron.2018.08.043.

Distinct and Dynamic ON and OFF Neural Ensembles in the Prefrontal Cortex Code Social Exploration

Bo Liang[#1], **Lifeng Zhang**[#1], **Giovanni Barbera**[1], **Wenting Fang**[1,2], **Jing Zhang**[1,3], **Xiaochun Chen**[2], **Rong Chen**[4], **Yun Li**[1,5,8,*], and **Da-Ting Lin**[1,6,*]

[1]Intramural Research Program, National Institute on Drug Abuse, National Institutes of Health, 333 Cassell Drive, Baltimore, MD 21224, USA

[2]Department of Neurology, Union Hospital, and Fujian Key Laboratory of Molecular Neurology, Fujian Medical University, Fuzhou, Fujian 350001, China

[3]Fujian Institute of Geriatrics, Union Hospital, and Fujian Key Laboratory of Molecular Neurology, Fujian Medical University, Fuzhou, Fujian 350001, China

[4]Department of Diagnostic Radiology and Nuclear Medicine, University of Maryland School of Medicine, 100 N Greene Street, Baltimore, MD 21205, USA

[5]Department of Zoology and Physiology, University of Wyoming, 1000 E University Avenue, Laramie, WY 82071, USA

[6]The Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, MD 21205, USA

[#] These authors contributed equally to this work.

Abstract

The medial prefrontal cortex (mPFC) is important for social behavior, but the mechanisms by which mPFC neurons code real-time social exploration remain largely unknown. Here, we utilized miniScopes to record calcium activities from hundreds of excitatory neurons in the mPFC while mice freely explored restrained social targets, in the absence or presence of the psychedelic drug phencyclidine (PCP). We identified distinct and dynamic ON and OFF neural ensembles that displayed opposing activities to code real-time behavioral information. We further illustrated that ON and OFF ensembles tuned to social exploration carried information of salience and novelty for social targets. Finally, we showed that dysfunctions in these ensembles were associated with

[*]Correspondence: yli30@uwyo.edu (Y.L.) and da-ting.lin@nih.gov (D.-T.L.).

[8]Lead Contact

Author Contributions:

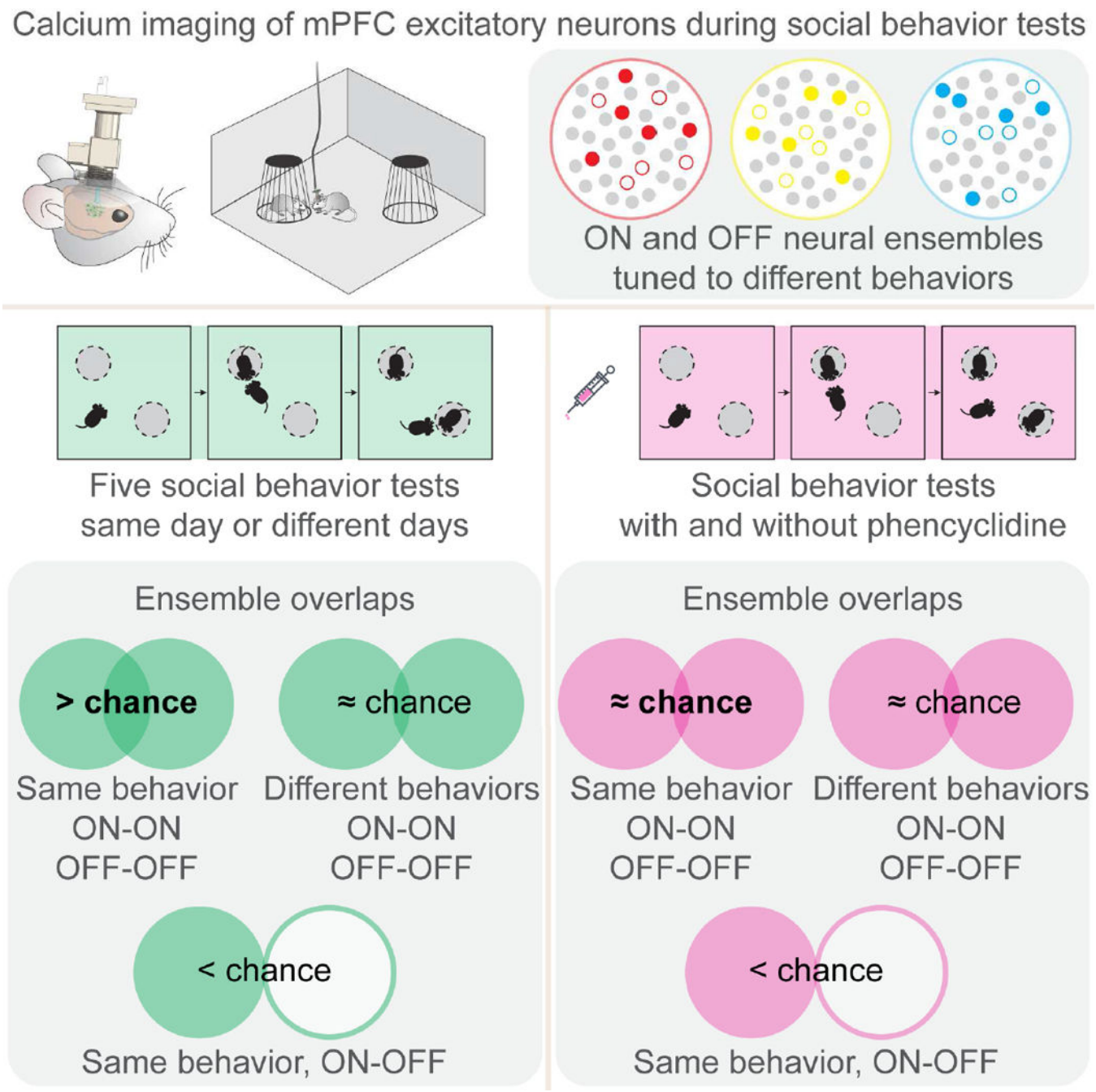
DTL and YL conceptualized the project. YL designed all the experiments. BL, GB, LFZ and YL constructed the experimental setup and built the custom imaging system. LFZ and YL performed surgery and data acquisition. WF and JZ performed ablation experiment. XC provided data analytic tools and supervised WF and JZ. BL, YL, GB and RC performed data analysis. DTL supervised all aspects of the project. YL, BL, and DTL wrote the manuscript with inputs from all other authors.

Competing Financial Interests Statement:

The authors declare no competing financial interests.

abnormal social exploration elicited by PCP. Our findings underscore the importance of mPFC ON and OFF neural ensembles for proper exploratory behavior including social exploration, and pave the way for future studies elucidating neural circuit dysfunctions in psychiatric disorders.

eTOC Blurb



Liang et al. employed miniScope to image calcium activities from principle neurons in the medial prefrontal cortex when mice freely explored restrained social targets. They identified distinct and dynamic ON and OFF neural ensembles code social exploration.

Introduction

Social behavior is one of the most fundamental complex behaviors. Deficits in social behavior represent common pathological features for a variety of psychiatric disorders including schizophrenia and autism spectrum disorder (Chevallier et al., 2012; Green et al., 2015). Multiple brain regions are involved in social cognition, including the amygdala (Adolphs, 2009; Hong et al., 2014), nucleus accumbens (NAc) (Dolen et al., 2013), hypothalamic nuclei (Anderson, 2016; McHenry et al., 2017), midbrain (Franklin et al., 2017; Gunaydin et al., 2014; Matthews et al., 2016; Molas et al., 2017), hippocampus (Okuyama et al., 2016), and prefrontal cortex (Bicks et al., 2015; Kim et al., 2015; Nakajima et al., 2014).

Several lines of evidence suggest that neural activity from the medial prefrontal cortex (mPFC) is important for rodent social behavior. For instance, bilateral lesion of mPFC neurons reduced sociability (Murray et al., 2015), and transient excitation of excitatory neurons in the mPFC disrupted sociability (Yizhar et al., 2011). In addition, mice lacking expressions of several molecules, including MeCP2 (Sceniak et al., 2016), β2 nicotinic receptor (Avale et al., 2011), insulin receptor substrate protein (IPSp53) (Chung et al., 2015) and Shank3 (Duffney et al., 2015; Lee et al., 2015), showed altered neural activity in the mPFC and displayed abnormal social behavior.

Previous studies demonstrated that some mPFC principle neurons enhanced their activities during mouse social behavior (Brumback et al., 2017; Kim et al., 2015; Lee et al., 2016). A recent study showed that prelimbic neurons projecting to the NAc carried both socially and spatially relevant information during social investigation (Murugan et al., 2017). Since activities of mPFC excitatory neurons are highly heterogeneous (Pinto and Dan, 2015; Rigotti et al., 2013; Shafi et al., 2007), the mechanisms by which mPFC excitatory neurons code real-time behavioral information remain largely unknown.

Here, we employed the miniature fluorescence microscope (miniScope) to simultaneously record *in vivo* calcium activity from excitatory neurons in the mPFC while mice freely explored restrained social targets. Our results showed that mPFC excitatory neurons formed two non-overlapping ensembles displaying opposing calcium activities (ON and OFF ensembles) to code real-time behavioral information. In addition, ON and OFF neural ensembles were highly dynamic, but the trial-to-trial stochasticity was restricted within specific boundaries. Moreover, social exploration tuned ON and OFF ensembles displayed preferential activity changes when mice explored the more salient and novel social target. Finally, systemic administration of the psychedelic drug phencyclidine (PCP) elicited deficits in social exploration, which was associated with reorganized and dysfunctional mPFC neural ensembles. Together, our results suggest that distinct and dynamic ON and OFF neural ensembles in the mPFC code exploratory behavior including social exploration.

Results

In vivo calcium imaging of mPFC when mice freely explored restrained social targets

To assess mouse social exploration, we set up a social behavior test based on the previously established protocol known as “Crawley’s sociability and preference for social novelty test” (Moy et al., 2004; Moy et al., 2007). The entire test was designed to gradually increase the complexity of social context and consisted of three consecutive 10-minute testing stages: “habituation”, “sociabili ty”, and “social novelty” (see STAR Methods). In our experimental setup, bilateral ablation of excitatory neurons (Figure S1A) or a brief optogenetic disturbance of excitatory neural activities (Figure S1B) in the mPFC abolished both sociability and preference for social novelty, suggesting that maintaining excitation/inhibition balance is important for proper social behavior (Yizhar et al., 2011).

To illustrate how mPFC excitatory neurons code real-time social exploration, we used our custom miniScope imaging system (Barbera et al., 2016) to concurrently record calcium activities from hundreds of excitatory neurons in the dorsal prelimbic cortex during social behavior test (see STAR Methods, Figure 1A-B and Movie S1). Mouse behaviors were manually annotated frame-by-frame into several categories: direct-exploration, proximity-exploration, irrelevant-exploration, ambulation, and self-directed activities (see STAR Methods and Movie S2). Behavioral analysis from 18 mice revealed several changes evoked by the presence of restrained social targets (Figure S1C). Subject mice spent more time on direct-exploration and less time on irrelevant-exploration and self-directed activities at the “sociability” and the “social novelty” stages than at the “habituation” stage (Kruskal-Wallis test and Dunn’s post hoc test, $p < 0.01$ to $p < 0.0001$). In particular, subject mice spent longer time on direct-exploration and proximity-exploration of stranger1 than those of the empty container at the “sociability” stage (Mann-Whitney test, $p < 0.0001$), and on direct-exploration and proximity-exploration of stranger2 than those of stranger1 at the “social novelty” stage (Mann-Whitney test, $p < 0.0001$ and $p < 0.01$; Figure 1C and Figure S1D). Both the duration and the frequency of direct-exploration were increased (Mann-Whitney test, $p < 0.05$ to $p < 0.0001$; Figure S1E). Together, these observations underscore direct-exploration as a reliable behavioral index reflecting sociability and preference for social novelty.

We detected and extracted calcium transients from active neurons in the mPFC of subject mice (see STAR Methods). On average, approximately 200 excitatory neurons were identified from each mouse, and ~ 80% of all imaged neurons stayed active during each 10-minute testing stage (Figure 1D). Calcium activities of individual excitatory neurons in the mPFC appeared highly heterogeneous (Figure 1E). At the population level, nearly half of the neurons decreased activity and one third of the neurons increased activity at the “sociability” and the “social novelty” stages (Figure 1F). Additionally, there were ~ 7% of the neurons exhibiting increased or decreased activity specifically at either the “sociability” stage or the “social novelty” stage.

Identification of mPFC neurons whose calcium activities tuned to specific behaviors

To identify neurons whose calcium activities tuned to specific behaviors, we performed a similarity comparison (Carrillo-Reid et al., 2015; Hamm et al., 2017), by calculating each neuron's observed calcium-behavior similarity and comparing with its own chance level of calcium-behavior similarity (see STAR Methods and Figure S2A-C). Three annotated behaviors were assessed: Direct-exploration was a behavioral index reliably reflecting sociability and preference for social novelty, while self-directed activities and irrelevant-exploration both served as controls. We selected neurons exhibiting high calcium-behavior similarity and neurons exhibiting high calcium-behavior dissimilarity, both beyond what would be expected by chance. We defined neurons with high calcium-behavior similarity as ON neurons, since they showed increased activity at the behavior onset and decreased activity at the end of the behavior epoch. Conversely, we defined neurons with high calcium-behavior dissimilarity as OFF neurons, since they displayed suppressed activity at the behavior onset and increased activity at the end of the behavior epoch (Figure S2D-E). We therefore identified a total of six subgroups of behaviorally tuned neurons at each 10-minute testing stage, including Direct-ON and Direct-OFF neurons for direct-exploration (Figure 2A), Self-ON and Self-OFF neurons for self-directed activities, Irrelevant-ON and Irrelevant-OFF neurons for irrelevant-exploration (Figure S2F). These ON and OFF neurons were spatially scattered and intermixed (Figure 2B, Figure S2G-H).

We first evaluated the reliability of a neuron's response across behavior epochs within each 10-minute testing stage. Considering each behavior epoch as a single trial, we found that ON and OFF neurons exhibited trial-to-trial stochasticity such that they were not engaged in all behavior epochs. For each individual behavior epoch, we labeled an epoch as "engaged" for an ON neuron when it exhibited increased activity. Conversely, we labeled an epoch as "engaged" for an OFF neuron when it exhibited suppressed activity. We calculated the percentage of "engaged" epochs for ON and OFF neurons across all behavior epochs, and found that they displayed an average of 54.9 ± 0.4% and 62.6 ± 0.5% consistency within each 10-minute testing stage, respectively (Figure 2C). In most behavior epochs, there were more than 30% ON neurons and more than 50% of OFF neurons actively participating.

We then calculated the averaged calcium activity across all behavioral epochs for each ON and OFF neuron. We sorted ON neurons based on the time of their peak activities, and OFF neurons based on the time of their minimal activities (Figure 2D). We found that a subset of ON and OFF neurons displayed transient activity changes before/at the onset of behavior, but the majority of ON and OFF neurons displayed transient activity changes only after the behavior onset.

We next calculated the calcium-behavior correlation for ON and OFF neurons, and compared the correlation coefficient using the averaged group calcium activity with those using individual calcium activities (Figure 2E and Figure S3A). We found that the averaged group activity of ON and OFF neurons exhibited correlation nearly twice as strong as those of individual neurons (Wilcoxon signed rank test, $p < 0.01$ to $p < 0.0001$), with ON neurons displaying a positive correlation and OFF neurons displaying a negative correlation. These results suggest that while calcium activities from individual neurons may vary across behavioral epochs, groups of behaviorally tuned neurons as a whole could compensate for

individual variations and become more reliable for coding. Alternatively, the coding strategies of mPFC neurons could be highly distributed, and differences in response latencies could correlate stronger with the binary behavioral vector when averaged together. Nevertheless, the behavior-relevant information was carried likely by a group of neurons rather than by a single neuron. We therefore described these specialized ON and OFF neurons as "neural ensembles" and propose that ON and OFF neural ensembles code real-time behavioral information with opposing activities. Each ON and OFF neural ensemble tuned to a specific behavior occupied only 3 ~ 11% of all imaged neurons at each testing stage (Figure 2F).

