Problem set 2 : Quantitative models of behavior

Elisa Lannelongue

April 4, 2019

1 Rescorla-Wagner Model

1.1 Introduction

The Rescrola-Wagner model allows us to reproduce classical conditioning experiments. Lets us assume that an animal aims to predict the presence of food reward events. The presence or absence of such a reward is denoted $r = 1$ when there is a reward, and $r = 0$ when there is not. The event of the stimuli, which may or may not predict the presence of the reward event is denoted by $u = 1$ when there is a stimuli and $u = 0$ when there is not. The animal aims to predict if a reward is present depending on whether there was a stimuli or not. The animal's prediction is denotes by $v = wu$ with w the parameter the animal needs to learn to accomplish the prediction task. After every trial, the parameter w is updated using the Rescorla-Wagner learning rule with ϵ the learning rate which can be understood as the associability of the stimulus and the reward, $\delta = r - v$ the prediction error and u the stimulus :

$$w \rightarrow w \ + \ \epsilon\delta u \tag{1}$$

1.2 Classical conditioning simulation

1.2.1 Presentation of the experiment

The learning rule is based on a linear prediction of the reward associated with a stimulus : we assume a linear relationship between the expectation v of the animal and its behavior. the value of w is established in order to minimize the average square error between the reward r and the prediction v.
If ϵ is small and $u = 1$ for every trial, we have that w converges toward the average value of r at which point the value of δ in average is close to 0. The learning process according to the Rescorla-Wagner rule comprises two steps, the phase of acquisition and the phase of extinction (figure 3) : for a total of 60 trials, for each of the 30 first trials, both the stimulus and the rewards were presents, and for the 30 following trials, only the stimulus was present, and the reward was removed (figure 1 and figure 2).

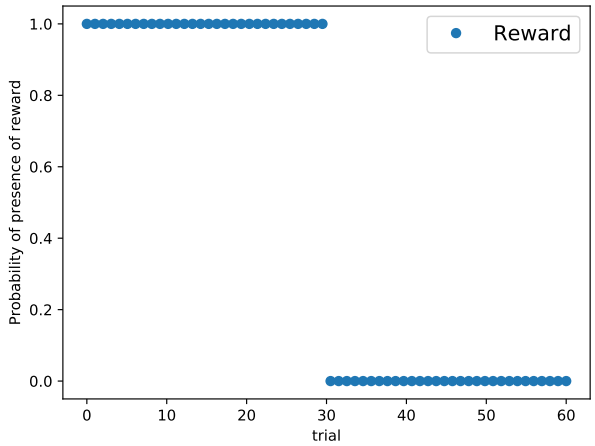

Figure 1: Presence of the reward thought the experiment

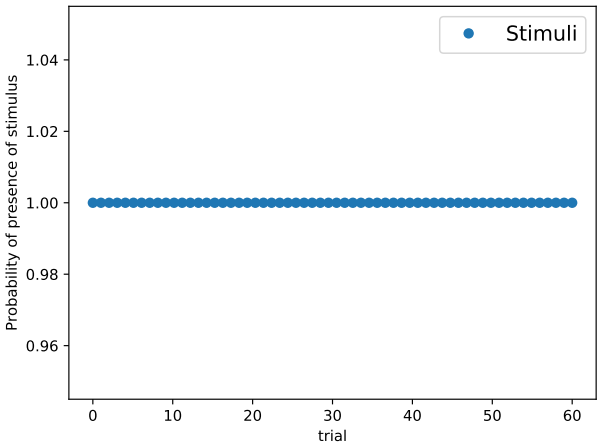

Figure 2: Presence of the stimulus through the experiment

1.2.2 The learning curve

Up until the 30^{th} trial, w grows exponentially and approaches the limit value of the reward $r = 1$. This corresponds to the conditioning phase of the training and from the 30^{th} trial to the 60^{th} trial, w decays exponentially and reached the value $w = 0$ on the last trials, which corresponds to the extinction phase of the training (figure 3).

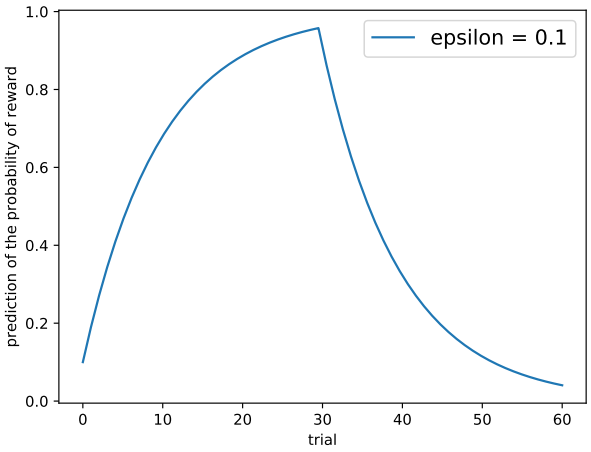

Figure 3: Evolution of v (estimate of the reward) as a function of the trial

When ϵ varies, the type of growth for v remains the same : during the conditioning phase, v which is proportional to w grows exponentially and during the extinction phase, v decays exponentially. However, the bigger ϵ is, the faster the animal learns, which is unsurprising considering that ϵ was introduced as the learning rate parameter of the model. Conversely, during the extinction phase, the greater ϵ is, the faster the extinction process is. The dependency of the learning curve on the learning rate is thus linear (figure 4).

