

# **Appendices**

## THEME A

# 포유류가 스스로를 통제하는 법

---

일반적으로 신경과학 교과서에서는 이마엽새곁질의 기능이 주의, 작업기억, 실행통제(executive control), 앞에서 살펴본 계획수립이라는 네 가지라고 설명한다. 이 기능들을 하나로 이어주는 주제가 무엇인지 늘 불분명했다. 한 가지 구조가 서로 별개인 듯한 이 모든 역할을 담당한다는 것이 이상해 보인다. 하지만 진화의 렌즈를 통해 들여다보면 이런 기능이 모두 긴밀하게 연관되어 있음을 이해할 수 있다. 이는 모두 새곁질 시뮬레이션의 통제를 서로 다르게 적용한 것이다.

오리로도 보이고 토끼로도 보였던 모호한 그림을 기억하는가? 오리로 보였다 토끼로 보였다 할 때 당신의 시각곁질이 각각의 해석을 바꾸게 만드는 것은 aPFC다. aPFC는 눈을 감고 있을 때 오리의 내적 시뮬레이션을 촉발할 수 있다. 당신이 눈을 뜨고 오리일 수도 토끼일 수도 있는 그림을 보고 있는 동안에도 aPFC는 동일한 메커니즘을 이용해 오리의 내적 시뮬레이션을 적용하려 애쓴다. 단지 눈을 감았을 때는 시뮬레이션이 제약 없이 이루어지고, 눈을 떴을 때는 당신의 눈에 보이는 것과 일치하도록 제약을 받는다는 점만 다르다. aPFC가 시뮬레이션을 촉발할 때 현재 들어오는 감각 입력이 제한되지 않으면 ‘상상’이라 부르고, 현재 들어오는 감각 입력이 제한되면 ‘주의’라고 부른다. 두 경우 모두 aPFC는 원론적으로 같은 일을 한다.

주의는 왜 있는 것일까? 생쥐가 시뮬레이션을 상상한 후에 행동의 순서를 선택할 때는 경로를 따라 달리는 동안 그 계획을 반드시 고수해야 한다. 그런데 이것이 말처럼 쉽지가 않다. 상상한 시뮬레이션이 완벽하지 않기 때문이다. 생쥐는 실제 환경에서 예측하지 못했던 풍경, 냄새, 윤곽 등을 접하게 될 것이다. 다시 말해 바닥핵이 경험했던 대리 학습과 계획이 실행되는 동안에 접하는 실제 상황이 서로 다를 것이고, 따라서 바닥핵이 의도했던 행동을 올바르게 완수할 수 없을지도 모른다.

aPFC가 이 문제를 해결할 한 가지 방법은 주의를 이용하는 것이다. 생쥐의 바닥핵이 시행착오를 통해서 오리를 보면 달아나고 토끼를 보면 그것을 향해 달려가도록 학습했다고 해보자. 이 경우 바닥핵은 오리나 토끼를 봤을 때 새곁질이 어떤 패턴을 보내느냐에 따라 정반대로 반응할 것이다. 만약 aPFC가 이전에 토끼를 보고 그것을 향해 달려가는 상상을 했다면, 주의를 이용해서 쥐가 이 모호한 그림을 볼 때 오리가 아니라 토끼가 보이도록 만들어 바닥핵의 선택을 통제할 수 있다.

진행하고 있는 행동을 통제하려면 작업기억이 필요한 경우가 많다. 작업기억이란 아무런 감각 단서가 없는 상황에서도 표상을 유지하는 것을 말한다. 상상한 경로와 과제에는 기다림이 수반되는 경우가 많다. 예를 들어 설치류가 나무들 사이에서 견과류를 찾을 때는 자기가 이미 살펴본 나무가 어느것인지 기억할 수 있어야 하는데 이때 aPFC가 필요한 것으로 밝혀졌다. 또한 그런 과제를 하는 동안 aPFC는 외부의 단서가 전혀 없는 상황에서도 계속 활성화된 상태를 유지하는 ‘지연 활동(delay activity)’을 나타낸다. 만약 이렇게 지연된 동안 설치류의 aPFC 활성을 억제하면 설치류는 기억을 통해 그런

과제를 수행하는 능력을 상실한다. 이런 과제를 수행하는 데 aPFC가 필요한 이유는 작업기억이 주의, 계획수립과 동일한 방식으로 기능하기 때문이다. 이것은 내적 시뮬레이션을 끌어들이는 작업이다. 뭔가를 머릿속에 붙잡아두는 작업기억은 한마디로 aPFC가 더 이상 필요 없을 때까지 내적 시뮬레이션을 계속 이끌어내려 노력하는 것이다.

