ABSTRACT

Title of Dissertation: **CORTICAL CONTRIBUTIONS TO A COMBINED APPETITIVE-AVERSIVE SOCIAL OUTCOME TASK**

Kevin Nicholas Schneider, Doctor of Philosophy, 2021

Dissertation directed by: Dr. Matthew Roesch, Professor, Department of Psychology, Neuroscience and Cognitive Science Program

Learning through the emotional states of others is a critical skill for navigating our complex social environments, which is why it is a focal point of investigation in social neuroscience. Significant advances have been made in recent years, highlighting cortical brain regions where the transfer and processing of socially-derived affective information may be taking place. Among these regions are the anterior cingulate cortex (ACC) and orbitofrontal cortex (OFC), known for their roles in attention and valuation during decision-making tasks, respectively. However, social decision-making studies have typically focused on either reward or punishment as the outcome valence, making it difficult to determine the social specificity of neural contributions observed. Using a social task that manipulates reward and shock

within the same experiment, I recorded single-unit activity from ACC and OFC in rats. I found that during the task, ACC activity shared responses for reward and shock outcomes, suggesting it encoded socially-derived information in the service of attention. OFC neurons showed responses to self and vicarious reward outcomes, consistent with previous work in primates. Interestingly, OFC also encoded the positive value of the rats' approach to their conspecific following foot-shock delivery, which leads to stress relief and a reduced fear response. Thus, in this task, ACC and OFC encoded other-related outcome information with respect to the self, in accordance with their nonsocial functions, suggesting that during social decision-making tasks, internal state goals are prioritized when outcomes to the self are at stake.

CORTICAL CONTRIBUTIONS TO A COMBINED APPETITIVE-AVERSIVE SOCIAL OUTCOME TASK

by

Kevin Nicholas Schneider

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
Professor Matthew Roesch, Chair
Professor Melissa Caras, Dean's Representative
Professor Anna Xuan Li
Professor Jens Herberholz
Professor Quentin Gaudry

Dedication

I would like to dedicate this dissertation first to my late grandfather Roberto, my first scientific inspiration and one of the reasons I set out on this journey. My memory of going to your lab as a child has stayed with me throughout my life.

I would also like to dedicate this dissertation to and thank my partner Zefanya for her incredible love and support throughout this endeavor. Your patience and company helped me stay grounded and overcome the toughest of times, every time. I would also like to thank my parents Mercedes and Jorge for their ever-present encouragement and moral support. I also want to thank family and friends too numerous to list here, but in particular my friend Luis for his support and patience throughout the stress and business. I love you all.

Acknowledgements

This dissertation would not exist without the many supportive mentors, advisors and colleagues I am lucky to have met. I want to thank my first mentors Dr. Graham Cousens and Dr. Steven Leiser, for introducing me into the amazing world of electrophysiology. Thank you for instilling in me and encouraging my curiosity for the scientific pursuit of knowledge, driving me to explore a career in neuroscience. I want to thank Dr. Andrew Delamater as well, who showed me the attentive creativity required for experimental design in behavior. I also wish to thank members of the Araneda lab, including Dr. Ricardo Araneda, Dr. Ruilong Hu and Pablo Villar del Río, all of whom contributed significantly to my development as an independent scientist. I give heartfelt thanks to my PhD advisor Dr. Matthew Roesch, for his invaluable mentorship while I explored the boundaries of this project, and his patient guidance when I seemed to lose my course. Your positivity and curiosity for neuroscience was in every conversation we shared and is a constant inspiration in my own pursuit of knowledge about the brain.

Last but certainly not least, many thanks go to current and past members of the Roesch lab, without whom this dissertation would not have been possible. In alphabetical order, I would like to thank: Dr. Adam Brockett, Heather Pribut, Xavier Sciarillo, Stephen Tennyson, Daniela Vázquez and Clarissa Xia.

Table of Contents

List of Figures

Chapter 1. Introduction

Perceiving, learning from and reacting to the environment is the perennial concern of all animals. The brain, even in its less complex iterations, serves as a yet another biological tool to solve the problems of safety, nourishment and reproduction. With social animals including humans, monkeys and rats, came new skills that allow us to learn from the experiences of others, in order to both avoid personal risk and determine friend from foe. In many cases these abilities are critical for survival, preventing us from eating poisonous food (that our friend may have unfortunately tried), predicting an aggressive encounter or seeking available mates in extremely tight time-windows. On other levels, success in social interactions usually leads to social acceptance as a member of the group, whereas failed interactions will lead to rejection, motivating the adaptation to and preference for certain social behaviors. To succeed in them, social interactions require, broadly: the accurate perception of social cues, the integration of information from these cues and their outcomes, and a response utilizing said information. Properly executing these steps is known as empathizing, and people who are particularly good at them are considered effective empathizers. Typically, our brains are able to perform these tasks well enough with some variability, but the impairment of social skills in disorders such as those in the Autism Spectrum (ASD; Bachevalier & Loveland, 2006; King et al., 2006; Lockwood, 2016) leads to deep disruptions in the lives of those living with them. Unfortunately, despite recent advances, the neural processes underlying these abilities are still not well-understood. Thus, current treatment options for these neuropsychiatric disorders are imprecise and hardly effective. Elucidating how each of the aforementioned

functional steps in empathizing take place in the brain is key to better understanding their etiology, hopefully leading to the development of more targeted therapeutic interventions. One inquiry is of particular importance towards these investigations, namely whether the brain developed social-specific mechanisms or repurposed existing ones in order to deal with social interactions (Gangopadhyay et al., 2020; Lockwood et al., 2020). Below I describe how empathetic processes have been defined for study. In subsequent chapters, I highlight significant gaps in the literature regarding the neural substrates of this empathizing process, and discuss two completed studies—each targeting a different region of interest in the brain—aimed at answering these gaps in knowledge to further characterize cortical contributions to social decision-making.

1a. Empathy as a set of social decision-making processes – perception, valuation & action

Generally, it is agreed that empathizing refers to "the ability to vicariously experience and to understand the affective state of other people" (Lockwood, 2016). Empathy as a skill is understood differently in different contexts, adding to the difficulty in operationally defining it as a process, but early academic models and more recent interpretations have helped in framing specific questions for investigation. Naturally, empathizing can be observed as a broad spectrum of behaviors, from sharing in an emotion to considering what the other might be thinking; operationalizing it may thus seem like an insurmountable task. However, a widely accepted set of divisions was put forth by de Waal (2008; Preston and de Waal, 2002), stating that empathy is fundamentally composed of two different systems, a "multi-level conceptualization of empathy." His theory puts the

emotional & cognitive descriptions on a complexity spectrum, which helps to consider the evolutionary timeline and separate animals by the extent they reach on the spectrum in terms of social cognition (de Waal, 2008; Meyza et al., 2017; Preston and de Waal, 2002). At the first level of empathy is emotional contagion, perceiving and adopting another's emotional state, which de Waal considers to be the most evolutionarily simple and a cornerstone for the more advanced levels. Contagion is followed by sympathetic concern, including the distress we may feel for another's state and our attempts to help them. Finally, empathetic perspective-taking describes the most complex level of empathetic cognition, our ability to attribute a state to another, instead of simply sharing in their state (de Waal, 2008). Panksepp & Panksepp (2013) more recently ascribed a similar three-level division to specialized brain circuits, in order to better visualize how our understanding of top-down processing can apply to social cognition. The authors argue that primary-processing structures, responsible for the "primal empathy" spontaneous phenomena of emotional contagion, are critical for learning and development in social cognition. Secondary-processing structures from basal ganglia and the limbic system integrate primary processes to modulate emotional responses in primary circuits, as well as inform the higher-level cognitive, tertiary processes. The latter are mainly cortical structures, which integrate the lower-level processes to culminate in cognitive control and decision-making processing that regulates social cognition and behavior (Panksepp & Panksepp, 2013). Therefore, there are well-conceived levels, with associated systems, to consider when studying the neural underpinnings of empathy. Ultimately, empathizing refers to a set of processes that culminate in social decision-making, which can be made tractable by their familiar roles in non-social contexts: social perception, valuation, and response (Gangopadhyay et al.,

2020). This operationalization of social cognitive processes makes them analogous to their nonsocial counterparts, providing leverage for studying them with classical behavioral tools. Moreover, given the evolutionary recentness of social cognition, the largest debate on the subject has pondered whether the brain developed new social-specific tools and strategies ('social specialization'), or whether established 'self-focused' processes are repurposed in social contexts ('social co-option'). A critical step towards shedding light on these questions is to examine how brain circuits, particularly those known to be involved in nonsocial learning and decision-making processes, are engaged during similar behavioral demands but within social contexts. Given their tractable toolkits and behavioral paradigms, rodents have been a popular model for the study of social cognition.

1b. Rodents as a model for studying social decision-making

The relatively recent use of rodent models in social cognition research has shown many empathic behaviors to exist in several species. In fact, both recent and early studies have noted rodents to exhibit prosocial behavior, which would be categorized as a higher order cognitive process for the theories previously described (Atsak et al., 2011; Church, 1959; de Waal, 2008; Panksepp & Panksepp, 2013; Sato et al., 2015). Rats will also show an aversion to inequitable outcomes in social contexts (Oberliessen et al., 2016), and prefer both rewards when the choice is available (Hernandez-Lallement et al., 2015, 2016a). Indeed, rodents have been a model of choice for research lately, thanks to their proficiency for tasks probing emotional contagion and forms of learning involving social cognition. Both rats and mice have been shown to be able to share in the emotional state of fear from a conspecific in a multimodal fashion (Kim et al., 2010; Knapska et al., 2006, 2010; Meyza

et al., 2017; Munger et al., 2010; Panksepp, 2013), which allows them to participate in vicarious learning of fear (also known as observational fear learning), whereby observer animals acquire fear conditioning by witnessing a conspecific partner undergoing classical fear conditioning (Atsak et al., 2011; Chen et al., 2009; Jeon et al., 2010; Kim et al., 2010). Further, they have been shown to exhibit conditioned responses to a neutral cue, after socializing with a previously fear-conditioned conspecific (Bruchey et al., 2010), which – combined with heightened arousal from interactions with a fear conditioned partner (Knapska et al., 2006, 2010; Meyza et al., 2017)– shows that part of their vicarious acquisition of fear must be socially-derived. Vicarious learning of fear has attracted great interest from researchers in recent years (Kim et al., 2019; Meyza et al., 2017), as it involves the tractable intensity of fear responses and invokes well-studied learning mechanisms in the brain, in addition to those unknown. Moreover, the performance of humans in the task correlated with overall measures of empathy (Kleberg, 2015). Thus, despite the previous lack of attention, as in other areas, animal models have proven to be an incredibly useful tool for the study of social cognition and neurophysiology.

1c. Social decision-making in the brain

Early findings in human neuroimaging studies, combined with more recent physiological work in animals have revealed many regions of interest associated with aspects of social cognition, such as our ability to recognize emotions in others and alter our behavior accordingly (Gangopadhyay et al., 2020; Kim et al., 2019). Many of these structures are already known to play significant roles in non-social functions in learning, decision-making and cognitive control (Brockett et al., 2020; Bryden & Roesch, 2015;

Bryden et al., 2018; Etkin et al., 2011; Gangopadhyay et al., 2020; Kennerly et al., 2009), which has reinforced emphasis on using social learning tasks based on classical learning and attention paradigms. The targeted nature of these tasks allow for a controllable approach to studying the individual contributions of relevant brain regions to social cognition. From this approach, two regions have appeared to be particularly relevant to the appraisal of affect in others, especially as it is involved in adapting goal-directed behavior: the anterior cingulate cortex (ACC) and the orbitofrontal cortex (OFC ; Allsop et al., 2018; Baez-Mendoza et al., 2013; Blair et al., 1999; Carrillo et al., 2019; Cox & Witten, 2019; Etkin et al., 2006; Gangopadhyay et al., 2020; Kim et al., 2019; Lockwood, 2016, 2020; Olsson et al., 2007).

A separate but relevant question in the field is whether social cognition relies on exclusive mechanisms in the brain. The social co-option hypothesis posits that cortico-limbic-striatal circuits have in part repurposed or adapted their function at the physical, algorithmic or computational level, to allow for the development of social-oriented processes in the brain (Gangopadhyay et al., 2020; Lockwood et al., 2020). On the other hand, the social-specialization hypothesis points to the existence of isolated circuits or processes aimed specifically at social cognition; they are either not involved with or even antagonistic to nonsocial processes. Evidence for both has been found, suggesting both might exist at different levels of application within the brain (Gangopadhyay et al., 2020; Lockwood et al., 2020). Neuronal ensembles specifically encoding socially-derived information may be integrated into greater circuits that are independent of social context, yielding socially-dependent behaviors. For example, social-specific neurons encoding

vicarious reward could be integrated into broader computations to determine how or whether they influence choice behavior.

Based on these questions and previous findings detailed above and below, the primary goal of my research was to test whether the nonsocial functions of cortical regions like ACC and OFC could account for the processing of social information, integrating it into existing circuits.

1d. Investigating the contributions of attention and valuation to social decision-making tasks

As described, ACC and OFC are cortical structures known to contribute to learning and decision-making in nonsocial contexts. ACC and OFC exist at the crossroads of many critical yet flexible executive functions, geared towards reacting and adapting to changes in the environment; to this end, they integrate across a wide spectrum of qualitative information, from the sensory to the more abstract, which makes them strong candidates for linking social information to behavioral output. In decision-making tasks, ACC and OFC share communication with each other as well as with BLA (Allsop et al., 2018; Barreiros et al., 2021; Lichtenberg et al., 2017; Janak and Tye, 2015; McDonald, 1998; Wassum and Izquierdo, 2015), among other critical regions, to guide goal-directed behavior as illustrated in Figure 1. Typically, in nonsocial contexts, internal (e.g., hunger state) and nonsocial external cues (e.g., neutral or predictive sound) drive processing in these regions (Fig. 1; black solid line). Social contexts introduce a second stream of information to integrate into decision-making processes (Fig 1, violet solid line). In the following chapters I discuss what is known about the overlap between the social and

nonsocial functions of ACC and OFC in decision-making contexts. I then present my results using a combined appetitive-aversive social task while recording activity from ACC (Chapter 3) and OFC (Chapter 4) neurons in rats. I found that in this task both regions functioned as they do in nonsocial contexts, but integrating socially-derived information when relevant to the task. Specifically, ACC encoded social information in service of attention, whereas OFC activity reflected the positive value of a social outcome or action if it was valuable. Through the work presented here, I aim to show that a dynamic balance between the two sources of information is what ultimately drives behavior in social decision-making tasks, depending on the goal of the self, like other task-related features (Fig. 1).

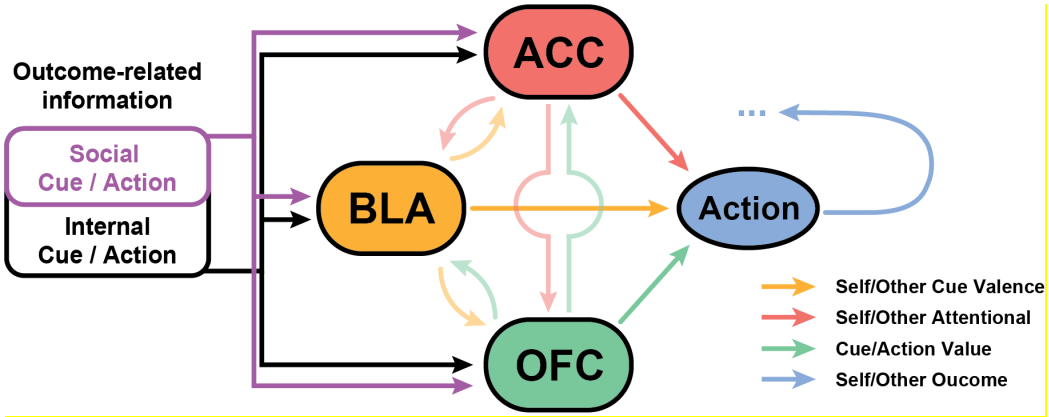


Figure 1: Information flow diagram for ACC-OFC roles in social decision-making. In social contexts, **BLA** receives socially-derived sensory information (perhaps already coded for valence), where it is integrated with internal state cues and corticolimbic modulation to determine social-outcome valence association. **ACC** integrates outcome/action-related sensory information, along with social cue information, to drive attentional modulation of cortico-limbic-basal ganglia goal-directed processing. **OFC** integrates internal state cue information, social/nonsocial outcome-related information and attentional signals to evaluate available sensorimotor options. Legend on the right describes the type of information represented by arrow style. Yellow – social/nonsocial valence information; Red, social/nonsocial attentional signals; Green, value information for outcome-related cues/actions; Blue, integrated cue-action-outcome information; Black solid line, information from internal or nonsocial cues (e.g., hunger, satiety); Violet solid line, information derived from social cues or actions, including cues relating to conspecific outcomes, or actions from or involving them (approach, directional light).

Chapter 2: Detailed Methodology

Most studies examining neural circuits in social decision-making tasks have focused on either appetitive or aversive contexts, limiting the variability of valence in emotions both exhibited and perceived and how each may differ in processing, respectively. Moreover, rodent studies in particular have tended to use tasks in which there are no outcomes delivered to the self (Allsop et al., 2018; Carrillo et al., 2019; Jeon et al., 2010; Keum and Shin, 2019; Kim et al., 2012), presenting no risk or reward, which is uncommon in social decision-making contexts. Together, these caveats complicate interpretation of behavioral or neural results since these regions process appetitive and aversive stimuli differently. The first objective of my research was thus to determine how socially-derived information influences social decision-making in contexts that present appetitive and aversive outcomes directed at either oneself or a conspecific. To this end, I used a Pavlovian social task in which a rat or its partner receive rewards or foot shocks depending on two contingent cues in each trial; an auditory cue predicting a reward, shock or no outcome, followed by a visual cue signaling which rat would receive the outcome. The behavioral task and neurophysiological tools I used, along with their respective analyses, are detailed below.

2a. Pavlovian social outcome task

I used a modified version of a task previously published (Lichtenberg et al., 2018) with a few modifications, which is run in a modified shuttle box chamber (Fig. 2A-C; 16 in x 6.25in x 8.375 in; WDH; Med Associates). A modified guillotine door with wire mesh covering the opening divided the chamber in two equal compartments. Rats could see,

smell and hear each other. Each trial began with illumination of a house light (Fig. 2B-D). Five seconds later, one of three auditory cues (the 'outcome cue') was emitted for 5 s (i.e., tone, white noise or clicker, counterbalanced across rats) gated by an Arduino. One auditory cue indicated that reward would be delivered (i.e., reward trial), a second cue signaled that shock would be administered (i.e., shock trials) and a third cue (i.e. neutral) indicated that neither reward nor punishment would occur. After 5 s, the auditory cue was terminated simultaneously with the illumination of one of the two directional lights. This 'directional' cue informed the rats which side of the cage (random 50/50) would lead to a positive (reward), negative (foot-shock) or neutral outcome (nothing). After 5 s, reward or punishment or nothing was administered to the side of the box that was illuminated by the directional cue. The shock consisted of two 250 ms shocks (0.56 mA) spaced 2 s apart. Reinforcement occurred on 80% of trials. This paradigm was completely Pavlovian, thus rats had no control over what outcomes would occur or which rat would receive them. The directional light turned off 5 s after the delivery of outcomes, followed 5 s later by the houselights turning off and a final 5 s ITI before the start of the next trial.

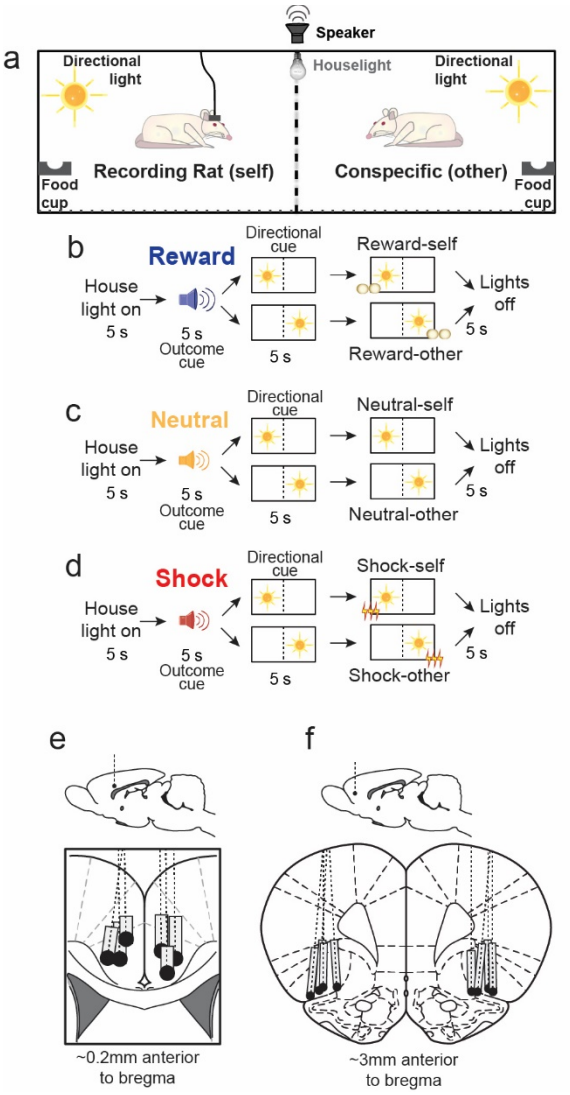


Figure 2: Pavlovian Social Outcome Task. Schematic of behavioral chamber used in the task. Pairs of rats are placed in opposite sides of the ‘‘divider’’ within the same shuttle-box chamber, on a grid of conductive rods. The directional light sits on the panel opposite to the divider on each side, above the grid but low enough to be visible through the divider from the other side. The food cup is to the left of the directional light. Both sides are diagonally mirrored in their placements of the light and food cup. **B–D**) Flow diagram of reward (**B**), neutral (**C**), and shock (**D**) trials within the task, which begins with the onset of the house lights. After 5 s, an outcome cue (5 s) indicates the type of outcome delivered, followed by activation of one of the directional lights (10 s), indicating which side receives the outcome. 5 s after the onset of the directional light cue, the outcome is delivered to the designated side. In the following 10 s, the directional lights turn off and then the houselights turn off, 5 s apart each. (**B**)–(**D**) represent all 6 basic trial types: reward-self; reward- other; neutral-self; neutral-other; shock-self; and shock-other. There were 4 different trial blocks (60 trials per block; 10 trials per trial type), during which both rats received outcomes (RR; where ‘‘R’’ designates ‘‘reinforced’’; numerator, recording rat; denominator, conspecific), neither rat received outcomes (NN; ‘‘N’’ designates which rat was not reinforced), only the recording rat was reinforced (RN), or only the conspecific was reinforced (NR). During non-reinforced trials, all stimuli were presented but shocks and reward were not delivered. **E-F**) Location of recording sites in ACC (**E**) and OFC (**F**) recording experiments (Paxinos and Watson). Dashed line represents electrode placement, and boxes mark the extent of the recording locations.

Experimental sessions lasted 2 h, where rats underwent 4 different blocks of 60 trials (six trial types, 30 s/trial, 10 trials/type; Fig. 2B-D). Trials were presented in a pseudo-randomized order. The block types represent four possible combinations (context pairs) for

outcome delivery for each pair of rats: 'Both reinforced' (RR), where both rats are reinforced (i.e., receive outcomes); 'both not reinforced' (NN), where neither rat receives outcomes; 'only self (recording rat) reinforced' (RN), where only the recorded rat received outcomes; 'only conspecific reinforced' (NR), where the conspecific received outcomes while the conspecific did not. The four blocks in every session follow one of two sequences, which alternated daily. Two sequences were established in order to counterbalance the order in which self or other are extinguished during the task. Finally, every 6 sessions, recording rats trained in a session alone, as a control for social context. In these sessions, pellet outcomes to the other were delivered to an empty beaker, and shock deliveries to the other were delivered as normal, but to an empty side.

An infrared beam was placed at the entrance to the food cup on both sides of the cage. This beam was disrupted upon entry of the rat's nose into the food cup, and beam breaks served as a quantitative measure of reward seeking. In the Med Associates boxes, I sampled every 10 ms to determine if the beam in the food cup was broken throughout the entire trial. Finally, video was captured of recording rats, in order to more directly measure aversive responding (freezing, gate approach) during the task.

2b. Behavioral data analysis

For analysis of behavioral responding, infrared beam break data (10 ms sampling rate) were aggregated as proportions across 1-second bins (i.e. divided by the number of possible breaks per second to yield a percentage), collected from the MED-PC software (Med Associates). For video scoring of freezing and approach, cameras were positioned facing the recording rat. Video analysis, like IR and neural analyses, focused on four trial

epochs lasting five seconds in length each: auditory cue; directional light; outcome and post-outcome to houselights off. Freezing (sudden cessation of movement) and approach toward the mesh divider were assessed during these periods by 2-4 independent observers. Statistical procedures on the data were executed using MATLAB (MathWorks; Wilcoxon and Student's t-test) and Excel (Microsoft; Chi-squared)

2c. In-vivo single-unit electrophysiology

I recorded single-unit activity from targeted ROIs, from assigned recording ('self') rats as they performed in the task with their partners. Single-unit recordings during this procedure would allow me to probe the neural activity of these regions, in a social context that varied the valence of the outcome and the relative outcome expectancy between 'self' and 'other.' Electrodes were manufactured and implanted as in previous recording experiments (Brockett et al., 2020; Bryden et al., 2011, 2018; Bryden & Roesch, 2015). Rats had a drivable bundle of ten 25µm diameter FeNiCr wires (Stablohm 675, California Fine Wire, Grover Beach, CA) chronically implanted in the left or right hemisphere ACC (N = 6 rats; 0.2mm anterior to bregma, 0.5mm left [n = 3] or right [n = 3] of the midline, and 1mm ventral to the brain surface, according to Paxinos and Watson; see Fig. 2.1E for ACC recording tracts) or OFC (N = 8 rats; 3mm anterior to bregma, 3.2mm left [n = 4] or right [n = 4] of the midline, and 4mm ventral to the brain surface, according to Paxinos and Watson; see Figure 2.1F for OFC recording tracts). Immediately prior to implantation, wires were freshly cut with surgical scissors to extend ~1mm beyond the cannula and electroplated with platinum (H2PtCl6, Aldrich, Milwaukee, WI) to an impedance of ~300kOhms. Cephalexin (15mg/kg p.o.) was administered twice daily for two weeks post-

operatively to prevent infection. At the end of recording sessions, electrodes were driven deeper by 40-80 μm to find new units.

2d. Electrophysiological data analysis

Units were sorted via Offline Sorter software from Plexon Inc (Dallas, TX), using a template matching algorithm and analyzed in Neuroexplorer (Plexon) and MATLAB (MathWorks). Activity was examined during two different 5s epochs: Directional Light epoch = directional light to outcome delivery (5s); Outcome epoch: 5s after start of outcome delivery (i.e., 5s in length starting 5s after onset of directional light). Activity in population histograms was normalized by dividing by the maximal firing rate of each neuron. All statistical procedures were executed using raw firing rates or counts, in either MATLAB (Wilcoxon) or Excel (Chi-squared). Neural response index scores for shock or reward trials were computed by subtracting the averaged firing rate each neuron during neutral trials from their average firing rate for the corresponding trial type, then dividing by the sum of both (e.g., [shock-self – neutral-self] / [shock-self + neutral-self]). Significant increases or decreases in population activity were determined by testing the distribution of response index scores for all neurons for significant shifts, using rank-sum tests (Wilcoxon; $p < 0.05$). Neurons were classified as being reward- or shock-responsive by comparing reward to neutral and shock to neutral, respectively (Wilcoxon; $p < 0.05$).

