



Title	Role of midbrain dopamine neurons in reward-based learning
Author(s)	佐藤, 武正
Citation	大阪大学, 2005, 博士論文
Version Type	VoR
URL	https://hdl.handle.net/11094/1305
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

氏 名 ^さ佐 ^{とう}藤 ^{たけ}武 ^{まさ}正

博士の専攻分野の名称 博 士 (理 学)

学 位 記 番 号 第 1 9 5 6 2 号

学 位 授 与 年 月 日 平成 17 年 3 月 25 日

学 位 授 与 の 要 件 学位規則第 4 条第 1 項該当

基礎工学研究科システム人間系専攻

学 位 論 文 名 Role of midbrain dopamine neurons in reward-based learning
(報酬学習における中脳ドーパミンニューロンの役割)

論 文 審 査 委 員 (主査)

教 授 村上富士夫

(副査)

教 授 木村 実 教 授 山本 亘彦 教 授 野村 泰伸

論 文 内 容 の 要 旨

動物が適切な行動選択・環境への適応を獲得する過程において、行動結果の評価と報酬予測との差、報酬予測誤差情報が学習の教師信号として働き、報酬予測誤差がゼロに近づくように学習が進むことを強化学習理論では提唱している。先行研究では中脳ドーパミンニューロンが報酬予測誤差をコードすること、動機付けや注意などへの関与が示唆されてきた。しかし報酬を元にした行動選択の学習・実行過程でドーパミンニューロンがこれらの異なる情報をどのように統合しているのかについて明らかではなかった。

私は本研究で2頭のニホンザルを用い、異なる報酬確率のもとで行動選択を行わせる課題を訓練した。サルは3個のボタンの中から1個のボタンを選択して押すことを要求される。正解の場合には報酬を獲得するが、不正解の場合には報酬を得ることは出来ない。

課題遂行中の中脳ドーパミンニューロンより単一細胞外記録を行った。その結果、課題の最初に提示される視覚刺激(条件刺激)に対して約半数の細胞が有意な興奮応答を示した。条件刺激に対する反応の強さは報酬予測を正確には反映していなかったが、動機付けのレベルと関連があると考えられる反応時間と有意な負の相関があった。また行動選択の正解・失敗を示すピープ音(強化因子)に対しても6割の細胞が有意な反応を示し、定量的に正負の報酬予測誤差をコードしていた。一方、少数の細胞は試行ブロックの切り替えを指示する顕著な刺激に対しても興奮応答を示した。

数ヶ月にわたる行動課題の学習過程で、条件刺激に対する反応における動機付けの修飾のゲインに変化はなかったが、強化因子に対する反応における報酬予測誤差コードのゲインは増加した。さらに4割の細胞は条件刺激と強化因子の両方に対して有意な活動を示し、その条件刺激に対する反応と強化因子に対する反応との間には有意な正の相関があった。一方、顕著な刺激に対する反応は動機付けのレベルと関連すると考えられる反応時間と相関を示さなかった。

これらの結果は、動機付けの程度により報酬予測誤差信号が修飾され、学習の教師信号としての効率性を調節している可能性を示唆する。

論文審査の結果の要旨

動物が適切な行動選択・環境への適応を獲得する過程において、行動結果の評価と報酬予測との差、報酬予測誤差情報が学習の教師信号として働き、報酬予測誤差がゼロに近づくように学習が進むことを強化学習理論では提唱している。先行研究では中脳ドーパミンニューロンが報酬予測誤差をコードすること、動機付けや注意などへの関与が示唆されてきた。しかし報酬を元にした行動選択の学習・実行過程でドーパミンニューロンがこれらの異なる情報をどのように統合しているのかについては明らかではなかった。

本論文では実験動物として2頭のニホンザルを用い、異なる報酬確率のもとで行動選択を行わせる課題を訓練した。サルは3個のボタンの中から1個のボタンを選択して押すことを要求され、正解ボタンを選択した場合には報酬を獲得するが、不正解の場合には報酬を得ることは出来ない。正解の位置はランダムに変化するためサルは試行錯誤で正解のボタンを探索する必要がある。

課題遂行中の中脳ドーパミンニューロンより単一細胞外記録を行った結果、課題の最初に提示される視覚刺激（条件刺激）に対して約半数の細胞が有意な興奮応答を示した。条件刺激に対する反応の強さは報酬予測を正確には反映していなかったが、動機付けのレベルと関連があると考えられる反応時間と有意な負の相関があった。また行動選択の正解・失敗を示すビーブ音（強化因子）に対しても6割の細胞が有意な反応を示し、定量的に正負の報酬予測誤差をコードしていた。一方、少数の細胞は試行ブロックの切り替えを指示する顕著な刺激（フラッシュ刺激）に対しても興奮応答を示した。

数ヶ月にわたる行動課題の学習過程で、条件刺激に対する反応における動機付けの修飾のゲインに変化はなかったが、強化因子に対する反応における報酬予測誤差コードのゲインは増加した。さらに4割の細胞は条件刺激と強化因子の両方に対して有意な活動を示し、その条件刺激に対する反応と強化因子に対する反応との間には有意な正の相関があった。一方、顕著な刺激（フラッシュ刺激）に対する反応は動機付けのレベルと関連すると考えられる反応時間と相関を示さなかった。

これらの結果は、動機付けの程度により報酬予測誤差信号が修飾され、学習の教師信号としての効率性を調節している可能性を初めて示したものであり、博士（工学）の学位論文として価値のあるものと認める。