Social value unblocks Pavlovian reinforcement learning in male rats

Authors: Sander van Gurp*, Jochen Hoog, Tobias Kalenscher, Marijn van Wingerden

Social Rodent Lab, Department of Experimental Psychology, Heinrich-Heine-University, Düsseldorf, Germany

Correspondence concerning this article should be addressed to Sander van Gurp, Department of Experimental Psychology, Heinrich-Heine-University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany. Email: sander.gurp@hhu.de

Marijn van Wingerden, Jochen Hoog and Sander van Gurp were supported by the Volkswagen Stiftung Freigeist fellowship, AZ88216

Abstract

Many species, including humans, are sensitive to social signals and the valuation of these social cues is important in maintaining social interactions. Social value, derived from social cues, thus could act as a reinforcer and influence reward learning. Here, we introduce a novel task in rats that investigates if social value can drive reinforcement learning about novel stimuli. Using the blocking/unblocking paradigm originally developed in the (non-social) reinforcement learning literature, we found that when actor rats have fully learned a stimulus-reward association producing reward for themselves, adding a cue that predicted an additional reward delivery to a partner rat unblocked associative learning about this cue in the actor rats. In contrast, additional cues that did not predict additional reward remained blocked from acquiring associative value. In a control experiment where putative social cue exchange between the partnered rats was prevented, the normally unblocked cues now remained blocked as expected. Taken together, these results suggest that social value can drive reinforcement learning in rats, and that the transmission of social cues is necessary for this learning, driven by social value, to occur.

Author Contributions:

Conceptualization: MvW and SvG; Methodology: MvW, SvG and JH; Formal Analysis: SvG; Investigation: SvG and JH; Resources: MvW and TK; Writing – Original Draft: SvG; Writing – Review & Editing: MvW, TK and SvG; Visualization; SvG; Supervision, MvW; Funding Acquisition: MvW; Project Administration: SvG, MvW.

Introduction

The need for frequent and meaningful social interactions is a fundamental human motivation [1] inherent in social relationships and contributes to mental and physical health [2]. From an evolutionary perspective, the ability to acquire value from a positive social interaction may have stimulated the development of mutual cooperative social interactions [3], the preference of prosocial outcomes to familiar partners [4,5] and the helping of others in need [6]. One of the driving forces behind these complex social behaviours could be, among others, the positive value that is derived from social interactions [7]. Social interactions contain social signals that can carry value to oneself, such as friendly faces in humans [8], appetitive pictures of conspecifics in monkeys [9] and playback of positive 50 kHz calls in rats [10]. The rewarding aspects of social reward have furthermore been shown to reinforce Pavlovian learning about novel stimuli. Humans for example learn to respond faster to stimuli that become associated with positive social reinforcement [11] and monkeys choose stimuli that predicted a reward delivery to a conspecific more than the stimuli that predicted no reward delivery [12]. It has also been found, using operant instrumental conditioning paradigms, that rats can be conditioned to press a lever to access various degrees of social contact and interactions with another rat [13,14], and that rats are also willing to work for social play periods [15]. It however remains unclear whether in rats, next to humans and monkeys, social value can drive learning about novel stimuli in an associative stimulus-outcome learning setting. Previous research has indicated that vocalisations of rats in the 50 kHz range, thought to represent positive affect [16], lead to both an approach response [10] and dopamine release in the Nucleus Accumbens NAcc [17]: two important components that can drive reinforcement learning of new cues [18,19]. There is furthermore evidence suggesting that observing another rat being rewarded is rewarding by itself as it is accompanied by 50 kHz vocalisations and dopamine release in the NAcc of the observer rat [20].

In this paper, we examine the hypothesis that social value can act as a reinforcer and thus can also drive new learning. To investigate our hypothesis, we use a well-established behavioural paradigm in associative learning called blocking. Kamin [21] found in simple stimulus-outcome association tasks that if new stimuli are added to a stimulus that already fully predicts a reward, associative learning about those additional stimuli will be blocked. Reinforcement learning about additional stimuli can become unblocked, however, by an increase in reward value or a change in reward identity contingent on the presentation of the new stimuli. This additional or change in value is then thought to be associated to these new stimuli and thus alters their incentive value [22]. We hypothesise that rewarding social outcomes, such as sugar pellet deliveries to a partner rat, will also be capable to unblock learning about newly added stimuli, indicative of an increased value of mutual-rewards relative to own-rewards. We tested this hypothesis by adopting a task from McDannald et al. [23] where unblocking is operationalised by adding additional pellet deliveries conditional on a second cue presented in compound with a learned cue that already fully predicted reward. We modified this task in such a way that the second cue is now followed by a food reward delivery to a partner rat, rather than increasing one's own reward. In addition, a third control cue added in compound to the learned

cue (on different trials) was not followed by food reward delivery to a partner rat. Concretely, we thus hypothesized that the associative learning about the second stimulus would become unblocked by mutual-reward outcomes, that is, by adding social value. In contrast, the third cue should remain blocked from acquiring associative value. We indeed found that, when tested in extinction, the unblocked second cue had acquired more associative value, as indexed by conditioned responding at the food trough, in comparison to the third blocked cue. We thus conclude that social value can support reinforcement learning processes involving novel cues. This opens up possibilities to investigate behavioural aspects of the social-value driven reinforcement learning and its associated neural basis, processes that might be disturbed in psychiatric disorders marked by impaired reinforcement learning and/or social behavior such as autism [24] and schizophrenia [25].

Materials and Methods

Subjects 56 male Long Evans rats where housed in pairs of two and kept under an inverted 12:12 h light dark cycle, in a temperature (20 ± 2 °C) and humidity-controlled (approx. 60%) colony room. All rats had ad libitum access to food, except during the testing period. During behavioural testing, the rats where food restricted (20 grams on weekdays and 22 grams in the weekend) and maintained a body weight of about 90% of their free-feeding weight. All testing was performed in accordance with the German Welfare Act and was approved by the local authority LANUV (Landesamt für Natur-, Umwelt und Verbraucherschutz North Rhine-Westphalia, Germany, AZ 84-02.04.2016.A522)

Apparatus. Testing was conducted in 4 customised PhenoTyper (Noldus Information Technology) behavioural testing boxes (Fig. 1A) of 45 by 45 by 55 cm, supplemented with operant devices (Med Associates) and placed inside a custom-made sound- and lightproof ventilated box. The boxes where modified by adding a custom-made Plexiglas separation wall (Fig. 1A, left panel), which divided the box into two compartments, to allow the training of a pair of rats at the same time. The separation wall was equipped with a sliding door (dimensions: 20 by 20 cm, located at 7 cm from the left side of the Skinner box) and 4 rectangular interaction windows (Fig. 1A, left panel; size: 10 by 1.5 cm) that were positioned exactly in between the door and the wall holding the stimulation devices used for conditioning.

Both compartments of the box contained a food trough (Med Associates, ENV-254-CB) positioned in the middle on the right side. The food troughs were adapted in such a way that the detection photobeams were positioned at the entry point of the food trough. The food trough was connected to an automated pellet dispenser (PTPD-0010, Noldus Information Technology) that delivered sucrose pellets (20 mg dustless precision pellets, Bio-Serv, Germany). Operant devices were positioned on the right side of the box at the level of the separation wall: an LED Stimulus Light (Med Associates, ENV-211m) with green cover was positioned 10 cm above the ground and a house light (Med Associates, ENV-215m) 28 cm above the ground. A speaker (Med Associates, ENV-224am) was positioned at 20 cm above the ground for the playback of auditory stimuli (Fig. 1A, right panel). Auditory stimuli were played back at a loudness of 75 dB measured with a hand-held analyser (type 2250-S from Brüel and Kjaer) right in front of the speaker. In the top cover of the Skinner boxes, a camera (Basler, acA1300-60gc, GigE) was positioned to obtain videos of the behavioural experiment at 25 fps. Analyses of the recorded videos was performed with EthoVision XT 11.5 (Noldus Information Technology). Finally, a USV-microphone was positioned next to the camera for recording ultrasonic vocalisations using Ultra Vox XT (Noldus Information Technology).

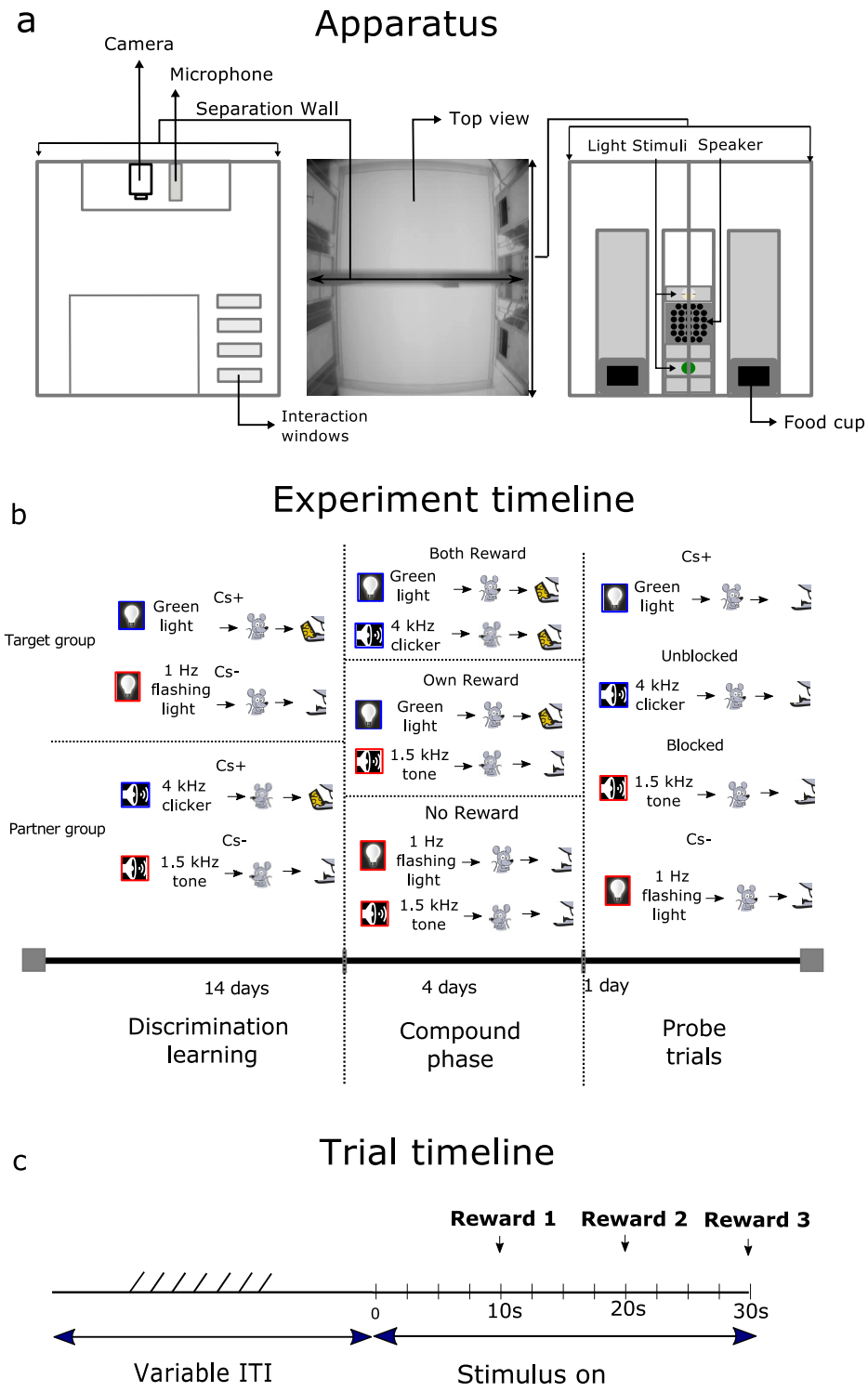

Figure 1: Behavioural apparatus, Experimental timeline and Trial timeline.