Chapter 3: ACC Signals Attention in a Social Paradigm that Manipulates Reward and Shock

3a. Anterior Cingulate Cortex and attending to relevant cues

The ACC participates in many higher-order, non-social cognitive functions, such as attention, error-detection and decision-making (Brockett et al., 2020; Bryden et al., 2011, 2018; Kennerley et al., 2009; Wallis & Kennerley, 2010, 2011). Previous work in the lab using the stop-change task, while recording from rats, has shown that the ACC is required for conflict resolution and subsequent behavioral adaptation (Brockett et al., 2020; Bryden et al., 2018). In this task, rats must inhibit a previously favored responding to GO signals (e.g., left) when STOP signals appear, and choose the opposite direction. The sudden appearance of a contradicting STOP signal conflicts with the GO cue promoting a habitual behavior, which requires executive control to overcome, directing behavior back towards the goal of the trial. Performance in the task was correlated with ACC activity, and impaired if the latter was inactivated during STOP trials. These results highlight how the attention functions of ACC have a role in learning & behavioral adaptation, supported further by findings that the development of action plans in dorsomedial striatum depends on ACC for performance during this task (Brockett et al., 2020).

3b. Anterior Cingulate Cortex – Social attention

Social cues can be an ambiguous source of information, compared to the simpler perceptual association of other external cues (lights, sounds) to different outcomes, but in most social contexts they provide critical information about the environment; in those cases, it may be necessary to attend to them over other salient cues. The attentional role that ACC plays in nonsocial contexts may also contribute to navigating outcome-related cues in social contexts. ACC has been shown to be engaged during social tasks involving learning and attention in humans, non-human primates (NHP) and rodents. In humans, the ACC was engaged during affect-based Stroop tasks and vicarious learning (Etkin et al., 2006; Olsson et al., 2007), requiring subjects to integrate socially-derived affect information to learn or exert attentional control over habitual responding. ACC activation was observed during both 'self' and 'other' shocks, suggesting a functional overlap (Olsson et al., 2007). Similar findings come from work in NHP, where ACC has been shown to have comparable non-social functions (Hayden et al., 2011; Wallis & Kennerley, 2010, 2011), as well as social ones (Chang et al., 2013; Gangopadhyay et al., 2020; Lindström et al., 2018; Lockwood, 2016). The advent of more invasive techniques has greatly aided in parsing the circuits of social decision-making. Investigators were able to show that ACC neurons encode reward outcome information about the self, the other or both in social contexts (Chang et al., 2013; Lindström et al., 2018; Noritake et al., 2018). In rodents, ACC is not only involved during observational fear learning, responding to self-directed and socially-derived cues during the task, but ACC neurons are also necessary for acquisition of the learned conditioning (Allsop et al., 2018; Carrillo et al., 2019; Jeon et al., 2010; Kim

et al., 2012). Optogenetic inactivation of BLA-projecting ACC neurons during vicarious fear conditioning resulted in impaired acquisition of observational fear conditioning, but not impaired expression of a learned response if they were inactivated during the test phase instead. Notably, optogenetic inhibition of these cells during classical fear conditioning, where animals directly experience cue-shock pairs, did not impair learning (Allsop et al., 2018), demonstrating the social relevance of ACC to the task. Taken together, these findings help strengthen the view that ACC plays a critical role in social cognition, potentially by driving attention to socially-derived cues and their relationship to outcomes for the self and the other, which is analogous to its functions in nonsocial learning. Nevertheless, the exact nature of ACC signals and the information they encode in social contexts remain unclear.

A major issue with previous literature is that most studies investigating rodent ACC (and other regions) in social cognition have focused heavily on fear learning as a framework. There is a rationale to this emphasis, since ACC is heavily involved in the affective sensation of pain, forming part of the medial pain system alongside anterior insula (Shackman et al., 2011; Xiao and Zhang, 2018). Based on its connectivity and sensitivity to both personal and vicarious pain stimuli, researchers have suggested that ACC may integrate pain and social stimuli through 'emotional mirror neurons.' The notion to consider a mirror neuron, similar to those originally found in premotor cortex (Gallese et al., 1996; Rizzolatti et al., 1996), focused instead on the emotional state of others is not a novel one (Baird et al., 2011; Preston & de Waal, 2002), but evidence was recently found for their existence in ACC for pain (Carrillo et al., 2019). The authors used a vicarious fear learning task while recording multi-unit activity from the ACC of observer rats, along with

behavioral measures gauging empathetic responding. Carrillo and colleagues (2019) found a subpopulation of ACC neurons that responded purely to pain stimuli in both the self and the other. Interestingly, these neurons were labeled 'pain mirror neurons,' while those that responded to both other pain and self fear were labeled as 'fear mirror neurons.' Lastly, the 'pain mirror neurons' showed response profiles to varying levels of pain in the other could be used to decode pain levels in the self, highlighting their physiological overlap. These findings would suggest social specificity in ACC, as unique cells tasked with encoding affect of both self- and other-pain or fear, but their conclusions are complicated by the focus on negative valence. ACC function in nonsocial learning contexts is not dependent on valence, given its involvement in many appetitive tasks (Brockett et al., 2020; Bryden et al., 2011, 2018; Chang et al., 2013; Hayden et al., 2011; Lockwood et al., 2015; Wallis & Kennerley, 2010), and rodents are able to gain appetitive information from social cues as well (Apps et al., 2016; Burgos-Robles et al., 2019; Munger et al., 2010; Oberliessen et al., 2016). As an alternative explanation, the 'pain mirror neurons' described in the study (Carrillo et al., 2019) may instead be neurons that are responding to self- or other-outcomes in the interest of attention, an already established nonsocial function of ACC. This interpretation, which suggests the co-option of existing circuits for social contexts, is parsimonious with a previously established function for ACC and releases the function of valence processing to other structures that may already have that role in nonsocial settings. Indeed, recent work found that the social transfer of either pain or fear may depend more on the connectivity between ACC and BLA (fear) or nucleus accumbens (pain; Smith et al., 2021), suggesting that these downstream targets are more relevant to valence-specific processing in social contexts. Therefore, the omission of positive valence ultimately makes

it difficult to interpret Carrillo and colleagues' findings with respect to the social specificity of ACC. Given the nonsocial functions ACC has been previously shown to exert, it is more likely that in social contexts ACC drives attention to salient, relevant social cues in the environment, regardless of their valence.

3c. Results

In order to interrogate the function of ACC during the combined appetitive-aversive task, I recorded single-unit activity from the ACC of rats experiencing the task with a partner. Based on previous findings in the literature, but given the issues I raised in Section b, I predicted that if ACC encoded the emotional valence of outcomes delivered to the self and other, neural responses would differ between reward and shock, showing increases for one but decreases for the other and vice versa (Fig, 3.1A). Instead, if ACC signals attentional changes during the task, as it does in nonsocial contexts, the response profiles of ACC neurons would be shared across outcome valence. That is, increases in firing in response to reward would be met with similar firing for shock outcomes (Fig. 3.1B). I first discuss the behavioral results from the task, followed by the neural data from ACC recordings.

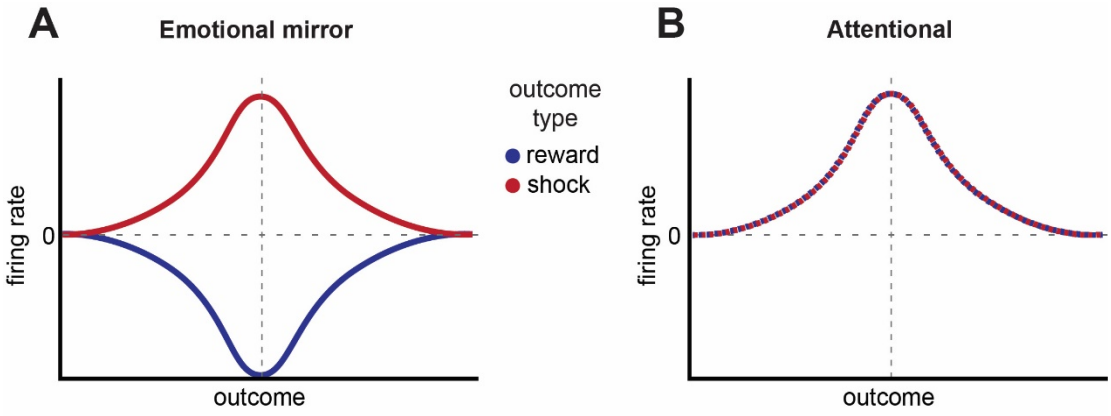


Figure 3.1: Hypotheses for social outcome encoding in ACC. A) Predicted average firing of ACC neurons if the valence of outcomes for self and other are encoded. Increases in firing for one type of valence outcome (e.g., negative/shock) would be met with decreases in response to the opposite valence. **B**) Predicted average firing of ACC neurons if ACC signals attentional changes during the task, showing responses based on outcome salience rather than valence (i.e., similar response profiles to both positive and negative outcomes).

Rats correctly internalize auditory, directional light cues, and social outcome context

Because the task was Pavlovian, I used food cup bream breaks and video scoring to determine if rats understood the task. Figure 3.2A-H shows the average beam break into the food cup over trial time for the three trial-types averaged over all recordings.

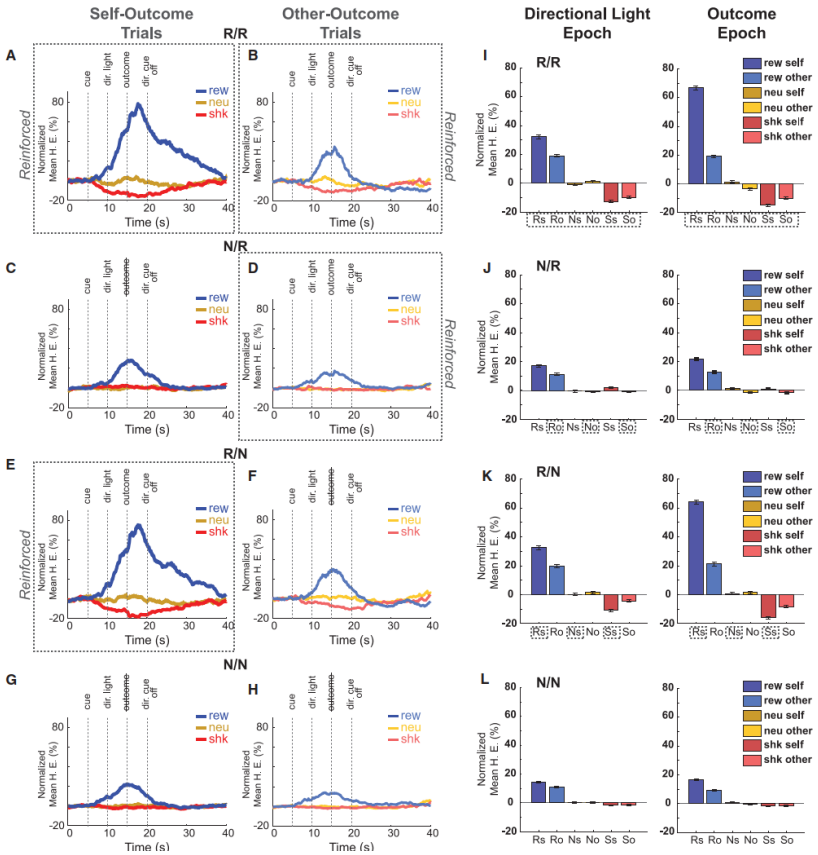


Figure 3.2. Rats learn the predictive value of both outcome and directional cues, modulated by different outcome contexts. Average beam breaks from food cup entry as a percentage of trial time for each outcome type (reward – blue, neutral – orange, shock – red), across each block type (Dotted line boxes indicate whether trials were "Reinforced") for self- (**A,C,E,G**) and other- (**B,D,F,H**) outcome trial. N = 139 sessions (6 rats). Vertical dotted lines indicate task-related events (outcome cue, directional light (dir. light), outcome delivery, directional light off). **I-L**) Averaged food cup entry during the directional light (left) and outcome (right) epochs (5 s). Reinforcement context for each row corresponds to the block type of the same row in **A-H**. Trial types: reward-self (Rs, solid blue), reward-other (Ro, light-blue), neutral-self (Ns, solid orange), neutral-other (No, light-orange), shock-self (Ss, solid red) and shock-other (So, light-red). Statistics are reported in the text. There were 4 different trial blocks (60 trials per block; 10 trials per trial-type), during which both rats received outcomes (R/R; where 'R' designates which rat was reinforced; numerator = recording rat; denominator = conspecific), neither rat received outcomes (N/N; 'N' designates 'not reinforced'), only the recording rat was reinforced (R/N) or only the conspecific was reinforced (N/R). During non-reinforced trials all stimuli were presented but shocks and reward were not delivered. 1st row of figures (A,B,I) = R/R trials; 2nd row of figures (C,D,J) = N/R trials; 3rd row of figures (E,F,K) = R/N trials; 4th row of figures (G,H,L) = N/N trials.

Bar graphs in Figure 3.2I-L show average beam breaks during two trials epochs: the 'directional light epoch' (5s after onset of the directional cue) and the 'outcome epoch' (5s after outcome delivery). Let us first consider the blocks of trials where both the recording rat and the conspecific received outcomes (i.e., 'R/R trial blocks; Fig. 3.2, top row). As was observed previously, prior to outcome delivery, beam breaks increased and decreased on reward-self (blue) and shock-self (red) trials relative to neutral (orange) trials (Fig. 3.2A), respectively (Lichtenberg et al., 2018). After the presentation of the directional light cue (i.e., the cue that informed the rat which animal would receive the outcome), there was a significant increase in food cup entries for reward-self compared to reward-other trials, demonstrating that rats anticipated the receipt of reward before its delivery (Fig. 3.2I, dark vs pale blue; Wilcoxon; $z = 5.326$, $p < 0.001$). During shock trials, there was a significant decrease in beam breaks during both the directional light and outcome epochs compared to neutral, for both shock-self and shock-other trials, and the effect was stronger for shock-self (Fig. 3.2I; red; Wilcoxon; shock-self: DL: $z = -8.193$, $p < 0.001$; Out: $z = -9.738$, $p < 0.001$; shock-other: DL: $z = -7.189$, $p < 0.001$; Out: $z = -4.472$, $p < 0.001$). Thus, in trial blocks where both rats were reinforced (R/R) food cup entries were higher and lower for reward and shock trials compared to neutral, respectively, and were stronger when the recording rats were personally going to receive the outcome.

These results demonstrate that recording rats understood the meaning of auditory and directional cues. Importantly, these effects were highly dependent on whether the recording rat was being reinforced in a given block of trials. That is, during N/R and N/N blocks increases and decreases in food cup entries relative to neutral trials were reduced relative to R/R and R/N trial blocks (Fig. 3.2; first and third rows vs. second and fourth

rows). Most interestingly, this was true during shock-other trials even in blocks where the conspecific was still receiving shock (i.e., N/R; Fig. 3.2D, J; Wilcoxon; DL: $z = -8.723$, $p < 0.001$; Outcome: $z = -8.124$, $p < 0.001$). This suggests that suppression of food cup responding reflects behavioral reactions due to potential harm to oneself, not to the conspecific. This argument is also supported by the observation that food cup entries were significantly suppressed during shock-other trials even during trial blocks when the conspecific was not being shocked, but the recording rat was (i.e., R/N; Fig. 3.2F,K; Wilcoxon; DL: $z = -6.466$, $p < 0.001$; Outcome: $z = -7.759$, $p < 0.001$). Overall, these results suggest that changes in behavior of the recording rats that occurred when conspecifics received shock reflected concern for oneself, as opposed to empathetic concern for the other.

Freezing and approach during shock trials

Above I show that rats understand the meaning of cues and exhibit increases and suppression of food cup entries, during reward and shock trials relative to neutral trials, respectively. To better understand the nature of these data, especially as they relate to shock trials, I scored video for freezing and approach. Figure 3.3A-D represents average freezing of recorded animals for each block, during the 5s-long outcome cue (‘cue’), directional cue (‘dir. cue ON’), outcome delivery (‘outcome’) and directional cue off (‘dir. cue OFF’) epochs of the task. Freezing was defined as the sudden absence of movement except for respiration.

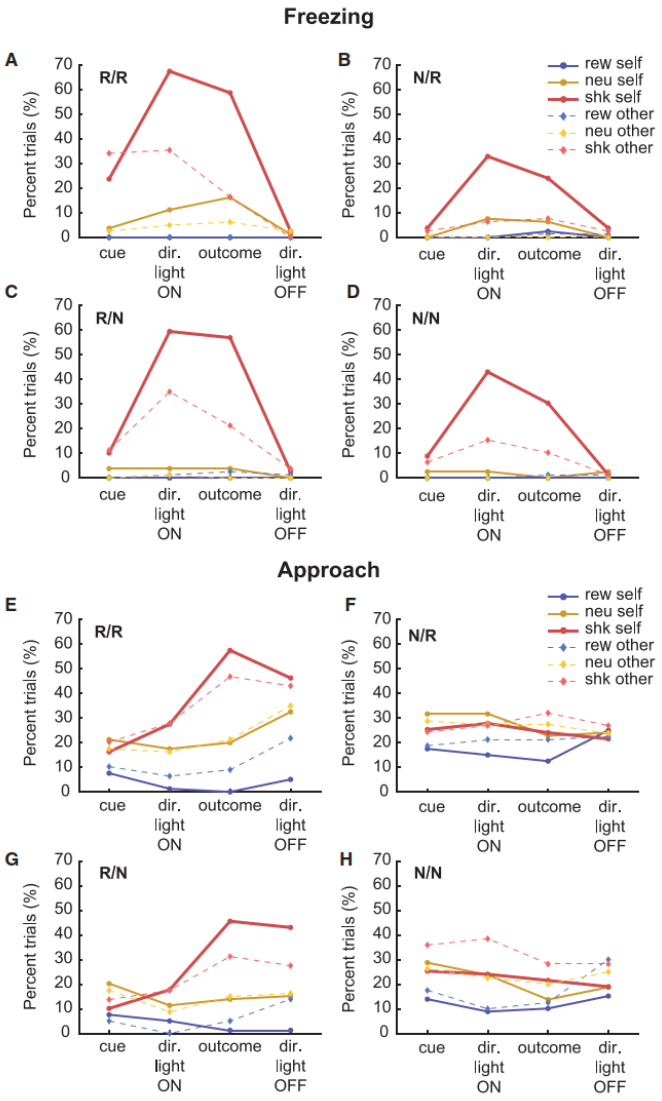


Figure 3.3: Rats show increased freezing and conspecific approach during shock-self and shock-other trials. A-D) Percentage of trials recorded rats froze during each epoch, for each trial and block type. **E-H)** Percentage of trials recorded rats approached the gate, defined as moving towards or actively interacting at the 'divider'. One session for each rat that contributed neural data was scored. Sessions were concatenated and counts (i.e., froze or did not freeze; approached or did not approach) were taken during each of the 5 s epochs. Counts were compared via chi-squared tests but graphs reflect percent over trials. Trial types: reward-self (solid blue), reward-other (dotted light-blue), neutral-self (solid orange), neutral-other (dotted light-orange), shock-self (solid red) and shock-other (dotted light-red). There were 4 different trial blocks (60 trials per block; 10 trials per trial-type), during which both rats received outcomes (R/R; where 'R' designates 'reinforced'; numerator = recording rat; denominator = conspecific), neither rat received outcomes (N/N; 'N' designates which rat was not reinforced), only the recording rat was reinforced (R/N) or only the conspecific was reinforced (N/R). During non-reinforced trials all stimuli were presented but shocks and reward were not delivered. Statistics are reported in the text.

Consistent with the analysis of food cup entries, it was clear from the recording rats' freezing behavior that they understood the meaning of the cues during trial blocks where both animals were reinforced (Figure. 3.3A, R/R); rats froze more in shock-self and shock-other trials compared to their neutral counterparts ('dir. cue ON': shock-self, χ^2 =

32.000, $p < 0.001$; shock-other, $\chi^2 = 17.850$, $p < 0.001$; Outcome: shock-self, $\chi^2 = 19.154$, $p < 0.001$; shock-other, $\chi^2 = 3.467$, $p = 0.0593$), and froze more often on shock-self trials compared to shock-other trial types during the directional light and outcome epochs ('dir. cue ON': $\chi^2 = 8.181$, $p = 0.004$; Outcome: $\chi^2 = 19.154$, $p < 0.001$). Freezing was most apparent during the directional cue light epoch indicating that rats anticipated shock delivery (Fig. 3.3A-D).

As previously reported, these results suggest that rats exhibit 'empathetic' behavior. However, the results were highly dependent on whether the recording rat was receiving shocks during that trial block. During trial blocks where the recording rat was not shocked, but the conspecific was (Fig. 3.3B, N/R), freezing was significantly reduced (Shock-other in R/R vs N/R: 'dir. cue ON': $\chi^2 = 15.89$, $p < 0.0001$; 'Outcome': $\chi^2 = 12.80$, $p = 0.0003$), suggesting that when rats did not anticipate first-hand harm, they did not express behavioral reactions associated with conspecific distress. This interpretation is further supported by the observation that freezing on shock-other trials was high during trial blocks where the recording rat, but not the conspecific, received shock (Fig. 3.3C, R/N shock-other vs neutral-other, 'dir. cue ON': $\chi^2 = 24.952$, $p < 0.001$; Outcome: $\chi^2 = 16.801$, $p < 0.001$).

Along with freezing, approach to the conspecific side was scored. (Fig. 3.3E-H). Approach was defined as the movement and investigation of the recording rat in the direction of the conspecific, which has been suggested to be a measure of attention, concern, and consolation (Atsak et al., 2011; Ben-Ami Bartal et al., 2011, 2014; Burkett et al., 2016; Lungwitz et al., 2014; Meyza et al., 2017). During trial blocks where both rats received shock (Fig. 3.3E, R/R), the recording rat approached the conspecific more on shock-self and shock-other trials compared to neutral trials during the directional light

epoch and after the outcome, with the strongest approach being observed after the shock was delivered ('dir. cue ON': self, $\chi^2 = 1.734$, $p = 0.182$; other, $\chi^2 = 2.263$, $p = 0.128$; Outcome: self, $\chi^2 = 14.420$, $p < 0.001$; other, $\chi^2 = 7.334$, $p = 0.006$). Notably, increases in approach were not observed on shock trials during trial blocks where reinforcement for the recording rat was omitted, even though the conspecific was still receiving shock (Fig. 3.3F, N/R; 'dir. cue ON': self, $\chi^2 = 0.179$, $p = 0.662$; other, $\chi^2 = 0.019$, $p = 0.879$; Outcome: self, $\chi^2 = 0.022$, $p = 0.869$; other, $\chi^2 = 0.179$, $p = 0.662$). That is, recording rats did not approach the conspecific while it was being shocked in trial blocks where there was no first-hand threat. However, increases in conspecific approach were present in blocks where the recording rats were receiving shock but the conspecifics was not (Fig. 3.3G, R/N; 'dir. cue ON': self, $\chi^2 = 1.044$, $p = 0.297$; other, $\chi^2 = 2.267$, $p = 0.127$; Outcome: self, $\chi^2 = 13.192$, $p < 0.001$; other, $\chi^2 = 4.4989$, $p = 0.033$), suggesting that it is the threat of personal shock that promoted approach on shock-other trials.

In summary, behavioral results demonstrate that the recording rats understand the task structure. Specifically, recording rats reacted more on 'self' versus 'other' trials, thus they understood the significance of the directional light. Rats entered the food-cup the most on reward trials, and the least on shock trials, thus they learned to discriminate between predictive auditory outcome cues. In addition, recording rats froze to cues and approached the conspecific on shock trials. These results demonstrate that both reward and shock trials have opposite valence, but are both arousing and drive behavior (shock elicits freezing and conspecific approach; reward elicits food cup entries). Lastly, my results suggest that the recording rats' reactions during shock-other trials were highly dependent on the potential for receiving shock first-hand. That is, food cup response suppression, freezing, and

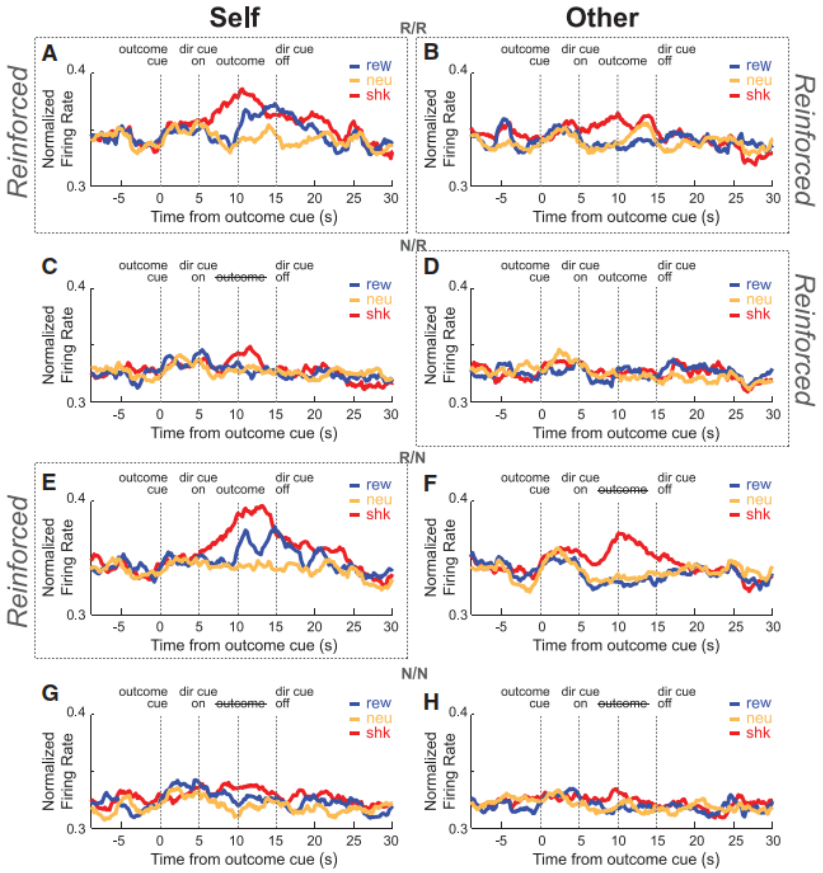
approach were stronger for both reinforced and non-reinforced shock-other trials during blocks of trials when the recording rat was reinforced (R/R and R/N) and they were not different from neutral trials when the recording rat was not receiving shock (N/R).