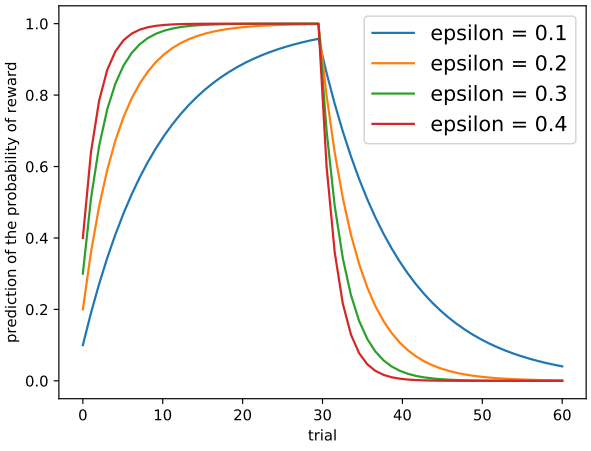

Figure 4: Learning curves for different values of ϵ

1.3 Partial conditioning

1.3.1 Presentation of the experiment

In the case of partial conditioning, the reward is associated with the stimulus on a random number of trials. The extent of the association between the stimulus and a behavior is weaker than when the reward is presented every time the stimulus is present, which is to be expected considering the form of the delta learning rule. In this case, the average value of w is smaller than $r = 1$ (here, $w \approx 0.4$). In the case considered here, the reward is presented with probability 0.4 (figure 5) : the presence of the reward is modeled by a binomial law with probability of success (the reward is present) coded by 1 such that $p = 0.4$, and the probability of failure (the reward is not present) coded by 0 with probability $1 - p = 0.6$.

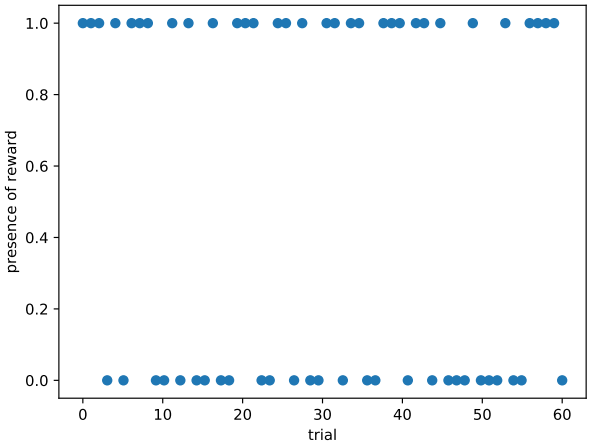

Figure 5: Presence and absence of the reward as a function of the trial index.

1.3.2 Learning curve of w given a stimulus associated with a random reward

In that case, w initially increases until it reaches the proximity of the average of r (here $\langle r \rangle = 0.4$) and then varies randomly around $\langle r \rangle$ (figure 6) : an interpretation of this result is that the animal predicts the average reward and the learning parameter is thus used to average out the noise caused by the random apparition of the reward.

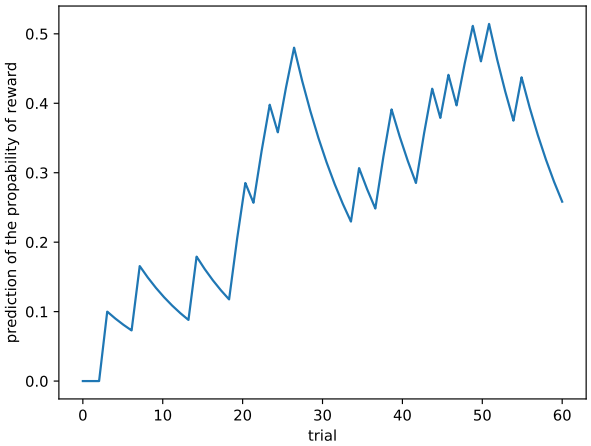

Figure 6: Learning curve when the stimulus is associated with the reward on a random number of trials

1.4 Blocking

1.4.1 Presentation of the stimulation

In the case considered here, the Rescorla-Wagner rule is modified in order to account for cases in which several stimuli are used in association with a reward. $w = (w_1, w_2)$ is a vector of weights, and $u = (u_1, u_2)$ a vector of variables representing the presence or the absence of a stimulus at a given trial. The expected reward v is given by :

(2) $v = w.u$ (with "." the dot product)

The delta learning rule for multiple stimuli is the following :

(3) $w \to w + \epsilon\delta u$ with $\delta = r - v$

Let's consider two stimuli u_1 and u_2. Both stimuli are associated with the same reward, but in this case, for the first 30^{th} trials, only the first stimulus is presented in association with the reward. For the following 30^{th} trials, both stimuli are presented with the reward (figure 7).