Mouse behavior can be reliably decoded using calcium activities from ON and OFF neural ensembles in the mPFC

To determine if ON and OFF neural ensembles code behavior-relevant information, we performed decoding analyses to test if mouse behavior could be predicted using calcium activity of ON and OFF neurons. We found that the calcium activity from either ON or OFF neural ensemble by itself can be used to predict mouse behavior with similar accuracy (Figure 3A, Friedman test and Dunn's post hoc test, $p > 0.05$). Decoding performance using ON and OFF neural ensembles together was similar to those using ON or OFF ensembles alone (Friedman test and Dunn's post hoc test, $p > 0.05$). Importantly, decoding accuracy using ON and OFF neural ensembles together was close to the maximal accuracy achieved using all imaged neurons, and the differences were not statistically significant (Friedman test and Dunn's post hoc test, $p > 0.05$).

We next determined functional connectivity within and between ensembles. For each annotated behavior, we calculated the pairwise correlation coefficient of calcium activities within ON ensembles, within OFF ensembles, and between ON and OFF ensembles (see STAR Methods). As controls, we calculated the pairwise correlation coefficient among other neurons not tuned to the three annotated behaviors, and between ON or OFF ensembles with other neurons (Figure 3B). We found a similar pattern of correlation for each annotated behavior, such that the correlation was strongest within ON ensembles, followed by within OFF ensembles, and finally between ON and OFF ensembles tuned to the same behavior. In addition, the correlation strength within ON ensembles, within OFF ensembles, and between ON and OFF ensembles tuned to the same behavior, were consistently higher than controls that were close to zero (Kruskal-Wallis test and post hoc Dunn's multiple comparison test, $p < 0.001$). These results indicate the existence of functional connectivity between ON and OFF ensembles tuned to the same behavior.

Dynamic population coding with ON and OFF neurons in the mPFC

Neural ensembles tuned to the same behavior identified at different testing stages were not identical, yet they shared common neurons across stages (Figure S3B). We therefore defined the across-stage overlap between ON-ON and OFF-OFF neural ensembles tuned to the same behavior as "ON-ON and OFF-OFF intra-behavior" overlap (Figure 4A), whereas the across-stage overlap between ON-ON and OFF-OFF neural ensembles tuned to different behaviors as "ON-ON and OFF-OFF inter-behavior" overlap (Figure 4A). We found that the ON-ON and OFF-OFF intra-behavior overlaps were much higher than that predicted by

chance based on random sampling of all imaged neurons (Wilcoxon signed rank test, $p < 0.0001$), while the ON-ON and OFF-OFF inter-behavior overlaps were similar to that predicted by chance (Figure 4A and Figure S3C). Moreover, we defined the across-stage overlap between ON and OFF neural ensembles tuned to the same behavior as "ON-OFF intra-behavior overlap" (Figure 4B), and found that the degree of across-stage ON-OFF intra-behavior overlap was significantly lower than that predicted by chance (Wilcoxon signed rank test, $p < 0.0001$; Figure 4B). The existence of the above-chance ON-ON and OFF-OFF intra-behavior overlap, and the below-chance ON-OFF intra-behavior overlap across testing stages, indicates that the recruiting strategy for mPFC neural ensembles is not a random process, instead controlled by specific rules to define boundaries for ON and OFF neural ensembles.

The across-stage ON-ON and OFF-OFF intra-behavior overlap was three times higher than the chance level. However, the absolute value (~ 12%) seemed surprisingly low. To minimize the possible influence from different social context on across-stage comparisons, we next split each 10-minute recording into two 5-minute halves, re-identified ON and OFF neurons tuned to direct-exploration during each of the two 5-minute halves, then calculated the ON-ON and OFF-OFF intra-behavior overlap between the two 5-minute halves (Figure 4C). We found that there was 17.6% to 27.2% within-stage overlap for Direct-ON ensembles and 10.3% to 13.9% within-stage overlap for Direct-OFF ensembles. These results suggest that Direct-ON and Direct-OFF neural ensembles are highly dynamic, and that the majority of identified neurons are different even between the two halves of the same stage.

Next, we generated a union ensemble by combining neurons identified from the two 5-minute halves (Half1 U Half2) for both Direct-ON and Direct-OFF neurons. We then measured the percentage of overlap between the union ensembles and the originally identified Direct-ON and Direct-OFF ensembles based on the full-length 10-minute recording (Figure 4D). We found that over 75% of the originally identified Direct-ON and over 60% of the originally identified Direct-OFF neurons belonged to the union ensembles of the two 5-minute halves.

To further determine if neurons tuned to specific behaviors could be repeatedly identified, we repeated *in vivo* calcium imaging during social behavior tests five times (Figure S4A). We compared the behaviorally tuned neurons identified from two sequential social behavior tests, performed on different days with stranger mice presented at the same locations (Figure 5A), performed on different days with stranger mice presented at different locations (Figure 5B), or performed on the same day with stranger mice presented at different locations (Figure 5C).

To this end, we detected and aligned all active neurons from each social behavior test (see STAR Methods and Figure S3D). We then identified ON and OFF neurons for each of the three annotated behaviors (i.e., direct-exploration, self-directed activities, irrelevant-exploration) at individual stages of each test, and determined how many neurons could be repeatedly identified for the same behavior during two independent social behavior tests. We first calculated for each annotated behavior, the ON-ON and OFF-OFF intra-behavior overlap across the same stages of the two tests (Figure S4B), and those across different

stages of the two tests (Figure S4F). We found statistically significant above-chance ON-ON and OFF-OFF intra-behavior overlap for “direct-exploration” and “irrelevant-exploration” in most cases of comparisons (Wilcoxon signed rank test, $p < 0.05$; Figure S4C-E, G-I). We then pooled all three annotated behaviors and calculated the averaged ON-ON and OFF-OFF intra-behavior overlap and the averaged ON-ON and OFF-OFF inter-behavior overlap across same or different stages (Figure S5A). We found statistically significant above-chance intra-behavior overlap across same and different stages in almost every case of comparison (Wilcoxon signed rank test, $p < 0.05$; Figure S5B-D). In contrast, the inter-behavior overlap was similar to the chance level.

Given that there was no difference in overlap comparisons across same or different stages of the two social behavior tests, we generated a union ON (or OFF) ensemble for each annotated behavior in each test by pooling together all ON (or OFF) neurons tuned to the same behavior from all three testing stages. We then compared the across-test overlaps among these union ensembles. For all three scenarios of repeated social behavior tests, we found that the across-test ON-ON and OFF-OFF intra-behavior overlaps were higher than what would be expected by chance, the across-test ON-ON and OFF-OFF inter-behavior overlaps were similar to chance levels, and the across-test ON-OFF intra-behavior overlap were lower than what would be expected by chance (Figure 5A-C). We calculated the cumulative ensemble size for Direct-ON and Direct-OFF neural ensembles across all five independent social behavior tests, and compared this to the predicted cumulative size based on random selection of the same number of neurons from all imaged neurons. We found that the cumulative ensemble size was smaller than what would be expected by random selection from all imaged neurons (Figure S5E).

Together, these results revealed the above-chance ON-ON and OFF-OFF intra-behavior overlap, below-chance ON-OFF intra-behavior overlap, and chance level ON-ON and OFF-OFF inter-behavior overlap. This observation suggests that the recruiting strategy for ON and OFF ensembles in the mPFC is not a random process. Moreover, the frequent occurrence of common neurons coding for the same behavior suggests the existence of specialized neural ensembles coding for distinct behaviors. We therefore proposed a “dynamic population coding model” (Figure 6A). In this model, two parallel neural ensembles (ON and OFF) in the mPFC are specialized to code a specific behavior *via* opposing activities. All neurons within each specific ensemble have the potential to be recruited to code behavior, but at any given time, only a portion of neurons from specific ON and OFF ensembles are actually recruited to code real-time behavioral information. Therefore, real-time behavioral information is carried by the “engaged subsets” of neurons selected from the ON and OFF neural ensembles.

We also computed within-stage overlap between ON and OFF neural ensembles tuned to different annotated behaviors, and found considerable ON-OFF inter-behavior overlap within individual testing stages (Figure S5F). Based on our data, we constructed an overlap illustration for six different ON and OFF neural ensembles tuned to the three annotated behaviors (Figure 6B). The ON and OFF ensembles tuned to the same behavior are mutually exclusive. The three ON ensembles tuned to different behaviors are largely non-overlapping. Similarly, the three OFF ensembles tuned to different behaviors are largely non-overlapping.

The ON and OFF ensembles tuned to different annotated behaviors display some degree of overlap.

Acute PCP administration reconstructed mPFC neural ensembles

We next repeated *in vivo* calcium imaging with the same 18 mice during social behavior test while they were under acute pharmacological influence of PCP (see STAR Methods and Figure 7A). PCP is a non-competitive N-methyl-D-aspartate (NMDA) receptor antagonist and an abused psychedelic drug (street name Angel Dust). Acute administration of PCP elicits schizophrenia-like symptoms in healthy human subjects (Javitt and Zukin, 1991; Luby et al., 1959), and PCP administered rodents are commonly used as an animal model to study pathogenic mechanisms for schizophrenia (Jentsch and Roth, 1999; Neill et al., 2010).

In the absence of PCP, the presence of restrained social targets during the “sociability” and the “social novelty” stages elicited longer time for the subject mice to spend on direct-exploration, and shorter time on irrelevant-exploration and self-directed activities compared to the “habituation” stage (Figure S1C). Interestingly, under acute PCP influence, the presence of restrained social targets no longer elicited such changes (Figure S6A). At the “sociability” stage under acute PCP influence, although mice still spent relatively longer time on direct-exploration and proximity-exploration of stranger1 than that of empty container (Mann-Whitney test, $p < 0.0001$; Figure 7B and Figure S6B-C), the amount of time spent on direct-exploration of stranger1 under acute PCP was only half of that in the absence of PCP. Moreover, at the “social novelty” stage under acute PCP influence, mice spent a similar amount of time on direct-exploration and proximity-exploration of stranger2 to those of stranger1 (Mann-Whitney test, $p > 0.05$). These observations indicate a diminished sociability and a completely abolished preference for social-novelty under acute PCP influence.

We next identified active neurons and extracted individual calcium transients. We found that acute systemic administration of PCP elicited a small but statistically significant increase in population calcium activity of the mPFC compared to that of saline injection (Kolmogorov-Smirnov test, $p < 0.001$; Figure S6D-E), consistent with results from previous *in vivo* electrophysiological studies (Katayama et al., 2007; Suzuki et al., 2002). Under the acute influence of PCP, we could still identify behaviorally tuned ON and OFF neural ensembles (Figure S7A-D). Many features of these ON and OFF ensembles identified under PCP influence remained similar to those in the absence of PCP, including the ensemble size, calcium-behavior correlation, and across-stage overlap within the social behavior test under PCP influence (Figure S7E-H).

Interestingly, between the two tests performed in the absence or presence of PCP, the ON-ON and OFF-OFF intra-behavior overlaps were similar to the ON-ON and OFF-OFF inter-behavior overlaps and those predicted by chance (Wilcoxon signed rank test, $p > 0.05$; Figure 7C), unlike comparisons between the two tests both performed in the absence of PCP (Figure 5). The ON-OFF intra-behavior overlap was still lower than what would be expected by chance (Wilcoxon signed rank test, $p < 0.05$). These observations suggest that PCP partially disrupts the boundaries of neural ensembles originally formed under physiological

conditions, such that a different set of neurons are recruited to code real-time behavioral information.

Direct-ON and Direct-OFF neural ensembles carried information relevant to social salience and novelty

In our social behavior test, the preferential exploration at the “sociability” and “social novelty” stages by the subject mice could be exploration of social salience or exploration of novel stimuli. To determine whether preferential exploration could be elicited solely by a novel stimulus using our behavioral setup, we implemented two controls of novel stimuli (i.e., with novel objects placed inside the containers) during our social behavior test (Figure S8A). We found that the presence of a novel object inside the container by itself could not elicit the preferential exploration from the subject mice. Therefore, the preferential exploration by the subject mice under our experimental conditions was induced by the combinatorial effect of both social salience and novel stimuli.

To determine if Direct-ON and Direct-OFF neural ensembles carry information of salience/novelty for the explored targets, we compared their calcium activities while mice explored the two different targets during each testing stage (Figure 8A). We reasoned that the two empty containers at the “habituation” stage would represent similar salience/novelty; at the “sociability” stage, stranger1 would represent higher salience/novelty than the empty container; at the “social novelty” stage, stranger1 and stranger2 would be equally salient, while stranger2 would represent higher novelty than stranger1. We found that Direct-ON and Direct-OFF ensembles exhibited similar activity when mice directly explored the two empty containers at the “habituation” stage (Mann-Whitney test, $p > 0.05$). At the “sociability” stage, the Direct-ON ensemble displayed preferentially enhanced activity, whereas the Direct-OFF ensemble displayed preferentially suppressed activity during direct-exploration of stranger1 (Mann-Whitney test, $p < 0.0001$). At the “social novelty” stage, only the Direct-ON ensemble displayed preferential activity changes for direct-exploration of stranger2 (Mann-Whitney test, $p < 0.05$). These results suggest that Direct-ON and Direct-OFF neural ensembles display preferential activity changes to code salience/novelty.

To further determine if Direct-ON and Direct-OFF neural ensembles respond solely to a novel stimulus, we performed *in vivo* calcium imaging during the novel object recognition (NOR) test (see STAR Methods). We found that subject mice spent slightly longer time on direct-exploration of the novel object compared to that of the familiar object during the “testing” stage (Figure S8B, Mann-Whitney test, $p < 0.01$). We identified Direct-ON and Direct-OFF neural ensembles during the “familiarization” and the “testing” stages of the NOR test, and compared their calcium activities while mice explored the two objects at each stage. We found that Direct-ON and Direct-OFF ensembles exhibited similar activity when mice explored the two identical objects at the “familiarization” stage. At the “testing” stage, on ly the Direct-ON ensemble displayed preferential activity changes for direct-exploration of the novel object (Figure S8B, Mann-Whitney test, $p < 0.01$). These results indicate that the Direct-ON ensemble, but not the Direct-OFF ensemble, is sensitive to novel stimuli.

Finally, we compared calcium activities of Direct-ON and Direct-OFF ensembles while mice explored the two different targets at each stage of the social behavior test under acute PCP

influence (Figure 8B). We found that Direct-ON and Direct-OFF ensembles still exhibited similar activity when mice directly explored the two empty containers at the “habituation” stage (Mann-Whitney test, $p > 0.05$). At the “sociability” stage under acute PCP influence, the Direct-ON ensemble still exhibited preferentially enhanced activity for direct-exploration of stranger1 (Mann-Whitney test, $p < 0.0001$), whereas the Direct-OFF ensemble displayed similar activity between direct-exploration of stranger1 and that of the empty container (Mann-Whitney test, $p > 0.05$). At the “social novelty” stage under acute PCP influence, both Direct-ON and Direct-OFF ensembles displayed similar activity between direct-exploration of stranger2 and that of stranger1 (Mann-Whitney test, $p > 0.05$). These results correlated well with the diminished sociability and completely abolished preference for social novelty on the behavioral level under acute PCP influence (Figure 7B).

Taken together, these results suggest that the Direct-ON neural ensemble codes both social salience and novelty, whereas the Direct-OFF neural ensemble codes only social salience but not novelty. Acute PCP administration selectively elicits dysfunctions in Direct-ON and Direct-OFF ensembles, such that the Direct-OFF ensemble no longer responds to social salience and the Direct-ON ensemble no longer responds to social novelty.

Discussion

We reported here that mPFC excitatory neurons formed distinct ensembles tuned to specific behaviors, including direct-exploration, irrelevant-exploration, and self-directed activities. For each specific behavior, there are two parallel neural ensembles using opposing activities (ON and OFF, respectively) to code behavioral information. We propose that the behaviorally tuned ON and OFF ensembles are highly dynamic, and that at any given time, only a portion of neurons from the ensemble are actively engaged in real-time behavioral coding. We further demonstrated that ON and OFF neural ensembles tuned to direct-exploration also carried information of social salience and novelty, and that the abnormal social exploration induced by PCP, a psychedelic drug, was associated with reorganized and dysfunctional mPFC ON and OFF ensembles.