(A) The PhenoTyper consisting of lower and upper compartment in which behavioural training took place is displayed in the middle. On the left the custom made separation wall is shown with interaction windows, camera and microphone. On the right, the right side of the PhenoTyper is displayed with the used operant devices and in both sides of the box the food cup. (B) An example experimental time line is displayed. Target rats learn to discriminate two visual cues in upper compartment while at a different time partner rats learned to discriminate two auditory cues in the lower compartment. In the compound phase target and partner rat are either both rewarded (Both Reward), target rat is rewarded while the partner is not rewarded (Own Reward) or both target and partner rat are not rewarded (No Reward). In the probe trials all learned cues are presented to the target rat without reward. (C) Here, a timeline is shown with the different components that make up a single trial throughout the discrimination learning, compound phase and probe trials.

Behavioural Training

Pavlovian discrimination task. Before the start of behavioural training, rats were put on food restriction to reduce their weight to 90% of their free-feeding weight. Within a pair of cage mates, one rat was assigned at random as the actor animal, and the other as the partner animals. As a first step, they were habituated to their pre-determined training side of the customised PhenoTyper for 3 days (15 minutes per day). During this period, they could retrieve 6 pellets that were put along the edges of their respective side of the box. Subsequently, the discrimination learning phase started. Here, the pairs of rat cage mates were divided into two groups; one group of rat pairs would learn a visual discrimination problem, and the other an auditory discrimination problem. The visual stimuli to be discriminated consisted of a houselights flashing at 1 Hz (0.1sec on, 0.9 sec off) and a steady green light; the auditory stimuli were made up by a 4.0 kHz clicker (0.1sec on, 0.9 sec off) and a 1.5 kHz (75 dB) steady tone. We counterbalanced the CS+ and CS- condition assignment to each stimulus across rat pairs. The different groups (Auditory vs Visual) were each trained alone either in the upper or lower compartment of the Skinner box, and the side assignments between actor and partner rats were counterbalanced between experiments (Fig. 1A). Each rat received 14 days of discrimination training. One daily session consisted of 40 trials, of which 20 trials were CS+ and 20 CS-. The order of CS+ and CS- trials was pseudo-randomized, with no more than 3 trials of one kind occurring in a row. Stimuli were presented for 30s and at every 10s (+ 0.1 to 0.4 sec jitter), a pellet was delivered (Fig. 1C). For N=24 out of the 40 animals in the experimental group and for all N=16 rats in the control experiment, a second pellet dispenser was placed outside of the behavioural box, on the opposite side of where the current rat was trained, delivering pellets outside of the box. This was done in order to have a uniform sound associated with the pellet delivery throughout the experiment avoiding any difference in pellet delivery related sounds influencing the conditioning to added cues. With the remaining 16 rats of the experimental group the second pellet dispenser was not placed outside of the behavioural box, on the opposite side of were the current rat was trained (see comparison of main effects below). The ITI in both groups was made up of a fixed 30s window supplemented with a randomized time window ranging from 5 to 100 seconds with steps of 5 ms, uniformly distributed. The ITIs were thus fully randomized, resulting in a total variable ITI with a mean of 80 ms. Ultrasonic vocalisations where recorded from 10s before cue onset to 20s after cue offset, for a total duration of 60s per trial. After completion of the discrimination phase, rats progressed to the compound conditioning stage.

Compound conditioning. After discrimination training was completed, rats in the visual discrimination group received 1 day of pre-exposure to the two novel auditory stimuli while the rats in the auditory discrimination group received 1 day of pre-exposure to the two novel visual stimuli. The pre-exposure session consisted of one session with 6 trials. The stimuli were presented in a randomized order with ITIs of 15, 30 45, 60, 75 and 90s. Pre-exposure was done to reduce possible novelty-induced akinesia [26,27] for each group. Cues were presented for 30s in the same manner as described for the discrimination period, only without reward. After pre-exposure, all rats received 4 days of compound

training. During compound conditioning, both rats were trained together in the same Skinner box, each in their own compartment. Three different conditions were implemented; Both Reward (BR), Own Reward (OR) and No Reward (NR, Fig. 1B, middle panel). Rats from one Target group were rewarded on BR and OR trials, but not on NR trials, while rats from the Partner group were only reward on BR trials, but not on OR and NR trials. During BR trials, both the (respective visual or auditory) CS+ of the Target group and the Partner group were simultaneously displayed and both rats were rewarded with 3 pellets. In the OR trials, the respective CS+ of Target group was simultaneously displayed with the CS- of the Partner group and only the Target group was rewarded. In the NR, both groups were presented their respective CS- and none of the rats were rewarded. Assignment of visual-discrimination-trained and auditory-discrimination-trained rats to either Actor or Partner groups was counterbalanced across experiments. A compound conditioning session consisted of 20 trials per condition. The Conditions BR, OR and NR were pseudo-randomized with every condition not being repeated more than 3 times in a row. ITI randomization, stimulus presentation and reward delivery were implemented as in the discrimination phase.

Probe trials. During probe trials, all rats were tested in isolation for one extinction session in their assigned box compartment. All stimuli were now presented in isolation, both the CS+ and CS- learned in the Pavlovian discrimination task as well as the two novel stimuli added in the compound phase, for which learning was hypothesised to become unblocked and blocked, respectively. Rats in both groups went through 14 trials for each of these 4 stimuli, presented in isolation and without reward delivery (Fig. 1B, right panel). The 4 stimuli were pseudo-randomized with every condition not being repeated more than 3 times in a row.

Control experiment. In the control experiment, 8 Target rats and 8 Partner rats went through the same three experimental conditions. The only difference here is that during the compound phase, the wall that separated the Skinner box compartments was rendered opaque by adding an additional black wall, to block contact between the Target and Partner rats. We hypothesised that if visual, and/or auditory and/or olfactory contact between the rats facilitated the social information transmission that would unblock reinforcement learning of compound cues, then obstructing these transmission cues should impair unblocking.

Statistical data analyses. Entries into the food trough were recorded as photobeam breaks. Raw data were processed in EthoVision XT 11.5 (Noldus Information Technology) to extract our dependent variables: time spent in the food trough and number of entries in the trough (food cup rate). Food cup directed behavior in the form of time spent in the food trough, food cup rate and latency to entry were analysed per trial and per condition for all stages of learning; further analysis and graph preparation was performed using custom-made scripts in MATLAB (version 2014b, MathWorks). Both the percentage of time spent at the food trough and rate of food trough responding where used (Burke et al., 2008; McDannald et al., 2011) as measures of conditioned responding. To assess the strength of learning during discrimination and compound conditioning and probe trials, only the first 10s of the

cue period was analysed to avoid the influence of reward delivery/omission feedback [23]. Discrimination learning performance was quantified by averaging responding to the cues over the last 4 days of training and comparing the mean between CS+ and CS- using a paired sample t-test. Performance in the compound phase was quantified using a 2 factor repeated measures ANOVA on the mean response rate per day across conditions (BR, OR and NR) and post hoc tests were performed to assess the significance of any differences between conditions, corrected for multiple comparisons. Simple contrasts were calculated to further compare the difference between conditions on the different days against the last day of compound conditioning. Performance in the probe trials was assessed by averaging responding over 5 bins (2 trials per bin) and running a two factor repeated measures ANOVA over these bins and the 4 stimuli types (CS+, unblocked, blocked and CS-). Differences between conditions and bins were assessed with post-hoc tests, again corrected for multiple comparisons. Here, simple contrasts were calculated, comparing the difference between conditions on the different bins against the last bin of extinction. Finally, also a non-parametric Friedman two-way ANOVA test on the differences among the latency to entries including all trials (excluding all non-entered trials) between the conditions in the probe trials were performed. Differences between conditions were assessed with post-hoc tests, again corrected for multiple comparisons.

Results

Experimental group

Discrimination learning. Target rats were trained on a visual or auditory discrimination task. All rats developed a conditioned response to the CS+, resulting in an increase in time spent in the food trough on CS+ trials, independent of cue modality or identity. Concurrently, they learned to expect no reward during CS- presentations, as witnessed by a steady decrease in time spent in the food trough on CS- trials (Fig. 2A). A paired samples *t*-test examining the mean responding over the last 4 days of conditioning was performed. We found a significant difference in time spent in the food trough between the CS+ (M = 58.76, SD=12,86) and CS- (M=21.19, SD= 13,21; $t(19) = 12.109$, $p < 0.001$ (Fig. 2A). Next to that, the food cup rate was significantly different between the CS+ (M = 3.07, SD = 1.70) and CS- (M = 0.90, SD = 0.55); $t\ (19) = 6.03$, $p < 0.001$ (Fig. 2D). Finally, rats also developed a faster mean latency to entry for the CS+ and a slower entry for the CS- over the course of behavioural training (Supplemental Fig. 1A). These results indicate that our rats discriminated between auditory cues that differed in tone frequency (1.5 kHz vs. 4 kHz) and temporal presentation (steady tone vs 1Hz clicker (0.1 ms per 1s on). The same held for the visual cues that differed in colour (green vs white) and temporal presentation pattern (steady vs 1 Hz flicker (0.1 ms per 1s on).

