

목표지향적 학습과 기억

Goal-Directed Learning and Memory

신연순* · 한상훈*†

Yeon Soon Shin* · Sanghoon Han*†

*연세대학교 심리학과

*Department of Psychology, Yonsei University

Abstract

Previous research on learning and memory has focused on how they are constructed through past experiences. Recent studies, however, have shed light on that such cognitive processes are in service of higher goals of maximizing future rewards. This review paper aims to introduce and discuss a related line of research. First, this paper introduces researches that show goal-directed model-based reinforcement learning, in which agents choose a behavior that does not necessarily bring immediate rewards but will allow future rewards, based on generalization and analogical extrapolation. It also reviews studies on neural substrates of goal-directed learning, and discusses that cognitive process implicated in striatal dopaminergic signals can also influence memory. Especially, memory is not a merely passive process of storing and retrieving past experiences homogeneously, but rather results of a decision-making process to serve higher goals. The body of research suggests that information on future rewards can have influence on current cognitive processing in a retrospective manner.

Key words: goal-directedness, learning, memory, dopamine, striatum

요 약

그동안 학습과 기억이 과거의 경험에 의해 구성된다는 측면이 강조되어왔으나, 최근의 연구들은 이들 인지과정이 미래의 보상물을 최대화하는 목표를 달성하기 위해 이루어짐을 조명하였다. 본 개관 논문은 이와 관련된 연구를 소개하고 목표지향적 학습과 기억에 대하여 논의하고자 한다. 먼저 강화 학습에서 내적 모형 기반 학습, 즉 상위 차원의 목표를 달성하기 위해 즉각적인 보상을 가져오지 않음에도 불구하고 특정한 행동을 취하는 과정이 이루어지고, 또한 직접적 강화를 받지 않은 대상으로의 일반화 및 유추가 일어나 미래의 적응적 행동을 가져옴을 보여준 연구들을 소개한다. 또한 위와 같은 목표지향적 학습 과정의 신경학적 기제를 탐색한 연구들을 개관하고, 선조체의 도파민 신호를 기반으로 한 과정이 기억 과정에 역시 영향을 미칠 수 있음을 논의한다. 특히, 기억이 과거의 경험을 모두 동일한 수준으로 부호화하고 인출하는 과정이 아니라, 상위 수준의 목표에 맞춘 의사결정과정의 결과임을 보여주는 연구들을 소개한다. 이러한 연구들은 미래에 얻게 될 보상 정보가 역향적으로 현재의 인지처리에 영향을 줄 수 있음을 시사한다.

주제어 : 목표지향성, 학습, 기억, 도파민, 선조체

* 이 논문은 2013년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구이며(NRF-2012SA5A8024689)

2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(#2012-0003882)

† 교신저자: 한상훈 (연세대학교 심리학과)

Tel: 02-2123-5436

E-mail: sanghoon.han@yonsei.ac.kr

1. 서론

인간은 수많은 경험으로 이루어진 개체이며, 인간 행동의 많은 부분은 기존의 경험이 차후의 의사결정에 시간적으로 순향적으로 영향을 미치는 방식으로 작용하며 이 과정에서 학습과 기억에 의존한다. 즉, 과거의 경험을 바탕으로 현재의 행동을 선택하거나 의사결정을 내리는 인지과정은 우리의 선택과 행동을 개선해나가는 데에 도움을 줄 수 있다. 동시에 기억을 더듬어 더 나은 결정을 내리기 위해 애쓰는 행위는 우리가 기억과정과 그에 따른 의사결정의 결과물에도 매우 민감하다는 사실을 보여준다. 즉, 더 나은 결과물, 보상, 만족감 등의 목표물을 정해, 이를 얻고자 하는 의욕적인 인지 과정이 우리 행동과 의사결정의 핵심이 될 수 있다. 실제 우리가 현재 수행하고 있는 행동을 결정하는 데에는 크든 작든 목표 혹은 결과물이라는 미래의 사실이 시간적으로 앞선 의사결정에 역향적으로 영향을 끼치게 되는데, 이렇게 특정 결과물과 목표를 획득하기 위해 유발된 행동을 '목표지향적 행동(Goal-directed behavior)'이라고 개념화한다. 최근 의사결정 및 기억 관련 연구에서 이와 같은 목표지향적 행동의 역할이 주목을 받고 있다. 본 논문에서는 우리의 적응적 행동에 영향을 미치는 목표지향적 인지요소들에 대한 연구들을 개관하고자 한다. 이를 위해, 먼저 목표지향적 행동에 대한 전통적인 관점(예, 학습이론)과 이를 설명하는 여러 강화학습모형에 대한 연구들을 소개하고, 신경영상학과 신경심리학적 연구들에서 발견한 신경기전의 특징을 논의한다. 그리고, 최근 그 중요성이 강조되고 있는 서술적 일화기억과정에 영향을 미치는 목표지향적 인지과정에 대해 논의하고자 한다.

1.1. 목표지향적 행동

우리의 행동과 의사결정이 객관적, 주관적 목표의 달성을 증진시킬 수 있는 방향으로 이루어진다는 생각은 교육학의 기본 개념이나 인사관리이론과 같은 경영학적 모델들을 구성하는데 매우 중요한 역할을 해왔다. 최근 심리학에서 다양하게 언급하고 있는 '목표지향적 행동'은 이들 분야에서 언급하고 있는 것보다 훨씬 세분화되고 구체적인 개념을 포함하고 있는

데, 목표지향(Goal-directed)이라는 표현은 얻고자 하는 결과물, 보상 등에 의해 동기가 발생하여 현재의 행동을 의욕적으로 수행한다는 의미이다. 행동이 목표지향적이라고 판단되기 위한 필수조건으로는 1) 보상과 같은 행동의 결과물이 존재하여야 하고, 2) 행위의 주체가 결과물을 얻고자하는 의식 혹은 무의식적 욕구와 추동이 필요하다. 이러한 요소를 바탕으로 3) 행동 혹은 반응과 결과물의 관계가 점증적으로 학습되는 도구적 연합관계가 수반되어야 한다. 즉, 목표에 대한 인식과 욕구가 행동의 결정에 선행된다.

하지만 처음 목표지향적 행동의 개념을 언급한 심리학자들은 매우 단순화된 행동 유발과정에 초점을 맞추었다. 초기 행동주의 심리학자들이 '행동'을 영어로 Behavior가 아닌 Action이라고 표현했던 이유도 이들의 연구가 비둘기나 쥐와 같은 단순 인지과정 (혹은 인지를 수반하지 않은 학습과정)을 수행하는 동물 연구에 국한된 경우가 대부분이었기 때문으로, Behavior라는 단어 안에 함축된 상대적으로 고차원적 인지과정을 동반한 의사결정, 추론보다는 '먹이나 보상물을 얻기 위해 스스로 행한 선택을 포함한 움직임과 반응선택' 정도의 단순한 의미를 내포하고 있었다. 이 때문에 때로는 결과물과 행동의 관계가, 즉 둘 간의 '연합정도'가 매우 단순하면서도 잘 공고화되어 있어, 자동적이고 즉흥적이면서, 의식적 노력의 개입이 없는, 반사적이고 습관적인 행동으로 인식되는 경우가 있지만 (예, 습관학습-Habitual Learning), 엄밀한 의미에서 목표지향적 행동은 결과물에 대한 인식을 토대로 이루어지는 의도적 행동까지 포함하는 포괄적 개념으로 보는 것이 더 정확하다.