1.4.2 The learning curve of w in the event of blocking

What happens here is that two stimuli are presented with a reward, but only after an association has already been formed between one of the stimuli and the reward. The pre-existing association between one of the stimulus and the reward blocks an association from forming between the other stimulus and the reward in the following trials. This result is due to the form of the delta rule : after the training association the stimulus u_1 with the reward r, we have that w_1 is close to the value of r (here, $r = 1$). When the training with the second stimulus begins, the weight w_2 starts out at 0 but the prediction $v = w_1u_1 + w_2u_2 = r$, which gives us a $\delta = 0$ which means that the weights are not modified from this point (figure 7).

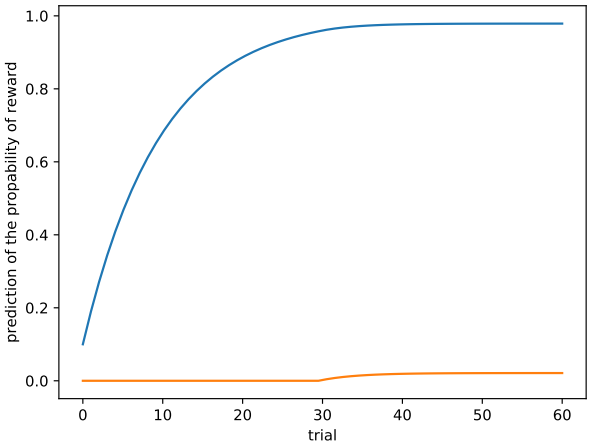

Figure 7: Learning curve when there are two parameters w_1 and w_2 to be associated with a single reward.

1.5 Overshadowing

1.5.1 Presentation of the stimulation

In this case, we consider a situation in which the delta rule is generalized so that both stimuli have different learning rates (u_1 has the learning rate $\epsilon = 0.1$ and u_2 has the leaning rate $\epsilon = 0.2$).

1.5.2 The learning curve in case of overshadowing

Different learning rates means that both stimuli have unequal associability with the reward.
When applied, the delta rule gives us that $v = w1 + w2 = r$. However, if the learning rates are different, the prediction is shared unequally between the

stimuli. Weight modification stops when $= 0$, at which point the faster growing weight will be larger than the slower growing weight (figure 8). Hence, the presence of two stimuli is equivalent to the presence of one stimuli whose learning rate is the sum of the two stimuli's respective learning rates. When we look at the learning parameter evolution. Quantitatively speaking, our simulation show that by the 50^{th} trial, $w_1 \approx 0.3$ and $w_2 \approx 0.7$ and thus that $w_1 + w_2 \approx 1 = r$: the learning parameters of two simultaneously present stimuli add up to one stimulus presented alone and their learning parameters are proportionally distributed according to their learning rates. The following figure is not correct, since in that case, both stimulus are correctly associated with the reward, which should not be the case.

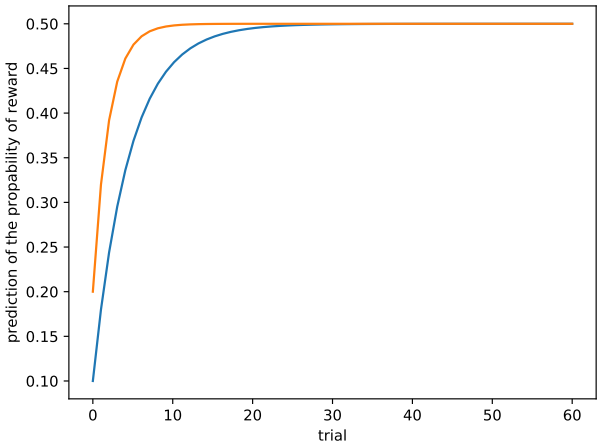

Figure 8: Learning curve for two different stimulus associated with a single reward and different learning rates ϵ.

1.6 Conclusion

The Rescorla-Wagner rule based on a linear reward prediction mechanism is a gross simplifications of animal learning behavior, yet it accounts for much of the classical conditioning experimental data. In the previous stimulation, we explored the role of the learning rate ϵ, which models the animal's capacity to adapt its predictions to the actual presence or absence of the reward : the bigger ϵ is, the faster the adaptation. In the case of blocking and overshadowing, both the presence of both stimuli interact and contribute to the prediction which is taken into account by adapting the delta learning rule.