ACC firing is stronger during threat of first-hand shock on self and other trials

The average over all recorded neurons (n = 139) across trial time for each trial-type and trial block is illustrated in Figure 3.4A-H. As previously reported (Carrillo et al., 2019), I saw increases in firing during shock-self (Fig. 3.4A; red) and shock-other (Fig. 3.4B; red) trials compared to neutral (orange) in trial blocks where both rats were shocked (R/R; top row). However, increased firing on shock-other trials was not present during trials blocks where the conspecific received shock, but the recording rat did not (N/R; Fig. 3.4D; red versus orange). Instead, there were increases in firing on shock-other trials relative to neutral-other trials during the trial blocks where the conspecific did not receive shock but the recording rat did (R/N; Fig. 3.4F). Thus, much like behavior, firing was stronger for both reinforced and non-reinforced shock-other trials during trial blocks when the recording rat was reinforced, and not different from

neutral trials when the recording rat was not receiving shock. Therefore, firing on shock-other trials cannot simply reflect that the conspecific is being shocked.

Figure 3.4: ACC population activity during Pavlovian Social Outcome Task. A-H) Normalized mean firing rate of all recorded neurons (n = 139), across each reinforcement block type for reward (blue), shock (red) and neutral (orange) trials. Each row shows neural firing for self- and other- outcome trials (as indicated above each column) for each block (Dotted line boxes indicate whether self or other trials were reinforced for that block). There were 4 different trial blocks (60 trials per block; 10 trials per trial-type), during which both rats received outcomes (R/R; where 'R' designates 'reinforced'; numerator = recording rat; denominator = conspecific), neither rat received outcomes (N/N;'N' designates which rat was not reinforced), only the recording rat was reinforced (R/N) or only the conspecific was reinforced (N/R). During non-reinforced trials all stimuli were presented but shocks and reward were not delivered. 1st row of figures (**A,B**) = R/R trials; 2nd row of figures (**C,D**) = N/R trials; 3rd row of figures (**E,F**) = R/N trials; 4th row of figures (**G,H**) =N/N trials. See Figure 3.8 for presentation of trial-types split into first and last half of trials per block.



To quantify these effects, for each neuron I computed the normalized difference between firing on shock and neutral trials (shock index = shock – neutral/ shock + neutral) independently for self and other trials during directional light and outcome epochs for each trial block. Distributions of these shock indices for all neurons are plotted in Figure 3.5. During both epochs shock index distributions were shifted above zero for 'self' and 'other' trials, indicating that the majority of ACC neurons fired higher during shock compared to neutral (Wilcoxon; R/R self, DL: $\mu = 0.031$, $p < 0.001$; Outcome: $\mu = 0.024$, $p = 0.024$; other, DL: $\mu = 0.030$, $p < 0.001$; Outcome: $\mu = 0.014$, $p = 0.0544$). Further, shock indices for self and other were positively correlated indicating that neurons that tended to fire more

or less strongly for shock-self trials, tended to fire more or less strongly for shock-other trials, respectively ($r^2 = 0.028$; $p = 0.046$).

Consistent with the population firing (Fig. 3.4), significant shifts in distributions on shock-other trials were not present when the conspecific was to receive shock, but there was no first-hand threat to the recording rat (i.e., N/R; Fig. 3.5C,D; Wilcoxon; DL: $\mu = 0.001$, $p = 0.4391$; Outcome: $\mu = 0.008$, $p = 0.196$). Instead, distributions were significantly shifted on shock-other trials during trial blocks where recording rats received shock, even when the conspecific did not (Fig. 3.5A,E,B,F; Wilcoxon; R/R self, DL: $\mu = 0.031$, $p < 0.001$; Outcome: $\mu = 0.024$, $p = 0.024$; other, DL: $\mu = 0.030$, $p < 0.001$; Outcome: $\mu = 0.014$, $p = 0.0544$; R/N self, DL: $\mu = 0.031$, $p < 0.001$; Outcome: $\mu = 0.024$, $p < 0.001$; other, DL: $\mu = 0.033$, $p < 0.001$; Outcome: $\mu = 0.026$, $p < 0.001$), suggesting that activity reflected behavioral reactions to the possibility of first-hand treat (i.e., suppression of food cup responding and increased freezing on shock-other during RN trials). Consistent with this hypothesis I found the firing rate shock indices for both self and other were correlated with suppression of food cup entries (self: $r^2 = 0.043$; $p = 0.013$; other: $r^2 = 0.030$; $p = 0.0403$).

In conclusion, ACC neurons tended to fire higher during ‘self’ and ‘other’ shock trials when there was a threat of first-hand shock, even during shock-other trials where the conspecific did not receive shock. Notably, these increases in firing on shock-other trials were not observed during sessions where the conspecific was not present (i.e., alone sessions; R/R and R/N; Wilcoxon; DL: $\mu = 0.015$, $p = 0.175$; Outcome: $\mu = 0.008$, $p = 0.401$), suggesting that conspecific presence was necessary for the observed increases in shock-other trials.

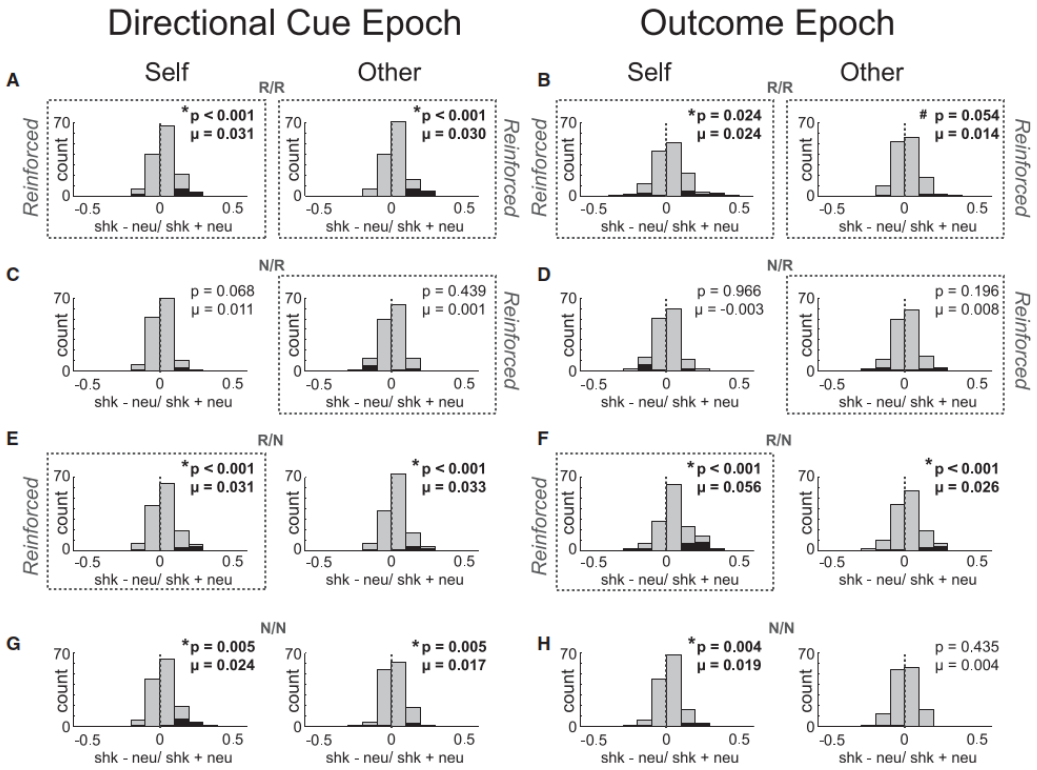


Figure 3.5: ACC neurons tend to fire more during Shock-self and Shock-other relative to neutral. For each neuron I computed the normalized difference between firing on shock and neutral trials (shock index = shock – neutral/ shock + neutral) independently for self (left columns under 'Self') and other (right columns under 'Other') trials during directional light (**A,C,E,G**) and outcome epochs (**B,D,F,H**) for each trial block (5 s epochs). Black bars represent neurons that showed significant within-session differences between shock and neutral trials (Wilcoxon; p's < 0.05). Distributions in trial blocks where threat of shock is likely (**A-B, E-F, G-H**) were significantly shifted in the positive direction, showing that many ACC neurons increased in firing during the directional light and outcome phases of shock trials, for self and other outcomes. Significant responding to other-outcome shock trials was not found in nonsocial contexts (data not shown). There were 4 different trial blocks (60 trials per block; 10 trials per trial-type), during which both rats received outcomes (R/R; where 'R' designates 'reinforced'; numerator = recording rat; denominator = conspecific), neither rat received outcomes (N/N; 'N' designates which rat was not reinforced), only the recording rat was reinforced (R/N) or only the conspecific was reinforced (N/R). During non-reinforced trials all stimuli were presented but shocks and reward were not delivered. 1st row of figures (A,B) = R/R trials; 2nd row of figures (C,D) = N/R trials; 3rd row of figures (E,F) = R/N trials; 4th row of figures (G,H) =N/N trials.

ACC firing was also elevated for reward delivered to the recording rat

The above behavioral and neural analysis suggests that ACC neurons are not signaling when shocks are to be delivered to a conspecific, but instead reflect attention paid to the conspecific (i.e., approach) on shock-other trials when there is a threat of personal shock. If this interpretation of the data is accurate and firing of ACC neurons on shock trials reflects attention, not valence, then reward trials should induce similar changes in firing (Fig. 3.1B).

Re-examination of Figure 3.4 reveals that average firing across the population is not only higher for shock compared to neutral trials, but is also higher during reward-self trials (blue). To quantify this effect and to elucidate the relationship between firing on reward- and shock-self trials, for each neuron I computed the normalized difference between firing rate on shock and neutral trials (shock index = shock – neutral/ shock + neutral) and between reward and neutral trials (reward index = reward – neutral/ reward + neutral) for self-outcome trials during the outcome epoch. For this analysis I combined data for 'self' trials from R/R and R/N blocks to double my sample within each session and because effects were present in both trial blocks (Fig. 3.4). Distributions of shock and reward indices are plotted in Figure 3.6 and counts of significant neurons are represented by black bars (Wilcoxons; p's < 0.05). Both reward and shock index distributions were significantly shifted above zero (shock: $\mu = 0.042$, $p < 0.001$; reward: $\mu = 0.028$, $p = 0.0043$) and counts of neurons that exhibited significantly higher firing for reward over neutral and

shock over neutral outnumbered those showing the opposite effect (Fig. 3.6; reward: 31 vs 14, $\chi^2 = 6.346$, $p = 0.011$; shock: 24 vs 4, $\chi^2 = 14.143$, $p < 0.001$).

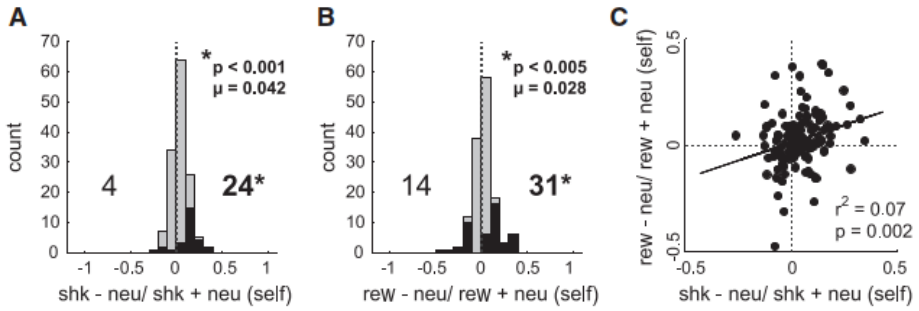


Figure 3.6: ACC neurons tend to fire similarly for reward and shock. A-B) Distributions of counts for neurons selective for shock- or reward-self and other during the outcome epoch, based on calculated index scores. Index scores were obtained as the normalized difference between (**B**) reward or (**A**) shock and neutral firing rates (i.e., shock index = shock - neutral/ shock + neutral; reward index: reward - neutral/ reward + neutral) during self-outcome trials. Counts of cells firing significantly greater than or less than neutral trials are represented by black bars (Wilcoxon; $p < 0.05$). Data was combined across R/R and R/N trial blocks. Wilcoxon tests report significant shifts in distributions and chi-squared tests report significant differences between greater and lesser counts of neurons. **C**) Correlation between reward and shock trials during self-trials.

Finally, I asked whether neurons that were responsive during reward trials were also responsive during shock trials (and vice versa) during both self and other trials. That is, did neurons that tended to fire more or less strongly for reward, tend to fire more or less strongly to shock, respectively? As aforementioned, a positive correlation would suggest population-level firing represented changes in attention associated with reward and shock (Fig. 3.1B), whereas a negative correlation would suggest that activity reflected valence or emotion associated with those stimuli (Fig. 3.1A). Lastly, no correlation would suggest that ACC neurons encode reward and shock independently. I found a significant positive correlation between reward and shock indices during both self- and other-trials for both

directional light and outcome epochs (DL: self: $r^2 = 0.102$, $p < 0.001$; other: $r^2 = 0.111$, $p < 0.001$; Outcome (Fig. 3.6C): self: $r^2 = 0.07$, $p < 0.001$; other: $r^2 = 0.122$, $p < 0.001$).

Although activity at the population level in ACC was elevated for both reward and shock – suggesting that overall function of ACC is more closely aligned with attention – this does not exclude the possibility that signals in ACC were heterogeneous or that some neurons in ACC did signal reward and shock independently. For example, I found that 22 (16%) and 12 (9%) neurons increased firing to reward and shock, without significant modulation during shock and reward, respectively (Fig. 3.7).

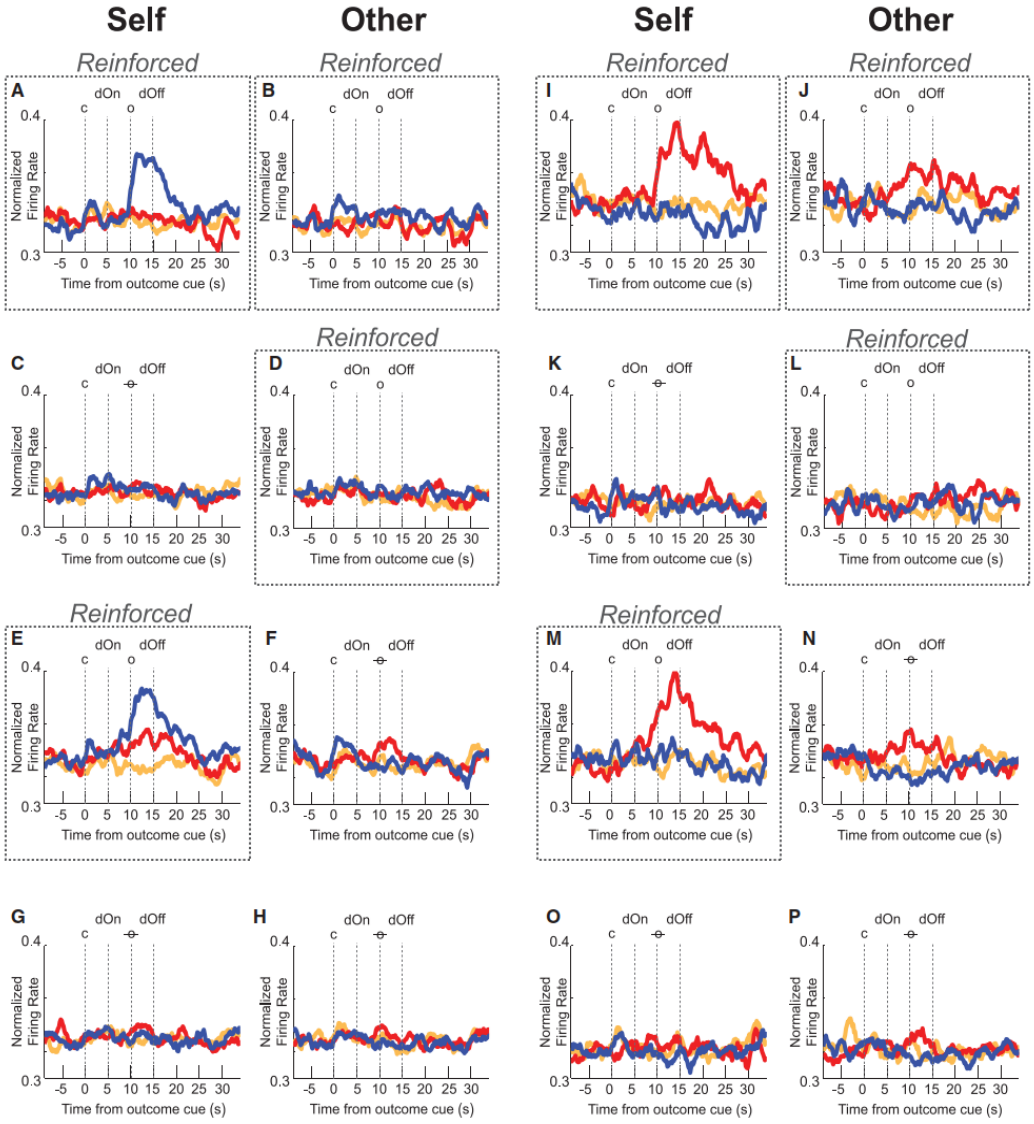


Figure 3.7: ACC neurons responsive to either shock or reward during the outcome epoch. Normalized mean firing rate of a population of neurons selective to shock (**I-P;** n = 12(9%) or reward (**A-H**; n = 22(16%)) trial-types during the outcome epoch (5s), across each reinforcement block type for reward (blue), shock (red) and neutral (orange) trials. Averaged over blocks where the recording rat was reinforced (i.e., R/R and R/N), these neurons showed significantly stronger firing (Wilcoxon; $p < 0.05$) for either reward-self or shock-self over neutral-self but not for both. Each row shows neural firing for self- and other-outcome trials (as indicated by column) for each block (Dotted line boxes indicate whether self or other trials were reinforced for that block). c = cue; dOn = directional light on; o = outcome; ~~o~~ = omitted outcome (i.e., non-reinforced); dOff = directional light off.

3d. Discussion

The current state of the social neuroscience field suggests that the ACC acts as an "emotional mirror neuron" system that allows an individual to perceive the emotions of another via neurons that signal both first-hand pain and the observed pain of others. This shared code is thought to underlie observational fear learning, consolation, empathy, harm aversion and pro-social behavior, which indeed appear to be ACC-dependent. While it is true that increased firing to both first-hand and observed pain might genuinely reflect a shared emotional state, it is equally possible that increases in activity reflect heightened arousal or attention associated with distress, whether it be to oneself or another. While both mechanisms might contribute to subsequent social behaviors - such as observational learning, harm aversion and pro-social behavior - the underlying mechanisms are completely different.

Here, I show that increases in activity reported during first-hand and observed distress can reflect increased social attention. Specifically, I show that increased firing to first-hand pain and a conspecific's pain are correlated with increased firing to reward delivery. Further, I show that rat behavior and ACC firing is only modulated when the recording rat was threatened with first-hand pain. That is, even in rats that have experienced shock, when they are safe, their behavior and firing in their ACC were not modulated by conspecific shock. Even more striking is the observation that firing increases during shock-other trials when the conspecific was not being shocked but the threat of first-hand shock was present. All this suggests that ACC is signaling attention in social contexts only when there was threat of personal harm.

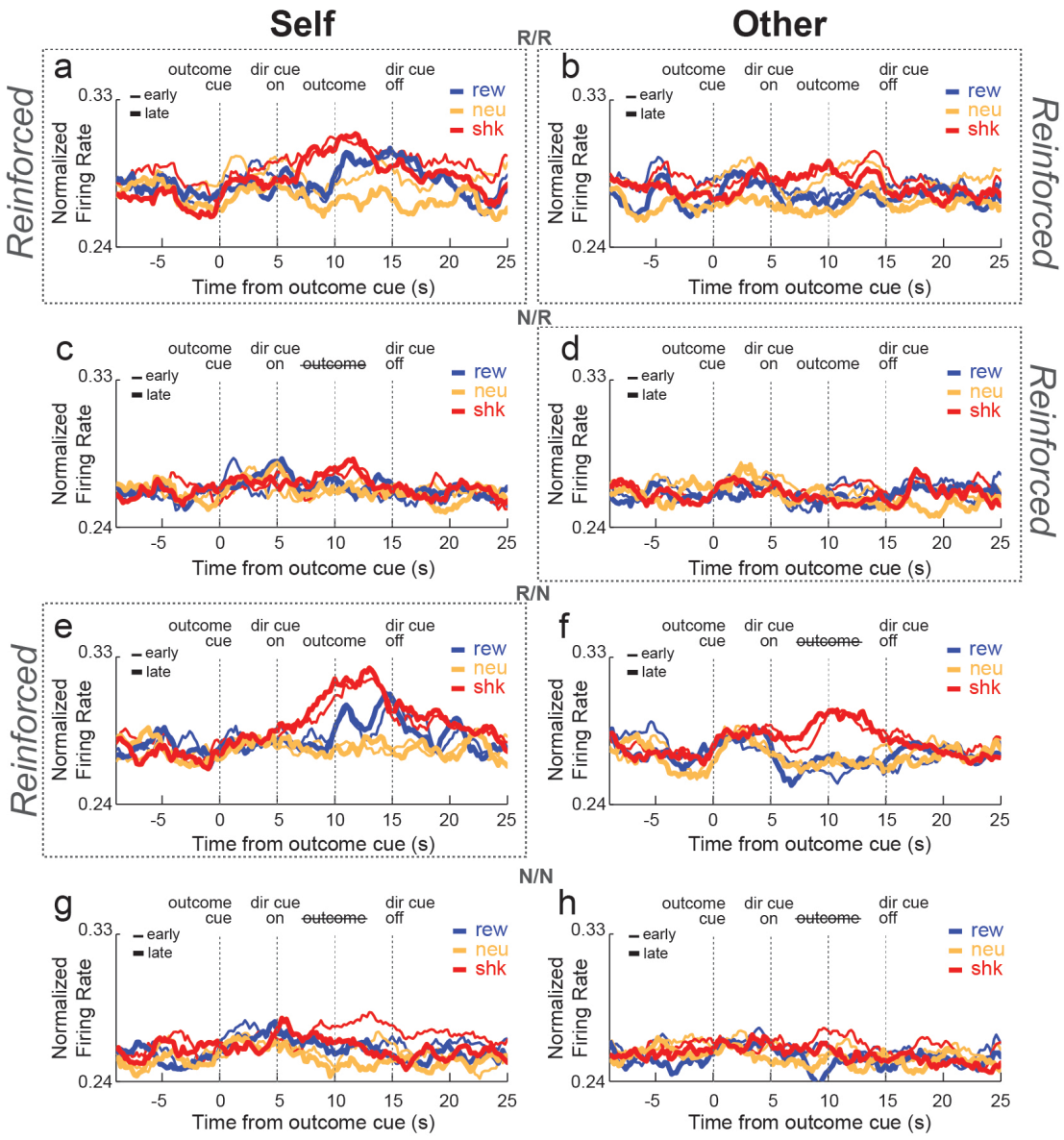


Figure 3.8: Adaptation of ACC activity across trial blocks. a-h) Normalized mean firing rate of all recorded neurons (n = 139), across each reinforcement block type for reward (blue), shock (red) and neutral (orange) trials. Each row shows neural firing for self- and other- outcome trials (as indicated above each column) for each block (Dotted line boxes indicate whether self or other trials were reinforced for that block). Thin lines denote averaged activity for the first half of trials of the block, whereas thick lines represent the latter half of trials. Block transitions result in small but observable carry-over effects from proceeding blocks; these effects are visible as selective responding during nonreinforced (N/N, N/R) blocks and strengthening of responses during R/N blocks. c = cue; dOn = directional light on; o = outcome; dOff = directional light off.

The data demonstrating that ACC is modulated during both shock-self and shock-other trials fits well with previous rodent work. In voles, ACC activity is high when animals console other stressed, previously shocked voles (Burkett et al., 2016). In mice, ACC inactivation or Ca^{2+} channel deletion in ACC impairs observational fear learning (Jeon et al., 2010; Keum and Shin, 2019) and inhibition of ACC projecting neurons to amygdala alters amygdala's representation of the aversive cue during observational conditioning (Allsop et al., 2018). Further, it has been shown that firing in ACC is synchronized with amygdala during observational learning (Keum and Shin, 2019) and that basolateral amygdala (BLA)-projecting ACC neurons preferentially encode socially derived aversive cue information (Allsop et al., 2018). Lastly, in rats, neurons in ACC have been characterized as 'emotional mirror neurons,' as they were found to increase firing to pain inflicted to the recording rat, as well as to a conspecific, according to a potential shared code that maps the distress of another animal onto that of the observer (Carrillo et al., 2019; Hernandez-Lallement et al., 2020).

Although these results are consistent with previous 'shock' work in rodents, the fact that I found very few neurons that increased during reward-other trials is inconsistent with 'reward' work previously reported in monkeys. In monkey ACC, neurons fire when reward is allocated to a conspecific, to oneself, or in both contexts (Chang et al., 2013). Although recent work has shown the influence of social cues on reward learning (van Gurp et al., 2020), to the best of my knowledge this has not been explored in rodents, thus it is possible that rodent ACC is not responsive to rewards delivered to others. However, I speculate that the mere presence of shock stimuli may have diluted neural effects due to low social engagement as evidenced by lower levels of approach during reward-other

compared to both neutral- and shock-other trials. Future work is necessary to better understand the role of rodent ACC in observation of appetitive events. With that said, I found a significant correlation between reward-self and reward-other trials ($r^2 = 0.085$, $p < 0.001$) suggesting that, albeit weak at the neuron level, population firing during reward-other trials also reflected how much attention that trial-type drew.

From previous research it has been clear that the ACC circuit is important for recognition of social distress and the utilization of socially-derived information to adapt behavior (Apps et al., 2016; Chang et al., 2013; Jeon et al., 2010; Kim et al., 2010, 2012). This work is significant because I add to this growing literature, by uncovering the potential nature of what is being encoded by ACC in response to conspecific reward and distress, simultaneously as opposed to separately. By manipulating both reward and shock, this work suggests – at least in the context of this task and the region of ACC that I recorded from – that ACC contributes more towards directing attention, and less to the evaluation of outcomes delivered to the conspecific or the emotional tags that they carry. Examining recording sites from previous studies suggests that more rostral and ventral regions of ACC might be involved in affective processing (Carrillo et al., 2019), while more caudal and dorsal regions may contribute in greater part to executive function, such as attention (Bush et al., 2000; Devinsky et al., 1995), leaving open the possibility that other regions in ACC might carry such information.

It might be argued that the main reason why, here, ACC seems to encode social attention but not vicarious emotion, is that rats performing the current task did not exhibit empathetic concern. This is certainly possible as I will discuss in the next paragraph, but it is important to point out that my rats did freeze, suppress food cup behavior, and approach

when the other rat froze during shock-other trials, which other studies have used as evidence for empathy in rodents (Atsak et al., 2011; Burkett et al., 2016; Carrillo et al., 2019; Jeon et al., 2010; Kashtelyan et al., 2014; Kim et al., 2010, 2012; Lichtenberg et al., 2018; Meyza et al., 2017; van Gurp et al., 2020). Moreover, also consistent with previous work, I show that when a rat is not experiencing shock they exhibit less 'empathetic' behavior (Allsop et al., 2018; Atsak et al., 2011; Kim et al., 2010; Meyza et al., 2017). Importantly, previous papers have concluded that shock naive rats don't freeze when another rat freezes because the observer rat is unable to fully empathize with what the other rat is feeling until it has experienced the pain itself (i.e., one cannot fully understand what another is going through unless they have experienced it themselves). By examining behavior in well trained animals and by manipulating shock and no-shock within the same session, I am able to show that in rats that are fully aware of what the shock is, their behavioral reactions (i.e., freezing, food cup suppression, approach) to the other rat being shocked are not because they are unfamiliar with the shock and can't empathize, but instead, it is because they don't feel threatened. Thus, I argue that my rats do show similar behavioral measures of empathy as found in previous work, and that under these circumstances neural activity in ACC correlates better with attention.