2. 목표지향적 행동으로서의 학습과정

2.1. 강화학습이론의 발전:

모형기반(model-based) 행동

경험에 기초한 행동의 변화는 과거 동물연구에서부터 인간에 이르기까지 자극(Stimulus)간 연합과 반응(Response)의 관계를 통하여 연구되어왔다. 가령 중성적이었던 자극이 공포 결과물과 연합된 이후 조건화된 공포 반응을 가져오는 고전적 조건형성(Classical Conditioning, Pavlov, 1927; Yerkes & Morgulis, 1909)이

나, 자극(Stimulus)에 따른 반응(Response)이 보상이라는 결과(Consequence)를 가져올 때 해당 반응이 반복될 가능성이 높아지고 부정적인 결과를 가져올 때 반복 가능성이 낮아지는 도구적 조건형성(Instrumental Conditioning 혹은 Operant Conditioning, Skinner, 1938; Thorndike, 1911)이 이에 속한다. 특히, Thorndike(1911)는 도구적 조건형성에서 긍정적 결과를 수반하는 반응은 증가하고 부정적 결과를 수반하는 반응은 감소하는 현상을 효과의 법칙(Law of Effect)라 명명하였다. 이와 같은 효과의 법칙은 행위-결과 수반성, 즉 행위자의 의사결정이 환경을 변화시킬 수 있다는 측면에서 목표에 따라 행동을 선택하는 목표지향적 행동을 이해하는 기초적인 토대를 마련하였다.

Thorndike(1911)의 효과의 법칙(Law of Effect)은 보상을 받은 행동이 강화를 받은 것과 동일한 방식으로 반복됨을 보여주었으나, 이후의 연구들은 단순히 보상과 연합된 행동이라고 해서 무조건적으로 같은 형태로 반복되지는 않음을 밝혔다. 예컨대, 쥐는 성공적으로 먹이를 찾는 목적을 달성하기 위하여 사전에 강화를 받은 경로를 그대로 따를 수도 있지만 점차 강화된 경로 대신 더 빠르게, 에너지를 절약하며 보상을 얻을 수 있는 효율적인 경로를 모색하게 된다(Tolman, 1948). 학습과 기억의 목적이 과거에 경험한 사건을 그대로 재생하는 것이 아니라 습득한 정보를 이후의 탐색 및 반응행동을 결정하는 데 활용하는 것에 있음을 생각해 볼 때, Tolman(1948)이 주장한 인지지도(cognitive map)와 같은 내적 모형을 활용하는 것은 분명히 이점이 있다. 이처럼, Tolman의 초기 연구는 행위자의 의사결정과정이 즉각적인 보상과의 연합보다 보상 가능성에 대한 연산 및 계획이라는 능동적 모형 구성에 바탕을 둔 가능성이 높음을 보여준다. 실제로 Tolman 이후 다수의 연구들은 내적 인지모형기반 학습을 강조하였다. 즉, 개체는 특정 시행에서 보상을 가져온 행동이 이후에도 보상을 가져올 수 있는지 여부를 상위 인지기능과 관련된 내적 모형에서 판단하게 되고, 이러한 모형에 기초하여 다음에 동일한 행동을 반복할 것인지 결정한다고 설명한다(Balleine & Dickinson, 1998; Daw et al., 2005; Daw et al., 2011; Doya, 1999; Niv et al., 2006; Redish et al., 2008).

2.2. 학습 모형과 신경기전

보상에 따른 강화학습의 신경학적 기제를 파악하고자 하는 노력은 원숭이 등을 대상으로 한 동물연구 및 인간의 병변 연구, 비침습적 신경영상학을 활용한 인간 참가자의 뇌 활성화 연구들을 통하여 지속적으로 이루어져왔다(Berns et al., 2001; Knutson et al., 2001; McClure et al., 2003; O'Doherty et al., 2003; O'Doherty et al., 2004; Pagnoni et al., 2002; Schultz et al., 1997; Niv, 2009 개관 참조). 그 결과 도구적 조건형성이 중뇌의 흑질(Substantia Nigra)과 복측피개영역(Ventral Tegmental Area)에서 분비된 도파민이 측핵(Nucleus Accumbens)을 포함한 선조체(Striatum)로 이르게 되는 도파민 경로를 기반으로 이루어진다는 사실이 밝혀졌다(Corbit et al. 2001; Yin & Knowlton, 2006).

특히 선조체에는 도파민 수용체가 다량 분포하고 있으며, 보상 정보를 계산하거나, 예측 기대치와 실제 보상가의 차이, 즉 오류에 대한 연산을 통해 이 차이를 줄여나가기 위한 방향으로 행동을 유도하는 신호를 다른 뇌 영역에 전달한다(Schultz, 2002; Schultz & Dickinson, 2000). 이와 같은 ‘보상 예측 오류’(RPE, Reward Prediction Error)는 강화 학습의 핵심 기전인데, 보상에 대한 기대가 형성되기 전에는 보상이 주어진 순간 도파민 뉴런이 발화하지만, 점차 기대가 형성되어나감에 따라 발화시점이 실제 보상이 주어지기 전, 단서가 나타나는 시점으로 이동해나가는 모습을 보인다(Schultz et al., 1997; Takikawa et al., 2004). 일례로, Schultz 등(1997)의 단일 세포 기록(Single Cell Recording) 연구는 원숭이의 도파민 뉴런에 전극을 꽂아, 특정 단서가 원숭이가 좋아하는 과일 주스 보상을 예측하는 학습이 일어나는 동안 도파민 뉴런의 발화 패턴을 관찰하였다. 이 때 도파민 뉴런은 조건형성이 일어나기 전, 다시 말해 예상하지 못한 보상을 받았을 때에는 주스가 주어진 시점에 발화하였다. 그러나, 조건형성(Conditioning)이 일어난 후에는 단서가 제시된 시점에 발화하였다. 또, 단서가 제시되었지만 실제 주스가 주어지지 않는 경우에는 오히려 불활성상태가 되었다. 즉, 결과가 주어졌을 때 확인한 결과물의 가치가 단서를 보고 예측했던 값보다 크다면 도파민 뉴런이 활성화되지만, 오히려 기대보다 작다면 도파민 뉴런 활성화가 기저선보다 낮은 수준으로 떨어지는 것이다(Schultz et al., 2007; Niv, 2009).

다양한 강화 학습 모형(Reinforcement Learning Model)은 이처럼 예측된 보상과 실제 보상의 차이를 연산함으로써 동기화되는 과정에 대한 신경생리학적 이해를 돕는다. 이들 모형은 의사결정 단계에서 보상이나 처벌을 받은 개별적인 사건들을 모두 회상하지 않더라도, 경험을 통해 예상하게 된 미래의 보상이나 도파민을 통하여 하나의 총합으로 표상됨으로써 의사결정을 인도하는 규범적 틀을 제공한다(Schultz et al., 1997). 이 중 특히 시간차 학습모형(Temporal Difference learning model, Sutton & Barto, 1990)은 한 시행 내에 여러 개의 시점을 설정하고, 앞서 원숭이의 도파민 뉴런을 통해 살펴본 바와 같이 학습이 일어남에 따라 도파민 발화 시점이 시행 초반으로 이동하는 과정을 효과적으로 설명한다(Schultz et al., 1997; Takikawa et al., 2004). 이 모형에서, 기대되는 미래의 보상가를 $V(t)$ 라고 가정하면 그 값은 다음과 같이 나타낼 수 있다.

$$V(t) = E[\gamma^0 r(t) + \gamma^1 r(t+1) + \gamma^2 r(t+2) + \dots]$$

여기서 E 는 기대가치, $r(t)$ 는 시점 t 에 주어질 것으로 기대되는 보상을 말하며, γ 는 보상이 시간상 더 가까울수록 더 큰 효과를 가지고 더 멀리 떨어져있을수록 더 적은 효과를 가지는 시간적 할인요인을 나타낸다. 즉, $V(t)$ 는 시간적으로 멀리 있는 보상일수록 더 적은 가치를 나타내는 미래 가치들의 총합이라고 규정할 수 있다. $V(t)$ 는 앞서 언급한 예측 오류에 의하여 실제 가치에 가깝도록 갱신되는데, 이 예측 오류를 $\delta(t)$ 라고 표현하면, 이는 현재 받은 보상 $r(t)$ 과 이전 시행에서 현재 시행에 대해 기대한 값 $\hat{r}(t)$ 의 차이에 해당한다. 이 때, $\hat{r}(t)$ 는 이전 시행에서 기대했던 미래 보상의 총합 $\hat{V}(t)$ 에서 여전히 미래에 대해 가지고 있는 기대가치 $\gamma \hat{V}(t+1)$ 를 뺀 값으로 계산된다. 즉,

$$\delta(t) = r(t) - [\hat{V}(t) - \gamma \hat{V}(t+1)]$$

으로 나타낼 수 있고 앞서 언급한 복측선조체의 측핵(Nucleus Accumbens) 영역은 바로 이러한 예측 오류 $\delta(t)$ 를 부호화한다고 알려져 있다(McClure et al., 2003).

