Table 1: **Decision Making**

Types	Value Function	Reward	Next Node [*]
sim	$Q- = \frac{expect-true}{expect}$ $value\_exp = \frac{e^{Q_i}}{\sum_{j=1}^{k} e^{Q_j}}$	–	$[expect \times value\_exp]$
rl	$Q^{new}(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha.[r_t + \gamma.\max_a Q(s_{t+1},a) - Q(s_t,a_t)]$ [**] $value\_exp = \frac{e^{Q_i}}{\sum_{j=1}^{k} e^{Q_j}}$	$r_t = log(\lvert expect - true \rvert)$ [***]	$[expect \times value\_exp]$
exp	$Q^{new}(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha.[r_t + \gamma.\max_a Q(s_{t+1},a) - Q(s_t,a_t)]$ $value\_exp = \frac{e^{Q_i} * expect_i}{\sum_{j=1}^{k} e^{Q_j} * expect_j}$	$r_t = log(\lvert expect - true \rvert)$	$[value\_exp]$
rl_true	$Q^{new}(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha.[r_t + \gamma.\max_a Q(s_{t+1},a) - Q(s_t,a_t)]$ $value\_exp = \frac{e^{Q_i}}{\sum_{j=1}^{k} e^{Q_j}}$	$r_t = true$	$[expect + \max_{\mathcal{N}}(value\_exp)]$ [****]

$$Q^{new}(s_t,a_t) \leftarrow \underbrace{Q(s_t,a_t)}_{\text{old value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \overbrace{\Bigg(\underbrace{\underbrace{r_t}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(s_{t+1},a)}_{\text{estimate of optimal future value}}}_{\text{new value (temporal difference target)}} - \underbrace{Q(s_t,a_t)}_{\text{old value}}\Bigg)}^{\text{temporal difference}}$$

[*] Choosing neighbour with maximum value of – .Epsilon-greedy with $\epsilon = 0.1$

[**] For this simulation, $\alpha = 0.1, \gamma = 0.95$

[***] If $expect = true, r_t = 0$

[****] (expected idleness + max($value\_exp$ of next neighbours))

Table 2: **Estimating Idleness**

Types	**Expression**
sim	AGENT MODEL Expected idleness = Time elapsed since last visit of that bot to that node
avg	OBSERVATION MODEL 1 Expected Idleness = Average of all true idleness observed on the previous visits $expect_n = \frac{1}{n-1} * \sum_{j=1}^{n-1} true_j$
ses [*]	OBSERVATION MODEL 2 Expected Idleness = Simple Exponential Smoothing (SES) of all true idleness observed on previous visits $expect_n = \alpha true_{n-1} + \alpha(1-\alpha) true_{n-2} + \alpha(1-\alpha)^2 true_{n-3} \ldots$

[*] $\alpha = 0.9$