Behaviorally tuned ON and OFF neural ensembles in the mPFC

The co-existence of principle neurons positively and negatively correlated with locomotion has been reported in both the premotor and the motor cortex (Garcia-Junco-Clemente et al., 2017; Peters et al., 2017). Existing evidence also supports the presence of bidirectional modulation of excitatory neuronal activities across various sensory cortices during specific task performance (Krupa et al., 2004; Kuchibhotla et al., 2017; Zhang et al., 2014), likely through a top-down influence from higher order brain regions such as the mPFC. However, most studies carried out in the mPFC have focused on the identification or manipulation of task-related neural assemblies with enhanced activities (Brumback et al., 2017; Dejean et al., 2016; Fujisawa et al., 2008; Kvitsiani et al., 2013; Murugan et al., 2017; Pinto and Dan, 2015).

Here we demonstrated the co-existence of behaviorally tuned ON and OFF ensembles among mPFC excitatory neurons. Both ON and OFF ensembles carried real-time behavioral information (Figure 3A). ON and OFF ensembles tuned to the same behavior were spatially

intermingled (Figure 2B and Figure S2G-H), but functionally correlated (Figure 3B). Specific rules were in place to minimize recruiting the same neurons into both ON and OFF ensembles coding for the same behavior (Figure 4B and Figure 5). The underlying mechanisms governing the recruitment of neurons into ON and OFF ensembles in the mPFC remain to be elucidated. It is possible that mPFC principle neurons receive different inhibitory and dis-inhibitory inputs from local interneurons (Garcia-Junco-Clemente et al., 2017; Hattori et al., 2017; Kvitsiani et al., 2013; Nakajima et al., 2014; Pinto and Dan, 2015; Zhang et al., 2016), such that ON and OFF neurons are functionally predetermined by their anatomical connectivity. It is also conceivable that ON and OFF neurons in the mPFC might project to different groups of downstream neurons for efficient top-down control.

Dynamic population coding model

Hebb's "cell assembly" hypothesis states that information is not represented by a single neuron but rather by a set of neurons (Hebb, 1949). Hebbian neural ensembles can have intersections, such that one neuron could belong to many ensembles. Many of our results fit Hebb's "cell assembly" hypothesis. For instance, the calcium-behavior correlation calculated using group calcium activity was twice as strong as that using individual calcium activities, indicating that behavioral information was better represented by a group of neurons rather than by a single neuron (Figure 2E and Figure S3A). Moreover, we showed that distinct ensembles were tuned to different annotated behaviors, and that one neuron can belong to the Direct-ON ensemble and simultaneously to the Self-OFF ensemble within the same testing stage (Figure S5F).

Importantly, we revealed new features of mPFC neural ensembles beyond Hebb's original cell assembly hypothesis. Our results showed that behaviorally tuned ON and OFF ensembles in the mPFC were stochastic (Figure 4C), for what is likely a combination of biological and analytical reasons. Epoch-to-epoch variability in activation of ON/OFF cells could be due to differences in the animal's behavior or sensory cues between behavior epochs. Our results also showed that the recruiting strategy was not a random process, such that the trial-to-trial stochasticity was restricted within specific boundaries (Figure 4A-B, Figure 5, Figure S4 and S5). We proposed a dynamic population coding model in the mPFC (Figure 6A), based on Hebb's cell assembly hypothesis (Hebb, 1949), while incorporating Edelman's concept of degeneracy (Edelman and Gally, 2001). Degeneracy refers to the ability of structurally different elements to perform the same function. It has been proposed as a ubiquitous biological property that leads to complexity and robustness in biological systems (Edelman and Gally, 2001). Our dynamic population coding model suggests degeneracy existing in prefrontal neural codes, such that different subsets of neurons from ON and OFF ensembles could code the same behavioral outcome (Figure 3A and Figure 6A). Degeneracy in dynamic population coding would offer complexity in neural codes to ensure balance between neural plasticity for learning and stability for long-term memory. For example, it was recently reported that neural representations in the parietal cortex and motor cortex underwent major reorganizations across days as mice stably performed a learned behavior (Driscoll et al., 2017; Peters et al., 2017), indicating possible degeneracy of neural codes in these brain regions. Degeneracy in dynamic population coding also would ensure robustness of prefrontal neural codes, such that mild neuronal loss during normal

aging is usually not associated with obvious cognitive dysfunction or behavioral abnormalities (West, 1999), whereas severe neuronal loss in neurodegenerative disorders such as frontotemporal dementia are mostly accompanied by drastic cognitive decline and behavioral deficits (Bang et al., 2015; Warren et al., 2013).

PCP elicited ensemble reorganization and disrupted coding of social salience/novelty information

Rodents treated with NMDA receptor antagonist are considered useful pharmacological models to study pathogenic mechanisms for schizophrenia-like symptoms (Adler et al., 1998; Breier et al., 1997; Hamm et al., 2017; Jackson et al., 2004; Javitt and Zukin, 1991; Jentsch and Roth, 1999; Katayama et al., 2007; Lahti et al., 1995; Suzuki et al., 2002; Takahata and Moghaddam, 2003). It has been proposed that systemic administration of NMDA receptor antagonist (PCP or MK801) in rodents induced elevated activities in mPFC pyramidal neurons, mainly through local inhibition on mPFC interneurons or through enhanced excitatory inputs from other brain regions such as the ventral hippocampus (Homayoun and Moghaddam, 2007; Jodo et al., 2005).

In our study, we applied PCP systemically at a dosage that elicited abnormal social exploration. We showed that in the presence of acute PCP, behaviorally tuned neural ensembles in the mPFC were reorganized (Figure 7C). We further showed that PCP selectively impaired the salience coding of the Direct-OFF ensemble and novelty coding of the Direct-ON ensemble (Figure 8B), which correlated well with partially diminished sociability and completely abolished preference for social novelty on the behavioral level under acute PCP influence (Figure 7B).

Social behavior requires proper coding for both internal states and external stimuli (Anderson, 2016). Coding of salience/novelty might be mediated *via* different neural transmitter signals. For instance, dopamine and oxytocin signals might be involved in social salience coding (Dolen et al., 2013; Gunaydin et al., 2014), while glutamatergic signals from the ventral hippocampus might be related to novelty coding (Okuyama et al., 2016). Under acute PCP influence, it seems that a new set of neurons were recruited into Direct-ON and Direct-OFF ensembles to code social exploration. It is possible that many of the newly recruited Direct-ON and Direct-OFF neurons are originally wired to code other external sensory stimuli or internal states under physiological conditions. Therefore, these newly recruited neurons not only are unable to respond to social salience or novel stimuli, but also might provide distractive information and lead to hallucination or other schizophrenia-like symptoms.

In sum, we identified distinct and dynamic ON and OFF neural ensembles in the mPFC tuned to specific behaviors. We proposed that Direct-ON and Direct-OFF neurons may also code more abstract information, such as salience and novelty of inanimate objects and conspecifics. Our approach could be applied to other mouse models of schizophrenia and autism spectrum disorder, paving the way for future studies in elucidating pathogenic mechanisms for social behavior deficits in these psychiatric disorders.

STAR METHODS

KEY RESOURCES TABLE

CONTACT FOR REAGENT AND RESOURCE SHARING

EXPERIMENTAL MODELS AND SUBJECT DETAILS

METHOD DETAILS

- Virus injection
- GRIN lens implantation
- miniScope
- Social behavior test
- Novel object recognition test
- Social behavior test with NOR control
- *In vivo* calcium imaging
- Ablation
- Immunostaining
- Optogenetics
- Behavior annotation
- Calcium image processing
- Neural activities in three testing stages
- PCP effects on population calcium activities
- Identification of behaviorally tuned neurons
- Neural ensemble correlation analysis
- Neural ensemble consistency analysis
- Neural ensemble functional connectivity
- Direct-exploration tuned ensemble activity
- Neural ensemble overlap analysis
- Behavioral variable decoding

QUANTIFICATION AND STATISTICAL ANALYSIS

- Statistics
- Sample size estimation
- Data inclusion and exclusion

DATA AND SOFTWARE AVAILABLILTY

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead contact, Dr. Yun Li (yli30@uwyo.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All experiments were conducted in accordance with the guidelines of the Institutional Animal Care and Use Committee, the Intramural Research Program, National Institute on Drug Abuse, National Institutes of Health. All mice were wild type C57BL/6J at 3 – 4 months of age and 25 – 30 gram of weight. Mice were maintained in a regular light cycle (7:00am – 7:00pm) and provided with food and water ad libitum. All mice used in this study were males.

METHOD DETAILS

Virus injection—We previously demonstrated that GFP protein driven by CamKII promoter in AAV1 virus is mostly expressed in excitatory neurons (i.e., only 1% GFP positive neurons were GABAergic neurons) once being injected into mouse prelimbic cortex (Zhang et al., 2017). In this study, for calcium imaging experiments, we unilaterally injected AAV1.CamKII.GCaMP6f.WPRE.SV40 virus (University of Pennsylvania Vector Core, 500 nl with a titer of 2.76e13 GC/ml) into the dorsal region of the mouse prelimbic cortex, using the stereotactic coordinates (A/P: +1.9 mm, M/L: +0.3 mm, D/V: −1.7 mm). For ablation experiments, we bilaterally injected AAV1 viral mix (AAV1.CamKII0.4.Cre.SV40 and AAV1.mCherry.Flex.dtA, 500 nl in a ratio of 1:1, with a title of 4.496e13 GC/ml and 2.06e13 GC/ml, respectively) into the dorsal region of the mouse prelimbic cortex. For optogenetic experiments, we bilaterally injected AAV1 viral mix (AAV1.CamKII0.4.Cre.SV40 and AAV1.CAGGS.Flex.ChR2-tdTomato.WPRE.hGH, 500 nl in a ratio of 1:1, with a title of 4.496e13 GC/ml and 1.38e13 GC/ml, respectively), or AAV1 viral mix (AAV1.CamKII0.4.Cre.SV40 and AAV1.EF1a.DIO.eYFP.WPRE.hGH, 500 nl in a ratio of 1:1, with a title of 4.496e13 GC/ml and 2.9e13 GC/ml, respectively) as controls, into the dorsal region of the mouse prelimbic cortex. Mice were anaesthetized with 2% isoflurane in oxygen at a flow rate of 0.4 liter/min and mounted on a stereotactic frame (Model 962, David Kopf Instruments). A hole was drilled on the skull using a 0.5-mm diameter round burr with a high-speed rotary micro drill (19007–05, Fine Science Tools). A total of 500 nl of virus solution were injected at a rate of 25 nl/min using a micro pump and Micro4 controller (World Precision Instruments). The injection needle was kept in the parenchyma for 5 minutes after injection. The hole on the skull was then sealed with bone wax, and skin was sutured.

GRIN lens implantation—One or two weeks after AAV injection, a 1-mm diameter GRIN lens (GRINTECH GmBH) was directly implanted into the dorsal region of the mouse prelimbic cortex under anesthesia with ketamine/xylazine (ketamine:100 mg/kg and xylazine:15 mg/kg). In brief, a 1.1mm-diameter craniotomy was first made at the coordinates (A/P: +1.9 mm. M/L: +0.7mm). Before GRIN lens implantation, brain tissue above the prelimbic cortex was aspirated using a 30-gauge blunted needle, along a direction of 10° laterally-shifted angle into a depth of −1.8 mm. The needle was attached to a custom-

constructed three-axis motorized robotic arm controlled by a MATLAB-based software (https://github.com/liang-bo/AutoStereota). The GRIN lens was then secured to the skull using dental cement (Metabond S380, Parkell).

MiniScope—Detailed information about the miniScope can be found in our previous study (Barbera et al., 2016). Design files and part list can be found in the Github repository (https://github.com/giovannibarbera/miniscope_v1.0). The miniScope consists of a blue LED (465 nm XLamp XP-E, Cree), a filter set (excitation filter, ET470/40, Chroma; dichroic mirror, FF495, Semrock; emission filter, EM525/50, Chroma), the relay optics (#83–605, #63–690, Edmund optics) and the CMOS sensor (MT9V022IA7ATM, Aptina). The miniScope housing was 3D printed (SLArmor Nickel-NanoTool, Proto Labs). The LED power was set to 0.1 – 0.3 mW at the miniScope focal plane. The image size was 400 × 400 pixels, acquired at 10-Hz frame rate. The pixel size of the acquired image was ~ 2.75 μm. Three to eight weeks after the GRIN lens implantation, the miniScope was mounted onto mouse head. After achieving the in-focus position for the entire field of view, the base of the miniScope was fixed on the skull using dental cement, and the main body of the miniScope was detached. Before each *in vivo* imaging experiment, the main body of the miniScope was connected and secured to the base on the mouse head through a side-locking screw. After data acquisition, the main body of the miniScope was detached from the base. We concurrently recorded calcium activities from hundreds of excitatory neurons in the dorsal prelimbic cortex, when mice freely performed the social behavior test (Movie S1).

Social behavior test—We set up a social behavior test based on the previously established protocol known as “Crawley’s sociability and preference for social novelty test” (Moy et al., 2004; Moy et al., 2007). Crawley’s sociability and preference for social novelty test was originally performed with a three-chamber apparatus. To facilitate *in vivo* calcium imaging, we modified the three-chamber apparatus by placing the subject mouse into one open square box (42 cm × 42 cm × 30 cm) with two small cylinder wire containers sitting at the opposite corners of a diagonal line (also see schematic illustration in Figure 1B). The behavior test was composed of three consecutive 10-minute testing stages: “habituation”, “sociability”, and “social novelty”. During the “habituation” stage, the subject mouse freely explored the apparatus with two empty wire containers. During the “sociability” stage, an age- and weight-matched never-before-met same-sex (i.e., male) conspecific (stranger1) was introduced into one container while the other container remained empty. During the “social novelty” stage, a second age- and weight-matched never-before-met same-sex conspecific (stranger2) was introduced into the previous empty container during sociability stage while the first conspecific remained in the same location. “Sociability” was defined as propensity f or a subject mouse to spend time with a conspecific rather than with an empty container during the “sociability” stage. “Preference for social novelty” was defined as propensity for a subject mouse to spend time with a previously un-encountered conspecific rather than with a relatively familiar conspecific during the “social novelty” stage.

Novel object recognition test—A novel object recognition (NOR) assay was designed to examine mouse exploratory behavior to a non-social novel stimulus (Leger et al., 2013). The NOR test contained two stages, a “familiarization” and a “testing” stage (also see

schematic illustration in Figure S8B). During the "familiarization" stage, mice freely explored two identical objects in a square box (42 cm × 42 cm × 30 cm) for 10 minutes. The two objects were placed at the same two corners of the diagonal line as that of the social behavior test. Following familiarization, mice were placed back to their home cages for 1 hour. During the "testing" stage, mice were re-introduced back into the same behavior apparatus – with one of the familiarized objects being replaced by a novel object. Mice explored these two objects for 10 minutes.

Social behavior test with NOR control—To further distinguish exploration of social salience versus exploration of novel stimuli, we performed the following social behavior test, incorporating two novel object controls (also see schematic illustration in Figure S8A). The entire behavior test was composed of five consecutive 5-minute stages, including one "habituation" stage (H stage) and four "testing" stages (T1 to T4 stages). During the H stage, mice freely explored two empty wire containers (E1 and E2) in a square box (42 cm × 42 cm × 30 cm). E1 and E2 were placed at the opposite corners of a diagonal line. Following the H stage, a novel object (object1) was placed into the previously empty container E1 during the T1 stage. During the T2 stage, object1 was replaced by a novel conspecific (stranger1). During the T3 stage, a second novel object (object2) was placed into the previously empty container E2. During the T4 stage, object2 was replaced by a second novel conspecific (stranger2).

***In vivo* calcium imaging—**A total of 18 mice were used for *in vivo* calcium imaging *via* miniScope during various behavior tests. These mice were unilaterally injected with AAV1 virus expressing calcium indicator GCamp6f mostly in excitatory neurons in the mPFC, and implanted with a 1-mm GRIN lens into the dorsal region of the prelimbic cortex. For the first batch of 9 mice, we performed *in vivo* calcium imaging with the NOR test first, followed by two social behavior tests. A week after the NOR test, the first social behavior test was performed in the absence of phencyclidine (PCP); while the second social behavior test was performed in the presence of PCP after another 5–14 days. PCP was intraperitoneally injected at a dose of 10 mg/kg, 30–50 minutes prior to the social behavior test. The objects in the NOR test and the strangers in the two social behavior tests were presented at the same locations (i.e., the same two corners of a diagonal line).

