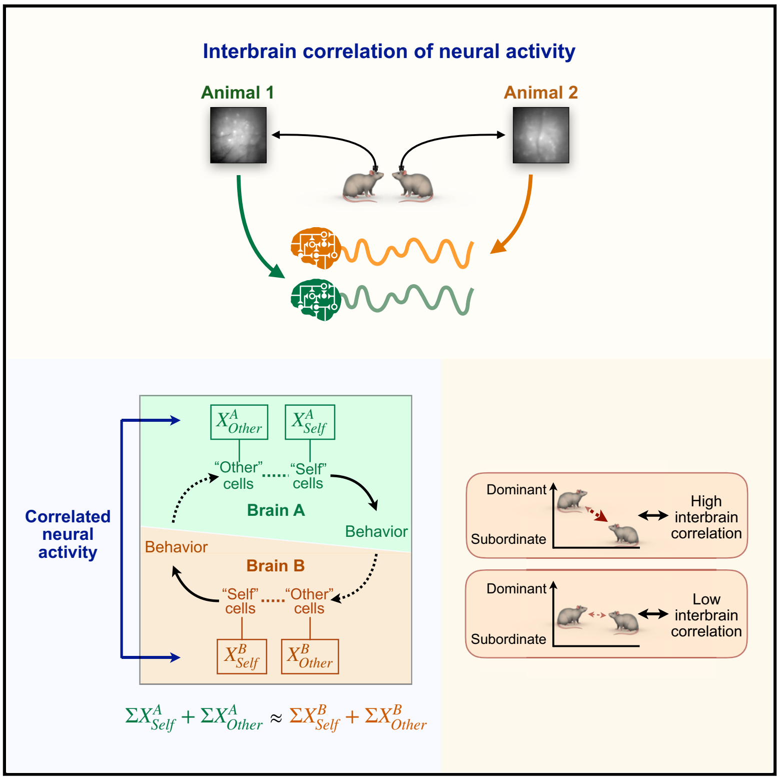
Cell

Article

Correlated Neural Activity and Encoding of Behavior across Brains of Socially Interacting Animals

Graphical Abstract


Interbrain correlation of neural activity
Animal 1
Animal 2
Correlated neural activity
X^A_{Other}
X^A_{Self}
"Other" cells
"Self" cells
Brain A
Behavior
Behavior
Brain B
"Self" cells
"Other" cells
X^B_{Self}
X^B_{Other}
$\Sigma X^A_{Self} + \Sigma X^A_{Other} \approx \Sigma X^B_{Self} + \Sigma X^B_{Other}$
Dominant
Subordinate
High interbrain correlation
Dominant
Subordinate
Low interbrain correlation

Highlights

- Simultaneous imaging of interacting mice reveals interbrain synchrony of activity
- Interbrain synchrony arises from ongoing social interaction between animals
- Synchrony emerges from neurons encoding behavior of oneself and the social partner
- Interbrain synchrony predicts future social decisions and dominance relationships

Authors

Lyle Kingsbury, Shan Huang, Jun Wang, Ken Gu, Peyman Golshani, Ye Emily Wu, Weizhe Hong

Correspondence

whong@ucla.edu

In Brief

When two animals interact, neural activity across their brains synchronizes in a way that predicts how they will behave and how they form social dominance relationships.

Kingsbury et al., 2019, Cell *178*, 429–446
July 11, 2019 © 2019 Elsevier Inc.
https://doi.org/10.1016/j.cell.2019.05.022

Cell

Correlated Neural Activity and Encoding of Behavior across Brains of Socially Interacting Animals

Lyle Kingsbury,[1,3] Shan Huang,[1,3] Jun Wang,[1] Ken Gu,[1] Peyman Golshani,[2] Ye Emily Wu,[1] and Weizhe Hong[1,4,*]
[1]Department of Biological Chemistry and Department of Neurobiology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA
[2]Department of Neurology and Department of Psychiatry & Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA
[3]These authors contributed equally
[4]Lead Contact
*Correspondence: whong@ucla.edu

https://doi.org/10.1016/j.cell.2019.05.022

SUMMARY

Social interactions involve complex decision-making tasks that are shaped by dynamic, mutual feedback between participants. An open question is whether and how emergent properties may arise across brains of socially interacting individuals to influence social decisions. By simultaneously performing microendoscopic calcium imaging in pairs of socially interacting mice, we find that animals exhibit interbrain correlations of neural activity in the prefrontal cortex that are dependent on ongoing social interaction. Activity synchrony arises from two neuronal populations that separately encode one's own behaviors and those of the social partner. Strikingly, interbrain correlations predict future social interactions as well as dominance relationships in a competitive context. Together, our study provides conclusive evidence for interbrain synchrony in rodents, uncovers how synchronization arises from activity at the single-cell level, and presents a role for interbrain neural activity coupling as a property of multi-animal systems in coordinating and sustaining social interactions between individuals.

INTRODUCTION

Social interactions involve some of the most complex decision-making tasks that animals must navigate to secure their survival and reproductive success (Chen and Hong, 2018), as individuals must integrate internal state with real time decisions of their social partners in a context-dependent manner. In interacting dyads, individuals thus become entrained as they attend to, predict, and react to each other's decisions (Figure S1A) (Rilling and Sanfey, 2011; Sanfey, 2007). To date, social neuroscience has mostly focused on behavior in individual animals to interrogate the neural computations underlying social decision-making. But a full understanding of the social brain requires a broader picture that reflects the dynamic nature of social interactions, as well as the emergent neural properties that arise from multiple individuals as a single integrated system (Adolphs, 2010; Chen and Hong, 2018; Ochsner and Lieberman, 2001; Schilbach et al., 2013).

In recent years, much effort has been made to explore how neural systems coordinate across individuals engaged in social interaction. Simultaneous recordings from multiple human subjects using non-invasive techniques (e.g., functional MRI [fMRI] and electroencephalography [EEG]) have revealed striking patterns of interbrain neural activity coupling during social engagement (Babiloni et al., 2006; King-Casas et al., 2005; Liu and Pelowski, 2014; Montague et al., 2002). Despite these remarkable findings, little is concretely known about how interbrain synchrony arises from social interactions. Moreover, it remains unclear how synchrony emerges from individual neurons and neuronal populations, in part due to the limited spatial resolution of recording techniques in humans, which cannot resolve single-cell activity. It is also unclear whether brain synchrony is unique to primates, or whether it is a general phenomenon present in other social species.

Competitive interactions are common among social species and play an important role in shaping social status hierarchies (Williamson et al., 2016) which influence the long-term health of individuals (Cooper et al., 2015; Sapolsky, 2004, 2005). Navigation of social interactions depends on circuitry in the medial prefrontal cortex (mPFC), which is implicated in the representation of social status (Utevsky and Platt, 2014; Wang et al., 2011; Zhou et al., 2017) and shapes social and motivational states (Franklin et al., 2017; Warden et al., 2012). However, while previous work has shown that mPFC neurons are active during social interaction (Liang et al., 2018; Murugan et al., 2017), it has not been clear how prefrontal ensembles encode behavioral decisions during real-time social engagements, such as social competition.

Here, we used microendoscopic calcium imaging to record from thousands of neurons in the dorsomedial prefrontal cortex (dmPFC) of pairs of mice engaged in social interactions. Our study provides conclusive evidence for interbrain activity correlations in interacting mice as well as a cellular level neural basis

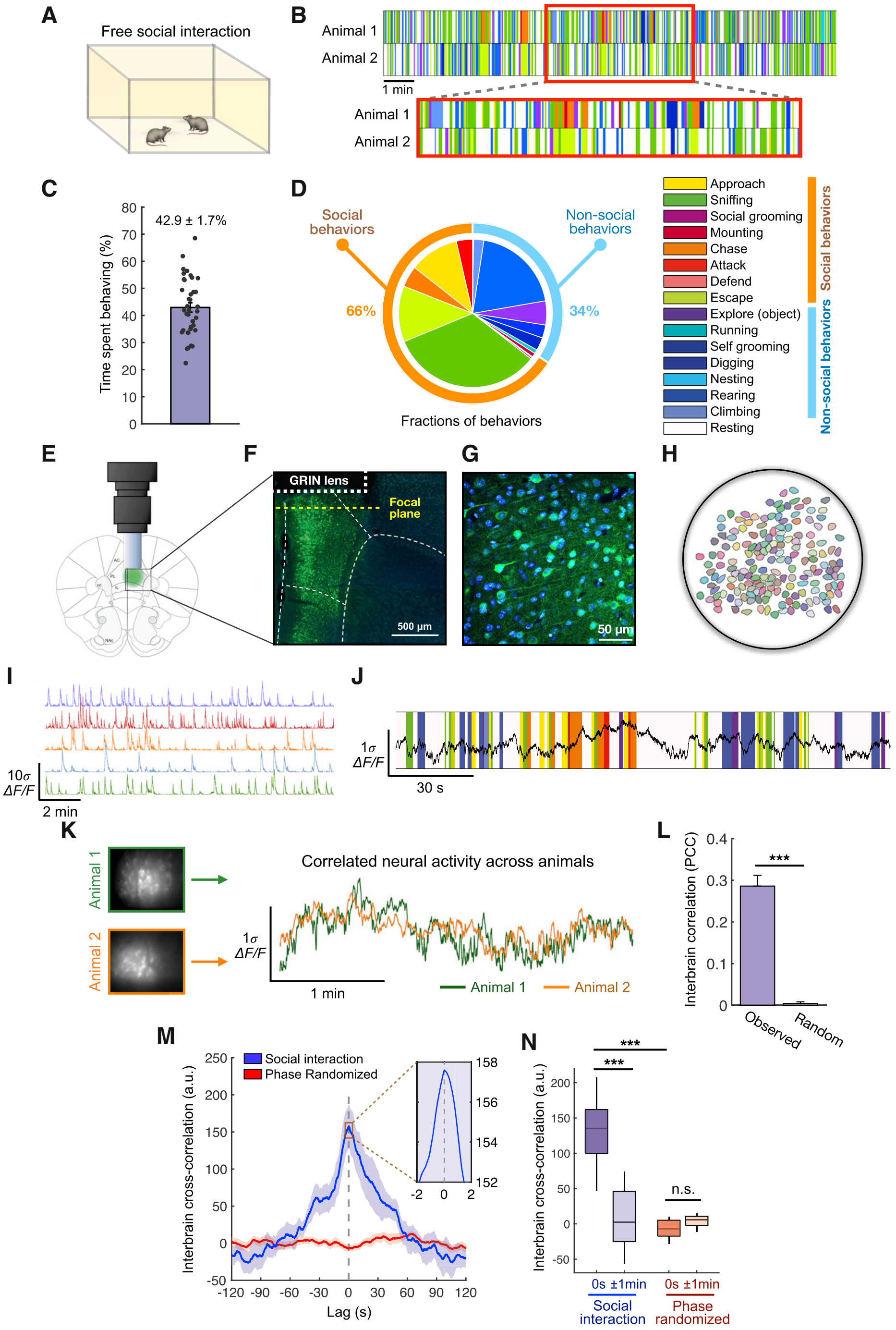


Figure 1. Correlated Neural Activity across Brains of Interacting Animals during Free Social Interaction

(A) Illustration of social interactions in the open arena.

(B) Behavior raster plot of two animals interacting in the open arena.

(C) Percentage of time animals engage in behavior in the open arena. Each dot represents one animal from one session.

(D) Distribution of behaviors mice display in the open arena interaction.

(E) Schematic of head-mounted microscope and GRIN lens implantation above dmPFC.

(F) Example image of injection site showing expression of GCaMP6f in dmPFC.

(G) Example image showing viral expression in dmPFC cell bodies. Green, GCaMP6f; blue, DAPI.

(H) Example imaging field of view with individual cell bodies.

(I) Example calcium traces recorded from one session.

(legend continued on next page)

underlying this phenomenon and identifies a critical role for interbrain synchrony in coordinating and facilitating social interaction.

RESULTS

Correlated Neural Activity across Animals during Free Social Interaction

During natural social encounters, animals exhibit a wide range of behavior that engage them in complex, often reciprocal interactions. To study neural dynamics across brains of socially interacting mice, we first examined naturally occurring behaviors during social interactions in an open arena, where two novel animals were permitted to freely interact (Figure 1A). We recorded the interaction using a video camera and annotated behaviors of both animals frame-by-frame (Figure 1B). Across all sessions, we identified 15 types of behaviors that included both social and non-social behavior. While animals spent about 43% of the time engaged in observable behavior (Figure 1C), the majority of this (~66%) was social behavior directed toward the interacting partner (Figures 1D and S2A). Thus, the open arena provides an unconstrained context where animals freely engage in highly diverse and naturalistic social interactions.

To investigate neural dynamics during the social interaction, we employed microendoscopic calcium imaging to simultaneously monitor activity from hundreds of dmPFC neurons in both individuals. To gain optical access to neurons below the cortical surface, we implanted a gradient refractive index (GRIN) lens above the dmPFC following injection of an AAV (adeno-associated virus) expressing the fluorescent calcium indicator GCaMP6f (Figure 1E). Lens placement and GCaMP6f expression were confirmed histologically (Figures 1F and 1G). Calcium fluorescence videos were processed using independent component analysis to identify putative cell bodies, which were used to extract calcium traces from single cells, expressed throughout as relative change in fluorescence ($\Delta F/F$) (Figures 1H and 1I; Video S1; STAR Methods). We analyzed a total of 7,535 dmPFC neurons in 19 pairs of animals engaged in open arena social interaction. Overall neural activity varied across different types of behaviors (Figure S2B), suggesting that activity in the dmPFC is differentially modulated by social behavior.

To explore how dmPFC neural dynamics were related across individuals, we computed the mean activity of neurons in each animal as aggregate signals that reflect the overall activity of the population (Figure 1J) and quantified correlations of activity (Pearson's correlation coefficient, PCC) across dyads in each session. Strikingly, dmPFC populations displayed highly correlated activity across animals, which far exceeded chance levels (Figures 1K, 1L, and S2C). To examine the timescale of interbrain correlations, we measured the cross-correlation of dmPFC activity across animals (Figure 1M); these showed a clear peak at 0.0 s, indicating precise synchrony of interbrain activity. This interbrain correlation was not due to autocorrelations in each signal, as the cross-correlation structure was abolished when traces were phase-randomized (Figure 1N). Together, these results establish that animals engaged in free social interaction exhibit highly correlated dmPFC activity.

Interbrain Activity Correlations Depend on Ongoing Social Interaction

Animals in a social environment are naturally inclined to engage with one another, but they occasionally exhibit periods of coordinated rest in which they are both quiescent (Figure S2D). To address whether interbrain correlations could be simply explained by concurrent behavior or rest periods, we removed epochs in which both animals did not exhibit observable behavior and compared interbrain correlations during these epochs with those of full sessions. Activity after discounting periods of coordinated rest was as correlated as activity during full sessions (Figure 2A), suggesting that bouts of concurrent rest cannot account for activity correlations.

Although animals do not tend to exhibit the same behaviors at the same time (Figure S2D), interacting animals do sometimes behave concurrently. To determine whether overall concurrent behavior could explain interbrain synchrony, we compared interbrain correlations during epochs with low versus high levels of concurrent behavior (measured by correlation of overall behavior, Figure 2B). Again, interbrain correlations during these epochs were not different and were equally disrupted upon phase randomization of activity traces.

To explore the relationship between interbrain synchrony and social interaction, we next compared the degree of interbrain correlation during social versus non-social behavior. Correlations were significantly higher during social behavior (Figure 2C), suggesting dependence on social interaction. However, because animals are in the same environment, there is a possibility that correlated activity reflects shared sensory inputs such as ambient noise or lighting rather than social engagement. To rule this out, we separated the animals within the same physical environment using a barrier (Figure 2D). Abolishing social interaction significantly reduced interbrain correlations among dmPFC neurons (Figures 2E, 2F, and S2E), suggesting that correlated activity is not due to shared sensory input but actually depends on ongoing interaction between the pair. Indeed, when we recorded from pairs of animals that naturally displayed low levels of social interaction, a lower degree of correlation was observed (Figure S2F).

(J) Example trace showing overall dmPFC activity (mean activity of all cells) in one animal during social interaction overlaid with behavior annotations.

(K) Example calcium traces showing overall dmPFC activity from two animals engaged in social interaction.

(L) Interbrain correlations of overall dmPFC activity in animals, compared with those of temporally permuted traces.

(M) Cross-correlation of dmPFC activity traces from interacting animals compared with that of phase-randomized traces.

(N) Quantification of cross-correlations shown in (M) at 0 s or ± 60 s.

***$p < 0.001$, $p > 0.05$, not significant (n.s.). (C, L, M) Mean ± SEM. In (L)–(N) and Figure 2, pairs with a relatively high degree of social interaction were analyzed (STAR Methods).

See also Figures S1 and S2 and Video S1.

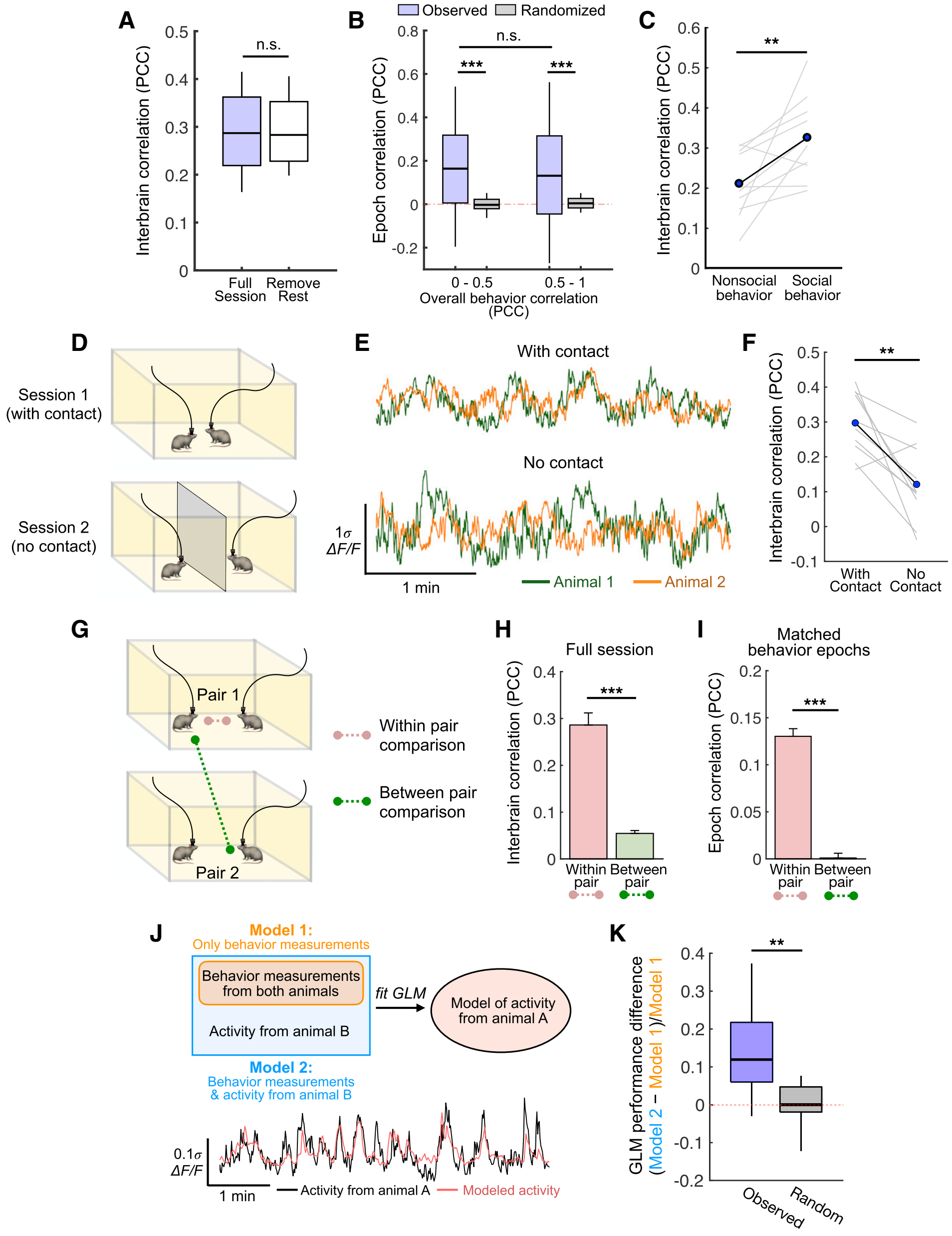


Figure 2. Interbrain Correlations Depend on Ongoing Social Interaction

(A) Interbrain correlations of dyads during full open arena sessions or correlations after removing epochs of concurrent rest, defined as when both animals display no observable behavior.