반면 선조체의 예측오류 신호를 단지 예상치 못한 보상에 수동적으로 반응하는 신호라고만 이해하는 것은 적절하지 않다. 이는 도구적 조건형성의 예측오류 신호에서 분명해지는데, 예측오류 신호는 보상을 획득하기 위해 적절한 행동을 선택하는 데 영향을 미치게 된다. 선조체의 예측오류 신호는 행동을 선택하는 도구적 학습 과정에서 역시 핵심적 기능을 하는 것으로 보이며(Knutson & Cooper, 2005; O'Doherty et al., 2004; Pagnoni et al., 2002), 파킨슨씨병 환자 등 선조체 손상 환자의 경우 보상 피드백을 통한 학습 능력이 현저히 저하된다(Frank et al., 2004; Shohamy et al., 2005). 이처럼 행동이 보상을 받을 수 있는 방향으로 변화해가는 과정을 설명하는 모형으로는 이익 학습 모형(advantage learning model; Baird, 1994, Dayan & Balleine, 2002)이 있다. 이 모형에서 이익 오류 신호 $\delta^A(t)$ 는 자극에 대한 기댓값 $\hat{V}(t)$ 대신, 특정 행동 a 를 취할 경우 얻을 수 있는 미래 보상의 기댓값 $\hat{Q}(t, a)$ 에 대한 오류를 계산한다. 이를 수식으로 나타내면 다음과 같다.

$$\delta^A(t) = r(t) - [\hat{Q}(t, a) - \gamma \hat{V}(t+1)]$$

이렇게 계산된 이익 오류 신호는 해당 시행에 행동 a 를 택할 확률 $p(t, a)$ 를 결정하게 된다. 예를 들어, 단서가 되는 문양을 보고 연합된 가상의 날씨를 예측하는 과제(Weather Prediction Task, Knowlton et al., 1994)를 수행할 때 맑은 날씨 혹은 흐린 날씨를 예측할 때 얻을 수 있는 미래의 총 보상을 각각 비교하여 기대 보상이 더 높은 선택을 하게 되는 것이다. 이러한 선택의 변화는 기초적인 수준의 목표지향성을 보여주는데 보상이라는 목표를 달성하기 위해 행동이 변화해나가는 것이다. 한편, 강화를 받은 선택이 점차 습관적이 됨으로써 결과적으로는 습관적인 선택을 내리는 것이 목표를 달성하도록 돕는다는 점에서, 이와 같은 피드백기반 학습은 목표를 설정하고 달성하기 위한 최적의 경로를 찾아나가는 과정과는 구별된다. 다시 말해 위에서 언급된 전통적 모형들은 학습의 주체가 도식화된 내적 모형을 가질 것으로 가정하지 않으며, 따라서 직접적으로 강화나 처벌을 받지 않은 행동의 빈도가 상위차원의 목표에 따라 증감하는 현상을 설명하기에는 불충분하다.

실제 그 차이는 최근 밝혀진 신경학적 기제들에서

도 나타나는데, 전통적인 강화학습에 의한 습관적 행동이 중뇌, 선조체 등의 피질하 영역에 기반을 둔 것에 비해, 목표지향적 행동은 주로 신피질, 특히 전두엽에 의지하는 것으로 보인다(Glascher et al., 2010; Killcross & Coutureau, 2003). 재밌는 사실은, 최근의 신경영상학적 연구들을 통해 습관학습에 주로 관여한다고 여겨지는 피질하영역이 내적 모형을 기반으로 한 목표지향적 학습에도 마찬가지로 관여할 수 있음을 발견하였다는 점이다(Daw et al., 2011).

이러한 목표지향적 학습은 마코프 의사결정 과정(Markov Decision Process, MDP)을 통하여 탐색된다. MDP는 여러 단계에 걸친 의사결정 끝에 보상을 획득하도록 구성되어있는데, 이 때 보상 가능성을 극대화하기 위해서는 과거에 직접적으로 보상을 받았던 행동 대신 보상을 받을 확률이 가장 높은 다른 행동을 취해야 하는 경우가 존재한다. 그리고 이를 효과적으로 수행하기 위하여, 학습자는 의사결정의 단계가 전환되는 확률에 대한 내적 모형을 구성해야 한다. FORWARD 학습자 모형은 이러한 상태 전환 확률에 대한 매트릭스 $T(s, a, s')$ 를 가정하는 모형이다. 즉, 행동 a 를 취할 경우 상태 s 에서 s' 로 전환될 확률을 예측하는 것이다. 이러한 확률에 대한 예측오류를 보상예측오류와 구별되는 개념으로 상태예측오류(State Prediction Error, SPE)라 한다(Daw et al., 2011; Glascher et al., 2010). 상태예측오류는 아래와 같이 계산된다.

$$\delta_{SPE} = 1 - T(s, a, s')$$

이렇게 계산된 상태예측오류는 다시 학습률을 결정하는 자유계수 η 와 곱해져 확률 $T(s, a, s')$ 를 업데이트하며, 이 때 선택되지 않은 다른 상태 s'' 의 확률은 줄어들게 된다.

흥미롭게도, 보상예측오류(RPE) 신호뿐 아니라 내적 모형이 가정하는 상태예측오류(SPE) 신호 역시 복측선조체에서 관찰되었다(Daw et al., 2011). 다시 말해, 선조체에 목표지향적 학습과 습관적 비모형 학습을 공통적으로 뒷받침하는 도파민 기반 신경기전이 존재하며, 피질하영역에서 두 종류의 학습이 상호작용하는 것이다. 내적 인지모형에 기반한 목표지향적 학습 역시 선조체를 중심으로 한 도파민 경로에 기대어 일어난다는 사실은 선조체가 단순히 현재 보상을

받은 행위에 대한 정보처리 뿐 아니라 목표를 달성하기 위한 전반적인 인지적 과정에 있어 핵심적인 기능을 수행할 가능성을 시사한다.

앞서 살펴본 강화학습 모형들 외에도, 학습의 목표지향성은 다른 상황에 활용할 수 있는 내적 모형 혹은 지식이 축적된 정도를 반영하는 영역의 존재를 통하여 탐색해볼 수 있다. 관찰할 수 없는 지식의 양을 정답률의 변화를 통하여 추산하는 모형이 상태공간모형(State-Space Model, Smith & Brown, 2003; Smith et al., 2004)인데, 관찰식(Observation equation)과 상태식(State equation), 두 수식으로 구성되어 있다. 관찰식은 학습자의 선택행동에 대한 베르누이 확률 함수이며, 해당 선택이 나타날 확률 p_k 을 보여준다.

$$P_r(n_k | p_k, x_k) = p_k^{n_k} (1 - p_k)^{1 - n_k}$$

상태식은 가우시안 무선 모형으로, 관찰할 수 없는 숨겨진 학습 과정을 나타낸다. 이는 아래와 같이 표현되는데,

$$x_k = x_{k-1} + \epsilon_k$$

가우시안 변수가 무선적으로 변화할 때 실제 행동을 통해 얻어진 관찰식을 가장 잘 설명할 수 있는, 즉 관찰식과의 부합치(fit)가 가장 높은 상태식을 추산함으로써 정답률로 나타나는 축적된 지식의 양인 x_k 를 구하게 된다. 이후에 활용할 수 있는 형태의 지식을 축적함으로써 미래의 최적행동을 이끌어낼 수 있다는 점을 고려할 때, 이와 같은 변수를 구하는 것은 목표지향적 행동의 신경기제를 탐색할 효과적인 발판을 마련한다.