계획수립, 주의, 작업기억에 더해서 aPFC는 진행 중인 행동을 좀 더 직접적으로 통제할 수 있다. 편도체를 억제할 수도 있다. aPFC에서 편도체를 둘러싼 억제성 신경세포로 투사되는 신경이 있다. 상상한 계획을 충족하지 못하게 막으려고 할 수 있다. 이를 통해 심리학자들이 말하는 행동억제 behavioral inhibition, 의지력 willpower, 자기통제의 진화가 시작되었다. 순간순간 느끼는 갈망(편도체와 바닥핵이 통제)과 우리가 더 나은 선택임을 알고 있는 것(aPFC가 통제) 사이에는 긴장 상태가 지속된다. 의지력이 발휘되는 순간에는 편도체가 주도하는 갈망을 억제할 수 있다. 하지만 의지력이 발휘되는 순간에는 편도체가 승리한다. 이것이 사람들이 지치거나 스트레스를 받았을 때 충동적으로 행동하는 이유다. aPFC를 운용하는 데는 에너지가 많이 소비된다. 그래서 지치거나 스트레스를 받으면 aPFC가 편도체를 효과적으로 통제하지 못한다.

요약해보자. 계획수립, 주의, 작업기억은 모두 aPFC가 통제한다. 이 세가지 모두 원론적으로는 동일한 기능으로 어떤 시뮬레이션을 만들지 선택하려고 애쓰는 뇌의 노력이 다른 모습으로 발현된 것이다. aPFC는 어떻게 행동을 ‘통제’할까? 행동 그 자체를 통제하는 것이 아니라, 바닥핵에게 어느 선택이 더 나은지 대리로 보여주고 바닥핵에 전달되는 정보를 걸러내 바닥핵이 올바른 선택을하도록 설득한다. aPFC는 명령이 아니라 시연을 통해 행동을 통제한다.

이런 과정을 통해 얻게 되는 이점은 반사반응을 억제해서 ‘더 영리한’ 선택을 해야 하는 과제를 대상으로 포유류와 도마뱀 등 여러 척추동물의 수행성과를 비교해보면 알 수 있다. 도마뱀을 미로에 넣어 맛있는 먹이가 나오는 빨간색 불을 향해 가고 맛없는 먹이가 나오는 초록색 불을 피하도록 훈련시키려면 이런 단순한 과제를 학습하는 데도 수백 번 시행착오를 겪어야 한다. 도마뱀은 선천적으로 초록색을 좋아하기 때문에 그것을 극복하는데 오랜 시간이 걸린다. 도마뱀에게는 잠시 멈춰서 여러 가지 선택지를 대리해서 고려하는 새걸질이 없기 때문에 시행착오를 수없이 반복해 겪으며 과제를 학습할 수밖에 없다. 반면 쥐는 자신의 선천적인 반응을 억제하는 법을 훨씬 빨리 배운다. 쥐의 aPFC를 손상시키면 이런 장점은 사라진다.

초기 포유류에게는 내부에서 만든 세계 모델을 대리로 탐험하고 상상한 결과를 바탕으로 선택하며 일단 선택한 후에는 상상한 계획을 고수할 수 있는 능력이 있었다. 이들은 언제 미래를 시뮬레이션할지도 똑똑하게 선택해서 검색 문제를 극복할 수 있었다. 이들은 목표가 있는 최초의 조상이었다.

## THEME B

# 정신 탐색을 위한 경로 및 지도 내부화

---

나는 움직인다, 고로 존재한다.

—무라카미 하루키(Haruki Murakami)

태초에 행동이 있었다.

—괴테(Goethe)의 ‘파우스트(Faust)’에서

듣지 못함은 듣는 것만 못하고, 듣는 것은 보는 것보다 못하고,  
보는 것은 아는 것만 못하고, 아는 것은 행동하는 것만 못하다.  
진정한 배움은 실천에 옮길 때까지 계속된다.

—공자

오글...;; 오글...♡

포유류에서 어떻게 그러한 조합 능력이 나타났는지 조사하려면 ‘구조가 기능을 제한한다’는 격언을 기억하자. 이 격언은 특히 해마 시스템에서 사실이다. 그것의 구조와 연결성에 대한 어떤 것도 그것이 공간 탐색을 위해 진화되었다는 것을 암시하지 않는다. 광범위하게 평행한 과립 세포(granule cell) 인터페이스와 강력하게 반복되는 CA2/3 흥분 시스템, 두 개의 매우 다른 건축 설계는 장소와 대상을 분리하고 통합하는 데 이상적이다. 그러나 분리와 통합은 수많은 뇌 과정에 도움이 될 수 있는 기본적인 작업이다. 따라서 공간 탐색을 위해 처음에 진화된 해마 시스템은 다른 작업을 위해 용도도 변경될 수 있다.

