High Level of Mutual Cooperation using Iterated Prisoner's Dilemma

1 Introduction

Ever since Darwin's theory of evolution was proposed, cooperative traits in animals and humans were challenging to explain. In order to do so, Darwin's theory incorporates two novel kinds of extensions: the genetically kinship theory that Hamilton grounded mathematically (Hamilton,1964) and the reciprocal altruism theory (Trivers, 1971). Both appended theories are capable to explain social altruism behavior through natural selection. In evolutionary biology, reciprocal altruism is a behavior whereby an organism performs a costly act that benefits another organism that if this behavior is reciprocate by others organism later time, they increase a long-time reward or their fitness.

The kinship theory that after landed on kin Selection and Maynar Smith use in first time (Smith,64), talk about, on knowledge of the genetic relationships of the organism involved, how altruistic behavior between closely related individuals can be selected though natural selection. This theory does not considers the altruistic act among distantly related organisms whereas Triver's theroy does. The term "reciprocal altruism" was first used by Trivers to refer to this type of behavior and therefore explain how selection favors altruistic behaviors in the long run when there is reciprocity in the interaction. Some of the most relevant and reliable examples about this type of cooperation are the Wilkinson studies in reciprocal food sharing behavior in vampires bats (Desmodus rotundus), (Wilkinson, 1984). In these experiments, animals share a part of the harvested food only to a partner that previously shared with them.

Different conditions must be fulfilled to ensure that reciprocally altruistic behaviors will be selected. Wilkinson makes a clear summary of these conditions :

> (1) the behavior must reduce a donor's fitness relative to the selfish alternative, (2) the fitness of the recipient must be elevated relative to a non-recipient, (3) performance of the behavior must not depend on receipt of an immediate benefit, (4) a mechanism for detecting individuals who receive benefits but never pay altruistic costs has to exist, and (5) a large but indefinite number of opportunities to exchange aid must exist within each individual's lifetime (Wilkinson, 1987).

Research on cooperative behaviors between non-related organisms has gained a new impulse since Trivers connected reciprocal altruism behaviors with the famous mathematical Prisoner's Dilemma (PD) (originally developped by Merrill Flood and Melvin Dresher 1950). The PD game is one class of 2x2 game that involves two players who must choose between two options, generally called cooperation and defection. The size of the reward delivered is established according both player's choice. If both chose the cooperation option both get payed a certain amount R (reward), and if both chose defection they both get payed a smaller reward P (punishment) than R. However if one cooperates and the second player defects, the first one gets payed S (sucker) and other receives the best amount T (temptation). In iterated Prisoner's Dilemma (iPD), the Evolutionary Stable Strategies (ESS)(von Neumann and Morgenstern, 1944; Nash, 1949) predict what behavior (strategy) is likely to occur, for example if a ESS is adopt by the population, in such a way, no minority using a another strategy can invade (Smith, 1974). When the 2x2 PD is played only for one time the best ESS is defect strategy. Nevertheless, when the pay-off matrix employed meets $T > R > P > S$ and $2R > T + P$ and the game is played an indefinite numbers of times the best strategy is still mutual defection, but arise a strategy (in game theory is called *Pareto Efficient*) in which if both adopt it, both earn the big long-term reward. But if one stop reciprocity behavior the best option change to defect. Axelrod and Hamilton (1981) showed that some organism's symbiosis can be understood through the reciprocal altruism's model and if organisms can remember the outcome of at least one previous interaction and they are able to recognize different partners, then the strategies situation includes a much richer set of possibilities. They present a ESS for iPD that combines robustness and stability with initial viability, called TIT FOR TAT. This strategy arised as the winner strategy, submitted by Anatol Rapoport, in the Robert Axelrod's computer tournament (Axelrod, 1984), because it can survive invasions from other strategies. The highly simple strategy consists in cooperating on the first move and then doing whatever the opponent did on the preceding move.

Thus, many experimenters have tried to understand different aspects of reciprocal altruism behavior in animals and whether non-human animals with less cognitive abilities can solve iPD (Mendres & Waal,2000; Waal, 2000; Hause & colleagues, 2003; Taborsky, 2007, 2008, 2011 and 2015, Wood, 2016; Kalenscher, 2015). Green, Price and Hamburger (1995) assessed the iPD game on pigeons and observed that birds are very impulsive and prefer small immediate rewards rather than big, long-term, delayed rewards. Stephens (1995) trained blue jays (Cyanocitta cristata) using four different pay-off matrices, where one of them was the iPD matrix, with a special dual operant box. This study found little cooperation in PD treatment and this findings suggest that blue jays don't cooperate when immediate benefit is available (defect only), even if a long-term larger benefit may exist. Then Stephen and colleages (2002 and 2005) inspired by the low levels of cooperation observed (Gardner et al., 1984; Clements and Stephens, 1995; Flood et al., 1983; Green et al., 1995) proposed assess over iPD frame to the effect of a pay-off accumulation and temporal clumping treatment to iPD game using blue jays in an apparatus that consisted of side by side V-shaped compartments. They found that combining both accumulation and clumping treatment birds showed a level of cooperation higher than previous experiments of iPD. The Stephens' studies is timely because it forces behavioral ecologists, and other behavioral researcher, not only to rethink the potential importance of temporal discounting, but also to address a number of other issues (Mesterton-Gibbons, 2002). Indeed, we can explain from the point of view of Operant Conditioning, the improve on cooperation levels in Stephen's work by adding a new conditioning stimulus to the game, transparent food storage, and in this way the game become more easy to predict where is the maximum reward in the game. Other issue to keep in mind is the kind of reinforcement and punishment used in the experiments, such as a iPD study with rats (Viana et al., 2010) that used a mimicking a tit-for-tat strategy on the opponent in a double T-maze chamber with positive reinforcement as pellets and punishment as tails pinch. They achieved an apparent moderately high level of cooperation, but in ours further analyzed of their Markov chain diagram shown that the strategy performed by rats was more selfish than cooperative because it adopted an alternating decision's rule among sucker and temptation outcomes, basically they flees from punishment(undisclosed data). (siento la necesidad de tener que hacer algo oficial para que sea creible lo último que decimos).

Regarding the about, we ask ourselves Why iPD-based studies haven't found high levels of reciprocity? In the aforementioned previous studies, animals don't learn iPD probably due to the fact that the reinforcement and punishment are inadequate and stimulus contingency not match with the natural acquisition capability of animals, i.e. the rat maybe has the ability to achieve an approximated optimal solution but for example the length of time used to make the contingency is longer than what rats can keep in mind or maybe they don't realized the real difference within outcomes in the pay-off matrix.

The present study was conduct to assess reciprocal altruism behaviors using a particular structure that combine iPD matrix pay-off and timeout punishment with a tit for tat opponent's strategy. We assess a iterated PD in two phase: first normal iPD and then, in a second phase, a reversion iPD using the previous animals set. We found that this combination achieve high level of reciprocal altruism behavior in rats.

2 Material and Method

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.

2.1 Subjects and Housing

2.1.1 Subject

Two groups of male Long-Evans rats (300–330 g) of two months of age were provided by the IBYME-CONICET. Twelves males were subjects and six (Nstooge=6) identical rats were opponents. At weaning time all subjects were housed in groups of two rats per cage to allow social interaction, whereas each stooge was housed in single individual cages. We had 12 stainless-steel cages altogether with sawdust as bedding and metal lids. All rats were food restricted to maintain animals at between 90-95% or 80-85% of free feeding body weight for subjects and stooges respectively. All animals were kept in a well-ventilated, temperature-controlled room (22 $\pm$ 2 °C) with a 12/12 h light/dark cycle (lights on at 8 am). Tap water was available ad libitum. To ensure that animals obtained sufficient food to survive, we provided supplementary food (at 22hs) for any rats that obtained the average amount of pellet to maintain body weight.

The subject was named in successive order: 1A, 2A, 3A, 4A, 5A, 6A, 7A, 8A, 9A, 10A, 3B, 4B.

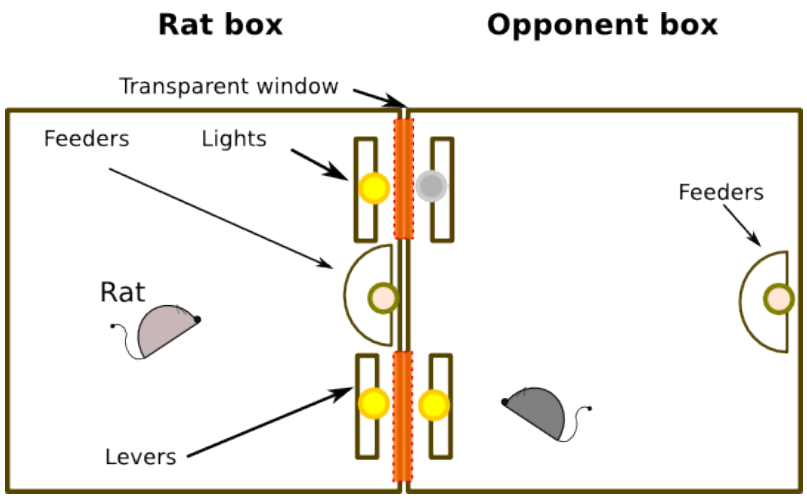


Fig. 1: Dual operant chamber equipped with levers, feeders, lights and windows. The size of the chamber was 75cm width and 30cm depth and 35cm height.

