Published in final edited form as:

Science. 2021 October 22; 374(6566): eabb4149. doi:10.1126/science.abb4149.

Social Agent Identity Cells in the Prefrontal Cortex of Interacting Groups of Primates

Raymundo Báez-Mendoza[1,*], **Emma P. Mastrobattista**[1], **Amy J. Wang**[1], **Ziv M. Williams**[1,2,3,*]

[1]Department of Neurosurgery, Massachusetts General Hospital, Harvard Medical School; Boston, MA, 02114; USA.

[2]Harvard-MIT Division of Health Sciences and Technology; Boston, MA, 02115; USA.

[3]Program in Neuroscience; Harvard Medical School; Boston, MA, 02115; USA.

Abstract

The ability to interact effectively within social groups is essential to primate and human behavior. Yet, understanding what neural processes underlie the interactive behavior of groups or by which neurons solve the basic problem of coding for multiple agents has remained a challenge. By tracking the interindividual dynamics of groups of three-interacting Rhesus macaques, we discover detailed representations of the groups' behavior by neurons in the dorsomedial prefrontal cortex, reflecting the other agents' identities, their unique interactions, social context, actions, and outcomes. We show how these cells represent not only interaction between specific group members, their reciprocation and retaliation but also their individual past behaviors. We also show how they influence the animals' upcoming decisions and their ability to form beneficial agent-specific interactions. Together, these findings reveal prefrontal neurons that code for the agency identity of others and a cellular mechanism that could support the interactive behavior of social groups.

One Sentence Summary:

Prefrontal neurons in primates code for the agency identity and group behavior of others.

Main Text:

Social groups play a foundational role in the behavior of most animal species. To interact effectively within social groups, individuals must be able to represent not only the identities of other group members but also their specific behaviors (1, 2). Without such

[*]To whom correspondence should be addressed, zwilliams@mgh.harvard.edu and rbaez-mendoza@mgh.harvard.edu.

Author contributions: R.B.M and Z.M.W. conceived and designed the study and wrote the manuscript. R.B.M, E.P.M., and A.J.W. performed the experiments. R.B.M. analyzed the data. Z.M.W. supervised all aspects of the research.

Competing interests: The authors declare no conflict of interest.

Supplementary Materials:
Materials and Methods
Figures S1–S8
Tables S1–S2
References (50–62)

representations, it would not be possible to understand how the actions and outcomes of specific individuals relate or how one's actions affect specific group members (3). It would also not be possible to develop mutually beneficial affiliations and avoid exploitation by others (1, 4, 5). Understanding how neurons in the brain represent the behavior of specific individuals or their interaction within groups, however, has remained a challenge. While prior investigations have revealed neurons in temporal regions that respond to the specific identities or facial features of others (6–9), they do not reveal how neurons encode another's behavior or their interaction. Other studies, by comparison, have identified neurons in associative brain regions that respond to another's behavior (10–15) but do not reveal how neurons represent their specific identities or group interactions.

The majority of primates live within social groups in which an individual's success relies on their ability to interact effectively with their conspecifics (1, 4, 16). Rhesus macaques, in particular, can recognize different individuals (6, 17), form non-kin interactions and long-lasting alliances (18–20). They also engage in mutually beneficial behavior based on reciprocation between specific individuals and keep track of others' behavior (19, 20). By studying the group behavior of Rhesus macaques, we can therefore begin to characterize how neurons in the primate brain represent interactions within small social groups and to explore how neurons code for the specific identities and behaviors of others.

Three-agent group interaction task in Rhesus macaques

We devised a three-agent task in which three adult male Rhesus macaques sit at a turntable, each of which could offer a food reward to either of the other two monkeys over successive trials (Fig. 1A–B). For each successive trial, one of the primates was assigned to be the "actor" and could use a handle on the turntable apparatus to offer a food reward to one of the other two agents ("recipient"). Further, the primate assigned to be the actor would alternate in a pseudo-random fashion from trial to trial (Fig. 1C). Thus, for example, monkey 1 could be the actor in one trial and may offer a reward to monkey 2. On the subsequent trial, monkey 2 could be the actor and may reciprocate that same offer of a reward to monkey 1 or, instead, offer a reward to monkey 3.

Next, to further dissociate the actor's movements from the specific individual receiving a reward, we set the apparatus such that either a clockwise or counterclockwise handle movement allowed the actor to offer a reward to the same animal on different trials (Fig. 1D, left and Fig. S1). To further limit the possibility that animals used simple conditioned responses, the trials also alternated such that the monkey receiving a reward may be different from the monkey offering reward as the actor the following trial (Fig. 1D, middle). Finally, to dissociate the location of reward from the specific monkey receiving it, we alternated the physical locations of the primates in relation to one another halfway through the session (Fig. 1D, right).

Therefore, taken together, the nature of this task aimed to mimic some of the basic ethological features that define interactions between primates within groups (e.g., offering and receiving reward or grooming and being groomed) but in a way that could be studied in a neurophysiologic setting. More importantly, it allowed us to examine interactions between specific individuals within the group (e.g., did monkey 1 or 2 receive a reward and, if so, was a reward given to them by monkey 2 or 3). All trial conditions were controlled in an automated fashion, and all events were recorded and analyzed offline at millisecond resolution (21). For each new session and day, a different triad of monkeys was selected from a possible four communally housed adult male macaques. The group performed an average of 105 ± 8.7 (mean ± standard error of the mean (SEM)) trials per session for a total of 22 sessions.

Tracking agent-specific interactions within the primate groups

Behaviorally, the primates reciprocated past offers of reward, suggesting that they kept track of their interaction with specific individuals in their group. Because the actors had to choose between two possible agents, we could examine the interaction between specific group members. Here, we find that the animals were significantly more likely than chance to reciprocate an offer of reward from another animal (9.2% ± 4.0% above chance; signed-rank test, Z = 2.3, P = 0.01; Fig. 1E, left; Fig. S2A). Therefore, if monkey 1 gave a reward to monkey 2 on a particular trial, for example, monkey 2 would be more likely to offer reward to monkey 1 on a subsequent trial (16, 22).

The animal’s behavior also reflected the specific type of interaction with the other group members. Whereas reciprocation of reward reflects a mutually positive interaction, retaliation reflects a negative one. For example, if monkey 1 gave reward to monkey 2 in the previous trial, monkey 3 retaliated against monkey 1 by offering a reward to monkey 2 in the following trial (Fig 1E, middle). Here, we find that the primates were significantly more likely to retaliate than expected by chance (10.2% ± 4.2%; signed-rank test, Z = 2.30, P = 0.01; Fig. 1E, middle; Fig. S2A). Moreover, when considering both positive and negative interactions together (i.e., ‘Tit-for-Tat’ strategy in which the current actor gives back to the previous actor if it received reward and withholds reward if they had not received reward) (16), we find that the animals were more likely than chance to display this behavior (6.1% ± 3.0%; signed-rank test, Z = 1.98, P = 0.023; Fig. 1E, right and Fig. S2A). All animals showed similar reciprocity and retribution across groups (all tests: Z>1.65, P<0.04; (21)). These results therefore suggested that the animals kept track not only of whom they interacted with but also how.