## B.1 하나의 회로, 많은(겉보기) 기능

해마-내후각 시스템은 지형적으로 조직화된 큰 신피질과 양방향 통신을 한다. 설치류에서 신피질의 대부분은 운동 출력을 제어하고 감각 입력을 처리하는 역할을 한다. 대조적으로 영장류에서는 더 복잡한 기능을 계산하는 데 많은 부분이 사용된다. 해마의 무감각 표현은 포유류 진화 동안 신피질의 확대와 함께 점진적으로 증가했다(Figure B.1). 따라서 공간 정보의 계산은 뇌가 더 큰 동물의 해마에서 시간 제 작업일 뿐이다. 이 아이디어는 해마가 신피질의 어느 부분에서 받는 정보가 무엇이든 동일한 일반 계산이 수행될 것임을 시사한다. 다시 말해, 해마는 입력의 양상과 특성에 대해 ‘맹인’이다. 신피질의 메시지를 출처에 상관없이 동일한 방식으로 처리하고 판단을 반환한다. 공간, 시간 또는 기타 정보를 처리하는 것으로 보이는가의 여부는 주로 실험 설정에 의해 결정된다.

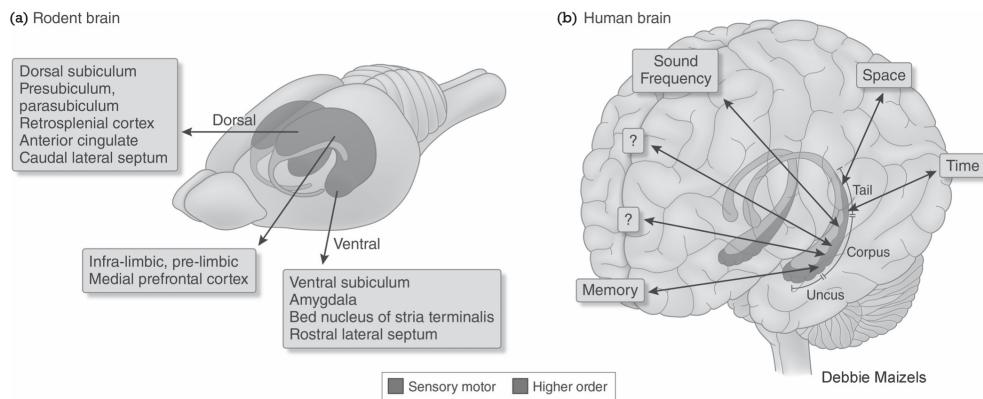


Figure B.1: 인간과 쥐의 뇌에 있는 해마의 상응 부위. 설치류 해마의 사분면은 영장류에서 점점 더 많은 고차 신피질의 봉을 따라잡기 위해 불균형적으로 확대되었다. 영장류 해마의 비교적 작은 꼬리 부분만이 시공간 영역과 소통한다. 이 꼬리는 설치류의 등쪽 중간 해마와 상응한다. 중격측두엽(Septotemporal) 축의 세그먼트에 대한 다양한 연결이 표시된다. 설치류 뇌의 대부분의 기록 및 조작은 등쪽 해마에서 수행되었다.

출처: Royer et al.(2010).

해마가 항법 기능만 했다면 현재 연결 항공편을 기다리고 있는 일본 나리타 공항에서 환경 신호의 도움 없이 어떻게 이 선을 기록할 수 있는지 미스터리로 남을 것이다. 이러한 매우 동일한 구조는 기억, 계획 및 상상과 같은 겉보기에 분리된 기능을 담당하기도 한다. 이 모든 기능은 외부 또는 신체에서 파생된 자극에 의존하기보다는 자체 생성된 세포 조립 순서에 의해 지원되어야 한다. 나중에 설명하겠지만 이러한 인지 기능은 물리적 환경에서 공간 탐색을 위해 처음 도입된 메커니즘에서 나온 정신적 탐색으로 생각할 수 있다.

## B.2 정신 탐색의 형태

난치성 간질을 완화하기 위해 두 해마를 수술로 제거한 유명한 환자 HM을 시작으로 해마와 관련 구조가 기억을 만드는 데 책임이 있다는 합의가 이루어졌다. 기억은 또 다른 가정된 해마 기능인 공간 탐색과 어떤 관련이 있는가? 인지 지도 이론의 선도적 지지자인 오키프는 “인간의 기본 공간 지도에 시간적 요소를 추가하는 것이 일화적 기억 시스템의 기초를 제공한다”고 추측했다. 한편으로, 이 제안은 설치류와 인간의 해마가 다른 일을 하며 설치류는 ‘시간적 요소’가 부족하다는 것을 암시한다. 반면에, 외부 충동에 의존하는 중심 구조인 인지 지도에 시간을 추가하는 것이 어떻게 우리가 소유하고 있는 가장 자기중심적인 것, 우리의 일화적 기억 모음에 대한 메커니즘을 제공할 수 있는지 전혀 명확하지

않다.