2 Problem 2 : Simple decision strategy for flower sampling by bees

2.1 Introduction

In the previous section, we considered a classical conditioning experiment, and rewards were associated with a stimulus. In the case we are now considering, rewards are associated with the action the animal takes and they develop policies meant to increase the reward. Another assumption of this model is that the reward follows the action immediately.

2.2 The exploitation-exploration trade-off

Let's consider a bee collecting nectar in a field from blue and yellow flowers. During one day, the bee samples 100 flowers. During the first day, blue flowers are associated with a reward $r_b = 8$ and yellow flowers are associated with the reward $r_y = 2$. On the second day, $r_b = 2$ and $r_y = 8$. On each trial, the bee chooses to land either on a blue flower or on a yellow flower, and this decision is based on the estimate of the reward associated with the action of land on a blue or a yellow flower. The bee's internal estimate of the reward associated with landing on a blue flower is m_b and the bee's internal estimate of the reward associated with the action of landing on a yellow flower is m_y.
The probability p_b of the bee landing on a blue flower is given by the following formula with β the exploitation-exploration trade-off parameter :

$$p_b = \frac{1}{1 + exp(\beta(m_y - m_b))} \tag{4}$$

The Exploitation-exploration trade-off parameter allows us to model the exploitation-exploration dilemma the bee faces when choosing between the blue and the yellow flower on each trial. Exploration is necessary, because it allows the bee to sample from the two colors of flowers to determine which is better, and keep sampling to make sure that the reward conditions have not actually changed. However, exploration can be costly since to make sure the reward conditions have not changed, the bee has to take actions that it believes to be associated with lesser rewards, to check if this is really the case. The model considered here ignores the spatial aspect of sampling : we consider that for each trial, the bee has to choose between a yellow or a blue flower which may seem unrealistic. The probability of the bee landing on a blue flower p_b is a non-linear function of $\beta(m_y - m_b)$ growing monotonically from 0 to 1 for $m_y - m_b$ varying between $-\infty$ and $+\infty$. The parameter β controls the variability of the bee's actions : the larger β is, the more rapidly the probability of an action rises to 1 or decreases to 0 as the difference between the rewards associated with the actions considered increases. If β is small, the probability of the action

considered either rises to 1 or falls to 0 more slowly (figure 9 and 10). When β is large, the bee prioritizes the exploitation of the resources on the basis of the current policy and when β is small, the bee prioritizes the exploration to determine whether it is possible to improve the current policy (figure 9 and 10).

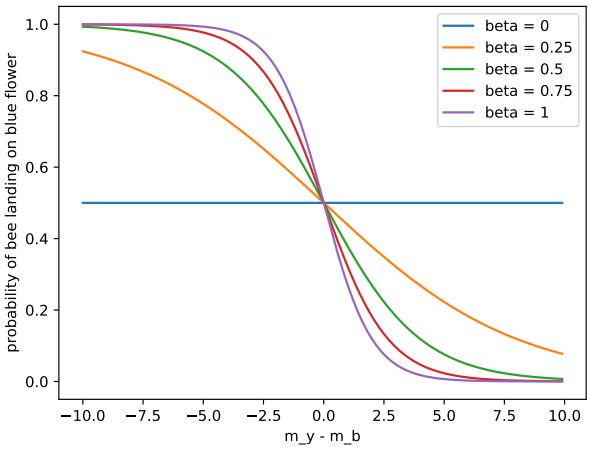

Figure 9: Probability of the bee landing on the blue flower (p_b) as a function of the difference $m_y - m_b$ with fixed β

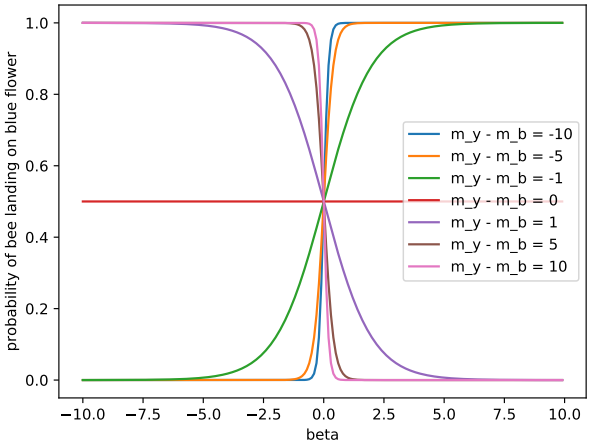

Figure 10: Probability of the bee landing on the blue flower (p_b) as a function of β with fixed difference $m_y - m_b$

2.3 A Dumb bee

In this case, the bee is dumb, and cannot learn from experience : on both days, the bee believes that landing on a yellow flower is associated with the reward $m_y = 5$ and that landing on the blue flower is associated with the reward $m_b = 0$. On each trial, the bee lands on the blue flower with the probability p_b and or it lands on the yellow flower with probability $1 - p_y$. The decision is based on the internal estimates of the bee m_b and m_y.