(B) Interbrain activity correlations during single epochs (1 min) with low or high behavior correlation (the PCC of binary vectors measuring the presence of any behavior), compared with correlations of phase-randomized signals.

(C) Interbrain correlations of epochs when one or both animals engaged in social versus non-social behavior.

(D) Schematic of the open arena interaction with social contact or with separation of animals with a barrier. Head-mounted microscopes were connected via an ultra-light cable that is long and flexible enough to prevent tangling during the course of social interactions.

(E) Example calcium traces of overall dmPFC activity (mean activity) in a dyad with or without social contact.

(F) Interbrain correlations in pairs with or without social contact.

(G) Schematic showing comparisons of correlations across pairs engaged in social interaction (within pair) and across animals that each interacted with a different animal (between pair).

(H) Comparison between interbrain correlations across interacting or non-interacting pairs.

(I) Interbrain correlations during single epochs (30 s) with concurrent behavior bouts in interacting pairs or those over behavior-matched epochs in non-interacting animals.

(legend continued on next page)

Given this observation, another hypothesis is that interbrain correlations reflect generic activity associated with social interaction, such as motivational state, regardless of whether animals are directly engaged. To rule this out, we examined neural activity across pairs of animals that each engaged in social interactions, but with separate animals and not with each other (Figure 2G). Activity correlations across animals from different sessions were significantly lower than those across interacting pairs (Figure 2H), confirming that directed engagement between two animals is necessary for interbrain coupling.

Moreover, it is possible that interbrain correlations could be purely explained by activity associated with individual coordinated behavior bouts at finer timescales. To address this, we computed correlations during epochs with coordinated behavior bouts, and compared them with correlations during behavior-matched epochs in non-interacting animals (Figure 2I). Activity from behavior-matched epochs across non-interacting pairs did not exhibit correlations; only those in socially interacting animals showed interbrain coupling (Figures 2I). This suggests that interbrain synchrony cannot be simply explained by overall concurrent behavior or individual coordinated behavior bouts but depends upon the context of a direct, ongoing social interaction. For example, the same type of behavior may be associated with different patterns of activity depending on social context (e.g., interactions over a longer timescale or specific social relationships).

Last, to further understand the relationship between dmPFC activity and behavior, we modeled activity in each animal as a function of behavior and activity recorded from the interacting partner. We constructed generalized linear models (GLMs) to model dmPFC activity from behaviors exhibited by both animals (Figure 2J; model 1; STAR Methods) and compared it to a second model fit using the overall activity from the interacting partner as an additional variable (model 2). We reasoned that, if neural activity in one animal did not contain information relevant to activity in the interacting partner beyond what is explained by individual behaviors, models that included partner activity (model 2) would not perform better than "behavior-only" models (model 1). In fact, model 2 performed significantly better than model 1 (Figure 2K), suggesting that activity in one animal contains additional information about activity in the other that cannot be fully explained by moment-to-moment behavior. This is consistent with the notion that interbrain coupling depends on the larger context of an ongoing interaction.

Behavioral Dynamics during a Competitive Social Encounter

To explore whether interbrain coupling was present in other contexts, such as competitive interaction, we adopted a social dominance assay (the tube test) that allowed us to examine competitive behavior and dominance relationships across dyads (Drews, 1993; Wang et al., 2014) (Figure 3A). In the tube test, mice are placed facing each other in a one-dimensional tube and allowed to push each other or retreat from conflict. Winning in the tube test (by pushing the other animal out of the tube) has previously been used to operationalize dominance behavior, as it correlates with other social status behavior in mice (Wang et al., 2011). Compared to the open arena, the tube test also offers an advantage of narrowing the animals' decisions to a set of well-defined behaviors, enabling a precise interrogation of the relationship between interbrain synchrony and single cell encoding of behavioral decisions.

To analyze behavioral dynamics during the tube test, we recorded the interaction using a video camera and developed an automated tracking algorithm using a convolutional neural network (Redmon and Farhadi, 2016) to track the positions of both animals (Figures 3B and 3C), which we validated by unbiased visual assessment (>99% accuracy; Figure S3A). We also manually annotated videos frame by frame to identify the onset and duration of behaviors in both animals. We observed that animals displayed three distinct types of behavior in the tube test: *approach*, a forward approach toward the opponent; *push*, a forceful push against the opponent sometimes resulting in forward movement; and *retreat*, a backward retreat away from the opponent. This parcellation, together with the position tracking, allowed us to examine how competitive interactions lead to gains or losses in territory for each animal.

On average, animals spent 23% of the time engaged in observable social interactions (Figure 3D), the majority of which (71%) was push behavior (Figure 3E). Although not all behavioral decisions lead to positional changes between the pair, position changes represent gains or losses of territory that result from competitive interaction. That is, each animal's position can be considered as a function of its individual decisions to approach, push, or retreat from conflict, thus characterizing its overall level of relative dominance. Within each pair, we identified the more dominant animal as the one who gained more territory on average (STAR Methods) and confirmed that dominant and subordinate animals exhibited large differences in tube position (Figure 3F).

In any complex social engagement, reciprocal interaction is a common feature. Indeed, dyads behaved reciprocally in a fraction of the total time (Figure 3G), indicating that their behavioral decisions partially depended on one another. To examine how animals reacted to each other, we analyzed how the probability of each behavior in one animal changed following opponent behavior (Figures 3H–3J). Overall, push behavior was followed by a probabilistic increase in retreat behavior in opponents, indicating that, while not all pushes result in opponent reactions, push and retreat behavior are sometimes linked (Figure 3H). There was also an increase in approach behavior following

(J) Schematic showing GLMs used to model dmPFC activity in one animal as a function of behaviors exhibited by both animals (model 1) or as a function of behaviors as well as activity in the other animal (model 2). An example of modeled dmPFC activity is shown in the pink trace.

(K) Relative difference in GLM model performance when activity from the interacting partner is included as additional explanatory variable (J), compared with that using phase randomized controls (STAR Methods; also see Figure S4G for the equivalent analysis in the tube test).

***p < 0.001, **p < 0.01, p > 0.05, n.s. (H and I) Mean ± SEM.

See also Figure S2.

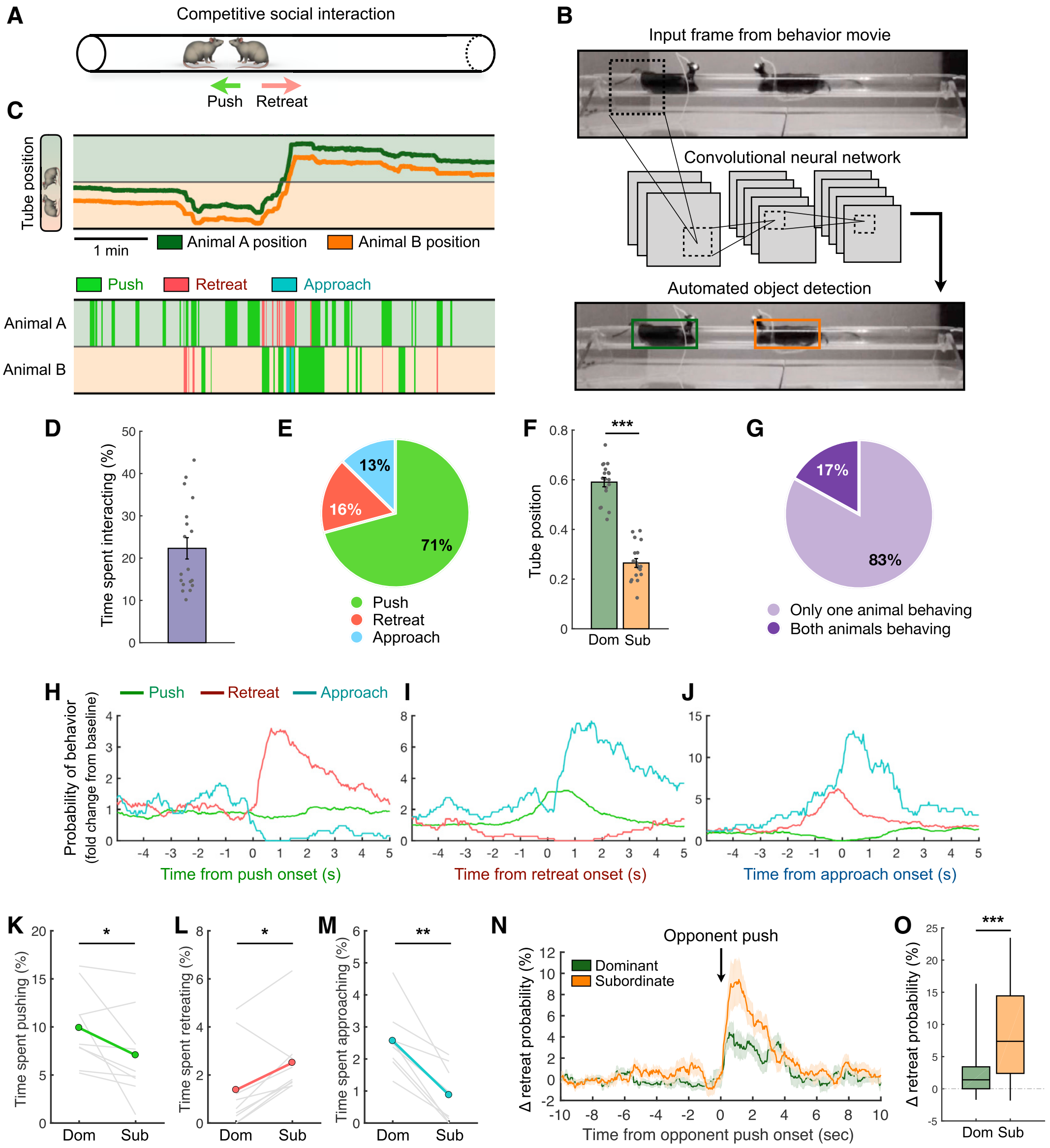


Figure 3. Dynamics of Social Behaviors during Competitive Interaction

(A) Cartoon of mice engaged in the tube test.

(B) Illustration of the neural network used for automated tracking of mice.

(C) Behavior annotations and position trajectories of a pair of mice in the tube test.

(D) Total percentage of time animal pairs (either animal) engaged in social interaction.

(E) Distribution of time animals displayed different behaviors.

(F) Average tube positions in dominant or subordinate animals (STAR Methods).

(G) Fraction of interaction time when only one or both animals are behaving.

(H–J) Change in probability of opponent animal behavior with respect to subject animal push (H), retreat (I), or approach (J).

(K–M) Percentage of time spent pushing (K), retreating (L), and approaching (M) in dominants or subordinates in pairs that displayed a large difference in dominance (STAR Methods).

(N) Change in relative probability of dominant or subordinate retreat behavior following opponent push.

(O) Probability of retreat in dominants or subordinates 1 s following opponent push.

***p < 0.001, **p < 0.01, p > 0.05, n.s. (D, F, and N) Mean ± SEM.

See also Figure S3.

opponent retreats (Figure 3I), suggesting that animals were generally motivated to engage with their opponent.

Because dominants and subordinates exhibit similar levels of behavior overall (Figure S3B), we reasoned that differences in tube position likely reflect differences in the distributions of displayed behavior. Indeed, dominants pushed more, retreated less, and approached more than subordinates (Figures 3K–3M). We found no differences between the per-bout durations of behaviors displayed by dominants and subordinates (Figures S3C–S3E), suggesting that differences in dominance (i.e., territory gained) depend mostly on the frequency of different social decisions.

Differences in overall dominance may also depend on how animals react to behavior from their opponent. To explore this, we constructed time courses of animals' change in retreat probability following opponent push behavior. While both dominants and subordinates showed a probabilistic increase in retreats following opponent pushes, subordinates were more likely to retreat reactively (Figures 3N and 3O). Collectively, this analysis shows that outcomes of dominance encounters between mice depend not only on different behavioral choices in each animal but also on how each animal responds to its opponent.

Animals Display Interbrain Correlations during a Competitive Social Encounter

To determine whether mice engaged in social competition also display interbrain coupling, we simultaneously imaged dmPFC activity using microendoscopes in animal dyads during the tube test (Figure 4A; Video S2). As in the open arena, overall dmPFC activity was highly correlated across interacting animals in the tube test, far exceeding chance levels (Figures 4B, 4C, and S4A–S4C).

We first ruled out the possibility that correlated activity in this context is due simply to concurrent behavior or rest: neural activity correlations were consistently higher than correlations of overall behavior (Figure 4D), removing coordinated rest epochs did not reduce neural correlations (Figure 4E), correlations remained high when only one animal was behaving (Figure 4F), and activity correlations were higher than chance even during epochs with a lower level of concurrent behavior (Figure 4G). These suggest that interbrain coupling is not simply due to concurrent behavior or rest.

To confirm, as in the open arena, that interbrain coupling is not due to similar sensory inputs from a shared environment, we separated animals inside the tube so that both could freely move but could not interact (Figure 4H). Activity correlations were significantly reduced after separation (Figures 4I, 4J, and S4E), indicating that brain coupling in a competitive context also depends on ongoing interaction. In addition, comparisons of activity correlations in interacting versus non-interacting pairs (Figure 4K) revealed that social engagement in the same encounter is necessary for correlated activity (Figure 4L). In support of this, while concurrent behavior epochs in interacting pairs had correlated activity, behavior-matched epochs in non-interacting pairs did not (Figures 4M and S4F).

As in the open arena, dmPFC activity from interacting animals also exhibited peak cross-correlation at 0.0 s (Figure 4N), indicating that interbrain activity is precisely synchronized. However, the cross-correlation was disrupted upon phase randomization and not significantly higher at zero time lag than at a lag of 30 s (Figure 4O), indicating a strong reduction in interbrain correlation.

Collectively, these results demonstrate that mice engaged in a competitive social encounter reliably display correlated activity across dmPFC neurons that depends on ongoing interactions in a larger social context and cannot be simply explained by overall concurrent behavior or individual coordinated behavior.

dmPFC Neurons Encode Distinct Social Behaviors during Competitive Interaction

Overall activity patterns of a brain region arise from individual cells, but a cellular-level basis for interbrain synchrony remains elusive. To explore how activities in single cells contribute to synchronous activity across animals, we first examined whether dmPFC neurons encode distinct social behaviors. dmPFC neurons as a whole exhibited time-locked excitation during push, retreat, and approach behavior (Figure 5A). However, this raises the question of whether behavioral decisions are associated with uniform activation of the dmPFC or are encoded uniquely by distinct subsets of dmPFC neurons.

To address this, we examined whether single cells responded during specific behaviors. Using a receiver operating characteristic (ROC) analysis (Figures 5B and S5A), we identified subsets of neurons that were excited or suppressed during push, approach, or retreat behavior (Figures 5C–5F and S5B). Of all recorded neurons, 35% encoded social behaviors (Figure 5D), and, among these, ~80% showed selective tuning to specific behaviors. Cells that were not identified as behavior-encoding (hereafter referred to as "neutral cells") were just as active, overall, as behavior cells (Figure S5C), indicating that behavior encoding was due to specific time-locked responses. Interestingly, while behavior cells included both excited and suppressed ones, the majority were excited (Figure 5E). Overall, we found no differences in the spatial distributions of behavior cells compared with neutral cells (Figures 5G and 5H), indicating that behavior cells are spatially intermixed. These results demonstrate that a substantial fraction of dmPFC neurons selectively encode social behaviors in the tube test.