With that said, it is entirely possible that what ACC encodes during this social task is task-dependent. For example, in primate studies, monkeys have to choose between delivering reward to the conspecific and oneself or between the conspecific and an empty bottle (Chang et al., 2013; Noritake et al., 2018).

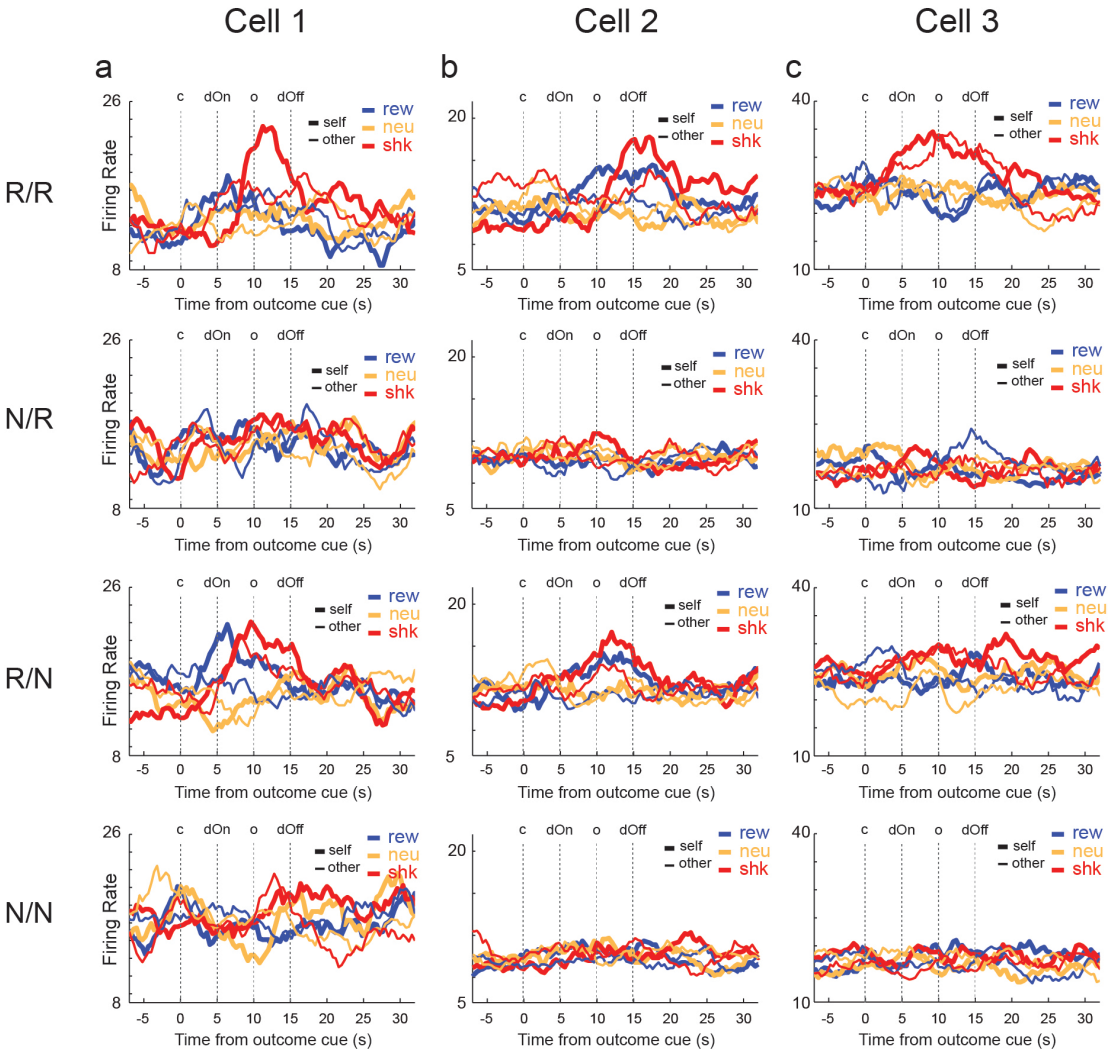


Figure 3.9: Example ACC neurons responsive to shock. a-c) Firing rate averaged across trials of each reinforcement block type for reward (blue), shock (red) and neutral (orange) trials. Thin lines represent averaged trials for other-outcome trials, whereas thick lines average over self-outcome trials. Each row shows neural firing for self- and other-outcome trials (as indicated above each column) for each block (Dotted line boxes indicate whether self or other trials were reinforced for that block). c = cue; dOn = directional light on; o = outcome; dOff = directional light off. There were 4 different trial blocks (60 trials per block; 10 trials per trial-type), during which both rats received outcomes (R/R; where 'R' designates 'reinforced'; numerator = recording rat; denominator = conspecific), neither rat received outcomes (N/N; 'N' designates which rat was not reinforced), only the recording rat was reinforced (R/N) or only the conspecific was reinforced (N/R).

This type of evaluation might require ACC to better encode the value that the animal places on these circumstances, by directing attention to socially-derived cues from the conspecific. Further, the nature of encoding in ACC might also be highly dependent on how the animal subsequently uses social information to alter its own behavior, which will consequently depend on the value that the animals places on outcomes delivered to the conspecific (e.g., aversive versus appetitive). Although many studies have shown rats to exhibit empathetic and pro-social behaviors (Atsak et al., 2011; Ben-Ami Bartal et al., 2011; Hernandez-Lallement et al., 2015a; Meyza et al., 2017; Sivaselvachandran et al., 2018), at least potentially to those of the same strain and in distress (Ben-Ami Bartal et al., 2014; Sato et al., 2015), I have found that rats can be rather 'self-interested.' This has been evident in previous studies examining dopamine (DA) release in nucleus accumbens core (NAc) in a version of the Pavlovian task described here (Kashtelyan et al., 2014; Lichtenberg et al., 2018). For example, previous work in the lab has shown that rats emit appetitive vocalizations and DA is released in NAc during rewards delivered to conspecific, but only early during learning. After rats experienced several trials of uneven outcomes, where the conspecific received reward and they did not, vocalizations became aversive and DA was inhibited during conspecific reward delivery (Kashtelyan et al., 2014). These results were consistent with previous work showing that rats can be sensitive to outcome inequity (Oberliessen et al., 2016, Oberliessen & Kalenscher, 2019). Further, the Roesch lab has shown that DA is released when the recording rat observes the conspecific receive shocks, suggesting that observation of the conspecific receiving shock, instead receiving shock itself, is an event that is better in value than expected (Lichtenberg et al., 2018).

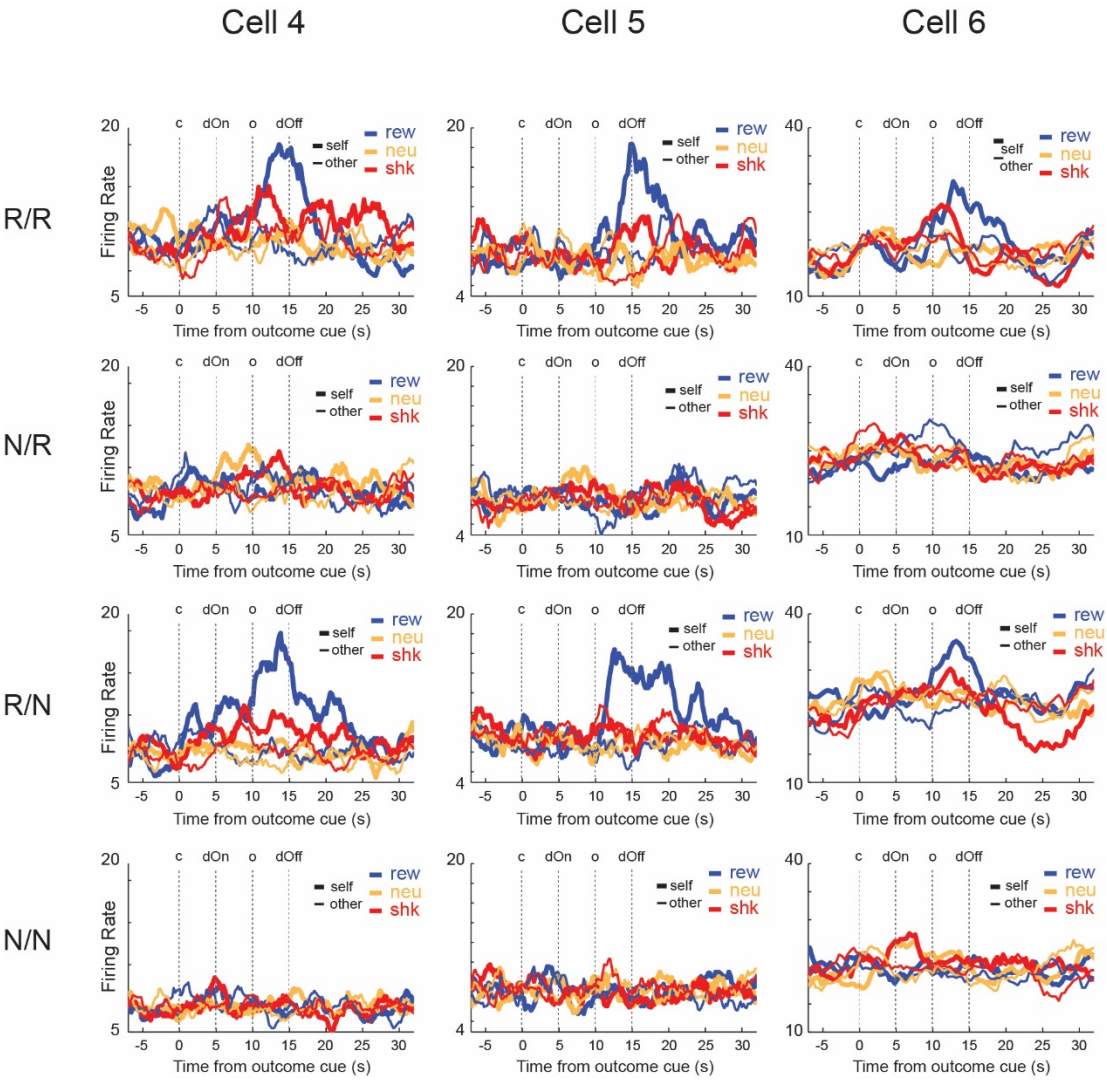


Figure 3.10: Example ACC neurons responsive to reward. a-c) Firing rate averaged across trials of each reinforcement block type for reward (blue), shock (red) and neutral (orange) trials. Thin lines represent averaged trials for other-outcome trials, whereas thick lines average over self-outcome trials. Each row shows neural firing for self- and other-outcome trials (as indicated above each column) for each block (Dotted line boxes indicate whether self or other trials were reinforced for that block). c = cue; dOn = directional light on; o = outcome; dOff = directional light off. There were 4 different trial blocks (60 trials per block; 10 trials per trial-type), during which both rats received outcomes (R/R; where 'R' designates 'reinforced'; numerator = recording rat; denominator = conspecific), neither rat received outcomes (N/N; 'N' designates which rat was not reinforced), only the recording rat was reinforced (R/N) or only the conspecific was reinforced (N/R).

Therefore, in tasks where rats are self-interested – such as in an appetitively/aversively competitive context – and circumstances are well-learned, ACC may contribute more to social attention. In contrast, when different task parameters (i.e., no cost for the observer) promote seemingly more empathetic and pro-social behaviors, then ACC activity might better reflect encoding of the affective information received from other rats. Given the evident influence of ACC-BLA interactions on vicarious learning and decision-making tasks (Allsop et al., 2018; Hernandez-Lallement et al., 2015b; Hernandez-Lallement et al., 2016; Jeon et al., 2010), differential ACC activity profiles in competitive versus non-competitive tasks may modulate downstream social decision making preferences for self-interested versus prosocial behavior, respectively.

In conclusion, here I replicate work showing that neurons in ACC respond to rewards and shocks delivered to oneself and others, but by varying valence within the same task and by omitting outcomes in different trial blocks, I demonstrate that while activity in ACC can represent specific attributes related to conspecific harm, its overall population activity reflects attention in social contexts when there is threat of personal harm.

Chapter 4: Orbitofrontal cortex encodes the value of conspecific approach following foot-shock

4a. Orbitofrontal Cortex and the valuation of task features

An additional but no less important process, engaged in nonsocial as well as social contexts, is valuation – calculating the expected values of stimuli, actions and behavioral strategies, based on their association with internal emotional states. As social animals, we tend to vicariously enjoy feel-good stories and dislike suffering in others, to varying degrees. For example, we may share in the joy when a friend shows us ample evidence of a great vacation they just had, or conversely share in their frustration over a career rejection. But what if we were burnt out and in dire need of a vacation ourselves? What if the career opportunity they were denied would have required a long-distance relocation, requiring our friend to move far away? Would we feel the same way about those outcomes? Social contexts often introduce complex situations that lead to differences in valuation, between outcomes for the self and for those around us. The same is true at more fundamental levels; a conspecific receiving a reward may not always be a positive outcome for a rat if they are urgently seeking food themselves. Similarly, if we are at risk of imminent pain from an expected source (e.g., shock), noticing that another received the punishment may signal temporary safety. Impairments in the ability to make these computations can lead to maladaptive behavior, highlighting the need to better understand the neurophysiology underlying social decision-making processes comparing outcomes between ourselves and others.

A highlighted candidate for evaluating outcomes in social settings is the OFC, which has a well-established role driving behavioral adaptation in response to changes in outcomes during both appetitive and aversive tasks, even if the fundamental mechanism has been a subject of disambiguation (Roesch et al., 2006; Rudebeck & Rich, 2018; Schoenbaum et al., 1998, 2009; Takahashi et al., 2011; Wallis & Kennerley, 2011; Wilson et al., 2014). Primarily, OFC contributes to goal-directed decision-making by updating changes in the expected motivational value of various task-related states (stimuli, responses, choices, rulesets, risk and uncertainty, and internal state), an integrative process dependent on its connectivity with downstream targets, including BLA and the ventral tegmental area (VTA; Roesch et al., 2006; Rudebeck & Rich, 2018; Schoenbaum et al., 1998, 2009; Takahashi et al., 2011). Recent evidence suggested that the OFC may achieve this breadth of computation by evaluating task-related factors dynamically, depending on attentional shifts and the feature's relevance to the goal (Rich et al., 2018; Xie et al., 2018). The signals from OFC, relating to the expected value of responses and their associated outcomes, are critical for flexibility during learning, without which behavioral adaption may come slower or not at all. In fact, disruption of OFC, either through lesions or found in learning-deficit disorders, leads to impairments in extinction and other forms of adaptive learning (Blair et al., 2010; Bouton, 2004; Butter, 1969; Eagle & Baunez, 2010; Itami & Uno, 2002; Mobini et al., 2002; Pickens et al., 2003). As previously mentioned, in social contexts, socially-derived cues may provide an alternative to other salient cues in the environment. Since rats can learn associations from socially-derived cues, these cues must be evaluated with respect to outcomes during the task. If OFC contributes to goal-directed

decision-making in part by highlighting more valuable cues in the environment, how are social cues evaluated against other features, with respect to one's own goal?

4b. Orbitofrontal Cortex – The value of social outcomes

As previous researchers have, one might first consider whether social cues are in fact integrated into OFC processes. Remarkably, there is evidence to suggest OFC encoding is modulated by social contexts (Azzi et al., 2012; Chang et al., 2013; Machado and Bachevalier, 2006), resulting in altered value signals that could influence downstream behavior. Indeed, studies involving OFC disruption in humans (Bechara et al., 2000; Blair et al., 2010; Forbes & Grafman, 2010) and non-human primates (Machado and Bachevalier, 2006) report impairments in expressions of social behavior and development. OFC is also activated during mutual cooperation in a prisoner's dilemma task (Decety et al., 2004), along with nucleus accumbens (NAcc) and caudate nucleus, suggesting OFC contributes to the perceived value of cooperation. Studies disrupting rodent OFC function during social behavior are scarce, but they have previously associated it with increased aggression, suggesting it modulates the emotional regulation of social behaviors (Kuniishi et al., 2017; Jennings et al., 2019; Rudebeck et al., 2007; Wall et al., 2004).

Unlike ACC, which has been shown to respond to other-related outcomes and choices, the majority of OFC neurons appear to focus on self-directed rewards and choices. When monkeys were cued to receive reward either for themselves only, for their partner only or for neither, self-reward elicited the strongest firing whereas reward to the other was reduced (Chang et al., 2013). These findings suggest that OFC may primarily evaluate features of social tasks, including the prospect of rewards to another, with respect to the

self, which is consistent with its non-social functions. However, this does not mean that socially derived cues are generally less valuable. Despite lacking direct interrogation of OFC activity, some research that involves different social outcome contexts has seen behavioral results that suggest the competition between social cues and internal motivators to drive goal-directed behavior may be more complex. For example, although rats tend to avoid actions that harm conspecifics, Hernandez-Lallement and colleagues (2020) found that with enough of a cost (lever resistance) or reward size they will not mind, suggesting that there is a variable limit at which internal cues overcome social cues in value for guiding actions. Further, recent work in mice found ensembles in OFC that differentially encode social and feeding behavior, in an antagonistic relationship that suggests a difference in processing between social and internal goals (Jennings et al., 2019). Thus, OFC integration of social information is critical for complex social cognition, yet the exact function of OFC in social decision-making remains unclear. This is in large part because, despite the advancements made, the mechanisms by which social context modulates value-encoding in OFC are not well-understood.

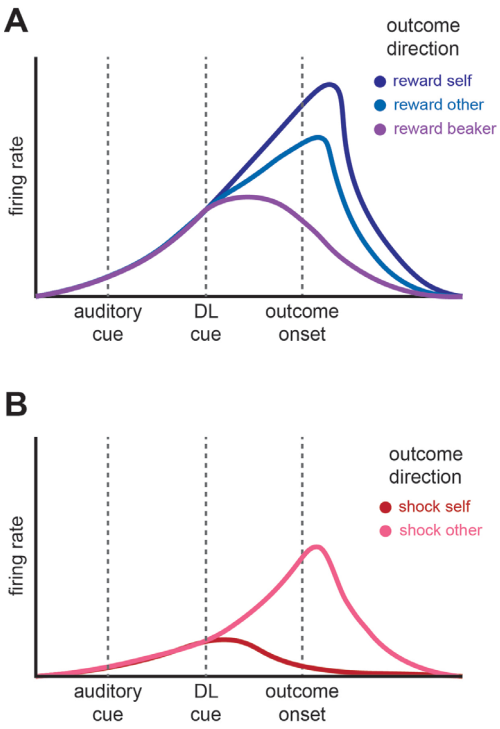


Figure 4.1: Predictions for social outcome encoding in OFC. A) Predicted average firing rates of OFC neurons during reward trials in task. Reward-self outcomes (indigo) will elicit the strongest firing, followed by reward-other (blue). Rewards to the empty beaker (reward-other) in social control sessions will elicit the weakest response (violet). **B**) Predicted average firing rates of OFC neurons during shock trials in task. Shock-self outcomes (red) will show little to no response, given its negative value for the self, while shock-other (pink) outcomes will elicit a strong response.

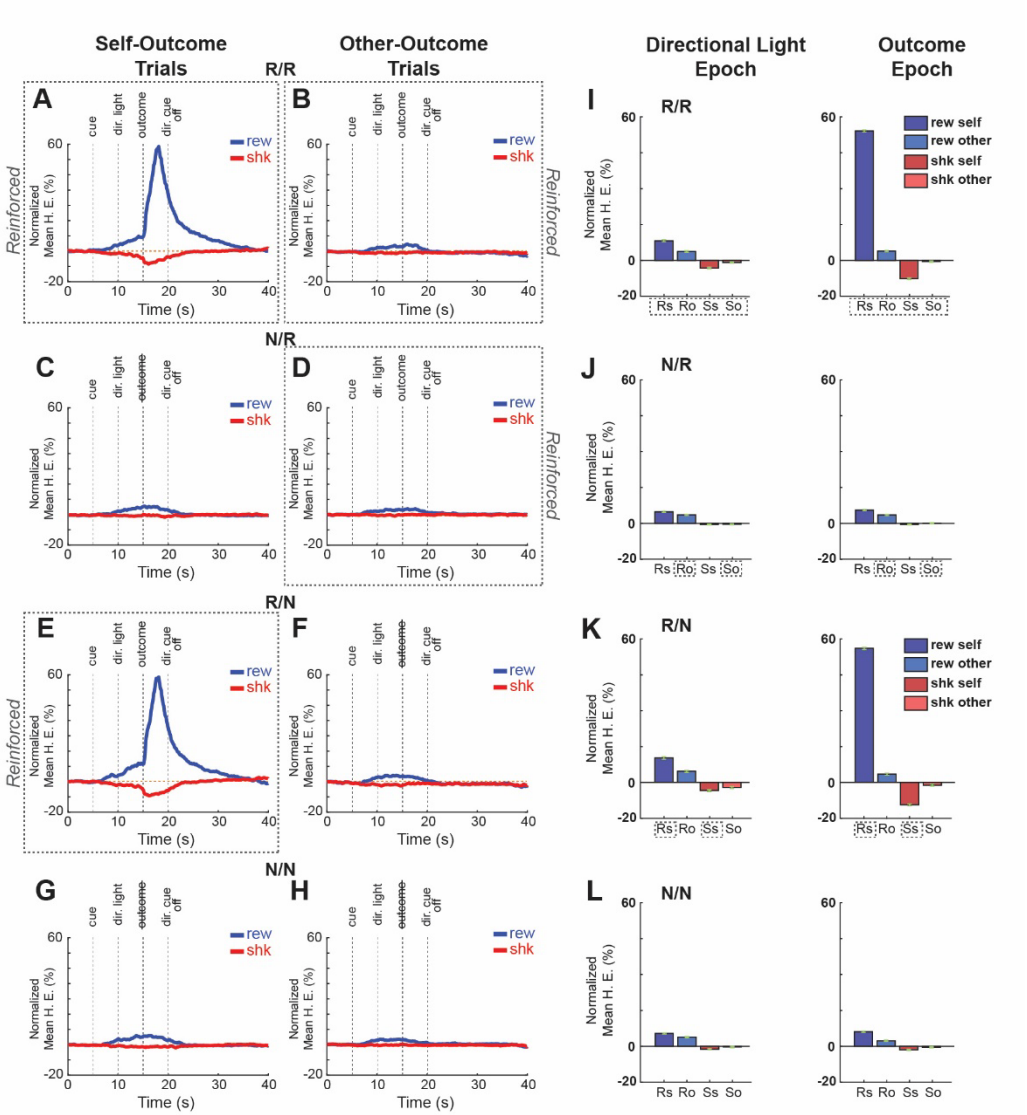
4c. Results

To determine whether rat OFC encoded the value of social outcomes as it does in nonsocial contexts, similar to previous findings in monkeys (Chang et al., 2013), I recorded from the OFC of rats experiencing the task with a conspecific. Based on previous findings regarding reward and shock outcomes in social tasks (Chang et al., 2013; Lichtenberg et al., 2018), I predicted that OFC would encode reward outcomes for the self and other, although outcomes to the other or to neither (equivalent to the beaker in this task) would be reduced in comparison (Fig. 4.1A). I expected shock-self trials to elicit a weak response, if any, since it represented a negative outcome for the self (Fig. 4.1B, 'shock self'). In contrast, shock-other outcomes would mean that the self avoided pain, representing a positive outcome (Fig. 4.1B, 'shock-other').

Rats correctly internalize auditory, directional light cues, and social outcome context

I used food cup beam breaks and video scoring to determine if rats understood the appetitive and aversive aspects of the task, respectively. Figure 4.2A-H shows the average beam breaks into the food cup over trial time for the reward and shock trial-types, normalized against neutral trials and averaged over all recordings. Bar graphs in Figure 4.2I-L show the beam breaks averaged across two trial epochs: the 'directional light epoch' (5s after onset of the directional cue) and the 'outcome epoch' (5s after outcome delivery).

Figure 4.2: Rats learn the predictive value of both outcome and directional cues, modulated by different outcome contexts. Average beam breaks from food cup entry as a percentage of trial time for each outcome type (reward – blue, neutral – dotted orange, shock – red), normalized to neutral responding across each block type (Dotted line boxes indicate whether trials were "Reinforced") for self- (**A,C,E,G**) and other- (**B, D, F, H**) outcome trial. N = 304 sessions (8 rats). Vertical dotted lines indicate task-related events (outcome cue, directional light (dir. light), outcome delivery, directional light off). **I-L)** Averaged food cup entry during the directional light (left) and outcome (right) epochs (5 s). Reinforcement context for each row corresponds to the block type of the same row in **A-H**. Trial types: reward-self (Rs, solid blue), reward-other (Ro, pale-blue), shock-self (Ss, solid red) and shock-other (So, pale-red). There were 4 different trial blocks (60 trials per block; 10 trials per trial-type), during which both rats received outcomes (R/R; where 'R' designates which rat was reinforced; numerator = recording rat; denominator = conspecific), neither rat received outcomes (N/N; 'N' designates 'not reinforced'), only the recording rat was reinforced (R/N) or only the conspecific was reinforced (N/R). During non-reinforced trials all stimuli were presented but shocks and reward were not delivered. 1st row of figures (**A,B,I**) = R/R trials; 2nd row of figures (**C,D,J**) = N/R trials; 3rd row of figures (**E,F,K**) = R/N trials; 4th row of figures (**G,H,L**) = N/N trials.



Consistent with previous results, in blocks where both animals were reinforced (Figure 4.2A-B R/R), percentage of time spent in the food cup was greater and less on reward-self (blue) and shock-self (red) trials, relative to neutral trials (Figure 4.2A), respectively (Lichtenberg et al., 2018; Schneider et al., 2020). After the presentation of the directional light cue (10s from trial start), there was a significant increase in food cup entries for reward-self compared to reward-other trials, demonstrating that rats anticipated the receipt of reward before its delivery (Figure 4.2I, dark vs pale blue; Wilcoxon; z =

6.629, $p < 0.001$). During shock trials, there was a significant decrease in beam breaks during both the directional light and outcome epochs compared to neutral, most strongly for shock-self trials, but also during the directional light cue of shock-other trials (Figure 4.2I; red; Wilcoxon; shock-self: DL: $z = -6.933$, $p < 0.001$; Out: $z = -9.088$, $p < 0.001$; shock-other: DL: $z = -4.033$, $p < 0.001$). A decrease was observed during the outcome epoch of shock-other trials, but it was not significant (Figure 4.2I; pale red; Wilcoxon; Out: $z = -1.689$, $p = 0.091$). Finally, recording rats also altered behavior according to block context in that increases and decreases to rewards and shocks, and their anticipation were reduced in blocks of trials where outcomes were omitted (Fig. 4.2C,G). Thus, food cup entry was strongly regulated by the appetitiveness (reward/shock) and social (self/other) factors of the task across blocks of trials.