2.1.2 Housing

All behavioral procedures were performed during the light phase of the light/dark cycle, using a standard operant chambers (MED associates Inc., USA) with Med PC IV software suit (Product SOF-735) and PCI Operating Package for up to Eight Chambers (Product MED-SYST-8) equipped with control of multiple devices SmartCtrl (Product DIG-716B) and pedestal mount pellet dispenser for rat, 45mg (Product ENV-203-45) and standard operant chambers (Product ENV 008). White noise with a flat power spectral density was used to reduce sensibility to ambient noise. The training was perform on a single standard operant chamber equipped with tree retractable levers and tree stimulus light over each lever, and in the back wall placed a illuminated feeder. The experiment was conducted in special ad hoc dual chamber. We placed two Med association standard operant chambers facing each other in such manner that each rat could make olfactory and eye contact through metal windows (see fig. 1). Each standard chamber was equipped with: two not retractable levers at the sides of the same wall and two stimulus light were placed over each levers, and in the center put a illuminated feeder. The boxes were faced by the wall that has levers and stimulus light. To allow the contact among rats we placed a aluminum rectangular windows below each levers. The window's height were such that the lever's height was 80% of maximum height of the forepaws while rearing (F. Cabrera et al., 2013). The subject's feeder placed in same wall and the opponent's feeder placed in back wall. Each feeders were equipped with a stimulus light that turn on when foods is coming.

2.2 Procedures

During the whole duration of experiment, every actor were trained for 2 session per day regardless of treatment (handing, habituation, training and iPD). Each typical trial began in the darkness during 5 second inter-trial interval(ITI) and then the stimulus light were illuminated for either 45 second or until a lever was press. Before deliver any reward the feeder light was 1 second turn on. The standard experiment session had 30 trials.

2.2.1 Training procedure

Handing At weaning time when the animals were moved to housing room started a handing procedure to decrease the stress by experimenter manipulation and finished when animals had 60 days old.

Habituation The animals were habituated to the single and double operant chamber for a days in session of 3 minute. In this treatment, there were either no stimulus-reward contingencies or reward, only a back-light that marked the beginning and the end of sessions.

Magazine To learn the place in which the food is given, the rat were exposed for a days to a "*Magazine*" procedure in the single chamber. It consisted in variable ratio reinforcement scheduler.

Tab. 1: Prisioner's Dilemma Pay-off Matrix

	Opponent choice	
	C	D
Subject choice		
C	1 pellet	8" delay
D	2 pellets	4" delay

Shaping The pressing lever training on rats was done through a successive approximations procedure, "*Shaping*" (Mazur, 1994, page 122) on a single chamber. The chamber was equipped as described above and was added a external lever. In this training, the external lever was manipulated by the experimenter to give reward as the rat go to the goal and finally when press lever. When the rat learned to press the lever after light onset, the operator left to gave reward with his external lever. It training finished when the rat made the task for at least two session. Trials procedure: each trial began when the center stimulus light was turned on and the lever below it was ejected, then if either inner or outer levers wasn't pressed and 45 second had elapsed, the trials finished and not reward was given. But nevertheless, if either inner or outer levers was pressed, the light turn-off and the feeder's light was turned on, one second later a 45mg pellet is automatically dispensed, when 5 second elapsed all light were turn off and the lever was retracted, finally 5 second ITI started. The procedure was performed until the rats met task goal or up to 3 days. When the rats reached the criteria (80% trials with reward), we exposed the rat to the same task but now the center lever during the session never more was retracted, the lever ejected at first trials and retracted at the end of last trial. In this phase, when the rat pressed lever on ITI was punished with 2 second delays that was added to ITI. This procedure went on for tree days or until pressing rate was below one per trials.

Basic toggle task for subject The subject rats group was exposed to a training called "toggle 8x8" in dual iPD chamber. The purposed of the training was to incorporate two lever and balance their lever's preference. The experiment was performed for 2 days in which at first day a rat use chamber 1 and at day 2 used chamber 2. This last allow not placed preference. The chamber communication windows were closed and any contact was suppressed. Each rat had two lever option and two light stimulus. When session started the first to eighth trials only one pair stimulus light and lever worked and the next 8 trials(9-16) the another pair light-lever worked. What pair light lever started was randomly and in the second day was maintained in the same way. The trials procedure was the same as that used in shaping treatment.

Mimic tit for tat for opponent The opponent learned to develop a mimic tit for tat strategy by to learn a basic operant task. The opponent rat group was exposed to a fixed ratio (FR) schedule in dual operant chamber and the pair light-lever was randomly assigned in each trials along the session. In this training, the rat must to press the lever that had light turned on. The other lever did not work during the trials and on the next trials the light-lever pair that worked was pseudo-randomly assigned. Pseudo-randomly means that the assigned was random but not allowed the same assigned for more than four time. The training was developed during five days and the FR schedule was increased step by step: day 1^{o} the FR=2 and day 2^{o} to 3^{o} the FR=3, then day 4^{o} the FR=4 and finally day 5^{o} the FR=5. If the rat presses the lever on ITI time, it received 2 second as punishment that was added to ITI.

2.2.2 Iterated Prisoner's Dilemma Procedure

The main experiment was designed to study the reciprocal altruism in a *iterated Prisoner's Dilemma* game(iPD). We used a pay-off matrix in which the iPD states(outcomes) meets that $T > R > P > S$ (Temptation, Reward, Punishment and Sucker respectability) and also $4R > T + 3P$ and $4R > 2T + 2S$. The T and R outcomes were positive reinforcement and P and S was punishment. The pay-off matrix is shown in table 1. Thus, if both choose cooperation option (C) their pay-off was 1 pellets for each other, If opponent choose C and the subject choose C or defection option (D), the subject received one pellet or 2 pellet, respectability. When the Opponent choose D and the subject choose C or D, the subject received 8" or 4" delay of punishment and no pellet in this trial. The opponent's pay-off was alway 1 pellets if either both make responses or the opponent meets the Fixed Ratio scheduler. The FR scheduler was used to make a time interval in which opponent stays opposite the selected lever and window. The opponent strategy was conducted by turn on a light stimulus over the target lever, e.i. the opponent didn't choose lever freely, it had to develop a Fixed-ratio scheduler on the side where light was turn on to be paid. The amount pellets preference was previously tested on discrimination treatment and the rat shown more preferences for 2 pellet than 1 pellet.

In iPD experiment used a dual chamber without neither center lever and center light. The windows cover was removed. The rats had only two lever option and the contact between chamber was restored. There was one chamber for the subject and one for the opponent fixed during all the experiment, see the figure 1.

The main experiment was conducted for at least 10 days and 2 session per day. The session consisted in 30 trials. For individual subject, the experiment was stopped, if the average behavior didn't change along ten consecutive sessions. To some subject was exposed more than 20 days when rise its behavior tendency.

2.2.3 Typical *i*PD Trial Structure

The sequence of events within a single trial was as follows. (1) The stimulus light was turned on stating that trail was started. The focal rats had both lights turned on (this was done throughout the iPD experiment) and the opponent only one for the tit for tat strategy. (2) After both rats done their response, the light was turned off and the rewards computed. (3) The feeder's light was lighting one second before a reward was delivered. If the opponent pressed the lever always received one pellet irrespective the focal rat's choice. If punishment was assigned, the focal rat feeder's light remained off and the punishment time started, but the opponent's feed was lighted and its one pellet reward was delivered. (5) After either five second eating time expired or punishment time complete, all light was turned off. Then, the focal and opponent wait the end of fixed ITI time in the darkness. The next trials started.

2.3 Statistical Analysis

Al statistical analysis were performed using statistic library from octave open source software.

2.3.1 Stability of cooperation

To analysis the choice stability of rats in the iterated game, we used the last 10 session as a pool of data due to the rat's behaviors had reached a plateau. The mean of cooperation per rat was obtained by count the number of times a rat chose the cooperation lever per session over the last ten session.

2.3.2 Probability of cooperation after outcomes

The animals probability to cooperate after each outcomes was assess using a chi square test. The probability of cooperation given specific outcomes was compared from chance, 0.5. We performed a Chi-square goodness of fit test with bonferroni corrected value of 0.05/n.

2.3.3 Rate of difference outcomes

We used Friedman's ANOVA test and multiple post-hoc pairwise comparisons using the Nemenyi's procedure/Two-tailed tests were performed, with the Bonferroni corrected a value of 0.0125.

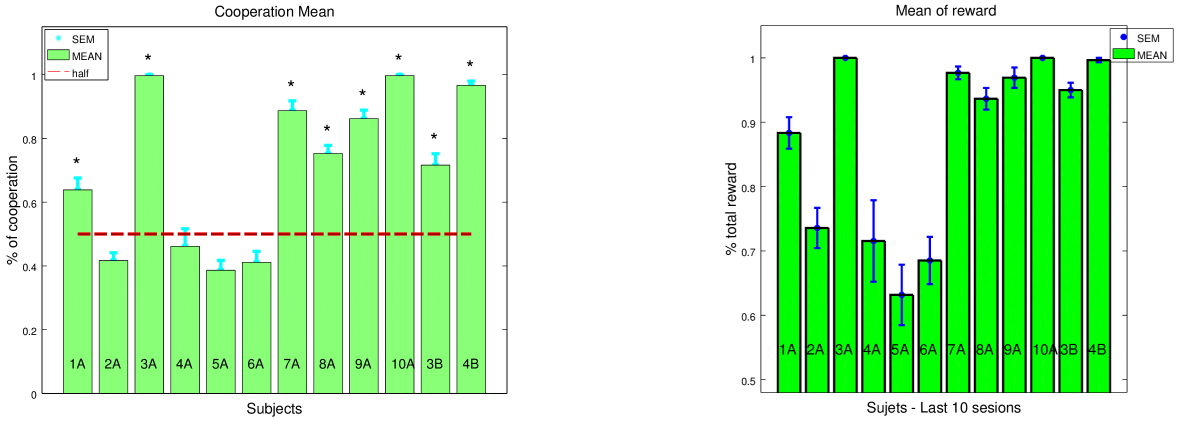
2.3.4 Markov chain diagram

Reciprocal altruism has been widely tested by iPD, but the individual's decision rules can be described by a transition vector t, r, p, s that reflects the probability of cooperation when the previous trials resulted in outcomes of R(reward), T(temptation), S(sucker) or P(punishment), respectively (Stevens and Stephens,2002). If every component of this vector is 0.5, the agent's decision rule is random irrespective of the last outcome. Also we used diagram of probability of transition between outcome to computed a complete making decision rule of each rats.