Information can be more robustly encoded at the population level than among single, highly tuned cells (Pouget et al., 2000). We next investigated whether neurons in the dmPFC formed stable activation patterns encoding social behaviors that could be read out at the population level. We examined how population response dynamics differed between types of behaviors using the Mahalanobis distance between behavior-evoked responses and baseline activity (Figures 5I and S6A). Again, we found that all behaviors elicited time-locked responses. Interestingly, push and approach elicited stronger response patterns than retreat (Figure 5J), consistent with the idea that distinct behaviors are encoded differentially rather than as an aggregate of ensemble activity. To analyze the separability of population dynamics during behavior, we visualized population responses using principal-component analysis (PCA); this revealed a clear separation of activity clusters based on behavior type (Figures 5K and S6B–S6D). Further, the

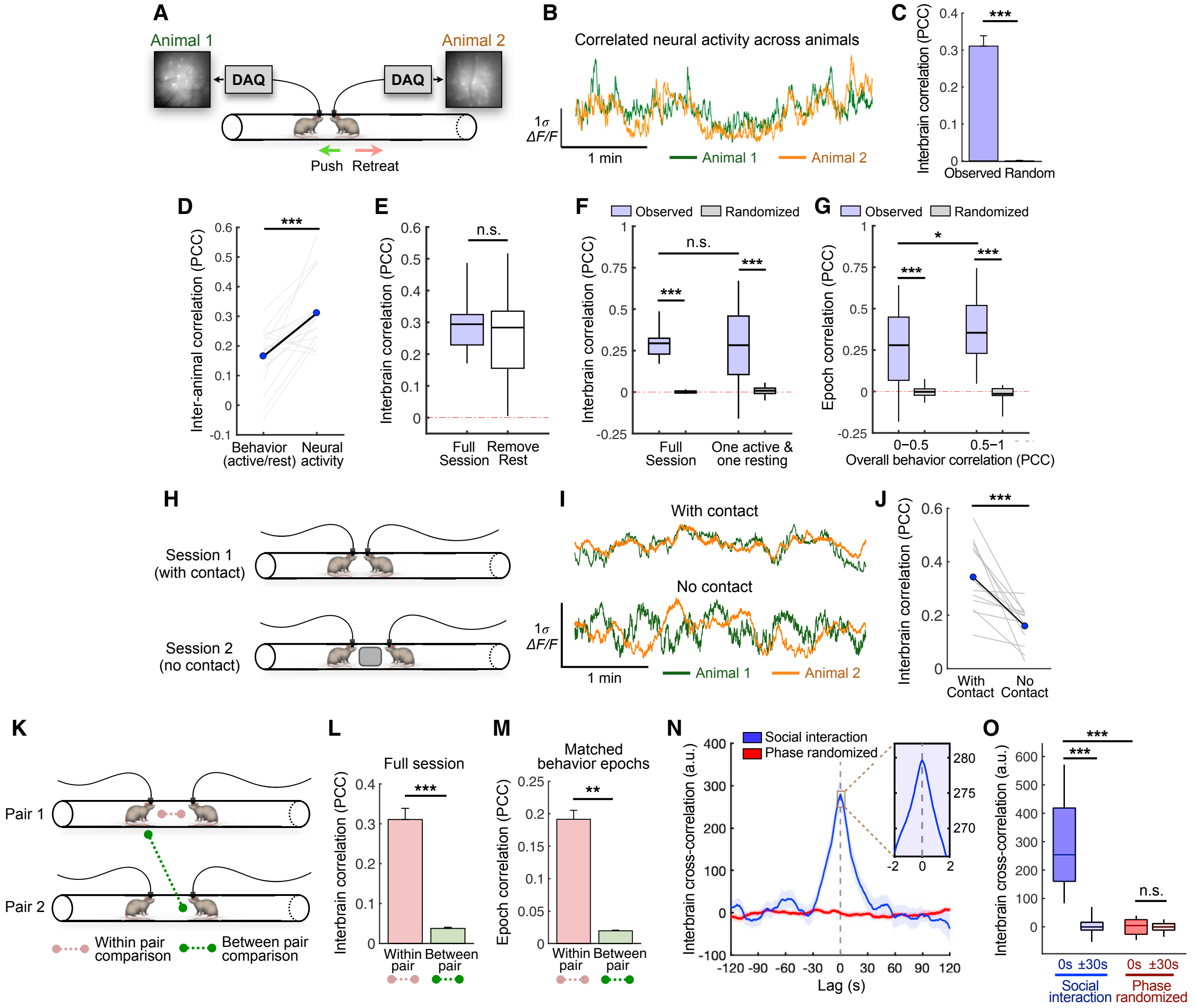


Figure 4. Correlated Neural Activity across Animals during Competitive Social Interaction

(A) Cartoon showing simultaneous imaging of two mice during the tube test.

(B) Example traces of overall dmPFC activity (mean of all neurons) from two animals during the tube test.

(C) Interbrain correlations in interacting pairs or correlations of randomly permuted traces.

(D) Comparison of correlations of behavior (PCC of binary event vectors) across animals versus correlations of dmPFC activity.

(E) Interbrain correlations during the tube test or after removing concurrent rest epochs when both animals display no observable behavior.

(F) Interbrain correlations during tube test sessions or during epochs (≥1 min) when one animal is behaving while the other is resting (displaying no observable behavior), compared with phase randomized controls.

(G) Interbrain correlations during single epochs (1 min) of low or high overall behavior correlation (PCC of binary vectors measuring the presence of any behavior), compared with those of phase-randomized traces.

(H) Schematic showing introduction of a separator in the tube test to abolish social contact.

(I) Example traces showing dmPFC activity across two animals with or without social contact.

(J) Interbrain correlations with or without social contact.

(K) Schematic showing pairs engaged in social interaction (within pair) or pairs that each interact with a different animal (between pair).

(L) Interbrain correlations across interacting or non-interacting animals.

(M) Interbrain correlations during single epochs (30 s) with concurrent behavior bouts in interacting pairs or during behavior-matched epochs in non-interacting animals.

(N) Cross-correlation of dmPFC activity from pairs of mice in the tube test and that of phase-randomized controls.

(O) Quantification of cross-correlations shown in (N) at 0 s versus ± 30 s.

***$p < 0.001$, **$p < 0.01$, $p > 0.05$, n.s. (C and L–N) Mean ± SEM.

See also Figure S4 and Video S2.

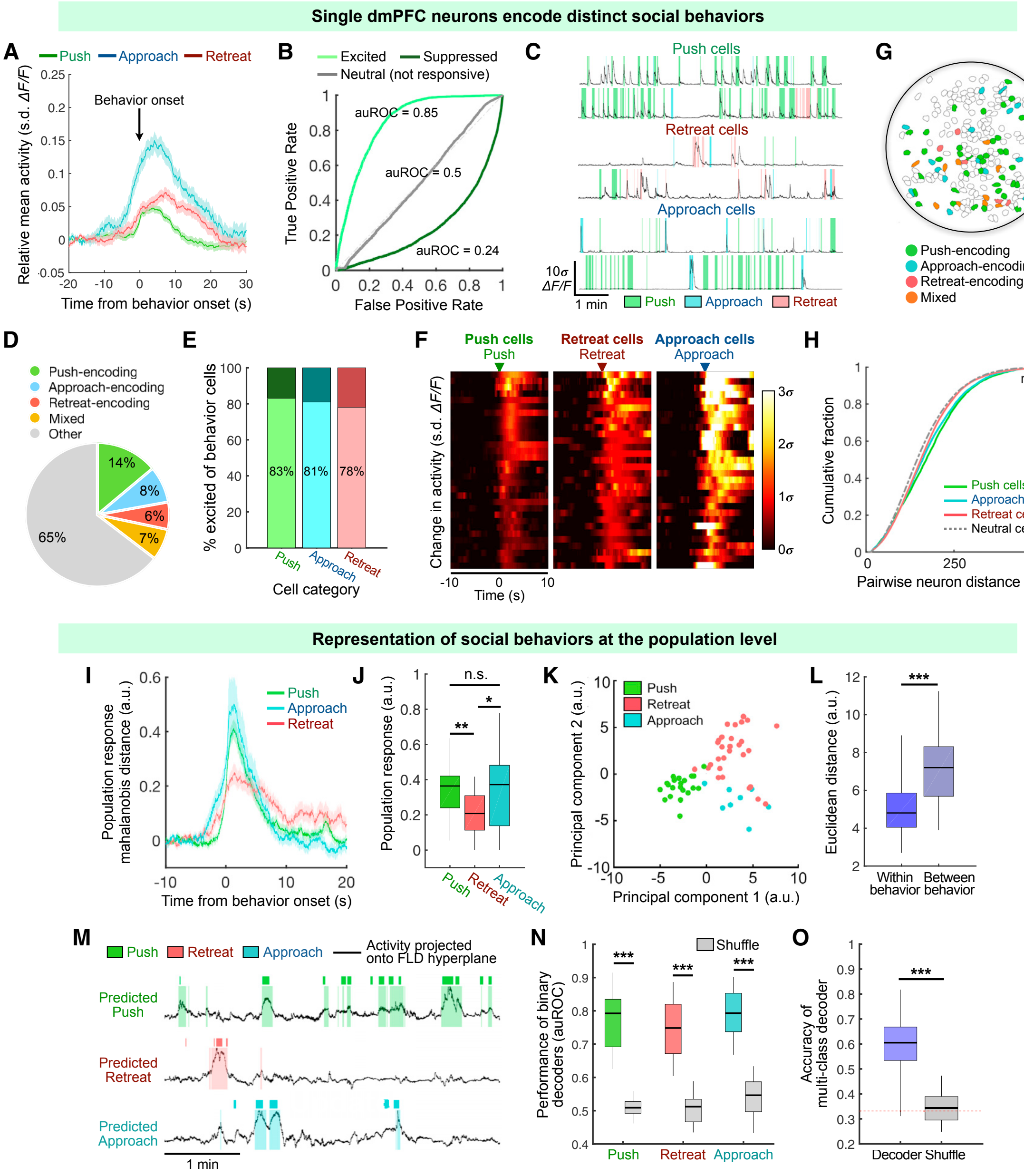


Figure 5. dmPFC Neurons Encode Social Behaviors during Competitive Interaction

(A) Mean trial-averaged response of dmPFC neurons (normalized to the 15 s preceding behavior) centered at onset of social behaviors.

(B) ROC curves from example neurons for push behavior.

(C) Examples of single cells that selectively encode different behaviors.

(D) Distribution of behavior-encoding neurons.

(E) Distribution of excited and suppressed cells within each behavior category.

(F) Trial-averaged responses of example behavior cells.

(G) Example field of view showing spatial distribution of behavior cells.

(H) Cumulative fraction of pairwise distances among different subsets of behavior cells, compared with neutral cells (Kolmogorov-Smirnov test).

(I) Population responses during behavior events (Mahalanobis distance between trial population vectors and baseline activity), averaged across sessions (STAR Methods).

(J) Population responses (as in I) during different behaviors over 3 s following behavior onset.

(K) Principal component (PC) separation of behavior-evoked population responses from one session; each dot is the mean response from one behavior bout.

(L) Euclidean distance between PC-projected population vectors within or between behavior types, averaged within each session.

(M) FLD decoders trained to predict different behaviors from rest using population activity. Plots: projections of population activity onto the linear discriminant; dark patches: annotated behavior; light patches: frame-by-frame predictions of example classifiers.

(legend continued on next page)

distance between different behaviors was significantly larger than within-behavior distances (Figure 5L), indicating that the separation of responses is not due to trial variability but reflects unique patterns of activation that distinguish social behaviors.

Finally, to explore the robustness of behavior representations, we constructed decoders using Fisher's linear discriminant (FLD) to predict the occurrence of behavior events based on population activity. Each behavior could be predicted by decoders (Figures 5M and S6E–S6G), which significantly outperformed models constructed using randomized training data (Figure 5N). Moreover, multi-class decoders trained to predict specific behaviors among push, retreat, and approach achieved significantly higher performance than chance (Figure 5O), again indicating that neural representations are distinct and stable. Taken together, these results show that dmPFC neurons encode social behaviors at both the single-cell and population levels.

Interbrain Activity Correlations Depend on Cells Encoding Social Behavior

To determine how interbrain coupling depends on activity in individual cells, we next examined whether interbrain correlations arise from uniform dmPFC activation or specific subsets of cells (e.g., behavior cells). Removal of behavior cells resulted in a marked reduction in the activity correlation across animals (Figure S7A), and this was driven specifically by behavior-excited cells (Figure 6A), as removal of behavior-suppressed cells did not affect interbrain correlations (Figure S7B). Moreover, interbrain correlations were equally disrupted upon removal of behavior cells in only one animal, indicating that brain coupling requires encoding of social information in both animals simultaneously. In contrast, removing neutral cells did not reduce activity correlations. This was not due to neutral cells being unresponsive, as their overall activity was as high as that of behavior cells (Figure S5C). Instead, this suggests that correlated brain activity depends on subsets of cells encoding social information, rather than uniformly distributed neural dynamics.

Following this, we next examined correlations between specific subpopulations of behavior-encoding cells. Indeed, certain categories of behavior cells exhibited elevated interbrain correlations (Figures 6B–6D). In particular, push-versus-retreat subpopulations were more highly correlated across animals than were neutral cells, consistent with our observation that these behaviors are sometimes coupled (Figure 3H). Interestingly, the synchronization of push and retreat cells was unidirectional across dyads, such that only push cells in dominants, but not in subordinates, were more correlated with retreat cells in the opponent. This suggests that interbrain correlations not only depend on specific subsets of cells, but that neurons encoding specific behavior interactions contribute preferentially to brain coupling.

Interbrain Activity Correlations Arise from Single-Cell Dynamics

To gain more insight into how interbrain correlations emerge from single dmPFC neurons, we constructed GLMs to express the overall dmPFC activity in one animal as a function of single cells in the interacting opponent (Figure 6E). These GLMs performed significantly better than chance (Figure 6F), suggesting that a weighted combination of individual cell activities in one animal could provide a good model of overall activity in the opponent. Moreover, behavior cells had significantly higher weight contributions in the models than neutral cells (Figure 6G), consistent with our results that interbrain correlations depend on behavior cells.

We next constructed GLMs using single cells from one animal to model subsets of behavior cells in the other (Figure S7C) and found that these models performed significantly better than models of neutral cells (Figures 6H–6J and S7D). Examination of subpopulation models in dominants and subordinates revealed further asymmetries that mirrored unidirectional behavior interactions displayed by the dyads (Figures 6K–6M): while the push-encoding population in dominants was best explained by subordinate retreat cells, the retreat-encoding population in subordinates was better modeled by dominant push cells. This further suggests that interbrain correlations in dmPFC arise from unique subpopulations in each animal that preserve individual differences in behavior.

Last, we investigated whether interbrain correlations were related to correlations between single pairs of cells across animals. Interacting animals contained more highly correlated cell pairs than expected by chance (Figures S7E and S7F), and the fraction of highly correlated cell pairs in each dyad was itself correlated with the degree of overall brain coupling between them (Figure 6N), supporting the notion that correlated activity at the population level arises from subsets of single cells. Moreover, behavior cells were enriched among more highly correlated cell pairs (Figure 6O). In particular, in dominants, a larger fraction of push and approach cells were highly correlated with cells in subordinates, possibly reflecting a greater influence of behaviors of dominants on opponent responses (Figure 6P).

Taken as a whole, these results show that interbrain correlations in the dmPFC arise from specific subsets of cells encoding distinct behaviors in both animals and reflect ensemble correlations that extend to the single-cell level.

Interbrain Correlations Depend on Cells Encoding Behaviors of the Social Partner

The observation that interbrain coupling depends on subsets of behavior-encoding neurons raises the possibility that correlated activity could be completely explained by activity in these cells. However, our findings that (1) the degree of activity correlation consistently exceeds behavior correlations (Figure 4D); (2) activity correlations cannot be explained simply by concurrent or

(N) Performance of FLD decoders exemplified in (M), compared with models constructed using shuffled class labels.

(O) Performance of 3-way multi-class FLD decoders trained to distinguish between push, approach, and retreat behavior. Red line: expected chance level in the three-way decoder.

***$p < 0.001$, **$p < 0.01$, $p > 0.05$, n.s. (A and I) Mean ± SEM. (A, C, and F) $\Delta F/F$ calcium traces are presented in units of SD.

See also Figures S5 and S6.

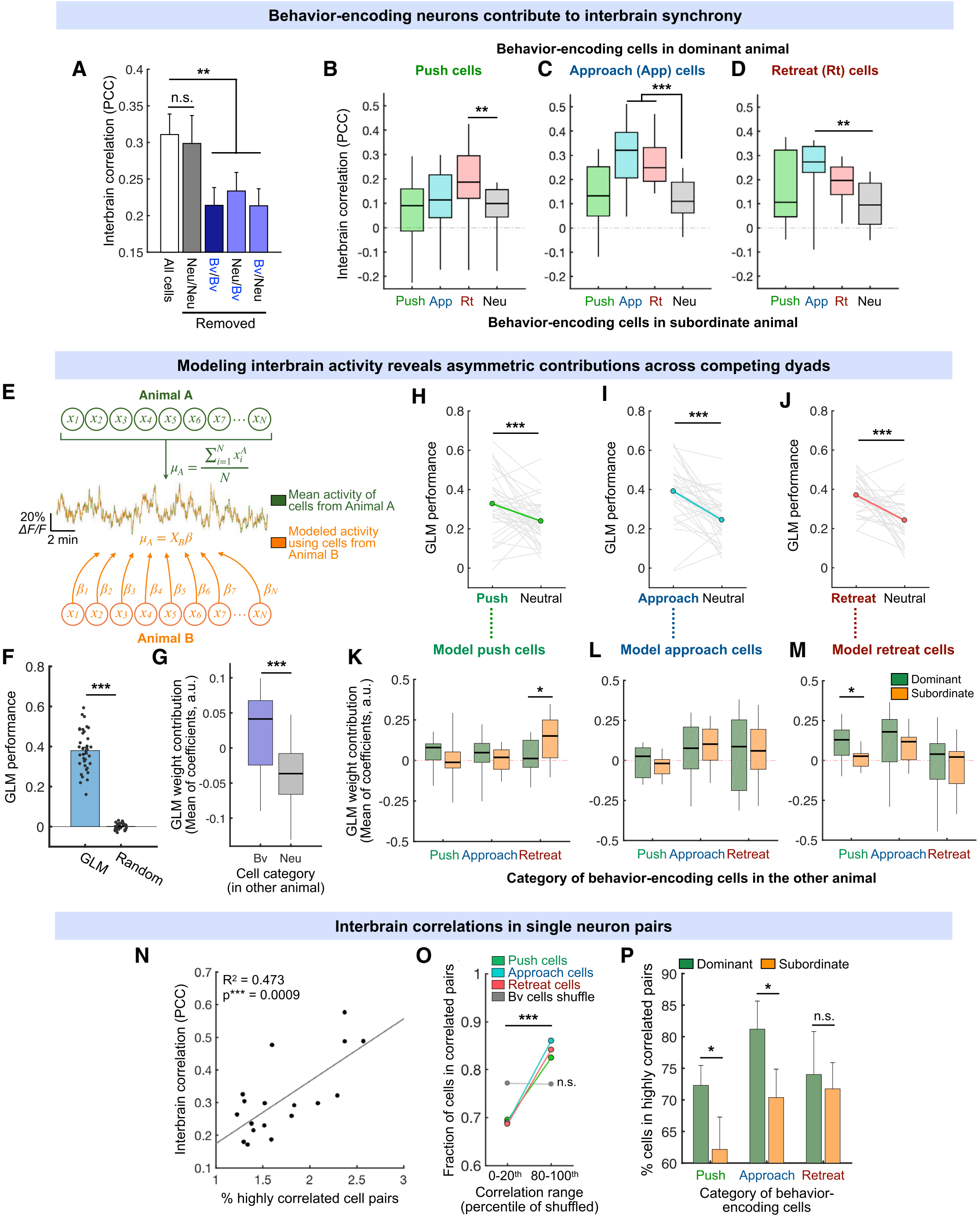


Figure 6. Interbrain Correlations Depend on Neurons Encoding One's Own Social Behavior

(A) Interbrain activity correlations after removal of behavior-excited (Bv) or neutral (Neu) cells from both animals.

(B–D) Interbrain correlations between the mean activity of subsets of push- (B), approach- (C), and retreat- (D) excited cells (the top 15 cells based on area under the ROC curve [auROC] values).

(E) Schematic of models of interbrain activity across animals. The mean activity of all neurons in one animal (top) is modeled as a function of single-cell activities in the interacting partner (bottom) using a GLM.

(F) Performance (cross-validated PCC) of GLMs to predict activity in one animal using single-cell activities from the other, compared with that of models using randomly permuted controls.

(G) Weight contributions of behavior (Bv) and neutral (Neu) cells in GLMs of overall activity in (F), computed as the average of *Z*-scored coefficients fit to Bv or Neu cells in each model.

(legend continued on next page)

coordinated behavior bouts (Figure 4M); and (3) activity correlations persist when only one animal is behaving (Figure 4F) raise the alternative possibility that other information in the circuit also contributes to interbrain coupling. In particular, one hypothesis is that some correlated activity arises from subsets of dmPFC neurons that encode the behavior of the interacting partner.

