High mutual cooperation rates in rats learning reciprocal altruism: the role of payoff matrix

Guillermo Ezequiel Delmas[1�явел], Sergio Eduardo Lew[1‡], Bonifacio Silvano Zanutto[1,2¤],

1 Instituto de Ingeniería Biomédica, Facultad de Ingeniería, Universidad de Buenos Aires, Buenos Aires, Argentina
2 Instituto de Biología y Medicina Experimental (IBYME-CONICET), Laboratorio de Biología del Comportamiento, Ciudad de Buenos Aires, Buenos Aires, Argentina, Instituto de Ingeniería Biomédica, Universidad de Buenos Aires, Ciudad de Buenos Aires, Buenos Aires, Argentina

☯ Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Validation, Visualization, Writing–Original Draft, Writing – Review & Editing.
‡ Formal Analysis, Validation, Visualization, Writing – Review & Editing.
¤ Conceptualization, Funding Acquisition, Investigation, Methodology, Resources, Supervision, Validation, Visualization, Writing–Original Draft, Writing – Review & Editing.
☯ gedelmas@gmail.com

Abstract

Cooperation is one of the most studied paradigms for the understanding of social interactions. Reciprocal altruism -a special type of cooperation that is taught by means of the iterated prisoner dilemma game (iPD)- has been shown to emerge in different species with different success rates. When playing iPD against a reciprocal opponent, the larger theoretical long-term reward is delivered when both players cooperate mutually. In previous experiments, rats showed low mutual cooperation rates. In this work, we trained rats in iPD against an opponent playing a Tit for Tat strategy, using a payoff matrix with positive and negative reinforcements, that is food and timeout respectively. We showed for the first time, that experimental rats were able to learn reciprocal altruism with a high average cooperation rate, where the most probable state was mutual cooperation (85%), but if subjects defected the most probable behavior was to go back to mutual cooperation. When, we modified the matrix by increasing temptation rewards (T) or by increased cooperation rewards (R), the cooperation rate decreased. In conclusion, we observe that an iPD matrix with large positive reward improves less cooperation than one with small rewards, shown that satisfying the relationship among iPD reinforcement was not enough to achieve high mutual cooperation behavior. Therefore, using positive and negative reinforcements and an appropriate contrast between rewards, rats have cognitive capacity to learn reciprocal altruism. This finding allows to infer that the learning of reciprocal altruism has appeared early in evolution.

Author summary

The reciprocal altruism is achieved when an individual makes a costly act in benefit of another and later it is benefited by the other in return. Subjects have to learn to

maximize long term profitability using this reciprocal behavior instead of selfish behavior (getting a long-term reward). In human beings and some animals (as monkeys) this behavior had been observed in laboratory conditions, but in animals with less cognitive abilities (as rats or birds) cooperation has been poorly seen. We have studied if it is due to cognitive abilities or due to other reasons. The reciprocal altruism used to be studied in paradigms where an experimental subject faces an opponent repeatedly having two possible options: cooperate or defect, when the opponent uses a reciprocal strategy (TitForTat: start cooperating and then copy the other's last choice). In this protocol, the best theoretical strategy is to cooperate. Using a matrix with positive and negative reinforcements, we found for the first time that rats developed high reciprocal altruism behaviour. Rats learned to cooperate mutually, and when they chose not to cooperate, they returned to cooperate in the following trial. In conclusion, rats learned the benefit of choosing the larger long-term reinforcement instead of an immediate, showing that even animals with less cognitive abilities are able to learn reciprocal altruism.

Introduction

Altruism is a kind of behaviour by which individuals choose to favour others in detriment of their own benefit. It is not obvious at the first sight how could Darwin's theory natural selection explain altruistic behaviour and many proposals appeared in order to account for altruist behaviours: kin selection [1], group selection and reciprocal altruism [2] among others. In the reciprocal altruism theory, the loss an individual get from being altruist returns later as part of the group gain. Thus, in the long term, being altruist become the most useful strategy. In this regard, Triver's theory of reciprocal altruism is able to explain how natural selection favours reciprocal altruism between non-related individuals. Perhaps the most insightful example of such behaviour are the observed among vampire bats, where individuals share blood with others who have previously shared their food [3]. Since 1971, Iterated Prisoner's Dilemma (iPD) has been proven a useful tool to study reciprocal altruism [4]. In the iPD, two players must choose between two possible behaviours: to cooperate or to defect. Rewards and punishments are defined in a 2x2 payoff matrix. When the game is played indefinitely, which is its iterated version, mutual cooperative behaviours are favoured. When played once, to defect is the best strategy [5]. However, when the game run indefinitely, evolutionary stable strategies (ESS) emerge ([6], [7]) and, under certain constraints imposed to the payoff matrix, mutual cooperation appears as the best strategy whenever reciprocity is maintained (*Pareto Optimum*). Among a huge number of reciprocal strategies, Tit-For-Tat is one of the most simple and robust [8]. Tit-For-Tat is based on two simple rules: to start cooperating and to do what the other player (opponent) did in the last trial.