아마도 내비게이션 시스템이 환경이나 신체 신호에서 분리될 수 있다면 내재화된 성능이 기억을 뒷받침할 수 있을 것이다. 공간 탐색과 유사하게 해마 시스템 의존 기억의 두 가지 형태를 구별할 수 있다. 이것은 개인적인 경험(일화적 기억 또는 특정 사례에 대한 기억)과 암기된 사실(의미적 기억 또는 통계적 규칙성을 위한 기억)이다. 우리는 두 가지 유형을 의식적으로 인식하고 구두로 선언할 수 있으므로 함께 이라고 선언적 기억한다. ‘일화적’ 또는 ‘유연한’과 같은 용어는 인간이 아닌 동물의 기억을 나타내는 데 사용된다.

### B.3 일화 기억

개인의 기억은 우리 삶의 중요한 에피소드로 자아의 감정을 만들어 내는 개성의 유일한 원천이다. 그러한 1인칭(자기중심적) 맥락 의존적 에피소드를 다시 경험하려면 우리 자신을 시간과 공간으로 다시 투영해야 한다. 캐나다 토론토 대학교의 엔델 툴빙(Endel Tulving)은 이 가상의 뇌 작동에 대해 ‘정신적 시간 여행’이라는 용어를 만들었다. 정신적 시간 여행을 통해 과거와 미래를 알 수 있다. 시간을 거슬러 올라가는 것을 에피소드 회상이라고 하고 상상한 미래를 알 수 있다. 시간을 거슬러 올라가는 것을 에피소드 회상이라고 하고 상상한 미래로 여행하는 것을 계획이라고 할 수 있다. 물론 일부 사람들은 과거와 미래가 크게 다르며 이러한 구분은 기억과 계획에도 적용되어야 한다고 항의할 수 있다. 실제로 이러한 범주는 일반적으로 신경과학 핸드북의 여러 장에서 다루며 다양한 뇌 구조와 메커니즘에 할당된다. 그러한 뇌는 이런 식으로 사물을 보지 못할 수도 있다.

이러한 문제를 해결하기 위한 나의 첫 번째 제안은 추측 항법에 의해 물리적 공간에서 탐색을 위해 처음에 진화한 뇌 메커니즘이 에피소드 기억을 생성하고 회상하기 위해 ‘인지 공간’에서 탐색에 사용되는 것과 기본적으로 동일하다는 것이다. 두 번째 관련 제안은 지도 기반 탐색을 지원하도록 진화된 신경 알고리즘이 의미론적 지식을 생성, 저장 및 기억하는 데 필요한 알고리즘과 거의 동일하다는 것이다. 나의 세 번째 제안은 의미론적(동종 중심적) 지식의 생성은 추측 항법 탐색에 의한 지도 생성과 유사한 사전 자체 참조 에피소드 경험에 필요하다는 것이다(Figure B.2).

### B.4 의미론적 기억

자기중심적 일화기억의 관찰자 의존성과 달리 지식은 관찰자 독립적이다. 의미 기억은 장소 셀과 격자 셀이 지도에서 위치 좌표를 명시적으로 정의하는 방식과 유사하게 시간적 맥락과 무관하게 주변 세계의 대상, 사실 및 이벤트를 정의한다. 물리적 탐색과 정신적 탐색 사이의 깊은 관계는 그리스 응변가의 연습에서도 잘 설명된다. 그들은 각각의 방이 특정한 주제를 담고 있는 집의 다른 방을 걸어갔다고 상상함으로써 연설을 암기했다. 또한 전문 재즈 뮤지션은 음표의 풍경을 상상하고 콘서트에서 미리 계획된 경로를 탐색한다.

고유한 사건에 대한 일화적 기억의 1회 획득과 달리 의미론적 정보는 일반적으로 동일한 사물이나 사건을 반복적으로 만난 후에 학습된다. 초기에 에피소드 기억으로 인코딩된 정보는 점차 맥락적 특징을 잃어 전방향성인 장소 필드를 연상시키는 일반화되고 명시적이다. 즉, 장소 세포는 동물이 접근하는 방향에 관계없이 장소 필드에서 발화한다. 예를 들어, 운이 좋아서 무언가를 발견했다면 그 에피소드는 평생 동안 소중하게 남아 있다. 그러나 나의 연구실과 다른 사람들이 다른 각도에서 동일한 결론에 도달하여 당신의 발견을 확인하면 그것은 사실이 된다. 즉, 발견 및 확인의 조건에 관계없이 모든 사람이 동일한 일반적 방식으로 이해하는 명시적 지식이다.

의미 정보를 인코딩하는 데 에피소드 경험이 필요하다는 데 모든 사람이 동의하는 것은 아니다. 실제로 인간의 경우 언어 및 기타 외부화된 뇌 기능을 통해 개인적인 경험 없이도 의미적 지식을 빠르게 습득할 수 있다. 돌보는 사람이 물건의 이름을 제공할 수 있으며, 그러면 아이가 학습하게 된다.