With a second batch of 3 mice and a third batch of 6 mice, we performed social behavior tests in the absence of PCP multiple times (i.e., 4 times and 5 times, for the second and third batch of mice, respectively), a NOR test, and lastly a social behavior test in the presence of PCP. The timeline of the five social behavior tests (Test1 to Test5) was illustrated in Figure S4A. Specifically, one week after social behavior Test1, social behavior Test2 was performed with strangers presented at the same locations as in Test1; one week after social behavior Test2, social behavior Test3 was performed with strangers presented at different locations (i.e., two corners on the opposite diagonal line from the Test2); one to two weeks after social behavior Test3, social behavior Test4 and Test5 were performed on the same day (1 hour apart) with strangers presented at different locations. NOR test was performed between social behavior Test1 and Test4. After 3–14 days of social behavior Test4 and Test5, the social behavior test was performed again in the presence of PCP. PCP was

intraperitoneally injected at a dose of 10 mg/kg, 30–50 minutes prior to the social behavior test. The second batch of 3 mice did not perform the social behavior Test3, while the rest of social behavior tests (i.e., Test1–2, Test4–5) were similar to those in the third batch of 6 mice. In all our social behavior tests, stranger mice were never-before-met conspecifics of the same gender (i.e., male) with matched age and weight.

Ablation—In the ablation experiment, we examined if focal ablation of excitatory neurons bilaterally in the mPFC would induce any abnormality in our social behavior test. After a pre-ablation social behavior test, we bilaterally injected AAV1 viral mix (AAV1.CamKII0.4.Cre and AAV1.mCherry.Flex.dtA) into the mPFC to express diphtheria toxin fragment A selectively in excitatory neurons in the mPFC. Five months after viral injection, a post-ablation social behavior test was performed with new pairs of strangers.

Immunostaining—After the post-ablation social behavior test, mPFC ablated mice and age-matched control mice were anesthetized with an overdose of ketamine (150 mg/kg) and xylazine (22.5 mg/kg), perfused with phosphate buffer solution (PBS) first and followed by a fixation buffer containing 4% paraformaldehyde (PFA) in PBS. Mice brains were dissected and post-fixed with 4% PFA in PBS overnight at 4°C. Fixed mice brains were sectioned *via* a Vibratome (LEICA VT1000S) and 30 μm coronal brain slices were collected. To evaluate the ablation efficiency, immunostaining of NeuN, a neuronal marker, was performed following a protocol for free-floating brain slice. In brief, floating brain slices were first washed in PBS (5 minutes for 3 times) and incubated in a blocking buffer (4% normal goat serum, 1% bovine serum albumin, 0.3% triton X-100 in PBS) with gentle rocking for two hours at room temperature. Brain slices were then incubated with a monoclonal mouse-anti-NeuN antibody (1:1000 diluted in blocking buffer, ab104224, Abcam) with gentle rocking overnight at 4°C. After washing in PBS (5 minutes for 3 times), brain slices were subsequently incubated with a secondary antibody (Alexa Fluor 488 conjugated donkey-anti-mouse IgG, 1:300 diluted in blocking buffer, Cat#715–546-150, Jackson ImmunoResearch) with gentle rocking for 1 hour at room temperature. After washing in PBS (5 minutes for 3 times), brain slices were mounted on slides for imaging. Fluorescent images were obtained through an Olympus VS120 scanner. For quantification, fluorescence intensity of NeuN staining was measured at the ablated region and an adjacent cortical region with a similar size. The ratio was calculated and compared between the ablated mice and controls.

Optogenetics—In the optogenetic experiment, we examined if transient disturbance of excitatory neural activity in the mPFC would induce any abnormalities in mice social exploration using our social behavior setup. We bilaterally injected AAV1 viral mix (AAV1.CamKII0.4.Cre and AAV1.CAGGS.Flex.ChR2-tdTomato) into the mPFC to express a light-gated cation-selective channel, channelrhodopsin-2 (ChR2), specifically in excitatory neurons in the mPFC. As a control, we bilaterally injected AAV1 viral mix (AAV1.CamKII.Cre and AAV1.EF1a.DIO.eYFP) into the mPFC. Two weeks after viral injection, customized optical fibers were bilaterally implanted into the dorsal region of the mouse prelimbic cortex (A/P: +1.9 mm, M/L: ±0.3 mm, D/V: −1.7 mm). Four months later, we first performed the social behavior test under a laser-ON condition. Specifically,

following habituation, mice received a brief light stimulation from a 488-nm laser (10-ms long pulse at 20 Hz, laser ON for 20 seconds, ~ 7 mW at the fiber tip), right before the "sociability" and "social novelty" testing stages. One week later, we repeated the social behavior test with new pairs of strangers under the laser-OFF condition. Mice were connected with the optical fiber with the laser off and received a sham stimulation to mimic the auditory stimulation from the laser shutter (Vincent Associates). Specifically, following habituation, the shutter control was run for 20-seconds (10-ms long pulse, 20 Hz) right before the "sociability" and "social nove lty" testing stages.

Behavior annotation—The mouse social behavior test and NOR test were recorded using a top-view camera and a side-view camera (BFLY-PGE-12A2C-CS, Point Gray). The behavior images were synchronized with the calcium images by an external TTL signal from the miniScope controller. The behavior cameras had the same frame rate as the miniScope. Mouse behavioral activities were analyzed manually with frame-by-frame annotation based on an open-source MATLAB graphical user interface tool (https://github.com/pdollar/toolbox) (Lin et al., 2011). Five distinct behavioral activities were annotated, including direct-exploration, proximity-exploration, irrelevant-exploration, ambulation and self-directed activities. Criterial for these annotated behaviors are listed below: (1) Direct-exploration labelled the scenario when the subject mouse proactively contacted the wire container through nose-poking, biting and paw-grabbing. (2) Proximity-exploration labelled the scenario when the head of the subject mouse was inside of the adjacent areas (within 5 cm of distance) of the wire container, but without any physical contact. (3) Irrelevant-exploration labelled the scenario when the subject mouse physically contacted walls of the apparatus through nose-poking or paw-touching. (4) Ambulation was defined as when the head of the subject mouse was physically outside of the adjacent areas of the wire container and moving about with measurable position (distance) changes. (5) Self-directed activities grouped several behavioral categories, including immobility, self-grooming and other self-directed subtle-movements. These behaviors were referred to as "self-directed activities" because mice presumably received less external information while performing these activities. Particularly, "immobility" was defined as when the subject mouse kept the same posture for more than 10 frames (1 second) at any place of the apparatus; "Self-grooming" was defined as previously reported (Hong et al., 2014), when the subject mouse performed head-washing, nose- or face- grooming, head- or body- scratching, body-, leg-, paw- or genital- licking at any place of the apparatus; Other self-directed subtle-movement was defined as when the head of the subject mouse was physically outside of the adjacent areas of the wire container and performing subtle body posture changes without a measurable position (distance) change. Please also see Movie S2 for behavior annotation from a representative mouse.

For each behavior, a binary behavior vector was constructed to represent the behavior periods (behavior is occurring: 1, behavior is not occurring: 0). The direct-exploration vector was further decomposed to direct-exploration of subject1 (H: E1; SB: S1; SN: S1) and direct-exploration of subject2 (H: E2; SB: E2; SN: S2). Similarly, the proximity-exploration vector was further decomposed to proximity-exploration of subject1 (H: E1; SB: S1; SN: S1) and proximity-exploration of subject2 (H: E2; SB: E2; SN: S2). Abbreviations H, SB

and SN referred to habituation, sociability and social-novelty stage. E1 and E2 referred to the two empty containers, and S1 and S2 referred to stranger1 and stranger2, respectively.

Calcium image processing—Calcium images were processed and analyzed using custom scripts in MATLAB (Mathworks). Raw images were stabilized by determining the peak of the cross correlation with their neighboring images (Guizar-Sicairos et al., 2008) to compensate for the brain position translations caused by mouse movements. To align the calcium images collected from different behavior tests, e.g. test 1 and test 2, we first calculated the maximum-projection image of all frames collected for each test. The maximum-projection image of test 1 was used as the reference image. We then used the MATLAB function "imregdemons"(Image processing toolbox) to estimate the displacement field between any two maximum-projection images. We generated the registered image by applying the estimated displacement field to the maximum-projection image of test 2. Subsequently, we visually inspected the overlaid image containing both the reference and registered images with complementary colors to ensure that the estimated displacement field was appropriate (Figure S3D). We then applied the estimated displacement field to all the frames collected during the social behavior test 2.

Cell identification was performed using the algorithm previously described (Barbera et al., 2016). For each mouse, a cumulative cell map including all active neurons identified from all behavior tests was generated and manually verified. Calcium trace of each active neuron was extracted using detailed procedures below: (1) For a given neuron, a region of interest (ROI) was assigned with a circular region corresponding to the size of soma (diameter = 15 μm) and the centroid indicated the detected cell position. The minimum fluorescence of image-stack was used as the baseline fluorescence of the background (referred to as F_b). The fluorescence of ROI (F_{ROI}) was calculated as the averaged fluorescence after background subtraction from the raw image fluorescence (F_{raw}), i.e., $F_{ROI} = \langle F_{raw} - F_b \rangle_{ROI}$, brackets indicate the average over ROI pixels. (2) A commonly used procedure (Barbera et al., 2016; Chen et al., 2013; Kerlin et al., 2010; Pinto and Dan, 2015) was then applied to correct the potential contamination from out-of-focus neurons and neuropil fluorescence, i.e., $F_{sig} = F_{ROI} - F_{con}$, where F_{sig} was the true fluorescence signal, F_{con} was the contamination fluorescence. F_{con} was calculated over background subtracted images, using the averaged fluorescence from an annular zone (d1 = 20 μm, d2 = 30 μm) surrounding the soma ROI but avoiding other ROI regions. The slowly variable baseline due to background/neuropil activity was also removed in this procedure. (3) Each calcium trace ΔF/F was calculated as F_{sig}/F_0, where F_0 was the baseline fluorescence over the neuron ROI. (4) The baseline noise level for each calcium trace was estimated from the power density spectrum at a high-frequency band (Pnevmatikakis et al., 2016). For each testing stage, active neurons were defined only if the maximum value of the calcium trace was 3 times greater than its baseline noise level. The calcium traces were smoothed with a 500-ms sliding window. For each social behavior test, a neuron needed to be active at least once during the three testing stages to be identified as active neuron.

Neural activities in three testing stages—We calculated the neural calcium activity changes in the presence of restrained social targets (Figure 1F). Data of social behavior test

1 (n = 18 mice) were used in this calculation. For each neuron, calcium activities at the three social behavior testing stages were compared first by the Kruskal-Wallis test. A neuron was defined as "unchanged" if the p-value of the Kruskal-Wallis test was greater than 0.05. If the p value of the Kruskal-Wallis test was less than 0.05, multiple comparisons with Tukey-Kramer correction were performed subsequently. A neuron was defined as "SB&SN increased" if neural activities at both the "sociability" and "social novelty" stages were significantly higher than that of the "habituation" stage ($p < 0.05$, two-side). A neuron was defined as "SB increased only", if the neural activity at the "sociability" stage was significantly higher than that of the "habituation" stage ($p < 0.05$, two-side), while the neural activity at the "social novelty" stage was not significantly different from that of the "habituation stage" ($p > 0.05$, two-side). A neuron was defined as "SN increased only", if the neural activity at the "social novelty" stage was significantly higher than that of the "habituation" stage ($p < 0.05$, two-side), while the neural activity at the "sociability" stage was not significantly different from that of the "habituation" stage ($p > 0.05$, two-side). Neurons defined as "SB&SN decreased", "SB decreased only" or "SN decreased only" followed the same principle.

PCP effects on population calcium activities—To evaluate the PCP effect on calcium activities at the neural population levels, *in vivo* calcium imaging during a 10-minute habituation test was performed on a group of 5 mice injected with GCaMP6f virus and implanted with a GRIN lens, with the same apparatus set up as the habituation stage of the social behavior test as described above. These mice were subsequently intraperitoneally injected with 10 mg/kg PCP, and 30–50 minutes later, *in vivo* calcium imaging was performed again during a second 10-minute habituation test. As control, *in vivo* calcium imaging during the habituation tests with intraperitoneal saline injection was performed on a different group of three mice injected with GCaMP6f and implanted with a GRIN lens. Calcium images were processed using the same methods described above. Active neurons were identified and calcium traces were extracted. For each neuron, the averaged activity of the 10-minute recording before and after PCP or saline injection was calculated and denoted as C_{pre} and C_{post}, respectively. Averaged calcium activity of all the neurons from all mice were pooled together for comparisons between C_{pre} and C_{post} of PCP and saline injection, respectively. The Kolmogorov–Smirnov test was used to compare distributions of the averaged activity before and after PCP (Figure S6D) or saline injection (Figure S6E). We also calculated the activity change index by $(C_{post} - C_{pre})/ (C_{post} + C_{pre})$, for before and after PCP or saline injection, respectively (Figure S6D-E). The value of activity change index was bound between −1 and 1. Positive value indicated increased activity. Negative value indicated decreased activity.

Identification of behaviorally tuned neurons—For each testing stage, we identified behaviorally tuned neurons corresponding to each of the three annotated behaviors (i.e. direction-exploration, self-directed activities and irrelevant-exploration), following the same analytical steps as described below (also see Figure S2A-C): (1) For any given neuron n, we first calculated the similarity between calcium trace (ΔF/F) C_n and behavior vector B. Similarity was defined as normalized inner product of two vectors, $2B \cdot C_n/(|B|^2+|C_n|^2)$ (Carrillo-Reid et al., 2015; Hamm et al., 2017). The value of similarity was bound between 0

and 1. A value of 1 meant these two vectors were identical; while a value of 0 meant they were completely different. (2) We then randomly shuffled the behavior epochs and calculated the similarity between the shuffled behavior vector and the calcium trace C_n for a given neuron. We repeated this permutation process 5,000 times to generate a similarity distribution representing the similarity distribution histogram predicted by chance for a given neuron. (3) A neuron was defined as an "ON neuron" only if its actual similarity was greater than the 99.17% quantile of chance similarity distribution histogram generated by random shuffling. Consequently, ON neurons were those exhibiting significantly higher activities than that predicted by chance during the annotated behavior epochs. Conversely, a neuron was defined as an "OFF neuron" only if its actual similarity was lower than the 0.83% quantile of chance similarity distribution histogram generated by random shuffling. Therefore, OFF neurons were those exhibiting significantly lower activity than that predicted by chance during the annotated behavior epochs. Neurons with similarity falling between the 0.83% and 99.17% quantiles of chance similarity distribution histogram generated by random shuffling, were defined as "Other neurons" (also see Figure S2B-C). For each mouse, we applied the calculations described above to all active neurons to identify the ON and OFF neurons correlated with each of the three annotated behaviors during different testing stages. We defined each group of ON neurons (or OFF neurons) as one "neural ensemble". Consequently, we identified a total of 9 pairs of mPFC neural ensembles for each mouse, including Direct-ON_H and Direct-OFF_H, Self-ON_H and Self-OFF_H, Irrelevant-ON_H and Irrelevant-OFF_H during the H stage; Direct-ON_{SB} and Direct-OFF_{SB}, Self-ON_{SB} and Self-OFF_{SB}, Irrelevant-ON_{SB} and Irrelevant-OFF_{SB} during the SB stage; Direct-ON_{SN} and Direct-OFF_{SN}, Self-ON_{SN} and Self-OFF_{SN}, Irrelevant-ON_{SN} and Irrelevant-OFF_{SN} during the SN stage (see Figure 2 and Figure S2).