To examine this hypothesis, we first asked whether any dmPFC neurons contained information about opponent behavior. Using ROC analysis, we identified a fraction of dmPFC neurons that responded specifically during opponent behavior, but not during subject behavior (Figures 7A and 7B), which constituted 13% of all recorded cells. On the other hand, 22% responded only during subject, but not opponent, behavior. We hereafter referred to neurons that only encoded opponent behavior as ‘‘opponent cells’’ and neurons that only encoded subject behavior as ‘‘subject cells.’’ Of the opponent cells, the majority (93%) responded selectively to single categories of behavior (Figures 7C and S8A), with response characteristics that were comparable to those of subject cells (Figures S8B and S8C). Subject and opponent cells were spatially intermixed within the population (Figures 7D and 7E). Interestingly, we also identified a comparable fraction of cells that encoded behavior of the interacting partner during free social interaction in the open area (Figures S8D and S8E), suggesting that behavior of social partners is encoded in multiple social contexts.

Opponent cells showed responses to specific opponent behaviors but did not appear to respond during the subject’s own behavior (Figures 7F and 7G). To confirm that these cells were selectively active during opponent behavior, we compared their mean activity during opponent push, retreat, or approach with activity during subject behaviors (Figures 7H–7J). Opponent cell activity during opponent behaviors (when the subject is not behaving or moving; Figures S3F and S3G) was significantly higher than baseline, while activity during subject behavior was not, confirming that opponent cells selectively encode opponent behavior.

To further explore the population encoding of opponent behavior, we constructed decoders to classify the identities of subject versus opponent behaviors (Figure 7K; STAR Methods) and found that discrimination was significantly higher than chance levels (Figure 7L), indicating that neural responses during subject and opponent behavior form distinct population-level representations.

To test whether opponent cells also contribute to brain coupling, we next examined the effect of removing subsets of opponent cells on interbrain correlations. As with removal of subject cells (Figure 6A), removal of opponent cells, even in only one animal, markedly decreased correlated activity (Figure S8F), an effect that was driven specifically by opponent-excited cells (Figures 7M and S8G). Conversely, examining interbrain correlations only among subject and opponent cells, we found that they displayed even higher correlations than the whole population, and that replacing these with neutral cells in either animal drastically reduced interbrain correlations (Figure 7N). Interestingly, we also observed that removing opponent cells had a stronger effect (~63% more) on reducing interbrain correlations than subject cells, suggesting that they contribute relatively more, cell for cell, to synchronized activity (Figures 7O and 7P).

Taken together, these results indicate that correlated brain activity depends not only on subject cells encoding one’s own behavior, but also on a separate subset of neurons in each animal that encode the behavior of the interacting partner (Figure 7Q). As each brain represents a common behavior repertoire consisting of both animals’ behavior, overall neural activity becomes synchronized across dyads. This offers an explanation for why interbrain synchrony cannot be fully explained by coordinated rest or concurrent behavior, and why it can be observed even when only one animal behaves.

Dominant Animals Exert a Greater Influence on Interbrain Correlations Than Subordinates

Next, to explore whether cells in dominants and subordinates encode subject and opponent information differently, we constructed GLMs to model the activity of each neuron as a function of the behaviors of both animals and their positions in the tube (Figures 8A and S8I). Overall, ~30% of all cells in both dominants and subordinates were well modeled (Figure S8J), and the majority of these were significantly fit by only subject behavior, opponent behavior, or a combination of both (Figure 8B). Moreover, a substantial fraction were fit with significant coefficients to specific opponent behaviors (Figures 8C and S8K), again indicating that activity in some dmPFC neurons is selectively modulated by opponent behavior.

Intriguingly, models of cells in dominants placed higher weight on the subject’s own behavior, whereas opponent behaviors had a stronger weight contribution to cells in subordinates (Figures 8D and S8L). This indicates that, while cells in dominants respond more to subject behaviors compared to cells in subordinates, cells in subordinates respond more to opponent behaviors compared to cells in dominants. This possibly reflects stronger engagement of attention in subordinates toward dominant animal behavior.

These observations led us to hypothesize that dmPFC neurons might exhibit stronger interbrain correlations when dominants

(H–J) Performance of GLMs modeling the mean activity of subsets of push (H), approach (I), and retreat (J) cells, as in (B)–(D), compared with that of GLMs modeling the mean of neutral cells.

(K–M) Weight contributions of behavior (push-, approach-, or retreat-excited) cells fit to models of push (K), approach (L), and retreat (M) cells in (H)–(J), computed as the average of *Z*-scored coefficients for each cell type.

(N) Correlation between the percentage of highly correlated single-cell pairs (>99th percentile of random distribution, see Figures S7E and S7F) and the interbrain activity correlation across pairs.

(O) Fraction of behavior cells that belong to an interbrain cell pair with low (bottom 20% of random distribution) or high (top 20% of random distribution) correlations.

(P) Fraction of behavior cells in highly correlated (>99th percentile of random distribution) cell pairs in dominants and subordinates.

***p < 0.001, **p < 0.01, p > 0.05, n.s. (A, F, and P) Mean ± SEM.

See also Figure S7.

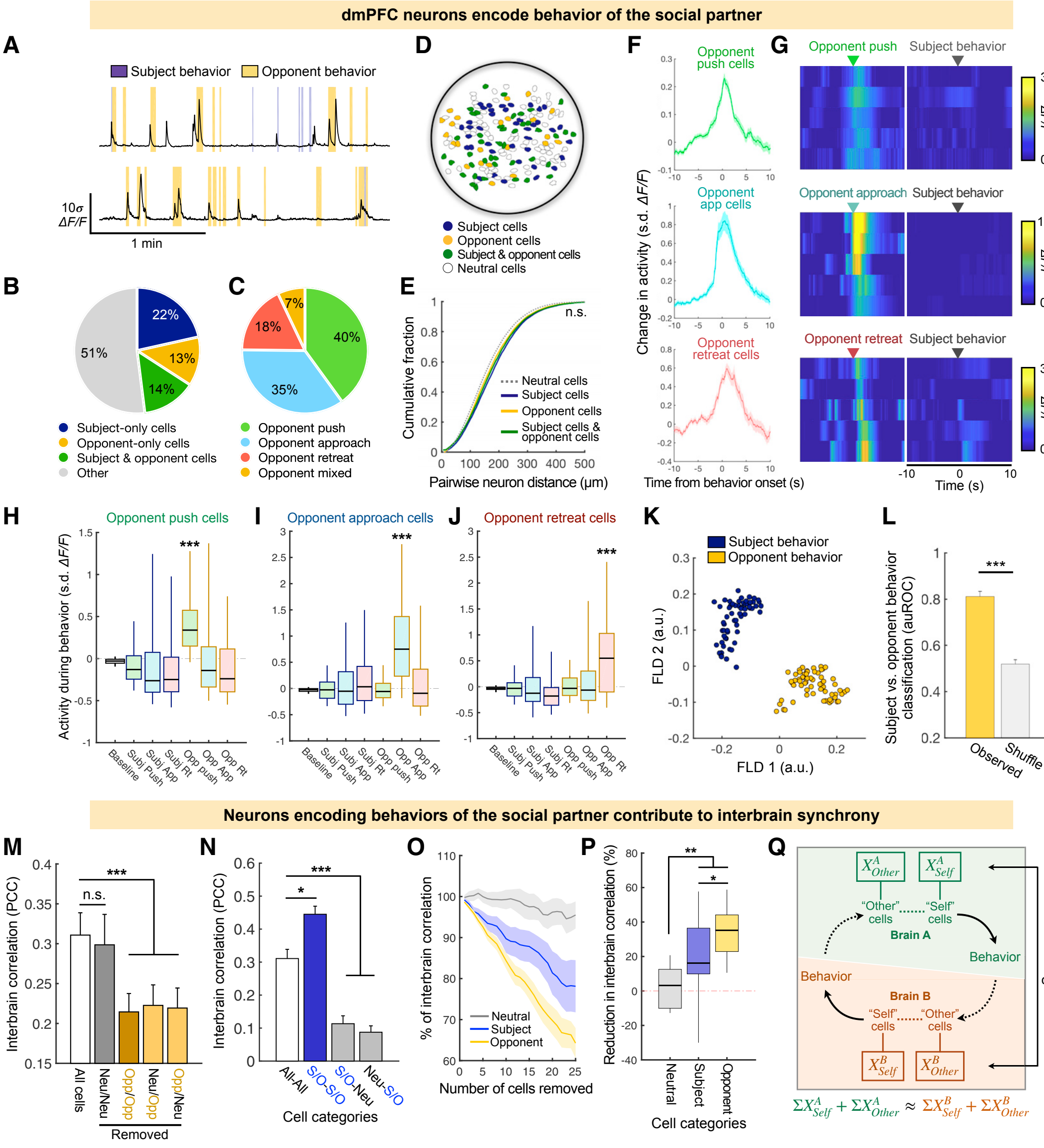


Figure 7. Neurons Encoding Behavior of the Social Partner Contribute to Interbrain Correlations

(A) Example traces from dmPFC neurons that respond during opponent behavior.

(B) Fraction of neurons that are significantly responsive during subject, opponent, or both types of behavior based on ROC analysis.

(C) Distribution of opponent-encoding neurons that selectively respond during specific behaviors.

(D) Example field of view showing the spatial distribution of subject and opponent cells.

(E) Cumulative fraction of pairwise distances among different subsets of cells, compared with neutral cells (Kolmogorov-Smirnov test).

(F) Trial-averaged responses of behavior-selective opponent cells.

(G) Trial-averaged responses of example opponent push-, approach-, and retreat-excited neurons during opponent or subject behavior.

(H–J) Mean activity of opponent push- (H), approach- (I), and retreat (J)-excited cells during each type of subject or opponent behavior. Behavior bouts that overlapped across subject and opponent were excluded to ensure that activity during opponent behavior was not contaminated by subject behavior. During opponent behaviors used for this analysis, the subject animal did not exhibit any behavior or positional change (see Figures S3F and S3G).

(K) Population responses during subject and opponent behavior (from a cross-validation test set) projected onto the first two FLD dimensions.

(L) Performance of FLD decoders to distinguish between subject and opponent behavior based on population activity.

(M) Interbrain activity correlations after removal of opponent-excited (Opp) or neutral (Neu) cells from both animals.

(N) Interbrain activity correlations between subsets of subject and opponent (S/O) or neutral (Neu) cells (the top 25 cells based on rank-ordered auROC values; STAR Methods).

(O) Interbrain correlation upon removal of different numbers of subject, opponent, or neutral cells from each animal.

(P) Reduction in interbrain correlation after removing 25 subject, opponent, or neutral cells from each animal, as in (O).

(legend continued on next page)

behave compared to subordinates. To test this, we examined interbrain correlations during epochs when one animal, but not the interacting partner, was behaving. Strikingly, activity correlations were higher during dominant than during subordinate behavior (Figure 8E), suggesting that interbrain correlations are driven more strongly by dominant animals (Figure 8F).

Interbrain Correlations Predict Social Interactions and Dominance Relationships across Dyads

The observation that dominant animals more strongly drive brain coupling suggests a more direct relationship between interbrain correlations and social interaction. To explore this more deeply, we first asked whether interbrain correlations could predict behavior interactions. We constructed time courses of the probability of behavioral response in one animal as a function of time following partner behavior (Figure 8G). Decisions in one animal preceded by highly correlated activity were more likely followed by a behavioral reaction from the opponent. Moreover, the probability of behavioral response following partner behavior was positively correlated with the degree of synchrony preceding the interaction (Figure 8H), suggesting that correlated activity not only arises during social interaction but actually predicts future interactions. As expected, correlations among subsets of subject and opponent cells in each animal also predict future interactions (Figure 8I). However, this relationship was abolished when considering correlations with neutral cells (Figure 8I), again highlighting the dependency of activity synchrony on neurons encoding social information.

Given that the overall dominance relationship between animals is a consequence of individual social interactions, we hypothesized that the degree of activity correlation across a dyad, which predicts their interactions, may reflect their difference in overall dominance levels. Using average tube positions of animals as a dominance metric (i.e., territory gained), we compared interbrain correlations across dyads with their difference in relative dominance. Strikingly, we observed a significant positive correlation across all pairs (Figure 8J). In particular, subsets of neurons encoding social behaviors of self and others significantly predicted differences in dominance behavior, while replacement with neutral cells in either animal abolished this relationship (Figure 8K).

Since brain coupling predicted future social interactions, we also asked whether correlations during only the initial phase of the encounter could predict dominance outcomes. Interestingly, the degree of interbrain correlation in just the first 2 min of each session predicted differences in dominance across the whole session (Figure 8L). Again, this relationship depended critically on behavior cells in both animals (Figures 8M–8O). Despite this, the degree of overall behavior correlation in the first 2 min was unrelated to differences in dominance (Figure S8M), suggesting that activity correlations may be a better predictor of dominance outcomes than behavior itself. Taken as a whole, these results demonstrate that activity correlations predict social interaction on timescales ranging from seconds to minutes, suggesting a functional role for brain coupling as an emergent property of multi-animal systems in coordinating social interactions and facilitating the development of social relationships (Figures S1B and 8P).

DISCUSSION

Interbrain Correlated Neural Activity during Social Interaction

Previous research on interbrain synchrony has illuminated the capacity for neural circuits to coordinate across individuals during social engagement (Liu and Pelowski, 2014; Montague et al., 2002). However, it has been largely unclear how region-wide interbrain correlations arise from activity patterns at the circuit or single-cell levels. Using simultaneous large-scale recordings in interacting animal dyads, we provide conclusive evidence that mice exhibit interbrain correlations of neural activity in the dmPFC that arise from ongoing social interaction. We observed correlated activity in an unconstrained environment, as well as during dominance competitions in the tube test, suggesting that social brain coupling is a general phenomenon present in multiple contexts. Importantly, interbrain correlations could not be simply explained by activity associated with concurrent or coordinated behavior. Rather, the coupling of brain activity likely reflects specific types of meaningful engagement, as well as attentional entrainment across pairs of animals embedded in a larger social context. As interbrain coupling has only previously been observed in humans and non-human primates, this finding strongly suggests generality and conservation of the phenomenon across a wide range of animal species.

Importantly, rather than reflecting uniform changes in the firing patterns of cell populations, we find that activity synchrony depends specifically on subsets of neurons that separately encode behaviors of the subject animal and those of the interacting partner. These cells allow each brain to represent a common repertoire of behavior (i.e., behavior of both interacting animals), such that activity across separate brains becomes synchronized. The existence of opponent-encoding cells in part explains why interbrain synchrony is not simply accounted for by coordinated rest and concurrent behavior and highlights the complexity of mechanisms underlying synchrony that invite deeper investigation at the circuit level.

Encoding of One's Own and the Social Partner's Behavior in dmPFC Neurons

In many social species, including humans, social interactions between individuals are shaped by status relationships and dominance competitions (Williamson et al., 2016). Recent work has begun to investigate the neural mechanisms underlying the expression of dominance behavior (Stagkourakis et al., 2018; Zhou et al., 2018). In our study, we identified a substantial fraction of neurons in the dmPFC that encode distinct social

(Q) Schematic showing that interbrain correlations arise from the collective contributions of neurons encoding subject and opponent behavior in both animals. As these neurons in each brain represent a common behavior repertoire (i.e., behavior of both animals), overall neural activity becomes synchronized across dyads. ***$p < 0.001$, **$p < 0.01$, $p > 0.05$, n.s. (F and L–O) Mean ± SEM. (A, F, and G) $\Delta F/F$ calcium traces are presented in units of SD.
See also Figure S8.

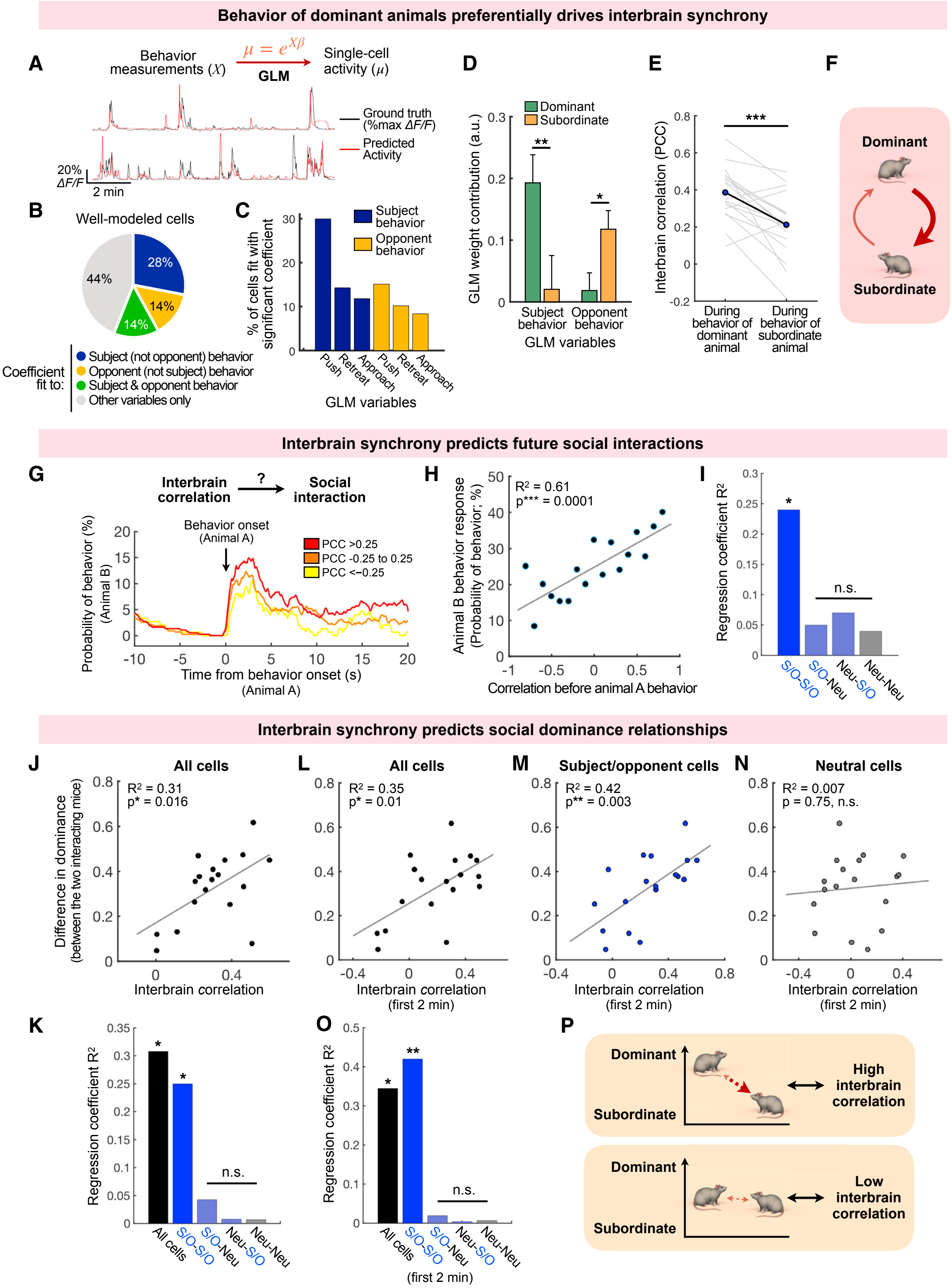


Figure 8. Interbrain Correlations Predict Future Social Interactions and Dominance Relationships

(A) Examples of neurons with activity modeled by GLMs using positions and behavior of both animals.