The monkeys’ behavior did not reflect simple conditioned responses. An important feature of the task was that either a clockwise or counterclockwise movement by the actor could deliver the reward to the same individual on different trials (Fig. S1; (21)). Moreover, the pseudorandom sequence in which the role of actor was assigned meant that there was no guarantee that the reward recipient would be the actor on the following trial (Fig. 1C, (21)). Here, we find that the animals were significantly more likely to reciprocate past offers of reward to a specific individual both when the offer occurred one or two trials back (7.1% ± 3.2%; signed-rank test, Z = 2.04, P = 0.02). By contrast, the monkeys displayed no evidence of “Win-Stay-Lose-Switch” behavior (i.e., simply responding to receipt of reward independently of the specific agent offering it; signed-rank test, Z = 0.4, P = 0.68); together suggesting that the animals responded to the past actions of specific individuals within the group rather than simply the last location from which they received a reward.

Social context dependency and specificity of group interactions

To further confirm that the monkeys kept track of their interactions with specific individuals in the group, we switched their physical seating positions in relation to one-another halfway through the session. Using this manipulation, we found that none of the primates displayed a systematic reward assignment preference to a particular location either before or after the switch (signed-rank test, $Z = 0.32$, $P = 0.74$). More notably, they continued to display reciprocity with specific animals that had offered them reward on past interactions regardless of seating arrangement (signed-rank test, $Z = -0.89$, $P = 0.37$; before vs. after change).

Next, we examined the influence that the history of past interactions and social context played in the animal's behavior. Prior studies have shown that the past behavior or 'reputation' of specific individuals and their social dominance status can markedly influence how group members interact with them (23–25). Here, we find that difference in the other's reputation based on past interactions (i.e., how likely they were to reciprocate over the past 20 trials) had a significant effect on the animal's choices (Odds Ratio (OR) = 1.54; $t = 9.2$, $p = 3.5\times10^{-20}$; Fig. S2C). Furthermore, all animals developed transient duopolies (i.e., consistent runs of reciprocation) at probabilities that were significantly higher than expected from chance (permutation test; $P < 0.05$; Fig. 1F). While social dominance did not have an independent effect on the animal's choices (i.e., on the current trial; reciprocity, $Z = 0.41$, $P = 0.68$; retaliation, $Z = -0.061$, $P = 0.95$; tit-for-tat, $Z = 0.68$, $P = 0.49$; Wilcoxon rank-sum test, Fig. S2B), it did play a role when considering the animals' past interactions (i.e., whether past interactions were with a more dominant or subordinate animal; OR = 1.14; $t = 2.1$, $p=0.035$; Fig. S2C).

We also confirmed the social-context dependency of the animals behavior by replacing the other two primates with distinct inanimate totems while yoking trials from past sessions (Fig. S2E; (21)). Here, we find that replacing the other group members with totems led to a loss of reciprocation (signed-rank test, $Z = 0.42$, $P = 0.34$; Fig. S2F) and tit-for-tat behavior (signed-rank test, $Z = 1.05$, $P = 0.15$). Together, these results suggest that the animals kept track of who they previously interacted with and that their choices were dependent on the social context of their interaction.

Finally, we used two additional ethological metrics to evaluate the animals' interactions (Fig. S3A–E). Consistent with prior field studies demonstrating that primates are more likely to look at the individuals they interact with (25–29), we find that the monkeys look first (58.9% ± 4.4% vs. 50% chance; $\chi^2(1) = 5.35$, $P = 0.021$) and longer (42.6% ± 3.4% vs. 57.3% ± 3.5%, recipient vs. non-recipient; $t(17) = 2.12$, $P = 0.049$, paired t-test) at the monkey receiving reward (Fig. S3C). We also examined whether differences in facial expressions may have affected the animals' choices. Of the trials tested ($n = 450$), we find that the most common facial expression displayed by the animals (when they were potential recipients of reward) prior to the actor's choice was affiliative (83.8%, $n = 78$; Fig. S3D). These expressions, however, did not alter the overall likelihood that the actor would reciprocate with reward to the expressing monkey ($\chi^2(1) = 1.36$, $P = 0.24$), that the actor would retaliate ($\chi^2(1) = 0.53$, $P = 0.46$) or engage in Tit-for-Tat strategy ($\chi^2(1) = 0.004$, $P = 0.94$, Fig. S3E).

Single neuronal representations of group behavior and receipt of reward

Based on these findings, we next investigated the relationship between neuronal activity and the real-time interaction dynamics between animals in these groups. Together, we recorded from 521 neurons in the primates' dorsomedial prefrontal cortex (dmPFC; Brodmann's area 24) along the dorsal anterior cingulate sulcus (Fig. S4A) — an area previously implicated in social cognition in both monkeys (10, 30–32) and humans (33–35). Only units with a high degree of signal-to-noise, adequate refractory period, and stable waveform morphology were used (Fig. 2A, inset, S4B). Here, for neuronal analysis, we defined the primate from which neuronal activity was recorded as 'Self' and the other two agents as 'Other Monkey 1' and 'Other Monkey 2' (Fig. 2A).

We first asked whether certain neurons in the population responded to the reward outcome of specific individuals within the group. Because each recorded animal interacted with two other agents, we could importantly examine not only whether another monkey received a reward, but also which specific monkey received it. Focusing on the reward period, we found that 19.9% (n = 104) of the neurons displayed a change in their activity when any of the other animals received a reward (two-way ANOVA with *post-hoc* testing corrected for repeated comparison across the three agents, $P < 0.01$; (21)). More notably, 9.6% (n = 50) of neurons displayed a significant change in their activity only when a specific other individual received reward (i.e., the neurons 'preferred' other monkey; Fig. 2D top, Fig. S5A); a proportion that was significantly higher than expected by chance given the number of neurons recorded (permutation test, $P < 0.0001$; Fig. 2B bottom). Figure 2B,C illustrates representative neurons recorded from the same animal that responded uniquely to reward received by oneself, any other agent, or a specific-other agent as well as their population dynamic.

Neurons that responded to receipt of reward by specific other agents were largely distinct from those that responded to the animal's own receipt of reward. Overall, 26.4% (n = 138) of the neurons displayed a change in their firing activity when reward was received by the recorded animal itself. However, most of these displayed little response to the other's reward, with only 14 neurons displaying a change in their activity to both self-reward and specific-other reward ($\chi^2(1) = 9.7$, $P = 0.001$; Fig. 2B bottom); results that were largely consistent across statistical analyses (Table S1). The responses of these neurons to receipt of reward, by comparison, did not reflect more generalized processes such as a negative reward prediction error. Because any of the three agents could function as actors, we could dissociate signals that reflected another agent's observed receipt of reward from its expectancy (i.e., the animals had no expectancy of reward and therefore held no reward prediction error when they were the actor) (36). Consistently, we found that neurons which responded to another specific agent's reward displayed no difference in response based on whether a putative negative reward prediction error was present (n = 50; rank-sum test, $Z = 0.86$, $P = 0.38$; Fig. 2D bottom), confirming that they responded selectively to the specific agents receiving reward.