Among many reciprocal behaviours, reciprocity and reciprocal altruism were well documented in several species. Although cooperation is needed in order to success in both reciprocity and reciprocal altruism, the latter add the possibility of getting reward by defecting an opponent. Some works had assessed reciprocal altruism behaviour by means of iPD paradigm in differents ways, but the experiments results were either low levels of cooperation [9] or depended on a treatment that enhance the cooperation preference (mutualism matrix) [10], [11], [12]. Direct reciprocity, which is established between two individuals, has been observed in monkeys [13] [14] [15] and in rats [16], [17], [18], [19]. There, while food quality seemed to impact on the cooperative behaviour, a key factor to obtain reliable cooperation levels was the opponent behaviour. In this sense, individuals tended to be more cooperative with opponents that had cooperated in the past. However, when reciprocal altruism is studied, differences

between species come to light. Thus, while reciprocal altruism has been proven in monkeys, birds and rats failed to reach high levels of cooperation, even for complex combinations of rewards and punishments in the payoff matrix and treatments to induce preference [9] [20] [12] [21] [10] [22] [23]. The reasons why some species do not learn reciprocal altruism remain obscure. A possible explanation is that animals are not able to discriminate low contrast reward contingencies. Indeed, it was shown that rats failed to discriminate the amount of reward when the number of reward units was bigger than three [24], [25], [26]. ~~Here, we used the iPD to test how payoff matrix components promote or disrupt altruistic behaviors. In that sense, positive rewards (food) have been combined with negative ones (time-out) in a game where the opponent was trained to execute a Tit for tat strategy.~~ Here, we design a iPD setup to test learning of reciprocal altruism based on a combination positive (food) and negative (timeout) reinforcement, where our goal was to choose the amounts of pellets in order both to maximization contrast among pay-off get by each iPD states and to minimized the amount of food that each rat gets in long term. These maybe help the rats to discriminate the amount of reward at long term. In order to evaluate if animals developed ALLC strategy by place preference or by reward maximization, we applied a reversion treatment. Finally, we evaluated how payoff matrix components promotes or disrupt altruistic behaviours.

Materials and methods

Subject

We used thirty male Long-Evans rats (weight 300–330g and two months old) provided by the IBYME-CONICET. Divided in two experiments, in the first one of eighteen rats, twelve experimental and six opponent, and in the second, six experimental and six opponent. Experimental subjects were housed in pairs (to allow social interaction), and opponent rats were housed individually. All rats were food restricted and maintained at 90-95% for experimental subjects, and 80-85% for opponents of free feeding body weight. And with tap water available ad libitum. The housing room was at $22°C \pm 2°C$ and 12/12 h light/dark cycle (with lights on at 9 am). Pre-training was performed on a single standard operant chamber (MED associates Inc., USA) equipped with two stimulus light and retractable levers below the light and feeders. Also the chambers were inside an anechoic chamber with white noise (with a flat power spectral density). The iPD experiments were performed in ad hoc dual chamber equipped with levers, lights and feeders (fig. 1A). The chambers were connected by windows in order that the rat could make olfactory and eye contact. The lever's height was 80% of maximum height of the forepaws while rearing [27]. The dual chamber is shown in supplementary material, see fig. . At the end of daylight, supplementary food was provided in order that rats get the amount of pellets necessary to maintain body weight.

Pre-experimental training

All rats had a shaping procedure to learn the response (press a lever) to get a reinforcement (pellets). To prevent animals from choosing a lever place over the other, they learned to get reward from both sides by changing the side of conditioned stimulus. The side was changed after eight trials. All rats learned to press the correct lighting lever after four sessions. Each rat was trained in 2 sessions per day, each trial began with the inter-trial interval (ITI) during 5 seconds, it was followed by the conditioning stimulus (light) for either 45 seconds or until a lever was pressed. One second before food is delivered, the feeder was lighted. In the opponent's training they learned to press the lever when the light was on. In the task, the side of the active lever was

chosen pseudo-randomly (allowing the same side no more than four times). The opponent subject had to perform a fix ratio treatment up to FR=5 to get rewards.