(B) Distribution of single-neuron GLMs with statistically significant ($p < 0.05$) coefficients fit to subject (but not opponent) behavior, opponent (but not subject) behavior, subject and opponent behavior, or other variables only.

(C) Distribution of cells with significant coefficients for specific subject and opponent behaviors.

(D) Weight contributions (the average of *Z*-scored coefficients) in single-neuron GLMs for subject and opponent behavior in dominants and subordinates.

(legend continued on next page)

dominance behaviors during a competitive encounter. These single-cell responses collectively formed stable representations of push, retreat, and approach behavior, suggesting a role for dmPFC neurons in regulating multiple, and sometimes opposing, behavioral strategies.

In addition to coordinating one's own behavior, social interactions also require animals to anticipate and react to the decisions of their social partners. However, it is not well understood how neural systems represent observed behavior. Studies in humans and non-human primates report that prefrontal, motor, and parietal regions can respond to actions displayed by other individuals (Hardwick et al., 2018; Ogawa and Inui, 2011; Rozzi and Fogassi, 2017; Tseng et al., 2018). Yet many of these studies were done in the context of passive and unidirectional behavioral observation. It is largely unclear how representations of self and others' behavior arise during dynamic interactions where animals must simultaneously observe and respond within seconds. We find that a fraction of dmPFC neurons in mice encode specific behaviors of the interacting partner and collectively form a neural response pattern that distinguishes opponent and subject behavior. The presence of these neurons in the rodent dmPFC suggests conservation of function across diverse species and sets the groundwork for deeper investigation using a genetically tractable animal model.

We also explored whether encoding of partner's behavior is shaped by dominance status. Interestingly, while subject behavior was more strongly encoded in dominants than in subordinates, opponent behavior was more robustly encoded in subordinates than in dominants, suggesting an asymmetry in the computational structure of the dmPFC circuit based on social status. Moreover, synchrony was consistently higher during dominant animals' behavior than during subordinate animals' behavior. These suggest that during competitive interactions, subordinates may be more attentive to dominants. Indeed, in primates, subordinates pay more attention to the actions and gazes of dominant individuals (Deaner et al., 2005; Klein et al., 2009). In rodents, this feature of directed social attention could be instantiated in the activity of dmPFC neurons.

Animals also have the capacity to encode other information about conspecifics, such as their physical location or emotional state (Allsop et al., 2018; Danjo et al., 2018; Panksepp and Panksepp, 2013). How these processes are related to the encoding of volitional behavior of others is unclear and remains an exciting topic for future study.

Interbrain Correlations Predict Social Interactions and Dominance Relationships

Beyond providing a neural basis for how interbrain synchrony arises from individual cells, our study also functionally links it to the coordination of social interactions—stronger interbrain correlations across dyads predict future social interaction. While interbrain coupling originates from activities in individual brains, it represents a state of multi-individual systems that operates at the level of the system itself and is not accessible to each brain to directly influence one's own decisions. Instead, this state reflects one or several underlying neural processes within each brain that operate to shape animal behavior. Given the role of opponent-encoding neurons in interbrain synchrony, correlated activity may in part reflect attentional engagement between animals, effectively coupling their decisions and increasing their behavioral reciprocity. As interbrain coupling both arises from and predicts dyadic behavior, the behavioral interaction and its interbrain neural correlate may form a bidirectional feedback loop that serves to facilitate and sustain ongoing interaction (Figure S1B).

In addition to our observation that dominants drive stronger responses from subordinates, we also found that the degree of interbrain correlation across each pair predicted dominance relationships, whereas correlations between their behavior could not. This echoes previous reports in humans that brain coupling can predict leader-follower relationships, even before leadership roles are manifested (Jiang et al., 2015; Konvalinka et al., 2014; Sänger et al., 2013). Our results suggest that synchrony across individuals with unequal status relationships depends on circuitry that encodes actions of social partners and, in such contexts, may reflect the directed engagement of ''followers'' toward more dominant individuals leading an interaction.

Collectively, our results shed new light on the neural basis and functional role of interbrain synchrony in coordinating social interactions. More importantly, they set the groundwork for a more incisive investigation of the emergent neural properties of multi-individual systems, which may yet reveal a richer and deeper understanding of the social brain as it is embedded in a truly social world.

(E) Interbrain correlations during behaviors of dominants versus subordinates.

(F) Schematic showing greater influence on interbrain synchrony by dominant animals.

(G) Time courses showing the probability of behavior in one animal as a function of time following behavior onset in the interacting partner, color coded based on the interbrain correlation over the preceding 30 s.

(H) Correlation between the interbrain activity PCC preceding behavior in one animal and the response probability of the interacting partner.

(I) Regression coefficients (R^2) for the linear relationship shown in (H) using subsets of neurons (S/O, subject and opponent cells; Neu, neutral cells).

(J) Correlation between the interbrain activity PCC across pairs and the differences in their mean tube position.

(K) Regression coefficients (R^2) for the linear relationship shown in (J) using subsets of neurons.

(L–N) Correlation between interbrain activity PCC during the first 2 min of interaction and overall difference in tube position over the session using all cells (L), only subject- and opponent-encoding cells (M), or only neutral cells (N).

(O) Regression coefficients (R^2) for the linear relationships between interbrain activity correlations during the first 2 min of interaction and dominance difference using subsets of neurons.

(P) Schematic showing that interbrain coupling is higher when one animal is significantly more dominant than its opponent, and lower when two animals have similar levels of dominance.

***$p < 0.001$, **$p < 0.01$, $p > 0.05$, n.s. (D) Mean ± SEM.

See also Figure S8.

STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - ○ Viral injections and GRIN lens implantations
  - ○ Histology
  - ○ Behavior Assays
  - ○ Analysis of animal behavior
  - ○ Extraction of Calcium Signals
  - ○ Analysis of Single Cell Responses During Behavior
  - ○ Analysis of population dynamics during behavior
  - ○ Behavior decoding based on population activity
  - ○ Generalized linear models of single-neuron and population activity
  - ○ Modeling neural activity using both neural activity and behavioral variables across animals
  - ○ Analysis of interbrain neural activity correlations
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cell.2019.05.022.

ACKNOWLEDGMENTS

We thank D. Buonomano, A. Silva, E. Marcus, M. Basso, S. Masmanidis, and M. Yartsev for critical comments and D. Aharoni and P. Zhao for technical support on microendoscopes. This work was supported in part by NIH grants (T32 NS058280, F31 MH117966, R01 MH105427, R01 MH101198, U54 HD87101, and P50 DA005010), NSF 1700408, an ARCS Fellowship, a Marion Bowen Postdoctoral Grant, a Whitehall Foundation grant, a NARSAD Young Investigator grant, a Sloan Research Fellowship, a Searle Scholars Award, a Klingenstein-Simons Fellowship, a Brain Research Foundation grant, and a Packard Foundation Fellowship.

AUTHOR CONTRIBUTIONS

L.K., S.H., and W.H. designed the study. L.K. and S.H. performed experiments. L.K. performed most of the computational analysis. J.W. provided technical assistance. K.G. and Y.E.W. performed animal tracking. P.G. provided support on microendoscopes. L.K. and W.H. wrote the manuscript with input from S.H. and Y.E.W. W.H. supervised the entire study.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 3, 2018
Revised: February 11, 2019
Accepted: May 9, 2019
Published: June 20, 2019

REFERENCES

Adolphs, R. (2010). Conceptual challenges and directions for social neuroscience. Neuron *65*, 752–767.

Allsop, S.A., Wichmann, R., Mills, F., Burgos-Robles, A., Chang, C.J., Felix-Ortiz, A.C., Vienne, A., Beyeler, A., Izadmehr, E.M., Glober, G., et al. (2018). Corticoamygdala Transfer of Socially Derived Information Gates Observational Learning. Cell *173*, 1329–1342.

Babiloni, F., Cincotti, F., Mattia, D., Mattiocco, M., De Vico Fallani, F., Tocci, A., Bianchi, L., Marciani, M.G., and Astolfi, L. (2006). Hypermethods for EEG hyperscanning. Conf. Proc. IEEE Eng. Med. Biol. Soc. *1*, 3666–3669.

Chen, P., and Hong, W. (2018). Neural Circuit Mechanisms of Social Behavior. Neuron *98*, 16–30.

Cooper, M.A., Clinard, C.T., and Morrison, K.E. (2015). Neurobiological mechanisms supporting experience-dependent resistance to social stress. Neuroscience *291*, 1–14.

Cunningham, J.P., and Yu, B.M. (2014). Dimensionality reduction for large-scale neural recordings. Nat. Neurosci. *17*, 1500–1509.

Danjo, T., Toyoizumi, T., and Fujisawa, S. (2018). Spatial representations of self and other in the hippocampus. Science *359*, 213–218.

Deaner, R.O., Khera, A.V., and Platt, M.L. (2005). Monkeys pay per view: adaptive valuation of social images by rhesus macaques. Curr. Biol. *15*, 543–548.

Drews, C. (1993). The Concept and Definition of Dominance in Animal Behaviour. Behaviour *125*, 283–313.

Driscoll, L.N., Pettit, N.L., Minderer, M., Chettih, S.N., and Harvey, C.D. (2017). Dynamic Reorganization of Neuronal Activity Patterns in Parietal Cortex. Cell *170*, 986–999.

Franklin, T.B., Silva, B.A., Perova, Z., Marrone, L., Masferrer, M.E., 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.

Hardwick, R.M., Caspers, S., Eickhoff, S.B., and Swinnen, S.P. (2018). Neural correlates of action: Comparing meta-analyses of imagery, observation, and execution. Neurosci. Biobehav. Rev. *94*, 31–44.

Jiang, J., Chen, C., Dai, B., Shi, G., Ding, G., Liu, L., and Lu, C. (2015). Leader emergence through interpersonal neural synchronization. Proc. Natl. Acad. Sci. USA *112*, 4274–4279.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C.F., Quartz, S.R., and Montague, P.R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. Science *308*, 78–83.

Klein, J.T., Shepherd, S.V., and Platt, M.L. (2009). Social attention and the brain. Curr. Biol. *19*, R958–R962.

Konvalinka, I., Bauer, M., Stahlhut, C., Hansen, L.K., Roepstorff, A., and Frith, C.D. (2014). Frontal alpha oscillations distinguish leaders from followers: multivariate decoding of mutually interacting brains. Neuroimage *94*, 79–88.

Li, Y., Mathis, A., Grewe, B.F., Osterhout, J.A., Ahanonu, B., Schnitzer, M.J., Murthy, V.N., and Dulac, C. (2017). Neuronal Representation of Social Information in the Medial Amygdala of Awake Behaving Mice. Cell *171*, 1176–1190.

Liang, B., Zhang, L., Barbera, G., Fang, W., Zhang, J., Chen, X., Chen, R., Li, Y., and Lin, D.-T. (2018). Distinct and Dynamic ON and OFF Neural Ensembles in the Prefrontal Cortex Code Social Exploration. Neuron *100*, 700–714.

Liu, T., and Pelowski, M. (2014). A new research trend in social neuroscience: Towards an interactive-brain neuroscience. PsyCh J. *3*, 177–188.

Montague, P.R., Berns, G.S., Cohen, J.D., McClure, S.M., Pagnoni, G., Dhamala, M., Wiest, M.C., Karpov, I., King, R.D., Apple, N., and Fisher, R.E. (2002). Hyperscanning: simultaneous fMRI during linked social interactions. Neuroimage *16*, 1159–1164.

Mukamel, E.A., Nimmerjahn, A., and Schnitzer, M.J. (2009). Automated analysis of cellular signals from large-scale calcium imaging data. Neuron *63*, 747–760.

Murugan, M., Jang, H.J., Park, M., Miller, E.M., Cox, J., Taliaferro, J.P., Parker, N.F., 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.

Ochsner, K.N., and Lieberman, M.D. (2001). The emergence of social cognitive neuroscience. Am. Psychol. *56*, 717–734.

Ogawa, K., and Inui, T. (2011). Neural representation of observed actions in the parietal and premotor cortex. Neuroimage *56*, 728–735.

Panksepp, J., and Panksepp, J.B. (2013). Toward a cross-species understanding of empathy. Trends Neurosci. *36*, 489–496.

Pnevmatikakis, E.A., and Giovannucci, A. (2017). NoRMCorre: An online algorithm for piecewise rigid motion correction of calcium imaging data. J. Neurosci. Methods *291*, 83–94.

Pouget, A., Dayan, P., and Zemel, R. (2000). Information processing with population codes. Nat. Rev. Neurosci. *1*, 125–132.

Redmon, J., and Farhadi, A. (2016). YOLO9000: Better, Faster, Stronger. arXiv, arXiv:1612.08242. https://arxiv.org/abs/1612.08242.

Remedios, R., Kennedy, A., Zelikowsky, M., Grewe, B.F., Schnitzer, M.J., and Anderson, D.J. (2017). Social behaviour shapes hypothalamic neural ensemble representations of conspecific sex. Nature *550*, 388–392.

Rilling, J.K., and Sanfey, A.G. (2011). The neuroscience of social decision-making. Annu. Rev. Psychol. *62*, 23–48.

Rozzi, S., and Fogassi, L. (2017). Neural Coding for Action Execution and Action Observation in the Prefrontal Cortex and Its Role in the Organization of Socially Driven Behavior. Front. Neurosci. *11*, 492.

Runyan, C.A., Piasini, E., Panzeri, S., and Harvey, C.D. (2017). Distinct timescales of population coding across cortex. Nature *548*, 92–96.

Sanfey, A.G. (2007). Social Decision-Making: Insights from Game Theory and Neuroscience. Science *318*, 598–602.

Sänger, J., Müller, V., and Lindenberger, U. (2013). Directionality in hyperbrain networks discriminates between leaders and followers in guitar duets. Front. Hum. Neurosci. *7*, 234.

Sapolsky, R.M. (2004). Social Status and Health in Humans and Other Animals. Annu. Rev. Anthropol. *33*, 393–418.

Sapolsky, R.M. (2005). The Influence of Social Hierarchy on Primate Health. Science *308*, 648–652.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., and Vogeley, K. (2013). Toward a second-person neuroscience. Behav. Brain Sci. *36*, 393–414.

Stagkourakis, S., Spigolon, G., Williams, P., Protzmann, J., Fisone, G., and Broberger, C. (2018). A neural network for intermale aggression to establish social hierarchy. Nat. Neurosci. *21*, 834–842.

Tseng, P.-H., Rajangam, S., Lehew, G., Lebedev, M.A., and Nicolelis, M.A.L. (2018). Interbrain cortical synchronization encodes multiple aspects of social interactions in monkey pairs. Sci. Rep. *8*, 4699.

Utevsky, A.V., and Platt, M.L. (2014). Status and the brain. PLoS Biol. *12*, e1001941.

Wang, F., Zhu, J., Zhu, H., Zhang, Q., Lin, Z., and Hu, H. (2011). Bidirectional Control of Social Hierarchy by Synaptic Efficacy in Medial Prefrontal Cortex. Science *334*, 693–697.

Wang, F., Kessels, H.W., Hu, H., Schjelderup-Ebbe, T., Wilson, E.O., Sapolsky, R.M., Hand, J.L., and Lindzey, G. (2014). The mouse that roared: neural mechanisms of social hierarchy. Trends Neurosci. *37*, 674–682.

Warden, M.R., Selimbeyoglu, A., Mirzabekov, J.J., Lo, M., Thompson, K.R., Kim, S.-Y., Adhikari, A., Tye, K.M., Frank, L.M., and Deisseroth, K. (2012). A prefrontal cortex-brainstem neuronal projection that controls response to behavioural challenge. Nature *492*, 428–432.

Williamson, C.M., Lee, W., and Curley, J.P. (2016). Temporal dynamics of social hierarchy formation and maintenance in male mice. Anim. Behav. *115*, 259–272.

Zhou, T., Zhu, H., Fan, Z., Wang, F., Chen, Y., Liang, H., Yang, Z., Zhang, L., Lin, L., Zhan, Y., et al. (2017). History of winning remodels thalamo-PFC circuit to reinforce social dominance. Science *357*, 162–168.

Zhou, T., Sandi, C., and Hu, H. (2018). Advances in understanding neural mechanisms of social dominance. Curr. Opin. Neurobiol. *49*, 99–107.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
AAV1-Syn-GCaMP6f-WPRE-SV40	Penn vector core	Cat# 37087
Experimental Models: Organisms/Strains		
Mouse: C57BL/6J	Jackson Laboratories	Stock 000664, RRID: IMSR_JAX:000664
Software and Algorithms		
MATLAB	Mathworks	https://www.mathworks.com/products/matlab.html
ImageJ	NIH	https://imagej.nih.gov/ij/index.html; RRID: SCR_003070
YOLOv2 (You Only Look Once) software	Redmon and Farhadi, 2016	https://pjreddie.com/darknet/yolov2/
Miniscope Controller	UCLA Miniscope	https://github.com/daharoni/Miniscope_DAQ_Software
NoRMCorre	Pnevmatikakis and Giovannucci, 2017	https://github.com/flatironinstitute/NoRMCorre
CellSort	Mukamel et al., 2009	https://github.com/mukamel-lab/CellSort
Other		
Microendoscope	UCLA Miniscope	http://miniscope.org
Nanoinjector	World Precision Instruments	Cat# Nanoliter 2000
Superfrost Plus slides	Fisher Scientific	Cat# 22-037-246

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, Weizhe Hong (whong@ucla.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All experiments were carried out in accordance with the NIH guidelines and approved by the UCLA institutional animal care and use committee (IACUC). All subject mice were male C57BL6/J mice ordered from Jackson Laboratories at 8-10 weeks of age and 25-30 g of weight. Mice were maintained in a 12 h:12h light/dark cycle (lighted hours: 10:00 pm – 10:00 am) with food and water *ad libitum*. All mice were individually housed for three weeks prior to imaging and behavior experiments. All experiments were performed during the dark cycle of the animals.

METHOD DETAILS

Viral injections and GRIN lens implantations

For all surgical procedures, mice were anaesthetized with 1.0 to 2.0% isoflurane. We bilaterally injected 300 nL (on each side) of AAV1.Syn.GCaMP6f.WPRE.SV40 virus (titer: 4.65 × 10^{13} GC per ml, Penn Vector Core) at 30 nL min^{-1} into the dorsomedial prefrontal cortex (dmPFC; also prelimbic cortex, PL) using the stereotactic coordinates (AP: +2.0 mm, ML: ± 0.3mm, DV: −1.8mm to bregma skull surface). 30 minutes after injection, a 1.9mm diameter circular craniotomy was centered at the coordinates (AP: +2.0 mm, ML: 0.0 mm), and the GRIN lens (Edmund Optics; 1.8mm) was implanted above the injection site at a depth of −1.6mm ventral to the bregma skull surface and secured to the skull using super glue and dental cement. Mice were given one subcutaneous injection of Ketoprofen (4mg/kg) on the same day of surgery and Ibuprofen in drinking water (30mg/kg) starting on surgery day for 4 days. Mice were individually housed after surgery for two weeks. Then, the microscope together with a plastic baseplate were placed on top of the lens. We adjusted the position of the microscope until the cells and blood vessels appeared sharp in the focal plane and secured this position using dental cement. Left and right dmPFC were counterbalanced when choosing the field of view. The subjects included two mice that received a unilateral viral injection and were implanted with a 1 mm GRIN lens (Inscopix) above the right dmPFC. All mice were handled and habituated for at least 4 days before experiments. We did not observe any alterations in self-directed or social behavior in implanted animals.