Specificity of neuronal responses to social context and identity

The responses of these neurons to receipt of reward by other specific agents were also robust to differences in their physical locations. It could be argued, for example, that neurons that responded to the other agents may have simply encoded the location of reward rather than the specific agents receiving it. Therefore, to control for this possibility, we switched the locations of the other two animals halfway through the sessions as the primates continued to perform the task. However, we find that only 4 of the neurons which displayed selectivity to the agents receiving reward also displayed selectivity to reward location ($\chi^2(1) = 23.7$, $P = 1.13 \times 10^{-6}$). More notably, the neurons that displayed selective responses to particular agents before the switch continued to show similar responses to those same agents afterward (rank-sum test, $Z = 2.86$, $P = 0.004$; Fig. 2E right, Fig. S3B). Figure 2E left illustrates the responses of one such neuron before vs. after the switch, with the vertex of the triangle representing maximal neuronal activity for a specific monkey.

We also considered the possibility that these neurons may have responded to lower-level sensory features such as the others' faces (7, 37) independently of their social interaction. For example, simply looking at the other agents may have elicited similar responses. Therefore, to test for this, we examined a separate inter-trial control period in which no task was performed but in which the primates were allowed to gaze freely at the other two monkeys (21). We find, however, that even when directly viewing the other animals under this control, only 6.1% (n =15 of 244) of the neurons distinguished between which specific agent the primates were looking at (i.e., based on the recorded animal's eye positions; Fig. 2F). More notably, only 2 of these neurons overlapped with those that responded to the specific agents receiving reward ($\chi^2(1) = 21.6$, $P = 3.4 \times 10^{-6}$), and the degree to which the recorded animal's gaze modulated these neurons' activities was negligible (rank-sum test, $Z = -4.89$, $P = 9.7 \times 10^{-7}$; Fig. 2F). Therefore, unlike interconnected areas such as the temporal lobe (8, 9, 38, 39), neurons in this area did not reflect information about the others' face.

Finally, to confirm that neuronal responses to the other agents reflected the social context of their interaction, we recorded from an additional 403 neurons while the recorded primates performed the non-social control (12). As before, the two other animals were replaced with distinct inanimate totems while we yoked the distribution of reward from a past session (Fig. S2E). Unlike the main task, however, we find that 0.6% (n = 3) of the neurons changed their activity based on which specific totem was given reward and at a proportion was significantly lower from that observed before ($\chi^2(1) = 40.6$, $P = 1.8 \times 10^{-6}$; Fig. 2G, inset). Moreover, these differences in neuronal response were not associated with a change in mean neuronal activity (rank-sum test, $P > 0.5$) and, as noted above, we observed no change in behavioral reaction and movement times to suggest a difference in engagement or attention. Lastly, 0% (n = 0 of 83) of neurons were modulated when a specific other monkey received reward, but when no actor offered it (Reward Dissociation Control; Methods), together suggesting that the activities of these neurons were indeed dependent on the social context of the animals' interactions.

Neuronal representations of agent-specific actions and group interaction

In order for the primates to effectively interact within these groups, it was necessary for them not only to know who received the reward but also who was the actor that offered it. In our task, any of the three primates could be the actor on a given trial (if they were not the actor on the previous trial) and, in turn, could offer a reward to either of the other two agents. Thus, for example, Other Monkey 1 may be the actor in one trial and could choose to offer a reward to Other Monkey 2, or Other Monkey 2 may be the actor and could choose to offer a reward to Other Monkey 1. Here, we find that 8.8% (n = 46) of the neurons distinguished between whether Other Monkey 1 or Other Monkey 2 was the actor (Fig. 2C, 3A), meaning that they responded differently based on which agent offered reward. Moreover, when considering their group interactions, we find 11.1% (n = 58) of neurons changed their activity based on which specific animal the actor offered reward to (Fig. S5B–C). Figure 3A bottom illustrates such a representative cell; displaying a difference in activity based on whether Other Monkey 1 or Other Monkey 2 offered reward to the recorded animal but displaying little or no difference in activity for any other interaction.

Next, given these observations, we asked whether and to what degree these neural populations were predictive of interactions within the group and the identities of the specific agents involved on a *per*-trial basis. Here, we trained multi-class decoders on the neuronal responses of 80% of matched-trials from all recorded cells and tested the model's performance in the held-out sample (performed in 1-second windows advanced in 0.1-second intervals; (21)). We find that the identity of the specific actor could be decoded with an accuracy of 81.7% ± 2.9% (mean ± 95% CI) prior to choice selection (Fig. 3B, left), meaning that these neurons could be used to predict which specific agent offered reward. As the trial progressed, however, prediction accuracy increased for the specific identity of the reward recipient; with a decoding accuracy of 70.1% ± 2.8% once the actor made their choice. More notably, both the actor (72.8% ± 3.1%) and recipient of reward (72.75% ± 3.3%) could be accurately decoded even when confining our analyses to interactions between Other Monkey 1 and Other Monkey 2 (i.e., excluding the recorded animal as the agent; Fig. S6A).

Collectively, the activities of these neurons held detailed representations about specific interactions within the group. Peak decoding accuracy for agent-specific interactions was 40.4% ± 0.78% and significantly higher than chance shortly before reward was acquired (chance = 16.6% given the number of possible actor-recipient combinations, $P < 0.01$, permutation test; Fig. 3B right). The highest decoding performances were for interactions that specifically resulted in reward for the recorded animals (52%); meaning that they were predictive of who specifically offered a reward to them. Decoding accuracy for the agent to whom the recorded animals offered reward was slightly lower at 44%. Similar decoding performances were also observed for 'mixed-selectivity' neurons (40) that encoded information about both the specific actor and recipient of reward (Fig. S6B–C; (21)) as well as when comparing decoding performances across spatial locations (Fig. S7A and Supplemental Material). Decoding accuracy for more basic sensorimotor variables such as movement direction ($p > 0.2$ permutation test; Fig. S7B) or the direction of gaze ($p > 0.2$ permutation test), on the other hand, was at chance. The activities of these neurons, therefore, appeared to hold detailed information about which specific individuals in the group interacted with whom.

Effect of past interactions on neuronal response and upcoming decisions

Last, we asked how the neural population responses may relate to the animal's own decisions. To interact effectively, the actor had to take into consideration past interactions with other agents when making their decisions. Here, we find that 15.2% (n = 79) of the neurons displayed a difference in response on the current trial (t) based on who was the specific actor on the prior trial (t-1; $P < 0.01$) whereas 6.7% (n = 35) displayed a difference in response based on the past trials' specific recipient of reward (Fig. 3C right). From all population neurons, we could decode information about past (t-1) interactions within the group on trials (t) in which the monkey was the current actor with an accuracy of 52.8% ± 1.9% (H_0 = 25% chance; $P < 0.01$, permutation test; Fig. 3C left). When further accounting for the actor's own current decisions, these neural populations could predict the animal's upcoming choices contingent on the other agent's past actions with an accuracy of up to 49.5 ± 1.0% (H_0 = 25% chance, permutation test, $p = 0.005$; Fig. 4A). In other words, the activities of these neurons could be used to accurately predict whether the recorded animal will reciprocate or retaliate in response to the other's past choices (i.e., rather than simply based on any receipt of reward; irrespective of which social agent offered it). Overall, peak decoding accuracy for whom the recorded animals will offer a reward before their motor responses was 74.3 ± 1.4% and significantly higher than chance (H_0 = 50% chance, Fig. 4B). Figure 4C further illustrates these decoding performances across the different group interactions and how they relate the animal's own choices. Taken together, these dmPFC neurons therefore appeared to predict the animal's upcoming decisions based on past interactions with specific agents in their group.