Neural ensemble correlation analysis—For each identified ON and OFF neuron, z-scored calcium activity was calculated by $(C_{\mathbf{n}} - \mu)/\sigma$, where μ and σ were the mean and standard deviation of C_n across H, SB and SN stages. For each ON and OFF neural ensemble, the averaged z-scored calcium activity was calculated by pooling the z-scored calcium activities from all identified ON and OFF neurons within the ensemble (Figure 2A and Figure S2F). We also calculated the averaged z-scored calcium activity across all behavior epochs (with an interval of −4 s to +4 s) for each ON and OFF neuron, and sorted all identified ON neurons based on the time of their maximal peak activities, and OFF neurons based on the time of their minimal activities (Figure 2D). To compare calcium dynamics around the behavior onset and the behavior ending, for each annotated behavior, we calculated the averaged z-scored calcium activity by pooling all ON or OFF neural ensembles tuned to the same behavior but identified at different testing stages (Figure S2D). We then compared the averaged calcium activity from all 18 mice, pre- (−2 s to 0 s) and post- (0 s to +2 s) behavior onset and ending for each annotated behavior (Figure S2E). To study the correlation between calcium activity of neural ensembles and specific behaviors, we calculated both (1) the Pearson correlation coefficient between the behavior vector and the averaged calcium activity of all neurons from a given neural ensemble (referred to as "group-correlation coefficient" in Figure 2E and Figure S3A), and (2) the average of the individual Pearson correlation coefficients between the behavior vector and each individual

neural activity (referred to as “individual-correlation coefficient” in Figure 2E and Figure S3A).

Neural ensemble consistency analysis—We evaluated the reliability of ON and OFF neuron’s responses across all behavioral epochs (Figure 2C). For an ON neuron, we defined an “engaged” behavior epoch when the ON neuron exhibited activity 3 times higher than the baseline noise; we defined a “not engaged” behavior epoch when the ON neuron did not exhibit activity 3 times higher than the baseline noise. By contrast, for an OFF neuron, we defined an “engaged” behavior epoch when the OFF neuron did not exhibit activity 3 times higher than the baseline noise; we defined a “not engaged” behavior epoch when the OFF neuron exhibited activity 3 times higher than the baseline noise. We evaluated the consistency of ON and OFF neuron’s responses through calculating the percentage of “engaged” behavior epochs across all behavior epochs within each testing stage. We also calculated for any given behavior epochs, the percentage of ON or OFF neurons being “engaged” within each ON and OFF ensemble.

Neural ensemble functional connectivity—Functional connectivity within and between neural ensembles was evaluated by calculating the averaged pairwise Pearson correlation coefficient of the calcium traces within and between behaviorally tuned neural ensembles (Figure 3B). The connectivity of an ON ensemble tuned to a specific behavior was evaluated by the averaged pairwise Pearson correlation coefficient of the calcium traces within the ON ensemble. The connectivity of an OFF ensemble tuned to a specific behavior was evaluated by the averaged pairwise Pearson correlation coefficient of the calcium traces within the OFF ensemble. The connectivity between ON and OFF ensembles tuned to the same specific behavior was evaluated by the averaged pairwise Pearson correlation coefficient of the calcium traces between ON and OFF ensembles. “Other” neurons were defined as neurons not belonging to ON or OFF ensembles tuned to any of the three annotated behaviors. The connectivity of “other” neurons was evaluated by the averaged pairwise Pearson correlation coefficient of the calcium traces within these neurons. We also evaluated the connectivity between ON (or OFF) neurons and “other” neurons by the averaged pairwise Pearson correlation coefficient of the calcium traces between ON neurons (or OFF neurons) and “other” neurons.

Direct-exploration tuned ensemble activity—Calcium activity of Direct-ON and Direct-OFF ensembles during direct-exploration of the two targets at each testing stage was calculated (Figure 8). The calcium trace of each neuron from a given neural ensemble was first normalized, using $C_{norm} = (C - C_{min})/(C_{max}-C_{min})$, where C_{max} and C_{min} were the maximum and minimum values of C, respectively. Subsequently, for each Direct-ON (or Direct-OFF) neuron, the average of the normalized calcium activity during direct-exploration on each of the two targets were calculated, respectively. Lastly, the averaged activity of all Direct-ON (or Direct-OFF) neurons for each mouse at each testing stage was calculated for quantification.

Neural ensemble overlap analysis—For a given pair of ensembles, A and B, the degree of observed overlap was calculated by the Sørensen–Dice index: $2N_{A\cap B}/(N_A+N_B)$,

with N_A indicating the neuron number of ensemble A and N_B indicating the neuron number of ensemble B, and $N_{A\cap B}$ indicating the neuron number of overlaps between ensemble A and B. The overlap degree was 0 if both A and B were empty. The observed overlap was then compared to what would be predicted by chance. The probability of one neuron to be randomly assigned into ensemble A was N_A/N, where N is the total imaged neuron number. Similarly, the probability of one neuron to be randomly assigned into ensemble B was N_B/N. If both assignments were random and independent, the probability of one neuron to be simultaneously assigned into both ensemble A and B was N_AN_B/N^2. Thus, the degree of overlap predicted by chance was calculated by $2N_AN_B/(N(N_A+N_B))$.

(A) Overlap within the same testing stage (Figure 4C-D): To evaluate the dynamic feature of neural ensembles, we split each 10-minute recording (one testing stage) into two 5-minute halves and re-identified ON and OFF neurons tuned to direct-exploration (Direct-ON and Direct-OFF neurons). We calculated the overlap between Direct-ON neurons of the two halves, and the overlap between Direct-OFF neurons of the two halves (Figure 4C). Next, to evaluate the reliability of our neuron identification process, we generated a union ensemble (Half1 U Half2) by pooling together all the identified Direct-ON (or Direct-OFF) neurons from the two 5-minute halves and calculated the percentage of Direct-ON (or Direct-OFF) neurons originally identified based on the full-length 10-minute recording belonging to the union ensembles.

(B) Overlap across different stages within one social behavior test (Figure 4A-B, Figure S3C): To clarify the across-stage overlaps between ON and OFF ensembles identified at different stages of a social behavior test, we defined across-stage overlap between ON-ON and OFF-OFF neural ensembles tuned to the same behavior as “ON-ON and OFF-OFF intra-behavior” overlap (e.g., neurons identified as Direct-ON at the “habituation” stage and remained as Direct-ON at the “sociability” or “social novelty” stage); across-stage overlap between ON-ON and OFF-OFF neural ensembles tuned to different behaviors as “ON-ON and OFF-OFF inter-behavior” overlap (e.g., neurons identified as Direct-ON at the “habituation” stage but Self-ON at the “sociability” stage); and across-stage overlap between ON-OFF neural ensembles tuned to the same behavior as “ON-OFF intra-behavior overlap” (e.g., neurons identified as Direct-ON at the “habituation” stage but Direct-OFF at the “soci ability” stage). The ON-ON intra- and inter-behavior overlaps between neural ensembles were illustrated in Figure 4A left panel. The OFF-OFF intra- and inter-behavior overlaps followed the same principle. The ON-OFF intra-behavior overlap was illustrated in Figure 4B left panel.

For the quantification of ON-ON and OFF-OFF intra- and inter-behavior overlaps, we calculated the average of all the possible pairwise overlaps for each mouse (9 pairs per mouse for ON-ON or OFF-OFF intra-behavior overlap, and 18 pairs per mouse for ON-ON or OFF-OFF inter-behavior overlap, as illustrated in Figure 4A left panel). We also quantified ON-ON and OFF-OFF intra-behavior overlaps for each annotated behavior, by calculating the average of 3 possible pairwise overlaps for each mouse (Figure S3C). For quantification of ON-OFF intra-behavior overlap for each annotated behavior, we calculated the average of all the possible pairwise overlaps for each mouse (6 pairs per mouse for each annotated behavior, as illustrated in Figure 4B left panel). The corresponding overlap to be

predicted by chance was calculated as described above. (C) Overlap between two different social behavior tests (Figure 5, Figure S4 and S5): We performed four different types of overlap comparisons for ON and OFF ensembles from all the social behavior tests (See *In vivo* calcium imaging section and Figure S4A): (1) Social behavior Test1 versus Test2, performed on different days with social targets presented at the same locations; (2) Social behavior Test2 versus Test3, performed on different days with social targets presented at the different locations; (3) Social behavior Test4 versus Test5, performed on the same days with social targets presented at different locations; (4) Social behavior test in the absence of PCP versus social behavior test in the presence of PCP, performed on different days with social targets presented at the same locations.

For the first three comparisons (i.e., Test1 versus Test2, Test2 versus Test3, Test4 versus Test5), **(a)** we first calculated for each annotated behavior, the ON-ON and OFF-OFF intra-behavior overlap across the same testing stages of the two tests (e.g., Direct-ON at the "sociability" stage of Test1 versus Direct-ON at the "sociability" stage of Test2, as illustrated in Figure S4B). For quantification, there was only one pairwise comparison per mouse for each same-stage overlap of each annotated behavior (Figure S4C-E). **(b)** We then compared for each annotated behavior, the ON-ON and OFF-OFF intra-behavior overlap across same testing stages versus across different testing stages of the two tests (e.g., Direct-ON at the "habituation" stage of Test1 versus Direct-ON at the "sociability" stage of Test2, as illustrated in Figure S4F). For quantification, we pooled all possible pairwise comparisons of same-stage overlaps (3 pairs per mouse) or different-stage overlaps (6 pairs per mice) for each annotated behavior. **(c)** Next, we grouped all three annotated behaviors to compare ON-ON and OFF-OFF intra-behavior overlap versus ON-ON and OFF-OFF inter-behavior overlap, both across same testing stages and across different testing stages of the two tests (as illustrated in Figure S5A). Therefore, there were four types of ON-ON (or OFF-OFF) ensemble overlaps in this comparison: the overlap of "intra-behavior, same-stage", the overlap of "intra-behavior, different-stage", the overlap of "inter-behavior, same-stage", and the overlap of "inter-behavior, different-stage". Specifically, the ON-ON overlap of "intra-behavior, same-stage" was defined as the overlap of the two ON ensembles tuned to the same behavior identified in the same stages of the two tests (9 pairs per mouse). The ON-ON overlap of "intra-behavior, different-stage" was defined as the overlap of the two ON ensembles tuned to the same behavior identified in different stages of the two tests (18 pairs per mouse). The ON-ON overlap of "inter-behavior, same-stage" was defined as the overlaps of the two ON ensembles tuned to different behaviors identified in the same stages of the two tests (18 pairs per mouse). The ON-ON overlap of "inter-behavior, different-stage" was defined as the overlaps of the two ON ensembles tuned to different behaviors identified in different stages of the two tests (36 pairs per mouse). The four types of OFF-OFF overlaps were defined following the same principle. For quantification of the four types of overlaps, we calculated the average of all possible pairwise overlaps for each mouse (Figure S5B-D). **(d)** Lastly, we generated a union ON ensemble for each annotated behavior in each test, by pooling together all ON neurons tuned to the same behavior identified from all three testing stages. We also generated union OFF ensembles for each annotated behavior in each test using a similar strategy. We then compared the across-test ON-ON and OFF-OFF intra-behavior overlap (e.g., neurons identified as Direct-ON during Test1 and remained as Direct-

ON during Test2), ON-ON and OFF-OFF inter-behavior overlap (e.g., neurons identified as Direct-ON during Test1 but Self-ON during Test2), and ON-OFF intra-behavior overlap among these union ensembles (e.g., neurons identified as Direct-ON during Test1 but Direct-OFF during Test2, See illustration in Figure 5A). The across-test ON-ON intra-behavior overlap was defined as the overlap between ON ensembles tuned to the same behavior (3 pairs per mouse) identified in different tests; while the across-test ON-ON inter-behavior overlap was defined as the overlap between ON ensembles tuned to different behaviors (6 pairs per mouse) identified in different tests. The across-test OFF-OFF intra- and inter-behavior overlap followed the same principle. The across-test ON-OFF intra-behavior overlap was defined as the overlap between ON and OFF ensembles tuned to the same behavior (1 pair per mouse for each annotated behavior) identified in different tests. For quantification, we calculated the average of all possible pairwise overlaps for each mouse (Figure 5A-C).

For the last comparison (i.e., social behavior tests in the absence and presence of PCP), we generated a union ON ensemble and a union OFF ensemble for each annotated behavior in each test as described above. We then compared the across-test ON-ON and OFF-OFF intra-behavior overlap, ON-ON and OFF-OFF inter-behavior overlap, and ON-OFF intra-behavior overlap among these union ensembles identified in the absence or presence of PCP (See illustration in Figure 7C). All definitions and calculations were the same as described above in **(d)** of the across-test overlap comparisons.

Behavioral variable decoding—We performed binary decoding analyses to test if the mouse behavior variable (whether a behavior was occurring or not) could be reliably predicted using calcium activities of the identified ON and OFF neurons. Neural decoding is formulated as constructing a predictive model (classifier) f: C → b, where C is a set of calcium imaging derived features (calcium traces), and b is a behavior variable. For each mouse and a given behavior variable, five types of decoding tasks were conducted using five different types of calcium trace dataset: (1) calcium traces of all neurons; (2) calcium traces of the behavior tuned ON and OFF neurons; (3) calcium traces of the behavior tuned ON neurons and the randomly shuffled calcium traces of the behavior tuned OFF neurons; calcium traces of the behavior tuned OFF neurons and the randomly shuffled calcium traces of the behavior tuned ON neurons; (5) randomly shuffled calcium traces of the behavior tuned ON and OFF neurons. We used the widely used non-parametric decision tree classifier (Classification And Regression Trees, CART) (Hastie et al., 2009). CART was implemented using the R package caret (https://CRAN.R-project.org/package=caret). Model parameters were tuned by the resampling-based method which only used the training dataset. Some behavioral variables were imbalanced. That is, the behavior only occurred on a small fraction of the total testing frames, thus one class outnumbers other classes by a substantial proportion. The down-sampling method (Chawla, 2005) was used to correct the imbalance data problem. We used 10-fold cross-validation to assess classification performance. In the 10-fold cross-validation approach, the original dataset is randomly divided into 10 subsets. We then trained the model using 9 of the subsets (the training dataset) and tested the model using the 1 left-out subset (the test dataset). We repeated this process 10 times and calculate the average performance. Cross-validation provides a reliable assessment of model

performance because the model never uses any cases from the test dataset. Model quality metric is the balanced accuracy, which is the arithmetic mean of sensitivity (true positive rate) and specificity (true negative rate). The balanced accuracy is in the range [0, 1]. The balanced accuracy = 1 represents perfect classification. The balanced accuracy = 0.5 represents the chance level. The decoding analysis was repeated for each stage of the social behavior test (test1), with a separate classifier trained for each decoding task. For the quantification of the decoding performance of each behavior variable (Figure 3A), we used the balanced accuracy of all three testing stages from 18 mice (54 cases).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistics—All reported sample numbers represented biological replicates. All data were presented as mean ± SEM unless otherwise specified. In all the histograms in the figures, unless otherwise specified, dots indicated individual mice and bars represented group averages. In all the figures, **** represented $p < 0.0001$; *** represented $p < 0.001$; ** represented $p < 0.01$, * represented $p < 0.05$; ns represented not significant ($p > 0.05$). All statistical analyses were performed using Graphpad Prism (Graphpad) or MATLAB (Mathworks). Non-parametric tests were used in all of our statistical analysis, including Wilcoxon matched-pairs signed rank test, Friedman's test and Dunn's post-hoc test, Mann Whitney test, and Kolmogorov-Smirnov test. Pearson's correlation was used for calculating correlation coefficient, and random permutations were used for determining chance level. All tests were two-sided, and statistical significance was defined as $p < 0.05$. Statistical details of all experiments can be found in Table S1.

Sample size estimation—Our estimates of animal use for *in vivo* calcium imaging recording was based on past experiences and those presented in the literatures.

Data inclusion and exclusion—For *in vivo* calcium imaging experiments, we did not exclude any mice during our behavior data acquisition and analysis. Occasionally we could not reliably identify any behaviorally tuned neurons from a mouse in the absence or the presence of PCP, during a given testing stage for a given behavior, which would explain the reason that for neural ensemble related analyses, n numbers were less than 18, or not equal to the multiples of 18, in figure legends and in Table S1.

DATA AND SOFTWARE AVAILABILITY—All datasets and custom MATLAB scripts will be available upon request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

Research was supported by NIH/NIDA/IRP. We would like to thank the Genetically-Encoded Neuronal Indicator and Effector (GENIE) Project and the Janelia Research Campus of the Howard Hughes Medical Institute for generously allowing the use of GCaMP6 in our research. We would also like to thank Dr. Yavin Shaham, Dr. Deon Harvey and Casey Moffitt for critical reading of the manuscript.