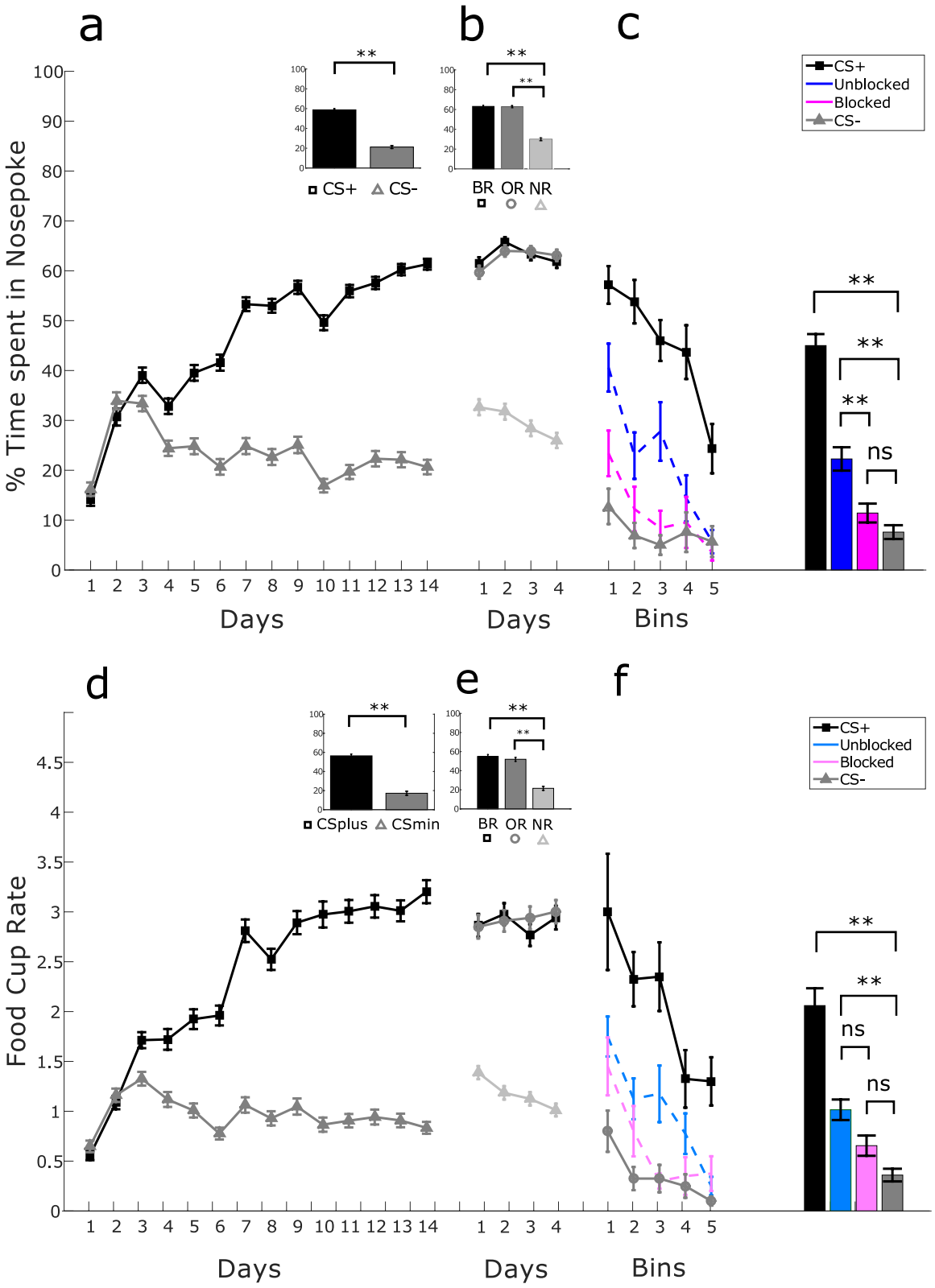

Figure 2. Conditioning per experimental phase.

(A) Percentage of time spent in food cup for discrimination learning between CS+ and CS- over days. (B) Percentage of time spent in food cup for the compounds BR, OR and NR over days. (C) Percentage of time spent in the food cup during the probe trials over 10 trials. Averaged time spent in the food cup over 10 probe trials (D) Food cup rate for discrimination learning between CS+ and CS- over days. (E) Food cup rate for the compounds BR, OR and NR over days. (F) Food cup rate during the probe trials over 10 trials. Averaged Food cup rate over 10 probe trials. Error bars indicate SEM.

Social learning. We aimed to show the influence of adding cues onto learned discriminated cues (forming a compound) with the following social outcomes: Both Reward (BR), Own Reward (OR) and No Reward (NR; Fig. 1B). Rats Pavlovian conditioned responses to these compound cues are shown and a direct comparison of these responses to the original CS+ and CS- cues is made (Fig. 1B), both indexed by time spent in the food cup and the food cup rate. In the subsequent analysis, only the behavior of the Target rats is reported. A 2-factor repeated measures ANOVA with condition and day as factors and the time spent in food trough as dependent variable was performed. We found a significant main effect of condition type ($F_{(2, 38)}$ = 417,04 $p < 0.01$). This effect was explained by a greater responding on BR and OR trials compared to NR trials (both $p < 0.01$, Bonferroni-corrected; Fig. 2B) and no difference between responding on the BR compared to the OR condition ($p = 1.00$). Furthermore, a paired samples *t*-test revealed that responding for BR (M = 61.8, SD = 7.08) did not significantly differ from the average response rate over the last 4 days of discrimination learning for the CS+ (M = 58.76, SD = 12.85,21); $t(19) = 0.907$, $p = 0.376$). A similar patterns was found for OR trials (M = 63.1, SD = 6.93) in comparison to CS+ trials (M = 58.76, SD = 12.85,21); $t(19)=1.334$, $p = 0.198$). We found no main effect of day ($F_{2,38}$ = 2.248, $p = 0.092$) but did find a significant interaction between condition and day ($F_{(6,114)}$ = 2.795, $p=0.014$). Simple contrasts revealed that the difference between condition BR and NR ($F_{(1,19)}$ = 7,238, $p = 0.014$) and the difference between conditions OR and NR ($F_{(1,19)}$ = 6.898, $p = 0.017$) was significantly higher for day 4 compared to day 1 (Fig. 2B). The increase in differences between BR and NR trials and between OR and NR trials, when contrasting day 1 and day 4, is furthermore driven by a decrease of responding on the NR trials over days, as witnessed by a significant decrease when directly comparing day 1 (M = 32.63, SD= 8.47) to day 4 (M = 25.97, SD= 7.68); $t(19) = 2.898$, $p = 0.009$ (Fig. 2B, light grey line). A paired samples *t*-test revealed that responding on NR trials was significantly higher on day 1 (M = 32.64, SD = 8.47) than responding, averaged over the last 4 days of discrimination learning, for CS- trials (M = 21.18, SD = 13.21); $t(19)= 3.218$, $p = 0.005$. Responding on NR day 4 however (M = 25.97, SD = 7,68;) did not differ anymore from responding to CS- averaged over the last 4 days of discrimination learning; $t(19)= 1.523$, $p = 0.114$. Additionally, we observed a faster mean latency to entry on BR and OR trials compared to NR trials throughout the compound phase (Supplemental Fig. 2B). We conclude from these results that adding a second cue, which predicts either a reward (BR) or no reward (OR) to the partner, to the target rats CS+ did not change the target rat responding on BR and OR in comparison to its originally learned CS+ response at the end of discrimination. However, adding a second cue, which predicts no reward (NR) to the partner, on a learned CS- response did change the target rats responding on the NR condition. Here, a transient increased responding was found that over 4 days decreased back to the CS- responding found at the end of discrimination learning. Most importantly, both BR and OR trials, including the CS+ as part of the compound, still elicited more responses than NR (CS- compound) trials.

A 2-factor repeated measures ANOVA with condition and day as factors was also performed for the food cup rate. We found a main effect of condition type on food cup rate ($F_{(2, 38)}$ = 685.349, $p < 0.01$).

This effect was explained by a higher food cup rate on BR and OR trials compared to NR trials (both $p < 0.01$; Fig. 2E). Furthermore, a paired samples *t*-test revealed that food cup rate did not differ for BR trials (M = 2.68, SD = 0.28) in comparison to the average of the last 4 days of responding for the CS+ (M = 3.07, SD = 1.63); $t(19) = -0.571$, $p = 0.575$), or for OR trials (M = 2.85, SD = 0.54); $t(19) = -0.554$, $p = 0.586$). We did not find an effect of days on food cup rate ($F_{(3, 57)} = 0.640$, $p = 0.592$) but did find an interaction between day and condition ($F_{(3, 57)} = 2.960$, $p=0.010$). Simple contrasts revealed that the difference between condition BR and NR ($F_{(1,19)} = 8.266$, $p = 0.010$) and the difference between conditions OR and NR ($F_{(1,19)} = 5.565$, $p = 0.014$) was, here, also significantly higher for day 4 compared to day 1. However there was also a significant contrast indicating that the difference between BR and NR was larger on day 4 then on day 3 ($F_{(1,19)} = 1.653$, $p = 0.031$; Fig. 2B). A paired samples *t*-test furthermore revealed that responding to NR on day 1 (M = 1.38, SD = 0.34) was significantly higher than responding on the last 4 days of CS- (M = 0.89, SD = 0.43); $t(19) = 4.349$, $p = 0.000$). Again, like for the time spent in the food trough, responding on NR day 4 however (M = 1.01, SD = 0.32); $t(19) = 881$, $p = 0.389$) was not significantly higher than for CS- trials averaged over the last 4 days of discrimination training. Here as well, the increase in difference between the BR and NR and difference between OR and NR when contrasting day 1 and day 4 is driven by a further decrease of the value of the NR over days leading to a significant difference when directly comparing day 1(M = 1.38, SD= 0.34) to 4 (M = 1.01, SD= 0.31); $t(19)= 3.279$ $p = 0.001$ (Fig. 2E, light grey line). We conclude from these results that adding a second cue, which predicts either a reward (BR) or no reward (OR) to the partner, to the target rats CS+ did also not change the target rat food cup rate on BR and OR in comparison to its originally learned CS+ response at the end of discrimination. Furthermore, adding a second cue, which predicts no reward (NR) to the partner, on a learned CS- response changed the target rats responding on the NR condition. Here, a similar increased food cup rate was found that over 4 days decreased back to the CS- responding found at the end of discrimination learning. Most importantly, both BR and OR trials, including the CS+ as part of the compound, still elicited a higher food cup rate than NR (CS- compound) trials.

Probing associative value. In the probe trials, we aimed to show the effect of social reinforcement learning of compound cues with different social outcomes by presenting the cues in isolation, omitting reward, as indexed by the time spent in the food cup, the food cup rate and the latency to entry. We show the conditioned responding of the target rats to 14 presentations each of the CS+ and CS-, the unblocked cue associated with an added reward to the partner (BR) and the blocked cue associated with a no reward to the partner (OR) and no reward to self (NR). A two factor repeated measures ANOVA with stimulus type and bin as factors and the time spent in the food trough as the dependent variable was performed. We found a significant main effect of probe trial condition on time spent in food trough ($F_{(3, 57)} = 83.287$, $p < 0.01$).