2.3. 목표지향적 행동의 목표: 일반화와 유추

최근 연구들은 모형기반 학습 뿐 아니라 다양한 차원에서 학습의 목표지향적 특성을 보여주는데, 특히 학습 과정에서 다양한 방식의 일반화가 일어남을 지지하는 실험적 증거들이 축적되고 있다(Kumaran et al., 2009; Shohamy & Wagner, 2008; Wimmer & Shohamy, 2012; Wimmer et al., 2012). 또한 신경영상학 연구들은 이러한 일반화 과정에서 학습을 뒷받침하는 선조체와

여타 뇌영역, 특히 해마의 상호작용이 목표지향적 학습에 도움을 준다는 증거를 제시한다(Poldrack et al., 2001; Shohamy & Wagner, 2008; Wimmer & Shohamy, 2012).

경험을 차후에 활용하는 데 있어 가장 합리적이고 적합한 형태는 개별적 자극과 상관없이 새로운 상황에 적용할 수 있는 개념적 틀을 형성하는 방법일 것이다. 우리는 학습이 일어나는 동안 개별 경험을 아우르는 일반화된 개념적 구조를 형성하고, 이후에 새로운 문제 상황이 주어졌을 때 앞서 형성된 구조에 입각하여 의사결정을 내리게 된다. Kumaran 등(2009)은 해마가 이와 같은 구조화 과정에 관여함을 보여주었다. 이 실험은 피드백 기반 학습을 유도하는 전통적 패러다임인 날씨 예측 과제(Weather Prediction Task, Knowlton et al., 1994)를 변형하여 카드와 특정 위치, 혹은 두 개의 카드가 복합적으로 날씨를 예측하도록 하였다. 기존의 날씨 예측 과제에서 참가자는 카드에 대한 정보가 전무한 상태에서 맑은 날 혹은 흐린 날을 예측하여 맞으면 보상을 받는 시행을 여러 회기 수행하고, 이에 따라 점차 정보를 축적하여 정답률을 향상해나가는 것이다. 변형된 패러다임은 피드백을 통한 점진적인 수행 향상을 유도한다는 점에서는 기존과 동일하였으나, 하나의 카드만으로는 정확한 예측을 내릴 수 없으며 두 개 이상의 정보(즉, 카드가 나타난 위치, 함께 제시된 카드)를 통합하는 개념적 지식을 형성해야 정답을 맞힐 수 있도록 하였다. 상태공간 모형(SSM)을 통하여 추산된 지식의 양을 반영하는 뇌영역을 탐색한 결과, 시행이 진행되면서 지식의 양이 상승해나가는 만큼 해마의 활성화가 증가하여, 해마가 개념적 지식 형성을 추적하는 것으로 나타났다.

나아가 개념적 구조는 동일하게 유지되되 학습한 경험이 없는 새로운 개별 자극을 사용함으로써 개념적 지식을 일반화하여 활용하도록 유도한 시행에서, 일반화된 지식을 활용하여 높은 수행률을 보인 참가자일수록 해마([x,y,z]=[-21, -30, -6])의 활성화가 높았다(Kumaran et al., 2009). 이는 피드백을 받는 경험이 점진적으로 축적됨으로써 학습이 일어나는 경우에도, 이후에 유연하게 활용할 수 있는 구조적 지식을 형성해나가는 데 선조체 이외의 영역인 해마가 주요한 역할을 수행함을 시사한다.

해마는 유추에도 주요한 역할을 하는데, 특히 해마와 도파민경로의 기능적 연결성은 효과적 유추를 예

측한다. Shohamy와 Wagner(2008)의 연구에서, 참가자들은 연합 학습(Associative learning) 이후 직접적으로 연합되지 않은 자극에 대하여 판단할 때 유추를 사용하는 것으로 나타났다. 이 때, 유추를 효과적으로 사용한 우수 학습자의 경우 학습 진행에 따른 해마 활성화 증가폭 및 중뇌와 해마 활성화의 상관성이 불량 학습자에 비해 큰 것으로 나타났다. 또한 Wimmer와 Shohamy(2012)는 연합된 두 개의 자극 중 하나의 자극에 대해 강화학습을 실시하면 강화를 받지 않은 나머지 자극에까지 기억에 기초한 일반화를 통하여 가치 전이가 일어남을 보였다. 이 실험에서 참가자들은 두 자극 모두 직접적으로 강화를 받지 않았음에도 불구하고, 보상을 예측하지 않는 단서와 연합된 자극에 비해 보상을 예측하는 단서와 연합된 자극을 선호하였다. 이러한 일반화 과정에 역시 해마가 관여하였는데, 강화 학습이 일어나는 동안 후측해마가 활성화된 정도가 이후의 가치 전이 정도를 예측하였다. 또한 뇌영역간의 상호작용이 존재하여, 강화 학습 단계에서 해마와 선조체가 기능적으로 강하게 연결될수록 가치 전이 정도가 크게 나타나 학습에 관여하는 신경 기제와 기억에 관여하는 신경기제의 상호작용이 기억을 활용한 의사결정을 뒷받침한다는 가설을 지지하였다. 위와 같은 유추와 가치 전이는 보상의 극대화라는 목표에 적합한 형태로 학습이 조정된다는 효과적인 증거를 제시한다. 더불어, 학습에 있어 필수적으로 여겨지는 선조체 이외의 영역, 즉 해마가 목표지향적 의사결정을 보조한다는 점에서 주목할 만하다.

목표지향적 행동의 학습기전을 요약하자면 첫째, 특정연합이론이나 모형이 가정되지 않는 습관적 조건형성과과정과, 이에서 더 나아가 인지모형을 기반으로 목표달성에 필요한 효율적 과정을 탐색해나가는 강화 학습이 행위주체의 의도적이고 보상지향적인 행동들을 결정한다. 다양한 강화학습 모형이 보상에 대한 예측, 결과물, 그리고 행동 간의 관계를 유연하게 설명해주며 신경전달물질인 도파민의 개입과 같은 신경학적 기제를 밝히는 데에도 사용되어 왔다. 최근에는 행위와 보상결과물간의 인과적 관계에 대한 지식을 쌓고 이를 일반화하거나 개념적 구조로 습득할 수 있음을 보여주는 목표지향적 행동의 고차원적 인지과정을 밝힌 연구들이 등장하고 있으며 이 과정에서 보상학습의 중추라고 알려진 선조체와 외현적 지식과 기억

습득 중추인 해마영역간의 상호작용에 대한 탐색도 증가하고 있다.

아래에서는 보상-행동간 연합과정인 목표지향적 행동이 외현기억과정에도 영향을 미칠 수 있음을 보여주는 연구를 소개하고, 그 이전에 강화학습모형에서 언급했던 선조체와 도파민 작동이 관여할 수 있음을 논의하고자 한다. 더불어 목표지향적 행동이 단순 연합과정을 넘어 인간의 고위인지과정, 특히 정보의 입력-인출이라는 순향적 인지과정이 핵심인 일화기억과정에도 영향을 줄 수 있음을 논의한다.

3. 목표지향적 행동으로서의 기억과정

앞서 언급하였듯이, 이전의 심리학, 인지신경과학 및 신경심리 연구들은 '목표지향적인 행동'의 의미를 고전 학습심리 연구에서 발견한 조건형성(Conditioning)과정을 통해 설명하였고 (특히, 도구적 조건형성), 행동의 결과로서 보상이 주어지는 반응에 대해 점차 더 많이, 강하게, 지속적으로 반응하는 학습이 일어난다고 설명하였다. 더불어 이 학습 과정에서 뇌 피질 하부영역에 위치한 선조체의 개입을 강조하고 있는데, 이 영역은 특히, 중뇌 영역의 신경핵에서 분비되는 신경전달물질인 도파민(DA) 전달 경로의 중심에 있으면서 신경전달 물질이 뇌 전반, 특히 의사결정을 담당하고 있는 (전)전두엽 영역으로 전달되도록 하는 기능을 담당한다.