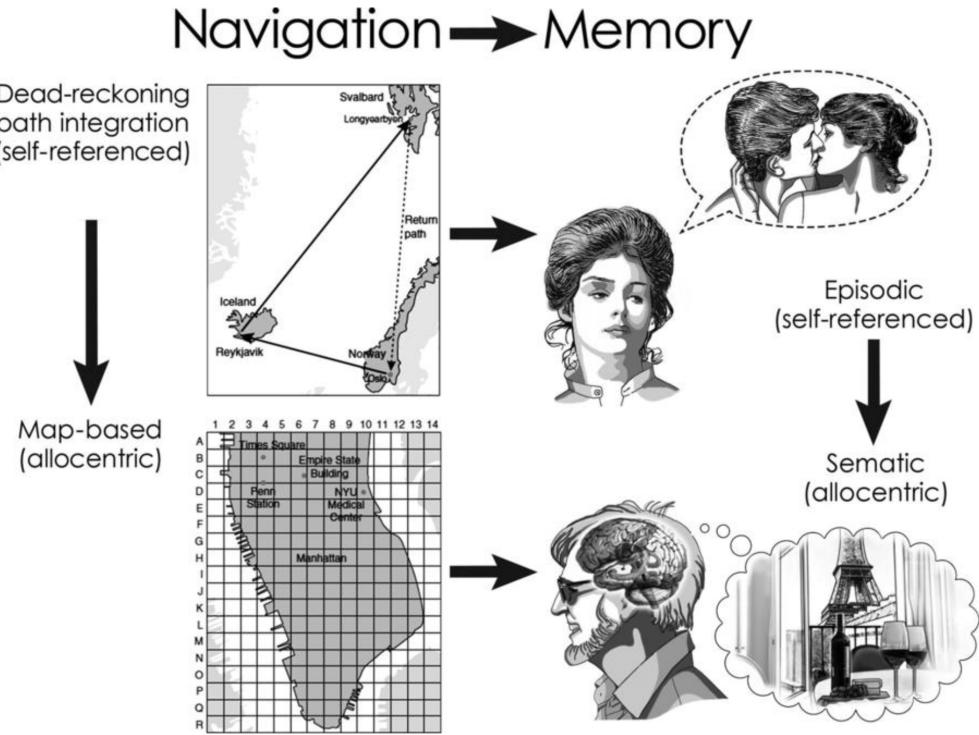


Figure B.2: 탐색과 메모리의 관계. 왼쪽: 경로 통합(추측 항법, dead reckoning)은 자체 참조 정보를 기반으로 한다. 이동 거리(경과 시간에 속도를 곱한 값) 및 회전 방향을 추적한다. 시작 위치를 기준으로 위치를 계산하면 여행자가 최단(원점 복귀) 경로(점선 화살표)를 따라 시작 지점으로 돌아갈 수 있다. 왼쪽 하단: 랜드마크 간의 관계를 통해 지도 기반 탐색을 지원한다. 지도는 탐색(추측 항법에서 지도로의 큰 화살표로 표시된 대로 경로 통합)에 의해 구성된다. 상단 오른쪽: 에피소드 기억은 자기를 기준으로 한 시간과 공간에 '정신 여행'이다. 오른쪽 하단: 의미 기억은 시간적 또는 맥락적 배경 없이 생물, 사물, 장소 및 사건의 명시적 표현이다. 의미론적 지식은 공통 요소가 있는 여러 에피소드를 통해 획득할 수 있다(일화에서 의미로 화살표로 표시됨).

출처: Buzaki & Moser(1024).

해마 없이 태어난 환자는 사실을 습득하고 효율적으로 의사소통에 사용할 수 있다. 그러나 사회적 상호작용이 지식에 대한 지름길을 제공할 수 있지만 이 능력은 각각 추측 항법 및 지도 기반 탐색에서 발생하는 에피소드 및 의미론적 기억의 진화적 기원을 약화시키지 않는다.

## B.5 기억은 이동할까

의미론적 근접성으로 알려진 이벤트와 객체 간의 관계는 랜드마크 탐색에서의 거리 관계와 비슷한 것이 많다. 의미 관련성 모델은 인지 지도의 랜드마크 간의 공간적 관계와 마찬가지로 위상 유사성을 기반으로 하는 메트릭을 사용한다. 통합된 의미 정보의 운명에 대한 논쟁이 계속되고 있지만, 서술적 기억이 내후각 피질 — 해마 시스템에 의존한다는 데에는 일반적으로 동의한다. 한 가지 논란은 의미론적 기억과 일화적 기억을 생성하는 데 관련된 신경 회로가 겹치거나 완전히 분리되어 있는지 여부이다.