Experiment

To study the reciprocal altruism in an iterated Prisoner's Dilemma game (iPD), we used a payoff matrix with positive and negative reinforcements. Positive reinforcements were pellets (Bio-Serv 45 mg Dustless Precision Pellets) and negative reinforcement was timeout (a fix delay in starting a new trial). The payoff of the experimental subject was according to the matrix, and the opponent's payoff was 1 pellet when the correct lighted lever was pressed. The iPD game have four possible occupancy state where experimental and opponent individual behaviour can be as follows: both cooperate (mutual cooperation, R), both do not cooperate (mutual defection, P), experimental subject does not cooperate when the opponent cooperates (T), and experimental cooperates when the opponent does not cooperate (S). The amount of pellets preference was previously tested on a discrimination test, showing that rats prefer 2 pellets rather than 1 pellet (data not showed). We performed two sessions per day and each session had 30 trials. Each experimental subject was trained with the same opponent. The training was finished after five consecutive sessions with no changes in the cooperation rate. We defined cooperation (C) and defection (D) lever in the iPD box. The single iPD trial procedure was as follows: (1) ITI time, (2) then, the light (CS) was turned on, (3) after that both rats made their responses, the light was turned off and the reinforcement was delivered according to a payoff matrix, (4) if positive reinforcement was assigned, the feeder's light was turned on, and a second later a reward was delivered. The opponent Conditioned Stimulus (light) was controlled following a Tit for Tat strategy.The opponent received a pellet after pressing three times the lever (FR=3, so as to be enough time in front of the window until the experimental subject choose a lever). If negative reinforcement (timeout) was assigned, delay time started, and the opponent subject got a pellet reward. (5) After either five seconds eating time expired or timeout was completed, a new trial started. In the first experiment the payoff matrix was: 1 pellet for mutual cooperation ($P_R = 1$), 2 pellets when the experimental subject defected and the opponent cooperated ($P_T = 2$), 4 seconds of timeout for mutual defected ($P_P = 4 seconds$), and 8 seconds of timeout when the experimental subject cooperated and the opponent defected ($P_S = 8$). At the end of these experiment, the four rats with the best performance in cooperation were trained in a reversion treatment (see Fig. 1F). The reversion consist on inter-change sides of C and D lever in both subject and opponent chamber. In that sense, if a animals has a place preference behavior, he will not learn the new side that maximize reward. In the second experiment we used six naive experimental rats on a different payoff matrix with greater temptation ($P_R = 1$, $P_T = 3$, $P_P = 4$, $P_S = 8$). After training, we divided rats in two groups, depending on the cooperation levels. The first group with high cooperation rate was trained with the payoff matrix ($P_R = 1$, $P_T = 5$, $P_P = 4$, $P_S = 8$) with greater temptation for T state. The other group (with low cooperation rate) was trained with the matrix ($P_R = 2$, $P_T = 3$, $P_P = 4$, $P_S = 8$) that enhances cooperative behaviour (in comparison with ($P_R = 1$, $P_T = 3$, $P_P = 4$, $P_S = 8$), but with low contrast between positive rewards. All experimental procedures were approved by the ethics committee of the IByME-CONICET and were conducted according to the NIH Guide for Care and Use of Laboratory Animals.2.1 Subjects and Housing.

Statistic

All statistical analysis were performed using statistics library from open source software Octave and MATLAB. We pooled the data from the last five sessions where cooperation

rate was stable (to calculate cooperation rate we counted the number of times a rat   132
chose the cooperation lever per session). We compared individual's means of   133
cooperation along treatment using a two-sided Wilcoxon rank sum test. To test whether   134
the probability of cooperation after each outcome (T,R,P or S) was different from   135
chance (0.5), we performed a Chi-square goodness of fit test with Bonferroni corrected   136
value of 0.05/n. To compare mean rate of the different outcomes for each game, we   137
performed an ANOVA two tails test. When significant $\alpha = 0.05$, multiple post-hoc   138
pairwise comparative tests were performed with Bonferroni corrected value of   139
$\alpha = 0.0125$. The individual's decision rules can be described by the components of   140
transition vectors and Markov Chain diagram. The transition vector was made up of   141
probabilities of cooperation when the previous trials resulted in state R(reward,   142
p(c—R-1)), T(temptation, p(c—T-1)), S(sucker, p(c—S-1)) or P(punishment,   143
p(c—P-1)) respectively. If every component of this vector is 0.5, the agent's decision   144
rule is random mode. Markov Chain diagram show the graphic representation of the   145
complete decision making rule for each rat.   146