As expected, time spent in food trough was higher for CS+ than CS- trials ($p < 0.01$; Fig. 2C). Critically, pairwise comparisons revealed that time spent in food trough was higher in unblocked cues than blocked cues across the 5 bins of extinction ($p = 0.003$; Fig. 2C). Furthermore, we found a significantly

higher responding to the unblocked cue compared to the CS- cue ($p < 0.01$) while responding to the blocked cue was not higher than the CS- cue ($p = 0.362$; Fig. 2C). Additionally, we found a main effect of bin number on time spent in food trough ($F_{(4, 76)} = 18.809$, $p < 0.01$), reflecting the extinction process, and finally, we found an interaction between condition type and bin number on the time spent in food trough ($F_{(12,228)} = 2.930$, $p = 0.001$). Simple contrasts revealed that the difference between unblocked and blocked was significantly higher in bin 1 ($F_{(1,19)} = 4.94$, $p = 0.039$) and bin 3 ($F_{(1,19)} = 7.127$, $p = 0.015$), and trending for bin 2 ($F_{(1,19)} = 3.365$, $p = 0.082$) in comparison to that difference observed in bin 5. Simple contrasts furthermore revealed that the difference in responding for the Unblocked and CS- cues was significantly higher in bin 1 ($F_{(1,19)} = 42.316$, $p = 0.000$), bin 2 ($F_{(1,19)} = 10.857$, $p = 0.004$) and bin 3 ($F_{(1,19)} = 15.530$, $p = 0.001$) than the difference observed in bin 5. The difference between the blocked cue and the CS- on the other hand was significantly higher only in bin 1 compared to bin 5 ($F_{(1,19)} = 4.856$, $p = 0.040$). We furthermore found a significant main effect of probe trial condition type on food cup rate ($F_{(1.676, 31.853)} = 24.940$, $p < 0.01$). Pairwise comparisons revealed that the food cup rate was descriptively higher across the 5 bins of extinction in response to unblocked cues than to blocked cues ($p = 0.084$) and significantly higher than to CS- cues ($p< 0.01$, Fig. 2F). In contrast, responding to the blocked cue did not differ from responding to the CS- ($p = 0.190$). We furthermore found a main effect of bin number on food cup rate ($F_{(4, 76)} = 18.629$, $p < 0.01$) as well, in line with extinction processes and did not find a interaction between condition type and bin number on the food cup rate ($F_{(3.953,75.116)} = 1.630$, $p = 0.176$). A non-parametric Friedman two-way ANOVA test on the differences in the latency to entry for all trials (excluding all non-entered trials) between the conditions CS+ (median = 3.32), unblocked (median = 3.56), blocked (median = 6.8) and CS- (median = 9.58) during extinction was performed. We found a significant effect of probe trial condition on the latency to entry ($X^2_{(3, N = 62)} = 23.733$, $p < 0.001$). Post hoc comparisons revealed that latencies to entry were not significantly shorter for unblocked compared to blocked ($p = 0.309$) and CS- ($p = 0.203$) or when comparing blocked and CS- ($p = 1.000$). CS+ latencies were, however, significantly shorter than both blocked ($p < 0.001$) and CS- ($p < 0.001$) but not the unblocked cue ($p = 0.203$). Considering that the blocked cue was associated with a significantly slower reaction time than the CS+, but not the unblocked cue, the unblocked cue did show shorter median latencies to entry and a more concentrated peak in its density function compared to the blocked cue. These results suggest that the latency to entry for the unblocked cue is more closely related to the CS+ than the blocked cue (Supplemental Fig. 2C).

Taken together, these results show that the rats showed more food cup directed behavior for the unblocked cue than both the CS- and blocked cue over 10 trials of extinction. This means that when actor rats have fully learned a stimulus-reward association producing reward for themselves, adding a cue that predicted an additional reward delivery to a partner rat unblocked associative social value learning. We observe here that the unblocking of the novel cue last for approximately 6 trials. In contrast, rats did not spent more time in the food cup for the blocked cue than the CS-, suggesting that additional cues that did not predict additional reward remained blocked from acquiring associative social reward value.

Control group

Discrimination learning. During discrimination learning, the experimental conditions were identical to the Experimental group. We replicated the discrimination learning results from the experimental group showing that rats again developed a conditioned response to the CS+, resulting in an increase in time spent in the food trough on CS+ trials but not to the CS- as witnessed by a steady decrease in time spent in the food trough on CS- trials (Fig. 3A). A paired sampled t-test revealed a significant difference in time spent in food trough between the CS+ (M = 56.40, SD =12.89) and CS-(M =17.20, SD = 10,18); $t(7) = 11.54$, $p < 0.001$ (Fig. 3A). We also found a significant difference for the food cup rate between the CS+ (M = 2.68, SD=1.39) and CS-(M = 0.66, SD = 0.38); $t(7) = 5.02$, $p < 0.01$ (Fig. 3D) looking at the last 4 days of conditioning. Finally, control rats also developed a shorter latency to entry for the CS+ and a larger latency to entry for the CS- over training (Supplemental Fig. 2D).

Social learning. In the compound phase, rats were again trained side by side on the compound cues associated with the social outcomes Both Reward (BR), Own Reward (OR) and No Reward (NR; Fig. 1B). This time, however, we aimed to show the influence of blocking visual and auditory information exchange on Pavlovian conditioned responses to the compound cues and to a direct comparison of these responses to the target rats original CS+ and CS- (Fig. 1B). A 2-factor repeated measures ANOVA with condition and day as factors and the time spent in food trough as dependent variable was performed. Mauchly's test of sphericity was significant for both the main effect of the factor condition and days. Therefore, for these effects the Greenhouse-Geisser corrected values are reported. As under the experimental conditions, we found a significant effect of condition type on time spent in food trough ($F_{(1.175, 8.228)} = 46.407$, $p < 0.01$). This effect was again explained by a greater responding on BR and OR trials compared to NR trials ($p = 0.01$, Fig. 3B). Furthermore, a paired samples *t*-test revealed that responding for BR trials did not differ on day 1 (M = 58.45, SD = 9.40) in comparison to the average of the last 4 days of responding for the CS+ (M=56.40, SD= 12.89); $t(7)=0.398$, $p = 0.702$) and when comparing OR (M = 50.38, SD= 15.84) with the CS+ (M=56.40, SD= 12.89); $t(7)=-1.779$, $p = 0.118$). We found no effect of days on time spent in the food trough ($F_{(1.294, 9.059)} = 1.501$, p = 0,262) but did find an interaction between day and condition ($F_{(6, 42)} = 2.873$, p= 0.019). Contrasts revealed that the difference between condition BR and NR ($F_{(1,7)} = 13.043$, $p = 0.009$) and the difference between conditions OR and NR ($F_{(1,7)} = 7.00$, $p = 0.033$) was significantly higher for day 4 compared to day 2 but found no difference between day 1 and 4.

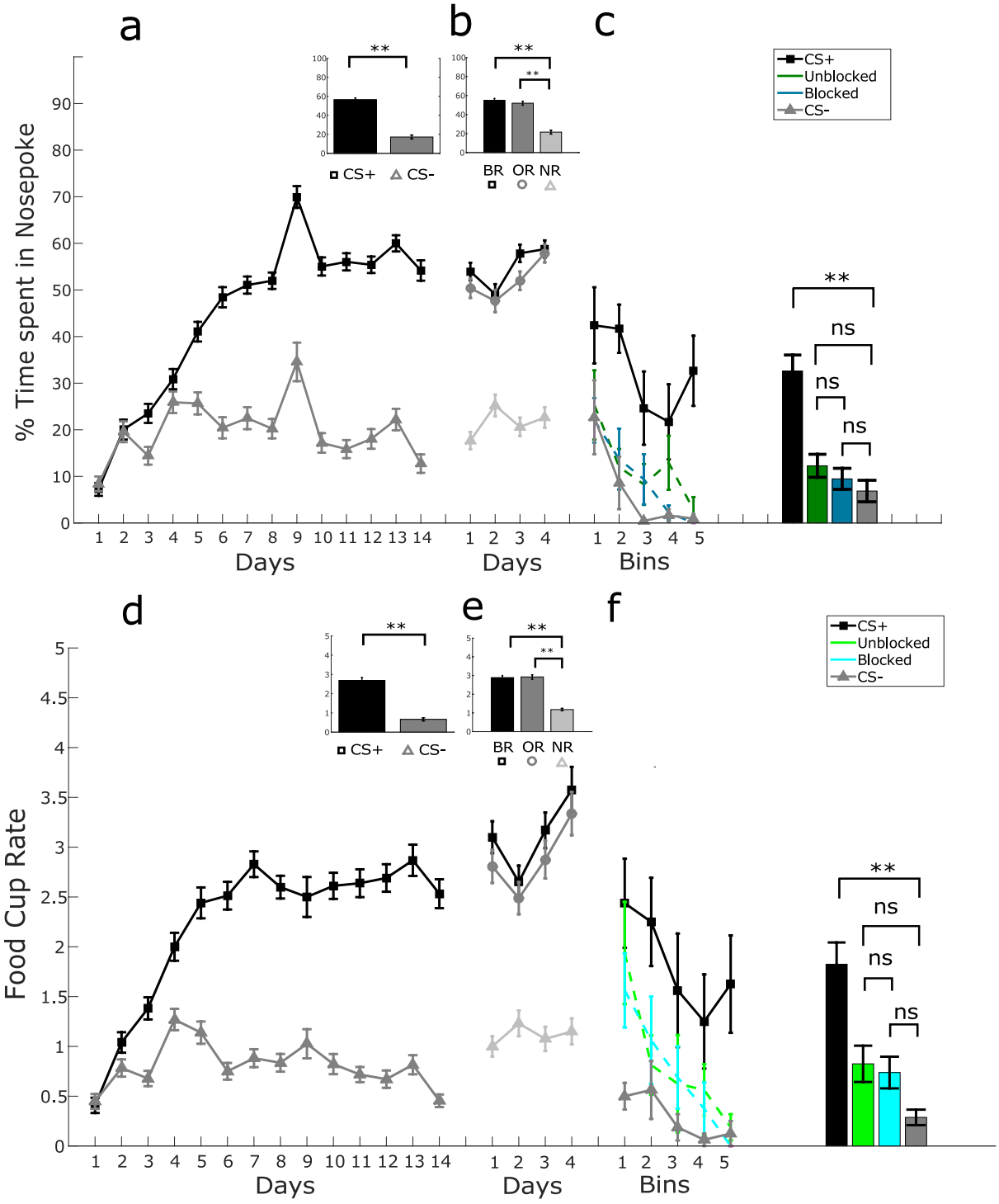

Figure 3. Control group conditioning per experimental phase.