References:

Adler CM, Goldberg TE, Malhotra AK, Pickar D, and Breier A (1998). Effects of ketamine on thought disorder, working memory, and semantic memory in healthy volunteers. Biol Psychiatry 43, 811–816. [PubMed: 9611670]

Adolphs R (2009). The social brain: neural basis of social knowledge. Annu Rev Psychol 60, 693–716. [PubMed: 18771388]

Anderson DJ (2016). Circuit modules linking internal states and social behaviour in flies and mice. Nat Rev Neurosci 17, 692–704. [PubMed: 27752072]

Avale ME, Chabout J, Pons S, Serreau P, De Chaumont F, Olivo-Marin JC, Bourgeois JP, Maskos U, Changeux JP, and Granon S (2011). Prefrontal nicotinic receptors control novel social interaction between mice. FASEB J 25, 2145–2155. [PubMed: 21402717]

Bang J, Spina S, and Miller BL (2015). Frontotemporal dementia. Lancet 386, 1672–1682. [PubMed: 26595641]

Barbera G, Liang B, Zhang L, Gerfen CR, Culurciello E, Chen R, Li Y, and Lin DT (2016). Spatially Compact Neural Clusters in the Dorsal Striatum Encode Locomotion Relevant Information. Neuron 92, 202–213. [PubMed: 27667003]

Bicks LK, Koike H, Akbarian S, and Morishita H (2015). Prefrontal Cortex and Social Cognition in Mouse and Man. Front Psychol 6, 1805. [PubMed: 26635701]

Breier A, Malhotra AK, Pinals DA, Weisenfeld NI, and Pickar D (1997). Association of ketamine-induced psychosis with focal activation of the prefrontal cortex in healthy volunteers. Am J Psychiatry 154, 805–811. [PubMed: 9167508]

Brumback AC, Ellwood IT, Kjaerby C, Iafrati J, Robinson S, Lee AT, Patel T, Nagaraj S, Davatolhagh F, and Sohal VS (2017). Identifying specific prefrontal neurons that contribute to autism-associated abnormalities in physiology and social behavior. Mol Psychiatry

Carrillo-Reid L, Miller JE, Hamm JP, Jackson J, and Yuste R (2015). Endogenous sequential cortical activity evoked by visual stimuli. J Neurosci 35, 8813–8828. [PubMed: 26063915]

Chawla NV (2005). Data Mining for Imbalanced Datasets: An Overview. In Data Mining and Knowledge Discovery Handbook, Maimon O, and Rokach L, eds. (Boston, MA: Springer US), pp. 853–867.

Chen TW, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA, Orger MB, Jayaraman V, et al. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature 499, 295–300. [PubMed: 23868258]

Chevallier C, Kohls G, Troiani V, Brodkin ES, and Schultz RT (2012). The social motivation theory of autism. Trends Cogn Sci 16, 231–239. [PubMed: 22425667]

Chung W, Choi SY, Lee E, Park H, Kang J, Park H, Choi Y, Lee D, Park SG, Kim R, et al. (2015). Social deficits in IRSp53 mutant mice improved by NMDAR and mGluR5 suppression. Nat Neurosci 18, 435–443. [PubMed: 25622145]

Dejean C, Courtin J, Karalis N, Chaudun F, Wurtz H, Bienvenu TC, and Herry C (2016). Prefrontal neuronal assemblies temporally control fear behaviour. Nature 535, 420–424. [PubMed: 27409809]

Dolen G, Darvishzadeh A, Huang KW, and Malenka RC (2013). Social reward requires coordinated activity of nucleus accumbens oxytocin and serotonin. Nature 501, 179–184. [PubMed: 24025838]

Driscoll LN, Pettit NL, Minderer M, Chettih SN, and Harvey CD (2017). Dynamic Reorganization of Neuronal Activity Patterns in Parietal Cortex. Cell 170, 986–999 e916. [PubMed: 28823559]

Duffney LJ, Zhong P, Wei J, Matas E, Cheng J, Qin L, Ma K, Dietz DM, Kajiwara Y, Buxbaum JD, et al. (2015). Autism-like Deficits in Shank3-Deficient Mice Are Rescued by Targeting Actin Regulators. Cell Rep 11, 1400–1413. [PubMed: 26027926]

Edelman GM, and Gally JA (2001). Degeneracy and complexity in biological systems. Proc Natl Acad Sci U S A 98, 13763–13768. [PubMed: 11698650]

Franklin TB, Silva BA, Perova Z, Marrone L, Masferrer ME, Zhan Y, Kaplan A, Greetham L, Verrechia V, Halman A, et al. (2017). Prefrontal cortical control of a brainstem social behavior circuit. Nat Neurosci 20, 260–270. [PubMed: 28067904]

Fujisawa S, Amarasingham A, Harrison MT, and Buzsaki G (2008). Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. Nat Neurosci 11, 823–833. [PubMed: 18516033]

Garcia-Junco-Clemente P, Ikrar T, Tring E, Xu X, Ringach DL, and Trachtenberg JT (2017). An inhibitory pull-push circuit in frontal cortex. Nat Neurosci 20, 389–392. [PubMed: 28114295]

Green MF, Horan WP, and Lee J (2015). Social cognition in schizophrenia. Nat Rev Neurosci 16, 620–631. [PubMed: 26373471]

Guizar-Sicairos M, Thurman ST, and Fienup JR (2008). Efficient subpixel image registration algorithms. Opt Lett 33, 156–158. [PubMed: 18197224]

Gunaydin LA, Grosenick L, Finkelstein JC, Kauvar IV, Fenno LE, Adhikari A, Lammel S, Mirzabekov JJ, Airan RD, Zalocusky KA, et al. (2014). Natural neural projection dynamics underlying social behavior. Cell 157, 1535–1551. [PubMed: 24949967]

Hamm JP, Peterka DS, Gogos JA, and Yuste R (2017). Altered Cortical Ensembles in Mouse Models of Schizophrenia. Neuron 94, 153–167 e158. [PubMed: 28384469]

Hastie T, Tibshirani R, and Friedman J (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second edition edn (Springer series in statistics).

Hattori R, Kuchibhotla KV, Froemke RC, and Komiyama T (2017). Functions and dysfunctions of neocortical inhibitory neuron subtypes. Nat Neurosci 20, 1199–1208. [PubMed: 28849791]

Hebb DO (1949). The organization of behavior (New York: Wiley).

Homayoun H, and Moghaddam B (2007). NMDA receptor hypofunction produces opposite effects on prefrontal cortex interneurons and pyramidal neurons. J Neurosci 27, 11496–11500. [PubMed: 17959792]

Hong W, Kim DW, and Anderson DJ (2014). Antagonistic control of social versus repetitive self-grooming behaviors by separable amygdala neuronal subsets. Cell 158, 1348–1361. [PubMed: 25215491]

Jackson ME, Homayoun H, and Moghaddam B (2004). NMDA receptor hypofunction produces concomitant firing rate potentiation and burst activity reduction in the prefrontal cortex. Proc Natl Acad Sci U S A 101, 8467–8472. [PubMed: 15159546]

Javitt DC, and Zukin SR (1991). Recent advances in the phencyclidine model of schizophrenia. Am J Psychiatry 148, 1301–1308. [PubMed: 1654746]

Jentsch JD, and Roth RH (1999). The neuropsychopharmacology of phencyclidine: from NMDA receptor hypofunction to the dopamine hypothesis of schizophrenia. Neuropsychopharmacology 20, 201–225. [PubMed: 10063482]

Jodo E, Suzuki Y, Katayama T, Hoshino KY, Takeuchi S, Niwa S, and Kayama Y (2005). Activation of medial prefrontal cortex by phencyclidine is mediated via a hippocampo-prefrontal pathway. Cereb Cortex 15, 663–669. [PubMed: 15342431]

Katayama T, Jodo E, Suzuki Y, Hoshino KY, Takeuchi S, and Kayama Y (2007). Activation of medial prefrontal cortex neurons by phencyclidine is mediated via AMPA/kainate glutamate receptors in anesthetized rats. Neuroscience 150, 442–448. [PubMed: 17935894]

Kerlin AM, Andermann ML, Berezovskii VK, and Reid RC (2010). Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. Neuron 67, 858–871. [PubMed: 20826316]

Kim Y, Venkataraju KU, Pradhan K, Mende C, Taranda J, Turaga SC, Arganda-Carreras I, Ng L, Hawrylycz MJ, Rockland KS, et al. (2015). Mapping social behavior-induced brain activation at cellular resolution in the mouse. Cell Rep 10, 292–305. [PubMed: 25558063]

Krupa DJ, Wiest MC, Shuler MG, Laubach M, and Nicolelis MA (2004). Layer-specific somatosensory cortical activation during active tactile discrimination. Science 304, 1989–1992. [PubMed: 15218154]

Kuchibhotla KV, Gill JV, Lindsay GW, Papadoyannis ES, Field RE, Sten TA, Miller KD, and Froemke RC (2017). Parallel processing by cortical inhibition enables context-dependent behavior. Nat Neurosci 20, 62–71. [PubMed: 27798631]

Kvitsiani D, Ranade S, Hangya B, Taniguchi H, Huang JZ, and Kepecs A (2013). Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. Nature 498, 363–366. [PubMed: 23708967]

Lahti AC, Holcomb HH, Medoff DR, and Tamminga CA (1995). Ketamine activates psychosis and alters limbic blood flow in schizophrenia. Neuroreport 6, 869–872. [PubMed: 7612873]

Lee E, Rhim I, Lee JW, Ghim JW, Lee S, Kim E, and Jung MW (2016). Enhanced Neuronal Activity in the Medial Prefrontal Cortex during Social Approach Behavior. J Neurosci 36, 6926–6936. [PubMed: 27358451]

Lee J, Chung C, Ha S, Lee D, Kim DY, Kim H, and Kim E (2015). Shank3-mutant mice lacking exon 9 show altered excitation/inhibition balance, enhanced rearing, and spatial memory deficit. Front Cell Neurosci 9, 94. [PubMed: 25852484]

Leger M, Quiedeville A, Bouet V, Haelewyn B, Boulouard M, Schumann-Bard P, and Freret T (2013). Object recognition test in mice. Nat Protoc 8, 2531–2537. [PubMed: 24263092]

Lin D, Boyle MP, Dollar P, Lee H, Lein ES, Perona P, and Anderson DJ (2011). Functional identification of an aggression locus in the mouse hypothalamus. Nature 470, 221–226. [PubMed: 21307935]

Luby ED, Cohen BD, Rosenbaum G, Gottlieb JS, and Kelley R (1959). Study of a new schizophrenomimetic drug; sernyl. AMA Arch Neurol Psychiatry 81, 363–369. [PubMed: 13626287]

Matthews GA, Nieh EH, Vander Weele CM, Halbert SA, Pradhan RV, Yosafat AS, Glober GF, Izadmehr EM, Thomas RE, Lacy GD, et al. (2016). Dorsal Raphe Dopamine Neurons Represent the Experience of Social Isolation. Cell 164, 617–631. [PubMed: 26871628]

McHenry JA, Otis JM, Rossi MA, Robinson JE, Kosyk O, Miller NW, McElligott ZA, Budygin EA, Rubinow DR, and Stuber GD (2017). Hormonal gain control of a medial preoptic area social reward circuit. Nat Neurosci 20, 449–458. [PubMed: 28135243]

Molas S, Zhao-Shea R, Liu L, DeGroot SR, Gardner PD, and Tapper AR (2017). A circuit-based mechanism underlying familiarity signaling and the preference for novelty. Nat Neurosci

Moy SS, Nadler JJ, Perez A, Barbaro RP, Johns JM, Magnuson TR, Piven J, and Crawley JN (2004). Sociability and preference for social novelty in five inbred strains: an approach to assess autistic-like behavior in mice. Genes Brain Behav 3, 287–302. [PubMed: 15344922]

Moy SS, Nadler JJ, Young NB, Perez A, Holloway LP, Barbaro RP, Barbaro JR, Wilson LM, Threadgill DW, Lauder JM, et al. (2007). Mouse behavioral tasks relevant to autism: phenotypes of 10 inbred strains. Behav Brain Res 176, 4–20. [PubMed: 16971002]

Murray AJ, Woloszynowska-Fraser MU, Ansel-Bollepalli L, Cole KL, Foggetti A, Crouch B, Riedel G, and Wulff P (2015). Parvalbumin-positive interneurons of the prefrontal cortex support working memory and cognitive flexibility. Sci Rep 5, 16778. [PubMed: 26608841]

Murugan M, Jang HJ, Park M, Miller EM, Cox J, Taliaferro JP, Parker NF, Bhave V, Hur H, Liang Y, et al. (2017). Combined Social and Spatial Coding in a Descending Projection from the Prefrontal Cortex. Cell 171, 1663–1677 e1616. [PubMed: 29224779]

Nakajima M, Gorlich A, and Heintz N (2014). Oxytocin modulates female sociosexual behavior through a specific class of prefrontal cortical interneurons. Cell 159, 295–305. [PubMed: 25303526]

Neill JC, Barnes S, Cook S, Grayson B, Idris NF, McLean SL, Snigdha S, Rajagopal L, and Harte MK (2010). Animal models of cognitive dysfunction and negative symptoms of schizophrenia: focus on NMDA receptor antagonism. Pharmacol Ther 128, 419–432. [PubMed: 20705091]

Okuyama T, Kitamura T, Roy DS, Itohara S, and Tonegawa S (2016). Ventral CA1 neurons store social memory. Science 353, 1536–1541. [PubMed: 27708103]

Peters AJ, Lee J, Hedrick NG, O'Neil K, and Komiyama T (2017). Reorganization of corticospinal output during motor learning. Nat Neurosci 20, 1133–1141. [PubMed: 28671694]

Pinto L, and Dan Y (2015). Cell-Type-Specific Activity in Prefrontal Cortex during Goal-Directed Behavior. Neuron 87, 437–450. [PubMed: 26143660]

Pnevmatikakis EA, Soudry D, Gao Y, Machado TA, Merel J, Pfau D, Reardon T, Mu Y, Lacefield C, Yang W, et al. (2016). Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. Neuron 89, 285–299. [PubMed: 26774160]

Rigotti M, Barak O, Warden MR, Wang XJ, Daw ND, Miller EK, and Fusi S (2013). The importance of mixed selectivity in complex cognitive tasks. Nature 497, 585–590. [PubMed: 23685452]

Sceniak MP, Lang M, Enomoto AC, James Howell C, Hermes DJ, and Katz DM (2016). Mechanisms of Functional Hypoconnectivity in the Medial Prefrontal Cortex of Mecp2 Null Mice. Cereb Cortex 26, 1938–1956. [PubMed: 25662825]

Shafi M, Zhou Y, Quintana J, Chow C, Fuster J, and Bodner M (2007). Variability in neuronal activity in primate cortex during working memory tasks. Neuroscience 146, 1082–1108. [PubMed: 17418956]

Suzuki Y, Jodo E, Takeuchi S, Niwa S, and Kayama Y (2002). Acute administration of phencyclidine induces tonic activation of medial prefrontal cortex neurons in freely moving rats. Neuroscience 114, 769–779. [PubMed: 12220577]

Takahata R, and Moghaddam B (2003). Activation of glutamate neurotransmission in the prefrontal cortex sustains the motoric and dopaminergic effects of phencyclidine. Neuropsychopharmacology 28, 1117–1124. [PubMed: 12700703]

Warren JD, Rohrer JD, and Rossor MN (2013). Clinical review. Frontotemporal dementia. BMJ 347, f4827. [PubMed: 23920254]

West MJ (1999). Age-Related Neuronal Loss in the Cerebral Cortex. In Cerebral Cortex: Neurodegenerative and Age-Related Changes in Structure and Function of Cerebral Cortex, Peters A, and Morrison JH, eds. (Boston, MA: Springer US), pp. 81–88.