Rats show greater freezing and gate approach when at risk for self-shock

Recording rats' understanding of the task is further demonstrated in their aversive response behavior during shock trials, measured through behavioral scoring of video data. Figure 4.3A-D shows freezing observed, as a percentage of total trials for recorded animals for each block, during the 5s-long outcome cue ('cue'), directional light cue ('dir. light ON'), outcome delivery ('outcome') and directional light cue off ('dir. light OFF') epochs of the task. Freezing was defined during the scoring as the sudden absence of movement except for respiration. In accordance with the suppression of food cup entry described above, in R/R blocks recording rats froze more often in shock-self and shock-other trials compared to their neutral counterparts (Figure 4.3A-D; 'dir. light ON': shock-self, $\chi^2 = 35.423$, $p < 0.001$; shock-other, $\chi^2 = 25.806$, $p < 0.001$; 'outcome': shock-self, χ^2

= 61.652, $p < 0.001$; shock-other, $\chi^2 = 7.201$, $p = 0.007$), and froze more often on shock-self trials compared to shock-other trial types during the directional light and outcome epochs ('dir. light ON': $\chi^2 = 10.176$, $p < 0.002$; 'outcome': $\chi^2 = 47.667$, $p < 0.001$). Freezing observed during the directional light cue epoch indicates that rats anticipated shock delivery before the actual outcome. Similarly consistent with my previous results, expression of behavior associated with conspecific distress was dependent on whether the recording rat was at risk of shock (Schneider et al., 2020). During trial blocks where the recording rat was not shocked, but the conspecific was (Fig. 4.3B, N/R), freezing was significantly reduced during the directional light epoch (shock-other in R/R vs N/R: 'dir. light ON': $\chi^2 = 10.760$, $p = 0.001$), replicating previous data. These results suggest that the recording rats expressed vicarious distress less frequently if the likelihood of being shocked themselves was zero. Conversely, in blocks where only the recording rat received shocks (R/N), freezing during shock-other trials was still higher than neutral (Fig. 4.3C, R/N shock-other vs neutral-other, 'dir. light ON': $\chi^2 = 28.267$, $p < 0.001$; 'outcome': $\chi^2 = 17.801$, $p < 0.001$).

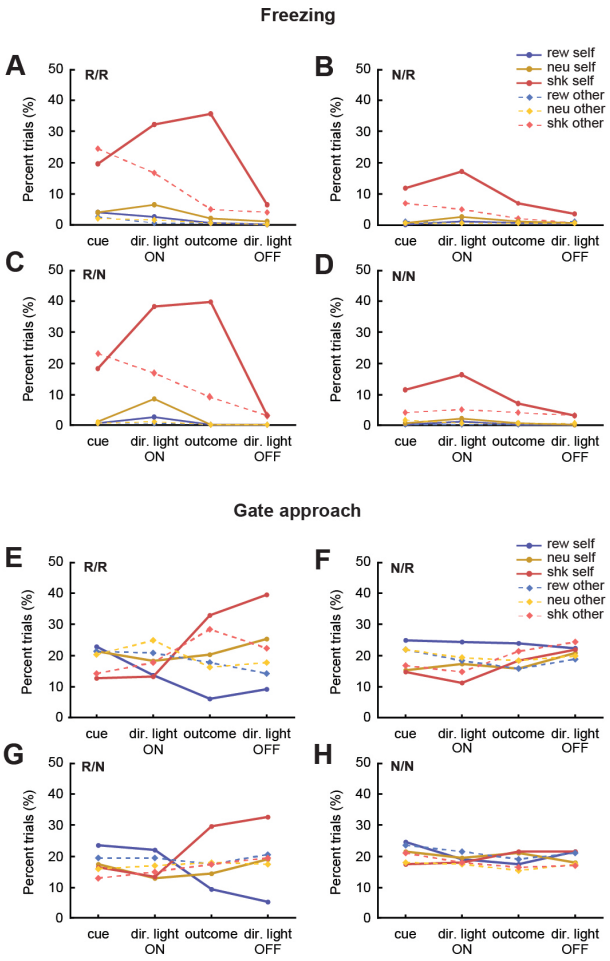


Figure 4.3: Rats show increased freezing and gate approach during shock-self and shock-other trials. A-D) Percentage of trials recorded rats froze during each epoch, for each trial and block type. **E-H)** Percentage of trials recorded rats approached the gate, defined as moving towards or actively interacting at the 'gate,' for each trial and block type. Two sessions were scored for each rat that contributed neural data. Instances of different behavioral events (i.e., froze or did not freeze; approached or did not approach) were counted during each of the 5 s epochs, across both sessions. Recorded rats showed greater freezing in trials where they were reinforced (**A**, **C**), and especially during the directional light cue epoch, for both shock-self and shock-other trials. Gate approach of recorded rats was similarly dependent on reinforcement block, as they were more likely to approach their conspecific when they were reinforced (**E**, **G**). However, they showed increased approach behavior for shock-other trials only when they themselves were at risk for shock (**E** vs **F**).Trial types: reward-self (solid blue), reward-other (dotted light-blue), neutral-self (solid orange), neutral-other (dotted light-orange), shock-self (solid red) and shock-other (dotted light-red). There were 4 different trial blocks, during which both rats received outcomes (R/R; where 'R' designates 'reinforced'; numerator = recording rat; denominator = conspecific), neither rat received outcomes (N/N; 'N' designates which rat was not reinforced), only the recording rat was reinforced (R/N) or only the conspecific was reinforced (N/R). During non-reinforced trials, all stimuli were presented but shocks and reward were not delivered. Statistics are reported in the text.

After experiencing shock, social mammals such as rats tend to approach a conspecific, which has been suggested to provide a social form of relief from their stress (Burkett et al., 2016; Gutzeit et al., 2020; Kiyokawa and Takeuchi, 2017; Kiyokawa et al., 2019; Lungwitz et al., 2014; Mikami et al., 2020). Given its potential social value perceived

during the task, gate approach was scored for all recording rats (Fig. 4.3E-H), defined as the movement and investigation of the recording rat in the direction of the mesh divider. In line with previous findings, when both rats received shock (Fig. 4.3E, R/R), the recording rat approached the conspecific more often on shock-self and shock-other trials compared to neutral trials, most significantly during the outcome and directional light off epochs (Fig. 4.3E, R/R; 'dir. light ON': self, $\chi^2 = 4.763$, $p = 0.03$; other, $\chi^2 = 0.101$, $p = 0.75$; 'outcome': self, $\chi^2 = 15.019$, $p < 0.001$; other, $\chi^2 = 16.563$, $p < 0.001$; 'dir. light OFF': self, $\chi^2 = 17.625$, $p < 0.001$; other, $\chi^2 = 13.041$, $p < 0.001$).

In summary, the behavioral results show that on shock-self trials recording rats froze during the directional light cue on and outcome epochs, and approached the conspecific during the outcome and directional light cue off epochs. These effects were stronger in blocks of trials when the recording rat was at risk of receiving shock. Notably, social control sessions – described below – revealed that recording rats show reduced freezing and higher approach behavior when performing together with the conspecific, while appetitive responding was largely the same.

Average OFC activity increases in anticipation of reward and after self shock

To examine the population activity of OFC during the task, I averaged the normalized firing rate of all recorded neurons (n= 304) across trial time for each trial-type and trial block, illustrated in Figure 4.4A-H. As a population, OFC neurons showed strong increases in firing during reward-self trials especially in blocks of trials where the recording rat was reinforced (Figure 4.4 A,E). They also increased firing for reward outcomes to the other rat (Figure 4.4 B,F), even when the self was not reinforced (Figure 4.4D). Shock

outcomes elicited increases during self-shock trials in blocks where the self was reinforced, but only following shock delivery.

In order to determine the proportion of neurons driving population activity, I characterized all recorded neurons based on their response to reward or shock trials compared to neutral, illustrated in Figure 4.5A-D. To achieve this, for each neuron I computed the normalized difference in firing rate between shock and neutral trials (shock index = shock – neutral/ shock + neutral) and between reward and neutral trials (reward index = reward – neutral/ reward + neutral). Data was collapsed across R/R and R/N blocks for self trials and R/R and N/R for other trials, indices were then calculated as described above for each social and epoch variable (e.g., neural response for reward-self during DL epoch, reward-other during outcome, etc.). Lastly, Wilcoxon tests were used to determine the significance of shifts in the distributions. For reward trials (Fig. 4.5B,D), distributions were positively shifted for 'self' and 'other' trials during both DL and outcome epochs, showing that most OFC neurons increased their firing in response to reward for either self or other (Wilcoxon; R/R+R/N self, DL: $\mu = 0.011$, $p < 0.05$; Outcome: $\mu = 0.092$, $p < 0.001$; other, DL: $\mu = 0.010$, $p < 0.005$; Outcome: $\mu = 0.020$, $p < 0.001$). Examining counts of neurons that showed significant within cell effects (e.g., black bars), the counts of neurons that increased firing for

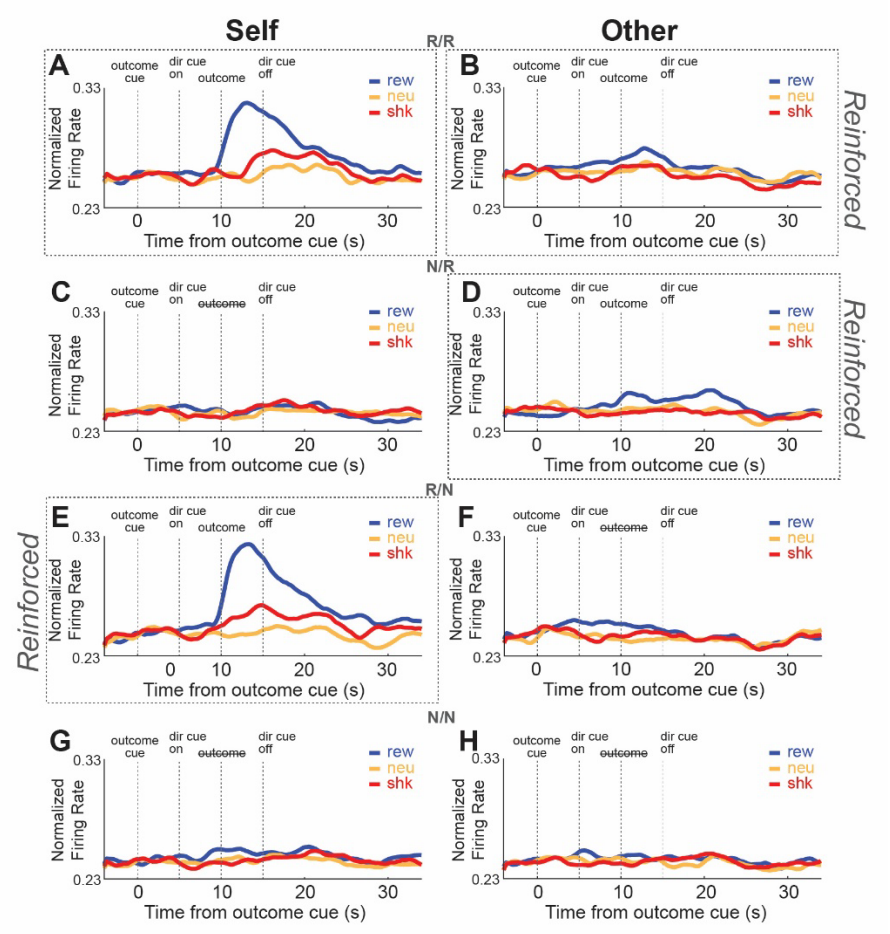


Figure 4.4: OFC population activity during Pavlovian Social Outcome Task. A-H) Normalized mean firing rate of all recorded neurons (n = 304), across each reinforcement block type for reward (blue), shock (red) and neutral (orange) trials. Each row shows neural firing for self- and other-outcome trials (as indicated above each column) for each block (Dotted line boxes indicate whether self or other trials were reinforced for that block). As a population OFC neurons responded strongly to reinforced reward trials (**A**, **E**), including responding to other-outcomes (**B**, **D**), primarily during the outcome epoch (10-15s). Activity increased at times during the directional light cue epoch, as visible even in nonreinforced reward trials (**C**,**G-H**), but it did not ramp up to the higher responding seen in self-reinforced blocks. Firing was also modulated by shock delivery, but only during shock-self trials (**A**, **E**) and later into the outcome epoch, compared to responding in reward trials. Block types: R/R; where 'R' designates 'reinforced'; numerator = recording rat; denominator = conspecific), N/N;'N' designates which rat was not reinforced), R/N and N/R. During non-reinforced trials all stimuli were presented but shocks and reward were not delivered. 1st row of figures (**A,B**) = R/R trials; 2nd row of figures (**C,D**) = N/R trials; 3rd row of figures (**E,F**) = R/N trials; 4th row of figures (**G,H**) =N/N trials.

reward over neutral outnumbered those showing the opposite effect during both the directional light and outcome epoch for both reward-self and reward-other (R/R+R/N self, DL: 12 vs 33, $\chi 2 = 9.71$; $p = 0.002$; Outcome: 20 vs 131, $\chi 2 = 81.45$; $p < 0.001$; other, DL: 2 vs 13, $\chi 2 = 7.92$; $p = 0.005$; Outcome: 6 vs 25, $\chi 2 = 11.52$; $p = 0.0007$).

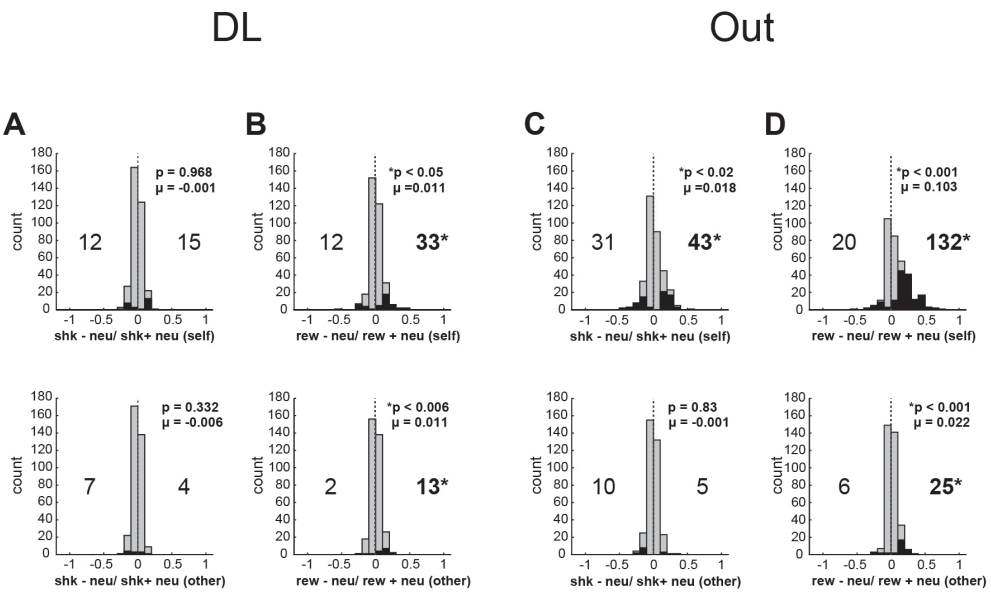


Figure 4.5: OFC neurons tend to fire more during Reward-self and Reward-other trials relative to neutral. (**A-D**) Distributions of counts for neurons selective for shock- or reward-self and other during the outcome epoch, based on calculated index scores (n = 304). Index scores were obtained as the normalized difference between reward (**B, D**) or shock (**A, C**) and neutral firing rates (i.e., shock index = shock - neutral/ shock + neutral; reward index: reward - neutral/ reward + neutral) during directional light cue (DL) and outcome (Out) epochs of self- or other-outcome trials. Counts of cells firing significantly greater than or less than neutral trials are represented by black bars (Wilcoxon; $p < 0.05$). Data was combined across R/R and R/N trial blocks. Wilcoxon tests report significant shifts in distributions and chi-squared tests report significant differences between greater and fewer counts of neurons. Distributions were significantly shifted in the positive direction in reward trials (**B**, **D**) for self- (top) and other-outcomes (bottom), during both the directional light cue and outcome epochs, showing that many OFC neurons most frequently responded to reward outcomes for themselves and their conspecific. Shock trials did not elicit the same response in general (**A**, **C**), but there was a significant number of cells responding to shock outcomes to the self (**C**, top).

Unlike reward trials, for shock trials, distributions were significantly shifted only during the outcome epoch of shock-self trials (Figure 4.5C, top; Wilcoxon; R/R+R/N self,

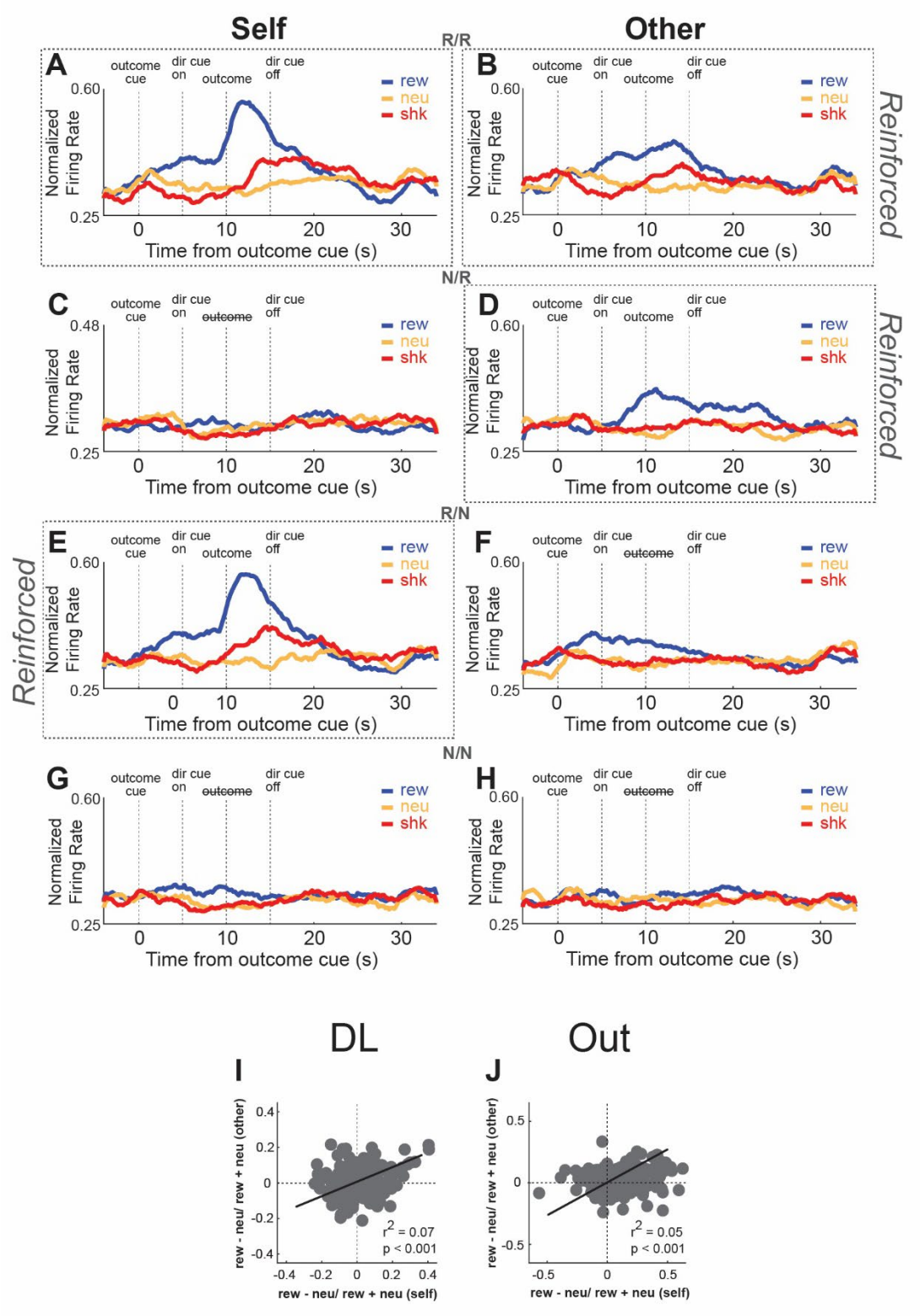
Outcome: $\mu = 0.016$, $p < 0.02$). During the outcome epoch 43 neurons significantly increased firing during shock-self trials relative to neutral-self trials, whereas 31 neurons exhibited activity that was significantly decreased during shock-self trials. Both frequencies are more than one would expect from chance alone (43 out of 304: $\chi^2 = 53.33$, $p < 0.0001$; 31 out of 403: $\chi^2 = 12.37$, $p < 0.0001$). Overall, these results are consistent with previous work demonstrating that single neurons in OFC increasing firing to self-reward. Further, many neurons also decreased firing to shock (31 out of 304), possibly reflecting value in that these neurons also tended to in increase firing during reward trials (Supplemental figure 2).

Here, for the first time in rats, I show that average activity across the population of recorded OFC neurons and at the level of single neurons was stronger on reward-other trials compared to neutral-other trials, similar to primates (Chang et al., 2013). To examine these neurons further I plotted their activity (n = 25) in Figure 4.6. As defined before, a strong increase can be observed in response to rewarded outcomes for the conspecific (Fig. 4.6B). Notably, this activity was present on reward-other trials (Fig. 4.6D) even when the recording rat was not being rewarded, and during non-rewarded reward-self trials (i.e., N/R blocks; Fig. 4.6C) these neurons did not increase firing as observed during R/R and R/N blocks. Thus, consistent with previous reports in primates, OFC neurons appear to genuinely signal rewards delivered to the conspecific even when the availability for self-reward is absent.

These results demonstrate that neurons selected for responsivity on reward-other trials also fired strongly in anticipation and delivery of reward for oneself. To determine if

the was true across the entire population of OFC, I examined the correlations between self and other indices over all neurons (Fig. 4.6I-J). I found that OFC neurons tended to fire similarly for reward, whether they were to 'self' or 'other,' during the directional light (Fig. 4.6I) and outcome epochs (Fig. 4.6J), demonstrating that neurons that tended to fire more or less for reward-self tended to fire more or less strongly for reward-other, respectively.

Figure 4.6: OFC subpopulation responds strongly to reward outcomes for the conspecific, including contexts where the self is not rewarded. A-H) Normalized mean firing rate of a subset of neurons responsive to reward outcomes during 'other' trials (n = 25), across each reinforcement block type for reward (blue), shock (red) and neutral (orange) trials. Each row shows neural firing for self- and other-outcome trials (as indicated above each column) for each block (dotted line boxes indicate whether self or other trials were reinforced for that block). Compared to population activity (**Figure 4**), these cells showed a stronger increase in firing during the directional cue and outcome epochs of reward-other trials (**B**, **D**, **F**). They also displayed the strong responses to reward-self seen in the overall population activity. This was true regardless of whether only the recording rat (**F**; RN) or only the conspecific (**D**; NR) was being reinforced for that block. **E-H**) Correlation scatter plots showing comparisons between index scores of self- versus other-reward trials, during the directional light cue (E-F) and outcome (G-H) epochs. OFC neurons tended to respond similarly between self- and other-reward.


Self
Other
R/R
N/R
R/N
N/N
Reinforced
Reinforced
Reinforced
A
B
C
D
E
F
G
H
outcome cue
dir cue on
outcome
dir cue off
rew
neu
shk
Normalized Firing Rate
Time from outcome cue (s)
DL
Out
I
J
rew - neu/ rew + neu (other)
rew - neu/ rew + neu (self)
r2 = 0.07
p < 0.001
r2 = 0.05
p < 0.001

Social approach is stronger while freezing is reduced when rats are together compared to alone

Thus far I have shown that OFC activity increased during the anticipation and delivery of reward for both self and other. As previously reported, I interpret this signal to reflect the value of reward. Surprisingly, however, I also found that activity in OFC increased after self-shock. Indeed, even in neurons selected for higher firing during reward (Figure 4.6) there was higher activity after self-shock. This signal could be interpreted in a number of different ways. First, it could be argued that OFC is not signaling value but instead reflects attention or arousal that is associated with both appetitive and aversive events. This seems unlikely because increases in activity were not present during the anticipation of shock, even though they were attention-grabbing and arousing as evidenced by the elevated freezing during the auditory cue and directional light epochs. A second interpretation of this data is that activity during shock-self trials also reflects value, as it does for reward trials, possibly reflecting the termination of shock (i.e., release from pain is pleasurable) or the value of positive social actions that promote coping or reduced fear (Burkett et al., 2016; Gutzeit et al., 2020; Kiyokawa and Takeuchi, 2017; Kiyokawa et al., 2019; Lungwitz et al., 2014; Mikami et al., 2020). To dissociate the two, I examined behavior and neural firing during 'alone' sessions. If rats were engaged in positive social behavior during 'together' sessions, then they should show higher freezing and less approach when alone. Further, if increased firing during shock-self trials reflects the value of social approach, then firing in OFC to shock-self trials should be eliminated when the conspecific was not present.

To address this issue, I scored freezing and approach behavior during 'alone' sessions and compared it to results from 'together' sessions, as shown in Figure 4.7. Indeed, compared to sessions 'together,' when alone, recorded rats exhibited overall increased freezing during the directional light cue epoch of shock-self trials, particularly when both rats were reinforced (Fig. 4.7A,C; alone Ss vs together Ss; R/R: 'dir. light ON': $\chi^2 = 4.537$, $p = 0.033$; R/N: 'dir. light ON': $\chi^2 = 0.388$, $p = 0.533$). Thus, rats froze less when they were in the presence of the conspecific. Accompanying reduced freezing on shock-self together trials, I found that gate approach was more pronounced when the conspecific was present. Gate approach was less frequent in sessions alone for shock-self trials in R/R (Fig. 4.7E; alone Ss vs together Ss; 'dir. light OFF': $\chi^2 = 3.43$, $p = 0.064$) and R/N (Fig. 4.7G; alone Ss vs together Ss; 'dir. light OFF': $\chi^2 = 6.13$, $p < 0.002$). When combined, gate approach when the recording rat was reinforced is significantly higher in sessions when together with the conspecific (alone Ss vs together Ss; 'dir. light OFF': $\chi^2 = 9.68$, $p < 0.002$). Approach behavior was also higher for shock-other together trials, during the outcome and directional light cue off epochs of blocks where the other was reinforced (Fig. 4.7E-F; alone So vs together So; RR: 'outcome': $\chi^2 = 11.065$, $p < 0.001$; 'dir. light OFF': $\chi^2 = 2.961$, $p = 0.085$; N/R: 'outcome': $\chi^2 = 4.941$, $p = 0.026$; 'dir. light OFF': $\chi^2 = 10.227$, $p < 0.002$; combined R/R, N/R: 'outcome': $\chi^2 = 16.206$, $p < 0.001$; 'dir. light OFF': $\chi^2 = 12.21$, $p < 0.001$). In conclusion, these results show that when animals experienced the task 'together,' they approached the gate more often, and froze less.