Results   147

We trained twelve rats in iPD against an opponent that plays Tit for Tat strategy.   148
Tit-For-Tat is based on two simple rules: to start cooperating and to do what the other   149
player (opponent) did in the last trial. Fig 1A shows a schema of the different choices a   150
subject can do in each trial. Thus, when the subject cooperates, it receives one pellet   151
(P_R) or eight seconds timeout (P_S) depending on whether the opponent choice was to   152
cooperate or to defect. On the other hand, when the subject defects, it receives 2 pellets   153
(P_T) or four seconds timeout (P_P), according to whether the opponent choice was to   154
cooperate or to defect respectively. The criteria for cooperation was an established the   155
preference for pressing C lever (cooperation) over D lever (defection) in more than 60%   156
of the trials for five or more consecutive sessions. Eight out of the twelve animals   157
learned to cooperate (cooperation rate 0.86 ± 0.05, mean $\pm$ s.e.m), reaching criteria in   158
30 ± 4 sessions, mean $\pm$ s.e.m. In fig 1B, we show the mean cooperation levels for those   159
animals during the last twenty three sessions before reaching criteria. The inset in fig   160
1B shows the mean cooperation level for each animal during the last five training   161
sessions. As a consequence of the increase in cooperation levels, the average total   162
timeout per session decreased as training progressed (0.23 ± 0.08, mean $\pm$ sem, see fig   163
1C). Due to the fact that many sequences of lever pressing can give the same amount of   164
reward and/or timeout, independently of the cooperation level. we analyzed the   165
relationship between total reward and timeout for each animals respect to a simulated   166
population. Thus, for each animal we compared those values with a simple regression   167
model fitted to a population of 100.000 simulated individuals when their cooperation   168
level was set to 60%, see fig 1D. Each simulated individual had a different complete   169
strategy and each strategy was a combination of thirty C and D choice (session length).   170
An individual that play a iPD game with a 60% of her choices in C will be near to the   171
line regardless of her strategies. As can be seen in the figure, the higher the cooperation   172
levels, the larger the total reward and the lower the total timeout. Accordingly the fig.   173
1D, the group of cooperator's rats developed a complete different behavior respect both   174
line regression with 60% of cooperation and the no cooperator group. It means that not   175
exist any strategy with low level of cooperation that get high level of reward and small   176
timeout as in the cooperative group. We then built one Markov model for the group of   177
cooperative animals (see fig 1E) averaging occupancy state rate and transition   178
probabilities in the group. In the iPD there are four possible occupancy state where   179
experimental and opponent individual behaviour can be as follows: R (both cooperate   180
or mutual cooperation), P (both do not cooperate or mutual defection), T   181

(experimental subject does not cooperate when the opponent cooperates), and S (experimental cooperates when the opponent does not cooperate). For the group of non-cooperative animals see fig 1S (supplementary materials). The cooperative group showed that the permanency in R state was high in cooperative animals and, whenever the animal defects (states T and P), it returns to cooperate in most of the cases. All conditional probabilities to cooperate given a previous outcome was near 1. Besides, the probability of R state was the highest and other states near zero. The probability of R state was significantly different to other states ($p = < 1e^{-8}$, ANOVA two-way test, n=8). On the contrary, in the group of non-cooperative animals, any states were significantly different to the other $p > 0.05$, F=0.353, ANOVA two-way test, n=4) and the probability to cooperate given a previous states did not evidence preference for any defined strategy. See the table X

To discard the fact that animals had have a preference for one of the levers and, in consequence, their behaviour biased independently of the training paradigm, we select the best four cooperators and apply a reversal procedure immediately after cooperation was reached. All animals learned to cooperate after reversal (0.87 ± 0.04, mean $\pm$ sem), see fig 1F.

Fig 1. High level of cooperation in iPD. (**A**) Dual operant box diagram and the matrix with positive(blue) and negative(red) reinforcement is shown. The iPD game had four possible states: R(reward) mutual cooperation, P(punishment) mutual defection, T(temptation) in which subject defected and opponent cooperated and S(sucker) subject cooperated and opponent defected. The opponent´s light was driven in order to perform a Tit for Tat strategy. (**B,C**) Time-course of cooperation and timeout rate along the last 23 games sessions. In the last 5 sessions, the mean $\pm$ sem of cooperation was 0.86 ± 0.05 and timeout was 0.23 ± 0.08. (**D**) Total reward versus timeout for all animals (color bar means cooperation mean). Each animal was compared with the theoretical reward-timeout function when the cooperation level was set to 60% (black continuous line). The higher the cooperation levels, the larger the total reward and the lower the total timeout.(**E**) Markov Chain diagram shows the probabilities of transition between states ($p(c|T_{-1}) = 0.76$, $p(c|R_{-1}) = 0.85$, $p(c|S_{-1}) = 0.93$, $p(c|P_{-1}) = 0.87$). The arrow represents transitions: driven by cooperation in blue, and driven by defection in red (the arrow thickness is proportional to transition probability). The size of circles is proportional to the state occupancy ratio. Below, bars show the occupancy ratio when the cooperation reaches stability. The probabilities were: $p(R) = 0.76$, $p(T) = 0.1$, $p(P) = 0.04$, $p(S) = 0.1$. Asterisks denote significant differences from multiple comparisons using one-way ANOVA and Bonferroni correction. (**F**) Evolution cooperation rate before and after reversion treatment. Graphs show a moving average with samples of 3 sessions (the mean and sem from reversion on the last five sessions was 0.87 ± 0.04).