If we consider a small exploitation-exploration trade-off parameter β ($\beta = 0$), the probability of the bee landing on the flower with the largest estimated reward is close to the probability of the bee landing on the flower associated with the smallest estimated reward ($p_b \approx p_y \approx 0.5$) on both days. In this case, the probability of the bee landing on the blue flower is modeled by a binomial law for which the success is encoded by 1 and corresponds to the bee landing on the blue flower associated with probability $p = \frac{1}{1 + exp(\beta(m_y - m_b))}$ and the failure is encoded by 0 and corresponds to the bee landing on the yellow flower with probability 1_p (figure 11). In terms of the Exploitation-exploration dilemma, the bee's behaviour is highly exploratory . This behaviour is costly since on the second day, when the bee's estimate is closest to the actual rewards, it does not favor the flower it believes rightly to carry the greatest reward.

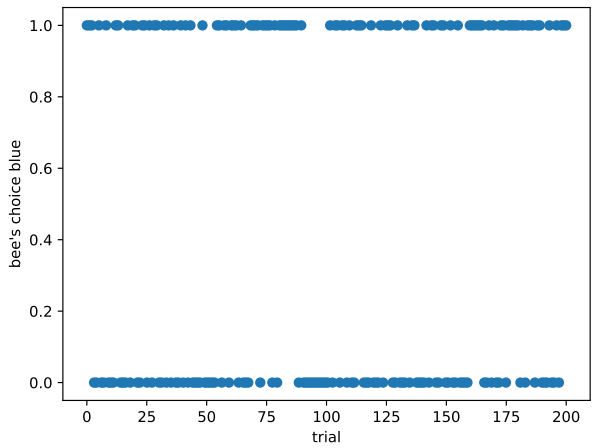

Figure 11: Evolution of the bee's behaviour for a parameter $\beta = 0$

If we consider a large parameter β ($\beta = 0.75$) (figure 12), the probability of the bee choosing to land on the flower with the largest estimated reward is much higher than the probability of the bee choosing to land on the flower associated with the smallest reward on both days (figure 12). The probability

of the bee landing on the blue flower is modeled by a binomial law for which the success is encoded by 1 and corresponds to the bee landing on the blue flower associated with probability $p = \frac{1}{1 + exp(\beta(m_y - m_b))}$ and the failure is encoded by 0 and corresponds to the bee landing on the yellow flower with probability 1_p (figure 12). In this case, the bee adopts a highly exploitative behaviour which is sub optimal on the first day when the blue flower carries a reward of nectar $r_b = 8$ and the estimate $m_b = 0$ is smaller than the estimate of the reward carried by yellow flowers $m_y = 5$ when it is actually the case that $r_y = 2$.

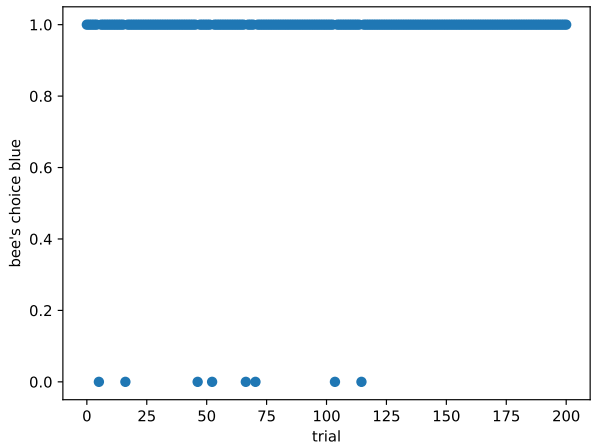

Figure 12: Evolution of the bee's behaviour for a parameter $\beta = 0.75$

2.4 A Smart bee

Let's assume the bee is smart and can learn form experience : when it lands on a blue flower, it updates the estimated reward according to the following update rule :

$$m_b \rightarrow m_b + \epsilon(r_b - m_b) \tag{5}$$

And when it lands on a yellow flower, it updates the estimated reward according to the following rule :

$$m_y \rightarrow m_y + \epsilon(r_y - m_y) \tag{6}$$

Let's assume in this case that the learning parameter $\epsilon = 0.2$ and the bee's

initial internal assumptions about the flower rewards are such than $m_b = 0$ and $m_y = 5$. On day one, the flower rewards are such that $m_b = 8$ and $m_y = 2$ and on day two, the flower rewards are such that $m_b = 2$ and $m_y = 8$.
The smart bee's blue flower reward estimate initial value being lower than r_b, it increases exponentially on the first trials until it reaches the value of r_b, on approximately the 10th trial ($m_b \approx 7.9$). The value of m_b remains stable until the 100^{th} trial at which point the second day begins and the flower reward value is changes (figure 13). For the next 100 trials, $r_b = 2$, and the bee's estimate of the blue flower reward decays exponentially until it reaches the value of $r_y = 2$ on the 113^{th} trial ($m_y \approx 2.0$) and thus stabilizes.