(A) Percentage of time spent in food cup for discrimination learning between CS+ and CS- over days. (B) Percentage of time spent in food cup for the compounds BR, OR and NR over days. (C) Percentage of time spent in the food cup during the probe trials over 10 trials. Averaged Time spent in the food cup over 10 probe trials. (D) Food cup rate for discrimination learning between CS+ and CS- over days. (E) Food cup rate for the compounds BR, OR and NR over days. (F) Food cup rate during the probe trials over 10 trials. Averaged food cup rate over 10 probe trials. Error bars indicate SEM.

The increase in difference between the BR and NR and difference between OR and NR when contrasting day 2 and day 4 was not driven by a decrease in responding to the NR from day 1 (M = 17.68, SD = 7.56) to day 4 (M = 22.63, SD = 10.59; $t(7)$= -1.041, p = 0.333) as was the case in the experimental group results.

Paired samples *t*-test revealed furthermore that the time spent in the food cup for NR on day 1 (M = 17.68, SD = 7.56) was not significantly higher than responding on the last 4 days of CS- (M = 17.20, SD = 10.17); $t(7)$ = 0.137, p = 0.895) and NR day 4 (M = 25.97, SD = 7,68;); $t(7)$= 1.523, p = 0.114) was not significantly higher. Additionally, we again observed a clear shorter mean latency to entry in the BR and OR trials compared to NR trials throughout the compound phase (Supplemental Fig. 2B). These results indicate that adding a second cue that predicts either reward or no reward to the partner when social information exchange is impeded does not influence time spent in the nose poke for a self-reward. Both BR and OR trials, including the CS+ as part of the compound, still elicited a higher time spent in the food cup than NR (CS- compound) trials. However when social information exchange was blocked, we found that observing a partner not being rewarded did not result in an increase in the time spent in the food cup for the NR on day 1. NR value is comparable to CS- value throughout compound conditioning as indexed by the time spent in the food cup and thus no further modulation through learning processes became apparent.

A 2-factor repeated measures ANOVA with condition and day as factors with the food cup rate as dependent variable was performed. We again found a significant effect of condition type on food cup rate ($F_{(1.019, 7.134)}$ = 21.781, p = 0.002). This effect was also explained by a greater responding on BR and OR trials compared to NR trials ($p < 0.01$, Fig. 3E). Furthermore, a paired samples t-test revealed that food cup rate was similar for BR (M = 3.1, SD= 1.41) in comparison to the average of the last 4 days of responding for the CS+ (M = 2.68, SD= 1.39); $t(7)$= 1.302, p = 0.234) and when comparing OR (M = 2.80, SD = 1.11) with the CS+ (M = 2.68, SD= 1.39); $t(7$ = 0.452, p = 0.665). Next to this, we found no effect of days on food cup rate (F (1.31, 9.168) = 1.024, p = 0,402) and no interaction between day and condition (F (1.577, 11.042) = 1.618, p= 0.166). These results indicate that adding a second cue that predicts either reward or no reward to the partner when social information exchange is impeded does not influence food cup rate for a self-reward. Both BR and OR trials, including the CS+ as part of the compound, still elicited a higher food cup rate then NR (CS- compound) trials. NR value as indexed by food cup rate does not decrease throughout compound conditioning phase and thus no further learning is apparent.

Probing associative value. In the control experiment probe trials, we aimed to show the effect of impeding social information exchange on social reinforcement learning of compound cues with different social outcomes. We again show the conditioned responding of the target rats to 14 presentations each of the CS+ and CS-, the unblocked cue associated with an added reward to the partner (BR) and the blocked cue associated with a no reward to the partner (OR) and no reward to self (NR). A two factor repeated measures with condition and bin as factors and the time spent in the food cup as dependent variable was performed. We found a significant main effect of probe trial condition type on time spent in food trough ($F_{(3, 21)} = 15.327$, $p < 0.01$). We furthermore find a main effect of bin number on time spent in food trough ($F_{(1.540, 10.781)} = 18.809$, $p < 0.01$) Finally, we found no interaction between condition type and bin number on the time spent in food trough ($F_{(12,84)} = 1.225$, $p = 0.280$). Pairwise comparisons revealed that time spent in food trough was higher in CS+ cues than CS- cues across the 5 bins of extinction ($p = 0.021$, Fig. 3C). We furthermore found that the time spent in the food trough was not different in response to unblocked cues compared to blocked cues across the 5 bins of extinction ($p = 1.000$, Fig. 3C). Crucially, in this control experiment, time spent in the food trough for both the blocked cue and the unblocked cue was not different from the CS- ($p= 0.622$; $p=1.00$). We further performed a two factor repeated measures with condition and bin as factors and the food cup rate as dependent variable. We found a significant main effect of probe trial condition type on food cup rate ($F_{(1.292, 9.042)} = 10.979$, $p < 0.01$). We furthermore find a main effect of bin number on time spent in food trough ($F_{(2.186, 15.304)} = 10.335$, $p < 0.001$). Finally, we found no interaction between condition type and bin number on the time spent in food trough ($F_{(4.476 ,31.335)} = 1.229$, $p = 0.235$). Pairwise comparisons revealed that the food cup rate was marginally higher in CS+ cues compared to CS- cues across the 5 bins of extinction ($p = 0.057$, Fig. 3C). We furthermore found that the food cup rate was not different in unblocked cues in comparison to blocked cues across the 5 bins of extinction ($p = 1.000$, Fig. 3C). A non-parametric Friedman two-way ANOVA test on the differences in the latencies to entry over all trials (excluding all non-entered trials) between the conditions CS+ (median = 3.64), unblocked (median = 5.70), blocked (median = 5.86) and CS- (median = 11.30) during extinction was performed for the control experiment as well. We did not find a significant effect of probe trial condition on the latency to entry ($X^2_{(3, N = 26)} = 7.24$, $p = 0.064$). Also descriptively the latencies to entry distributions for unblocked and blocked cue probe trials look similar while at the same time the median latency to entry is still faster for CS+ compared to the CS- (Supplemental Fig. 2C).

These results show that, contrary to the findings for the experimental group, the rats in the control experiment did not show more food cup directed behavior for the unblocked cue than both the CS- and blocked cue over 10 trials of extinction. This means that when actor rats have fully learned a stimulus-reward association producing reward for themselves, adding a cue that predicted an additional reward delivery to a partner rat does not unblock associative social value learning when social information exchange is blocked. These rats also did not spent more time in the food cup for the blocked cue than the CS-. This shows that additional cues that did not predict additional reward remained blocked from acquiring associative social reward value when information exchange is

blocked. Overall, these results are consistent with the idea that the exchange of social information is necessary for the social unblocking effect observed in the experimental group.

Control for secondary reinforcement. Besides a contribution of social value, i.e. the vicarious experience of reward, other confounding factors could drive learning/unblocking in our paradigm. Most notably, sources of secondary reinforcement should be excluded as drivers of learning. During discrimination learning, the actor rat is conditioned to receive pellets contingent on its CS+. Afterwards, in the compound phase, the rat is presented with an auditory –visual compound. Instead of one pellet drop (self-reward), now, on some trials, two pellets drop simultaneously (Both Reward trials). It is possible that the additional pellet delivery sound acted as a third CS+ in the compound, delivering occasional pellets to the partner rat. Because the sound of the pellet dispenser is associated with the CS+ of the target rat, it is possible that the appetitive value associated with this cue enhanced the value of the added novel cue leading to its unblocking. To control for this possible source of secondary reinforcement, we added a non-rewarding pellet dispenser aimed outside the box during the discrimination phase. Here, we compare results between these conditions. In the first experimental group (N=8), rats were trained on the discrimination paradigm without added pellet dispensers and therefore secondary reinforcement of the added pellet dispenser during compound phase is likely. A two factor repeated measures with condition (unblocked and blocked) and trial (trial 1 to 6) as factors and time spend in the food cup as dependent variable was performed on the experiment where only one active pellet dispensers was used in the discrimination phase. For the subsequent probe phase, we found a significant main effect of probe trial condition type on time spent in food trough ($F_{(1, 7)}$ = 26.026, $p < 0.01$). We furthermore found no main effect of trial number on time spent in food trough ($F_{(5, 35)}$ = 0.900, $p = 0.492$). We did not find an interaction between condition type and bin number on the time spent in food trough ($F_{(5, 35)}$ = 1.877, $p = 0.124$). Pairwise comparisons revealed that time spent in food trough was higher for unblocked cues than blocked cues across the 6 trials of extinction ($p = 0.001$; Supplemental Fig. 2A).

In the second experimental group (N = 12), rats were trained on the discrimination paradigm with added pellet dispensers to reduce the potential influence of sources of secondary reinforcement during compound phase. A two-factor repeated measures ANOVA with condition (unblocked and blocked) and trial (trial 1 to 6) as factors and time spent in the food cup as dependent variable was performed on the experiment where 2 active dispensers were used in the discrimination phase. For the subsequent probe phase, we found a significant main effect of probe trial condition type on time spent in the food trough ($F_{(1, 11)}$ = 5.185, $p = 0.044$). We furthermore found a main effect of trial number on time spent in food trough ($F_{(5, 55)}$ = 4.977, $p = 0.001$). We did not find an interaction between condition type and bin number on the time spent in food trough ($F_{(5, 55)}$ = 0.803, $p = 0.552$). Pairwise comparisons revealed that time spent in the food trough was higher for unblocked cues than blocked cues across the 6 trials of extinction ($p = 0.044$; Supplemental Fig. 2B). In the control group (N = 8), rats were trained on the discrimination paradigm without added pellet dispensers and therefore secondary reinforcement of the added pellet dispenser during compound phase is also likely. A two factor

repeated measures with condition (unblocked and blocked) and trial (trial 1 to 6) as factors and time spend in the food cup as dependent variable was performed on the control experiment where only one active pellet dispensers was used in the discrimination phase. For the subsequent probe phase, we did not find a significant main effect of probe trial condition type on time spent in food trough ($F_{(1, 7)}$ = 0.002, p = 0.969). We furthermore found no main effect of trial number on time spent in food trough ($F_{(5, 35)}$ = 2.390, p = 0.058). We also did not find a interaction between condition type and bin number on the time spent in food trough ($F_{(5, 35)}$ = 43.279, p = 0.976). Pairwise comparisons revealed that time spent in food trough was not higher for unblocked cues than blocked cues across the 6 trials of extinction (p = 0.969; Supplemental Fig. 2C). We can conclude that secondary reinforcement of pellet sounds in the compound phase influenced the strength of the observed unblocking effect, as the effect size of unblocking vs. blocked cues is higher for the condition without added pellet dispensers than for the condition pellet when dispensers were added during the discrimination learning. However, when controlling for this source of potential secondary reinforcement, the social unblocking effect persisted.