We then asked how the ratio in the amount of positive reinforcement of R and T states affects cooperation learning and maintenance. We defined a contrast index CI that measures the relationship between the amount of reward in R and T as follows:

$$CI = \frac{T - R}{T + R}$$

Thus, in the experiment shown in fig 1 CI was $\frac{1}{3}$, which is the maximum contrast level constrained to a payoff matrix that favours cooperation, that is, $2R > T + S$, assuming that S becomes a negative stimulus induced by timeout. We trained six animals with a payoff matrix ($R = 1$, $T = 3$, $P = 4$, $S = 8$) and found that three animals learned to cooperate (0.88 ± 0.01, mean $\pm$ sem, see fig 2A), while others did not (0.64 ± 0.13,

mean ± sem, see fig 2B. The last group was no cooperator since both their conditional probabilities to cooperate and occupancy R state ratio were near chance. For details see table 1. Then we changed the amount of reward in order to increase/decrease CI in the cooperative/non-cooperative groups. As it can be seen, a high value of $CI = \frac{2}{3}$ disrupts cooperation in cooperative group, fig 2A. The cooperation was 0.604 ± 0.102, mean ± sem whereas before 0.88 ± 0.01). When a lower value of $CI = \frac{1}{5}$ was applied for not cooperator group and the matrix empowers the cooperation in two out of three animals, see fig 2B (0.711 ± 0.04, mean ± sem, whereas before 0.64 ± 0.13. See table 1.

We analyzed how these changes in strategies impact on the amount of received reward and timeout penalties. In the group of cooperative animals, the change in T (3 pellets to 5 pellets) increased both timeout and a bit a reward, as expected when states T, P and S become more probable. The occupancy states ratio before and after matrix change had significant difference among all states, p¡0.05 (wilcoxon ranksum test). See fig 2C and 2E. It is worth noting however that the amount of received reward is not the maximum allowed, which would be delivered in the case of an animal that alternates from state T to S indefinitely. On the other hand, when we applied a matrix with a lower contrast $CI = \frac{1}{5}$ to the group of non-cooperative animals, they enhance significantly her cooperation level, receiving more reward without significant changes in total timeout, see fig 2D. In fig 2F, we show the state occupancy probabilities for this group before and after the change in the payoff matrix. It can be seen that the occupancy state ratio of R had significantly increased after the change in the payoff matrix. It can be observed a significant difference in R and P states, ($p_R < 0.008$ and $p_P < 0.048$, wilcoxon rank-sum test). We showed that when the contrast index increased using a matrix to favor the cooperation the animals learned to cooperate, but when the index increased and the matrix favor defection the animals left to cooperate.

Fig 2. Effect of changes in the amount of positive reinforcement of R and T. (A) The rats were pre-trained by pay-off matrix [$R = 1$, $T = 3$, $P = 4$, $S = 8$ and contrast $CI = \frac{1}{2}$] (filled dots) and the cooperation was strongly affected by change of temptation payoff, decreasing when T payoff increased and matrix with changed to [$R = 1$, $T = 5$, $P = 4$, $S = 8$ and contrast $CI = \frac{2}{3}$] (open circles). There was a significantly difference (red circle) in two animals with $p < 9.8e^{-06}$ (wilcoxon rank-sum test) and the other not modify her behavior in spite of matrix change. **(B)** The cooperation enhances when the matrix changed to [$R = 2$, $T = 3$, $P = 4$, $S = 8$ and $CI = \frac{1}{5}$] (open circles) and the difference was statistically different ($p < 0.0062$) in two of three subjects, because one had no significant difference after matrix change, $p > 0.05(0.7063)$. **(C)** In the 3D plots were related cooperation, reward and timeout. In the group of cooperative animals (filled dots), the change in T (3 pellets to 5 pellets) increased both timeout and reward in order to decrease cooperation (open circles). The comparison between cooperation mean of both group was significantly different, $p < 0.05(6.7e^{-05})$.**(D)** In the group of non-cooperative animals (filled dots), they learned to cooperate (open circles) by receiving more reward without significant changes in total timeout. The cooperation was significantly different, $p > 0.05(0.081)$. **(E,F)** The mean of occupancy state rate graph (last five sessions) from cooperative (left) and non-cooperative (right) groups (Mean ± sem). Asterisks denote significant difference, after matrix changed, among T, R, P or S state occupancy and dash line indicates the level of equal rate in each state (that corresponds to a strategy with strongly random component). Before changes (filled dots) and after changes (open circles).

From the results shown in figs 1 and 2, it is reasonable to ask whether a fine tuning in contrasted reward encourages cooperative behaviour. We have shown that eight out of twelve (66%) animals acquired a cooperative behaviour when CI was $\frac{1}{3}$ while three out of six (50%) succeeded when CI was $\frac{1}{2}$, as expected when temptation payoff