Effect of micro-stimulation on group decisions and agent-specific interactions

Next, based on these observations, we asked whether and what causal role the dmPFC may have played in the animal's decisions during these group interactions. As noted above, the primates reciprocated past offers of reward from specific individuals to enact mutually beneficial interactions. They also displayed evidence of retaliation against individuals who did not; behaviors that are often naturally seen within primates' groups (41–43). Therefore, to further study this question, we used event-triggered stimulation delivered bilaterally to the dmPFC (200 Hz, 0.1 mA over 2 seconds; from lock on to trial start; Fig. 5A) as the primates performed the same task as before. To allow for control comparison, stimulation was given to the animals on randomly interleaved trials divided equally between those in which the stimulated primate was the actor and observer (21).

Before proceeding with the main task, we confirmed that stimulation did not have nonspecific effects on the animal's motoric behavior. Overall, we find that the animals displayed similar reaction times (sign rank test, $Z = 0.07$, $P = 0.94$) and a similar likelihood of selecting one direction over another (sign rank test, $Z = 0.04$, $P = 0.96$) when comparing stimulated vs. non-stimulated trials. We also confirmed that stimulation did not disrupt the monkey's ability to make appropriate choices. Here, in a separate control task, we allowed the animals to deliver reward to themselves by moving the turntable handle (i.e., without other agents; Fig. S1D; (21)), but find that stimulation did not affect the animals' performances (100% correct performance for both stimulated, n = 27, and non-stimulated, n = 24, trials). Next, we considered the possibility that stimulation might affect their memory of past events or more complex strategic behaviors by evaluating the animal's likelihood of enacting 'Win-Stay-Lose-Switch' strategies. These canonical strategies represent decisions in which the animal repeated the last choice made on the prior trial if they received a reward irrespective of who offered it (21). However, we again find that stimulation did not affect the animals' likelihood of enacting this strategy (F(2,15) = 0.67, P = 0.52; Fig. S8A). Lastly, we verified that stimulation did not affect the animals likelihood of looking at particular animals after receiving reward (F(2,119) = 0.145, P = 0.86, for stimulation condition; Fig. S8B); together confirming that the effect of stimulation was specific.

Finally, based on these findings, we considered the primates' interactions with the other group members. Evaluating the primates' behavior on trials in which they were the actor during the main task, we find that stimulation led to an 11.8% drop in their likelihood of reciprocating past offers of reward from another specific agent (F(2,15) = 4.8, P = 0.02). In other words, stimulation diminished their propensity to offer reward to the specific agent from which they received a reward on a previous trial, an effect that was consistent when examined across the different agents (post-hoc test, P = 0.013; Fig. 5B). By contrast, stimulation had little effect on the animal's likelihood of retaliating in response to past negative interactions or use tit-for-tat strategy (F(2,15) = 0.43, P = 0.65; F(2,15) = 0.91, P = 0.42; respectively, Fig. 5B) and had no effect on response variability (rank-sum, P > 0.2 for reciprocity, retaliation and both strategies) to suggest a generalized disruption of behavior. More notably, stimulating when the animal observed the other's choices did not affect the animal's likelihood of reciprocating past offers of reward (post-hoc test, P = 0.80); together suggesting that stimulation had a temporally selective effect on the primates' ability to enact mutually positive interactions with specific agents in their groups.

Discussion

Most animals, including humans, live within social groups in which they interact with many other group members. The basic cellular processes that precisely underlie group behavior or by which neurons represent specific group interactions, however, have remained poorly understood. Here, we identify neurons in the primate dmPFC that responded selectively to the actions and outcomes of specific group members. By recruiting different subsets of cells to represent the specific actions and outcomes of each individual, these neural populations encoded information not only about the behavior of individuals but also the directionality of the interactions between them, even when the recorded animals themselves were not involved. Together, these findings identify cells in the primate dmPFC that encode the 'agency identity' of others, meaning that they encode information about the behavior of specific individuals. Such computations are essential for effective social behavior.

Another notable finding was that many of the neurons encoded information not only about the actions and outcomes of specific individuals within the group but also their past behaviors. Moreover, neural predictions of the animal's own upcoming decisions were modulated by the other agent's past actions; suggesting that animal's upcoming decisions to reciprocate or retaliate were influenced by their past interactions with specific group members. Consistent with these observations, stimulation of the dmPFC had a selective effect on the animal's ability to reciprocate past favorable interactions with specific individuals while having little effect on other aspects of their decisions, social viewing preferences, or motoric responses. These observations that stimulation of the dmPFC impacted social choices but not social orienting behaviors indicate potentially different parallel systems underlying these behaviors. Together, they also suggest that neuronal activity in this area is necessary for mediating mutually beneficial interactions with specific individuals within these social groups.

Collectively, these findings begin to elucidate the neuronal computations that underlie social group interactions and the role that the dmPFC may play in this process. They also identify neurons capable of encoding the actions, outcomes, and past behavior of specific agents. Given its broad connectivity with temporal regions such as the fusiform gyrus, amygdala that are known to respond to the identities and facial features of others (6, 37, 44), the dmPFC may be particularly well suited for holding representations of specific group members and mediating mutually favorable interactions. Moreover, the rich representation of agency-specific action and reward recipient in dmPFC could be potentially used for monitoring the consequence of social actions between specific group members, a hypothesized function of the medial prefrontal cortex (36, 45, 46). Together with other areas proposed to be involved in social cognition (47, 48), the dmPFC may play a core role in orchestrating the interactive social behavior of groups.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We thank Mohsen Jamali, Yoav Kfir, Arjun Khanna, Maimon Rose, Michael Yartsev, Fabian Grabenhorst and Yuriria Vazquez for comments on the manuscript; Kejia Hu and Jacob Grondin for animal training.

Funding:

R.B.M is funded by NARSAD YI Award 25066, and the MGH Fund for Medical Discovery. A.J.W is an HHMI Medical Research Fellow. Z.M.W is funded by NIH R01HD059852, NIH R01NS091390 and NIH U01NS123130.

Data and materials availability:

Data will be deposited on the Figshare repository (49).

References and Notes:

1. Clutton-Brock T, Cooperation between non-kin in animal societies. Nature 462, 51–57 (2009). [PubMed: 19890322]
2. Sliwa J, Duhamel J-R, Pascalis O, Wirth S, Spontaneous voice–face identity matching by rhesus monkeys for familiar conspecifics and humans. Proceedings of the National Academy of Sciences 108, 1735–1740 (2011).

3. Hamilton WD, The evolution of altruistic behavior. The American Naturalist 97, 354–356 (1963).

4. Nowak MA, Sigmund K, Tit for tat in heterogeneous populations. Nature 355, 250 (1992).

5. Nowak MA, Five rules for the evolution of cooperation. Science 314, 1560–1563 (2006). [PubMed: 17158317]

6. Meyers EM, Borzello M, Freiwald WA, Tsao D, Intelligent information loss: the coding of facial identity, head pose, and non-face information in the macaque face patch system. J. Neurosci 35, 7069–7081 (2015). [PubMed: 25948258]

7. Chang L, Tsao DY, The Code for Facial Identity in the Primate Brain. Cell 169, 1013–1028 e1014 (2017). [PubMed: 28575666]

8. Landi SM, Freiwald WA, Two areas for familiar face recognition in the primate brain. Science 357, 591–595 (2017). [PubMed: 28798130]

9. Landi SM, Viswanathan P, Serene S, Freiwald WA, A fast link between face perception and memory in the temporal pole. Science, (2021).