3 Results

3.1 First Phase: Iterated PD with opponent

The experiment result are show below. To assessing whether a subject has ability to solve the iPD game, we first need to know whether either the subjects adopt some strategy or play the game in random mode. The total session per rats was: (1A/23; 2A/33; 3A/23; 4A/23; 5A/23; 6A/31; 7A/23; 8A/50; 9A/31; 10A/29; 3B/33; 4B/23). The full data set used had difference length, because after recorded 20 sessions on each rats the experiment stop when the means of cooperation reached a plateau at least in the last ten sessions, *i.e.*, the end of experiment for each rat was extended beyond 20 session in order to the rat was progressively finding some strategy. We compute the rat behaviors from last ten



(a) Cooperation mean

(b) Reward mean

Fig. 2: 2a)**Mean and χ^2test**: The rats with 95% of free feeding body weight played under a matrix pay-off T=2, R=1,P=4"delay, S=8"delay against a TFT opponent. We show means of the numbers of times rats chose the cooperate option ($mean \pm s.e.m.$). The Asterisk denote when strategy adopted by a rat had significant difference from chance strategy (χ^2 goodness of fit test with bonferroni corrected, p>0.125). The rats without significant difference did not surpass 0.5 probability of cooperation. 2b) **Means of reward.** Bar line shows the mean of obtained reward per session over the last 10 session ($mean \pm s.e.m.$). The rats with random strategy (not chi-square significant) obtained the lowest level of reward, below 75% of total reward.

session of the full data set. A strategy was defined by a transition vector that reflects the probability of cooperation when the previous trials resulted in outcomes of R, T, S or P, respectively ($v = [p(c|t), p(c|r), p(c|p), p(c|s)]$). The rat that maintained a random strategy beyond 20 session was discarded. To test if rat strategy differs from chance strategy (v=[.5 .5 .5 .5]), we perform a χ^2 goodness of fit test on last ten session, see **??**. After compute the parameter $\chi^2$with bonferroni correction for each rat we obtained two group of subject. Those who adopted some strategy and those who did not. The figure 2a shows the mean of cooperation for all rats and the asterisks marks denoted subject that had significant difference respect the chance. The subject 2A, 4A, 5A, 6A were removed from data pool because they not had a fixed strategy (See Supplementary materials 6.2 the graph of markov chain with probabilities of transition). Since the table **??** we can see that there is a group of 8 rats with had a fixed no random strategy (blue color text).

In figure 2b we show the mean of reward per subject. The rats with not random strategies got many more reward than the removed group. But nevertheless, we show that the iPD game gives to the random strategy group a amount that between 60% to 70% of total reward.

Theoretically, when a subject play iPD against TFT opponent, its the best strategy is All C and not TFT, because TFT against TFT when one defect both remain in this state and never change. Thus, if the rats make a wrong choice, defect, then its probability to cooperate must be high. Analyzing the not random strategy groups behaviors, we found that each rats had all components of strategic vector markedly above 0.5 suggesting that the rat's behaviors trend to be highly cooperative, see table **??**. We calculated the mean and s.e.m. of cooperation choice in the groups, 0.852 ± 0.0482 and the average strategy, $p(c|T) = 0.794$, $p(c|R) = 0.856$, $p(c|P) = 0.801$, $p(c|S) = 0.890$, tested by χ^2 goodness of fit with theory frequency 0.5, figure 4a. In figure 4b, the rate outcomes suggest that R outcomes(i.e. both subject and opponent chose cooperation lever option in the trial) had very high incidence in the experiment over the last 10 sessions into the group. We found that exist significant difference between outcomes rate, a Friedman's ANOVA test was performed with bonferroni correction ($\alpha = 0.0125, p > 3.33e-5; \chi^2 = 23.4$) and Multiple pairwise comparison using Nemeyi's procedure point out that all outcomes levels are different.

In the figure 3 we show the means of cooperation choice per session per rats and the mean of the group with non-random strategy per sessions. Since the number of sessions was different along the rats, we aligned the last session of each rats and make a pooled of data from the 23 last sessions. The means and standard error of the means of last ten session, that represent the plateau of the curve, was a$mean \pm sem = 0.853 \pm 0.0681$.

In figure 4c and 4d we show diagrams of transition probability given the last outcomes, respect four and two state. The diagrams represent the probability to cooperation (blue arrow) or defect lever (red arrow) given the actual state (T, R, P, S or C, D). The arrow width is proportional to the probability of transition, i.e., cooperate or not cooperate. In both diagrams the big blue arrow predominate and these indicate a strong trend to mutual

Tab. 2: We show the mean of cooperation and the probabilities of cooperation given each outcomes (outcomes: T, R, P and S). The χ^2goodness of fit with bonferroni correction was performed used a theorical frecuency of 0.5. The significance value show the subject that developed a specific strategy. The subject underlined and in blue text color had significant different respect random strategy.

	Mean	Strategies					
Subject	cooperation	$p(c\|T)$	$p(c\|R)$	$p(c\|P)$	$p(c\|S)$	$\chi^2_{bonferroni}$	$p < 0.0125$
1A	0.671	0.599	0.721	0.664	0.687	17.15	$6.583e^{-4}$
2A	0.417	0.469	0.408	0.446	0.333	8.02	0.0456
3A	0.997	1	0.996	0	1	199.30	0.0000
4A	0.461	0.5	0.625	0.348	0.403	9.60	0.0223
5A	0.386	0.359	0.408	0.351	0.459	10.42	0.0152
6A	0.411	0.493	0.37	0.381	0.394	8.45	0.0375
7A	0.886	0.778	0.904	0.857	0.885	103.23	0.0000
8A	0.762	0.638	0.75	0.682	0.864	49.38	$1.081e^{-10}$
9A	0.862	0.781	0.851	0.778	1	105.91	0.0000
10A	0.997	1	0.996	0	1	199.30	0.0000
3B	0.715	0.742	0.687	0.842	0.705	50.51	$6.217e^{-11}$
4B	0.966	0.889	0.97	1	1	174.45	0.0000

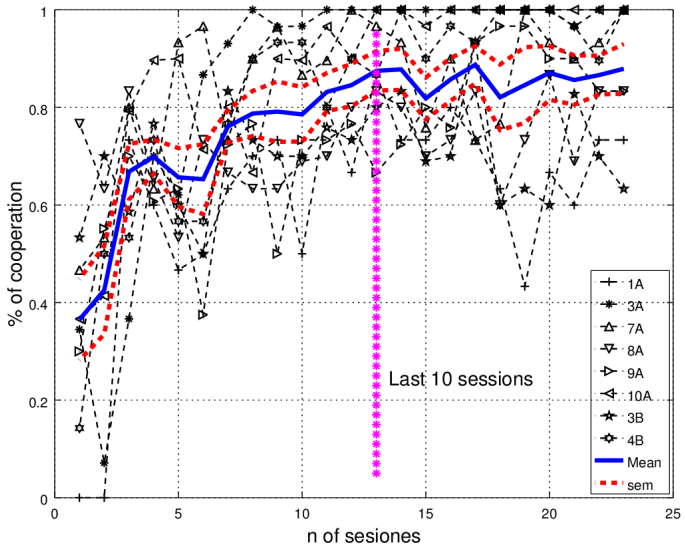
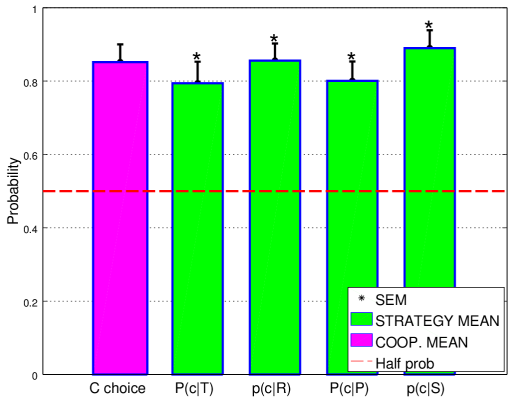
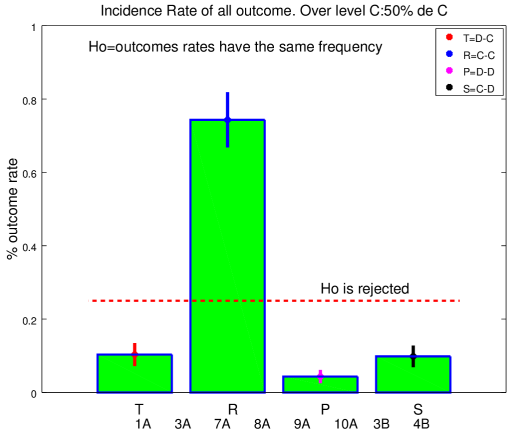


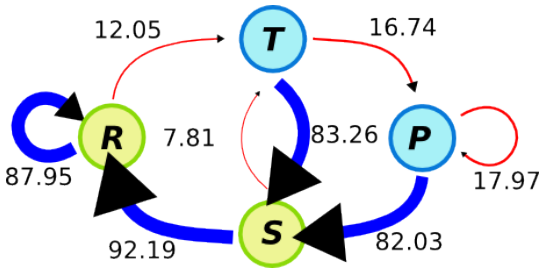
Fig. 3: Evolution of cooperation choice from cooperator group data set with the last 23 sessions. The blue and continuous and thicker line is the means per session of the group and the dotter line is the standard error of the mean ($mean \pm sem = 0.853 \pm 0.0681$over the las ten sessions). The vertical dotter line mark the pool of data that was used to analyse the strategies adopted by the rats.



