Theoretical Neuroscience II
Teacher: Prof. Jochen Braun
Assistant Teacher: Ehsan Kakaei
Name: Viviana Villafañe Barraza & Tasnuba Tazeen & Lee Po Shing
Date: 24th May, 2018

Exercise 5: Rescorla-Wagner Rule

Introduction

For over many decades, rewards and punishments have been employed in various learning paradigms. Pavlovian conditioning, instrumental conditioning and reinforcement learning are noteworthy among such paradigms. The Rescorla-Wagner rule is commonly used to model how animals learn to expect a reward.

In Rescorla-Wagner rule, a stimulus *u* and a reward *r* are introduced in order to postulate a causal heuristic between the stimulus and the reward. According to Rescorla-Wagner rule, the difference between the total expectations towards the stimulus and the actual outcome is the drive of learning, and with the difference the expectation is adjusted each time. Thus for each stimulus, a specific reward expectation is learned which is used to predict future rewards. The difference arises between the actual and expected total reward was called the prediction error δ. The rule can be formulated as follow:

$$\delta_i = r_i - w_i \cdot u_i \qquad\qquad w \to w + \epsilon\delta u$$

In this assignment, the Rescorla-Wagner rule was used to understand how different conditioning paradigms affect learning. In particular, the reward expectations of two stimuli and also the prediction error associated with the reward expectations were determined. Finally, the development of reward expectations over time was plotted.

Method

In this exercise, a situation consisting two stimuli, *A* and *B* were considered to understand the Rescorla-Wagner rule. The stimuli u_A and u_B, with a strength of either 1 or 0, were independently present with probabilities $P(u_A = 1)$ and $P(u_B = 1)$ (P_A and P_B in the code) of $1/3$ and $1/5$ respectively. In order to formulate the stimuli with the same statistics, a vector *u* (u_11, u_10, u_01, u_00 in the code) was employed with elements of 0 or 1 to indicate the values the presence or absence of stimuli *A* (first element) and *B* (second element). Next the correlation matrix $\boldsymbol{Q}$ and its inverse matrix $\boldsymbol{Q}^{-1}$ were computed both analytically and computationally.

A Matlab function file called *RescorlaWagner.m* was used in order to compute the reward vector (r in the code) according to the type of conditioning. Three different types of conditioning, namely fully conditioning, partial conditioning and inhibitory conditioning were studied with reward paradigms as in later sessions. In addition, the joint probability of reward and stimulus ($\langle ru \rangle$, P_ru in the code), the conditional probability of a reward given a stimulus ($\langle r|u \rangle$, P_r_u in the code) and the specific reward expectation as predicted by the Rescorla-Wagner rule (w_{ss}, w_ss in the code) were also calculated in this function. The Rescola-Wagner rule was then applied iteratively to learn specific reward expectations. For each trial, the prediction error was computed and increment reward expectations were computed as:

$$\delta_i = r_i - w_i \cdot u_i \qquad\qquad w \to w + \epsilon\delta u$$

The learning rate ϵ was taken as 0.05 and initial weights were 0 for both stimuli. The provided MatLab file *ShowSequence*.m was used to plot the development of reward expectations over trials.

Results & Discussion

Correlation matrix was the same when calculated analytically and with Matlab

Considering the probabilities $P(u_A = 1)$ and $P(u_B = 1)$ for the presence of the two independent stimuli, the joint probability table of said stimuli would be:

$P(u_A \cap u_B)$	$u_A = 1$	$u_A = 0$	$P(u_B)$
$u_B = 1$	$P(u_A = 1)P(u_B = 1)$	$(1 - P(u_A = 1))P(u_B = 1)$	$P(u_B = 1)$
$u_B = 0$	$P(u_A = 1)(1 - P(u_B))$	$(1 - P(u_A = 1))(1 - P(u_B = 1))$	$1 - P(u_B = 1)$
$P(u_A)$	$P(u_A = 1)$	$1 - P(u_A = 1)$	1

By replacing $P(u_A = 1)$ and $P(u_B = 1)$ by $1/3$ and $1/5$, the table would become:

$P(u_A \cap u_B)$	$u_A = 1$	$u_A = 0$	$P(u_B)$
$u_B = 1$	$1/15$	$2/15$	$1/5$
$u_B = 0$	$4/15$	$8/15$	$4/5$
$P(u_A)$	$1/3$	$2/3$	1