Histology

Three weeks after imaging experiments, mice were transcardially perfused with 4% paraformaldehyde (PFA), followed by 24 h post-fixation in the same solution. 60-μm coronal sections were obtained using a cryostat. Finally, sections were stained with DAPI (1:5,000 dilution) and mounted on slides. Images were acquired using a Nikon A1 confocal microscope to confirm the position of lens implantation and GCaMP6f expression.

Behavior Assays

Free social interaction in the open arena

Two novel male mice were simultaneously placed in an open arena (32 × 20 cm) which allows for free social interaction. During each imaging session (10-15 minutes), calcium fluorescence videos from both animals and their behavior were simultaneously recorded using microendoscopes and a video camera, respectively. The microendoscopes were connected to a digital acquisition device (DAQ) through a flexible, ultra-light coaxial cable. The long cable length prevented cables from becoming tangled during interaction between animals, ensuring that the social interaction was not affected by the presence of cables. For each pair, the social interaction assay was followed by a 10-minute separation assay (without removing the microscopes) where a solid, opaque board was inserted at the midline of the arena to prevent subjects from engaging in social interaction. All animals were habituated to being exposed to the open arena individually and to wearing the miniature microscope for at least 4 days before experimentation. We imaged from 10 pairs of animals that naturally displayed a high level of mutual social interaction (> 15% of total time) and 9 pairs of animals that displayed a low level of mutual social interaction (< 15% of total time). Here, mutual social interaction is defined as moments when both animals engaged in social behavior. A total of 8 implanted animals were used. Pairs that naturally displayed high levels of social interaction were used in further analyses of neural dynamics in Figures 1L–N, 2, S2C, and S2E. When we recorded from pairs of animals that naturally displayed lower levels of social interactions, relatively lower interbrain correlation was observed (Figure S2F), consistent with the notion that interbrain synchrony depends on ongoing social interaction.

Competitive social interaction in the tube test

Animals were placed in a closed acrylic tube (length 60 cm; circumference 2.5 cm) with a 1.1 cm channel cut lengthwise down the center to allow movement of the head-mounted microscope. The microendoscopes were connected to a DAQ through a flexible, ultra-light coaxial cable. During each imaging session (12-15 minutes), subjects faced a novel male conspecific and were permitted to freely engage the opponent mouse by approaching, pushing, or retreating. Tube tests have previously been implemented using shorter tubes with individual trials lasting only seconds (Zhou et al., 2017). Here, the longer tube (60 cm) allowed us to perform longer sessions in order to permit each animal ample time to exhibit its full range of volitional behavior, and to respond dynamically to its opponent over the course of the encounter. Each session was typically broken into 2-5 trials, and the same pair of animals were manually reset to their respective end of the tube prior to each trial. All animals were habituated to engagement with a (different) novel male conspecific while wearing the miniature microscope for at least 4 days before behavior experiments. 18 pairs of mice were imaged using a total of 13 animals, and all pairs were used in further analyses of neural dynamics.

For simultaneous recording with and without social contact, 10-minute imaging sessions were performed in 13 pairs of mice using 6 animals (from the same cohort that were used in the other tube test experiments), immediately followed (without removing the microscope) by another 10-minute session after introduction of a translucent plastic separator in the center of the tube. Animals were free to move at will but were not in physical contact with one another.

Analysis of animal behavior

For both the open arena and the tube test experiments, behavior videos were recorded with a video camera at 20 frames per second (fps) and manually annotated frame by frame to identify onset and offset times for behavior of both animals. Behavior annotations were converted into a binary vector for each type of behavior that denotes precisely when animals are engaged in behavior ("1" indicates presence of a given behavior, and "0" indicates absence of that behavior). Epochs when animals engaged in no observable behavior or movement were considered to be "rest" epochs. During the "rest" epochs, the animal could observe the interacting partner, but was not actively behaving.

For the open arena experiments, a total of 15 social and non-social behaviors were annotated. Social behaviors included attacking, approaching, chasing, escaping, sniffing, social-grooming, defending, and mounting. Non-social behaviors included running, self-grooming, digging, exploration, rearing, climbing, and nesting. The level of social interaction for each pair was measured using the percentage of total time that both animals were engaged in social behaviors. 19 pairs of animals were used for basic behavior analyses shown in Figures 1B–1D.

For the tube test experiments, the positions of both animals were tracked automatically using a supervised learning algorithm. For position tracking, we employed YOLOv2 (You Only Look Once), a convolutional neural network (CNN) framework optimized for high accuracy object detection (Redmon and Farhadi, 2016). We trained the CNN to detect and report bounding boxes around mice in each frame based on hundreds of example images. Accuracy for the automated tracking algorithm was confirmed by comparing the detected mouse positions with ground truth assessments in random samples of movie frames (> 99% accuracy, Figure S3A). For this analysis, individual scorers were blind to the identities of pairs and mice. Position vectors denoting the coordinates of each mouse were extracted and normalized to the length of the tube to obtain the relative tube position of each animal on a range from 0 (the starting end) to 1 (the opponent's end).

In order to quantitatively assess the relative dominance levels of each animal within each pair, we calculated their average position in the tube over the entire session. Previous reports have associated push and retreat behavior, as well as winning in the tube test, with overall social dominance status among male mice (Wang et al., 2011). Because positional changes in the tube test correspond to gains or losses of territory that result from approach, push, or retreat behavior, each animal's average tube position can be considered as a measure of its overall dominance level within the pair. We confirmed that animals defined in this way (one dominant and subordinate in each pair) had significantly different average tube positions (Figure 3F). To assess whether dominant and subordinate animals exhibited different levels of push, retreat, and approach behavior, we compared dominant and subordinate animals in pairs that displayed large differences in dominance (having a tube position difference greater than 20% of the length of the tube) (Figures 3K–3M). Indeed, pairs with large differences in tube position exhibited significantly different levels of push, retreat, and approach behavior, suggesting that the tube position metric corresponds to meaningful differences in behavioral repertoire that are consistent with previous studies.

Extraction of Calcium Signals

Motion-correction and preprocessing

During behavior experiments, calcium fluorescence videos from both animals were simultaneously recorded using customized miniature microscopes (UCLA miniscope) at 30Hz through custom-written data acquisition software. Raw videos from each imaging session were first processed using a MATLAB implementation of the NoRMCorre algorithm to correct for motion-induced artifacts across frames (Pnevmatikakis and Giovannucci, 2017). In order to normalize image frames prior to cell sorting, (*F-F0)/F0* (*ΔF/F*) was applied to each frame, where *F0* was the de-trended mean image from the entire movie. *ΔF/F* normalized videos were de-noised using an FFT spatial band-pass filter through a custom-written script in ImageJ (U.S. National Institutes of Health), and spatially down-sampled by a factor of 2 prior to ROI identification and cell sorting.

Segmentation and ROI Identification

In order to identify putative cell bodies for extraction of neural signals, we employed an automated ROI detection algorithm that uses principal (PCA) and independent component analysis (ICA) to extract spatial filters based on spatiotemporal correlations among pixels (Mukamel et al., 2009). Independent components were manually inspected to remove components that did not represent cell bodies, and binary thresholding was applied to remove contributions from pixels outside the bounds of putative neurons. Spatial filters were then applied to the *ΔF/F* movie to extract the calcium traces. All traces from recorded cells were manually inspected to ensure quality signals. Specifically, putative neurons that had abnormally shaped cell bodies (abnormally large or small), or that had calcium transients with low signal-to-noise ratio (< 2 standard deviations above the mean) were excluded from further analysis. Less than 5% of all putative neurons were removed based on these criteria. This approach ensured that the cells we included in our analyses had signal that reflected real neural activity and was robust enough for downstream analyses.

For open arena experiments, a total of 7535 (mean ± SEM = 198 ± 5) single neurons were analyzed. For tube test experiments, a total of 6728 (mean ± SEM = 187 ± 10) single neurons were analyzed. Here, a single neuron refers to one calcium trace extracted from an ROI, identified as described above, from one recording session.

Analysis of Single Cell Responses During Behavior

Prior to downstream analysis, all *ΔF/F* calcium traces were z-scored and are presented throughout in units of standard deviation (s.d.) unless otherwise specified. Responses of single neurons during behavior events (push, retreat, and approach) were quantified using an ROC (receiver operating characteristic) analysis, a commonly used approach that has previously been applied to calcium imaging data to characterize neural responses during social investigation (e.g., Li et al., 2017). Upon application of a binary threshold to the *ΔF/F* signal and comparison with a binary event vector denoting behavior bouts, behavior event detection based on neural activity can be measured using the true positive rate (TPR) and the false positive rate (FPR) over all time-points. Plotting the TPR against the FPR over a range of binary thresholds, spanning the minimum and maximum values of the neural signal, yields an ROC curve that describes how well the neural signal detects behavior events at each threshold. We used the area under the ROC curve (auROC) as a metric for how strongly neurons are modulated by each behavior. For each neuron/behavior category (for both subject and opponent behaviors), the observed auROC was compared to a null distribution of 1,000 auROC values generated from constructing ROC curves over randomly permuted calcium signals (that is, traces that were circularly permuted using a random time shift). A neuron was considered significantly responsive ($\alpha = 0.05$) if its auROC value exceeded the 95th percentile of the random distribution (auROC < 2.5th percentile for suppressed responses, auROC > 97.5th percentile for excited responses). Throughout, ''neutral cells'' refer to neurons that were not identified as responsive during subject or opponent behaviors.

While the significance of the auROC values for single cells can be analytically determined by performing a Mann-Whitney U test, the test statistic from the U test carries a caveat of being highly influenced by group sample sizes. Because of the kinetics of the calcium fluorescence signals, treating individual frames (sampled here at 30Hz) as independent samples for a U test would inappropriately inflate the power of the statistical test. Instead, we chose to use the permutation-based resampling method described above in order to test for statistical significance, as this approach is not sensitive to this particular sampling issue.

For comparison of response characteristics across subject and opponent cells (Figures S8B and S8C), the response strength for each neuron and each behavior was calculated as the average z-scored *ΔF/F* activity during all behavior epochs of a given type. Response probability for each neuron and each behavior was calculated as the percentage of behavior events with average neural activity that exceeded 110% of the local baseline (increased by more than 10% above baseline), taken over the 10 s preceding behavior onset.

In order to ensure that opponent cell responses to opponent behaviors were not contaminated by activity associated with overlapping subject behavior, we analyzed the mean activity of opponent cells during isolated subject and opponent behavior bouts (Figures 7H–7J). For this analysis, all events that overlapped across subject and opponent (within 2 s) were removed. We confirmed that subject animals did not display observable behavior, and did not exhibit changes in movement along the tube, during opponent behaviors used for this analysis (Figures S3F and S3G).

For analysis of cells responding during opponent behavior in the open arena assay (Figures S8D and S8E), ROC analysis was performed using binary behavior vectors denoting all pooled social behaviors from the opponent that do not overlap with subject behavior and rest. Observed auROC values were compared with null distributions based on randomly permuted calcium traces (as described above, $\alpha = 0.05$). The mean activity of open arena opponent cells was computed over non-overlapping subject and opponent behavior, or baseline epochs.

Mean activity of opponent cells in the tube test was found to be significantly higher during social contact than after introduction of a separator to abolish contact (Figure S8H), suggesting dependence on social context and interaction with another individual for opponent cell firing.

Analysis of population dynamics during behavior

Principal Component Analysis

To visualize population responses during social behavior, we applied principal component analysis (PCA) to obtain components that capture the covariance of the neural population during behavior events (Cunningham and Yu, 2014). After binning neural traces into 1 s bins, trial-averaged responses were computed over a time window of 40 s (20 s prior to 20 s after event onset) for each neuron/behavior event, and concatenated across event types (e.g., approach, push, and retreat). Responses for each neuron were formed into a matrix which was used to perform PCA. Population vectors were then averaged over individual behavior bouts and projected onto the first 2 principal components for visualization (Figure 5K). For comparison of population responses to different behavior types (Figure 5L), we calculated the pairwise Euclidean distances between PC-projected population vectors (using the first 3 principal components) within or across different behaviors.

Mahalanobis Distance

In order to visualize population response dynamics during behavior, we used the Mahalanobis distance, which provides a measurement of the separation between two population vectors while accounting for the covariance structure of the underlying distribution. This provides a way to quantify the strength of specific population response patterns, as opposed to simply measuring the average response of all neurons (Figure S6A; see Li et al., 2017; Remedios et al., 2017). Average population vectors were constructed over frames from different behavior categories or over all baseline frames. The Mahalanobis distance between two vectors is computed as:

$$D_{\mathrm{MAH}}(x_1, x_2) = \sqrt{(x_1 - x_2)^{\mathrm{T}} S^{-1} (x_1 - x_2)}$$

where x_k is the mean population vector over all frames for event type k, and S is the covariance matrix computed over all baseline frames. For population response time-courses (Figure 5I), the Mahalanobis distance was measured between individual frame population vectors from a given class k and the average population vector over all baseline frames.

Behavior decoding based on population activity

In order to measure population-level encoding of social behaviors among dmPFC neurons, we constructed statistical models to predict behavior events based on population activity. For classification of individual behaviors, we used binary Fisher's linear discriminant (FLD) classifiers, and to distinguish between behavior types, we used a multi-class (3-way) Fisher's discriminant.

For all classifiers, training sets were constructed using population vectors during behavior bouts and negative training data was sampled from baseline (rest) frames. In order to measure the performance of FLD models, we split the data into training and tests sets and performed cross-validation. For each cross-validation fold, the test set represented 10% of the data drawn from 10 uniformly distributed 1% segments, and the remaining 90% training set was used to construct the model. For each fold, model performance was measured using the area under the ROC curve (auROC) for test data projected onto the Fisher discriminant. Overall model performance for each animal/session was calculated as the average over 50 folds where the training and test sets were randomly redrawn. Models were compared with null models constructed using training data with randomly shuffled class labels. Sessions with fewer than 5 bouts of the modeled behavior were not considered for this analysis. For frame-by-frame classification and visualization of the FLD projection (Figure 5M), frames were sampled uniformly every second over the entire session and used to construct training data to fit models. Population activity over the session was then projected onto the discriminant, and class predictions for each frame were evaluated.

For multi-class decoding of push, retreat, and approach behavior (Figure 5O), 3-way FLD models were constructed from population data using behavior vectors to define class labels, and cross-validation was performed as described above. Predictions were determined by taking the minimum Euclidean distance between test points and the mean of each class' training set after projection onto the first 2 FLD components. Performance for each fold was measured using the average accuracy for each class (weighted by the number of examples in the test set), and overall model performance was taken as the average over 50 folds (as described above). Models were compared with null models constructed using training data with randomly shuffled class labels.

For discrimination of subject versus opponent behaviors (Figures 7K and 7L), behavior bouts within each animal were pooled together. Behavior frames that overlapped (concurrent subject and opponent behaviors) were removed from the analysis, and the remainder were used to construct training and test sets using the same cross-validation method as described above. Dimension reduction was first performed on the training data using partial least-squares regression (PLS), and FLD components were computed from the training data after projection onto the first 10 PLS dimensions. For visualization (Figure 7K), population vectors from the test set from one example session/fold were projected onto the first two FLD components. For each model, ROC analysis was performed to quantify discriminability of subject and opponent population responses, and auROC values were averaged over the holdout partitions for each session. Overall model performance was quantified using the average of auROC values over all sessions (Figure 7L), and was compared with null models constructed using training data with randomly shuffled class labels.

Generalized linear models of single-neuron and population activity

Modeling neural activity across brains of interacting animals

In order to gain deeper insight into correlations of dmPFC neurons across animals in the tube test (Figure 6), we constructed Gaussian-residual generalized linear models (GLM) to express the mean activity of all neurons in one animal as a function of individual activities of neurons in the opponent. After binning calcium data from both animals into 1 s bins, GLMs were fit as:

$$\mu = X\beta + \varphi$$

where μ is the predicted mean activity in animal A, X is the matrix containing all normalized (to maximum) calcium traces from animal B, β is a vector of coefficients fit to each neuron in X, and φ is an error term. In order to validate the predictive power of GLMs, we performed 10-fold cross validation by withholding 10% of the data, sampled uniformly in 1% segments, from model fitting. Full predicted activity traces were constructed by concatenating test predictions from each fold, and the overall performance of the model was evaluated using the Pearson's correlation coefficient (PCC) between the predicted activity μ and ground truth. Model performance was compared to the performance of null models constructed using randomly permuted calcium data—97.2% of the mean activity models individually exceeded chance levels (the 95th percentile of the null distribution). Cross-validated R^2 was also used as an alternative performance metric to confirm model significance and validated the mean activity models. Coefficients β from full models were z-scored before being pooled with those from other models (Figure 6G). For models of subpopulation activity (Figures 6H–6M), the response variable μ was the mean activity of the top 15 behavior-excited neurons based on their rank-ordered auROC values for a given behavior type, and z-scored coefficients were averaged within each session according to cell identity before comparison across sessions/groups.

Modeling neuron activity using behavioral variables

To analyze the contributions of subject and opponent behaviors to the activity of individual neurons, we constructed GLMs using the behaviors and positions of both animals. Single-neuron GLMs were fit using a Poisson model with a log link function:

$$\ln(\mu) = X\beta + \varphi$$

where μ is calcium activity from one cell and X is a matrix of behavior and position vectors. The use of a log link function for single neuron models was based on the assumption that a Poisson distribution best characterizes the calcium data used to fit the model, as has been made in previous studies (Driscoll et al., 2017). Binary behavior vectors were smoothed with an exponential decay function ($\tau = 3$ s). Position vectors for each animal were projected onto four Gaussian functions centered at four positions $(P_1,\ P_2,\ P_3,\ P_4)$ that uniformly tiled the length of the tube. In total, 14 variables were used to model activity: 6 behavior vectors (corresponding to push, approach, and retreat for both animals) and 8 position vectors (corresponding to the four tube positions for both animals). Model performance was quantified following 10-fold cross validation using the Pearson's correlation coefficient (PCC) of predicted and observed activity, and was compared to a distribution of null models fit using randomly permuted calcium data. Models were only considered significant and used for downstream analysis if their performance exceeded the 99th percentile of the null distribution. Significance testing for individual coefficients (Figures 8B and 8C) was based on a likelihood ratio test ($\alpha = 0.05$) which compares model performance with the associated variable against a null model without it. For comparisons of coefficients between dominant and subordinate animals, coefficients were z-scored and averaged within each animal/session. Results of coefficient analyses shown in Figures 8C and 8D were also consistent with analyses performed with models identified using R^2 as a performance metric (Figures S8K and S8L).

