Algorithm Algorithm Title

Initialize each agent's Q network; (Q_1, Q_2)
while Train Episode **do**
    Take action and queue the sample to each replay buffer
    Obtain random sample batch (s, a, r, s')
    Calculate target Q
    if With 0.1 probability **then**
        $y = r + Q_i(s', a')$ where $a' = \arg\max(Q_j(s'))$
    else
        $y = r + Q(s', a')$ where $a' = \arg\max(Q(s'))$
    end if
    Update each Q according to basic DQN update rule
        $Q \leftarrow Q + \alpha(\gamma y - Q)$
end while