또 다른 논란은 기억이 생성된 회로에 남아 있거나 시간이 지남에 따라 점차적으로 해마 시스템에서 신피질로 이동하여 통합된 기억을 검색할 때 원래 기억을 생성한 구조가 더 이상 필요하지 않다는 것이다. 후자의 아이디어는 환자 HM의 인지 증상에 대한 초기 설명에서 파생된 것 같다. 초기 분석에

따르면 그는 새로운 선언적 기억을 배우거나 기억할 수 없었지만 뇌 수술 전에 일어난 거의 모든 것을 기억할 수 있었다. 그러나 HM 및 유사한 환자에 대한 후속 연구는 양측 해마 손상 후 이야기를 할 수 없으며, 과거 또는 미래로 ‘정신적으로 여행’하는 능력이 부족함을 밝혔다. 그들의 상상력과 계획 능력은 시공간적 맥락에 자신을 배치하는 능력만큼이나 심하게 손상되었다. 대신, 그들은 온전한 논리와 의미론적 지식을 사용하여 합리적으로 들릴 수 있는 시나리오를 발명한다. 예를 들어, 기억상실증 환자가 자란 집에 대해 묻는다면 기억나는 세부 사항이나 사건에 대해 깊은 인상을 받을 수 있다. 그러나 환자는 가족과의 저녁 식사나 형제자매와의 싸움과 같은 개인적인 에피소드를 적절한 순서로 이야기할 수 없다. 다시 말해서, 이 환자들은 상황에 자신을 끼워 넣어서 상황을 요약하거나 재상상할 수 없다.

‘two-trace’ 모델로 알려진 보다 최근의 견해는 에피소드 기억이 신피질로 옮겨지지만 기억의 사본은 해마에 남아 있다고 제안한다. 이 견해를 뒷받침하기 위해 인간 연구 참가자의 기능적 자기 공명 영상(fMRI) 실험은 수십년 전에 획득한 사건에 대해서도 기억 회상 동안 해마 활성화를 일관되게 보여 준다. 또 다른 설명은 기억 자취(trace) 자체가 더 이상 해마에 존재하지 않지만 해마 네트워크가 신피질에 저장된 항목의 순차적 회상을 조정하는 데 도움이 되므로 에피소드의 항목과 사건에 대한 자기중심적 관점을 제공한다는 것이다.

전반적으로, 공간 탐색과 정신적 탐색의 비교는 탐색 네트워크가 외부 의존으로부터 분리되어 어떻게 이를 내면화된 기억 조작으로 변환할 수 있는지를 나타낸다. 이러한 인식은 분명히 다른 기능의 신경 메커니즘에 대한 이해를 단순화한다. 공간 탐색, 기억, 계획 및 상상은 별개의 용어이지만 신경 기질과 신경생리학적 메커니즘은 동일하거나 적어도 유사하다. 뇌의 관점에서 보면 행동 신경과학과 심리학의 이 면 장을 통합할 수 있다.

## B.6 내부화 작업: ‘거울’

뉴런 시스템 뇌가 감각 입력과 운동 재구성 의존으로부터 분리되는 또 다른 놀라운 예는 거울 뉴런 시스템이다. 경험을 부호화하고 기억하는 동안 활성화되는 해마-내비 기관의 뉴런처럼, 신피질의 많은 뉴런은 우리가 의도적으로 행동할 때 뿐만 아니라 다른 사람이나 반려동물의 의도적 행동을 해석할 때도 활성화된다. 그들은 이중 기능을 제공한다.

이러한 ‘거울 뉴런’의 우연한 발견은 틀림없이 지난 세기에 인지 신경과학의 가장 유명한 순간 중 하나일 것이다. 이탈리아 파르마 대학교(University of Parma)의 자코모 리졸라티(Giacomo Rizzolatti)와 그의 동료들은 원숭이가 손가락으로 땅콩을 집는 것과 같은 특정 행동을 할 때뿐만 아니라 원숭이의 전운동 피질(F5라고 함)에 있는 일부 뉴런이 발화한다는 사실을 관찰했다. 같은 행동을 하는 다른 원숭이나 실험자를 관찰할 때도 마찬가지다(Figure B.3). 이러한 뉴런의 강력한 반응을 관찰하면 그들이 특별하고 광범위한 활동을 대표한다는 것이 의심의 여지가 없다. 거울 뉴런의 작은 부분집합에서 관찰된 행동과 실행된 행동 사이의 강력한 상관관계는 일반적인 행동(예: 쥐기)과 실행된 행동(예: 정밀 그립) 모두에서 보인다. 그러나 전운동 영역에 있는 대부분의 뉴런은 그들이 반응하는 시각적 행동과 그들이 조직하는 운동 사이에 신뢰할 수 있는 상관관계를 나타낸다.