10. Haroush K, Williams ZM, Neuronal prediction of opponent’s behavior during cooperative social interchange in primates. Cell 160, 1233–1245 (2015). [PubMed: 25728667]

11. Rudebeck PH, Buckley MJ, Walton ME, Rushworth MF, A role for the macaque anterior cingulate gyrus in social valuation. Science 313, 1310–1312 (2006). [PubMed: 16946075]

12. Báez-Mendoza R, Harris CJ, Schultz W, Activity of striatal neurons reflects social action and own reward. Proc. Natl. Acad. Sci. U. S. A 110, 16634–16639 (2013). [PubMed: 24062436]

13. Livneh U, Resnik J, Shohat Y, Paz R, Self-monitoring of social facial expressions in the primate amygdala and cingulate cortex. Proc. Natl. Acad. Sci. U. S. A 109, 18956–18961 (2012). [PubMed: 23112157]

14. Grabenhorst F, Baez-Mendoza R, Genest W, Deco G, Schultz W, Primate Amygdala Neurons Simulate Decision Processes of Social Partners. Cell 177, 986–998 e915 (2019). [PubMed: 30982599]

15. Sliwa J, Freiwald WA, A dedicated network for social interaction processing in the primate brain. Science 356, 745–749 (2017). [PubMed: 28522533]

16. Axelrod R, Hamilton WD, The evolution of cooperation. Science 211, 1390–1396 (1981). [PubMed: 7466396]

17. Perrodin C, Kayser C, Logothetis NK, Petkov CI, Auditory and visual modulation of temporal lobe neurons in voice-sensitive and association cortices. J. Neurosci 34, 2524–2537 (2014). [PubMed: 24523543]

18. Maestripieri D, Macachiavellian intelligence: how rhesus macaques and humans have conquered the world (University of Chicago Press, 2008).

19. Testard C et al. , Rhesus macaques build new social connections after a natural disaster. Curr. Biol 31, 2299–2309.e2297 (2021). [PubMed: 33836140]

20. Kapsalis E, Berman CM, Models of Affiliative Relationships among Free-Ranging Rhesus Monkeys (Macaca mulatta)II. Testing Predictions for Three Hypothesized Organizing Principles. Behaviour 133, 1235–1263 (1996).

21. Materials and methods are available as supplementary materials at the Science website.

22. West SA, Griffin AS, Gardner A, Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. J. Evol. Biol 20, 415–432 (2007). [PubMed: 17305808]

23. Deaner RO, Khera AV, Platt ML, Monkeys pay per view: adaptive valuation of social images by rhesus macaques. Curr. Biol 15, 543–548 (2005). [PubMed: 15797023]

24. Klein JT, Deaner RO, Platt ML, Neural correlates of social target value in macaque parietal cortex. Curr. Biol 18, 419–424 (2008). [PubMed: 18356054]

25. Ballesta S, Duhamel J-R, Rudimentary empathy in macaques’ social decision-making. Proceedings of the National Academy of Sciences 112, 15516–15521 (2015).

26. Flombaum JI, Santos LR, Rhesus monkeys attribute perceptions to others. Curr. Biol 15, 447–452 (2005). [PubMed: 15753039]

27. Hayashi T et al. , Macaques Exhibit Implicit Gaze Bias Anticipating Others’ False-Belief-Driven Actions via Medial Prefrontal Cortex. Cell Rep 30, 4433–4444 e4435 (2020). [PubMed: 32234478]

28. Dal Monte O, Noble P, Costa VD, Averbeck BB, Oxytocin enhances attention to the eye region in rhesus monkeys. Frontiers in Neuroscience 8, (2014).

29. Krupenye C, Kano F, Hirata S, Call J, Tomasello M, Great apes anticipate that other individuals will act according to false beliefs. Science 354, 110–115 (2016). [PubMed: 27846501]

30. Yoshida K, Saito N, Iriki A, Isoda M, Representation of Others' Action by Neurons in Monkey Medial Frontal Cortex. Curr. Biol 21, 249–253 (2011). [PubMed: 21256015]

31. Yoshida K et al. , Single-neuron and genetic correlates of autistic behavior in macaque. Sci Adv 2, e1600558 (2016). [PubMed: 27679817]

32. Chang SW, Gariepy JF, Platt ML, Neuronal reference frames for social decisions in primate frontal cortex. Nat. Neurosci 16, 243–250 (2013). [PubMed: 23263442]

33. Rilling J et al. , A neural basis for social cooperation. Neuron 35, 395–405 (2002). [PubMed: 12160756]

34. Behrens TE, Hunt LT, Woolrich MW, Rushworth MF, Associative learning of social value. Nature 456, 245–249 (2008). [PubMed: 19005555]

35. Tomlin D et al. , Agent-specific responses in the cingulate cortex during economic exchanges. Science 312, 1047–1050 (2006). [PubMed: 16709783]

36. Matsumoto M, Matsumoto K, Abe H, Tanaka K, Medial prefrontal cell activity signaling prediction errors of action values. Nat. Neurosci 10, 647–656 (2007). [PubMed: 17450137]

37. Perrett D et al. , Neurones responsive to faces in the temporal cortex: studies of funcional organization, sensitivity to identity and relation to perception. Human Neurobiol 3, 197–208 (1984).

38. Carmichael S, Price JL, Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. J. Comp. Neurol 363, 615–641 (1995). [PubMed: 8847421]

39. Sallet J et al. , The Organization of Dorsal Frontal Cortex in Humans and Macaques. The Journal of Neuroscience 33, 12255–12274 (2013). [PubMed: 23884933]

40. Rigotti M et al. , The importance of mixed selectivity in complex cognitive tasks. Nature 497, 585–590 (2013). [PubMed: 23685452]

41. de Waal FBM, Coping with social tension: Sex differences in the effect of food provision to small rhesus monkey groups. Anim. Behav 32, 765–773 (1984).

42. Jensen K, Call J, Tomasello M, Chimpanzees are vengeful but not spiteful. Proc. Natl. Acad. Sci. U. S. A 104, 13046–13050 (2007). [PubMed: 17644612]

43. Jensen K, Punishment and spite, the dark side of cooperation. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 365, 2635–2650 (2010). [PubMed: 20679108]

44. Hitti FL, Siegelbaum SA, The hippocampal CA2 region is essential for social memory. Nature 508, 88–92 (2014). [PubMed: 24572357]

45. Ito S, Stuphorn V, Brown JW, Schall JD, Performance monitoring by the anterior cingulate cortex during saccade countermanding. Science 302, 120–122 (2003). [PubMed: 14526085]

46. Yoshida K, Saito N, Iriki A, Isoda M, Social error monitoring in macaque frontal cortex. Nat. Neurosci 15, 1307–1312 (2012). [PubMed: 22864610]

47. Gangopadhyay P, Chawla M, Dal Monte O, Chang SWC, Prefrontal–amygdala circuits in social decision-making. Nat. Neurosci 24, 5–18 (2021). [PubMed: 33169032]

48. Wittmann MK, Lockwood PL, Rushworth MFS, Neural Mechanisms of Social Cognition in Primates. Annu. Rev. Neurosci 41, 99–118 (2018). [PubMed: 29561702]