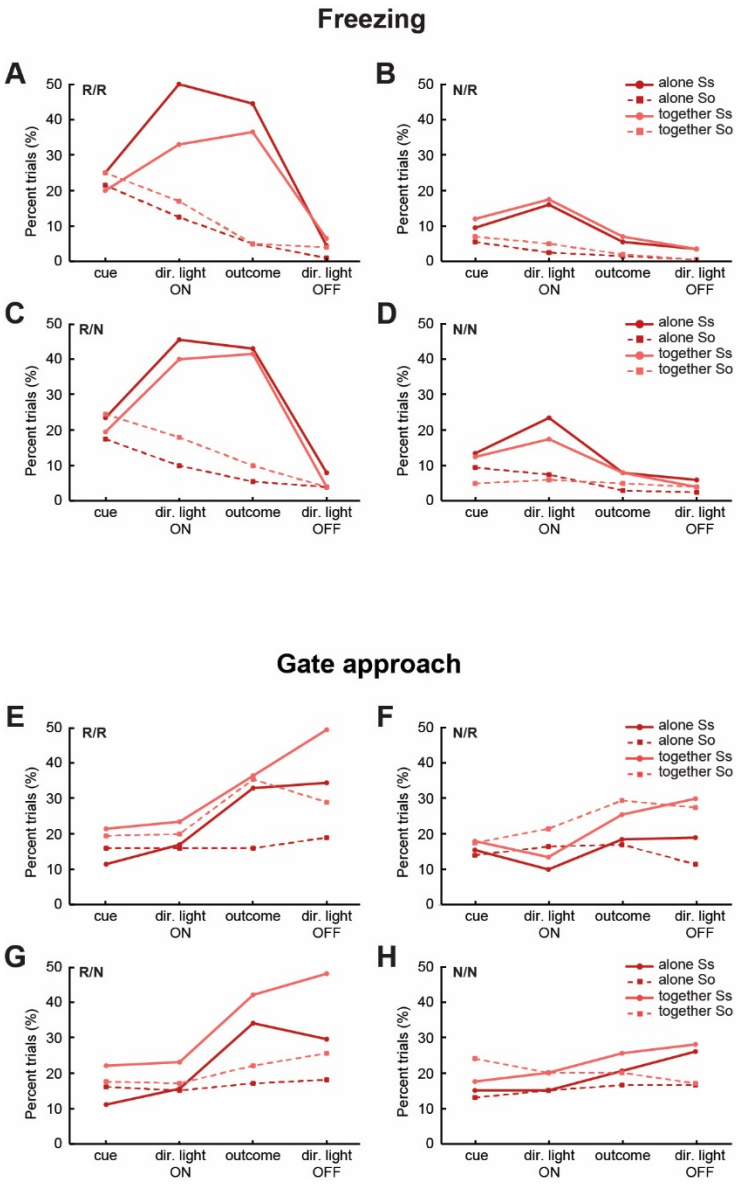


Figure 4.7: Rats freeze more often but approach the divider gate less frequently in task sessions alone. A-H) Percentage of trials recorded where rats froze (**A-D**) or approached the gate (**E-H**) in either *together* (light-red) or *alone* (dark-red) sessions, for each epoch of shock-self (Ss) or shock-other (So) trials across the different block types. Two sessions were scored for each rat that contributed neural data. Instances of different behavioral events (i.e., froze or did not freeze; approached or did not approach) were counted during each of the 5 s epochs, across both sessions. When performing together with their partner, recorded rats displayed reduced freezing during the directional light cue and outcome epochs of Ss trials (**A,C**; solid light-red), compared to when performing alone (solid dark-red), in the R/R block. (E-H) Rats approached the gate less often overall when alone, compared to when performing together with conspecifics. Statistics are reported in the text.

OFC neurons do not fire during shock-self trials when the recording rat is alone

The behavioral results, showing reduced freezing and enhanced social approach during together trials, support my hypothesis that activity during shock-self trials might reflect the value associated with social approach. If true, then activity in OFC should not be high when the conspecific was not present during 'alone' trials. To answer this question,

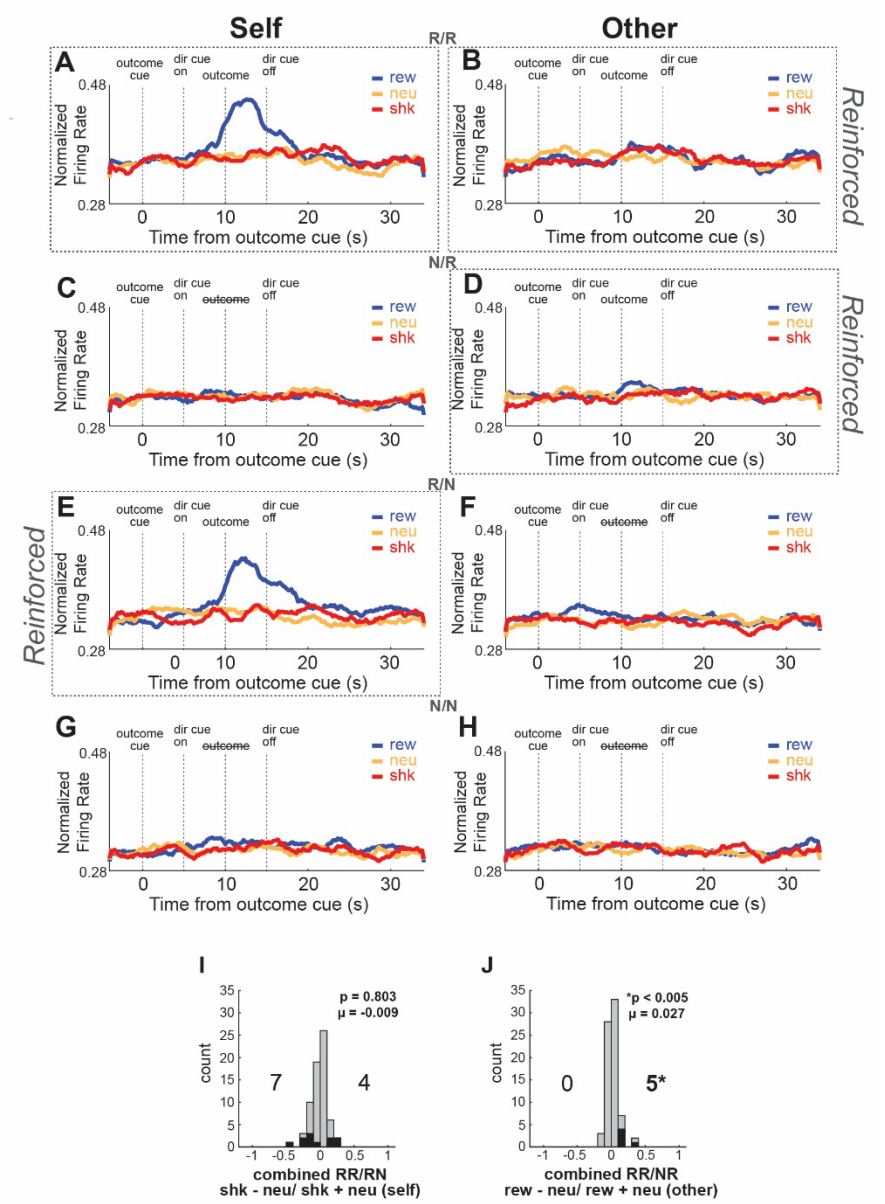
I examined the firing of OFC neurons during the task in 'alone' sessions. The average over all recorded neurons during 'alone' sessions (n = 67), across trial time for each trial-type and trial block is illustrated in Figure 4.8A-H. During 'alone' sessions, I found that OFC activity was similar to what I observed in partnered sessions. That is, neurons increased in firing to the anticipation and delivery of reward (Figure 4.4A,E). Remarkably, an increase in firing for shock-self trials, visible during the outcome phase of 'together' sessions (Figure 4.4A, 4E), appeared to be absent in 'alone' sessions (Figure 4.8A,E). Indeed, the distributions of shock indices (shock-self minus neutral-self/ shock-self plus neutral-self) during the outcome epoch was not significantly shifted when rats were alone (Wilcoxon; $\mu = -0.003$, $p = 0.736$). Only, 6% of neurons recorded during alone trials increased firing on shock-self trials, which was not significantly more than chance (4 out of 67; $\chi^2 = 0.11$; $p = 0.72$).

These results demonstrate that removing the social component of the paradigm also removed increases in firing observed during shock-self trials. Although this was true for shock-self trials, interestingly, this was not the case for increases in firing observed on reward-other trials. The distributions of reward indices (reward-other minus neutral-other/ reward-other plus neutral-other) were still significantly shifted above zero (Wilcoxon; $\mu = 0.024$, $p < 0.005$) and the counts of neurons whose activity was significantly higher on reward-other compared to neutral trials were still in the majority during the outcome epoch (5 vs 0; $\chi^2 = 4.8$; $p = 0.028$). Together, these results suggest that activity in OFC related to self-shock was socially dependent whereas the reward-related responses to self and other were not.

4d. Discussion

As in most nonsocial decision-making environments, the underlying processes to adapt choice behavior in social contexts is critical for achieving both social and homeostatic goals. A large focus of social neuroscience has been to study the neural mechanisms driving prosocial behavior in animals. The expression of prosocial behavior suggests that social information regarding the internal state of conspecifics is integrated in decision-making processes, informing choice of actions. However, it is

Figure 4.8: OFC response following shock-self outcomes is socially-dependent, whereas reward-other response is not. A-H) Normalized mean firing rate of all recorded neurons during social control sessions (n = 67), across each reinforcement block type for reward (blue), shock (red) and neutral (orange) trials. Each row shows neural firing for self- and other-outcome trials (as indicated above each column) for each block (dotted line boxes indicate whether self or other trials were reinforced for that block). In social control sessions, the recorded rat performed the task alone while the 'other' side of the chamber was empty, and food pellets delivered to the 'other' were instead dropped into a bottle to signal receipt to the recorded rat. When compared to population activity in sessions with a conspecific (**Figure 4.4**, **A**, **E**), an increase in firing observed late into the outcome phase of shock-self trials was absent in social control sessions (**A**, **E**). **I-J**) Distributions of counts for neurons selective for shock-self (**I**) or reward-other (**J**) during the outcome epoch, based on calculated index scores. Index scores were obtained as the normalized difference between reward or shock and neutral firing rates (i.e., shock index = shock - neutral/ shock + neutral; reward index: reward - neutral/ reward + neutral) during sessions where recorded rats performed together (**left**; n = 304) with their partner or alone (**right**; n = 67). Counts of cells firing significantly greater than or less than neutral trials are represented by black bars (Wilcoxon; $p < 0.05$). Data was combined across R/R and R/N trial blocks. Wilcoxon tests report significant shifts in distributions and chi-squared tests report significant differences between greater and fewer counts of neurons. Distributions were significantly shifted in the positive direction during the outcome epoch of self-shock trials when recorded rats performed together with conspecifics (**I, left**), but not when alone (**right**). Reward-other trials (**J**) elicited responses in a small but significant number of neurons even while performing alone, when compared to neutral trials.



unclear whether positive outcomes to the other are consistently perceived as positive for the self, leading at times to prosocial behavior, or whether they are contingent on other task-related factors. The OFC is a remarkable candidate for this discerning function, being highly connected with limbic, thalamic and other cortical regions essential for decision-making (Barreiros et al., 2021; Izquierdo, 2017; Price, 2007), and has been extensively shown to be in involved in value-encoding in nonsocial contexts (Roesch et al., 2006;

Rudebeck & Rich, 2018; Schoenbaum et al., 1998, 2009; Takahashi et al., 2011). Although OFC has been studied in social contexts, primarily in nonhuman primates, the means by which it contributes to social functions are not completely understood.

Given its known function evaluating among different rewards, previous research in primates has examined how OFC encoding is modulated by social outcome context, such as during rewards to the self, the partner, to both or to neither. In accordance with OFC function in nonsocial contexts, a majority of neurons were selective to self reward, with a smaller minority signaling reward to other or both (Chang et al., 2013). Specifically, Chang and colleagues found that OFC neurons were largely selective for self reward, showing stronger responses over rewards to the other or even to the both of them, consistent with the well-studied role of OFC in evaluating rewards to the self. However, OFC representation of value has also been shown to be modulated by social context, as a reward shared with others corresponded with a reduction in the firing of value encoding neurons (Azzi et al., 2012). These findings suggest that OFC may primarily evaluate features of social tasks, such as familiarity of the partner, equity of outcomes and even the prospect of outcomes to another, all with respect to the self. Even so, the exact relationship between the value encoding of nonsocial versus social cues remains unclear, especially given that ensembles in OFC may be differentially engaged by these two types of cues (Jennings et al., 2019).

As described previously, I show that OFC activity signaled reward to the conspecific, even when the self was not expecting reward. Further, as previously reported, changes in activity under reward-self trial were more pronounced compared to reward-other (Chang et al., 2013). Unlike previous work, however, I found that removing the

conspecific did not eliminate the reward-other signal, suggesting that in this task the OFC signal related to vicarious reward may partially reflect delivery of reward to adjacent chamber, which may signal benefit to the conspecific when present, but is not entirely dependent on that factor. Interestingly, increased activity in OFC on reward-other trials is in contrast to dopamine (DA) release in nucleus accumbens core (NAc) in rats experiencing a paradigm that consistently delivered reward to the other within the same block, decreasing over trials (Kashtelyan et al., 2014), but it is consistent with DA release in a task similar to the one used (Lichtenberg et al., 2018). Together, these findings suggest that when reward outcome direction (self/other) is ambiguous, the presence of reward may be represented even when nobody can obtain it.

The most novel and surprising finding is the pronounced increase in firing after rats experienced shock. At first, this signal could have been interpreted as reflecting attention or arousal associated with the appetitive and aversive outcomes. However, the response to shock was not present for the anticipation of its delivery, when freezing behavior was most often observed, suggesting it is unlikely to signal the salience of these events. Instead, a more plausible explanation is that OFC activity increases during shock-self trials might reflect the positive value of the offset of shock delivery (pain relief) or of social behaviors that lead to coping mechanisms reducing fear (Burkett et al., 2016; Gutzeit et al., 2020; Kiyokawa and Takeuchi, 2017; Kiyokawa et al., 2019; Lungwitz et al., 2014; Mikami et al., 2020). To dissociate these interpretations from each other I examined behavior and neural firing during alone trials. When alone, rats froze more and approached the adjacent chamber less often, and firing in OFC to shock-trials was not present at the population or single neuron level. These results suggest that the main component of this response is likely

due to the relief the recorded rat may have obtained from proximity to a conspecific, a phenomenon better known as social buffering (Kiyokawa and Hennessy, 2018). Stress coping via social proximity has previously been associated with reduced freezing in fear conditioning protocols (Gutzeit et al., 2020; Kiyokawa and Takeuchi, 2017; Kiyokawa et al., 2019; Mikami et al., 2020). In support of this hypothesis, recent work has found cells in OFC selective to social stimuli, including approach by a conspecific, (Jennings et al., 2019). Together with the results, these findings suggest that approach by others as well as self-initiated approach is evaluated in OFC, potentially promoting them when they have positive value.

OFC is thought to integrate across a wide spectrum of inputs, including sensory cues derived socially and from internal state processing (Gangopadhyay et al., 2021; Jennings et al., 2019; Padoa-Schioppa and Conen, 2017; Rich et al., 2018; Roesch et al., 2006; Rudebeck & Rich, 2018; Schoenbaum et al., 1998, 2009; Takahashi et al., 2011; Walton et al., 2011; Xie et al., 2018). The results presented here bolsters previous research suggesting that OFC evaluates socially-derived information with respect to the self, its internal state and goals, at least when their own reward or pain is at stake (Azzi et al., 2012; Chang et al., 2013; Gangopadhyay et al., 2021; Piva et al., 2019). In contexts where they are not expecting outcomes to themselves, such as during purely observational procedures, OFC may assign a higher value to vicarious rewards, being the more valuable outcome in that scenario. This form of cognitive flexibility is a critical function of OFC, which may achieve this in the complex sphere of social interactions thanks in part to its connections with ACC and basolateral amygdala (BLA), the latter of which shows extensive reciprocal connections with OFC (Barreiros et al., 2021; Lichtenberg et al., 2017; Janak and Tye,

2015; Wassum and Izquierdo, 2015). This interconnectivity may point to the function OFC plays in mediating social choice behavior that leads to a reduced stress response. Recent studies in rodents have found strong links between the BLA and the processing of social information and behavior. In particular, BLA has been found to mediate the response to social affective stimuli (Song et al., 2021), as well as participate in the establishment and expression of social buffering (Gutzeit et al., 2020; Jung et al., 2021; Minami et al., 2019). My findings present the possibility that, via communication with BLA and other mPFC regions, in this task the OFC may promote social approach behavior by assigning positive value to it when it leads to a reduction in fear. Although the exact site of social valence processing remains a focal point of debate, social information processing in BLA along with internal state cues may be integrated with such signals in OFC and ACC in order to determine which outcome or action to value the most in any given social context: pro-self or prosocial.

In conclusion, in a social task that manipulates both reward and shock outcomes between the self and other, as a population OFC encoded the anticipation and delivery of reward, regardless of the recipient, consistent with previous findings in similar contexts. As previously observed, subpopulations further dissected different outcome-related features and reflected value in some circumstances. Remarkably, increases in OFC population activity were associated with increased social approach and reduced freezing behavior, suggesting that during the task OFC also encoded the positive value of approaching the conspecific, which may help reduce the rat's shock-related stress response.

Chapter 5: General Discussion

5a. Summary of results

Successful navigation of the complex social interactions we engage in daily requires that we accurately perceive social cues, integrate them with the emotional states of ourselves and others, and choose an appropriate response depending on what outcome we seek. Conditions in which these skills are impaired often lead to severe difficulties with a variety of social decision-making skills (Bachevalier & Loveland, 2006; King et al., 2006; Lockwood, 2016). These difficulties may range from failing to notice or correctly interpret social cues, to reacting inappropriately to accurately perceived cues, which underscores the efforts both past and present in attempting to better understand how each step of the empathizing process is achieved in the brain (Gangopadhyay et al., 2020; Lockwood, 2016; Lockwood et al., 2020). A reinvigorated momentum, thanks to advanced tools and techniques, has built upon past landmark findings to yield significant advances in recent years, but the exact anatomical and physiological implementations of social cognitive processes remain largely unclear. Part of the difficulty in answering these questions is that, like other complex cognitive tasks, social decision-making engages a brain-wide circuit involving otherwise well-studied regions for guiding goal-directed behavior (Gangopadhyay et al., 2020). Naturally, this creates a more ambiguous, widespread search for social signals across several different brain regions, which has led some to focus (or perhaps broaden) the questions to examine how these processes may have been initially implemented in the brain (Gangopadhyay et al., 2020; Lockwood et al., 2020). Among others, two major possibilities have been proposed: social specialization, from which we would expect social cognition to arise from exclusive neuroanatomy and

physiology, developed specifically for social information processing; or social co-option, from which we would expect social information to be integrated into existing nonsocial mechanisms, applying a 'social' tag, as it were, based on other contingent mechanisms. As it often occurs in the brain, the truth may be more mixed and complex than one or the other, with social specific or co-opted implementations found depending on the mechanism or circuits adapting to social information (Lockwood et al., 2020). Nonetheless, attempts to disambiguate on this debate still have the potential to shed light on how social decision-making processes may be disrupted, making them a valuable scientific pursuit.

Perhaps without the debate in mind originally, studies examining classical decision-making structures have found evidence for social specialization and social co-option in the brain. Two particular candidates for guiding goal-directed behavior during social decision-making have been the ACC, a region known for its versatile encoding of outcome-related information, thought to drive attentional changes to salient, task-related cues (Brockett et al., 2020; Bryden et al., 2011, 2018; Kennerley et al., 2009; Wallis & Kennerley, 2010, 2011), as well as the OFC, which integrates across a broad variety of task-related sensorimotor features in order to promote those relevant to a given goal (Roesch et al., 2006; Rudebeck & Rich, 2018; Schoenbaum et al., 1998, 2009; Takahashi et al., 2011; Wallis & Kennerley, 2011; Wilson et al., 2014). In the ACC of rats, researchers have found neurons that encoded the pain of the conspecific as well as their own, seemingly separate from expressions of fear between the two, respectively (Carrillo et al., 2019). Further, Allsop and colleagues (2018) found projections from ACC to BLA that were critical for acquisition during vicarious –but not classical– fear learning. These findings would suggest that socially specific mechanisms underlie both emotional contagion and vicarious

learning, the ability to share the emotional state of another and learn from this information, respectively. In contrast, in primate OFC most neurons showed the strongest response for rewards given to the self, when compared to rewards for the partner, a result that is consistent with OFC activity in nonsocial contexts, encoding the most valuable option for the self (Chang et al., 2013). This would suggest co-option instead, as social information is integrated into computations against other nonsocial information to determine its influence on behavior. Indeed, when monkeys were pitted between giving reward for the other or none at all, they did prefer the former (Chang et al., 2013), suggesting that prosocial behavior may have been promoted as valuable in that reinforcement context. Thus, these findings would appear to pit both hypotheses against each other.

A caveat with both approaches, however, is that they focus on one valence for outcomes, complicating the interpretation of behavioral and neural results. For example, pain and fear may simply represent two types of salient negative valence outcomes (Carrillo et al., 2019), which the ACC encodes in nonsocial contexts. Without presenting similar outcomes with opposite valence, it is difficult to determine whether ACC activity was reflecting outcome salience or emotional valence. Additionally, most social tasks in rodents omit outcomes to the observer, eliminating the stakes that they personally have for their actions (Allsop et al., 2018; Carrillo et al., 2019; Jeon et al., 2010; Keum and Shin, 2019; Kim et al., 2012). Therefore, to address this gap in the literature and introduce a new paradigm, a task was needed that could manipulate positive and negative valence outcomes within the same experiment, to both a designated ‘self’ rat along with its conspecific partner.

To fill this gap, I used a combined appetitive-aversive social task as described above and in detail in Chapter 2 (Fig. 2). In Chapters 3c and 4c, I presented results replicating previous findings (Lichtenberg et al., 2018), showing that rats are able to internalize the auditory, directional light cues and social context of the outcomes of trials during the task. On rewarded trials, preemptive infrared beam-breaks that started with the auditory cue ramped up during the directional light cue and peaked during the delivery of reward (Fig. 3.2A-B, 'R/R'; Fig. 4.2A-B, 'R/R'). In contrast, neutral trials saw little change in behavior whereas shock trials led to a strong drop in food-seeking. Appetitive responding to other-outcomes followed a similar pattern but reduced in strength. Video scored aversive-related behaviors reflected the appetitive responding results; rats exhibited freezing behavior for self- and other-shock trials, with a clear emphasis for self-directed shock trials, a difference that was visible at the directional light cue onset (Fig. 3.3A, 'R/R'; Fig. 4.3A, 'R/R'). Similarly, approach to the conspecific was present for self- and other-shock outcomes, but it was greatest following shock to the self (Fig. 3.3E; Fig. 4.3E). Lastly, the task also divided trials into blocks that varied the social reinforcement context of outcomes during the block; that is, some blocks only reinforced either the self (R/N) or the other (N/R), or neither (N/N). In contexts where the self was not receiving outcomes (N/R, N/N), behavioral responding was greatly reduced, even when the partner was still receiving reward (Fig. 3.2D; Fig. 4.2D) or shocks (Fig. 3.3B,F; Fig. 4.3B,F). Overall, rats were able to learn whether outcomes were to be a reward, shock or neutral and whether they were to be the recipient, adapting their appetitive and aversive behavior accordingly. Further, vicarious expressions of behavior were reduced when the self was not at risk of shock or reward, indicating that rats also adapted behavior based on the outcome

expectancy of each social outcome context, suggesting empathetic expressions of behavior may instead reflect a projected concern for the self.

One of the goals of my research was to examine the contributions of ACC and OFC to a social task that could achieve the above. In Chapter 3, I showed my results from single-unit recordings in the ACC of rats as they experienced the task with a partner. If ACC encoded outcome valence between self and other outcomes, then it should differ in its firing responses to positive or negative valence outcomes (Fig. 3.1A). Instead, shared responses between valence outcomes would suggest that ACC encodes the salience of social outcomes (Fig. 3.1B), likely in service of attention as it does in nonsocial contexts. ACC activity showed strong increases in firing for both reward and shock outcomes to the self, with a noticeably stronger response to shock-self outcomes (Fig. 3.4A). While shock-other outcomes elicited a reduced response compared to self-shock, reward-other outcomes showed no responses (Fig. 3.4B). Interestingly, this pattern of activity was similar in contexts where the other was not receiving outcomes, including the increase in firing to shock-other trials despite no actual foot-shock to the conspecific (Fig. 3.4E-F). These effects were quantified in the distributions of recorded neurons based on their difference in firing between shock and neutral trials (Fig. 3.5), showing significant positive shifts during the directional light cue and outcome epochs of shock-self trials, including when the other was not at risk of shock while the recording rat was (Fig. 3.5E-F), but not if they were safe, even when the other received shocks (Fig. 3.5C-D). Firing distributions during reward-self outcomes of self-reinforced blocks also showed a significant positive shift (Figure 3.6B), indicating a significant population response for reward outcomes as well. I concluded by showing that shock and reward response indices for all recorded ACC neurons were

correlated, which means that ACC neurons tended to share response profiles for both reward and shock outcomes. Overall, these results are consistent with previous findings in ACC that show firing in response to both self- and other-shock outcomes (Burkett et al., 2016; Carrillo et al., 2019), although not with work in monkeys finding ACC neurons firing for self- and other-rewards (Chang et al., 2013). However, they also present two novel findings. First, vicarious encoding of shock-other outcomes was dependent on the risk of shock to the self, suggesting that the integration of socially-derived information may depend on self-reinforcement. Second, when both positive and negative valence outcomes are presented during a social task, as a population ACC encoded salient outcomes of either valence based on their relevance to the outcome context for the self, as it functions in nonsocial contexts.

Using the same task, my next goal was to determine how OFC evaluated social information when outcome valence, direction and context were modulated during decision-making. In Chapter 4b, I proposed that based on previous findings in monkeys (Chang et al., 2013) and rats (Lichtenberg et al., 2018), OFC activity would assign a higher value to self-reward over other-reward (Fig. 4.1A), whereas other-shock would be preferred as it means no shock for self (Fig. 4.1B). In Chapter 4c, I showed that OFC neurons encoded self- and other-reward outcomes, with a preference for self-reward (Fig. 4.4A-B), consistent with previous findings (Chang et al., 2013). Increases in firing for OFC neurons was significant during both the directional light and outcome epochs of rewarded trials for self and other, but more so for the outcome epochs (Fig. 4.5B,D). Remarkably, reward-other signals were still present when the conspecific was not being reinforced (Fig. 4.4C-D). A closer look at the neurons that were driving vicarious reward responses revealed a

subpopulation that showed stronger firing for both self- and other-rewards (Fig. 4.6A-H). Further, Response indices of all OFC neurons for self- and other-reward trials were correlated (Fig. 4.6I-J), indicating that as a population OFC activity tended to share responses to reward outcomes for self or other.