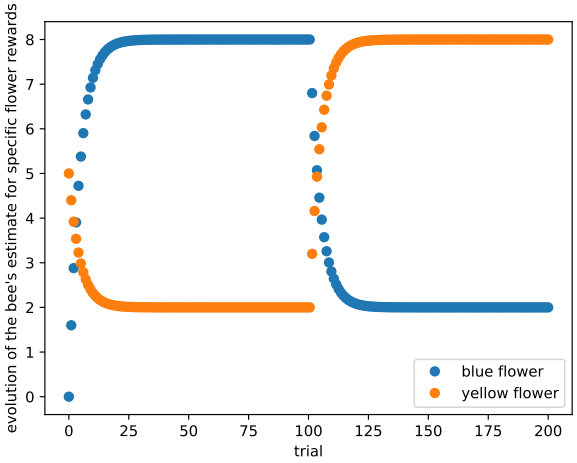

Figure 13: Evolution of the bee's flower reward estimation on both days for $\beta = 0.5$

As expected, curve tracing the evolution of the smart bee's internal estimate of the yellow flower reward is the symmetrical of the one blue flower estimate curve. The smart bee's yellow flower reward estimate initial value being higher than r_b, it decays exponentially on the first trials until it reaches the value of $r_y = 2$, on approximately the 10th trial ($m_y \approx 2.1$). The value of m_y remains stable until the 100^{th} trial at which point the second day begins and the flower reward value is changes (figure 13). For the next 100 trials, $r_y = 8$, and the bee's estimate of the yellow flower reward increases exponentially until it reaches the value of $r_y = 8$ on the 113^{th} trial ($m_y \approx 8.0$) and thus stabilizes.

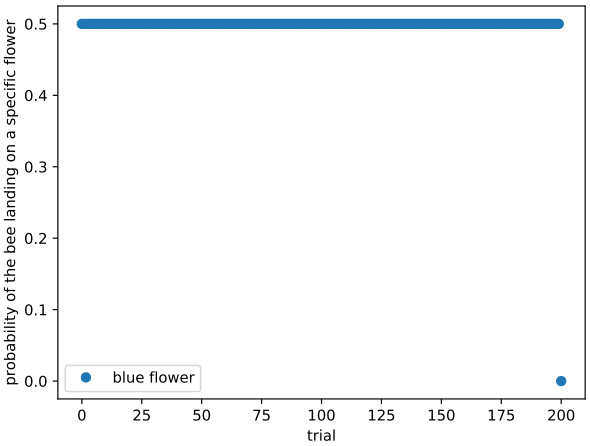

Figure 14: Probability of the bee landing on the blue flower with $\beta = 0$

In the case of purely explorative behavior ($\beta = 0$), given a learning parameter $\epsilon = 0.2$ and the same initial assumptions about flower reward as previously, let's simulate the bee's sequence of choices (figure 14). We can observe that the bee's choices are insensitive to the flower reward and to the bee's own estimates (figure 15).

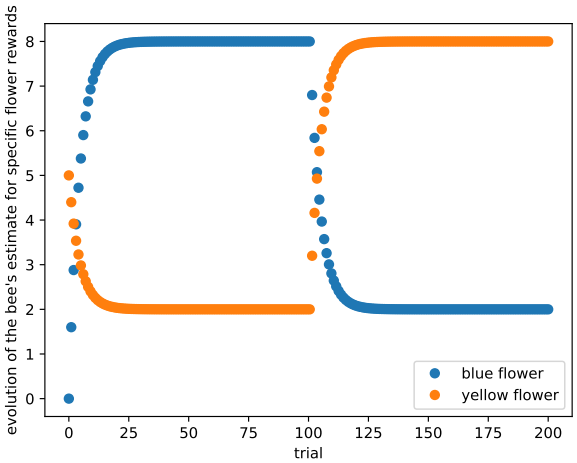

Figure 15: Evolution of the bee's estimates on both days as a function of time for $\beta = 0$

In the case of a strongly exploitative behavior ($\beta = 1$), the bee's choices

adapts to the shift of flower rewards on the second day (figure 16). In the first trials of the first day, the bee picks yellow flowers on several occasions since we started the simulation with $m_y = 5 > m_b = 0$. For the same reason, on the fist trials of the second day, the bee chooses blue flowers on several occasions, because after the first day, its blue flowers estimated reward is higher than its yellow flowers estimated reward. After a short period of time, the bee's choices are stabilized (figure 16).