Comparison between Experimental and Control conditions. To examine the influence of obstructing the exchange of social information during the compound phase, we applied a similar analysis on the difference scores for the comparison between the experimental group and the control group based on responding during the probe phase. In the probe phase, we calculated both the CS+/CS- contrast and the Unblocked/Blocked contrast. We examined the difference scores of the Unblocked/Blocked contrast with a two way repeated measures ANOVA with group (Experimental, control) and trial (1-6) as factors. We found no significant within subject main effect of trials $F_{(5, 130)}$ = 0.715, p = 0.613) and no interaction effect ($F_{(5,130)}$ = 0.554, p= 0,735. However a significant main (between subject) effect of group on the contrast score for unblocked – blocked ($F_{(1, 26)}$ = 6.823, p = 0.013) revealed that the percentage difference in responding between unblocked and blocked cues was higher for the experimental group then control group (Fig. 4E). We similarly examined the difference scores for the unblocked/CS- contrast with a two way repeated measures ANOVA. We found no significant (within subject) main effect of trial on group ($F_{(5,130)}$ = 0.589, p = 0.708) and no interaction effect ($F_{(5,130)}$ = 1.750, p= 0.128). But here as well, we found a significant main between subject effect of group when comparing the unblocked / CS- contrast score ($F_{(1, 26)}$ = 8.614, p =0.007, Fig. 3F).

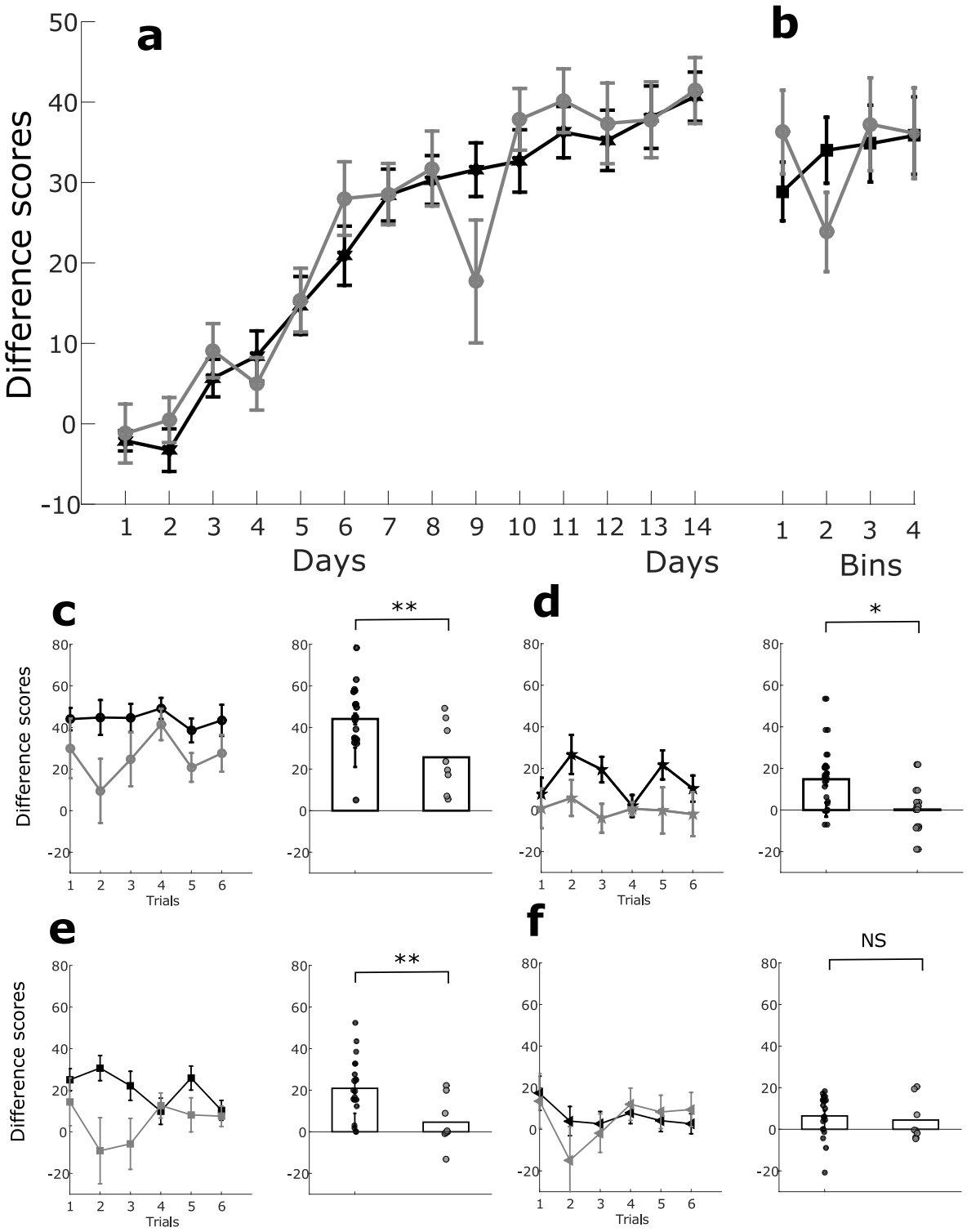

Figure 4. Difference scores for the experimental experimental (black) and control group (grey)

(A) Difference of the percentage of time spent in food cup for CS+ - CS- over days in the discrimination learning. (B) Difference of the percentage of time spent in food cup for BR – NR over days in the compound phase. (C) Difference of the percentage of time spent in food cup responding over 6 trials for the CS+ - CS- difference scores. White bars show average over 6 trials of the CS+ - CS- difference scores with dots showing the SEM per rat. (D) Difference of the percentage of time spent in food cup responding over 6 trials for the CS+ - CS- difference scores. white bars show average over 6 trials of the unblocked – blocked difference scores with dots showing the SEM per rat (e) Difference of the percentage of time spent in food cup responding over 6 trials for the unblocked - CS- difference scores. White bars show average over 6 trials of the CS+ - CS- difference scores with dots showing the SEM per rat (f) Difference of percentage of the time spent in food cup responding over 6 trials for the blocked - CS- difference scores. White bars show average over 6 trials of the CS+ - CS- difference scores with dots showing the SEM per rat. (*$p < 0.05$; **$p < 0.01$)

As expected, when comparing responding for the blocked / CS- contrast, we again found no significant (within subject) main effect of trial on group ($F_{(5,130)} = 1.425$, $p = 0.219$) and no interaction effect ($F_{(5,130)} = 0.597$, $p= 0.702$). Now, in addition, there was no significant main between subject effect of group ($F_{(1, 26)} = 0.221$, $p =0.643$, Fig. 3G).We similarly examined the difference scores for the CS+/CS- contrast with a two way repeated measures ANOVA with variables of group (Experimental, control) and trial (1-6). We found no significant (within subject) main effect of trial ($F_{(5,130)} = 1.133$, $p = 0.346$) and no interaction effect ($F_{(5, 130)} = 0.604$, $p= 0.697$). We did however find an unexpected significant main (between subject) effect of group when comparing the CS+/CS- difference score ($F_{(1, 26)} = 7.862$, $p =0.009$, Fig. 4D). These results further strengthen the conclusion that blocking the exchange of social information during the compound phase selectively affects learning about unblocked vs. blocked cues. Though we found a magnitude difference in the CS+/CS- contrast between experimental and control groups, we note that the CS+/CS- contrast was significant by itself in both groups as well, while this was not the case for the Unblocked/Blocked contrast.

Discussion

Summary Social valuation is crucial in forming and maintaining social relationships and, presumably, in experiencing the pleasurable and reinforcing aspects of social interaction. However, it remained unclear whether social value, which we define here as value attached to social signals associated with reward to another conspecific [cf. 29], could drive learning just as self-experienced value. If this would be the case, then vicariously experienced reward, e.g. social value should be able to reinforce behaviour in a formal Pavlovian learning paradigm. Here we addressed this question by introducing a novel social unblocking task. We find that social value, operationalized in this task as rewards delivered to social partners, can indeed drive learning about novel stimuli. After having fully learned that a specific cue predicted a self-reward, learning about a second cue delivered in compound was blocked, as predicted by learning theory [30,31], when no additional self or other reward was contingent on this cue. Learning was unblocked, however, by providing social reward, e.g. an additional reward delivered to the partner simultaneously with the fully predicted self-reward. Finally, preventing the exchange of social information in the compound learning phase impeded unblocking of the novel stimuli. These results suggest that social value can indeed drive learning processes, in line with formal behavioural learning theory.

Learning Theory Our results extend previous work by Peter Holland [22] and Geoffrey Schoenbaum [23,32] on unblocking in appetitive Pavlovian conditioning. These authors found that rats, after learning that distinct cues have specific food outcomes, can show unblocking of learning for cues added in compound, when self-rewards were altered by increasing reward value (e.g. an upward shift from 1 to 3 pellets) or a change in reward identity (same reward type but a shift in reward features such as flavour). In contrast, learning was blocked when no such reward change occurred (e.g. same reward amount or same identity). According to reinforcement learning theory, the upshift or change in identity led to a discrepancy between the expected reward (1 pellet) and the received reward (3 pellets), thus

producing a reward prediction error. The theory states that if the added cue reliably predicts the increase in reward/ identity outcome, it will acquire the value inherent in the reward itself [33]. The main indicator of learning about the value of a (novel) cue in our social unblocking paradigm is an increase in the time spent in the food cup in the probe (extinction) phase [cf. 23,32]. Indeed, we observed an increase in the time spent in the food cup for the unblocked cue, predicting mutual rewards to the actor rat and its conspecific, relative to the blocked cue, predicting own-rewards only, when presenting the cue in isolation without reward delivery.

From this result, we conclude that the added novel cue has acquired the positive social value inherent in the "Both Reward" delivery through reinforcement learning mechanisms. Analogous to the description above, our result mimics food cup responding that represent the appetitive US-directed 'wanting' response [18]. This indicates that the incentive salience of the unblocked cue can, in addition to being changed by appetitive reward, also be changed by social value. Further research is necessary to see whether socially reinforced cues can also act as a conditioned reinforcer for instrumental responses of rats, as has been found humans [34], in a similar way as has been found for appetitive cues [28,35,36]. Finally, it is important to investigate if social cues can guide rats' social choices. It has been found that rats choose a reward arm in a T-maze that leads to play behavior more than an arm leading to a social encounter where play was absent [37]. Furthermore, social play can induce a social place preference (CPP) [38] and rats are willing to lever press for social play reinforcement [15]. Our task indicates that the unblocked cue has attained the rewarding properties of social reward and it is therefore likely that a presented unblocked cue would be preferred over a blocked cue when tested in a choice task.