그동안 목표지향적 행동이 인간의 기억(특히, 일화 기억)과정에서 관심을 받지 못했던 이유는, 첫째로, 의식적이고 서술할 수 있을 만큼 구체적인, 즉 외현적인 일화기억과정의 성격과 달리, 위 학습기전에서 언급한 보상목표-행동의 반응 관계들이 주로 무의식적으로 습득되는 습관 행동이나 그 맥락을 반영하고 있다는 인식이 있어서이다. 구체적으로, 동일한 자극에 대해 반응-결과물간의 반복적 연합을 통해 점차 우세한 혹은 편향된 행동 반응패턴을 습득하는 보상지향적 학습은 동일자극에 대해 보상에 대한 수반성을 대개 무의식중에 습득하게 되는데 비해, 일반적으로 기억검사에서 매 시행 제시되는 자극은 동일한 자극이 아니며, 반복적으로 제시되는 경우도 극히 드물어 반응-보상의 관계를 정립하는 것이 쉽지 않다. 또한 옳은 반응에 대한 결과물 혹은 피드백이나 보상이 주어

지는 경우에도 이들은 행동반응자체에만 의존하는 것이 아니라, 우리가 얼마나 정확하게 기억 자극의 흔적을 가지고 있는지에 의해 결정된다는 데에서 큰 차이가 있다. 즉, 반응-결과물간의 수반성이라는 것이 대개 존재하지 않을 뿐만 아니라 그것을 파악하는 것(즉, "봤다(기억난다)" 혹은 "못 봤다(기억나지 않는다)"라고 어느 한 방향으로 일관되게 반응하는 것)이 큰 의미를 지니지도 않는다. 그 만큼 반응 자체 혹은 성향보다는 기억자극 정보, 기억 흔적이나 증거의 인출 여부가 중요시되어왔다.

둘째, 무엇보다도 우리의 일반적인 기억 과정은 오랫동안 정보처리 관점에서 순향적인 절차로 인식되어왔다. 즉, 기억정보의 처리는 1) 부호화, 2) 저장, 3) 인출이라는 시간적 흐름을 반영하는 인지과정인데, 마치 컴퓨터 키보드를 통해 정보가 입력되듯이 부호화 과정은 외부 감각정보를 받아들이고 이것이 기억에 남을 경험의 시작점 역할을 한다. 컴퓨터 시스템을 이용한 비유를 즐겨하는 인지과학자들은 이렇게 입력된 감각정보들이 컴퓨터 하드드라이브와 같은 저장소에 저장된다고 비유하는데, 인간의 정보 저장과정이 컴퓨터와 다른 한 가지 특징은 단순한 정보 유지 기능뿐 아니라 의미를 부여하고, 재조직하는 과정이 일어나는 능동적인 유지과정을 거친다는 점이다. 이 기간이 지나고 나면 컴퓨터 화면에 필요한 정보를 불러오듯이 유의미한 기억 정보들을 인출하게 된다. 이러한 시스템 모형은 우리 기억정보의 인지적 정보처리과정을 매우 쉽고 간결하게 설명해준다는 장점을 지니고 있지만, 기억 과정이 다음 단계인 고차원적 추론이나 문제해결과정을 지지해주는 수단으로써, 결과물로 얻게 되는 보상, 그에 따라 발생하는 동기에 의해 영향을 받을 수 있음은 간과하고 있다. 즉, 앞서 언급한 목표지향적 성격을 가지고 각종 의사결정 과정과 계획 수립 등을 위해 작용하는 과정은 설명하지 않고 있는데, 이는 순향적인 정보처리과정을 뒤집고, 목표와 결과물을 염두에 두고 현재의 정보처리를 수행하는, 역향적인 과정을 나타내는 과정이 인지적으로 부자연스럽게 여겨졌고 무엇보다 순향적 정보처리 과정에 대한 연구가 더많이 이루어져왔기 때문이다.

셋째, 피드백/보상정보를 기반으로 한 행동은 대개 외현지식보다는 학습행동에 민감한 것으로 알려진 동물들을 이용해 연구되었고, 선조체와 도파민의 기능이 저하된 Parkinson환자 등의 피드백기반 암묵학습,

혹은 절차적 학습과정에 대한 연구를 통해 그 특징을 살펴볼 수 있었다. 이 연구들에서는 보상이 반복적으로 주어짐에 따라 (1) 점증적으로 행동이 변화되고 (2) 암묵적으로 변화된 결과를 습득하게 되는 학습 과정으로서 목표지향적 행동의 특징을 설명해왔다(예, Gluck et al., 2002). 학습심리학과 인지심리학을 연구하는 두 다른 사조에서는 오랫동안 학습, 그리고 기억을 각각 기저핵(Basal Ganglia)과 해마(Hippocampus)에 기반한 지식습득으로 구분하기도 하며, 이 프레임 안에서 '암묵적', '외현적'이라는 표현을 사용할 만큼 두 기제의 독립성을 강조하였다(Poldrack et al., 2001).

하지만 최근 들어 '동기', '의욕', '의도'라는 심리개념을 매개로 보상학습과 외현기억의 연결고리를 탐색하는 연구가 늘어났고, 특히 뇌신경기제를 통한 연구도 상당수 진행되었고 기저핵의 선조체가 외현기억과정에 관여하고 있음을 발견한 연구가 늘어나고 있다(Cohn et al., 2010; Han et al., 2010; Long et al., 2010; Moustafa & Gluck, 2011; Poldrack & Foerde, 2008; Shohamy & Adcock, 2010). 본 개관을 통해 제시하는 보상정보처리와 외현기억 기제의 상호작용은 크게 세 가지가 있다. 첫째, 보상영역의 활성화는 도파민 신경전달물질의 분비를 자극하여 해마 등 정보처리 영역에서의 효율적 자극 부호화를 유도한다. 둘째, 선조체 영역의 활성화는 기억인출에 따른 목표인지행동의 달성을 유도한다. 셋째, 보상정보는 반응 경향성의 변화를 유도하여 더 많은 기억인출반응을 유발한다.

3.1. 보상정보의 효율적 기억 부호화 유도

Adcock등의 최근 신경영상학 연구는 외현적인 기억정보를 부호화할때 과제와 연결된 금전적 보상, 동기 부여가 기억 수행에 미치는 영향과 그 때의 뇌활성화 패턴을 탐색하였다(Adcock et al., 2006). 이 연구자들은 기억해야할 시각 자극들을 제시하기에 앞서 두가지 금전적 인센티브 조건을 참가자들에 알려주었는데, 한 조건에서는 실험 화면에 제시될 기억자극의 가치가, 즉 성공적으로 제시된 자극을 기억할 경우 받게 되는 돈의 액수가 큰 조건 (각 아이টে마다 \$5)이, 다른 조건에서는 상대적으로 액수가 작은 조건 (\$10)임을 예고하는 단서 (지폐, 동전 그림)가 기억 자극이 제시되기 직전 짧게 제시되었다. 참가자들은 높은 인센티브 조건 단서 후에 제시되는 자극들을 낮은 인센티브

조건에 비해 유의미하게 더 많이 기억하는 양상을 보였고 이는 보상과 연결된 정보에 대한 처리수준을 높여주는 '접근 동기'가 기억을 향상시킨 것이라고 볼 수 있다(Shah et al., 1998).