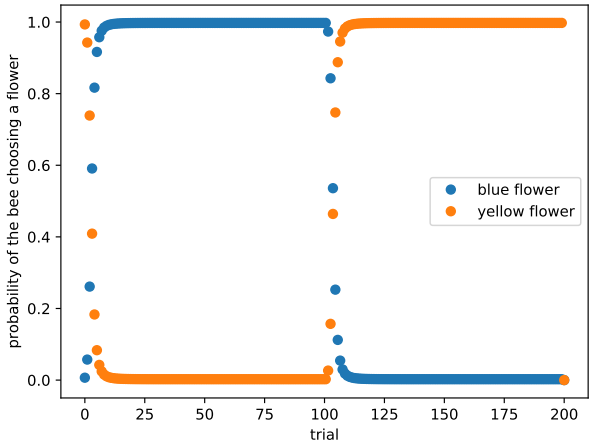

Figure 16: Probability of the bee landing on a specific flower with $\beta = 1$

If we consider the evolution of the bee's estimates of blue and yellow flowers (figure 17), the bee's estimate of each flower converges towards the actual flower reward values. In the following , the estimates of both flowers are updated every time a flower is picked, which should not be the case in the model as it is presented (the estimate of the flower reward should only be updated when the bee actually visits the flower).

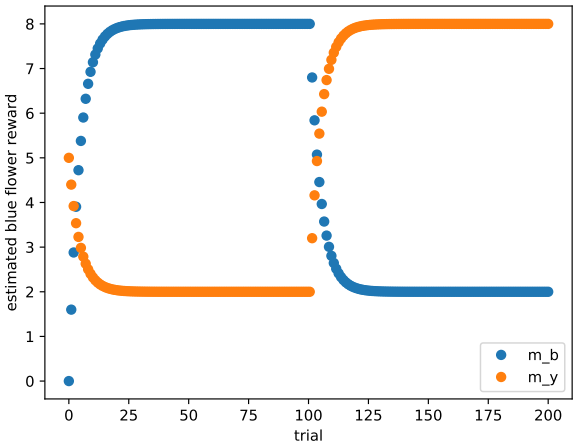

Figure 17: Evolution of the bee's estimates on both days as a function of time for $\beta = 1$

2.5 Conclusion

In this case, we departed from classical conditioning models to model a situation in which the reward is associated with a certain behavior on the animal's part. The type of strategies adopted in order to maximize the reward obtained at the issue of each trial depends on the exploitation-exploration parameter β which models the exploitation-exploration dilemma. It represents the animal's preference for exploitation or exploration : the bigger β is, the more exploitative the bee's behaviors are, and the closer to 0 β is, the more explorative they are. We also looked into an online update rule which allows our simulation to account for the animal's estimation update according to the rewards previously collected.

3 Problem 3 : The drift diffusion model of decision-making

3.1 Introduction

In the following section, lets consider a two-alternative forced choice task in which subjects are asked to choose between two actions. In this case, a subject receives a visual motion stimulus (a set of points n a screen moving in different directions) and they need to indicate whether the points are moving upward or downward. The task is all the more difficult when the stimulus is ambiguous. We assume that the motion stimulus continues until the subject has chosen. This scenario can be described with the "drift-diffusion model", in which the subject compares the firing rate m_A of the upward-motion sensitive neuron with

the firing rate of the downward-motion sensitive neuron m_B and integrates the difference between m_A and m_B, with $\eta(t)$ a noise term (Gaussian white noise) :

$$\dot{x} = m_A - m_B + \sigma\eta(t) \tag{7}$$

An ordinary differential equation can be solve using the Euler method, that is the following approximation :

$$x(t + \Delta t) = x(t) + \dot{x}\Delta t \tag{8}$$

In the case of stochastic differential equations, the Euler methods yields the following approximation for the drift-diffusion model :

$$x(t + \Delta t) = x(t) + (m_a - m_b)\Delta t + \sigma\eta(t)\sqrt{\Delta t} \tag{9}$$

If x is above some threshold μ, the subject decides in favor of A, and if it is bellow some threshold $-\mu$, then they decide in favor of B.

3.2 The drift-diffusion model

Let's assume for the first stimulation that $m_A = 1$ and $M_B = 0.95$, for $x(0) = 0$, a stepwidth $\Delta t = 0.1\ ms$, and a noise level $\sigma = \frac{1}{2}$. We iterate the Euler method described above over 10000 time steps up to time $t = 1s$ (figure 18).