Our findings that unblocking can occur through social value as well as Pavlovian appetitive value adds to the emerging literature showing that animals attach value to rewards delivered to conspecifics [7,20] and learn about cues that predict rewards delivered to others. Monkeys, for example, prefer cues that are associated with reward delivery to another monkey over cues that were associated with juice delivery to a chair with no monkey in it and this preference was absent, in the non-social condition, when there was only a juice bottle present [12]. By contrast, not all unequal reward structures lead to positive social value presumably motivating pro-social behavior. A recent study shows that macaques increase licking frequency in line with a higher probability of self-reward but decrease their anticipatory licking with increased probability of reward delivery to another monkey [39]. The authors interpret this decrease of anticipatory licking as an indicator of the negative affect associated with unequal disadvantageous reward pay-outs. Both monkeys and rats have been found to have a distaste for these unequal pay-outs [40,41]. It is thus clear that monkeys can learn to associate cues with both negative and positive outcomes delivered to others. Our results provide evidence that, next to monkeys, rats as well can learn about cues that are associated with reward delivered to others. The social reinforcement learning hypothesis [7] proposes that integration of the social signal expressed by partners cues can aid in making appropriate decision in a social context. Evidence for this hypothesis comes from the prosocial choice task (PCT) in which it was found that, rewards delivered

to oneself and to a partner are preferred over a reward delivery only to the actor himself in both monkeys [4] and rats [5,42]. In rats, it was found that this effect was modulated by the behavior displayed by the other rat and that this effect was impaired when the partner was replaced by a toy or when the display of the partner's preference was impeded. We found that that unblocking occurs when additional reward was delivered to the partner and that preventing the exchange of social information by an opaque wall impeded unblocking. These results also fit the cascade model of representation of cognitive (social) structures that predicts partial contributions that sum up over multiple modalities of social signals (Kalenscher et al, in press). We thus expect that associative learning of the actor rats in our task is driven by a similar social reinforcement learning process, dependent on the transmission of social signals between the actor and partner rat [43]. These signals could manifest in the following aspects; confounding factors, auditory transfer, visual observation and olfactory/tactile social interaction. In the following paragraph, these potential modes of social information transfer will be discussed.

Social Information Transfer

Potential confounding factors It is possible that the additional pellet delivery sound acted as a third CS+ in the compound and influenced the social unblocking effect we observe. We controlled for this by exposing a group of target actors already to the sound of a second pellet delivery apparatus delivering pellets during discrimination learning. Importantly, this dispenser released its pellets outside of the behavioural box on the (empty) partner' side. We found that, relative to a group rats trained without a second pellet dispenser in the discrimination phase, rats that were conditioned in the presence of an added pellet dispenser during discrimination show slightly lower, but still significant levels of unblocking (Supplemental Fig. 2). These results provide evidence that the sound of an active pellet dispenser indeed could entail secondary reinforcement value and can act a CS+ moderator on the observed unblocking effect and should be controlled for.

1. Auditory transfer of affective state/value. Our data show that the possibility for social information exchange was necessary to produce the social unblocking effect. One possibility is that social information was transferred through auditory transfer of affective state/value between rats. It has been found that rats observing conspecifics being rewarded produce 50 kHz vocalisations and show an increase in dopamine release [20] and Playback of 50 kHz vocalisation furthermore motivates rats to approach the location of playback and increase their USV production [10]. Importantly, during playback of 50 kHz calls, dopamine release in the Nucleus Accumbens is transiently increased [17]. Prosocial choices in rats, as found by Hernandez-Lallement in the PCT [5], are therefore likely to be driven by communicative social reinforcement signals [7], such as 50 kHz vocalisations expressed by the partner rat. Recently, it has been found that supressing VTA dopamine transients by means of optogenetics during the compound phase of a non-social blocking task impairs unblocking of novel cue value and features [44]. Furthermore, increasing VTA dopamine activity can drive learning about novel cue-reward associations in an unblocking paradigm [45]. Taken together, these findings suggest

that unblocking in our task could occur through social reinforcement by way of 50 kHz calls-associated dopamine release. Some evidence that 50 kHz can enhance the value of novel cues comes from a study by Saito et al. [46] that found that the playback of 50 kHz ultrasonic vocalisations, 20 min. prior to hearing a neutral tone, increased lever presses of rats during subsequent presentation of the neutral tone more so than in a control condition where there was no playback. The amount of lever presses became more similar to the number of lever presses observed in response to a previously conditioned rewarding stimulus (positive bias), indicating that the playback of 50 kHz USVs induced a positive affective state that transferred onto the neutral stimuli and thereby enhancing its value. In our task, positive reinforcement of the partner can be perceived as both surprising and rewarding to the actor rat and this could therefore increase the partner's and/or target's 50 kHz vocalisations. Future work will indicate whether the production of 50 kHz USVs (unpublished observations) during the BR trials in the compound phase can indeed contribute to dopamine release in the actor rat and thus drive learning i.e. unblocking of the additional compound cue that is associated with partner reward.

2. Visual observational learning. Another possibility is that social information was transferred through visual observational learning. Previous experiments in rats have shown that, during learning, animals can develop both goal tracking towards the location of the US and sign tracking behavior showing orienting and enhanced responding to the location of the CS [47]. This enhanced responding is indicative of an enhanced incentive salience of the cue. We observed that some animals in our task indeed showed coordinated cue-directed behavior during the compound phase. This conforms to the definition of stimulus enhancement in social learning: when observation of an action leads the observer to increase the proportion of its behavior directed towards the location or object of the demonstrator's [48]). Such coordinated putative sign-tracking could therefore facilitate learning in the target rats about the cues inducing cue-directed behavior in the partner rat and presumably increase sign tracking behavior for the actor rat towards the unblocked cue. As an example of such stimulus enhancement, Heyes et al. [49] showed that rats observing other rats performing a lever pressing discrimination task have a higher lever-pressing rate for a previously observed CS+ lever. In our task, responding at the food cup by actor rats might be enhanced through observational conditioning, i.e. by observing the conditioned approach of the partner on BR trials in contrast to OR/NR trials. Our experimental setup was not designed to gauge the influence of observational learning on unblocking because both animals approach the reward site simultaneously during BR conditions. Furthermore, the observed latency to entry was not different between conditions in the compound phase and probe phase. This implies though that a pure matching response is unlikely, but leaves room for a role for coordinated sign tracking during the both reward cue and subsequent observation of the partner's approach (on BR trials) or lack thereof (on OR/NR trials). Trial by trial computational modelling of the influence of the partner rat's sign tracking behavior and approach behavior on the actor rat's actions could lead to further insights regarding the putative role of stimulus enhancement and observational learning in this task.

3. Olfactory and tactile information exchange. Finally, social information exchange could have occurred through olfactory and tactile information exchange. It has been found that in rats, food preferences can be transferred from demonstrators to observers and that for this effect to occur, the scent of the food together with the presence of carbon disulphide (CS2) on the breath of the demonstrator are necessary [50]. Preliminary observations of the interaction between our rats in this task showed that facial touch occurs throughout all conditions in the interaction window (Fig. 1B). The presence of pellet consumption related olfactory cues and associated CS2 on the partner's breath could thus have aided the target rats in identifying BR trials, while an absence or decrease in these indicator smells might indicate OR/NR trials. In addition, it has been found that facial touch is associated with an increase in the production of 50 kHz calls [51]. Within a BR condition, where sniffing of dyadic olfactory cues might be enhanced, social face touch could potentially further enhance the production of 50 kHz calls and thus contribute to unblocking.

Neural substrates

Overall, these data provide evidence that social value can drive reinforcement learning in rats, and that the transmission of social cues is necessary for this learning. Further replication experiments should be conducted to reveal which mode(s) of social information processing are necessary and sufficient to drive unblocking through social value. Previous work on the neural mechanisms of unblocking has indicated that the Basolateral amygdala [52], Nucleus Accumbens (NAcc), Orbitofrontal cortex (OFC) [23] and VTA [44] are involved in unblocking. The amygdala has been found to signal and update the exact value of the outcome while the OFC is involved in coding the different identities of the cues that predict reward [53] with a specific role for the lateral OFC in coding upshifts in rewards [32]. Next to that it was found that OFC is activated by unexpected outcome signalling an increase in reward value and that the VTA was necessary for this learning to occur [54]. VTA dopamine activity furthermore has also been found to drive learning about novel cue-reward associations in unblocking [44,55]. We propose that social reinforcement learning of novel cues takes place within the same circuitry while using similar reinforcement learning mechanism as in appetitive learning. Humans for example, learn to respond faster to stimuli that become associated with positive social reinforcement [11] and this learning was directly correlated to activity in the ventral striatum and OFC. In monkeys it was furthermore found that OFC neurons show enhanced activity for joint rewards if the partner is preferred [56] and that the BLA codes the values of rewards delivered to self and [57]. In rats it has been shown that oxytocin release in the VTA enhances the rewarding aspects of social interaction [58] and that optogenetic enhancement of VTA activity increased social behavior and this effect was dependent on projections to the NAcc [59]. Finally, in a recent study it was found that there is a specific population of OFC neurons that was activated when interacting with a social stimulus [60]. It is thus likely that a social prediction error originating from the VTA, resulting in DA efflux could selectively enhance the value of novel cues leading to novel social cue-outcome associations located in the OFC and BLA. Our novel behavioural paradigm could be used to further explore how rats learn about social value and is well suited to probe the neural circuits involved in social reinforcement learning.