이 연구가 중요한 이유는 동기와 기억과정의 연결고리를 신경학적 기제를 보여줬다는 데 있는데, 보상정보에 민감한 중뇌, 복측선조체 등의 도파민 회로영역의 활성화와 기억중추로 알려진 해마영역의 활성화가 민감하게 연결되어 있음을 확인하였다. 특히 이 두 영역의 기능적인 연결성은 기억해야 할 자극이 물리적으로 화면에 제시되기도 전인, 인센티브정보 단서만 화면에 제시되는 순간부터 강화되기 시작하였는데, 심리적인 기대치, 보상에 대한 예상만으로도 기억영역인 해마의 활성화가 증가하고, 이는 부여된 동기가 정보처리의 효율성을 극대화하기 위해 기억중추를 준비시키는 것임을 보여주는 증거라고 할 수 있다.

실제 도파민이 기억중추인 해마의 기능을 조절한다는 증거는 동물을 사용한 해부학적 신경생리학적 연구들에서 지속적으로 발견되어왔다. 중뇌 도파민 뉴런은 직접적으로 해마와 주변 MTL영역을 자극하며(Gasbarri et al., 1994; Samson et al., 1990), 해마 또한 중뇌영역의 도파민분비 신호를 직접 자극하고 다시 도파민을 수용하게 되는 순환적 관계 또한 밝혀진바 있다(Lisman & Grace, 2005). 최근 fMRI연구는 해마영역이 휴지상태(Resting-State)에도 도파민 경로의 여러 영역들과 기능적으로 연결된 활성화를 보임을 밝힌바가 있다(Kahn & Shohamy, 2012).

접근 동기(Approach Motivation) 뿐 아니라, 동기의 다른 한 축인 회피 동기(Avoidance Motivation) 또한 외현기억 형성에 영향을 줄 수 있다는 최근 연구도 보고되었다(Murty et al., 2012). 이들은 기억자극을 보여주면서 보상대신 망각 시 주어질 수 있는 처벌(예, 불쾌감을 유발할 수 있는 전기 자극) 정보를 참가자들에게 제시하였는데, 이러한 정보는 참가자들로 하여금 처벌을 받지 않기 위해 보다 정확한 기억 부호화를 하도록 유도할 수 있음을 발견하였다. 또한, 동기화된 기억 과정은 비단 아이템 자체에 대한 기억뿐만 아니라 출처와 맥락에 대한 정보의 습득과도 연결되어 행동적으로 수행의 우세를 보이게 된다(Wittman et al., 2005). 보상과 동기에 의해 기억 부호화에 대한 의욕이 고취되고, 의도적인 기억 저장이 일어나는 과정은 사실 매우 자연스럽고 누구나 쉽게 이해하는 과정

이다. 실제 수많은 교육학 연구들이 학습수행을 증진시키기 위해 어떤 보상정보를 사용해야하는지에 대해 매우 오랫동안 연구해왔으며 유럽과 미국의 고등학교에서는 학업성적에 따른 금전적 인센티브를 학기초에 학생들에게 약속하기도 한다(Singer-Vine, 2008). 또한, 업무 수행 향상을 위한 인센티브 제도의 개발 또한 인사관리에 있어 큰 관심을 받고 있는 주제 중 하나이다.

3.2. 목표지향적 기억인출과 선조체

2000년대 초반 진행된 재인기억 연구에서는 성공적인 기억 인출과정이 도파민 회로의 일부이며 보상영역으로 알려진 선조체의 활성화를 유발한다는 사실을 발견하였다. 다양한 실험 패러다임을 사용한 기억과제들에서 반복적으로 보상회로 영역(Caudate, Putamen, Nucleus Accumbens)의 활성화가 관찰되었고(Achim & Lepage, 2005; Iidaka et al., 2006; McDermott et al., 2009; Spaniol et al., 2009; von Zerssen et al., 2001), 기억인출과정의 신경기제를 분석한 메타분석 연구들에서도 일관되게 이들 영역의 활성화가 발견되었다(Kim, 2013; Spaniol et al., 2009). 특징적인 부분은 이들 연구에서 실제 외적 보상정보가 주어진 경우는 없다는 것인데, 이에 대해 von Zerssen 등의 연구자들은 성공적인 기억인출에 따른 자기보상적 활성화라고 해석하였다(von Zerssen et al., 2001). 즉, 마치 우리가 어려운 퀴즈 문제의 정답을 맞혔을 때 느끼는 희열과 뿌듯함처럼, 기억을 더듬어 정보 인출에 성공하였을 때 느끼는 주관적인 만족감과 보상심리가 도파민 신경회로로 대표되는 뇌의 보상중추에서 나타나는 것이라고 추측하였다. 실제로 최근 fMRI 연구에서(Han et al., 2010) 이 영역이 내재된 주관적 보상가, 목표 달성에 대한 성취감을 반영한다는 보다 직접적인 결과를 보여주기도 하였다. Han 등의 연구에서는 각각 다른 회기에서 성공적인 인출 혹은 성공적인 기각에 따로 보상을 제시하였는데, 즉, 한 회기에서는 성공적인 인출("본적 있다") 반응을 내리는 것에, 다른 회기에서는 성공적으로 기각("본적 없다")하는 반응에 선택적으로 \$1씩의 보상을 지불하였다(Han et al., 2010). 두 회기 모두 기억자극의 부호화는 동일하게 이루어졌고, 기억 정보 입력단계에서는 보상정보에 대한 어떤 정보나 설명도 주어지지 않았기 때문에, 부호화 조건의 동기수준의 차이가, 그리고 그에 따른 부

호화의 차이가 기억 반응에 대한 활성화 수준의 차이를 유발하였다고는 볼 수 없다.

하지만 이미 정보들의 부호화, 입력이 끝난 후 동기와 보상정보가 기억자극을 인출해내는데 어떻게 영향을 줄 수 있을까? Han 등(2010)의 연구에서 참가자들은 보상이 주어지는 조건에서 더 우세하고 빠른 행동 반응성향을 나타내었는데, 목표(보상)와 연합된 반응에 대해 더욱 빠르게 의사결정을 내린 것으로 볼 수 있다. 이 연구에서는 선조체 영역의 활성화를 각 목표 반응(인출, 기각)에 따라 구분하였는데, 상부선조체인 미상핵(Caudate Nucleus)의 활성화가 각 회기의 목표 반응에 따라 선택적으로 크게 증가하는 모습을 보였다. 즉 보상영역의 활성화가 기억 인출 행동의 신경기전에도 밀접하게 연결되어 있음을 보여주는 결과라고 할 수 있다.

이러한 활성화는 잠재적으로 다음 기억인출행동을 결정하는 중요한 역할을 하는데, 최근 많은 행동 실험들에서 재인기억의 인출 과정에서 편향된 인지 피드백을 제공함으로써 사람들이 피드백에 수반된 보상정보나 목표반응정보에 민감한지, 그리고 그 결과 편향된 피드백 확률만큼 행동으로 나타나는 반응성향에도 변화가 유발되는지를 살펴보았다(Han & Dobbins, 2008; Han & Dobbins, 2009). 이 연구들에서는 기억 반응 중 특히 오반응, 즉 오경보(False alarm)와 탈루(Miss)에 대해서 선택적으로 일정 확률 이상의 편향된 피드백을 제공하였다. 보다 구체적으로 실험자들은 참가자들의 실수반응에 대해서도 긍정적인 피드백을 제공해 옳은 반응이라는 피드백을 반복해서 제공했는데, 오경보와 탈루에서 이러한 긍정피드백을 제공한 경우, 각각 '본적 있다', '본적 없다'의 기억반응을 더 많이 나타내는 편향된 반응 성향을 발견하였다. 이러한 편향된 피드백조작에 대해 참가자들은 의식적으로 인식하지 못하였는데, 즉, 참가자들은 자신의 기억 수행에 대하여 긍정적인 피드백이 높은 확률로 주어지는 반응을 암묵적이지만 점증적으로 더 많이 선택하는 점증적 학습패턴을 보였다.