The most surprising finding, however, was the increased OFC activity in response to shock-self outcomes (Fig. 4.4A,E). While at first it may have appeared to be signaling the end of shock, results from the social control sessions, when rats experienced the task alone, revealed a different possibility. When rats were together, the shock-self response observed coincided with an increase in gate approach (Fig. 4.7E,G, 'light-red') and a decrease in freezing behavior (Fig. 4.7A,C, 'light-red'), suggesting that the observed response may signal the positive value (reduced stress) obtained via approach to the conspecific. Indeed, when rats were alone and the conspecific was not available, gate approach (Fig. 4.7E,G, 'dark-red') was reduced whereas freezing was higher (Fig. 4.7A,C, 'dark-red'). Importantly, when rats were alone, the significant response to shock-self previously observed was no longer present (Fig. 4.8A,E,I). On the other hand, the vicarious reward response visible when rats were together was reduced but still present when the rat was alone (Fig. 4.8B,F,J). Together these findings agreed with work in monkeys showing that OFC encodes vicarious rewards, albeit as less valuable in comparison (Chang et al., 2013). However, vicarious reward responses were still present when the rat was alone, suggesting that this activity was not socially dependent, perhaps reflecting the general potential for reward delivery regardless of recipient. Surprisingly, the unexpected OFC response to shock-self outcomes I found was absent when rats were alone, suggesting that it is a socially-dependent signal. Combined with the socially-dependent differences in

freezing and gate approach between sessions together and alone, the results suggest this OFC activity reflects the positive value that social approach behavior has, when a conspecific is present, since it may lead to stress relief via proximity (Kiyokawa and Hennessy, 2018).

5b. Comparisons between ACC and OFC

The main goal of my research was to examine the contributions of ACC and OFC to social decision-making and determine the nature of these functions. Specifically, I asked whether these regions differentially encoded socially-derived information during the task, compared to nonsocial contexts. Since OFC and ACC are both prefrontal cortical regions highlighted for their potential functions in social decision-making contexts thanks to their well-studied nonsocial roles, part of the value of the work presented here is that I can directly compare my findings between ACC and OFC during the task. In Chapter 3, I found that in the task ACC neurons were activated by shocks to the rat or its partner, consistent with other findings (Carrillo et al., 2019). However, I found this shared response only when the rat was at risk of personal harm. While studies in primates and humans have found activation of ACC for vicarious reward (Chang et al., 2013; Gangopadhyay et al., 2021), I found no significant response in ACC for rewards to the other. Still, population activity shared similar responses to both reward and shock self outcomes. Thus, my findings suggested that in social decision-making contexts, ACC signals outcome-related cues in service of attention, akin to its nonsocial role. Unlike ACC, in Chapter 4 I found that OFC signaled rewards to both self and other, with relative strengths as observed in primates (Chang et al., 2013). Further, the OFC response to other rewards was still present when the

rat was alone, in agreement with findings that OFC evaluates outcomes with respect to self in social contexts (Chang et al., 2013). In line with this thinking, the OFC representation of social approach I found is likely to reflect the positive value of buffering received through conspecific proximity, rather than affiliative reward. OFC has been found to shift value encoding dynamically, depending on changes during the task (Xie et al., 2018), which may further explain why a previously valuable behavior (approach) is no longer promoted when it does not yield a desired goal (stress relief). Coupled with previous findings, the results presented in this work show that in this social task, ACC and OFC both appear to encode cues and outcomes with respect to goals for the self, including positive social interactions with a conspecific.

5c. Social contributions to nonsocial goals

In Chapter 1, I showed a general model for the integration of social information into the existing decision-making circuits in which ACC and OFC participate, along with other cortico-limbic-striatal structures such as BLA (Fig. 1). I suggested that alongside internal and external nonsocial input (Fig. 1, black solid line), social cues presented another stream of information to integrate into existing decision-making circuits (Fig. 1, violet solid line). Based on the diversity of findings in previous research, I proposed that ACC and OFC integrate social cues as they do nonsocial ones, with respect to the self, leading to a dynamic balance in influence between social and nonsocial information depending on what is ultimately relevant to the self (Fig. 5). Indeed, the results I described above demonstrate how ACC and OFC encoded outcome-related information relative to the self during this task. Some of these findings would appear to be in contrast with previous behavioral and neural results suggesting a focus on other-related information processing in ACC or OFC

during social tasks, especially fully or partially observational ones (Allsop et al., 2018; Carrillo et al., 2019; Jeon et al., 2010; Keum and Shin, 2019; Kim et al., 2012). However, as noted before, most of the tasks isolate either the valence of outcomes presented or the outcome expectancy of the observer / recorded animal. The findings presented here need not be in conflict with previous work if we consider that each social decision-making context used may have differentially modulated the balance in influence, between social and nonsocial information.

Observational learning tasks present a context in which the self is not at risk, rendering outcomes to the other more valuable but also making social cues the most informative for the goal for the self, whether that is for a partner to receive food or avoid pain. Therefore, in this context socially-derived information is likely to carry more weight in decision-making processes, driving behavior downstream (Fig. 5A). Conversely, in nonsocial contexts or perhaps for the demonstrators in an observational task, social outcomes and cues are neither relevant nor informative for the self goal (Fig. 5B). Although it can be simpler and tractable to separate social outcomes this way for certain questions, most social decision-making contexts are not as straightforward, presenting multifaceted calculations to be made between outcomes for ourselves or others. As shown in this work, competitive-like social tasks similar to the one I use introduce a more complex situation. In this social task, which presents highly salient and motivating outcomes for the self and other (e.g., food when hungry, pain to avoid), positive or negative outcomes for the conspecific are in conflict with internal, self-directed goals. Moreover, social cues or actions may be considered valuable or not based on how relevant they are for the current

self goal. An examination of the results presented here demonstrates how complex social decision-making contexts may result in a dynamic balance of information weight.

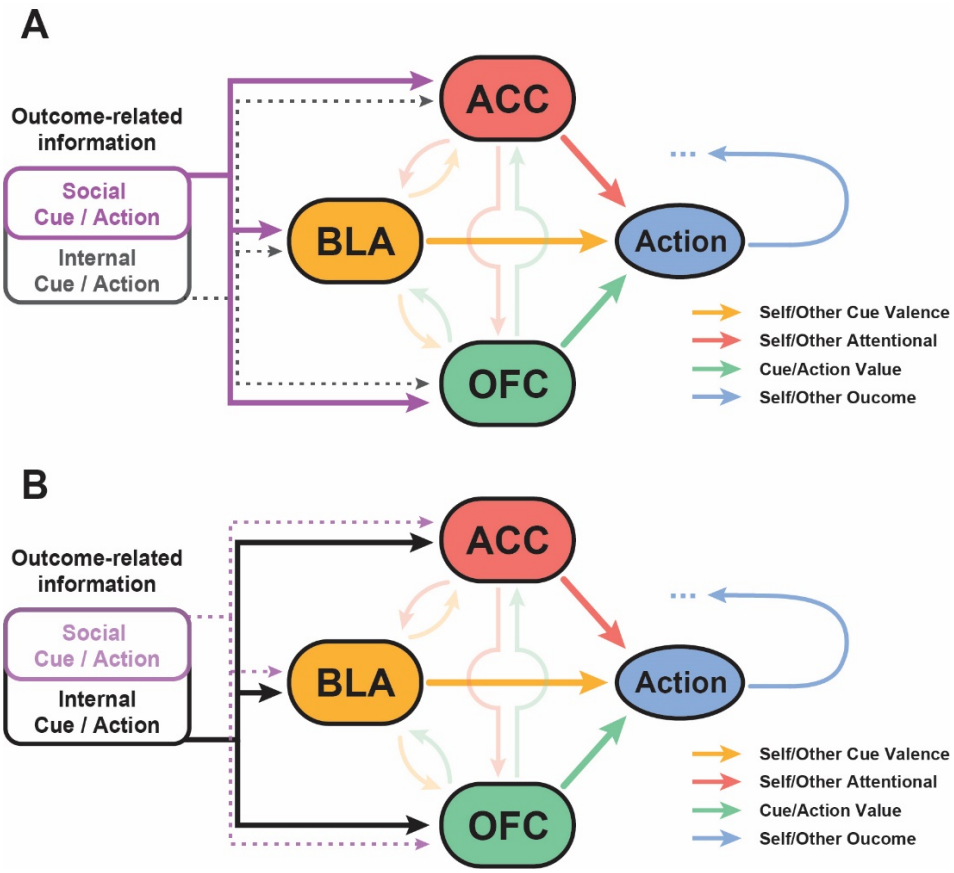


Figure 5: Model for balance of nonsocial vs social influence on decision-making processes in social contexts. A) When balance is weighted in favor of social inputs, socially-derived information is highlighted by **ACC**, prioritizing it for association in **BLA** and for subjective valuation in **OFC**, leading to socially-oriented behavior output from striatum (blue line; e.g., approach, vicarious freezing). **B**) When balance is weighted in favor of nonsocial or internal inputs, socially-derived information is now secondary to other sensorimotor cues to drive behavior instead (e.g., lack of freezing for conspecific fear in contexts where the self is freed from risk of shock), leading to self-focused behavior. In this case, social cues are either not highlighted by ACC, associated with internal cues in BLA, promoted in OFC or all of the above, resulting in little downstream influence. Legends describe the type of information represented by arrow style. Yellow – social/nonsocial valence information; Red, social/nonsocial attentional signals; Green, value information for outcome-related cues/actions; Blue, integrated cue-action-outcome information; Black solid line, information from internal or nonsocial cues (e.g., hunger, satiety); Violet solid line, information derived from social cues or actions, including cues relating to conspecific outcomes, or actions from or involving them (approach, directional light). Dashed lines represent loss of influence.

During the task, recording rats exhibited vicarious behavioral responses to conspecific shocks, where socially-derived information may drive downstream freezing or approach behavior. However, they also showed a behavioral bias in these responses, displaying them more strongly or only when the rats had a strong probability of receiving outcomes themselves (R/R, R/N). This pattern was reflected in the neural activity of ACC, suggesting that during the task, other-outcomes were attended to and drove behavior only when they provided relevant information for the self. As a population OFC encoded both self- and other-reward outcomes, including a subpopulation that appeared to focus on this overlap. This was true even when the self was not expecting rewards (N/R). However, a small but significant number of neurons responded to reward outcomes to an empty chamber, suggesting that instead of vicarious reward, this signal may reflect the anticipation and delivery of reward outcomes in general. During reinforced shock-self trials (R/R, R/N), partnered rats exhibited overall less freezing and more social approach behavior than when they were in the task alone, suggesting that conspecific presence led to a reduction in stress-related response behavior (Kiyokawa and Hennessy, 2018). Also during reinforced shock-self trials, an increase in OFC activity during shock-self trials coincided with the timing of increased approach behavior, but was absent when rats were alone, suggesting that it reflected the positive value of approach to conspecific as a behavioral choice. What these results show is that during this task, there were trials, outcomes or contexts in which social cues or actions were more meaningful, at times driving behavioral responses, but outcome-related information was generally encoded with respect to the self.

When combined with previous findings (Allsop et al., 2018; Carrillo et al., 2019; Chang et al., 2013), the results presented here suggest that in more complex social tasks, the balance between social and nonsocial may shift depending on whether the encoded feature is relevant to a given goal for the self. The results from the task I used suggest that when internal goals are at stake, benefits to the other are not prioritized. However, this balance may vary by task, outcome, familiarity, species, among other factors. Indeed, a boundary might exist between cost to the self and benefit to the other that is modulated by these factors (Hernandez-Lallement et al., 2020). Hunger and pain-aversion are both robust motivators, which may be enough to bias concern for the self during the task. Nevertheless, I believe that this work demonstrates the significance of manipulating outcome valence and context when interrogating neural circuits in social decision-making contexts.

5d. Socially-guided behavior via cortico-limbic-striatal circuits

The goal of my research was to investigate the encoding of social information in ACC and OFC, with the hypothesis that social and nonsocial information are in competition, as it were, for influence during decision-making in social contexts. Above, I discussed how the behavioral and neural results reflected this balance as social outcome contexts changed, as well as what the population activity suggests each region is signaling during the task. Unfortunately, what this work cannot presently address is the exact impact the observed activity of ACC and OFC have on other cortico-limbic-striatal structures. However, I can discuss how my results in ACC and OFC fit into the previous literature on the function of relevant structures during social decision-making.

A region that has gained considerable attention for its potential contributions to social cognition is the basolateral amygdala (BLA), which is also a well-studied, critical component in nonsocial learning and decision-making (Janak and Tye, 2015; Wassum and Izquierdo, 2015). As mentioned before, during observational fear conditioning, inhibition of BLA-projecting ACC neurons prevented vicarious learning during the task (Allsop et al., 2018). BLA recordings obtained during the procedure suggested that ACC inputs were modulating the baseline activity of BLA neurons, potentially facilitating association of perceived social cues during the task. Consistent with those findings, Song et al. (2021) more recently found that inactivation of BLA altered the preferred investigation of social cues in female rats, suggesting that it is required for different socially-related cues to be prioritized. The BLA is not the only structure within amygdala linked to social cognition and behavior, as the intercalated cell mass, central and medial amygdala subregions have been found to be involved in various social behaviors (Andraka et al., 2021; Hong et al., 2014; Hu et al., 2021; Minami et al., 2019). However, BLA neurons are thought to be upstream from more central amygdala structures, activating them differentially to achieve different behavioral outcomes (Janak and Tye, 2015). Together, previous nonsocial and social findings in BLA have spotlighted it as a potential place for the processing of social outcome valence (an emotional mirror perhaps?), especially since it may contain neural ensembles that specifically encode the affective perception of pain (Corder et al., 2019).

The connection between ACC and BLA has been mentioned previously, but the BLA also shares reciprocal connectivity with the OFC (Barreiros et al., 2021; Lichtenberg et al., 2017; Janak and Tye, 2015; Wassum and Izquierdo, 2015). While the relevance of information flow between BLA and OFC has not been directly studied in social contexts,

to my knowledge, it has been investigated in a nonsocial appetitive decision-making context (Lichtenberg et al., 2017). The authors found that BLA projections to OFC mediated the ability of reward-predictive cues to drive appetitive behavior, specifically when outcome expectancy was cue-triggered. This suggests that action-selection guided by OFC value encoding depends on BLA encoding of relevant salient cues. Put together, what has been found on the BLA suggests that during the task I used, ACC drives attention to salient cues that are relevant for a given goal, facilitating its association with internal states in BLA. In turn, attentional signals from ACC and emotional associations made in BLA are integrated in OFC, where available social and nonsocial cues and actions are evaluated to guide goal-directed behavior downstream (Fig. 1; Fig. 5A).

Another potential critical region for social decision-making is, naturally, one also essential for making decisions to be executed: the striatum. Striatum is considered to be a limbic-motor intermediary, broadly integrating reward-related information to implement the most appropriate of available actions (Baez-Mendoza et al., 2013; Cox & Witten, 2019). The dorsomedial striatum (DMS) subregion is itself labeled 'associative striatum,' as it has been found to encode rewarded actions, unrewarded actions, as well as track the value of available actions themselves (Baez-Mendoza et al., 2013; Kawagoe et al., 1998; Samejima et al., 2005). The DMS also receives inputs from both ACC and OFC (Mailly et al., 2013; Murphy and Deutch, 2018), and feeds back via thalamus (Cox & Witten, 2019; Fettes et al., 2017). In particular, manipulation of glutamatergic ACC projections to DMS modulated decision-making during distress (Friedman et al., 2015). In social contexts, DMS in monkeys was found to associate the actions of others with reward outcomes for self or other, distinctly from self actions (Baez-Mendoza et al., 2013; Joiner et al., 2017).

Besides the more typical striatal neurons coding own action and own reward, Baez-Mendoza and colleagues (2013) found neurons that only encoded reward if it was during the conspecific's turn making choices, and some neurons reflected whose turn it was in a given trial. Importantly, when controlling for social context (bottle for a partner), 50% of the aforementioned neurons stopped differentiating between self-other trials. Ventral striatum has also been associated with social information processing (Kashtelyan et al., 2014; Lichtenberg et al., 2018; Smith et al., 2021; Williams et al., 2020; Willuhn et al., 2014). Together, the literature places DMS as a likely candidate region for the integration of social decision-making processes discussed throughout this work.

In conclusion and considering what is known about the DMS, the aforementioned enhancement or attenuation in the influence of socially-derived information throughout the ACC-BLA-OFC circuit (Fig. 5) would be integrated in DMS to guide goal-directed behavior. Specifically, social cues, actions or outcomes may be integrated with nonsocial cues, actions and outcomes in order to yield the complex calculations required for behavioral adaptations required during dynamic or intricate social decision-making contexts (e.g., 'How did their action affect them? How did it make me feel? How did my action affect them?'). Therefore, based on the context and goal of the task, social or nonsocial information would be given priority in modulating how each of these decision-making steps are influenced, ultimately shaping whether behavior is prosocial or not in social contexts – dependent on the goal for the self.

5e. Future directions

The research I presented here sought to probe the neural encoding of social information in cortical regions known to play a role in nonsocial decision-making contexts.

To summarize, I found that during the task, ACC encoded social cues in service of attention (Chapter 3) whereas OFC encoded the value of expected rewards as well as social approach following foot-shock. While I was able to characterize ACC and OFC activity during a task that delivered both appetitive and aversive social outcomes, my results presented new significant questions to be answered, especially regarding the mechanistic impact of my findings.

Do attentional signals from ACC modulate social cue-action encoding in DMS?

To my knowledge, there is no research examining the role of rat DMS in dissociating between self and other actions and rewards, but the evidence found in analogous regions of nonhuman primates (Baez-Mendoza et al., 2013; Joiner et al., 2017), along with my own results from cortical regions that target striatum (ACC, OFC), suggests that rat brains may also share some of the neurophysiological properties seen in higher-order mammals. Properly responding and adapting to social cues is essential to succeed in daily social interactions. Together with my results, disruptions observed in ASD (Guillon et al., 2014; Del Bianco et al., 2021) and findings in monkeys where ACC lesions prevented the acquisition of a prosocial preference (Basile et al., 2020), suggest that attending to social cues is a vital component for adaptive social behavior. Thus, a better understanding of how attentional signals regulate the integration of social information is needed to elucidate the mechanisms underlying social neuropsychiatric conditions in turn.

Further experiments using the novel task shown here could attempt to tackle this problem, by including optogenetic manipulations to selectively inactivate ACC during socially-relevant time periods, such as the directional light cue and outcome epochs, while

recording from DMS neurons. If social cues are dependent on ACC-mediated attentional processes to gain influence over internal or nonsocial external cues for the self, then silencing it during socially-relevant epochs would disrupt encoding in DMS, leading to maladaptive behavior during the task. Given that freezing was lower for shock-other trials (Fig. 3.3) and that ACC activity was strongly modulated by shock (Fig. 3.4), conspecific shocks may provide social cues that in self-reinforced cases signal a reduced probability of shock for the self. Thus, I would predict that optogenetic inhibition of ACC during the directional light and outcome epochs would lead to increased freezing behavior despite the reduced expectation for self-shock.

Dorsal striatum has been found to contain neurons that encoded self-reward during other-focused trials (Baez-Mendoza et al., 2013), along with neurons encoding other combinations of task properties. Therefore, when ACC is functioning as normal during the task I would expect DMS neurons will similarly display a spectrum of task-related activity, including cells that encode outcome-related information during other-directed trials in the task. However, DMS responses between appetitive and aversive outcomes should be opposite or independent, since food-seeking and freezing reflect opposing movement strategies. Thus, DMS would contain neurons that show significant changes in firing during the directional light cue and outcome phases of the task, for either direction (self, other) but specifically including cells selective to reward-other and shock-other trials. These results alone would provide evidence for the social modulation of decision-making processes in DMS of rats, which has not been previously established in the literature. On the other hand, when ACC is inhibited during the task, DMS neurons that would otherwise be inactive or attenuated during other-outcomes would instead show stronger responses,

promoting freezing or food-seeking when shock or reward are not likely, respectively. If accurate, these results would support the view that ACC plays an important role in social contexts by facilitating integration of socially-derived information into decision-making processes, such as the coding of response-outcome associations seen in DMS, ultimately driving behavioral adaptation in social contexts.

Does OFC encoding of approach behavior promote social approach following shock?

In my experiment from Chapter 4, I found that OFC appeared to encode the positive value that approach to the conspecific may have for the rat after it received a shock. Specifically, OFC neurons showed a strong response during the outcome epoch of shock-self outcomes (Fig. 4.4) that was absent when the rat experienced the task alone (Fig. 4.8), suggesting that it was a socially-dependent signal. This signal also coincided with an increase in approach and reduction in freezing observed when rats were together (Fig. 4.3). Likewise, when rats were alone they showed overall greater freezing but less approach (Fig. 4.7). Based on these results, I concluded that rats approached their conspecific more often after receiving a shock, in order to obtain stress relief via social buffering (Kiyokawa and Hennessy, 2018). In turn, this behavior led to a reduction in cue-triggered fear responding, making it a choice that OFC assigns positive value to when the conspecific is present to provide a buffer.

To test whether this OFC signal influences behavioral choices, the OFC of rats could be optogenetically silenced during the task, particularly during and following shock-self outcomes. If OFC encodes the positive value of social approach behavior following shock to the self, promoting it as a choice for action-selection, then disruption of this signal

should prevent the behavior from being encouraged. As a result, I would expect this manipulation to eliminate the observed differences in approach and freezing behavior between sessions where rats were together versus alone. That is, approach behavior when rats are together would not be greater than sessions alone. With a reduction in approach to the conspecific, I would also expect freezing behavior when rats are together to rise to levels similar to sessions alone. These results would suggest that OFC encoding of valuable social behaviors, such as approach to a conspecific, is critical for making behavioral decisions based on social cues but directed to self goals, which in the case of social buffering can provide a distressed animal with some relief.

The proposed studies above have the potential to further advance the current findings on cortical contributions to a social task that manipulates both reward and shock outcomes. Future efforts could also modify the task to yield an operant approach to the features it employs, allowing for the interrogation of neural circuits with contingent self actions to match with social cues and outcomes presented. Additionally, some of the proposed experiments could be more incisive if the task variables are focused or honed on specific context combinations, especially once characterization of neural responses to both valence social outcomes is obtained. Lastly, as mentioned early on, in the broader pursuit for understanding the neural mechanisms underlying social decision-making, the ACC, OFC, BLA and DMS are among many other regions highlighted for their latent contributions to decision-making in social contexts. Combined with advanced neurophysiological tools and techniques, the social task described here has the potential to offer an improved characterization of other candidate regions and what they contribute to social decision-making processes in the brain.

References

1. Allsop SA, Wichmann R, Mills F, et al. Corticoamygdala Transfer of Socially Derived Information Gates Observational Learning. *Cell*. 2018;173(6):1329-1342.e18. doi:10.1016/j.cell.2018.04.004

2. Andraka K, Kondrakiewicz K, Rojek-Sito K, et al. Distinct circuits in rat central amygdala for defensive behaviors evoked by socially signaled imminent versus remote danger. *Current Biology*. 2021;31(11):2347-2358.e6. doi:10.1016/j.cub.2021.03.047

3. Apps MAJ, Rushworth MFS, Chang SWC. The Anterior Cingulate Gyrus and Social Cognition: Tracking the Motivation of Others. *Neuron*. 2016;90(4):692-707. doi:10.1016/j.neuron.2016.04.018

4. Atsak P, Orre M, Bakker P, et al. Experience Modulates Vicarious Freezing in Rats: A Model for Empathy. Ferrari PF, ed. *PLoS ONE*. 2011;6(7):e21855. doi:10.1371/journal.pone.0021855

5. Azzi JCB, Sirigu A, Duhamel J-R. Modulation of value representation by social context in the primate orbitofrontal cortex. *Proceedings of the National Academy of Sciences*. 2012;109(6):2126-2131. doi:10.1073/pnas.1111715109

6. Bachevalier J, Loveland KA. The orbitofrontal–amygdala circuit and self-regulation of social–emotional behavior in autism. *Neuroscience & Biobehavioral Reviews*. 2006;30(1):97-117. doi:10.1016/j.neubiorev.2005.07.002

7. Baez-Mendoza R, Harris CJ, Schultz W. Activity of striatal neurons reflects social action and own reward. *Proceedings of the National Academy of Sciences*. 2013;110(41):16634-16639. doi:10.1073/pnas.1211342110

8. Baird AD, Scheffer IE, Wilson SJ. Mirror neuron system involvement in empathy: A critical look at the evidence. *Social Neuroscience*. 2011;6(4):327-335. doi:10.1080/17470919.2010.547085

9. Barreiros IV, Panayi MC, Walton ME. Organization of Afferents along the Anterior–posterior and Medial–lateral Axes of the Rat Orbitofrontal Cortex. *Neuroscience*. 2021;460:53-68. doi:10.1016/j.neuroscience.2021.02.017

10. Basile BM, Schafroth JL, Karaskiewicz CL, Chang SWC, Murray EA. The anterior cingulate cortex is necessary for forming prosocial preferences from vicarious reinforcement in monkeys. *PLoS Biol*. 2020;18(6):e3000677. doi:10.1371/journal.pbio.3000677

11. Bechara A, Damasio H, Damasio AR. Emotion, decision making and the orbitofrontal cortex. *Cerebral cortex*. 2000;10(3):295-307.

12. Ben-Ami Bartal I, Decety J, Mason P. Empathy and Pro-Social Behavior in Rats. *Science*. 2011;334(6061):1427-1430. doi:10.1126/science.1210789

13. Ben-Ami Bartal I, Rodgers DA, Bernardez Sarria MS, Decety J, Mason P. Pro-social behavior in rats is modulated by social experience. *eLife*. 2014;3:e01385. doi:10.7554/eLife.01385

14. Bissonette GB, Powell EM, Roesch MR. Neural structures underlying set-shifting: Roles of medial prefrontal cortex and anterior cingulate cortex. *Behavioural Brain Research*. 2013;250:91-101. doi:10.1016/j.bbr.2013.04.037

15. Blair RJR. Psychopathy, frustration, and reactive aggression: The role of ventromedial prefrontal cortex. *br j psychol*. 2010;101(3):383-399. doi:10.1348/000712609X418480

16. Blair RJR, Morris JS, Frith CD, Perrett DI, Dolan RJ. Dissociable neural responses to facial expressions of sadness and anger. *Brain*. 1999;122(5):883-893. doi:10.1093/brain/122.5.883

17. Bouton ME. Context and Behavioral Processes in Extinction. *Learning & Memory*. 2004;11(5):485-494. doi:10.1101/lm.78804

18. Brockett AT, Tennyson SS, deBettencourt CA, Gaye F, Roesch MR. Anterior cingulate cortex is necessary for adaptation of action plans. *Proc Natl Acad Sci USA*. 2020;117(11):6196-6204. doi:10.1073/pnas.1919303117

19. Bruchey AK, Jones CE, Monfils M-H. Fear conditioning by-proxy: social transmission of fear during memory retrieval. *Behavioural brain research*. 2010;214(1):80-84.