To compute the vector of expected stimuli, it is the same as the vector of the probabilities of both stimuli, i.e. $\langle \mathrm{r}\, u\rangle = \begin{pmatrix} 1/3 \\ 1/5 \end{pmatrix}$. For the correlation matrix $\boldsymbol{Q}$, using the following formula, $\boldsymbol{Q} = \begin{pmatrix} 1/3 & 1/15 \\ 1/15 & 1/5 \end{pmatrix}$:

$$\boldsymbol{Q} = \langle u \cdot u^T \rangle = \begin{pmatrix} \langle \mathrm{u}_1^2 \rangle & \langle \mathrm{u}_1 \mathrm{u}_2 \rangle \\ \langle \mathrm{u}_1 \mathrm{u}_2 \rangle & \langle \mathrm{u}_2^2 \rangle \end{pmatrix} = \begin{pmatrix} \sum_{\mathrm{u_A}=0}^{1} \mathrm{P}(\mathrm{u_A}) \cdot (\mathrm{u_A})^2 & \sum_{\mathrm{u_A}=0}^{1} \sum_{\mathrm{u_B}=0}^{1} \mathrm{P}(\mathrm{u_A}) \cdot \mathrm{P}(\mathrm{u_B}) \cdot (\mathrm{u_A}) \cdot (\mathrm{u_B}) \\ \sum_{\mathrm{u_A}=0}^{1} \sum_{\mathrm{u_B}=0}^{1} \mathrm{P}(\mathrm{u_A}) \cdot \mathrm{P}(\mathrm{u_B}) \cdot (\mathrm{u_A}) \cdot (\mathrm{u_B}) & \sum_{\mathrm{u_B}=0}^{1} \mathrm{P}(\mathrm{u_B}) \cdot (\mathrm{u_B})^2 \end{pmatrix}$$

The inverse of $\boldsymbol{Q}$ can be calculated as $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. The determinant of $\boldsymbol{Q}$ is $14/225$, and by doing some algebra,

$$\boldsymbol{Q}^{-1} = \frac{225}{14} \cdot \begin{pmatrix} 1/5 & -1/15 \\ -1/15 & 1/3 \end{pmatrix} = \begin{pmatrix} 45/14 & -15/14 \\ -15/14 & 75/14 \end{pmatrix}$$

The calculated values were the same as the one shown in the code.

Calculated steady state of specific reward expectation was the same as the iterative calculation

Under a full conditioning situation, a reward of 1 is only received when *A* is present, i.e. $\mathrm{u_A} = 1$. The probability of receiving rewards would be as followed:

$\mathrm{P}(\mathrm{r}=1)$	$\mathrm{u_A}=1$	$\mathrm{u_A}=0$	$\mathrm{P}(\mathrm{u_B} \cap \mathrm{r}=1)$
$\mathrm{u_B}=1$	$\mathrm{P}(\mathrm{u_A}=1)\mathrm{P}(\mathrm{u_B}=1)$	0	$\mathrm{P}(\mathrm{u_A}=1)\mathrm{P}(\mathrm{u_B}=1)$
$\mathrm{u_B}=0$	$\mathrm{P}(\mathrm{u_A}=1)(1-\mathrm{P}(\mathrm{u_B}=1))$	0	$\mathrm{P}(\mathrm{u_A}=1)(1-\mathrm{P}(\mathrm{u_B}=1))$
$\mathrm{P}(\mathrm{u_A} \cap \mathrm{r}=1)$	$\mathrm{P}(\mathrm{u_A}=1)$	0	$\mathrm{P}(\mathrm{u_A}=1)$

Replacing the values of $\mathrm{P}(\mathrm{u_A}=1)$ and $\mathrm{P}(\mathrm{u_B}=1)$,

$\mathrm{P}(\mathrm{r}=1)$	$\mathrm{u_A}=1$	$\mathrm{u_A}=0$	$\mathrm{P}(\mathrm{u_B} \cap \mathrm{r}=1)$
$\mathrm{u_B}=1$	$1/15$	0	$1/15$
$\mathrm{u_B}=0$	$4/15$	0	$4/15$
$\mathrm{P}(\mathrm{u_A} \cap \mathrm{r}=1)$	$1/3$	0	$1/3$

The joint probability of reward and stimulus, or one may understand it as the expected reward of individual stimulus, was defined as $\langle \mathrm{r}\, u\rangle = \begin{pmatrix} \sum_{\mathrm{u_B}=0}^{1} \sum_{\mathrm{u_A}=0}^{1} \mathrm{P}(\mathrm{u_A})\mathrm{P}(\mathrm{u_B}) \cdot \mathrm{r}_{\mathrm{u_A},\mathrm{u_B}} \cdot \mathrm{u_A} \\ \sum_{\mathrm{u_B}=0}^{1} \sum_{\mathrm{u_A}=0}^{1} \mathrm{P}(\mathrm{u_A})\mathrm{P}(\mathrm{u_B}) \cdot \mathrm{r}_{\mathrm{u_A},\mathrm{u_B}} \cdot \mathrm{u_B} \end{pmatrix}$ and therefore equalled to $\begin{pmatrix} 1/15 + 4/15 \\ 1/15 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1/15 \end{pmatrix}$. To find the conditional probability of rewards given a stimulus, or the reward association, one can compute that with the expected reward of individual stimulus divided by the probability of the corresponding stimulus, and hence