49. Báez-Mendoza R; Mastrobattista EP; Wang AJ; Williams Z (2021): Social Agent Identity Cells in the Prefrontal Cortex of Interacting Groups of Primates. figshare. Dataset 10.6084/m9.figshare.15102027

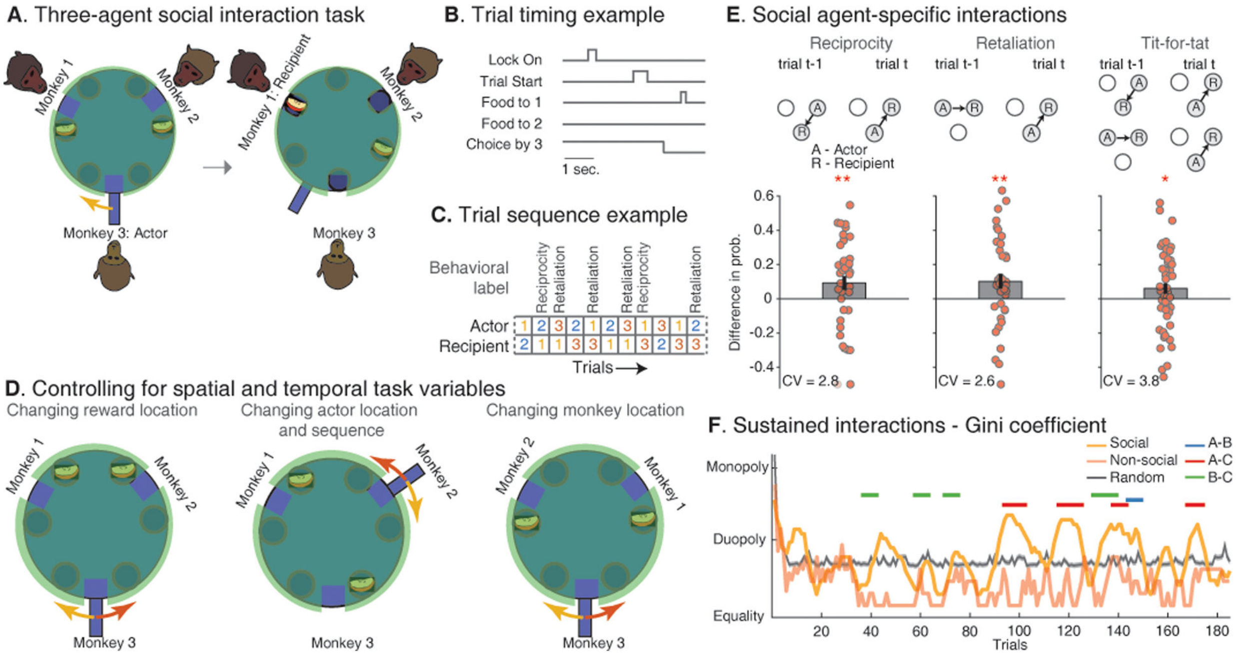

Fig. 1. Three-agent task for testing partner-specific interactions in Rhesus macaques.

(**A**) Groups of three monkeys sat around a custom-made turntable apparatus that allowed them to interact with each other through food allocation. All monkeys could observe the initial food location through a transparent cover (shown in green), the actor's choice, and the reward recipient. Turntable movement, together with food location, determined the reward recipient (Fig. S1). (**B**) Example timing of events for trial shown in A. In this trial, monkey 3 (referred to here as 'Choice by 3') offered a reward to monkey 2 (referred to here as 'Food to 2'). (**C**) The animals interacted with each other over multiple trials, with the actor on each trial being selected in a pseudo-random fashion. The actor could engage in reciprocity or retaliation based on what the previous actor chose. (**D**) Control measures were used to dissociate the identities of the different agents from variables such as the direction of movement (left panel), the role of each agent (center panel), or the monkeys' spatial locations (right panel). (**E**). Illustration of trial combinations in which the animals displayed reciprocation, retaliation, and tit-for-tat behavior. Arrows show who the actor offered a reward. The animals displayed reciprocation, retaliation, and tit-for-tat behavior with specific individuals at probabilities that were significantly higher than expected from chance (* $P < 0.05$, ** $P < 0.01$; Coefficient of Variation (CV)). Bars are the probability of reciprocating compared to not reciprocating ± SEM. Each point depicts an individual's probability within a particular session (Fig. S2). (**F**) Gini coefficient illustrates the distribution of reward (dark orange) during a representative session. The highlighted horizontal lines illustrate transient duopolies (permutation test, $p < 0.05$). For comparison, the distribution of reward expected from chance (gray) and in a representative non-social session (light orange) are displayed separately.

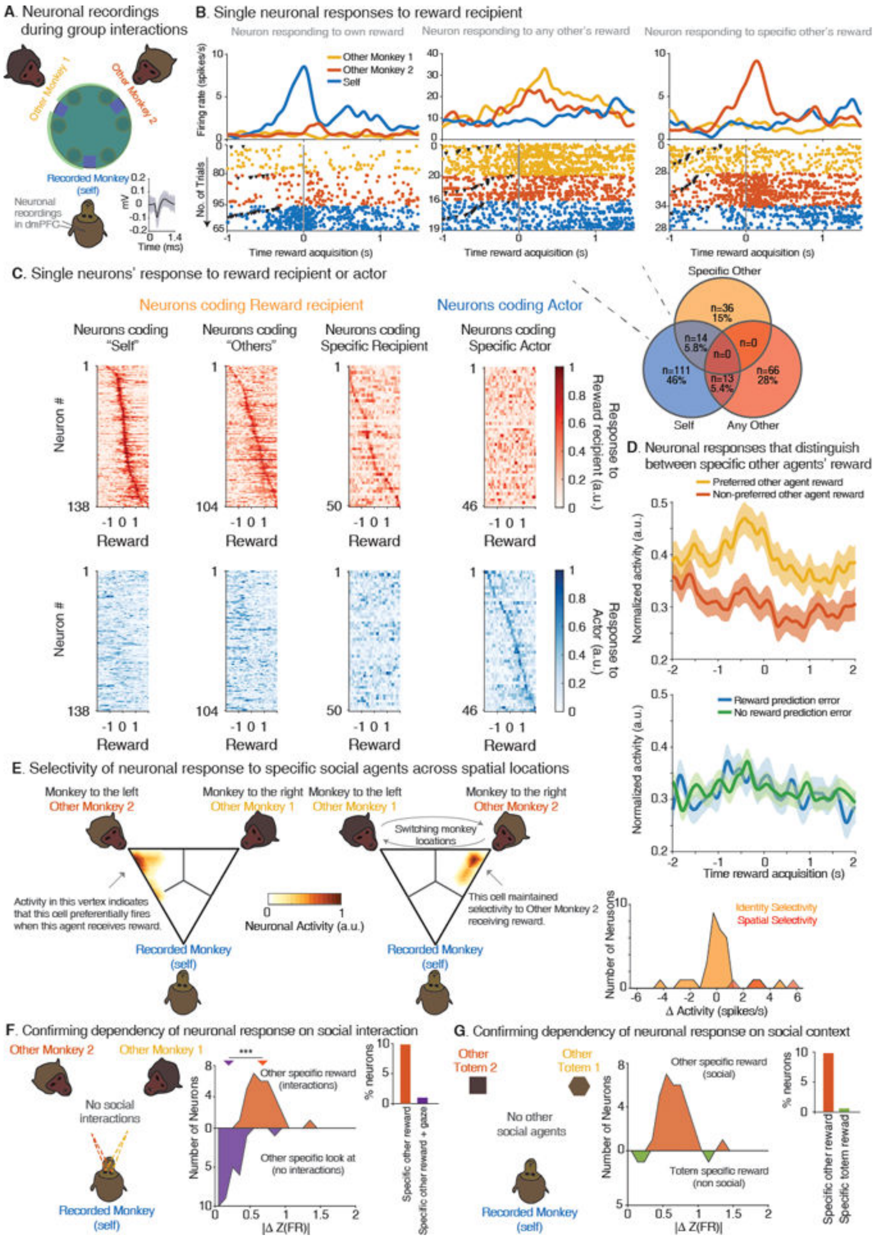

Fig. 2. Selectivity of neurons to specific social agents during group interactions.