Modeling neural activity using both neural activity and behavioral variables across animals

In order to examine whether interbrain correlations observed in the open arena and tube test experiments exceeded modulations that could be only explained by observable behavior variables, we compared the performance of mean activity GLMs fit using both animal's behavior (''behavior-only'' Model 1) with the performance of models that also included mean activity from the opponent animal as an additional explanatory variable (''interbrain'' Model 2) (Figure 2J; Runyan et al., 2017). For these analyses, GLMs were Gaussian residual models, behavior vectors were exponentially smoothed ($\tau = 3$ s), and behavior vectors and calcium activity were binned into 1 s bins prior to model fitting. Model performance was measured using cross-validated PCC with 10-fold cross-validation, as described above. We measured the change in model performance upon inclusion of opponent activity as *(*Model 2 – Model 1)/Model 1 (“GLM performance difference” in Figures 2K and S4G). Performance indexes were compared with those of models constructed using randomly-permuted opponent activity (behavior variables were not permuted).

Analysis of interbrain neural activity correlations

Correlation of neural activity across brains

Because previous hyperscanning studies have investigated correlations of aggregate, region-level activity patterns, we used the mean activity of all z-scored *ΔF/F* traces in each dmPFC population (mean *ΔF/F*, effectively their summed activity normalized by the number of recorded neurons) as a measure of overall neural activity. For both open arena and tube test experiments, interbrain correlations across mouse dyads were calculated using the Pearson’s correlation coefficient (PCC) of the overall neural activity across the entire session. To fairly compare interbrain correlations across sessions with different durations (Figures 2F and 4J), we cropped traces to the duration of the shortest session (10 min and 0 s for the open arena; 11 min and 16 s for the tube test). Interbrain correlations for each pair were compared to the 95th percentile of random permutation null distributions (Figures S4A and S4B). In order to confirm that changes in interbrain correlation when animals were separated were not due to changes in the autocorrelation of each signal, we also compared phase-randomized signals before and after separation in both the open arena (Figure S2E) and tube test experiments (Figure S4E). Phase-randomized surrogate signals (Figure S4D) were computed by independently randomizing the phase of each Fourier component, which disrupts the temporal structure of the signal but preserves its mean, variance, and autocorrelation. For comparison of overall correlations with dominance relationships (Figure 8J), interbrain correlations were measured over the first 5 minutes of each session to ensure a high degree of social interaction during each epoch.

Cross-correlations of neural activity across brains

In order to gain more insight into the timescale at which interbrain correlations occur, we performed a cross-correlation analysis using the neural activity from interacting animals in both the open arena and the tube test. We calculated the correlation between *ΔF/F* activity traces with different time shifts, ranging from −2 minutes to +2 minutes, and plotted the correlation as a function of time lag (Figures 1M and 4N). For interacting animals in both experiments, the peak of the average cross-correlation occurred precisely at 0.0 s lag. For both experiments, we also compared the correlation at the peak with the correlation at ± 60 or ± 30 s lag (based on the cross-correlation functions shown in Figures 1M and 4N). Cross-correlations were compared with those of phase-randomized signals (described above) to confirm that structure in the cross-correlation is not due to autocorrelations in each calcium trace (Figures 1N and 4O).

Interbrain correlations among subsets of neurons

To determine the contributions of subject-encoding and opponent-encoding neurons to interbrain correlations, we calculated correlations across animals after removing different types of cells from each neural population based on functional identity (e.g., behavior-excited or behavior-suppressed). While removal of behavior-excited cells resulted in a decrease in interbrain correlations, removal of behavior-suppressed cells did not (Figures 6A, S7A, S7B, 7M, S8F, and S8G). Neutral cells were neurons that were not identified as either subject-encoding or opponent-encoding by the ROC analysis. For subpopulation analyses in Figures 7N, 8I, 8K, and 8O, subsets of 25 cells from each animal were used to calculate interbrain activity correlations in order to control for differences in correlation that could result from unequal population sizes. Subsets of the top behavior-encoding were selected (with the largest auROC values) for modulation by subject or opponent behavior. Neutral cells were defined as described above, and were sorted (in ascending order) and selected by $|\text{auROC}_{sub} - 0.5| + |\text{auROC}_{opp} - 0.5|$, where auROC_{sub} and auROC_{opp} are the auROC values calculated from neural responses to pooled subject and opponent behavior, respectively. To assess the relative contributions of subject and opponent cells to interbrain correlations, we also removed fixed numbers (to ensure a fair comparison between subject and opponent cells) of subject, opponent, or neutral cells (ranging from 1 to 25) from each animal and computed interbrain correlations over these populations (Figures 7O and 7P).

Relationship between interbrain correlations and behavior interaction

In order to examine whether interbrain correlations could predict behavior interactions, we compared the degree of correlation prior to behavior in one animal to the probability of behavior response from the interacting partner (Figures 8G–8I). For each behavior event (pooled across behavior categories) in each tube test session, the PCC of interbrain activity across the two animals was taken over the 30 s prior to behavior onset. All behavior events with any behavior from the interacting partner starting in the 15 s prior to behavior onset were removed from the analysis to ensure that preceding correlations were not contaminated by preceding behavior bouts. For each range of PCC (e.g., 0.1 – 0.2), the probability of behavior response in the reacting animal was calculated by summing all behavior events from the reacting animal over 3 s following the onset of its opponent’s behavior for all epochs associated with that PCC range, and then dividing by total the number of epochs.

Interbrain correlations during matched behavior epochs across animal pairs

In order to address whether interbrain correlations could be accounted for simply by concurrent behaviors, we compared correlations of mean activity across animals during single epochs (30 s) of concurrent behavior (e.g., interacting animals A versus B), with behavior-matched epochs across pairs that did not interact (e.g., non-interacting animals A versus C) (Figures 2G and 4K). Specifically, we identified all epochs in which two interacting animals displayed behavior that have concurrent onset times (within 3 s), and computed interbrain activity correlations over these epochs (A versus B). Behavior epochs in one animal were then matched with behavior epochs in another non-interacting animal from a separate session (A versus C), such that the behavior types and onsets were identical to those in the epoch from the interacting pair (A versus B). Other types of behaviors immediately before and after the temporally aligned behavior were also matched, such that overall behavior transitions, as well as the onsets of the aligned behaviors, were the same. The associated interbrain correlations were then compared. For the analysis shown in Figure S4F, a single behavior bout in one animal was matched and aligned with an equivalent behavior bout from a separate non-interacting animal, and if multiple behavior bouts of the same type occurred within short intervals (1 s), they were considered as one bout. No other behavior bouts occurred during the epoch. For these analyses, lower PCC values are expected as interbrain correlation is lower in shorter temporal windows (Figures S2C and S4C).

QUANTIFICATION AND STATISTICAL ANALYSIS

All analyses for this study were conducted using custom routines in MATLAB (Mathworks), and are described in the respective Method Details, Results, and Figure Legends. All bar plots with error bars represent mean ± SEM; all box and whisker plots represent the median, interquartile range (box), and 5^{th} to 95^{th} percentile (whiskers) of the underlying distribution, unless otherwise specified. For all statistical tests throughout, normality of the data and equal variance of groups were not assumed, and non-parametric (Wilcoxon rank-sum and signed-rank) tests were used for unpaired and paired group comparisons, respectively. Statistical significance was defined with $\alpha < 0.05$ using two-tailed tests. For comparisons of proportions of binary-valued variables, Fisher's exact test was used. For comparisons of behavior bout length and cell pairwise distance distributions, two-sample Kolmogorov-Smirnov tests were used. Resampling methods based on temporally-permuted calcium traces were used to assess significance of auROC values for behavioral modulation of neural signals and performance of GLM models. Statistical significance of FLD classifiers was assessed by comparison with null models constructed using training data with shuffled class labels. The sizes of mouse groups were not pre-specified and approximated those of previous work.

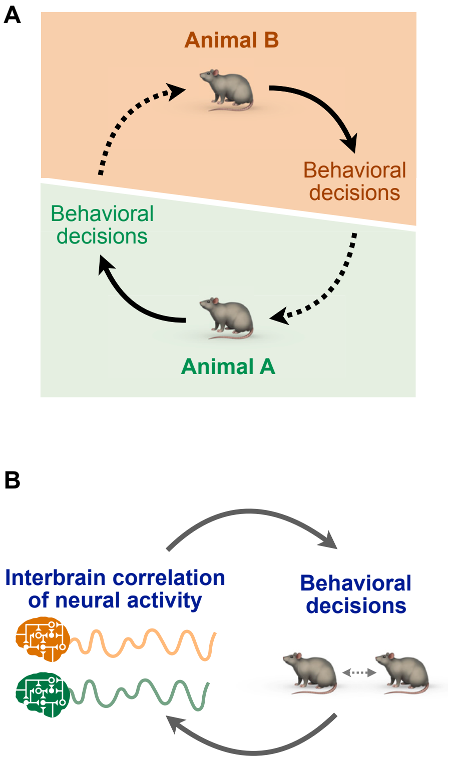


Figure S1. Social Behavior and Interbrain Coupling in Interacting Animals, Related to Figure 1

(A) Schematic showing social behavioral decisions of animals engaged in dyadic social interaction.

(B) Feedback loop between interbrain synchrony and social interactions. The coupling of activity between interacting animals facilitates and sustains ongoing social interaction.

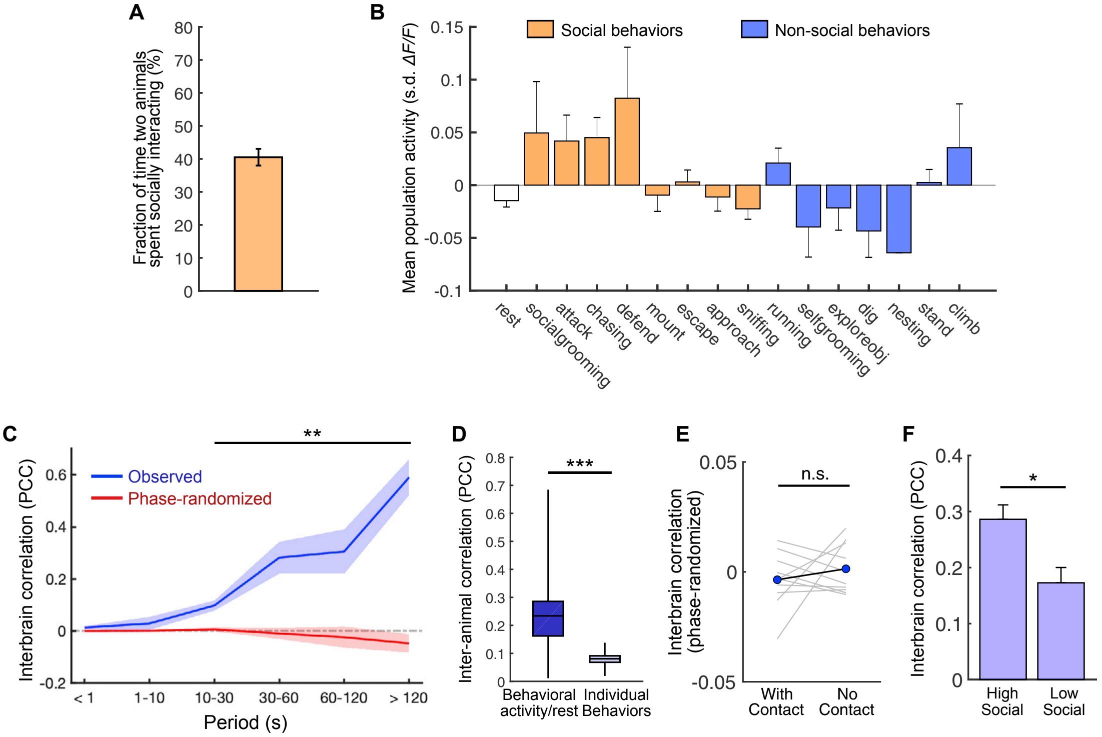


Figure S2. Analysis of behavior and Interbrain correlations in the open arena, Related to Figures 1 and 2

(A) Total time two animals spent interacting in the open arena (which includes time when a single animal or both animals are engaged in social behavior).

(B) Mean dmPFC activity during different types of social (orange) and non-social (blue) behaviors across all animals engaged in open arena interactions (mean ± SEM).

(C) Correlations of dmPFC activity (blue) and phase-randomized traces (red) across animal pairs at different timescales. Mean activity traces were decomposed into different frequency bands using a Fourier transform. Interbrain correlations are stronger at slower timescales, consistent with the notion that correlations depend on a larger context of continuous, ongoing interaction on a scale of seconds to minutes.

(D) Correlation of behavioral activity and rest across animals interacting in the open arena (all types of behavior pooled, left) and correlation of specific types of behaviors across animals (right) ($p^{***} < 0.001$). This suggests that, across animals, behavior activity and rest are somewhat correlated (left), whereas individual behaviors are not correlated (right).

(E) Correlations of phase-randomized activity traces across animals in the open arena with or without social contact ($p > 0.05$ – not significant).

(F) Comparison of interbrain correlations among animal pairs that naturally displayed high or low levels of mutual social interaction (STAR Methods). Pairs with a higher degree of social interaction showed higher interbrain synchrony, consistent with the notion that synchrony depends on ongoing interaction.

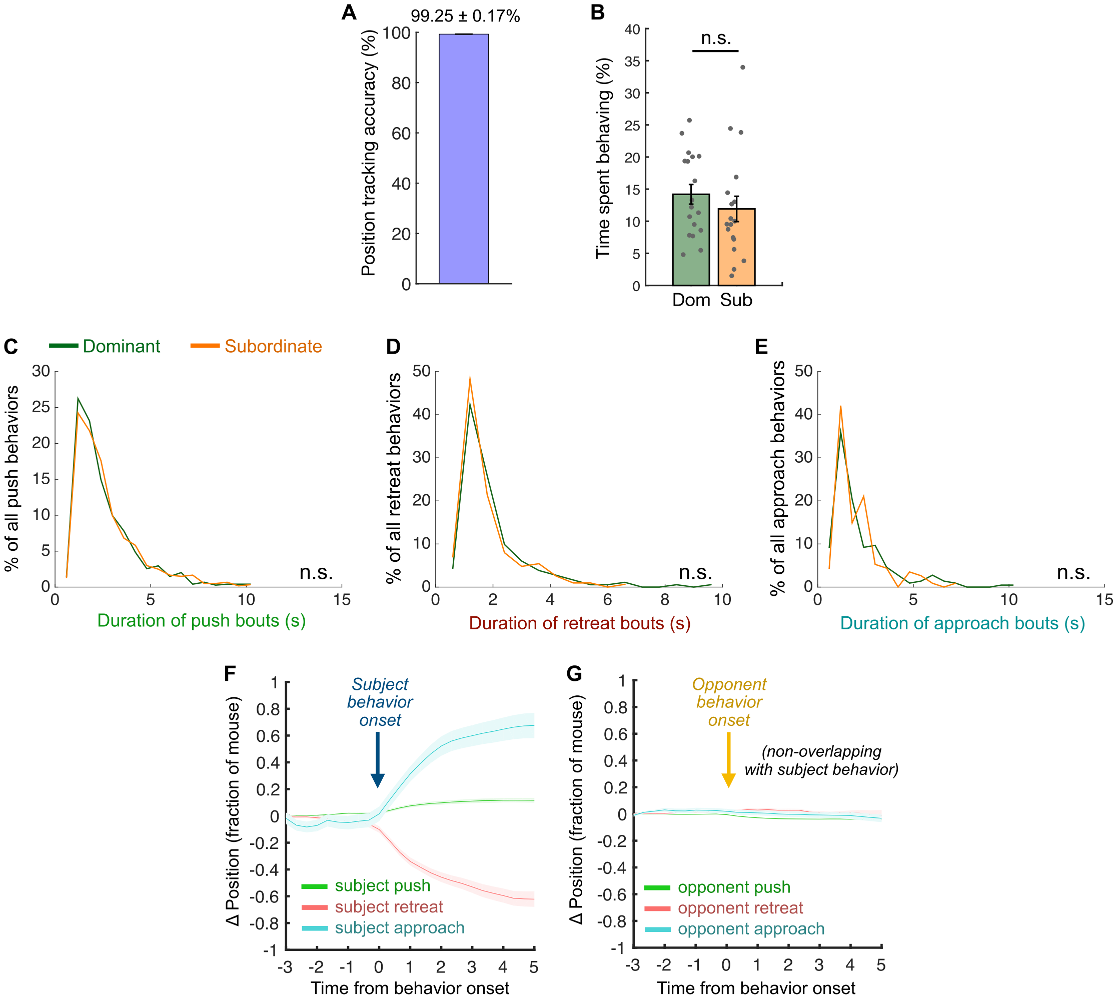


Figure S3. Automated Tracking and Analysis of Animal Behavior during the Tube Test, Related to Figure 3

(A) Performance of the convolutional neural network to automatically track the locations of interacting mice in behavior movies, measured by the accuracy of the algorithm to properly identify both mice and correctly determine their positions in a subset of randomly drawn frames, compared with ground truth assessment determined by an unbiased individual (mean ± SEM).

(B) Total percentage of time spent behaving among dominant and subordinate animals across all pairs. For each pair, the dominant animal is the one with the greater mean tube position (mean ± SEM, $p > 0.05$; not significant).

(C-E) Distribution of per-bout behavior durations for push (C), retreat (D), and approach (E) behavior in dominant or subordinate animals (Kolmogorov-Smirnov test, $p > 0.05$; not significant).

(F) Average change in position of mice during subject push, retreat, or approach behavior (mean ± SEM).

(G) Average change in position of mice during opponent push, retreat, or approach behavior (mean ± SEM; behavior bouts when subject and opponent behavior overlapped were removed from analysis).