$$\langle \mathrm{r}|\, u\rangle = \begin{pmatrix} (\sum_{\mathrm{u_B}=0}^{1} \sum_{\mathrm{u_A}=0}^{1} \mathrm{P}(\mathrm{u_A})\mathrm{P}(\mathrm{u_B}) \cdot \mathrm{r}_{\mathrm{u_A},\mathrm{u_B}} \cdot \mathrm{u_A})/\mathrm{P}(\mathrm{u_A}=1) \\ (\sum_{\mathrm{u_B}=0}^{1} \sum_{\mathrm{u_A}=0}^{1} \mathrm{P}(\mathrm{u_A})\mathrm{P}(\mathrm{u_B}) \cdot \mathrm{r}_{\mathrm{u_A},\mathrm{u_B}} \cdot \mathrm{u_B})/\mathrm{P}(\mathrm{u_B}=1) \end{pmatrix} = \begin{pmatrix} 1/3 \div 1/3 \\ 1/15 \div 1/5 \end{pmatrix} = \begin{pmatrix} 1 \\ 1/3 \end{pmatrix}$$

To calculate specific reward expectation at steady state, the following equation was used:

$$w_{ss} = \boldsymbol{Q}^{-1} \cdot \langle \mathrm{r}\, u \rangle$$

$$w_{ss} = \begin{pmatrix} {}^{45}/_{14} & {}^{-15}/_{14} \\ {}^{-15}/_{14} & {}^{75}/_{14} \end{pmatrix} \cdot \begin{pmatrix} {}^{1}/_{3} \\ {}^{1}/_{15} \end{pmatrix}$$

$$w_{ss} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Interestingly, w_{ss} was different from $\langle \mathrm{r} \,|u\rangle$ because, despite the reward association for u_B, it was dependent on u_A. Hence $\langle \mathrm{r} \,|u\rangle$ did not truly reflect the current full conditioning situation in which a reward of 1 was presented when u_A was 1, regardless of the presence of u_B. The value of the final specific reward expectation calculated above was the same as shown in the *Fig. 1*. In the upper panel, the red lines represented the presence of u_A; the blue lines, u_B; and the green lines, r. In the lower panel, the circles represented δ as in the Rescorla-Wagner rule; the red line, the total expected reward of u_A; the blue line, the total expected reward of u_B. As one can see, after 100 trials, the specific reward expectation of both stimuli developed as w_{ss}.

The prediction error was fluctuated with positive values, zero and negative values. δ was sometimes zeros, when the reward obtained in one trial matched the expectation. As a result, the reward expectation would not be updated. On the other hand, when the expected value differed from the reward received, there would be non-zero δ, and thus the expected value updated. When the reward was higher than the total reward expectation, δ became positive, and thus the total reward expectation increases, vice versa. The amplitude of prediction error started higher but was gradually getting lower towards zeros, which indicates the rate of learning also gradually decreased.

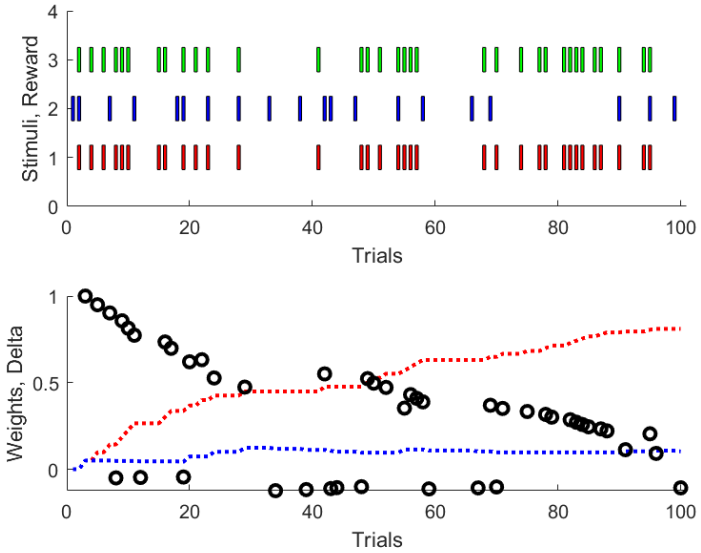

Fig. 1 Total reward expectation emerged towards the calculated steady state value in fully conditioning situation. In the upper panel, the presence of u_a in a trial was shown as the red bar, and that of u_B as the blue bar. Notably, the green representing the presence of reward aligned with u_A. With this set of data, the weights (i.e., total reward expectation) of each stimulus and the prediction error in each trial were shown in the lower panel as the red line, the blue line and the circles. The positive prediction error generally dropped while the red line reached almost 1. The weight of u_B fluctuated around 0. The number of trials was 100.