Yizhar O, Fenno LE, Prigge M, Schneider F, Davidson TJ, O'Shea DJ, Sohal VS, Goshen I, Finkelstein J, Paz JT, et al. (2011). Neocortical excitation/inhibition balance in information processing and social dysfunction. Nature 477, 171–178. [PubMed: 21796121]

Zhang S, Xu M, Kamigaki T, Hoang Do JP, Chang WC, Jenvay S, Miyamichi K, Luo L, and Dan Y (2014). Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing. Science 345, 660–665. [PubMed: 25104383]

Zhang W, Daly KM, Liang B, Zhang L, Li X, Li Y, and Lin DT (2017). BDNF rescues prefrontal dysfunction elicited by pyramidal neuron-specific DTNBP1 deletion in vivo. J Mol Cell Biol 9, 117–131. [PubMed: 27330059]

Zhang W, Zhang L, Liang B, Schroeder D, Zhang ZW, Cox GA, Li Y, and Lin DT (2016). Hyperactive somatostatin interneurons contribute to excitotoxicity in neurodegenerative disorders. Nat Neurosci 19, 557–559. [PubMed: 26900927]

Highlights

- mPFC excitatory neurons form distinct ensembles tuned to different behaviors
- Dynamic ON and OFF neural ensembles code real-time behavioral information
- Direct-exploration ensembles carry information of social salience and novelty
- PCP disorganizes mPFC ensembles and induces deficits in social exploration

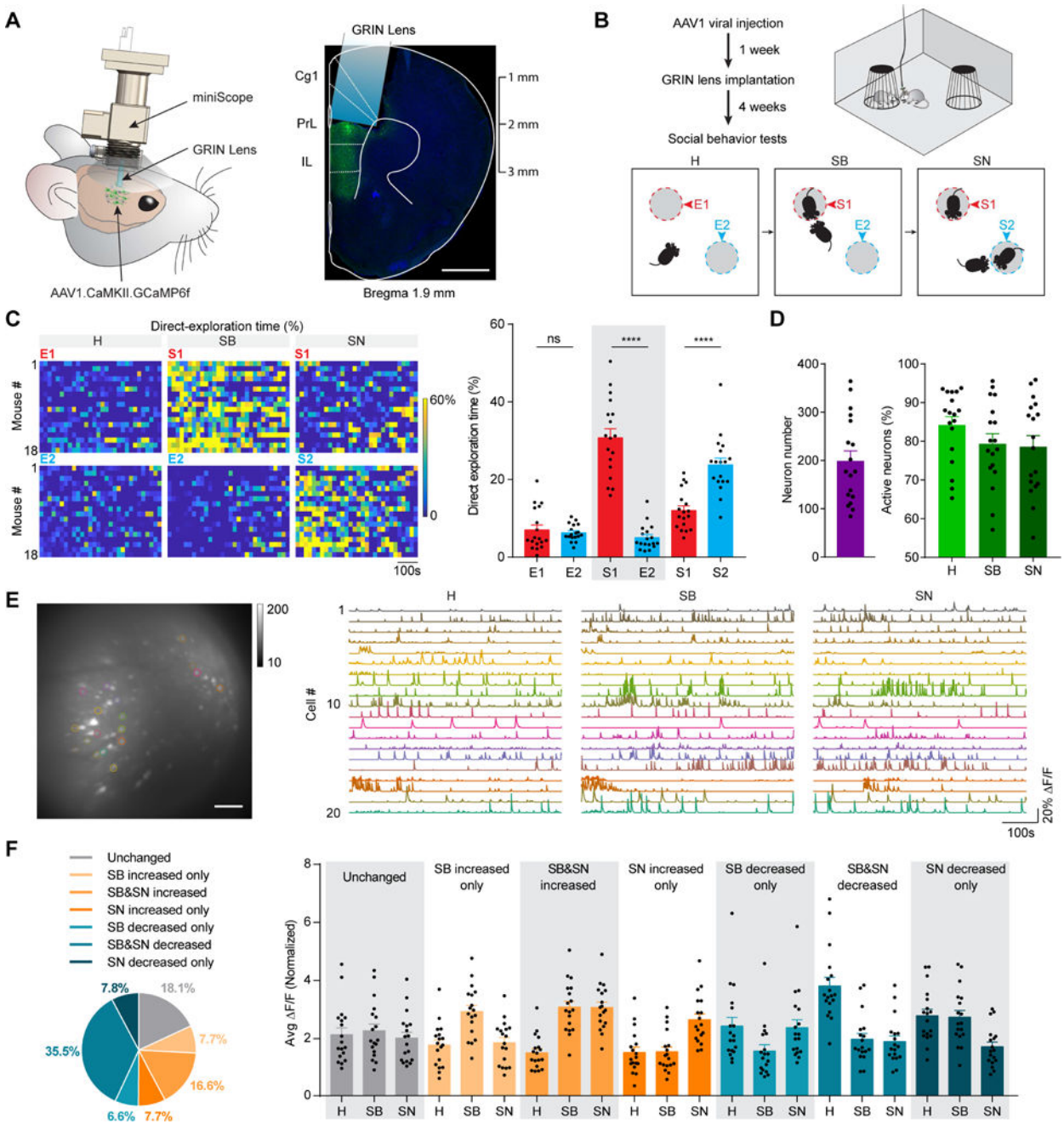


Figure 1: *In vivo* calcium imaging of mouse mPFC during social behavior test.

(**A**) Left panel: schematic diagram of experimental setup. Right panel: mouse coronal brain slice showing the anatomical position for GCamp6f expression and GRIN lens implantation. Scale bar: 1 mm. (**B**) Top left panel: experimental timeline. Top right panel: social behavioral test apparatus. Bottom panels: schematic paradigm for social behavior test: habituation (H), sociability (SB) and social novelty (SN). E1 and E2: empty containers; S1 and S2: stranger1 and stranger 2, respectively. (**C**) Left panels: raster plots for percentage of time (per 20 second bin) spent on direct-exploration of individual mouse toward different targets during each testing stage. Range indicators: 0–60%. Right histogram: statistical

results, n = 18 mice. (**D**) Left panel: total numbers of neurons detected from individual mice. Right panel: percentage of active neurons at each testing stage. (**E**) Left panel: representative maximum projection fluorescent image of mPFC excitatory neurons labeled with GCamp6f, scale bar: 100 μm. Right panels: representative calcium traces from twenty regions of interest (color matched) during each testing stage. (**F**) Left panel: pie chart shows the percentage of neurons displaying increased, decreased or unchanged activity at the SB or the SN stages, compared to the H stage. Right histogram: the averaged calcium activities of neural groups exhibiting different activity changes at each stage. Data were represented as mean ± SEM. See also Figure S1, Movie S1, Movie S2, and Table S1.

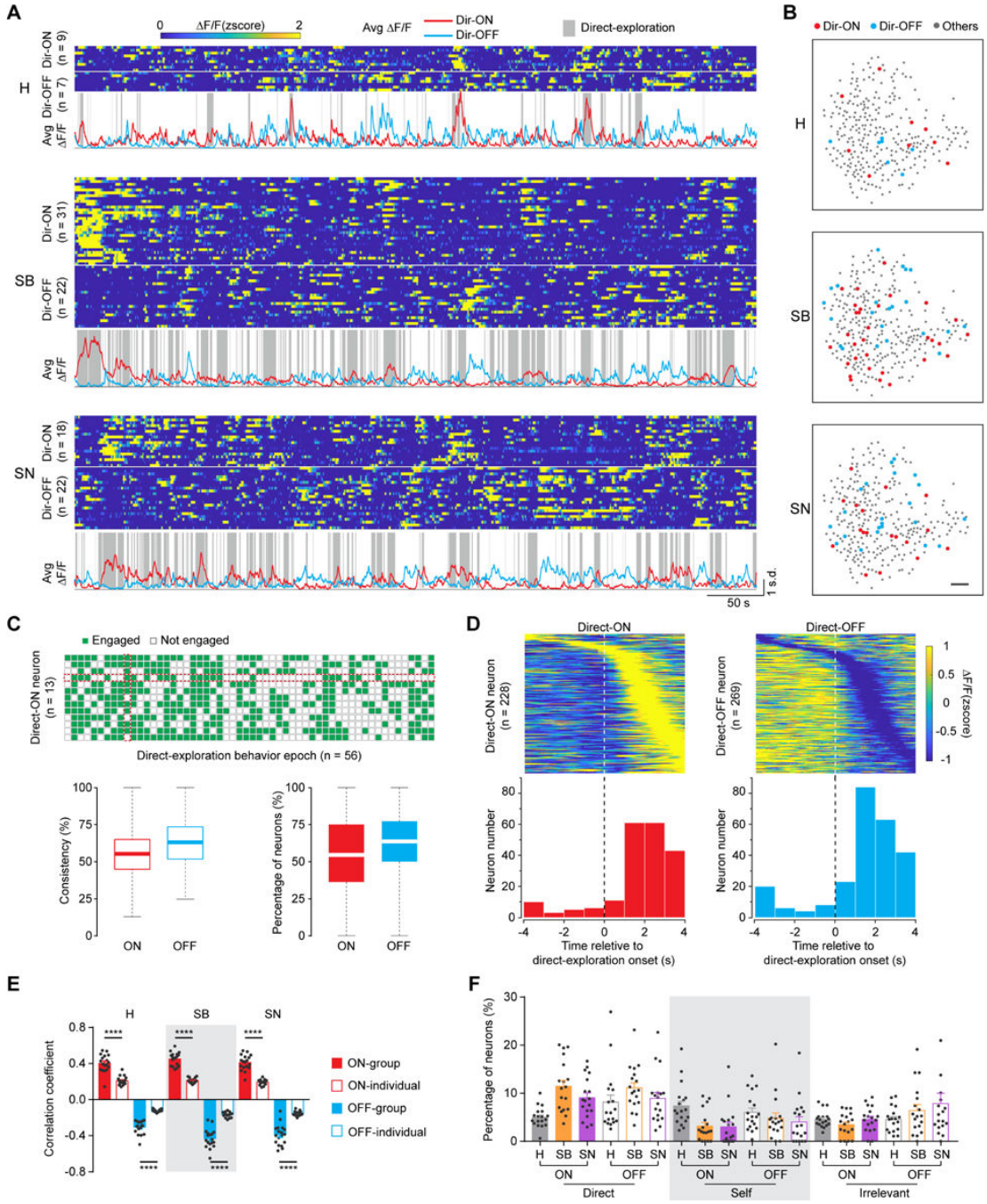


Figure 2: Identification of behaviorally tuned ON and OFF neural ensembles in the mPFC.

(**A**) Raster plots of individual neural calcium activities and calcium traces of the averaged group activity of Direct-ON and Direct-OFF neurons (Dir-ON and Dir-OFF) from a representative mouse, H: habituation; SB: sociability; SN: social novelty. Red traces: Dir-ON; Blue traces: Dir-OFF. Vertical gray shadows: individual direct-exploration behavior epochs. (**B**) Spatial distributions of Dir-ON and Dir-OFF neurons from the same representative mouse at different stages, scale bar: 100 μm. (**C**) Top panel: engagement of Direct-ON neurons from a representative mouse at one testing stage; Green: “engaged”

epochs; white: “not engaged “ epochs. Bottom left panel: percentage of engaged epochs for ON and OFF neurons across all behavior epochs indicated by the horizontal red box shown in top panel. Bottom right panel: percentage of engaged ON and OFF neurons at each individual behavior epoch indicated by the vertical red box shown in top panel. Box plots indicate Median and the 25–75th percentile with whiskers for maximal and minimal values. n = 9 mice. (**D**) Top panels: raster plot of the averaged calcium activity of individual Direct-ON and Direct-OFF neurons at the onset of direct-exploration (4 second before and 4 second after), sorted by the time of maximal activities (ON neurons) or minimal activities (OFF neurons). Bottom panels: distribution of Direct-ON neurons displaying maximal activity and Direct-OFF neurons displaying minimal activity at different time around the behavior onset. n = 9 mice. (**E**) Calcium-behavior correlation coefficients for Direct-ON and Direct-OFF neurons at different stages, n = 18 mice. (**F**) Sizes of behaviorally tuned ON and OFF neural ensembles in percentage of all imaged neurons at each testing stage, n = 18 mice. Data were represented as mean ± SEM. See also Figure S2, Figure S3 and Table S1.

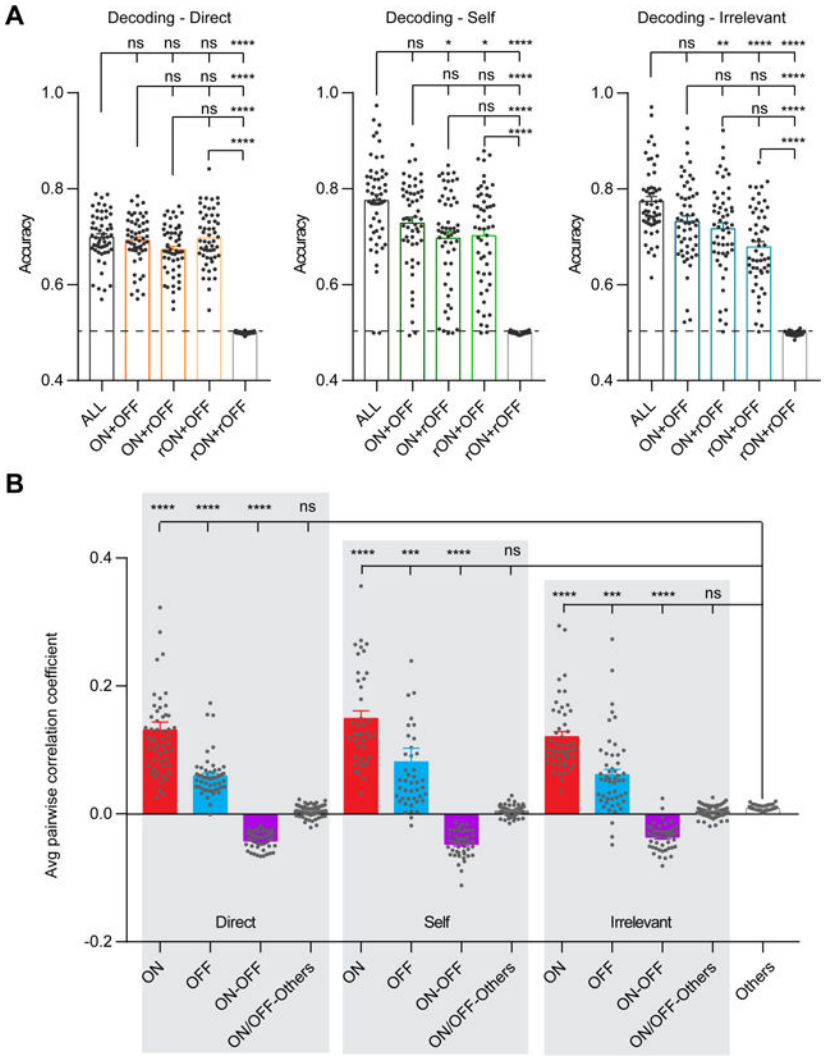


Figure 3: ON and OFF neural ensembles carry real-time behavioral information.

(**A**) Decoding analyses for each annotated behavior using calcium activity from different groups of neurons, n = 54 decoding cases from 18 mice. ALL: all imaged neurons; ON + OFF: ON and OFF neurons together; ON + rOFF: ON and shuffled OFF neurons; rON + OFF: OFF and shuffled ON neurons; rON + rOFF: shuffled ON and OFF neurons. Each dot represents one decoding result from one mouse for one behavior at one testing stage. Bars represent group averages. (**B**) Pairwise correlation coefficient of neural calcium activities within and between ensembles, n = 40 ~ 108 correlation analyses from 18 mice. ON: within ON ensembles; OFF: within OFF ensembles; ON-OFF: between ON and OFF ensembles; ON/OFF-Others: between ON or OFF neurons with other neurons not tuned to the three annotated behaviors; Others: among other neurons. Each dot represents one correlation analysis from one mouse for one behavior at one testing stage. Bars represent group averages. Direct: direct-exploration; Self: self-directed activities; Irrelevant: irrelevant-exploration. Data were represented as mean ± SEM. See also Table S1.