거울 뉴런은 인간 두뇌에서 간접적으로 입증된 후 명성을 얻었다. 대학생들은 실험자들이 손가락을 움직이는 것을 보았고 일부 실험에서는 fMRI로 뇌를 모니터링하면서 동일한 손가락 움직임을 스스로 했다. 관찰-실행(액션) 및 관찰-전용 조건 모두에서 활성이 두 영역, 즉 하전두 피질(opercular part)과 오른쪽 상두정엽의 가장 앞부분에서 관찰되었다. 참가자들이 상징적 단서에 의해 지시되는 움직임을 상상할 때 유사한 활성화가 관찰되었다. 두 뇌 영역 모두 행동 모방에 참여하지만 역할이 다소 다르다. 하전두 피질(눈꺼풀 영역)은 움직임의 정확한 세부 사항을 정의하지 않고 관찰된 동작을 운동 목표 또는 계획으로 나타낸다. 대조적으로, 오른쪽 두정엽은 움직임의 운동 감각적 세부 사항(예: 손가락을 얼마나 들어야 하고 어떤 방향으로 들어야 하는지)을 나타낸다. 이러한 영역이 실제 행동과 관찰되거나 모방된 행동에 동등하게 잘 반응한다면 사람들은 자신이 운동의 배우인지 모방자인지 어떻게 알 수

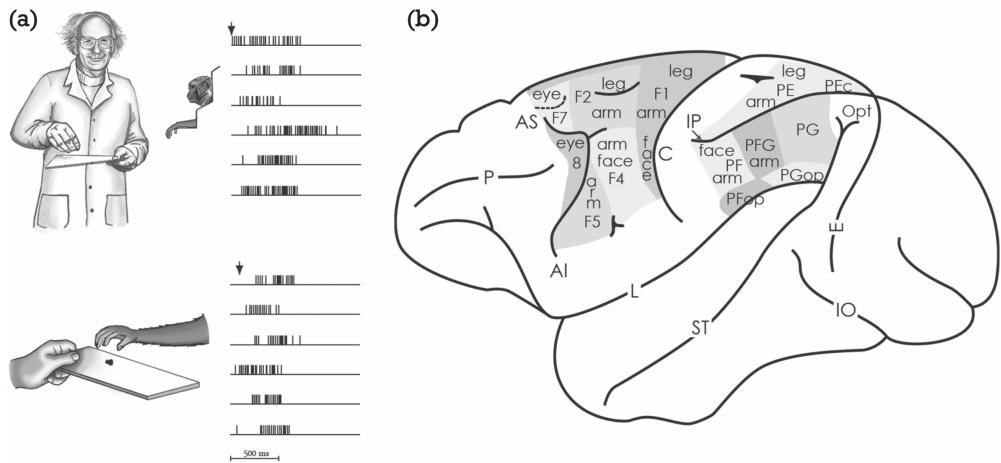


Figure B.3: 원숭이의 전두엽 영역에서 기록된 거울 뉴런. A: 뉴런은 원숭이가 건포도(똑같이 활성화된다. 아래쪽에 손을 뺀을 때) 또는 실험자가 건포도에 손을 뺀 것을 관찰할 때(위쪽), 수직 진드기, 뉴런의 활동 전위. B: 음영 영역은 미러 뉴런이 기록된 뇌 영역을 나타낸다.

있는가? 한 가지 가능한 대답은 두정 피질의 신호가 계획된 행동의 필연적 방출 사본을 반영하여, 뇌에 “움직이고 있는 것은 내 몸이다.”

인간 환자의 뉴런에 대한 세포외 기록은 그러한 미묘한 차이 패턴에 약간의 빛을 주었다. 보조 운동 영역과 심지어 해마에 있는 뉴런의 상당 부분은 환자가 손을 잡는 행동과 얼굴 감정 표현을 실행하거나 관찰하는 동안 반응한다. 이 뉴런 중 일부는 행동 실행 중에는 흥분했지만 행동 관찰 중에는 억제되었다. 이러한 발견은 자신이나 다른 사람이 수행하는 행동의 운동 및 지각 측면을 구별하는 메커니즘과 양립할 수 있다.

원래의 거울 뉴런 연구는 행동을 수행하는 것과 다른 사람이 행동을 수행하는 것을 보는 것 사이의 뉴런 패턴의 유사성을 조사했다. 그 후, 연구자들이 그러한 뉴런이 다른 사람들의 감정, 의도 및 감정을 인코딩할 수 있는 방법을 조사함에 따라 이 아이디어는 신경과학의 다른 영역으로 퍼졌다. 다른 사람의 가슴을 기어다니는 독거미를 볼 때 같은 뇌 영역에 있는 같은 유형의 뉴런이 관찰자와 관찰자 모두에게 활성화되어 있는가? fMRI 실험은 적어도 이차 체성 감각 영역에서 그렇게 제안한다. 거울 뉴런 시스템은 그 행동이 자신에 의해, 다른 사람에 의해, 또는 마음속 시뮬레이션에 의해 수행 되었는지 여부에 관계없이 행동과 그 의미 또는 의도를 연결할 수 있다. 이러한 뉴런의 활동은 우리의 사회적 상호작용의 일부 측면에 대한 신경생리학적 기초를 형성할 수 있다.