Development of steady states were slower under partial conditioning

To see the effect of partial conditioning, i.e. the reward was given half the time with u_A independent of u_B, calculations were made in Matlab iteratively. In the first place, 100 trials were used (*Fig. 2*(a)), but the expected w_{ss} of $\begin{pmatrix} {}^{1}/_{2} \\ 0 \end{pmatrix}$ was not observed. As a result, the same were done for 200 trials and then it is possible to see the calculated value developed gradually to the expected one. The vector of w_{ss} obtained for 100 and 200 trials were $\begin{pmatrix} 0.3376 \\ -0.0681 \end{pmatrix}$ and $\begin{pmatrix} 0.5678 \\ 0.0405 \end{pmatrix}$ respectively.

In this case, the conditioning was not as easy as the previous scheme to learn, as it required more trails to see the emerging pattern. The prediction error was varying between -0.5 and 0.5, indicating the probabilistic manner of the reward mechanism.

(a) (b)

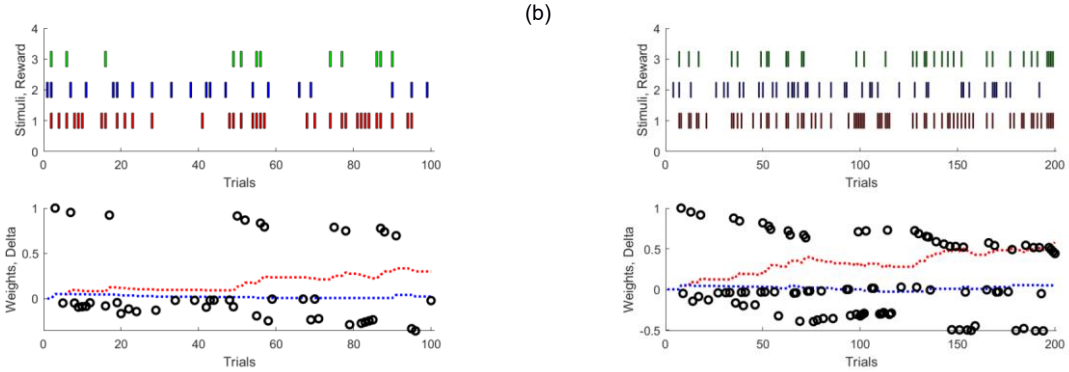

Fig. 2 Total reward expectation emerged towards the predicted steady state value with higher number of trials in partial conditioning situation. Same symbols were used as *Fig. 1*. In (a), the steady states of weights did not reach the predicted value with just 100 trials, but reached with 200 trials as in (b).

Development of synaptic weight based on uncorrelated inputs could only reach binocular

In *Fig. 3*, inhibition was shown to be a hard concept to learn since the fluctuation of δ was also large. Between the 40th- and 80th-trials, a clear pattern of learning that the presence of u_A provided reward was shown with the positive δ. The total reward expectation of the two weights approached the theoretical value of $\begin{pmatrix} 6/7 \\ -2/7 \end{pmatrix}$. The negative value for u_B indicated an association of negative reward for the stimulus, as the presence of this stimulus would inhibit the reward given with the presence of u_A.

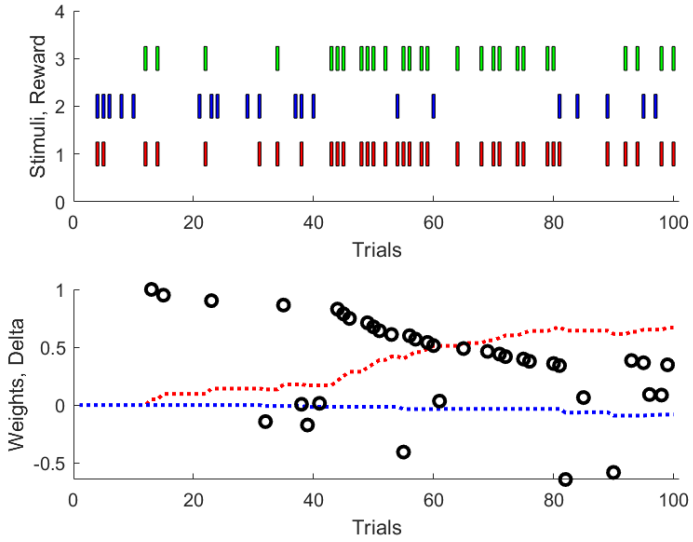

Fig. 3 Inhibitory conditioning predicted a negative weight for the inhibiting stimulus u_B. Same symbols were used as *Fig. 1*. Notably, the weight for u_B was approaching a negative value, yet the final value was expected to be lower.