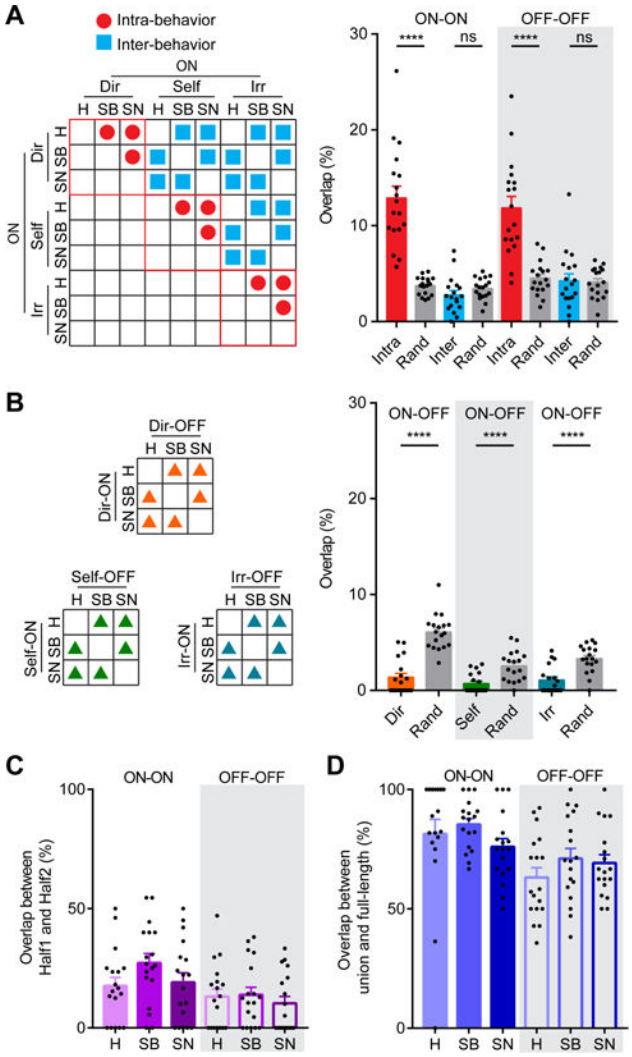


Figure 4: Across-stage overlap analyses between ensembles identified during different stages of one social behavior test.

(**A**) Left panel: a matrix illustrates definitions for ON-ON intra-behavior (red circles) and ON-ON inter-behavior (blue squares) overlaps. OFF-OFF intra-behavior and OFF-OFF inter-behavior overlaps follow the same principle. H, habituation; SB, sociability; SN, social novelty. Dir, direct-exploration; Self, self-directed activities; Irr, irrelevant exploration. Right panel: Statistical results showing across-stage ON-ON and OFF-OFF intra- and inter-behavior overlaps and those predicted by chance based on random selection (Rand), n = 18 mice. (**B**) Left panel: illustrations showing the definition for ON-OFF intra-behavior overlap (triangles) for each annotated behavior. Right panel: Statistical results showing across-stage ON-OFF intra-behavior overlap and those predicted by chance based on random selection (Rand), n = 18 mice. (**C**) Comparisons of Direct-ON and Direct-OFF neural ensembles identified from the two 5-minute halves of the same stage (n = 18 mice). (**D**) Comparisons of Direct-ON and Direct-OFF neural ensemble identified from a full-length 10-minute period versus the union ensemble of the two 5-minute halves, n = 18 mice. Data were represented as mean ± SEM. See also Figure S3 and Table S1.

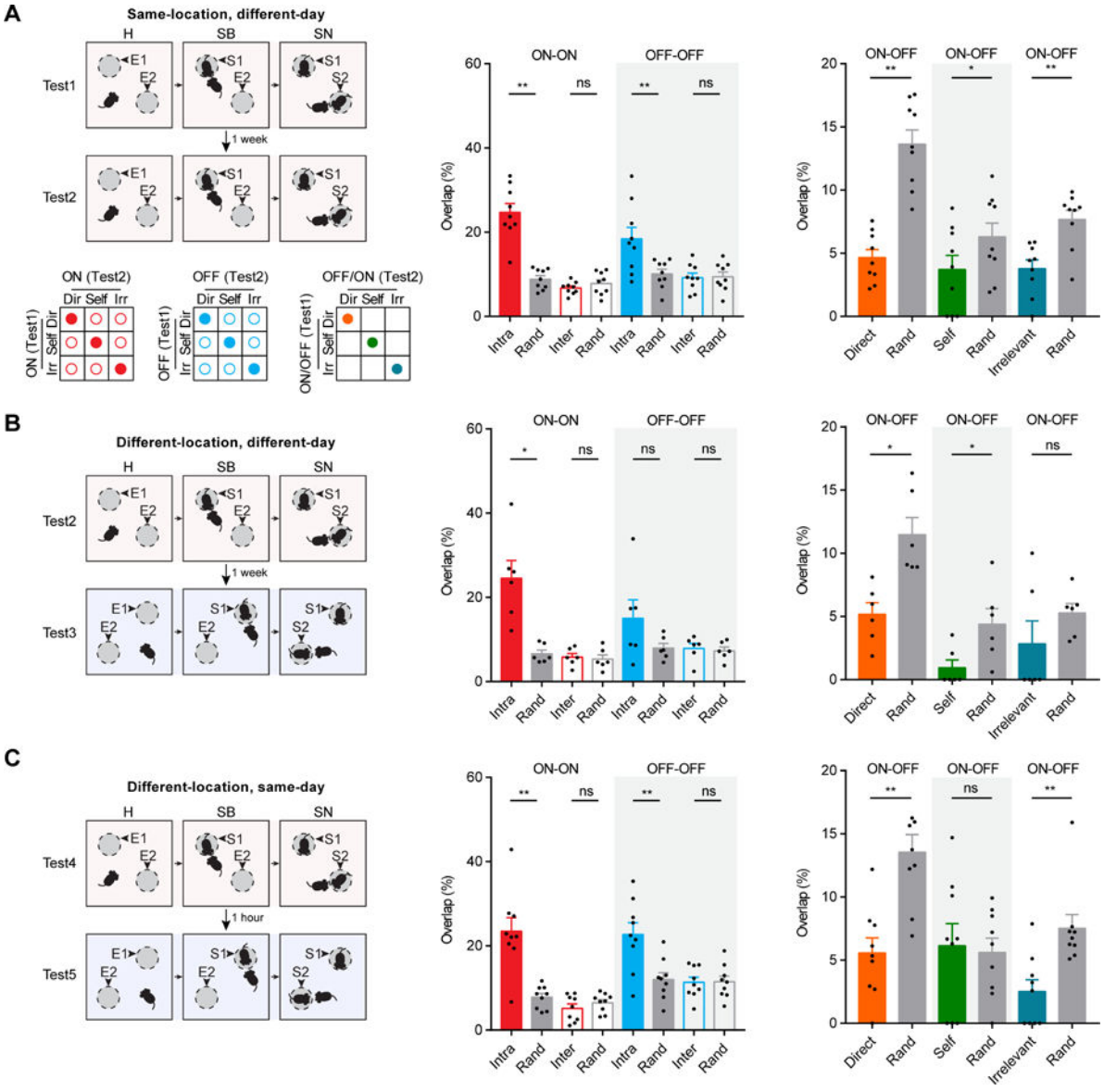


Figure 5: Across-test overlap analyses between ensembles identified during different social behavior tests.

(**A**) Top left panel: schematic illustration of the two social behavior tests performed at different days with strangers presented at the same locations. Bottom left panels: illustrations for ON-ON intra- and inter-behavior overlap, OFF-OFF intra- and inter-behavior overlap, and ON-OFF intra-behavior overlap for each annotated behavior as indicated by colored circles. Middle panel: across-test ON-ON and OFF-OFF intra- and inter-behavior overlaps and those predicted by chance based on random selection, n = 9 mice. Right panel: across-test ON-OFF intra-behavior overlap and those predicted by chance based on random selection, n = 9 mice. (**B**) Overlap analysis for two social behavior tests performed at different days with strangers presented at different locations, n = 6 mice. (**C**) Overlap analysis for two social behavior tests performed on the same day with strangers

presented at different locations, n = 9 mice. Data were represented as mean ± SEM. See also Figure S4, Figure S5 and Table S1.

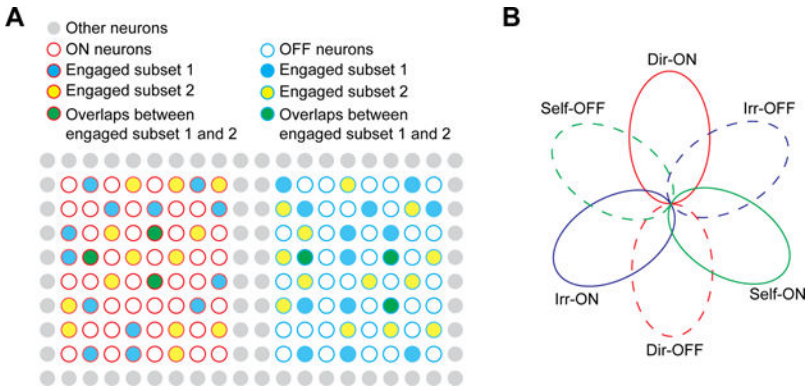


Figure 6: Dynamic population coding model.

(**A**) Schematic illustration for dynamic population coding model. Red circle: ON neurons; Blue circle: OFF neurons. “Engaged subset 1” and “Engaged subset 2” represent two subs ets of neurons engaged in behaviore coding at different points in time. (**B**) Schematic illustration of different ON and OFF neural ensembles coding for three annotated behaviors. Each oval represents a neural ensemble tuned to one annotated behavior. See also Figure S5F.

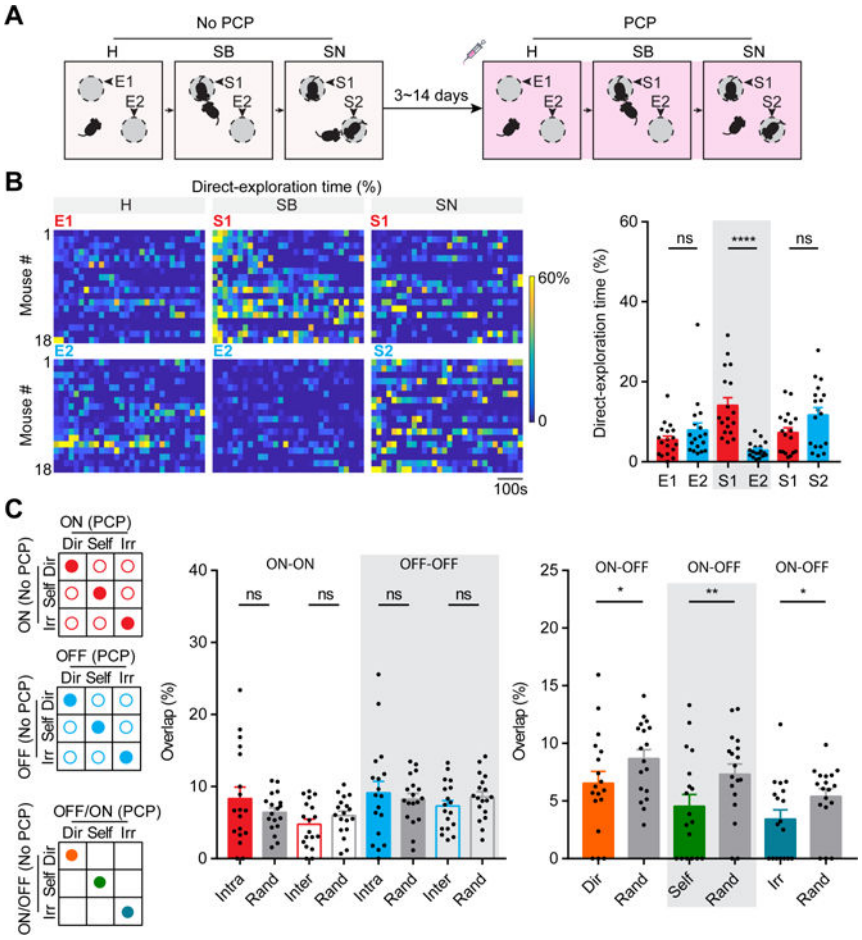


Figure 7. Acute PCP administration elicited deficits in social exploration and ensemble reorganization in the mPFC.

(A) Schematic illustration for social behavior tests performed in the absence and presence of PCP. H: habituation; SB: sociability; SN: social novelty. (**B**) Left panels: Raster plots for percentage of time (per 20 second bin) spent on direct-exploration toward different targets for individual mouse under acute PCP. Right panel shows the statistical results, n = 18 mice. (**C**) Ensemble overlap analyses between two social behavior tests performed in the absence and presence of PCP. Left panels: illustrations for ON-ON intra- and inter-behavior overlap (filled and unfilled red circles, respectively), OFF-OFF intra- and inter-behavior overlap (filled and unfilled blue circles, respectively), and ON-OFF intra-behavior overlap for each annotated behavior as indicated by colored circles. Middle panel: statistical results showing across-test ON-ON and OFF-OFF intra- and inter-behavior overlaps and those predicted by chance based on random selection, n = 18 mice. Right panel: statistical results showing across-test ON-OFF intra-behavior overlap and those predicted by chance based on random selection, n = 18 mice. Data were represented as mean ± SEM. See also Figure S6, Figure S7 and Table S1.

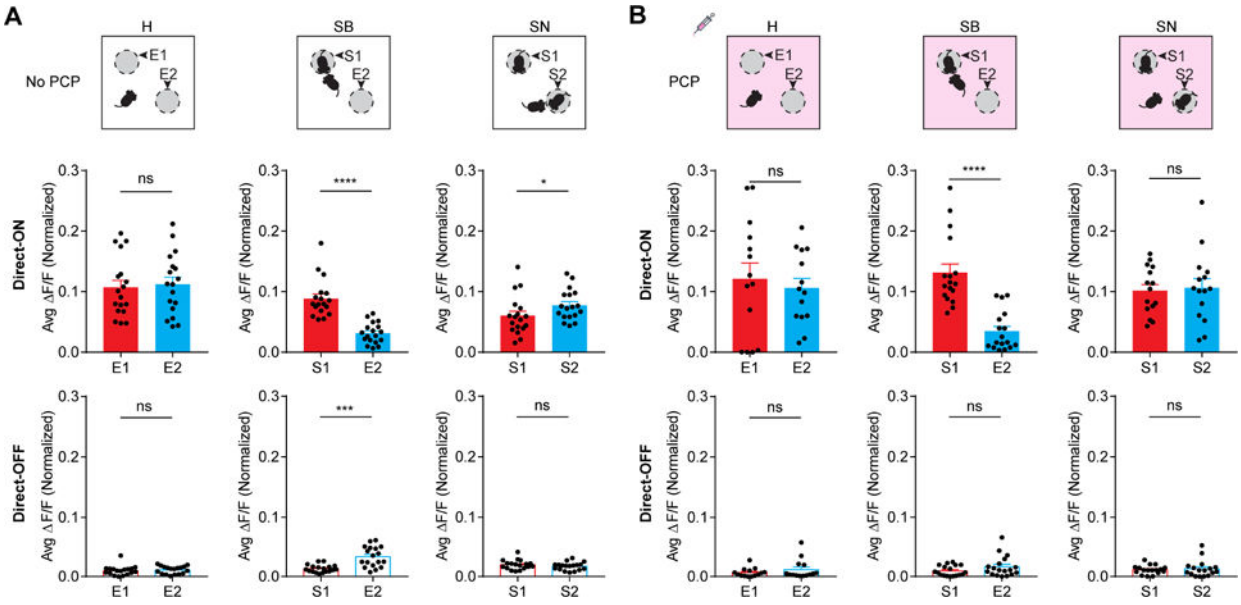


Figure 8: PCP disrupted information coding of social salience and novelty by Direct-ON and Direct-OFF ensembles.

(A) Top panels: schematic diagrams of the social behavior test performed in the absence of PCP. H, habituation; SB, sociability; SN, social novelty. E1 and E2 refer to empty containers; S1 and S2 refer to stranger1 and stranger 2, respectively. Middle and bottom panels: calcium activities of Direct-ON (middle panels) and Direct-OFF (bottom panels) ensembles responding to the two explorative targets at the H, SB and SN stages, n = 18 mice. (**B**) Top panels: schematic diagrams indicating the social behavior test under acute PCP influence. Middle and bottom panels: calcium activities of Direct-ON (middle) and Direct-OFF (bottom) ensembles responding to the two explorative targets at the H, SB and SN stages, n = 14 ~ 18 mice. Data were represented as mean ± SEM. See also Figure S8, Table S1.