Table 1. Caption

increases. In the same line of reasoning, animals that had learned cooperation under ²³⁶
CI=$\frac{1}{2}$ disrupted their cooperative behaviour when CI was increased to $\frac{2}{3}$, while those ²³⁷
that had not learned acquired a cooperative behaviour when CI was decreased to $\frac{1}{5}$. ²³⁸
Fig 3A exemplifies the occupancy and transition probabilities for an animal that ²³⁹
disrupted its cooperative behaviour when $CI = \frac{1}{2}$ was changed to $CI = \frac{2}{3}$. The ²⁴⁰
opposite can be seen in the example of fig 3B. A non-cooperative animal under a ²⁴¹
$CI = \frac{1}{2}$ became cooperative when CI was decreased to $\frac{1}{5}$. Figs 3C and 3D show ²⁴²
cooperation levels and normalized rewards. It can be seen that both variables follow an ²⁴³
inverted U profile as a function of contrast index CI, as expected when a delicate ²⁴⁴
balance between rewards at R and T is mandatory. ²⁴⁵

Fig 3. Markov chain diagrams and contrast index. Markov chain diagrams are
shown (the size of circle means of occupancy state rate and the arrow's width are
proportional to the probability of cooperate given (**A**) Occupancy state and transition
probabilities for an animal that disrupted its cooperative behaviour when contrast index
$CI = \frac{1}{2}$ was changed to $CI = \frac{2}{3}$ and pay-off matrix was changed
$[P_T,P_R,P_P,P_S]$=[3p,1p,4s,8s] to [5p,1p,4s,8s] (p=pellet and s=seconds). The line
thickness of blue arrows (conditional probabilities of cooperation) become more thin
after change. (for values see table 1. (**B**) The opposite situation can be seen,
non-cooperative animal becomes more cooperative when $CI = \frac{1}{2}$ was decreased to
$CI = \frac{1}{5}$ in a matrix that favours cooperation. The blue arrows become more thick after
change (for values see table 1. (**C**, **D**) show cooperation and timeout levels as a
function of CI. Here, it can be seen that both variables follow an inverted U profile in
correlation with the contrast index increase and if the payoff matrix favours or not the
cooperation behaviour.

Discussion and Conclusion ²⁴⁶

In this work, we study the contrasted role between reinforcements in reciprocal altruism ²⁴⁷
learning in rats. Traditionally, reciprocal altruism is achieved by playing the iterated ²⁴⁸
prisoner's dilemma game (iPD) when an experimental subject is confronted to a ²⁴⁹
reciprocal opponent. The payoff matrix used had positive and negative reinforcements ²⁵⁰
with highly contrasted between pairs, positive pairs and negative pairs and also using ²⁵¹
discriminable amount of reinforcements [24–26]. In our experiment, pellets were used as ²⁵²
positive reinforcements and timeout as negative reinforcements. In this way, the positive ²⁵³
and negative reinforcements acted as strengtheners of mutual cooperation behaviour ²⁵⁴
likelihood [28]. To our knowledge, results show for the first time high levels of ²⁵⁵
cooperation (86,11%) and mutual cooperation (76,32%) in iPD, see Fig 1B. A dynamic ²⁵⁶
system can be represented with Markov diagrams and its associated state transition ²⁵⁷
vector. In this case, each state (T, R, P, S, see Results section) will have two associated ²⁵⁸
conditional probabilities: to cooperate or not to cooperate given state. An individual ²⁵⁹
will adopt an altruist reciprocal behaviour if when playing with an opponent with a Tit ²⁶⁰
for Tat strategy, the cooperation probability is near 1, independently of the current ²⁶¹
occupancy state (T, R, P o S). And while the opponent perform a reciprocal behaviour, ²⁶²
the best strategy is to return to the mutual cooperation state, R. In the first experiment ²⁶³
in this work, we found that animals adopted two well defined strategies (fig 1D). On one ²⁶⁴

hand, a group of 8 animals proved to have learned a cooperative strategy while other 4 animals answered at random (see fig A Supporting information). The strategy of the first group (fig 1E) showed cooperation probabilities according to their occupancy state T, R, P or S in 0.760, 0.845, 0.929 and 0.870 respectively, and the second group showed not significantly different from random (see fig B Supporting information). In various works, results were presented with Markov diagrams and its associated transition vector [23] [10, 29] [11] and showed that conditional probabilities of cooperation were not high when facing a reciprocal opponent. In this protocol, with the matrix T=2p,R=1p,P=4s and S=8s, there are two theoretical strategies that maximize appetitive reinforcement: one is ALLC strategy and the other an alternating between cooperation (C) and defection (D) strategy. The latter, also maximizes positive reinforcement when alternating between cooperation and defection option, but it also increases negative reinforcement (timeout). In this case, ALLC strategy is the only one that maximizes positive reinforcement and minimizes the negative one (Pareto Optimum). Since negative reinforcement is timeout, ALLC strategy gives more food per unit of time. In this case, the role of the negative reinforcement appears.

In order to evaluate if animals developed ALLC strategy by place preference or by reward maximization, we applied a reversion treatment, see fig 1F, and we observed that animals relearn reciprocal altruism when they were exposed to a new lever's contingency.