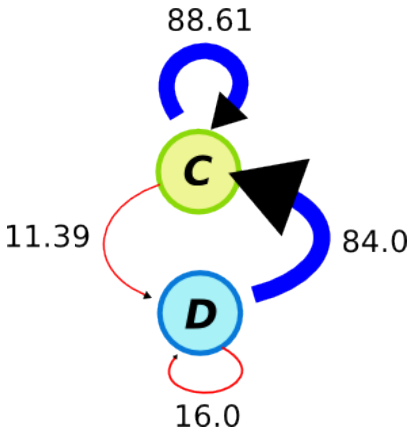
(a) Mean of cooperation and means of transition vector



(b) Outcomes rates over the last ten sessions.



(c) Markov chain graph of four state



(d) Markov chain graph of two state

Fig. 4: 4a) Mean of cooperation over not random strategy rats (8 rats) was performed and the measured is shown on the first magenta bar ($mean = 0.852$ and $s.e.m. = \pm 0.0482$). The next four green bars represent the mean strategy over the group and each bar agree with the transition vector,$v = [0.794, 0.856, 0.801, 0.890]$. The asterisk indicates significant difference respect the half probability of chose cooperation lever and we tested by χ^2 goodness of fit with theory frequency 0.5. 4b) The bar graph shows the outcomes rates and suggest that R outcomes had very high incidence in the experiment over the last 10 sessions. . We found significant difference between outcomes (Friedman's ANOVA test was performed with bonferroni correction ($\alpha = 0.0125, p > 3.3e-5; \chi^2 = 23.4$) and Multiple pairwise comparison's using Nemeyi's procedure). 4c and 4d) Graph of Markov chain of four and two state, in which the arrow represent the transition probability between outcomes. A blue arrow signalized when the subject chose cooperate lever given the las outcome and a red arrow signalized when it chose defect lever. The words mean temptation (T), mutual cooperation (R), punishment (P) and sucker (S). Pay atemption that the arrows width are proportional to the transition probability.

cooperate. This diagrams point out that the rats trend to stay on mutual cooperation or quickly come back to this state. In all cases, the probability to cooperate was over 80%

The reciprocate strategy is mainly defined by the probability of mutual cooperation. Thus we made a graph of Mutual cooperations versus cooperation and saw that exist a directly proportional relationship between them, if mutual cooperate does down then the cooperate choice goes down, the correlation coefficient was 0.996(*Pearson coefficient*). It result point out that both variable are related. See figure 5a. The color bar represent average accumulated punishment (in second) that each subject obtain per sessions in the last ten session and also is drive by same relationship respect both mutual cooperation or cooperation choice, the correlation coefficient was -0.998 and -0.935 respectively. When cooperation up, the punishment down.

In the figure, the square point represent theoretical simulated behaviors and was used to compare the real average rat behavior respect to these repetitive simulated agent. We created the follow behaviors: *1º)"switch CD"* in which the theoretical subject always alternated between cooperate (C) lever and defect (D) lever; 2º) "switch CCDD" is similar to the previous but the subject choose consecutively 2 time the same lever and then change to the other lever and repeat the action; 3º) "switch 3C3D" the subject choose consecutively 3 time the same lever and then change; 4º) "half C" the subject choose the same lever until the middle of session, then change and doesn't change at the end; 5º) "switch CCD" in which choose 2 time C lever and then choose only a time the D lever and go back to the previous C lever and repeat the action. 6º) "switch CCCD" is similar to the previous but the subject choose 3 time C lever. These simulations allow to visualize specific behavior and make assumptions about rats behaviors.

All rats got the high level of reward, over 80%, despite of the dispersion of mutual cooperation data, fig. 5c. Since the simulated agent (filled square) are used to understand the kind of strategy used by the rats, we observe that the rats after make a defect choice trend to learned that they have to back to cooperate choice. The rats of the group with lowest mutual cooperation level are near the simulated agent with "Switch CCD" and "Switch CCCD" behaviors.

Relating rewards with timeout punishment level we develop a coefficient that named *Preference*, and it can help to understand the preference for choose lever C, figure 5d. The red circles depict the simulated agent and blue circle the rats. The "CD" symbolize "switch CD" simulated agent that develop a *alternate* strategy. To the left side of "CD" are place the simulated agent and rats that have a cooperation level less than 50%. On the right side of "CD" represent the strategies in which choose C more than one time, consecutively. Holding the high level of cooperation no rats are close to the "CD" and all rats place on the right side of it. Since this picture we can infer that the behaviors of rats trend to be such that at least cooperate two time or more and then defect one time and quickly back to cooperate.

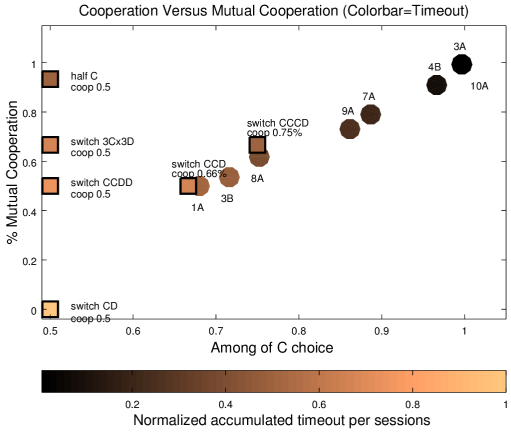
3.1.1 Second Phase: reversion

The second phase experiment was developed to evaluate if the behaviors exhibit by the rats was either learned through the experience into the iPD game or was simply the choice of one lever, cooperation, by chance. If was by chance and now levers position change, the rats is going to choose same lever and will not modify her behavior. We used four rats that got the best score in the previous experiment. In this phase, the levers order was inverted, i.e. if in previous experiment the lever meant cooperate, it became to defect lever. We procedures in the same way as in the first experiment. The sessions duration per rat were: 3A/52; 7A/52; 9A/52; 10A/24;. The means of cooperation from the last 10 session is showed in the figure 6a and the asterisk denote the rats with a some fix strategy that differs to the random strategy.

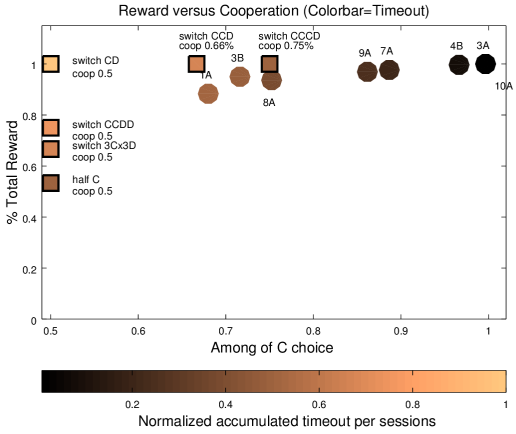
All rats reached the level of reward close to 1(i.e.100%, see fig. 6a) and the total punishment under 0.4 (40%, see fig. 6c). In the Table 3 is showed the means of cooperation and strategy vectors per rat, all subject overpass the 0.5(50%) probability and the strategy vectors are very different to chance. Only the 10A rat has a strategy that need a additional explanation about because the probabilities to cooperate given P or T outcomes are very low. But If we look at the frequency .

,

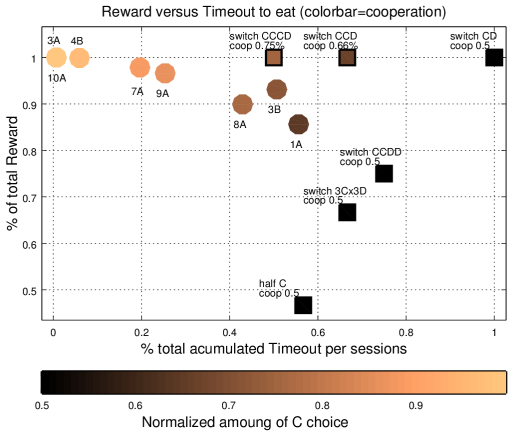
,



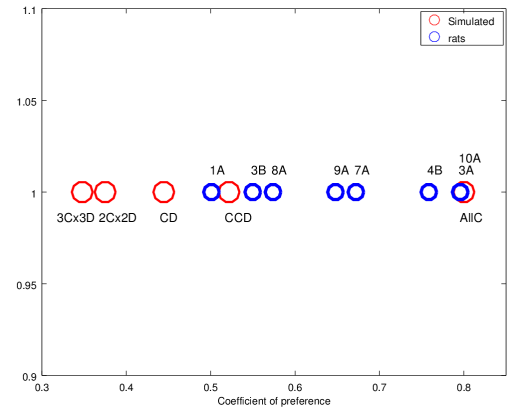
(a) Mutual Cooperation versus cooperation choice



(b) Total reward versus cooperation choice

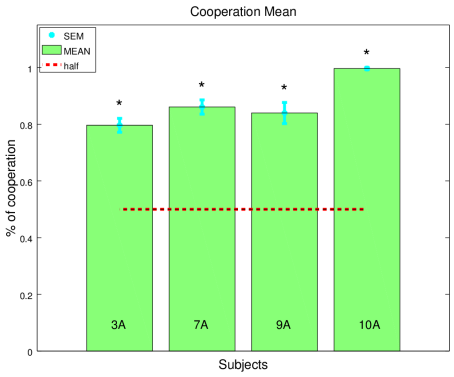


(c) Total Reward versus accumulated timeout

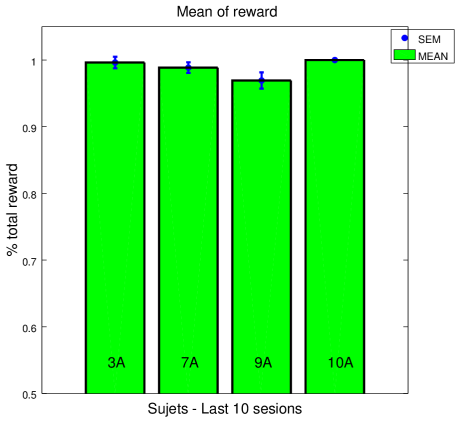


(d) Preference Coefficient.