(**A**). The monkey undergoing neuronal recordings from the dmPFC within each session was referred to as ‘Self’, and the two other monkeys as ‘Other Monkey 1’ and ‘Other Monkey 2’. Recorded neurons displayed stable waveform morphology (inset; Fig. S4). (**B**) Peri-event time histogram and raster examples of neurons that displayed changes in their activities when particular agents within the group received a reward. The inverted black triangles mark when the actor chose. Venn diagram of neurons that displayed response selectivity to reward recipient agency. (**C**). Heatmap of single neurons’ response to reward recipient (top, red) and actor (bottom, blue) aligned to the timing of reward acquisition. Only neurons with significant modulation are shown (ANOVA, P < 0.01). (**D**) Top, normalized population activity of neurons encoding ‘specific-other-reward’ to the preferred and the non-preferred other monkey. Bottom, the same neuronal population as above but parsed by the absence or presence of a possible reward prediction error for ‘self’. (**E**) The locations of Other Monkey 1 and Other Monkey 2 were switched halfway in the session to test the selectivity of neuronal responses to specific agents independently of their spatial locations. Heatmap of neuronal activities on a ternary plot before and after the switch of a representative neuron. Here, each vertex represents maximal neuronal activity for a particular monkey. The color code provides the density of activity across trials. The particular neuron displayed here responded almost exclusively to receipt of a reward by Other Monkey 2 both before and after switching its location relative to the recorded animal. Right, histogram of neurons that retained a preferential response to a specific agent (n = 34, orange), and neurons (n = 4, red) that signaled both reward receipt and location. (**F**) To test that neuronal responses were not explained by looking at others’ faces, we tracked the recorded animals’ eye positions during an inter-trial period. Middle, distribution of neurons’ activity displaying social agent-specific reward responses based on whether others received reward (top, orange) or the recorded animal looked at others during the inter-trial period (bottom, blue; normalized to the preferred animal). Right, proportion of cells (***, P < 0.0001). (**G**) The primates performed the same task but in the absence of social agents to test the effect of social context on neuronal responses. Middle, distribution of the absolute normalized difference in firing rates of individual-specific reward neurons based on social agents (orange), or non-social agents (green). Right, the proportion of neurons relative to the total number of recorded neurons on each task.

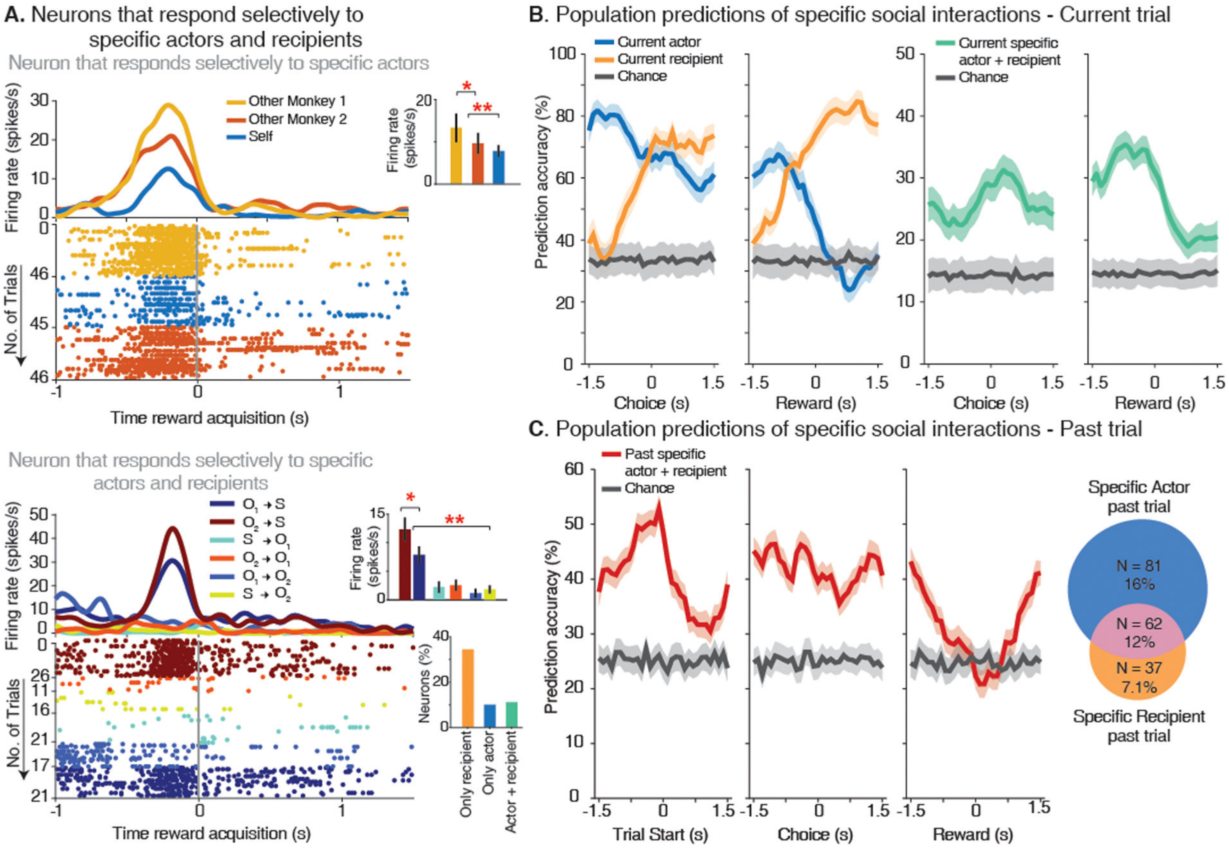

Fig. 3. Neural population predictions of specific interactions within the group.

(**A**) Top, a neuron that displayed a change in its activity based on whether Other Monkey 1 or Other Monkey 2 was the actor. Bottom, a neuron that displayed a change in its activity based on whether Other Monkey 1 or Other Monkey 2 were the actor but only when they specifically offered a reward to the recorded. Insets, average firing rate for each condition during a 1s time window centered at 0.3 s before the reward was acquired. Lower inset, proportions of neurons encoding specific agent receiving a reward, the specific actor offering a reward and the combination of the specific actor and recipient across all possible interaction types. (**B**) Left, decoding performance for specific actor and recipient, separately. Right, decoding performance for specific interactions in which both the actor and recipient of reward were decoded on a trial-by-trial basis. Multi-class one-vs-all decoders were trained with 80% of trials and tested on the remainder 20% trials (1 s window advanced in 0.1 s intervals). The colored curves indicate mean prediction accuracy on test trials (± 95% confidence interval). (**C**) Left, decoding performance for the combination of specific actor and reward recipient in the previous trial when the recorded animal is the actor in the current trial and, therefore, planning their choice. Right, Venn diagram of the number of neurons displaying selectivity for the specific actor (blue) and the specific recipient (orange) in the past trial (* $P < 0.05$, ** $P < 0.01$).