이러한 기억인출의 행동적 변화는 비단 재인기억과제에서만 나타나는 것이 아니라, 보다 많은 질적 정보인출을 필요로 하는 실험 패러다임에서도 나타났다(Han, 2009a).

흥미로운 사실은 기억 반응 성향을 결정하는데에 보상민감도를 나타내는 성격 특질과 같은 개인적 성향차

이가 큰 영향을 미친다는 것인데, 유사한 실험 패러다임을 사용해 진행된 기억편향 연구에서는(Han, 2009b), 편향된 피드백에 반응하여 판단의 방향을 선택하는 학습이 유도되는 과정에서 불안 회피/강화 추구 성향의 수준을 반영하는 BIS/BAS(Behavioral Inhibition System/Behavioral Approach System, Carver & White, 1994) 개인차가 반응편향의 학습정도를 유의미하게 예측할 수 있음을 보여주었다. 즉, 보상과 강화를 추구하는 참가자일수록 편향된 피드백 (틀렸음에도 맞았다고 긍정적인 피드백을 지속적으로 제공해 편향 유도)에 따라 반응성향을 변화시키는 정도가 크게 나타났다. 앞서 언급한 Han 등(2010)의 신경영상학 연구에서도 개인의 보상추구 성향 기질의 수준(BAS)과 보상영역의 뇌활성화 수준 간에 유의미한 상관관계가 나타났는데, 이는 보상정보가 기억인출에도 영향을 미칠 수 있음을 보여주는 것이다. 특히 흥미로운 점은 실제로는 아무런 보상정보가 주어지지 않았던 일반적인 재인지 역과제 중 성공적인 인출이 일어난 경우에도 이 활성화-기질특질 사이의 관계가 유지되었다는 점이다. 즉, 참가자들은 성공적으로 기억정보를 인출해내는 것에서 심리적으로 내적 보상감을 느끼고 이는 직접적으로 보상중추의 신경활성화 정도와 관련이 있음을 보여주는 것이다.

요약컨대, 성공적인 기억의 부호화와 인출은 수많은 정보 중에 중요한 정보를 선택적으로 골라내는 과정과 밀접한 연관이 있고, 정보의 중요성은 기억이라는 인지 과정을 통해 유발된 각종 의사결정과 문제해결 과정에서 얻게 되는 결과물, 특히 보상이 있는 결과물에 따라 부호화와 인출과정의 반응패턴도 달라지게 된다. 이는 앞서 언급한 전통적인 기억의 정보처리 과정 순서, 즉 순향적이고 시간의 흐름에 따른 처리과정과는 달리 미래에 얻게 될 정보가 역향적으로 현재의 인지처리에 영향을 줄 수 있다는 점에서도 흥미로운 관계라고 볼 수 있다.

4. 결론

인간의 학습과 기억이란 삶속에서 필수적인 행동으로 여겨지는 인지과정인데, 학습과 기억과정을 통해 과거의 경험을 인출하고 이를 바탕으로 현재의 행동을 선택하거나 중요한 의사결정을 내리며, 미래지향

적인 행동을 의도하고 계획할 수 있게 해주기 때문이다. 나아가 경험에 대한 기억은 우리의 계획과 의도된 행동들을 더 효율적인 것으로 개선하거나 공고화시켜주는, 즉 학습을 유발하는 데에도 중요한 기능을 담당하고 있다. 이처럼 기억을 더듬어 더 나은 결정을 내리려는 행동은, 우리 스스로가 행동의 결과물을 신경쓰고 있음을 보여주는 것인데, 더 나은 결과물, 보상, 만족감 등의 목표물을 정해, 이를 얻고자 하는 의욕적인 인지 과정이 우리 행동과 의사결정의 핵심이 될 수 있다.

실제 우리가 현재 수행하고 있는 행동을 결정하는데에는 크든 작든 목표 혹은 결과물이라는 미래의 사실이 영향을 끼치게 되는데, 본 논문에서는 이와 같은 '목표지향적 행동(Goal-directed action/behavior)'에 대한 개념을 소개하고, 목표지향적 행동에 대한 전통적인 학습이론과 이후 발전된 다양한 강화학습모델에 대해 개관하였다. 이와 더불어 최근 여러 신경영상학, 신경심리학 연구에서 발견된 신경기제를 함께 살펴보고, 이 기제가 인간의 고위인지과정인 일화기억과정에도 영향을 미칠 수 있음을 논의하였다.

목표지향적 학습과 기억행동은 과거 연합, 혹은 부호화된 경험정보가 미래의 적응적 행동을 도와줄 수 있다는 데에서 중요한데, 본 개관을 통해 살펴본 중요성은 크게 두 가지로 정리할 수 있다. 성공적인 기억의 부호화와 인출은 수많은 정보 중에 중요한 정보를 선택적으로 골라내는 과정과 밀접한 연관이 있고, 정보의 중요성은 기억이라는 인지 과정을 통해 유발된 각종 의사결정과 문제해결 과정에서 얻게 되는 결과물, 특히 보상이 있는 결과물에 따라 부호화와 인출과정의 반응패턴도 달라지게 된다. 이는 앞서 언급한 전통적인 정보처리 과정의 순서, 즉 순향적이고 시간의 흐름에 따른 처리과정과는 달리 미래에 얻게 될 보상 정보가 역향적으로 현재의 인지처리에 영향을 줄 수 있다는 점에서도 새로운 관점이라고 볼 수 있다. 또한 최근의 신경생물학적 연구의 결과들은 목표지향적 행동은 도파민을 매개로 선조체를 비롯한 보상중추와 고위인지기능 및 기억과정을 담당하는 해마 학습 기제간의 상호작용을 통해 적응적 행동이 유발, 향상될 수 있음을 지지한다. 이러한 관점은 학습과 기억 과정을 필요, 의지, 동기 등 심리적 요인에 따라 통제될 수 있고 변화할 수 있는 의지적 과정으로 이해하는데 중요한 틀을 제공할 것으로 본다.

REFERENCES

- Achim, A. M., Lepage, M. (2005). Dorsolateral prefrontal cortex involvement in memory post-retrieval monitoring revealed in both item and associative recognition tests. *Neuroimage*, 24, 1113-1121.
- Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., & Gabrieli, J. D. (2006). Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron*, 50(3), 507-517.
- Baird, L. C. (1993). *Advantage updating* (Tech. Rep. No. WL-TR-93-1146). Dayton, OH: Wright- Patterson Air Force Base.
- Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology*, 37(4), 407-419.
- Berns, G. S., McClure, S. M., Pagnoni, G., & Montague, P. R. (2001). Predictability modulates human brain response to reward. *The Journal of Neuroscience*, 21(8), 2793-2798.
- Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *Journal of Personality and Social Psychology*, 67(2), 319-333.
- Cohn, M., Moscovitch, M., & Davidson, P. S. R. (2010). Double dissociation between familiarity and recollection in Parkinson's disease as a function of encoding tasks. *Neuropsychologia*, 48, 4142-4147.
- Corbit, L. H., Muir, J. L., & Balleine, B. W. (2001). The role of the nucleus accumbens in instrumental conditioning: evidence of a functional dissociation between accumbens core and shell. *The Journal of Neuroscience*, 21(9), 3251-3260.
- Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69(6), 1204-1215.
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12), 1704-1711.
- Dayan, P., & Balleine, B. W. (2002). Reward, motivation, and reinforcement learning. *Neuron*, 36(2), 285-298.
- Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural networks*, 12(7), 961-974.
- Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in Parkinsonism. *Science*, 306(5703), 1940-1943.
- Gasbarri, A., Packard, M. G., Campana, E., & Pacitti, C. (1994). Anterograde and retrograde tracing of projections from the ventral tegmental area to the hippocampal formation in the rat. *Brain Research Bulletin*, 33(4), 445-452.
- Glascher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, 66(4), 585-595.
- Gluck, M. A., Shohamy, D., & Myers, C. (2002). How do people solve the "weather prediction" task?: individual variability in strategies for probabilistic category learning. *Learning & Memory*, 9(6), 408-418.
- Han, S. (2009a). Exploring episodic decision variable using adaptive biased feedback procedure: characteristic of decision axis during recollective judgment. *Korean Journal of Cognitive and Biological Psychology*, 21(1), 23-42.
- Han, S. (2009b). Trait individual difference of reinforcement-based decision criterion learning during episodic recognition judgments. *Korean Journal of Cognitive Science*, 20(3), 357-381.
- Han, S., & Dobbins, I. G. (2008). Examining recognition criterion rigidity during testing using a biased-feedback technique: evidence for adaptive criterion learning. *Memory & Cognition*, 36(4), 703-715.