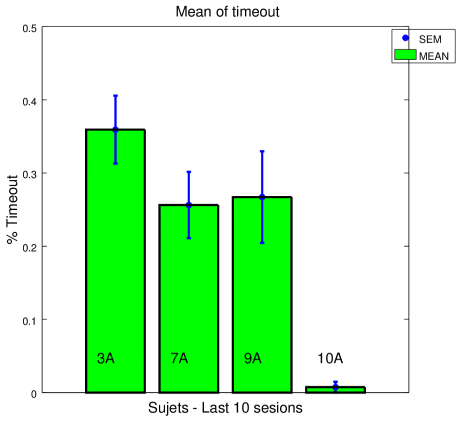
Fig. 5: There experiment's result was computed from last ten session pool data set. 5a) The graph of Mutual cooperations versus cooperation show that the subject (filled circle) had a directly proportional relationship, mutual cooperate less when cooperate less. The color bar represent average accumulated punishment (second) that each subject obtain per sessions and also is drive by same relationship respect . The square represent theoretical behaviors. 5b) The graph show the percentage of cooperation versus total reward. All animals obtain more that 80% of total reward per session. The simulated agent (filled square) help to understand the strategies used by animals. 5c) The figure show the relation between rewards and timeout punishment, the color bar point out the level cooperation choice. Since the graph we observe that the highest reward corresponds to the subject with less accumulated timeout punishment and they also have the highest cooperation choice level. 5d) The Coefficient of preference is the quotient between reward and punishment that each rat got as result to used one strategy. the blue circle are the coefficient of preference per rat and the size are the level of cooperation. The red circle are the coefficient for the simulated rats. All rats has a strategy more cooperative than the simulated rats with alternated "CD" strategy.



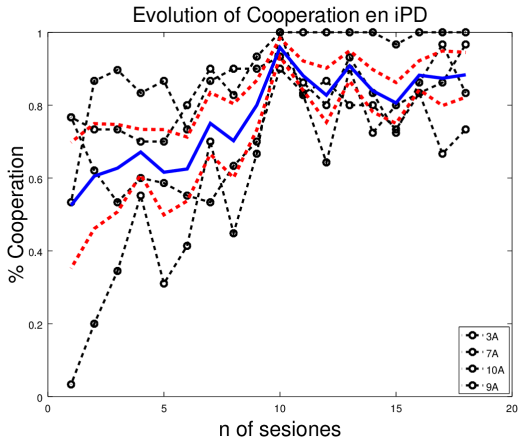
(a) Cooperation mean



(b) Reward mean



(c) Timeout mean

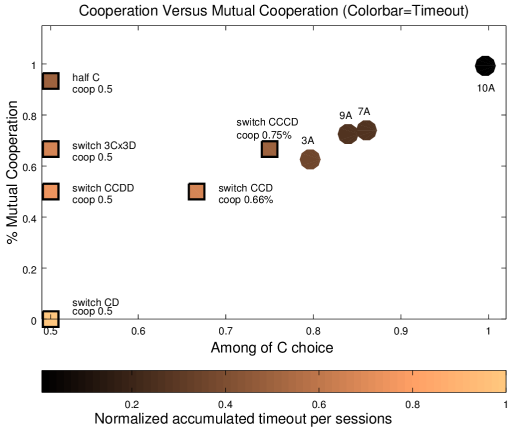


(d) Evolution of cooperation

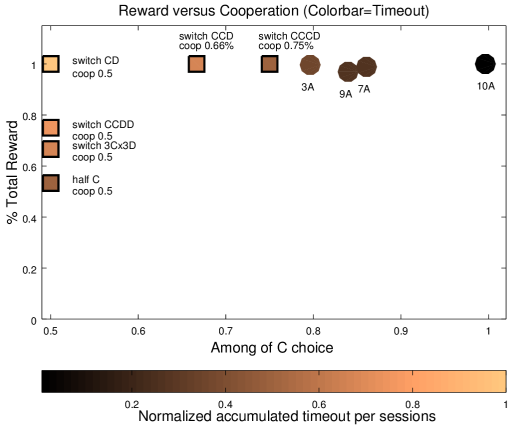
Fig. 6: 6a)**Mean and χ^2test**: The rats with 95% of free feeding body weight played under a matrix pay-off T=2, R=1,P=4"delay, S=8"delay against a TFT opponent. We show means of the numbers of times rats chose the cooperate option ($mean \pm s.e.m.$). The Asterix denote when the rat adopted a strategy that had significant difference from chance strategy (χ^2 goodness of fit test with bonferroni corrected, p>0.125). The rats without significant difference did not surpass 0.5 probability of cooperation. 6b) **Means of reward.** Bar line shows the mean of obtained reward per session over the last 10 session ($mean \pm s.e.m.$). The rats with random strategy (not chi-square significant) obtained the lowest level of reward, below 75% of total reward. 6c) Means of Timeout is the punishment that each rat got by develop a learned strategy. 6d) Evolution of cooperation choice from a data set with the last 18 sessions. The blue and continuous and thicker line is the means per session of the group and the dotter line is the standard error of the mean ($mean \pm sem = [0.866 \pm 0.015]$over the las ten sessions). The vertical dotter line mark the pool of data that was used to analyse the strategies adopted by the rats.

Tab. 3: We show the mean of cooperation and the probabilities of cooperation given each outcomes. The χ^2goodness of fit with bonferroni correction was performed used a theorical frecuency of 0.5. The significance value show the subject that developed a specific strategy. The subject underlined and in blue text color had significant different respect random strategy.

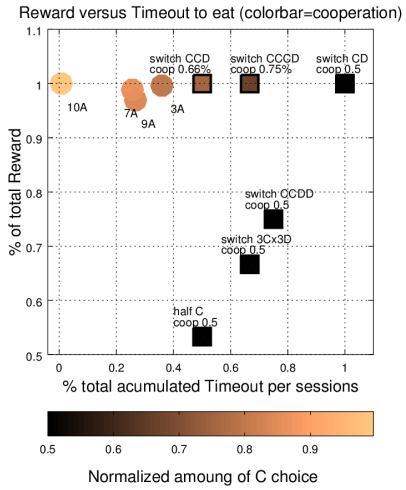
Subject	Mean cooperation	Strategies $p(c\|T)$	$p(c\|R)$	$p(c\|P)$	$p(c\|S)$	$\chi^2_{bonferroni}$	$p < 0.0125$
3A	0.787	0.868	0.786	1	0.769	113.336	0.0000
7A	0.865	0.885	0.891	1	0.714	117.942	0.0000
9A	0.849	0.697	0.877	1	0.781	98.041	0.0000
10A	0.963	0.333	1	0.250	0.660	200.000	0.0000



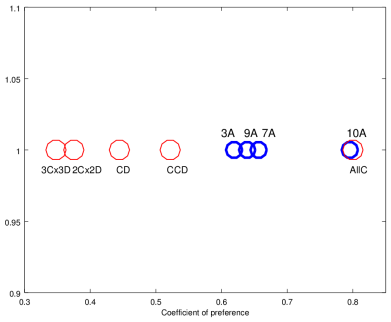
(a) Mutual Cooperation versus cooperation choice



(b) Total reward versus cooperation choice



(c) Total Reward versus accumulated timeout



(d) Preference Coefficient.

Fig. 8: There experiment's result was computed from last ten session pool data set. 8a) The graph of Mutual cooperations versus cooperation show that the subject (filled circle) had a directly proportional relationship, mutual cooperate less when cooperate less. The color bar represent average accumulated punishment (second) that each subject obtain per sessions and also is drive by same relationship respect . The square represent theoretical behaviors. 8b) The graph show the percentage of cooperation versus total reward. All animals obtain more that 80% of total reward per session. The simulated agent (filled square) help to understand the strategies used by animals. 8c) The figure show the relation between rewards and timeout punishment, the color bar point out the level cooperation choice. Since the graph we observe that the highest reward corresponds to the subject with less accumulated timeout punishment and they also have the highest cooperation choice level. 8d) The Coefficient of preference is the quotient between reward and punishment that each rat got as result to used one strategy. the blue circle are the coefficient of preference per rat and the size are the level of cooperation. The red circle are the coefficient for the simulated rats. All rats has a strategy more cooperative than the simulated rats with alternated "CD" strategy.

4 Discussion and conclusions

We found a high lever of cooperation on all rats that develop some strategy, because the mean of the group was over 85%, fig. 3, and the transition vector depict the probability to cooperation given each outcome uphold the trend(the mean of each component is on or over 80%). All animals learned that the best option wherever they are was cooperate or comeback to cooperate.