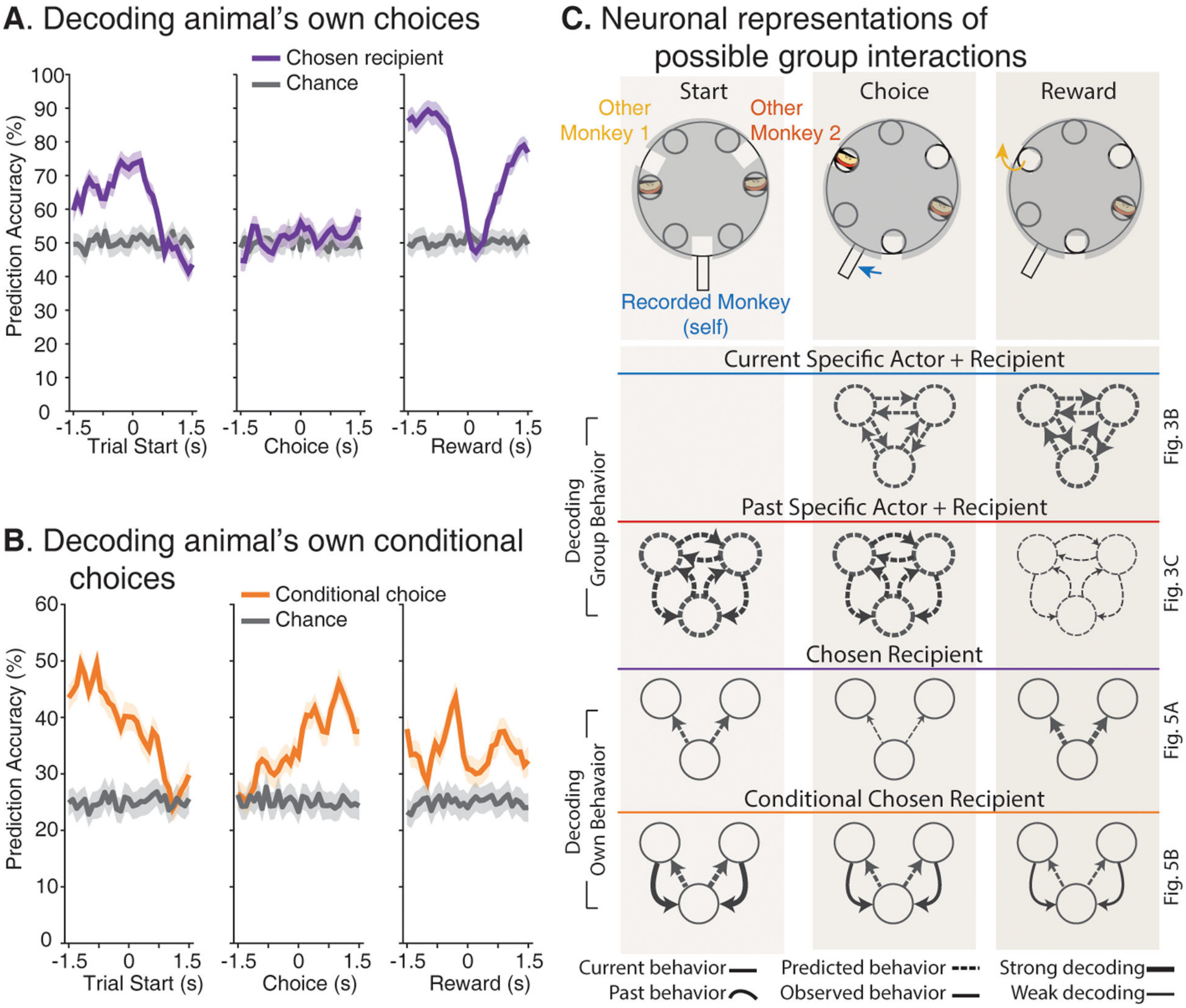

Fig. 4. Dependency between past interactions and predictions of upcoming choices.

(**A**) Neuronal responses accurately predicted the animal's own upcoming choices before making their motor selection. (**B**) The animal's past interactions modulated neuronal predictions of the animal's upcoming choice. By considering both the other monkeys' choices and the recorded animals' current choice combinations, the curves here reflect neuronal population predictions contingent on the other's past actions. (**C**) Summary of decoding results. Each column corresponds to one distinct epoch and each row to the relevant information decoded. The arrows reflect the actor (circle) offering a reward to another agent. Each set of arrows reflects the possible combinations of current/past behavior, predicted/observed behavior, and the relative strengths of decoding. Thus, for example, thick arrows indicate that those specific interactions could be highly accurately decoded from neural population response whereas thin arrows indicate that decoding accuracy for those interactions was poor when compared to chance. The relevant figures for each panel are shown on the right to allow for ease of comparison.

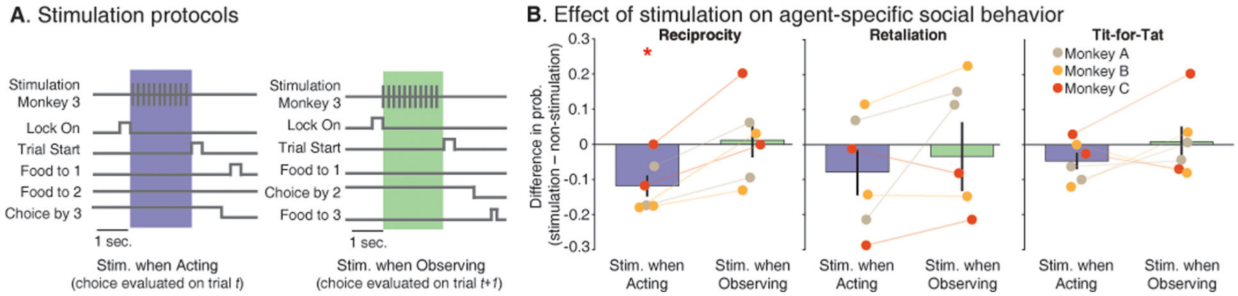

Fig. 5. Effect of stimulation in the dmPFC on group interactions and its selectivity.

(**A**) Brief event-triggered electrical stimulation was delivered bilaterally to the dmPFC (200 Hz, 0.1 mA over 2 seconds, given between locking of the apparatus and trial start) as the primates performed the same task as before. Stimulation was given either when the animal was the actor (blue background) or when they were the observer (green background) for control comparison. (**B**) The bar plot provides the mean difference in probability of reciprocating, retaliating, or using the tit-for-tat strategy on stimulated vs. non-stimulated (baseline) trials ± SEM. Each point depicts individual sessions color-coded by the animal receiving stimulation. Additional controls used to confirm that stimulation did not affect more basic motoric behavior or cognitive processes such as attention are described in the Main Text. For specific comparisons * P < 0.05.