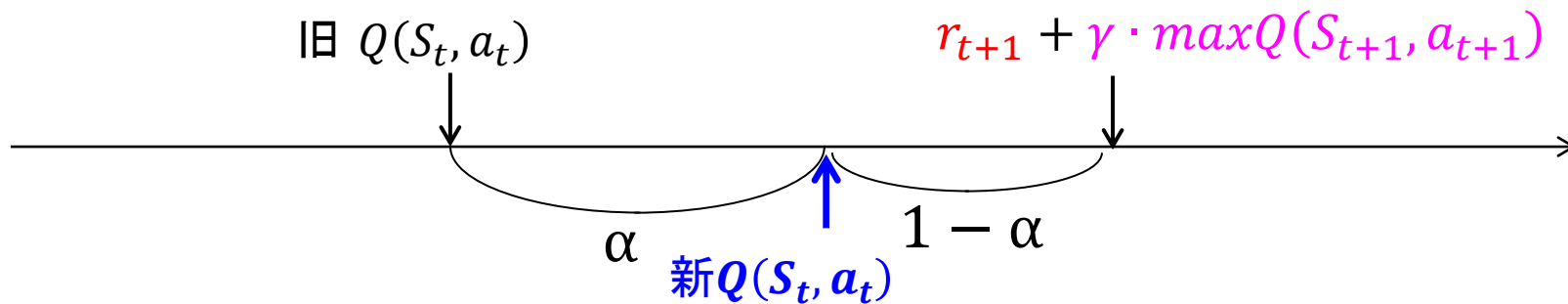
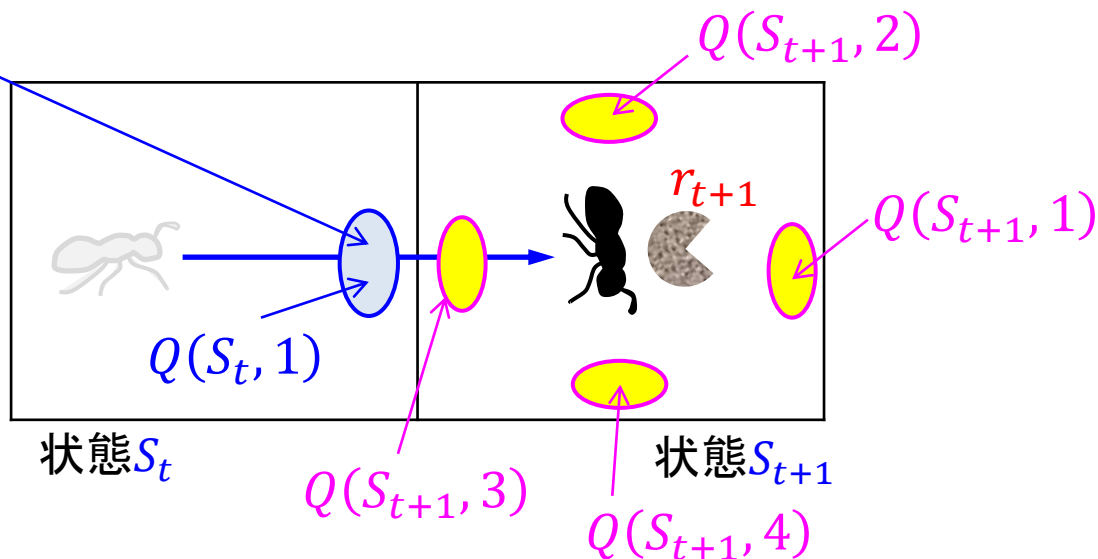


γ : 割引率 …… フェロモンが時間と共に薄れたりする分を割り引きます
 α : 学習率 …… 何度も往復して学習すると信頼度が上がってきます



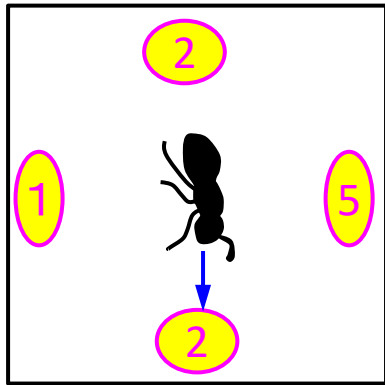
遅延報酬

新 $Q(S_t, a_t) \leftarrow Q(S_t, a_t) + \alpha \{ \overbrace{r_{t+1} + \gamma \cdot \max Q(S_{t+1}, a_{t+1})}^{\text{期待報酬}} - Q(S_t, a_t) \}$

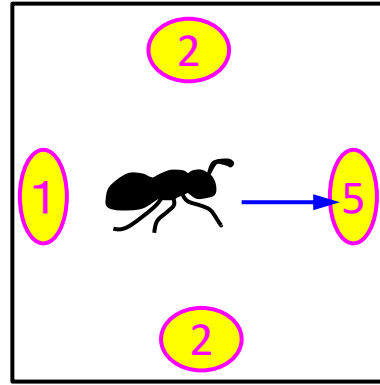


ϵ は蟻の冒険心の度合い

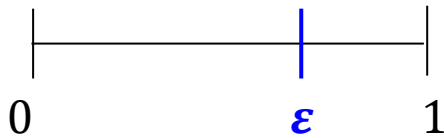
explore(探検する)



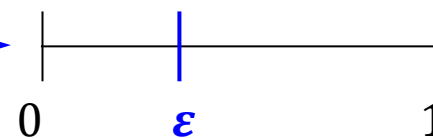
exploit(利用し尽くす)



学習するにつれ ϵ は小さく



最初は冒険する



冒険しない

修正 ϵ -greedy(欲張りな)法