In previous laboratory studies of iPD have tested on birds and rats, amongst anothers (Waal, 2000; Hause, 2003;St-Pierre et al., 2009), in diverse condition and the best results were achieved by Stephens with birds (Stephens, 2002) and Moita with rats (Moita, 2010). However, Danchin and colleages (2006) criticized the Steven's experiment arguing that each accumulation block the iPD payoff matrix becomes stag hung matrix (whereby the temptation outcome, T, leave to be the best reward, R>T≥P>S) because the bird quantify the effective amount after four trials. Further, they using a mutualist matrix to boost the cooperation level before each iPD session. About of Moita experiment we can conclude that low cooperation level, *near below 60%*, was as a result of the type of punishment used, because the rats adopted a behavior in order to avoid the aversive tails pinch. For this reason, we didn't use either any booster matrix or accumulation treatment and implement a timeout as a punishment for selfish behaviors in our experiment. Probably, the animals fail to solve iterated iPD satisfactory due to they fail to discern the difference in the reinforcements for their response. The contrast between rewards and also between punishments is the key for they can match dissimilar responses to different rewards. Moita and Stephens used similar contrast matrix because they used 4 pellets for R(mutual cooperation) and 6 pellets for T(temptation), these are both big and similar amoung that according Killen's *Incentive Model* are magnitud difficult to differentiate by the animals (Killeen, 1982a; Killeen, 1985; McDowell and Kessel, 1979). Indeed, we use a pay-off T as double of pay-off R and the punishment S as double of P. if a animal chose temptation and persist pushing *defect* option, it will receive only one big reward and then timeout, but if it prefer cooperate, will ear for each time small reward and any timeout. However, if the animal strategy is alternate between both lever (C or D) it receive alternately the biggest reward and the longest timeout. The payoff matrix used in the experiment gives the same amount of reward to one that always cooperator as one that always alternate between cooperate and defect options. Nevertheless, the matrix on iPD has the best strategy (*pareto optimum*), because the *always cooperate* strategy doesn't receive timeout punishment and the *alternate* does. Since the figure 5c we observe that the highest reward corresponds to the subject with the shortest accumulated timeout punishment and the highest cooperation choice level(light color). The simulated agent with *alternate* strategy got the maximum punishment. Although, we see that all rats stay on a thin scatter between 85 to 100% of reward and the cooperation level spread out 65 to 100%, this means that at these levels of cooperation the iPD game become very hard to understand whether the cooperation choice is still the best option and indeed the timeout maybe is who modulate what the rats choose. We found 5 of 8 rats were below 30% of punishment and surely the rats with more timeout's sensitivity will try to avoid more it and cooperate more. Certainly, the high level of cooperation was possible by the fact that high contract on both reward and punishment empower a drop in the range where discriminate response-reinforcemente in IPD become hard.

We can infer the strategy that each rats trend to use by a *coefficient of preference* and comparing positions among rats and simulated agent. The coefficient of preference graph showed that all rats were positioned themselves on the right side of *alternate (CD)* agent and this means that the frequency of cooperation option was higher than defect. Besides, we can estimate than all rats at least cooperate twice time before defect.

Then, the results of markov chains graphs confirm the fact. Since the Markov chains graph we see a very high value of transition to mutual cooperation and very low value of transition from S (sucker) to T (temptation). This is a main discrepancy respect to the Moita's markov chain graph (moita 2010) in which the animals trend to jump between cooperate and defect that is between T and S outcomes.

The reciprocal altruism is cooperation among unrelated individuals where in turn each individual help each other and the decision of cooperate is based on previous response and the partner's behavior. When reciprocity is towards the same partner that previous cooperate the behaviors is call *Direct* reciprocity and when is towards another is call *indirect* reciprocity. Our experiment was design to the rats has *direct* reciprocity interaction. Other authors as Taborsky and colleges (2007, 2008, 2012 and 2015) had tested reciprocal altruism in rats using a adaptated Waals' setup (Waal, 2000) to assess direct and general reciprocity in rats. The task consisted of pull a stick in such a way that a partner meets a food tray and then the roles is exchange. They show that rats are more likely to cooperate in direct than indirect or generalized reciprocity. However, from an operant conditioning perspective, interchange roles with a rats that not pulling is consistent with the extinction of the pulling rate behavior that previously was learned. To depict reciprocal altruism behaviors in laboratory experiment design must to meets all constraint together. Indeed, we keep in mind that not learned any cooperative behavior before the iPD treatment in order that the rats build their own strategies. Other example about is the work of wood and colleges (wood et. al., 2016) in which test iPD among same-sex conspecific and cage-mate norway rats not attained the constraint

that the individual must be unrelated (Trivers,1972).

In the second phase of the experiment we tested if the behavior develop by the rats with high performance were learned as a consequence of the interaction with the opponent strategy and amount of reward or solely by the lever preference.

Our result suggest that norway rats have the cognitive ability to reach the best strategy on iterated Prisoner's Dilemma when opponent is reciprocate and have good pay-off matrix and contingency parameters. The cooperation on iPD has been show in similar experiments, but the present experiment is the first time in which a high level of cooperation and a truly reciprocal behavior is reached. The rats not only develop a high frequency of cooperation, but also learn that is more suitable after one fail to cooperate come back to cooperate. The probability of cooperate given outcomes vector show that no matter what they chose previously, they always return to cooperate with high probability when have a reciprocate opponent.

After that these experiment were accomplished the need arose for test different opponent strategy and pay-off matrices to evaluate animals behaviors and now we are performing it.

5 Reference

6 Supplementary Material

6.1 Theorical subject statistics

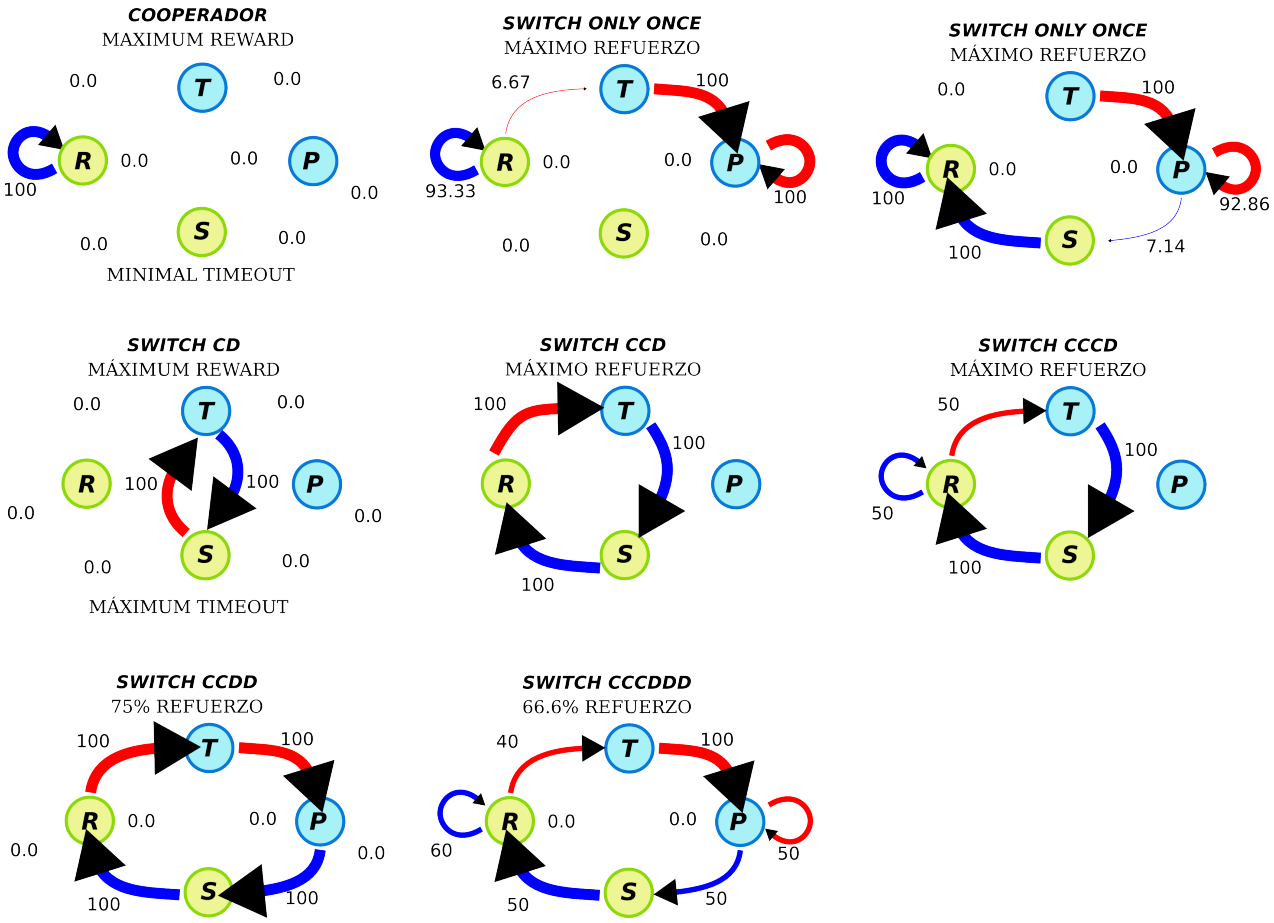
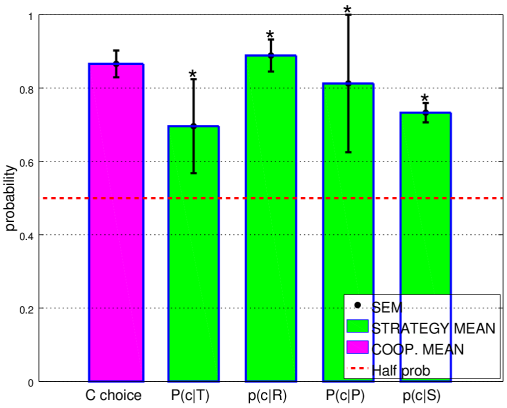


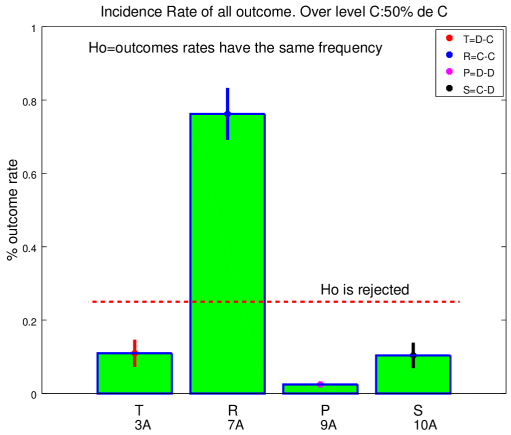
Fig. 9: Graph of markov chain with transition probabilities for all simulates agent used to compare behaviors.