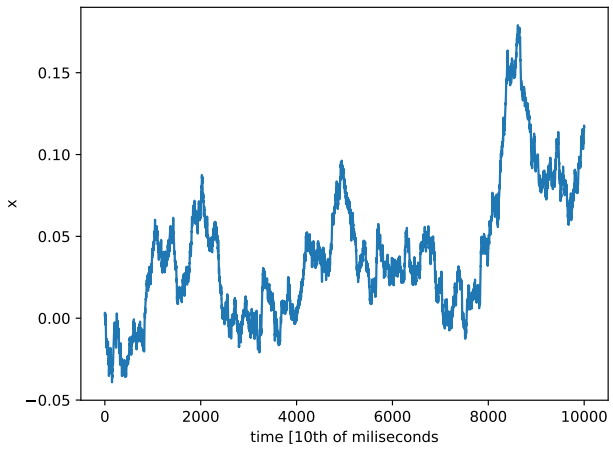

Figure 18: Evolution of x as a function of time over 10000 steps

Now let's run the model 1000 times, and store the outcome A or B as well as the time of threshold crossing t_t for each trial. The threshold μ was set at

0.1. If x is bellow the -0.1, then the subject decides in favor of outcome B, and if x is above 0.1, the subject decides in favor of outcome A. In this case, the subject decided in favor of outcome A 498 times, in favor of outcome B 445 times and the subject did not come to a decision in the timespan considered 57 times.

Let us define the reaction time of the subject RT as follows :

$$RT = 100 + t_t \tag{10}$$

The distribution of the reaction times for outcome A is the following (figure 19). In the cases when the subject choose the outcome A, it appears that most of the reaction times are comprised between 10 to 50 ms, and that the longer the reaction times, the fewer occurrences of the subject choosing outcome A. During all trials, the firing rate of the neurons rose quickly to an early peak, and then the firing rate continues to grow when the subject chooses outcome A, which in this case is the correct choice, or drops if the subject settles on outcome B. In the cases in which the subject choose the outcome B (figure 20), the reaction times distribution is similar, although slightly translated to the left, which means that most of the reaction times in this cases are longer than when the subject chooses outcome A.

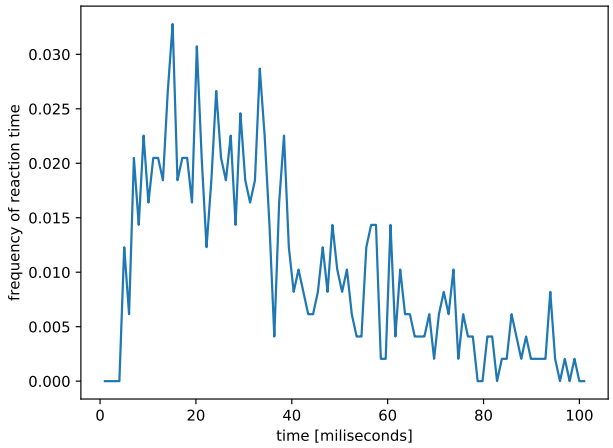

Figure 19: Distribution of the reaction times for outcome A

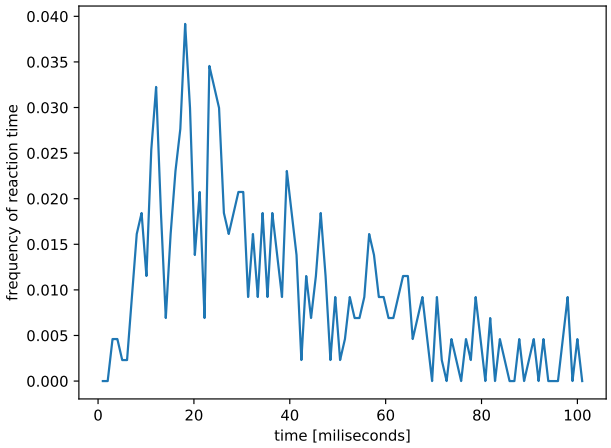

Figure 20: Distribution of the reaction times for outcome B

3.3 Comparison between empirical and analytical results

Let us denote the evidence for outcome A versus outcome B as $m_E = m_A - m_B$ and plot the probability of outcome A for different values of m_E ranging from -0.2 to 0.2 and compare the results of the stimulation with the following analytical formula with $\beta = \frac{2\mu}{\sigma^2}$:

$$p_b = \frac{1}{1+exp(\beta(m_A - m_B))} \tag{11}$$

If the analytical formula giving the theoretical probability of the subject choosing outcome A and the empirical probability of the subject choosing outcome A yielded by the stimulation are alike, the this would show the robustness of the model. If the curves thus obtained differ, then the nature of the difference between both curves would hint at the type bias present in the model. Due to code errors, I could not plot the last figure showing the difference between both curves.

3.4 Conclusion

In this section, we studied the drift-decision model, which based on stochastic differential equation allows us to study decision-making when the stimulus and probabilities influence the subject's behaviours in a predictable manner. In this case, the predictive power of this model is limited by the type of decision it can account for. For instance, it does not allow us to study decision making at an introspective level.