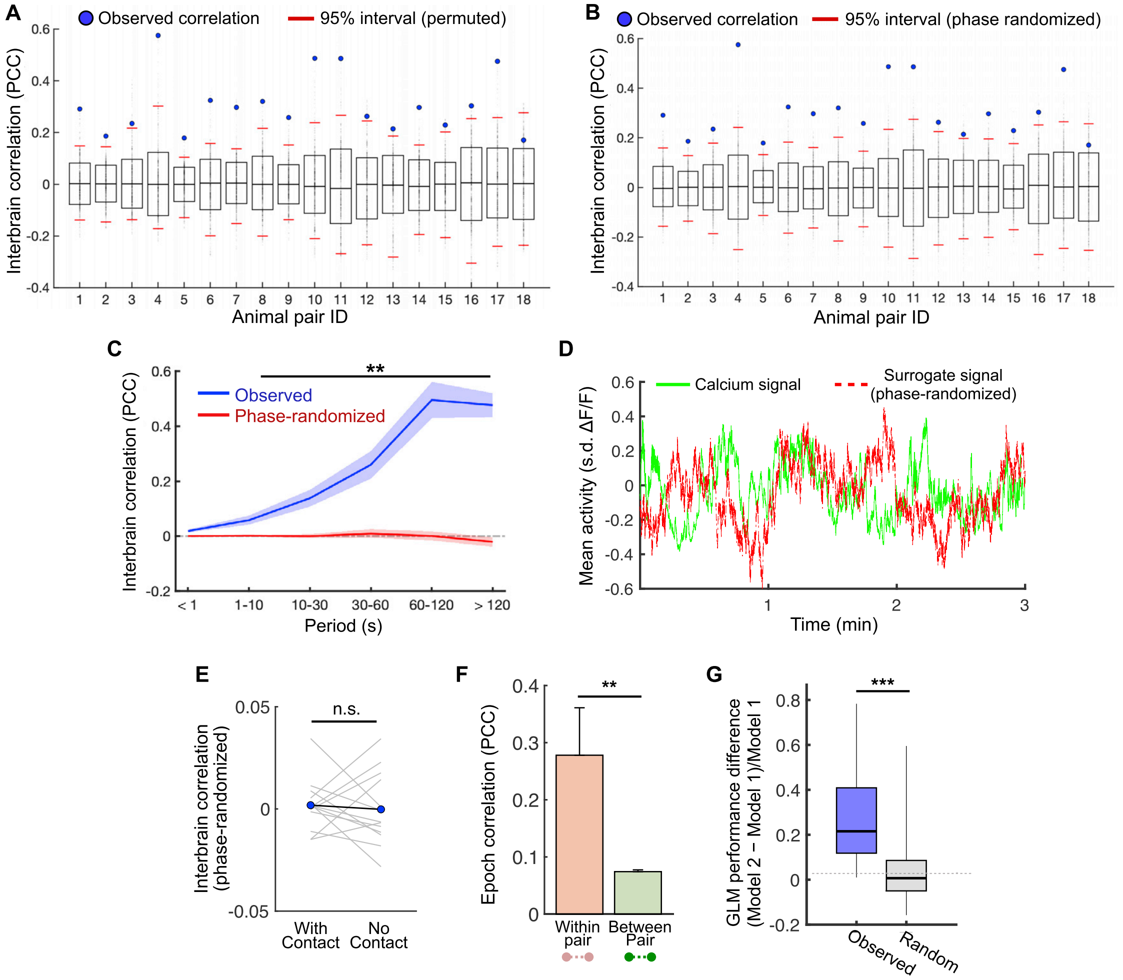


Figure S4. Analysis of Interbrain Correlations in the Tube Test, Related to Figure 4

(A, B) For each animal pair, the observed interbrain correlation (PCC; blue dots) shown against a null distribution of PCCs. Boxes indicate mean ± standard deviation of the null distributions; red lines indicate 95% intervals (2.5th and 97.5th percentile). (A) Null distributions are generated from temporally permuted traces. (B) Null distributions are generated from phase-randomized traces.

(C) Correlations of dmPFC activity (blue) and phase-randomized traces (red) across pairs at different timescales using Fourier decomposition of signals into different frequency bands. Interbrain correlations are stronger at slower timescales, consistent with the notion that correlations depend on a larger context of continuous, ongoing interaction on a scale of seconds to minutes.

(D) Example trace of the average activity of all dmPFC neurons in one animal (green), and a surrogate phase-randomized signal (red) with disrupted temporal structure but identical mean, variance, and autocorrelation as the original trace (STAR Methods).

(E) Interbrain correlations of phase-randomized traces from tube test experiments with or without social contact, as in Figure 4J ($p > 0.05$; not significant).

(F) Comparison of interbrain correlations during epochs with concurrent isolated behavior bouts (STAR Methods) in interacting pairs in the tube test (left), and behavior-matched epochs from non-interacting pairs (right) (mean ± SEM).

(G) The difference in performance of GLM models schematized in Figure 2J for animals engaged in the tube test, compared with that using phase-randomized activity from the interacting partner. The GLM performance difference quantifies the relative difference in model performance when activity from the interacting partner is included as a variable in addition to behavior variables (STAR Methods; p** < 0.01).

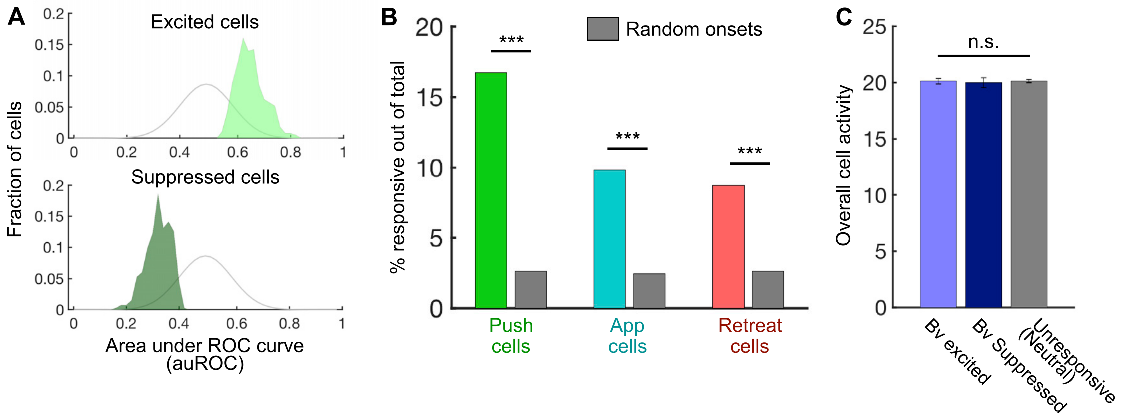


Figure S5. Activity and Spatial Intermixing of Behavior Cells in dmPFC, Related to Figure 5

(A) Distributions of auROC (area under the ROC curve) values for cells that are excited (top) or suppressed (bottom) during behavior. Significantly responsive cells were determined using permutation testing (see STAR Methods). Gray curve indicates the distribution of auROC values from neutral cells that do not respond during behavior.

(B) Comparison between the percentage of behavior-excited cells identified over all tube test sessions and the percentage expected by chance. Chance levels were determined by comparing auROC values of temporally permuted calcium traces against random null distributions ($p^{***} < 1.0e-10$, Fisher's exact test).

(C) Average cell activities for behavior-excited, behavior-suppressed, and neutral (behavior-unresponsive) cells. For each neuron, overall activity is measured as the percentage of time the calcium trace is above 10% of its maximum value ($p > 0.05$; not significant).

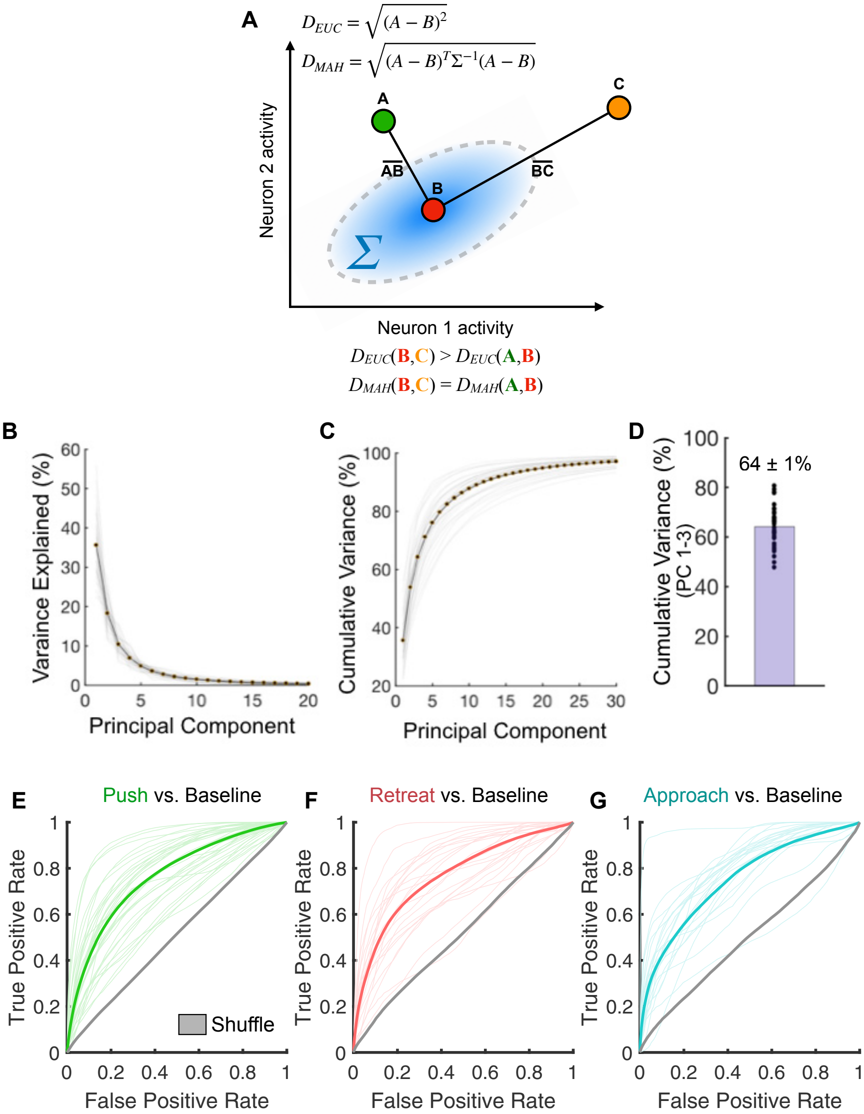


Figure S6. Separation of Population Responses Encoding Distinct Social Behaviors, Related to Figure 5

(A) Cartoon illustration of the Mahalanobis distance and Euclidean distance between pairs of points on a 2D plane. The Mahalanobis distance considers the shape of the underlying distribution of data by scaling dimensions based on their covariance (the correlational structure of the neural population). Although point C is further from B than A is in Euclidean terms, A and C are equidistant from B using the Mahalanobis distance.

(B) Percentage of the total variance of trial-averaged population activity during behavior in tube test sessions that is captured by principal components (gray curves); average over all sessions shown with black curve (see STAR Methods).

(C) The cumulative variance of trial-averaged population activity captured by principal components as a function of the number of components (gray curves); average over all sessions shown with black curve.

(D) Average variance in population activity captured by the first three principal components, as shown in (C) (mean ± SEM).

(E-G) ROC curves quantifying the performance of FLD decoders to predict push (E), retreat (F), and approach (G) behavior based on population activity. Thin color lines: performance for each session; dark color lines: average ROC curves taken over all sessions; gray lines: the average of chance decoders constructed using training data with randomly shuffled class labels.

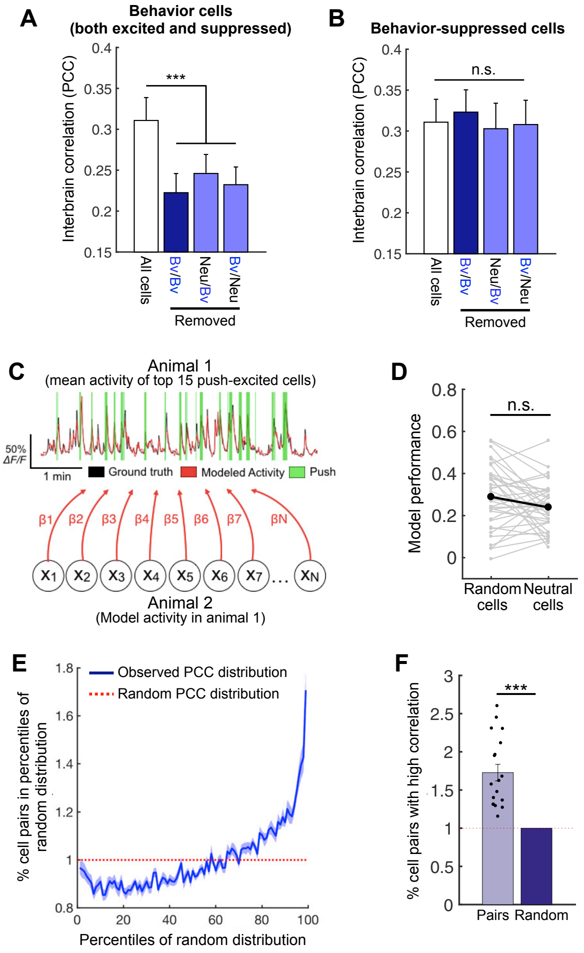


Figure S7. Single Behavior Cell Contributions to Interbrain Correlations, Related to Figure 6

(A) Interbrain activity correlations after removal of behavior cells (Bv; including both excited and suppressed cells) or neutral cells (Neu) from both animals (mean ± SEM).

(B) Interbrain activity correlations after removal of behavior-suppressed (Bv) or neutral (Neu) cells from both animals (mean ± SEM).

(C) Schematic of GLM fit to model mean activity of subsets of behavior-excited cells as shown in Figure 6H using single neuron activities from the interacting partner. Red line: modeled activity of the top 15 push cells from one animal/session. Black line: ground truth activity of the same group of cells.

(D) Comparison between the performance of GLMs constructed to model the mean of subsets (15 cells) of randomly selected or neutral cells ($p > 0.05$; not significant).

(E) Distribution of PCC of all single neuron pairs across interacting animals in the tube test (blue). Each bin represents one percentile of the random distribution (chance level of 1%, red) of correlations generated from calculating PCCs over temporally permuted calcium traces (mean ± SEM). This indicates that pairs of single cells across interacting animals exhibit a higher level of correlation than expected by chance.

(F) The percentage of single cell interbrain correlations that exceed the 99th percentile of null distributions generated from randomly permuted calcium traces, as in (E). Percentage of highly correlated cell pairs is compared with the chance level of 1% (mean ± SEM, $p^{***} < 10^{-5}$).

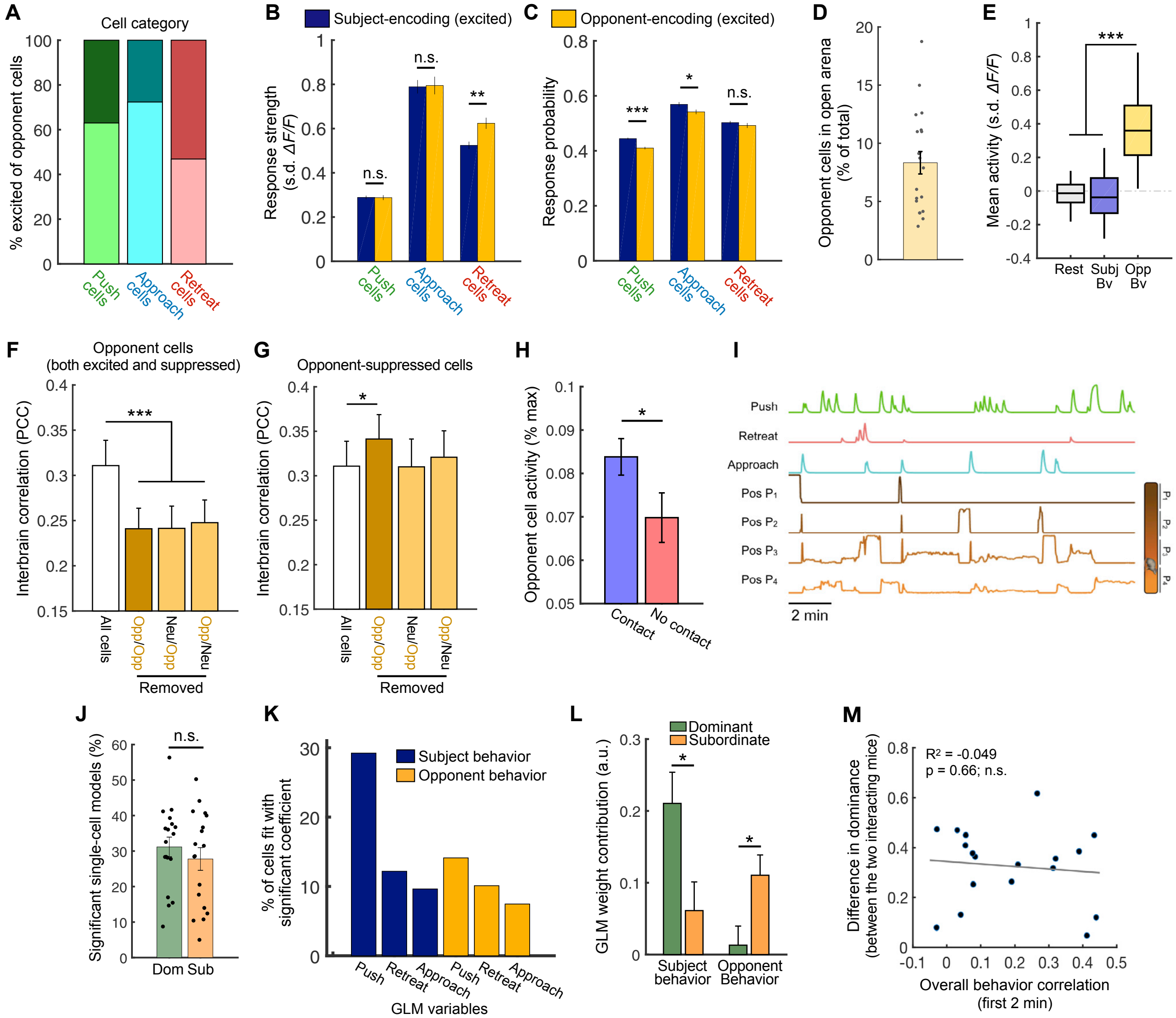


Figure S8. Analysis of behavior cell properties and single neuron models, Related to Figures 7 and 8

(A) Distribution of excited (light color) and suppressed (dark color) opponent cells within each behavior category.

(B, C) Response strength (B) and response probability (C) of subject behavior-excited and opponent behavior-excited cells for different behavior categories. The response strength for each cell is calculated as the mean activity over all behavior epochs. The response probability is calculated as the percentage of behavior events with neural activity exceeding 110% of the local baseline (STAR Methods).

(D) Percentage of neurons recorded during open arena interactions that respond selectively during opponent social behavior. Opponent cells in open arena interactions were identified using ROC analysis based on opponent behavior (not overlapping with subject behavior) and rest epochs.

(E) Mean activity of opponent cells during subject behavior, opponent behavior, or rest (when neither animal is behaving) in open area interactions.

(F, G) Interbrain activity correlations after removal of opponent cells (Opp) or neutral cells (Neu) from both animals. Opponent cells includes both excited and suppressed cells (F) or only suppressed cells (G).

(H) Activity (percent of the max activity value) of all behavior-excited opponent cells during the tube test with or without social contact.

(I) Illustration of the variables used to fit single neuron GLM models. Behavior vectors denoting social behavior of each animal are exponentially smoothed, and position coordinates for each animal are decomposed into four positions that tile the length of the tube (STAR Methods).

(J) Percentage of single neurons in each tube test session that are modeled well (exceed chance levels based on cross-validation) by a GLM fit to the behavior and positions of both animals.

(K) Percentage of cells fit with significant coefficients for individual subject and opponent behaviors, as in Figure 8C. Here, single-neuron GLMs were identified using cross-validated R^2 as an alternative performance metric.

(legend continued on next page)

(L) Contributions of coefficients in single neuron GLMs for subject and opponent behavior in dominant and subordinate animals. Weight contribution was calculated as the average of normalized coefficients over all cells in each animal, as in Figure 8D. Here, single-neuron GLMs were identified using cross-validated R^2 as an alternative performance metric.

(M) Relationship between overall behavior correlation across pairs during the first 2 min of interaction and their overall dominance difference over the session. Overall behavior correlations were measured by the correlation of the presence of behaviors of any types, which reflects the level of overall concurrent behavior.

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, $p > 0.05$, n.s. (B–D, F–H, J, L) Mean ± SEM.