6.2 Non cooperator subject Statistics

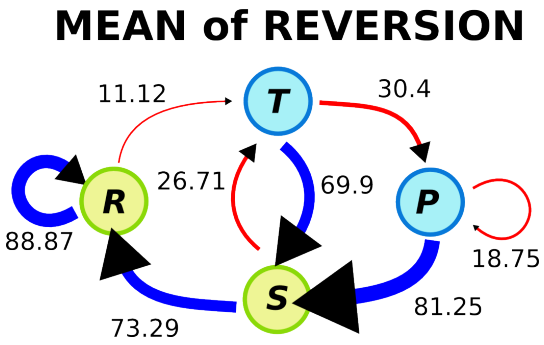
6.3 Cooperatos statistics



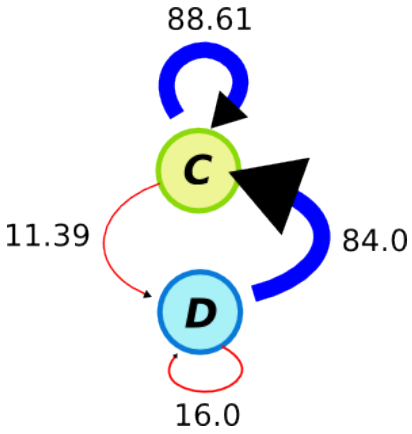
(a) Mean of cooperation and means of transition vector



(b) Outcomes rates over the last ten sessions.

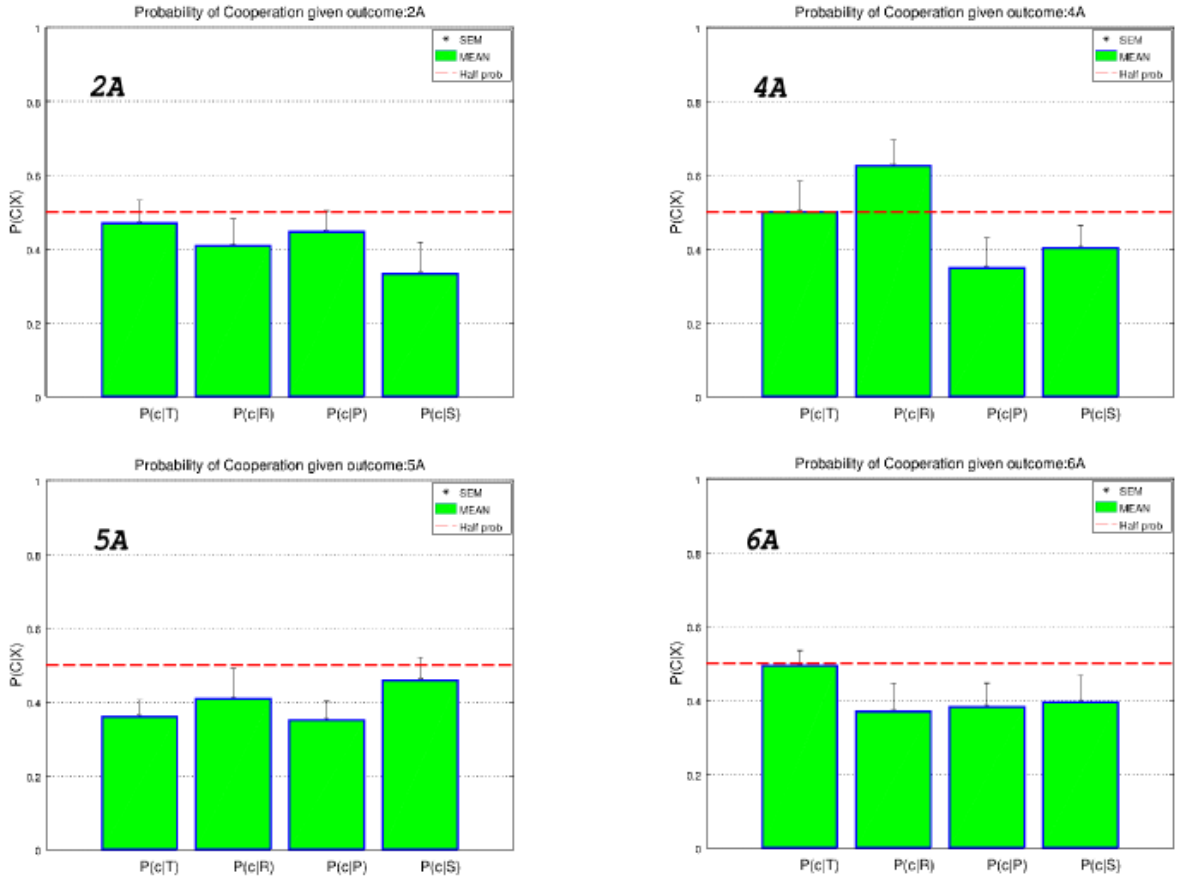


(c) Markov chain graph of four state

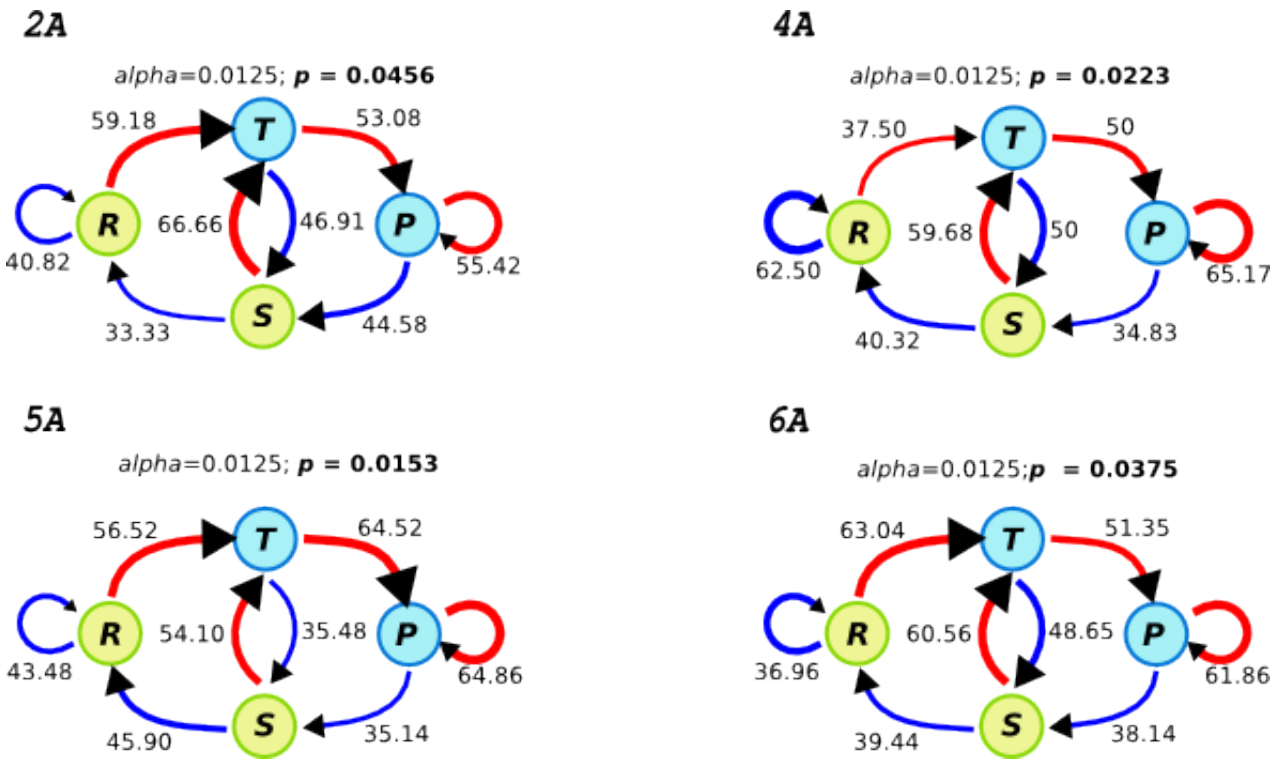


(d) Markov chain graph of two state

Fig. 7: 7a) Mean of cooperation over not random strategy rats (8 rats) was performed and the measured is shown on the first magenta bar ($mean = 0.866$ and $s.e.m. = \pm 0.036$). The next four green bars represent the mean strategy over the group and each bar agree with the transition vector,$v = [0.696, 0.889, 0.812, 0.733]$. The asterisk indicates significant difference respect the half probability of chose cooperation lever and we tested by χ^2 goodness of fit with theory frequency 0.5. 7b) The bar graph shows the outcomes rates and suggest that R outcomes had very high incidence in the experiment over the last 10 sessions. We found significant difference between outcomes (Friedman's ANOVA test was performed with bonferroni correction ($\alpha = 0.0125, p > 3.3e-5; \chi^2 = 23.4$) and Multiple pairwise comparison's using Nemeyi's procedure). 7c and 7d) Graph of Markov chain of four and two state, in which the arrow represent the transition probability between outcomes. A blue arrow signalized when the subject chose cooperate lever given the las outcome and a red arrow signalized when it chose defect lever. The words mean temptation (T), mutual cooperation (R), punishment (P) and sucker (S). Pay atemption that the arrows width are proportional to the transition probability.