Declaration of Interests

The authors declare no competing interests

References

1. Baumeister, R.F., and Leary, M.R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. Psychol. Bull. *117*, 497–529.
2. House, J., Landis, K., and Umberson, D. (1988). Social relationships and health. Science (80). *241*, 540–545.
3. Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., and Kilts, C. (2002). A neural basis for social cooperation. Neuron *35*, 395–405.
4. Horner, V., Carter, J.D., Suchak, M., de Waal, F.B.M., and Waal, F.B.M. De (2011). Spontaneous prosocial choice by chimpanzees. Proc. Natl. Acad. Sci. U. S. A. *108*, 13847–51.
5. Hernandez-Lallement, J., van Wingerden, M., Marx, C., Srejic, M., and Kalenscher, T. (2015). Rats prefer mutual rewards in a prosocial choice task. Front. Neurosci. *8*.
6. Ben-Ami Bartal, I., Decety, J., and Mason, P. (2011). Empathy and pro-social behavior in rats. Science *334*, 1427–30.
7. Hernandez-Lallement, J., van Wingerden, M., Schäble, S., and Kalenscher, T. (2016). A Social Reinforcement Learning Hypothesis of Mutual Reward Preferences in Rats. In Current Topic in Behavioral Neuroscience, pp. 159–176.
8. Spreckelmeyer, K.N., Krach, S., Kohls, G., Rademacher, L., Irmak, A., Konrad, K., Kircher, T., and Gründer, G. (2009). Anticipation of monetary and social reward differently activates mesolimbic brain structures in men and women. Soc. Cogn. Affect. Neurosci. *4*, 158–65.
9. Deaner, R.O., Khera, A. V, and Platt, M.L. (2005). Monkeys pay per view: adaptive valuation of social images by rhesus macaques. Curr Biol *15*, 543–548.
10. Wöhr, M., and Schwarting, R.K.W. (2007). Ultrasonic communication in rats: can playback of 50-kHz calls induce approach behavior? PLoS One *2*, e1365.
11. Jones, R.M., Somerville, L.H., Li, J., Ruberry, E.J., Libby, V., Glover, G., Voss, H.U., Ballon, D.J., and Casey, B.J. (2011). Behavioral and neural properties of social reinforcement learning. J. Neurosci. *31*, 13039–45.
12. Chang, S.W.C., Winecoff, A. a, and Platt, M.L. (2011). Vicarious reinforcement in rhesus macaques (macaca mulatta). Front. Neurosci. *5*, 27.
13. Angermeier, W.F. (1960). Some basic aspects of social reinforcements in albino rats. J. Comp. Physiol. Psychol. *53*, 364–367.
14. Evans, M.J., Duvel, A., Funk, M.L., Lehman, B., Sparrow, J., Watson, N.T., and Neuringer, A. (1994). Social reinforcement of operant behavior in rats: a methodological note. J. Exp. Anal. Behav. *62*, 149–56.
15. Achterberg, E.J.M., van Kerkhof, L.W.M., Servadio, M., van Swieten, M.M.H., Houwing, D.J., Aalderink, M., Driel, N. V, Trezza, V., and Vanderschuren, L.J.M.J. (2016). Contrasting Roles of Dopamine and Noradrenaline in the Motivational Properties of Social Play Behavior in Rats. Neuropsychopharmacology *41*, 858–868.
16. Knutson, B., Burgdorf, J., and Panksepp, J. (2002). Ultrasonic vocalizations as indices of affective states in rats. Psychol. Bull. *128*, 961–77.
17. Willuhn, I., Tose, A., Wanat, M.J., Hart, A.S., Hollon, N.G., Phillips, P.E.M., Schwarting, R.K.W., and Wöhr, M. (2014). Phasic dopamine release in the nucleus accumbens in response to pro-social 50 kHz ultrasonic vocalizations in rats. J. Neurosci. *34*, 10616–23.

18. Berridge, K.C. (2012). From prediction error to incentive salience: mesolimbic computation of reward motivation. Eur. J. Neurosci. *35*, 1124–1143.

19. Schultz, W. (2016). Dopamine reward prediction-error signalling: a two-component response. Nat. Rev. Neurosci. *17*, 183–95.

20. Kashtelyan, V., Lichtenberg, N.T., Chen, M.L., Cheer, J.F., and Roesch, M.R. (2014). Observation of reward delivery to a conspecific modulates dopamine release in ventral striatum. Curr. Biol. *24*, 2564–2568.

21. Kamin, L.J. (1969). Predictability, surprise, attention and conditioning. B. A. Campbell, R. M. Church Punishm. Aversive Behav., 279–296.

22. Holland, P.C. (1984). Unblocking in Pavlovian appetitive conditioning. J. Exp. Psychol. Anim. Behav. Process. *10*, 476.

23. McDannald, M.A., Lucantonio, F., Burke, K.A., Niv, Y., and Schoenbaum, G. (2011). Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. J. Neurosci. *31*, 2700–5.

24. Kohls, G., Chevallier, C., Troiani, V., and Schultz, R.T. (2012). Social “wanting” dysfunction in autism: neurobiological underpinnings and treatment implications. J. Neurodev. Disord. *4*, 10.

25. Fulford, D., Campellone, T., and Gard, D.E. (2018). Social motivation in schizophrenia: How research on basic reward processes informs and limits our understanding. Clin. Psychol. Rev. *63*, 12–24.

26. Honey, R.C. (1990). Stimulus Generalization as a Function of Stimulus Novelty and Familiarity in Rats. *16*, 178–184.

27. Channell, S., and Hall, G. (1983). Contextual effects in latent inhibition with an appetitive conditioning procedure. *11*, 67–74.

28. Burke, K.A., Franz, T.M., Miller, D.N., and Schoenbaum, G. (2008). The role of the orbitofrontal cortex in the pursuit of happiness and more specific rewards. Nature *454*, 340–4.

29. Ruff, C.C., and Fehr, E. (2014). The neurobiology of rewards and values in social decision making. Nat. Rev. Neurosci.

30. Rescorla, and Wagner (1972). Theory of Classical Conditioning.

31. Fanselow, M.S., and Wassum, K.M. (2016). The origins and organization of vertebrate pavlovian conditioning. Cold Spring Harb. Perspect. Biol. *8*, 1–28.

32. Lopatina, N., McDannald, M.A., Steyer, C. V, Sadacca, B.F., Cheer, J.F., and Schoenbaum, G. (2015). Lateral orbitofrontal neurons acquire responses to upshifted, downshifted, or blocked cues during unblocking. Elife *4*, e11299.

33. Sutton, R.S., and Barto, A.G. (1981). Toward a modern theory of adaptive networks: expectation and prediction. Psychol. Rev. *88*, 135–70.

34. Lehner, R., Balsters, J.H., Herger, A., Hare, T.A., and Wenderoth, N. (2017). Monetary, Food, and Social Rewards Induce Similar Pavlovian-to-Instrumental Transfer Effects. Front. Behav. Neurosci. *10*, 1–12.

35. Kruse, J.M., Overmier, J.B., Konz, W.A., and Rokke, E. (1983). Pavlovian conditioned stimulus effects upon instrumental choice behavior are reinforcer specific. Learn. Motiv. *14*, 165–181.

36. Rescorla, R.A. (1994). Control of Instrumental Performance by Pavlovian and Instrumental Stimuli. J. Exp. Psychol. Anim. Behav. Process. *20*, 44–50.

37. Humphreys, A.P., and Einon, D.F. (1981). Play as a reinforcer for maze-learning in juvenile rats. Anim. Behav. *29*, 259–270.

38. Calcagnetti, D.J., and Schechter, M.D. (1992). Place conditioning reveals the rewarding aspect of social interaction in juvenile rats. Physiol. Behav. *51*, 667–672.

39. Noritake, A., Ninomiya, T., and Isoda, M. (2018). Social reward monitoring and valuation in the macaque brain. Nat. Neurosci. *21*, 1452–1462.

40. Brosnan, S.F., and De Waal, F.B. (2003). Monkeys reject unequal pay. Nature *425*, 297–299.

41. Oberliessen, L., Hernandez-Lallement, J., Schäble, S., van Wingerden, M., Seinstra, M., and Kalenscher, T. (2016). Inequity aversion in rats, Rattus norvegicus. Anim. Behav. *115*, 157–166.

42. Márquez, C., Rennie, S.M.M., Costa, D.F.F., and Moita, M.A.A. (2015). Prosocial Choice in Rats Depends on Food-Seeking Behavior Displayed by Recipients. Curr. Biol. *25*, 1736–1745.

43. Nicol, C.J. (1995). The social transmission of information and behaviour. Appl. Anim. Behav. Sci. *44*, 79–98.

44. Chang, C.Y., Gardner, M., Di Tillio, M.G., and Schoenbaum, G. (2017). Optogenetic Blockade of Dopamine Transients Prevents Learning Induced by Changes in Reward Features. Curr. Biol. *27*, 3480–3486.e3.

45. Keiflin, R., Pribut, H.J., Shah, N.B., Janak, P.H., Keiflin, R., Pribut, H.J., Shah, N.B., and Janak, P.H. (2019). Ventral Tegmental Dopamine Neurons Participate in Article Ventral Tegmental Dopamine Neurons Participate in Reward Identity Predictions. 93–103.

46. Saito, Y., Yuki, S., Seki, Y., Kagawa, H., and Okanoya, K. (2016). Cognitive bias in rats evoked by ultrasonic vocalizations suggests emotional contagion. Behav. Processes *132*, 5–11.

47. Cleland, G.G., and Davey, G.C.L. (1983). Autoshaping in the rat: The effects of localizable visual and auditory signals for food. J. Exp. Anal. Behav. *40*, 47–56.

48. Spence, K.W. (1937). Experimental studies of learning and the higher mental processes in infra-human primates. Psychol. Bull. *34*, 806–850.

49. Heyes, C.M., Ray, E.D., Mitchell, C.J., and Nokes, T. (2000). Stimulus Enhancement: Controls for Social Facilitation and Local Enhancement. Learn. Motiv. *31*, 83–98.

50. Galef, Jr., B. (2001). Social Influences on Food Choices of Norway Rats and Mate Choices of Japanese Quail. Int. J. Comp. Psychol. *14*.

51. Rao, R.P., Mielke, F., Bobrov, E., and Brecht, M. (2014). Vocalization–whisking coordination and multisensory integration of social signals in rat auditory cortex. Elife *3*.

52. Chang, S.E., McDannald, M.A., Wheeler, D.S., and Holland, P.C. (2012). The effects of basolateral amygdala lesions on unblocking. Behav. Neurosci. *126*, 279–289.

53. Sharpe, M.J., and Schoenbaum, G. (2016). Back to basics: Making predictions in the orbitofrontal-amygdala circuit. Neurobiol. Learn. Mem.

54. Takahashi, Y.K., Roesch, M.R., Stalnaker, T.A., Haney, R.Z., Calu, D.J., Taylor, A.R., Burke, K.A., and Schoenbaum, G. (2009). The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes. Neuron *62*, 269–280.

55. Keiflin, R., Pribut, H.J., Shah, N.B., and Janak, P.H. (2019). Ventral Tegmental Dopamine Neurons Participate in Reward Identity Predictions. Curr. Biol. *29*, 93–103.e3.

56. Azzi, J.C.B., Sirigu, a., and Duhamel, J.-R. (2012). Modulation of value representation by social context in the primate orbitofrontal cortex. Proc. Natl. Acad. Sci. *109*, 2126–2131.

57. Chang, S.W.C., Fagan, N.A., Toda, K., Utevsky, A. V, and Pearson, J.M. (2015). Neural mechanisms of social decision-making in the primate amygdala.

58. Hung, L.W., Neuner, S., Polepalli, J.S., Beier, K.T., Wright, M., Walsh, J.J., Lewis, E.M., Luo, L., Deisseroth, K., Dölen, G., *et al.* (2017). Gating of social reward by oxytocin in the ventral tegmental area. Science (80-.). *357*, 1406–1411.

59. Gunaydin, L.A., Grosenick, L., Finkelstein, J.C., Kauvar, I. V., Fenno, L.E., Adhikari, A., Lammel, S., Mirzabekov, J.J., Airan, R.D., Zalocusky, K.A., *et al.* (2014). Natural neural projection dynamics underlying social behavior. Cell *157*, 1535–51.

60. Jennings, J.H., Kim, C.K., Marshel, J.H., Raffiee, M., Ye, L., Quirin, S., Pak, S., Ramakrishnan, C., and Deisseroth, K. (2019). Interacting neural ensembles in orbitofrontal cortex for social and feeding behaviour. Nature.