Finally, after animals adopted a strategy, we wanted to evaluate if a change in the payoff matrix could modify their behaviour, to do so we studied the effect of modifying positive reinforcements, see fig 2A and 2B. Animals were pre-trained with a payoff matrix where alternating between C and D strategy gives more positive reinforcements than with an ALLC strategy, and keeping the same negative reinforcement as in the first experiment. It was observed that only half of the animals learned to cooperate although all of them obtained the same mean amount reward (pellet), see fig 2C, 2D. Then, a matrix with an increased payoff T was applied to the cooperative group (fig 2A), and we observed that cooperative behaviour decreased. Animals reduced R frequencies and increased P frequencies, proving that they prefered a small-immediate option instead of a large-delayed option. This behaviour is similar to the one observed in birds ([30]). In the second group, we applied a matrix that keeps the proportions of reinforcements in T and R similar to the most common matrix (T=3p,R=2p equal proportion to T=6p,R=4). It was observed that animals modified their behaviour and became more cooperative (fig 2B). These results show that animals that learned to cooperate with an appropriate matrix, stop cooperating when a temptation payoff (T) was enough increased (matrix with high contrast index). However, if non-cooperative animals are trained with a matrix that favours cooperation (matrix with low contrast index), they become cooperators. In the latter case, cooperation levels achieved are comparable to results that are shared in diverse bibliography. The main differences with other research works are the levels of cooperation and mutual cooperation achieved. A possible explanation is that animals could not discriminate among the reinforcements obtained, preventing them from learning that in the long-term the large-delayed option provides more reinforcement and consequently they did not learn iPD. [31], [30], [4], [22], [20], [32], [33], [21], [34], [29], [9], [23]. We observe that if a iPD matrix uses large positive reward, it will improve less cooperation than one with small rewards, shown that satisfying the relationship among iPD reinforcement was not enough to achieve high mutual cooperation behavior. The reciprocal altruist behaviour in humans, monkeys and elephants has been studied in laboratories showing high levels of cooperation [[35], [36], [13], [15], [37], however in rats and birds those levels of cooperation were much lower. Our results showed that by using positive and negative reinforcements and an appropriate contrast index in order, to favour reinforcement discrimination, rats proved to have the cognitive capacity to learn reciprocal altruism.

Supporting information

317

S1 Fig. Non-cooperative rats. (**A**) Time-course of cooperation rate along the last 23 games sessions. In the last 5 sessions, the mean $\pm$ sem of cooperation was 0.36 ± 0.03. (**B**) Markov Chain diagram shows the probabilities of transition between states ($p(c|T_{-1}) = 0.44$, $p(c|R_{-1}) = 0.38$, $p(c|S_{-1}) = 0.32$, $p(c|P_{-1}) = 0.32$). The arrow represents transitions: driven by cooperation in blue, and driven by defection in red (the arrow thickness is proportional to transition probability). The size of circles is proportional to the state occupancy ratio. Below, bars show the occupancy ratio ($T = 0.25$, $R = 0.19$, $P = 0.33$, $S = 0.23$ and $p > 0.05$, F=0.353, ANOVA two-way test, n=4) and transition probabilities ($p(c|T_{-1}) = 0.43$, $p(c|R_{-1}) = 0.38$, $p(c|S_{-1}) = 0.32$ and $p(c|P_{-1}) = 0.31$) did not evidence preference for any defined strategy. Asterisks denote significant differences from multiple comparisons using one-way ANOVA and Bonferroni correction.

318
319
320
321
322
323
324
325
326
327
328
329

S2 Fig. Dual operant chamber. (**A**)

330

Acknowledgments

331

References

1. Smith JM. Group selection and kin selection. Nature. 1964;201(4924):1145–1147.

2. Trivers RL. The evolution of reciprocal altruism. The Quarterly review of biology. 1971;46(1):35–57.

3. Wilkinson GS. Reciprocal altruism in bats and other mammals. Ethology and Sociobiology. 1988;9(2–4):85–100.

4. Flood M, Lendenmann K, Rapoport A. 2× 2 Games played by rats: Different delays of reinforcement as payoffs. Systems Research and Behavioral Science. 1983;28(1):65–78.

5. Doebeli M, Hauert C. Models of cooperation based on the Prisoner's Dilemma and the Snowdrift game. Ecology letters. 2005;8(7):748–766.

6. Von Neumann J, Morgenstern O. Game theory and economic behavior. Joh Wiley and Sons, New York. 1944;.

7. Nash JF, et al. Equilibrium points in n-person games. Proceedings of the national academy of sciences. 1950;36(1):48–49.

8. Hamilton WD, Axelrod R. The evolution of cooperation. Science. 1981;211(27):1390–1396.

9. Wood RI, Kim JY, Li GR. Cooperation in rats playing the iterated Prisoner's Dilemma game. Animal behaviour. 2016;114:27–35.

10. Stephens DW, McLinn CM, Stevens JR. Discounting and reciprocity in an iterated prisoner's dilemma. Science. 2002;298(5601):2216–2218.