20. Bryden DW, Johnson EE, Tobia SC, Kashtelyan V, Roesch MR. Attention for Learning Signals in Anterior Cingulate Cortex. *Journal of Neuroscience*. 2011;31(50):18266-18274. doi:10.1523/JNEUROSCI.4715-11.2011

21. Bryden DW, Roesch MR. Executive Control Signals in Orbitofrontal Cortex during Response Inhibition. *Journal of Neuroscience*. 2015;35(9):3903-3914. doi:10.1523/JNEUROSCI.3587-14.2015

22. Bryden DW, Brockett AT, Blume E, Heatley K, Zhao A, Roesch MR. Single Neurons in Anterior Cingulate Cortex Signal the Need to Change Action During Performance of a Stop-change Task that Induces Response Competition. *Cerebral Cortex*. 2018;29(3):1020-1031. doi:10.1093/cercor/bhy008

23. Burgos-Robles A, Gothard KM, Monfils MH, Morozov A, Vicentic A. Conserved features of anterior cingulate networks support observational learning across species. *Neuroscience & Biobehavioral Reviews*. 2019;107:215-228. doi:10.1016/j.neubiorev.2019.09.009

24. Burkett JP, Andari E, Johnson ZV, Curry DC, de Waal FBM, Young LJ. Oxytocin-dependent consolation behavior in rodents. *Science*. 2016;351(6271):375-378. doi:10.1126/science.aac4785

25. Bush G, Luu P, Posner MI. Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*. 2000;4(6):215-222. doi:10.1016/S1364-6613(00)01483-2

26. Butter CM. Perseveration in extinction and in discrimination reversal tasks following selective frontal ablations in Macaca mulatta. *Physiology & Behavior*. 1969;4(2):163-171.

27. Carrillo M, Han Y, Migliorati F, Liu M, Gazzola V, Keysers C. Emotional Mirror Neurons in the Rat's Anterior Cingulate Cortex. *Current Biology*. 2019;29(8):1301-1312.e6. doi:10.1016/j.cub.2019.03.024

28. Chang SWC, Gariépy J-F, Platt ML. Neuronal reference frames for social decisions in primate frontal cortex. *Nat Neurosci*. 2013;16(2):243-250. doi:10.1038/nn.3287

29. Chen Q, Panksepp JB, Lahvis GP. Empathy Is Moderated by Genetic Background in Mice. Bartolomucci A, ed. *PLoS ONE*. 2009;4(2):e4387. doi:10.1371/journal.pone.0004387

30. Church RM. Emotional reactions of rats to the pain of others. *Journal of Comparative and Physiological Psychology*. 1959;52(2):132-134. doi:10.1037/h0043531

31. Corder G, Ahanonu B, Grewe BF, Wang D, Schnitzer MJ, Scherrer G. An amygdalar neural ensemble that encodes the unpleasantness of pain. *Science*. 2019;363(6424):276-281. doi:10.1126/science.aap8586

32. Cox J, Witten IB. Striatal circuits for reward learning and decision-making. *Nat Rev Neurosci*. 2019;20(8):482-494. doi:10.1038/s41583-019-0189-2

33. de Waal FBM, Preston SD. Mammalian empathy: behavioural manifestations and neural basis. *Nat Rev Neurosci*. 2017;18(8):498-509. doi:10.1038/nrn.2017.72

34. de Waal FBM. Putting the Altruism Back into Altruism: The Evolution of Empathy. *Annu Rev Psychol*. 2008;59(1):279-300. doi:10.1146/annurev.psych.59.103006.093625

35. Decety J, Jackson PL, Sommerville JA, Chaminade T, Meltzoff AN. The neural bases of cooperation and competition: an fMRI investigation. *NeuroImage*. 2004;23(2):744-751. doi:10.1016/j.neuroimage.2004.05.025

36. Del Bianco T, Mason L, Charman T, et al. Temporal Profiles of Social Attention Are Different Across Development in Autistic and Neurotypical People. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. 2021;6(8):813-824. doi:10.1016/j.bpsc.2020.09.004

37. Devinsky O, Morrell MJ, Vogt BA. Contributions of anterior cingulate cortex to behaviour. *Brain*. 1995;118(1):279-306. doi:10.1093/brain/118.1.279

38. Eagle DM, Baunez C. Is there an inhibitory-response-control system in the rat? Evidence from anatomical and pharmacological studies of behavioral inhibition. *Neuroscience & Biobehavioral Reviews*. 2010;34(1):50-72. doi:10.1016/j.neubiorev.2009.07.003

39. Etkin A, Egner T, Kalisch R. Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in Cognitive Sciences*. 2011;15(2):85-93. doi:10.1016/j.tics.2010.11.004

40. Etkin A, Egner T, Peraza DM, Kandel ER, Hirsch J. Resolving Emotional Conflict: A Role for the Rostral Anterior Cingulate Cortex in Modulating Activity in the Amygdala. *Neuron*. 2006;51(6):871-882. doi:10.1016/j.neuron.2006.07.029

41. Fettes P, Schulze L, Downar J. Cortico-Striatal-Thalamic Loop Circuits of the Orbitofrontal Cortex: Promising Therapeutic Targets in Psychiatric Illness. *Front Syst Neurosci*. 2017;11:25. doi:10.3389/fnsys.2017.00025

42. Finger EC, Marsh AA, Blair KS, et al. Disrupted Reinforcement Signaling in the Orbitofrontal Cortex and Caudate in Youths With Conduct Disorder or Oppositional Defiant Disorder and a High Level of Psychopathic Traits. *AJP*. 2011;168(2):152-162. doi:10.1176/appi.ajp.2010.10010129

43. Forbes CE, Grafman J. The role of the human prefrontal cortex in social cognition and moral judgment. *Annual review of neuroscience*. 2010;33:299-324.

44. Friedman A, Homma D, Gibb LG, et al. A Corticostriatal Path Targeting Striosomes Controls Decision-Making under Conflict. *Cell*. 2015;161(6):1320-1333. doi:10.1016/j.cell.2015.04.049

45. Gallese V, Fadiga L, Fogassi L, Rizzolatti G. Action recognition in the premotor cortex. *Brain*. 1996;119(2):593-609.

46. Gangopadhyay P, Chawla M, Dal Monte O, Chang SWC. Prefrontal–amygdala circuits in social decision-making. *Nat Neurosci*. 2021;24(1):5-18. doi:10.1038/s41593-020-00738-9

47. Guillon Q, Hadjikhani N, Baduel S, Rogé B. Visual social attention in autism spectrum disorder: Insights from eye tracking studies. *Neuroscience & Biobehavioral Reviews*. 2014;42:279-297. doi:10.1016/j.neubiorev.2014.03.013

48. Gutzeit VA, Ahuna K, Santos TL, et al. Optogenetic reactivation of prefrontal social neural ensembles mimics social buffering of fear. *Neuropsychopharmacol*. 2020;45(6):1068-1077. doi:10.1038/s41386-020-0631-1

49. Hayden BY, Heilbronner SR, Pearson JM, Platt ML. Surprise Signals in Anterior Cingulate Cortex: Neuronal Encoding of Unsigned Reward Prediction Errors Driving Adjustment in Behavior. *Journal of Neuroscience*. 2011;31(11):4178-4187. doi:10.1523/JNEUROSCI.4652-10.2011

50. Hernandez-Lallement J, Attah AT, Soyman E, Pinhal CM, Gazzola V, Keysers C. Harm to Others Acts as a Negative Reinforcer in Rats. *Current Biology*. 2020;30(6):949-961.e7. doi:10.1016/j.cub.2020.01.017

51. Hernandez-Lallement J, van Wingerden M, Marx C, Srejic M, Kalenscher T. Rats prefer mutual rewards in a prosocial choice task. *Front Neurosci*. 2015;8. doi:10.3389/fnins.2014.00443

52. Hernandez-Lallement J, van Wingerden M, Schäble S, Kalenscher T. A Social Reinforcement Learning Hypothesis of Mutual Reward Preferences in Rats. In: Wöhr M, Krach S, eds. *Social Behavior from Rodents to Humans*. Vol 30. Springer International Publishing; 2016:159-176. doi:10.1007/7854_2016_436

53. Hernandez-Lallement J, van Wingerden M, Schäble S, Kalenscher T. Basolateral amygdala lesions abolish mutual reward preferences in rats. *Neurobiology of Learning and Memory*. 2016;127:1-9. doi:10.1016/j.nlm.2015.11.004

54. Hong W, Kim D-W, Anderson DJ. Antagonistic Control of Social versus Repetitive Self-Grooming Behaviors by Separable Amygdala Neuronal Subsets. *Cell*. 2014;158(6):1348-1361. doi:10.1016/j.cell.2014.07.049

55. Hu RK, Zuo Y, Ly T, et al. An amygdala-to-hypothalamus circuit for social reward. *Nat Neurosci*. 2021;24(6):831-842. doi:10.1038/s41593-021-00828-2

56. Itami S, Uno H. Orbitofrontal cortex dysfunction in attention-deficit hyperactivity disorder revealed by reversal and extinction tasks. *Neuroreport*. 2002;13(18):2453-2457.

57. Izquierdo A. Functional Heterogeneity within Rat Orbitofrontal Cortex in Reward Learning and Decision Making. *J Neurosci*. 2017;37(44):10529-10540. doi:10.1523/JNEUROSCI.1678-17.2017

58. Janak PH, Tye KM. From circuits to behaviour in the amygdala. *Nature*. 2015;517(7534):284-292. doi:10.1038/nature14188

59. Jennings JH. Interacting neural ensembles in orbitofrontal cortex for social and feeding behaviour. Published online 2019:26.

60. Jeon D, Kim S, Chetana M, et al. Observational fear learning involves affective pain system and Cav1.2 Ca2+ channels in ACC. *Nat Neurosci*. 2010;13(4):482-488. doi:10.1038/nn.2504

61. Joiner J, Piva M, Turrin C, Chang SWC. Social learning through prediction error in the brain. *npj Science Learn*. 2017;2(1):8. doi:10.1038/s41539-017-0009-2

62. Jung T, Jang M, Noh J. Role of Medial Prefrontal Cortical Neurons and Oxytocin Modulation in the Establishment of Social Buffering. *Exp Neurobiol*. 2021;30(1):48-58. doi:10.5607/en20038

63. Kashtelyan V, Lichtenberg NT, Chen ML, Cheer JF, Roesch MR. Observation of Reward Delivery to a Conspecific Modulates Dopamine Release in Ventral Striatum. *Current Biology*. 2014;24(21):2564-2568. doi:10.1016/j.cub.2014.09.016

64. Kawagoe R, Takikawa Y, Hikosaka O. Expectation of reward modulates cognitive signals in the basal ganglia. *Nat Neurosci*. 1998;1(5):411-416. doi:10.1038/1625

65. Kennerley SW, Dahmubed AF, Lara AH, Wallis JD. Neurons in the Frontal Lobe Encode the Value of Multiple Decision Variables. *Journal of Cognitive Neuroscience*. 2009;21(6):1162-1178. doi:10.1162/jocn.2009.21100

66. Kikusui T, Winslow JT, Mori Y. Social buffering: relief from stress and anxiety. *Phil Trans R Soc B*. 2006;361(1476):2215-2228. doi:10.1098/rstb.2006.1941

67. Kim A, Keum S, Shin H-S. Observational fear behavior in rodents as a model for empathy. *Genes, Brain and Behavior*. 2019;18(1):e12521. doi:10.1111/gbb.12521

68. Kim EJ, Kim ES, Covey E, Kim JJ. Social Transmission of Fear in Rats: The Role of 22-kHz Ultrasonic Distress Vocalization. Chapouthier G, ed. *PLoS ONE*. 2010;5(12):e15077. doi:10.1371/journal.pone.0015077

69. Kim S, Matyas F, Lee S, Acsady L, Shin H-S. Lateralization of observational fear learning at the cortical but not thalamic level in mice. *Proceedings of the National Academy of Sciences*. 2012;109(38):15497-15501. doi:10.1073/pnas.1213903109

70. King JA, Blair RJR, Mitchell DGV, Dolan RJ, Burgess N. Doing the right thing: A common neural circuit for appropriate violent or compassionate behavior. *NeuroImage*. 2006;30(3):1069-1076. doi:10.1016/j.neuroimage.2005.10.011

71. Kiyokawa Y, Li Y, Takeuchi Y. A dyad shows mutual changes during social buffering of conditioned fear responses in male rats. *Behavioural Brain Research*. 2019;366:45-55. doi:10.1016/j.bbr.2019.03.024

72. Kiyokawa Y, Takeuchi Y. Social buffering ameliorates conditioned fear responses in the presence of an auditory conditioned stimulus. *Physiology & Behavior*. 2017;168:34-40. doi:10.1016/j.physbeh.2016.10.020

73. Kleberg JL, Selbing I, Lundqvist D, Hofvander B, Olsson A. Spontaneous eye movements and trait empathy predict vicarious learning of fear. *International Journal of Psychophysiology*. 2015;98(3):577-583. doi:10.1016/j.ijpsycho.2015.04.001

74. Knapska E, Nikolaev E, Boguszewski P, et al. Between-subject transfer of emotional information evokes specific pattern of amygdala activation. *Proceedings of the National Academy of Sciences*. 2006;103(10):3858-3862. doi:10.1073/pnas.0511302103

75. Knapska E, Mikosz M, Werka T, Maren S. Social modulation of learning in rats. *Learning & memory*. 2010;17(1):35-42.

76. Kuniishi H, Ichisaka S, Matsuda S, Futora E, Harada R, Hata Y. Chronic Inactivation of the Orbitofrontal Cortex Increases Anxiety-Like Behavior and Impulsive Aggression, but Decreases Depression-Like Behavior in Rats. *Front Behav Neurosci*. 2017;10. doi:10.3389/fnbeh.2016.00250

77. Lichtenberg NT, Lee B, Kashtelyan V, et al. Rat behavior and dopamine release are modulated by conspecific distress. *eLife*. 2018;7:e38090. doi:10.7554/eLife.38090

78. Lichtenberg NT, Pennington ZT, Holley SM, et al. Basolateral Amygdala to Orbitofrontal Cortex Projections Enable Cue-Triggered Reward Expectations. *J Neurosci*. 2017;37(35):8374-8384. doi:10.1523/JNEUROSCI.0486-17.2017

79. Lindström B, Haaker J, Olsson A. A common neural network differentially mediates direct and social fear learning. *NeuroImage*. 2018;167:121-129. doi:10.1016/j.neuroimage.2017.11.039

80. Lockwood PL. The anatomy of empathy: Vicarious experience and disorders of social cognition. *Behavioural Brain Research*. 2016;311:255-266. doi:10.1016/j.bbr.2016.05.048

81. Lockwood PL, Apps MAJ, Chang SWC. Is There a 'Social' Brain? Implementations and Algorithms. *Trends in Cognitive Sciences*. 2020;24(10):802-813. doi:10.1016/j.tics.2020.06.011

82. Lockwood PatriciaL, Apps MAJ, Roiser JP, Viding E. Encoding of Vicarious Reward Prediction in Anterior Cingulate Cortex and Relationship with Trait Empathy. *J Neurosci*. 2015;35(40):13720-13727. doi:10.1523/JNEUROSCI.1703-15.2015

83. Luk C-H, Wallis JD. Choice Coding in Frontal Cortex during Stimulus-Guided or Action-Guided Decision-Making. *Journal of Neuroscience*. 2013;33(5):1864-1871. doi:10.1523/JNEUROSCI.4920-12.2013

84. Lungwitz EA, Stuber GD, Johnson PL, et al. The Role of the Medial Prefrontal Cortex in Regulating Social Familiarity-Induced Anxiolysis. *Neuropsychopharmacol*. 2014;39(4):1009-1019. doi:10.1038/npp.2013.302

85. Machado CJ, Bachevalier J. The impact of selective amygdala, orbital frontal cortex, or hippocampal formation lesions on established social relationships in rhesus monkeys (Macaca mulatta). *Behavioral neuroscience*. 2006;120(4):761.

86. Mailly P, Aliane V, Groenewegen HJ, Haber SN, Deniau J-M. The Rat Prefrontostriatal System Analyzed in 3D: Evidence for Multiple Interacting Functional Units. *Journal of Neuroscience*. 2013;33(13):5718-5727. doi:10.1523/JNEUROSCI.5248-12.2013

87. McDonald AJ. Cortical pathways to the mammalian amygdala. *Progress in Neurobiology*. 1998;55(3):257-332. doi:10.1016/S0301-0082(98)00003-3

88. Meyza KZ, Bartal IB-A, Monfils MH, Panksepp JB, Knapska E. The roots of empathy: Through the lens of rodent models. *Neuroscience & Biobehavioral Reviews*. 2017;76:216-234. doi:10.1016/j.neubiorev.2016.10.028

89. Mikami K, Kiyokawa Y, Ishii A, Takeuchi Y. Social buffering enhances extinction of conditioned fear responses by reducing corticosterone levels in male rats. *Hormones and Behavior*. 2020;118:104654. doi:10.1016/j.yhbeh.2019.104654

90. Minami S, Kiyokawa Y, Takeuchi Y. The lateral intercalated cell mass of the amygdala is activated during social buffering of conditioned fear responses in male rats. *Behavioural Brain Research*. 2019;372:112065. doi:10.1016/j.bbr.2019.112065

91. Mobini S, Body S, Ho M- Y., et al. Effects of lesions of the orbitofrontal cortex on sensitivity to delayed and probabilistic reinforcement. *Psychopharmacology*. 2002;160(3):290-298. doi:10.1007/s00213-001-0983-0

92. Munger SD, Leinders-Zufall T, McDougall LM, et al. An Olfactory Subsystem that Detects Carbon Disulfide and Mediates Food-Related Social Learning. *Current Biology*. 2010;20(16):1438-1444. doi:10.1016/j.cub.2010.06.021

93. Murphy MJM, Deutch AY. Organization of afferents to the orbitofrontal cortex in the rat. *J Comp Neurol*. 2018;526(9):1498-1526. doi:10.1002/cne.24424

94. Noritake A, Ninomiya T, Isoda M. Suppl. Social reward monitoring and valuation in the macaque brain. *Nat Neurosci*. 2018;21(10):1452-1462. doi:10.1038/s41593-018-0229-7

95. Oberliessen L, Hernandez-Lallement J, Schäble S, van Wingerden M, Seinstra M, Kalenscher T. Inequity aversion in rats, Rattus norvegicus. *Animal Behaviour*. 2016;115:157-166. doi:10.1016/j.anbehav.2016.03.007

96. Oberliessen L, Kalenscher T. Social and Non-social Mechanisms of Inequity Aversion in Non-human Animals. *Front Behav Neurosci*. 2019;13:133. doi:10.3389/fnbeh.2019.00133

97. Olsson A, Nearing KI, Phelps EA. Learning fears by observing others: the neural systems of social fear transmission. *Social Cognitive and Affective Neuroscience*. 2007;2(1):3-11. doi:10.1093/scan/nsm005

98. Padoa-Schioppa C, Conen KE. Orbitofrontal Cortex: A Neural Circuit for Economic Decisions. *Neuron*. 2017;96(4):736-754. doi:10.1016/j.neuron.2017.09.031

99. Panksepp J, Panksepp JB. Toward a cross-species understanding of empathy. *Trends in Neurosciences*. 2013;36(8):489-496. doi:10.1016/j.tins.2013.04.009

100. Pickens CL, Saddoris MP, Setlow B, Gallagher M, Holland PC, Schoenbaum G. Different Roles for Orbitofrontal Cortex and Basolateral Amygdala in a Reinforcer Devaluation Task. *J Neurosci*. 2003;23(35):11078-11084. doi:10.1523/JNEUROSCI.23-35-11078.2003

101. Piva M, Velnoskey K, Jia R, Nair A, Levy I, Chang SWC. The dorsomedial prefrontal cortex computes task-invariant relative subjective value for self and other. *eLife*. 2019;8:e44939. doi:10.7554/eLife.44939

102. Preston SD, de Waal FBM. Empathy: Its ultimate and proximate bases. *Behav Brain Sci*. 2002;25(1):1-20. doi:10.1017/S0140525X02000018

103. Price JL. Definition of the Orbital Cortex in Relation to Specific Connections with Limbic and Visceral Structures and Other Cortical Regions. *Annals of the New York Academy of Sciences*. 2007;1121(1):54-71. doi:10.1196/annals.1401.008

104. Rich EL, Stoll FM, Rudebeck PH. Linking dynamic patterns of neural activity in orbitofrontal cortex with decision making. *Current Opinion in Neurobiology*. 2018;49:24-32. doi:10.1016/j.conb.2017.11.002

105. Rizzolatti G, Fadiga L, Gallese V, Fogassi L. Premotor cortex and the recognition of motor actions. *Cognitive brain research*. 1996;3(2):131-141.

106. Roesch MR, Esber GR, Li J, Daw ND, Schoenbaum G. Surprise! Neural correlates of Pearce-Hall and Rescorla-Wagner coexist within the brain: Neural correlates of RW and PH. *European Journal of Neuroscience*. 2012;35(7):1190-1200. doi:10.1111/j.1460-9568.2011.07986.x

107. Roesch MR, Taylor AR, Schoenbaum G. Encoding of Time-Discounted Rewards in Orbitofrontal Cortex Is Independent of Value Representation. *Neuron*. 2006;51(4):509-520. doi:10.1016/j.neuron.2006.06.027

108. Rudebeck PH, Izquierdo A. Foraging with the frontal cortex: A cross-species evaluation of reward-guided behavior. *Neuropsychopharmacol*. Published online August 18, 2021. doi:10.1038/s41386-021-01140-0

109. Rudebeck PH, Rich EL. Orbitofrontal cortex. *Current Biology*. 2018;28(18):R1083-R1088. doi:10.1016/j.cub.2018.07.018

110. Rudebeck PH, Walton ME, Millette BHP, Shirley E, Rushworth MFS, Bannerman DM. Distinct contributions of frontal areas to emotion and social behaviour in the rat: Emotion, social behaviour and the frontal cortex. *European Journal of Neuroscience*. 2007;26(8):2315-2326. doi:10.1111/j.1460-9568.2007.05844.x

111. Samejima K. Representation of Action-Specific Reward Values in the Striatum. *Science*. 2005;310(5752):1337-1340. doi:10.1126/science.1115270

112. Sato N, Tan L, Tate K, Okada M. Rats demonstrate helping behavior toward a soaked conspecific. *Anim Cogn*. 2015;18(5):1039-1047. doi:10.1007/s10071-015-0872-2

113. Schneider KN, Sciarillo XA, Nudelman JL, Cheer JF, Roesch MR. Anterior Cingulate Cortex Signals Attention in a Social Paradigm that Manipulates Reward and Shock. *Current Biology*. 2020;30(19):3724-3735.e2. doi:10.1016/j.cub.2020.07.039

114. Schoenbaum G, Chiba AA, Gallagher M. Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nat Neurosci*. 1998;1(2):155-159. doi:10.1038/407

115. Schoenbaum G, Roesch MR, Stalnaker TA, Takahashi YK. A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. *Nat Rev Neurosci*. 2009;10(12):885-892. doi:10.1038/nrn2753

116. Shackman AJ, Salomons TV, Slagter HA, Fox AS, Winter JJ, Davidson RJ. The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat Rev Neurosci*. 2011;12(3):154-167. doi:10.1038/nrn2994

117. Sivaselvachandran S, Acland EL, Abdallah S, Martin LJ. Behavioral and mechanistic insight into rodent empathy. *Neuroscience & Biobehavioral Reviews*. 2018;91:130-137. doi:10.1016/j.neubiorev.2016.06.007

118. Smith ML, Asada N, Malenka RC. Anterior cingulate inputs to nucleus accumbens control the social transfer of pain and analgesia. *Science*. 2021;371(6525):153-159. doi:10.1126/science.abe3040

119. Song Z, Swarna S, Manns JR. Prioritization of social information by the basolateral amygdala in rats. *Neurobiology of Learning and Memory*. 2021;184:107489. doi:10.1016/j.nlm.2021.107489

120. Takahashi YK, Roesch MR, Wilson RC, et al. Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nat Neurosci*. 2011;14(12):1590-1597. doi:10.1038/nn.2957

121. van Gurp S, Hoog J, Kalenscher T, van Wingerden M. Vicarious reward unblocks associative learning about novel cues in male rats. *eLife*. 2020;9:e60755. doi:10.7554/eLife.60755

122. Vázquez D, Pribut HJ, Burton AC, Tennyson SS, Roesch MR. Prior cocaine self-administration impairs attention signals in anterior cingulate cortex. *Neuropsychopharmacol*. 2020;45(5):833-841. doi:10.1038/s41386-019-0578-2

123. Wall PM, Blanchard RJ, Yang M, Blanchard DC. Differential effects of infralimbic vs. ventromedial orbital PFC lidocaine infusions in CD-1 mice on defensive responding in the mouse defense test battery and rat exposure test. *Brain Research*. 2004;1020(1-2):73-85. doi:10.1016/j.brainres.2004.06.008

124. Wallis JD, Kennerley SW. Heterogeneous reward signals in prefrontal cortex. *Current Opinion in Neurobiology*. 2010;20(2):191-198. doi:10.1016/j.conb.2010.02.009

125. Wallis JD, Kennerley SW. Contrasting reward signals in the orbitofrontal cortex and anterior cingulate cortex: Contrasting reward signals in the orbitofrontal cortex. *Annals of the New York Academy of Sciences*. 2011;1239(1):33-42. doi:10.1111/j.1749-6632.2011.06277.x

126. Walton ME, Behrens TEJ, Noonan MP, Rushworth MFS. Giving credit where credit is due: orbitofrontal cortex and valuation in an uncertain world: Orbitofrontal cortex: value assignment and comparison. *Annals of the New York Academy of Sciences*. 2011;1239(1):14-24. doi:10.1111/j.1749-6632.2011.06257.x

127. Wassum KM, Izquierdo A. The basolateral amygdala in reward learning and addiction. *Neuroscience & Biobehavioral Reviews*. 2015;57:271-283. doi:10.1016/j.neubiorev.2015.08.017

128. Williams AV, Duque-Wilckens N, Ramos-Maciel S, et al. Social approach and social vigilance are differentially regulated by oxytocin receptors in the nucleus accumbens. *Neuropsychopharmacol*. 2020;45(9):1423-1430. doi:10.1038/s41386-020-0657-4

129. Willuhn I, Tose A, Wanat MJ, et al. Phasic Dopamine Release in the Nucleus Accumbens in Response to Pro-Social 50 kHz Ultrasonic Vocalizations in Rats. *Journal of Neuroscience*. 2014;34(32):10616-10623. doi:10.1523/JNEUROSCI.1060-14.2014

130. Wilson RC, Takahashi YK, Schoenbaum G, Niv Y. Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron*. 2014;81(2):267-279. doi:10.1016/j.neuron.2013.11.005

131. Xiao X, Zhang Y-Q. A new perspective on the anterior cingulate cortex and affective pain. *Neuroscience & Biobehavioral Reviews*. 2018;90:200-211. doi:10.1016/j.neubiorev.2018.03.022

132. Xie Y, Nie C, Yang T. Covert shift of attention modulates the value encoding in the orbitofrontal cortex. *eLife*. 2018;7:e31507. doi:10.7554/eLife.31507