- Han, S., & Dobbins, I. G. (2009). Regulating recognition decisions through incremental reinforcement learning. *Psychonomic Bulletin & Review*, 16(3), 469-474.
- Han, S., Huettel, S. A., Raposo, A., Adcock, R. A., & Dobbins, I. G. (2010). Functional significance of striatal responses during episodic decisions: recovery or goal attainment? *The Journal of Neuroscience*, 30, 4767-4775.
- Iidaka, T., Matsumoto, A., Nogawa, J., Yamamoto, Y., Sadato, N. (2006). Frontoparietal network involved in successful retrieval from episodic memory. Spatial and temporal analyses using fMRI and ERP. *Cerebral Cortex*, 16, 1349-1360.
- Kahn, I., & Shohamy, D. (2012). Intrinsic connectivity between the hippocampus, nucleus accumbens, and ventral tegmental area in humans. *Hippocampus*, 23, 187-192.
- Killcross, S., & Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cerebral Cortex*, 13(4), 400-408.
- Kim, H. (2013). Differential neural activity in the recognition of old versus new events: an activation likelihood estimation meta-analysis. *Human Brain Mapping*, 34, 814-836.
- Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory*, 1(2), 106-120.
- Knutson, B., & Cooper, J. C. (2005). Functional magnetic resonance imaging of reward prediction. *Current Opinion in Neurology*, 18(4), 411-417.
- Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *The Journal of Neuroscience*, 21(16), RC159.
- Kumaran, D., Summerfield, J. J., Hassabis, D., & Maguire, E. A. (2009). Tracking the emergence of conceptual knowledge during human decision making. *Neuron*, 63(6), 889-901.
- Lisman, J. E., & Grace, A. A. (2005). The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron*, 46(5), 703-713.
- Long, N. M., Oztekin, I., & Badre, D. (2010). Separable prefrontal cortex contributions to free recall. *The Journal of Neuroscience*, 30, 10967-10976.
- McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, 38(2), 339-346.
- McDermott, K. B., Szpunar, K. K., & Christ, S. E. (2009). Laboratory-based and autobiographical retrieval tasks differ substantially in their neural substrates. *Neuropsychologia*, 47, 2290-2298.
- Moustafa, A. A., & Gluck, M. A. (2011). Computational cognitive models of prefrontal-striatal-hippocampal interactions in Parkinson's disease and schizophrenia. *Neural Networks*, 24, 575-591.
- Murty, V. P., LaBar, K. S., & Adcock, R. A. (2012). Threat of punishment motivates memory encoding via amygdala, not midbrain, interactions with the medial temporal lobe. *The Journal of Neuroscience*, 32(26), 8969-8976.
- Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3), 139-154.
- Niv, Y., Joel, D., & Dayan, P. (2006). A normative perspective on motivation. *Trends in Cognitive Sciences*, 10(8), 375-381.
- O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, 38(2), 329-337.
- O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, 304(5669), 452-454.
- Pagnoni, G., Zink, C. F., Montague, P. R., & Berns, G. S. (2002). Activity in human ventral striatum locked to errors of reward prediction. *Nature Neuroscience*, 5(2), 97-98.
- Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. London: Oxford Univ. Press.

- Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience and Biobehavioral Review*, 32, 197-205.
- Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Moyano, J. C., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, 414(6863), 546-550.
- Redish, A. D., Jensen, S., & Johnson, A. (2008). A unified framework for addiction: vulnerabilities in the decision process. *Behavioral and Brain Sciences*, 31(4), 415-436.
- Samson, Y., Wu, J. J., Friedman, A. H., & Davis, J. N. (1990). Catecholaminergic innervation of the hippocampus in the cynomolgus monkey. *Journal of Comparative Neurology*, 298(2), 250-263.
- Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, 36(2), 241-263.
- Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review of Neuroscience*, 23(1), 473-500.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306), 1593-1599.
- Shah, J., Higgins, E. T., & Friedman, R. (1998). Performance incentives and means: how regulatory focus influences goal attainment. *Journal of Personality and Social Psychology*, 74, 285-293.
- Shohamy, D., & Adcock, R. A. (2010). Dopamine and adaptive memory. *Trends in Cognitive Science*, 14, 464-472.
- Shohamy, D., & Wagner, A. D. (2008). Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron*, 60(2), 378-389.
- Shohamy, D., Myers, C. E., Grossman, S., Sage, J., & Gluck, M. A. (2005). The role of dopamine in cognitive sequence learning: evidence from Parkinson's disease. *Behavioural Brain Research*, 156(2), 191-199.
- Singer-Vine, J. (2008, Aug 21). When schools offer money as a motivator. *The Wall Street Journal*. Retrieved from <http://online.wsj.com/article/SB121928822683759447.html>
- Skinner, B. F. (1938). *The behavior of organisms*. New York, NY: Appleton-Century.
- Smith, A. C., & Brown, E. N. (2003). Estimating a state-space model from point process observations. *Neural Computation*, 15(5), 965-991.
- Smith, A. C., Frank, L. M., Wirth, S., Yanike, M., Hu, D., Kubota, Y., & Brown, E. N. (2004). Dynamic analysis of learning in behavioral experiments. *The Journal of Neuroscience*, 24(2), 447-461.
- Spaniol, J., Davidson, P. S., Kim, A. S., Han, H., Moscovitch, M., & Grady, C. L. (2009). Event-related fMRI studies of episodic encoding and retrieval: metaanalyses using activation likelihood estimation. *Neuropsychologia*, 47, 1765-1779.
- Sutton, R. S., & Barto, A. G. (1990). Time-derivative models of Pavlovian reinforcement. In M. Gabriel & J. Moore (Eds.) *Learning and computational neuroscience: Foundations of adaptive networks* (pp. 497-537). Cambridge: MIT Press.
- Takikawa, Y., Kawagoe, R., & Hikosaka, O. (2004). A possible role of midbrain dopamine neurons in short-and long-term adaptation of saccades to position-reward mapping. *Journal of Neurophysiology*, 92(4), 2520-2529.
- Thorndike, E. L. (1911). *Animal intelligence: Experimental studies*. New York: Macmillan.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55(4), 189-208.
- von Zerssen, G. C., Mecklinger, A., Opitz, B., von Cramon, D. Y. (2001). Conscious recollection and illusory recognition: an event-related fMRI study. *European Journal of Neuroscience*, 13, 2148-2156.
- Wimmer, G. E., & Shohamy, D. (2012). Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science*, 338(6104), 270-273.
- Wimmer, G. E., Daw, N. D., & Shohamy, D. (2012). Generalization of value in reinforcement learning by humans. *European Journal of Neuroscience*, 35(7),

1092-1104.

- Wittmann, B. C., Schott, B. H., Guderian, S., Frey, J. U., Heinze, H. J., & Düzel, E. (2005). Reward-related fMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron*, 45(3), 459-467.
- Yerkes, R. M., & Morgulis, S. (1909). The method of Pawlow in animal psychology. *Psychological*

Bulletin, 6(8), 257.

- Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, 7(6), 464-476.

원고접수: 2013.07.10

수정접수: 2013.08.13

게재확정: 2013.08.16