11. Kéfi S, Bonnet O, Danchin E. Accumulated gain in a Prisoner's Dilemma: which game is carried out by the players? Animal Behaviour. 2007;4(74):e1–e6.

12. St-Pierre A, Larose K, Dubois F. Long-term social bonds promote cooperation in the iterated Prisoner's Dilemma. Proceedings of the Royal Society of London B: Biological Sciences. 2009;276(1676):4223–4228.

13. De Waal FB. Attitudinal reciprocity in food sharing among brown capuchin monkeys. Animal Behaviour. 2000;60(2):253–261.

14. Mendres KA, de Waal FB. Capuchins do cooperate: the advantage of an intuitive task. Animal Behaviour. 2000;60(4):523–529.

15. Hauser MD, Chen MK, Chen F, Chuang E. Give unto others: genetically unrelated cotton-top tamarin monkeys preferentially give food to those who altruistically give food back. Proceedings of the Royal Society of London B: Biological Sciences. 2003;270(1531):2363–2370.

16. Rutte C, Taborsky M. Generalized reciprocity in rats. PLoS biology. 2007;5(7):e196.

17. Rutte C, Taborsky M. The influence of social experience on cooperative behaviour of rats (Rattus norvegicus): direct vs generalised reciprocity. Behavioral Ecology and Sociobiology. 2008;62(4):499–505.

18. Schneeberger K, Dietz M, Taborsky M. Reciprocal cooperation between unrelated rats depends on cost to donor and benefit to recipient. BMC evolutionary biology. 2012;12(1):41.

19. Dolivo V, Taborsky M. Norway rats reciprocate help according to the quality of help they received. Biology letters. 2015;11(2):20140959.

20. Green L, Price PC, Hamburger ME. Prisoner's dilemma and the pigeon: Control by immediate consequences. Journal of the experimental analysis of behavior. 1995;64(1):1–17.

21. Stephens DW, Anderson D. The adaptive value of preference for immediacy: when shortsighted rules have farsighted consequences. Behavioral Ecology. 2001;12(3):330–339.

22. Gardner RM, Corbin TL, Beltramo JS, Nickell GS. The Prisoner's Dilemma game and cooperation in the rat. Psychological Reports. 1984;55(3):687–696.

23. Viana DS, Gordo I, Sucena E, Moita MA. Cognitive and motivational requirements for the emergence of cooperation in a rat social game. PloS one. 2010;5(1):e8483.

24. Capaldi E, Miller DJ. Counting in rats: Its functional significance and the independent cognitive processes that constitute it. Journal of Experimental Psychology: Animal Behavior Processes. 1988;14(1):3.

25. Killeen PR. Incentive theory: II. Models for choice. Journal of the Experimental Analysis of Behavior. 1982;38(2):217–232.

26. Killeen PR. Incentive theory: IV. Magnitude of reward. Journal of the experimental analysis of behavior. 1985;43(3):407–417.

27. Cabrera F, Sanabria F, Jiménez ÁA, Covarrubias P. An affordance analysis of unconditioned lever pressing in rats and hamsters. Behavioural processes. 2013;92:36–46.

28. Mazur JE. Learning and behavior. Psychology Press; 2015.

29. Stevens JR, Stephens DW. The economic basis of cooperation: tradeoffs between selfishness and generosity. Behavioral Ecology. 2004;15(2):255–261.

30. Clements KC, Stephens DW. Testing models of non-kin cooperation: mutualism and the Prisoner's Dilemma. Animal Behaviour. 1995;50(2):527–535.

31. Baker F, Rachlin H. Teaching and learning in a probabilistic prisoner's dilemma. Behavioural Processes. 2002;57(2):211–226.

32. Márquez C, Rennie SM, Costa DF, Moita MA. Prosocial choice in rats depends on food-seeking behavior displayed by recipients. Current Biology. 2015;25(13):1736–1745.

33. Mesterton-Gibbons M, Adams ES. The economics of animal cooperation. Science. 2002;298(5601):2146–2147.

34. Stephens DW, McLinn CM, Stevens JR. Effects of temporal clumping and payoff accumulation on impulsiveness and cooperation. Behavioural processes. 2006;71(1):29–40.

35. Wedekind C, Milinski M. Human cooperation in the simultaneous and the alternating Prisoner's Dilemma: Pavlov versus Generous Tit-for-Tat. Proceedings of the National Academy of Sciences. 1996;93(7):2686–2689.

36. Kümmerli R, Colliard C, Fiechter N, Petitpierre B, Russier F, Keller L. Human cooperation in social dilemmas: comparing the Snowdrift game with the Prisoner's Dilemma. Proceedings of the Royal Society of London B: Biological Sciences. 2007;274(1628):2965–2970.

37. Plotnik JM, Lair R, Suphachoksahakun W, de Waal FBM. Elephants know when they need a helping trunk in a cooperative task. Proceedings of the National Academy of Sciences. 2011;108(12):5116–5121. doi:10.1073/pnas.1101765108.