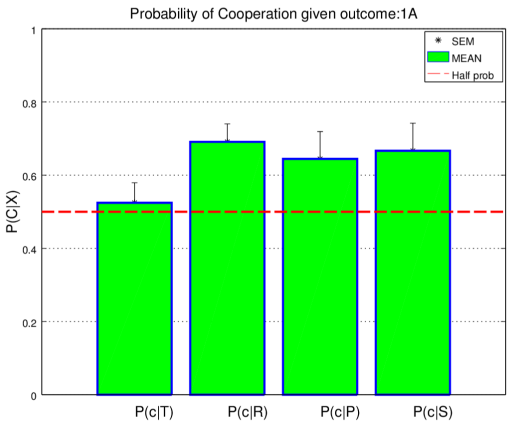
(a) Probabilities of cooperate given each outcomes



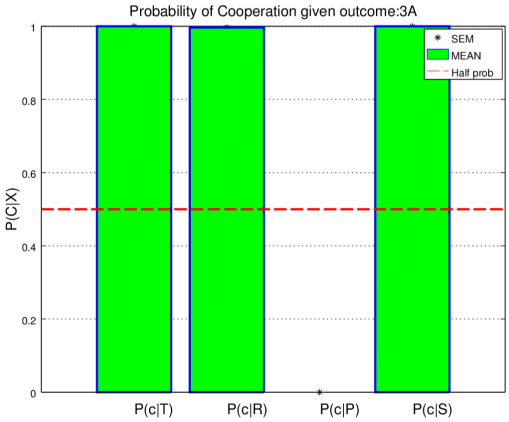
(b) Graph of markov transition probabilities per subject

Fig. 10: Statistic for the discarded subject. These subject had a random behavior because. 10a) All probabilities of cooperate given each outcome are near 50 percent and this means that their haven't any preference choice. See significant value of X^2test in table ??. 10b) The graph of markov chain show that the transition probabilities are very closed to 50%.

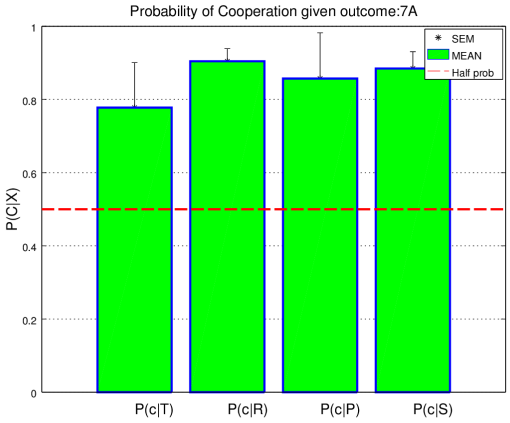
1A



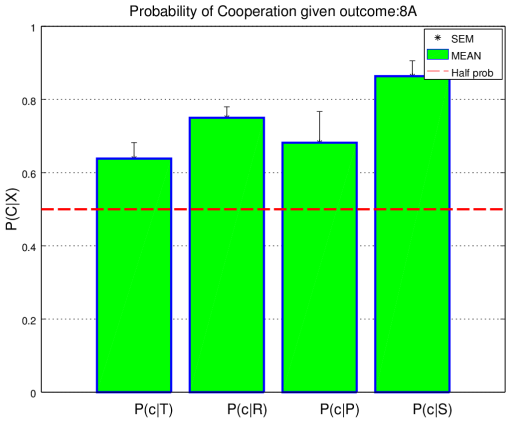
3A



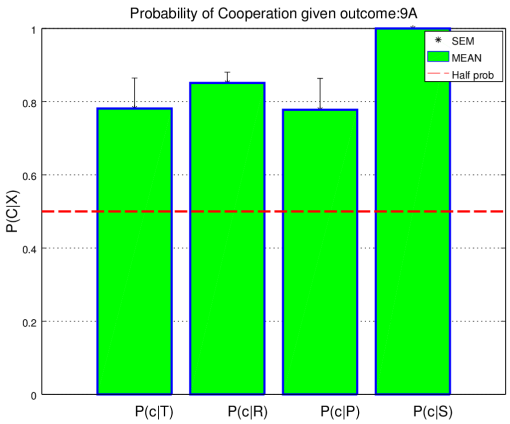
7A



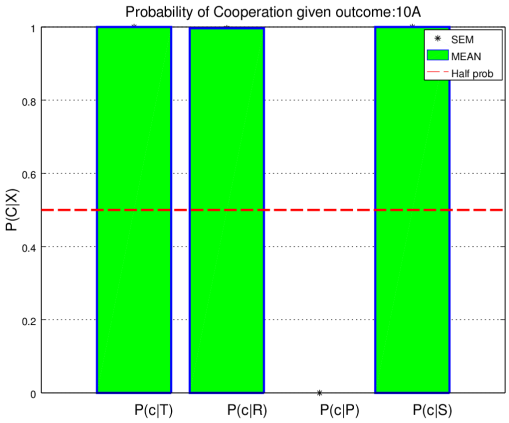
8A



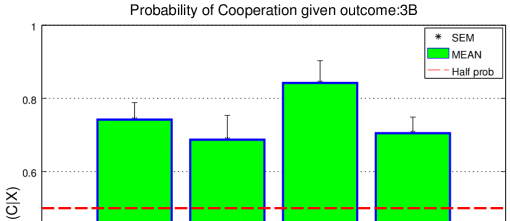
9A



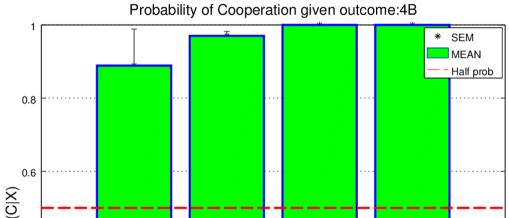
10A



3B



4B



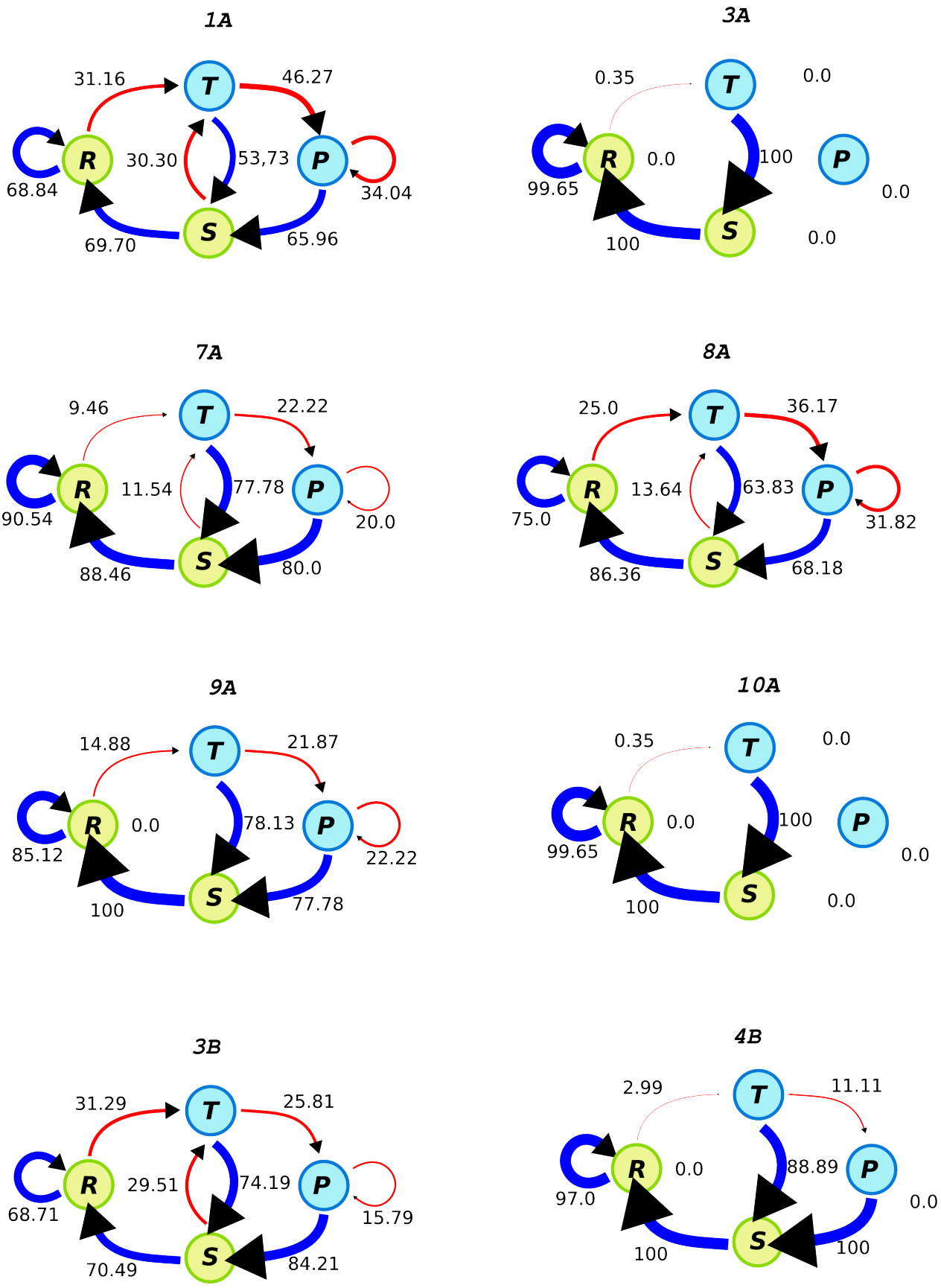


Fig. 12: Graph of markov chain transition probabilities per cooperative rats