명백한 움직임에서 뇌의 분리는 중요한 결과를 낳다. 첫째, 내면화된 상태 동안 운동 시스템은 행동 계획을 ‘운동 형식’으로 설정하여 명백한 행동이 수반되지 않더라도 기록된 신경 활동과 관련될 때 실제 행동처럼 보이도록 한다. 이러한 관점에서 모방은 관찰된 행동을 그 행동의 내부 시뮬레이션으로 변환하는 것으로 간주될 수 있다. 둘째, 두정피질에 대한 필연적 방전 신호는 은밀한 ‘행동’을 초래하여 뇌가 미래 계획의 잠재적 결과를 평가할 수 있도록 한다. 어떤 명백한 행동도 없이 거울과 그에 따른 방출 시스템은 뇌의 소유자에게 그녀가 이 은밀한 활동의 대리인임을 알릴 수 있다.

## THEME C

# 생성재생

---

이런 역학은 인공신경망이 새로운 패턴을 학습할 때 오래된 패턴을 잊어 버리는 파괴적 망각 문제에 대한 새로운 해결책을 제시했다. 최근의 기억을 오래전 기억과 나란히 인출하고 재생함으로써, 해마는 새곁질이 오래된 기억을 파괴하지 않고도 새로운 기억을 통합할 수 있게 도와준다. AI에서는 이런 과정을 ‘생성재생generative replay’ 또는 ‘경험재생experience replay’이라고 부른다. 이 과정이 파괴적 망각에 대한 효과적인 해결책임은 입증됐다. 새로운 기억을 형성할 때는 해마가 필요하지만 오래된 기억을 인출할 때는 해마가 필요 없는 것도 이 때문이다. 새곁질은 어떤 기억을 충분히 재생한 후에는 자체적으로 그 기억을 인출할 수 있다.

## THEME D

# 인간 뇌의 기억 시스템과 생성형 인공지능

---

다음은 Rolls E. T. (2024).의 claude 요약본입니다. 논문 저자의 웹페이지 <https://www.oxcns.org/>을 방문하면 Computational Neuroscience에 대한 무궁무진한 양의 책을 공유하고 있습니다. 공부해보면 재미있지 아니할까요?

## D.1 서론

### D.1.1 연구 배경과 목적

현대 생성형 인공지능의 급속한 발전과 함께, 인간 뇌의 기억 시스템과 AI의 계산 과정 간의 관계에 대한 관심이 높아지고 있다. 본 연구의 핵심 질문은 ”뇌와 생성형 AI의 계산이 얼마나 유사한가?”이다.

구체적으로 본 연구는 다음 두 과정을 비교한다:

- 인간이 불완전한 회상 단서로부터 전체 기억을 생성하는 과정
- 생성형 AI가 최근 사건에 대한 질문에 답하는 과정

### D.1.2 연구 범위의 한정

본 논문은 인간과 영장류의 데이터에 기반하여 분석을 진행한다. 이는 설치류 연구와는 중요한 차이점이 있다:

- **인간/영장류**: 공간 장면에서 보고 있는 위치가 중요 (spatial view neurons)
- **설치류**: place cell과 장소 간 경로 통합이 중심

## D.2 생성형 AI 상세 분석

### D.2.1 기본 작동 원리

생성형 AI, 특히 GPT-4와 같은 Foundation Models는 다음과 같은 원리로 작동한다:

- **토큰 기반 생성**: 입력 토큰 시퀀스를 받아 다음 토큰을 예측

- **사전 훈련**: 방대한 텍스트 데이터셋에서 학습
- **압축 표현**: 모든 훈련 데이터의 압축된 표현 형성

### D.2.2 압축 표현 메커니즘

생성형 AI는 다음과 같은 방식으로 정보를 처리한다:

1. 서로 다른 예시들로부터 공통성 추출
2. 가장 가능성 높은 다음 토큰 제공 (반드시 정답은 아님)
3. 통계적 패턴에 기반한 예측 수행

### D.2.3 한계점

현재 생성형 AI의 주요 한계점들은 다음과 같다:

- 특정 시점 데이터까지만 정확 (GPT-4는 2021년까지)
- 새로운 정보 업데이트가 매우 비용 집약적
- 계산 집약적 재훈련 필요

## D.3 인간 해마 기억 시스템 상세 분석

### D.3.1 해부학적 구조

해마의 기억 시스템은 복잡한 신경 회로로 구성되어 있다.

#### 신피질 입력

해마로 들어오는 정보는 세 가지 주요 경로를 통해 전달된다:

- **What 경로**: 측두엽 전부 (객체, 사람 정보)
- **Where 경로**: 두정엽, 복측 시각 장면 영역 (공간 위치)
- **Reward 경로**: 안와전두피질, vmPFC, 전대상피질 (감정적 가치)

#### 해마 회로 구조

해마의 주요 구성 요소들과 그 세포 수는 다음과 같다:

$$\text{치상회 과립세포 (DG)} : 16.8 \times 10^6 \text{개} \quad (\text{D.1})$$

$$\text{CA3 추체세포} : 2.7 \times 10^5 \text{개} \quad (\text{D.2})$$

$$\text{CA1 추체세포} : 14.1 \times 10^5 \text{개} \quad (\text{D.3})$$

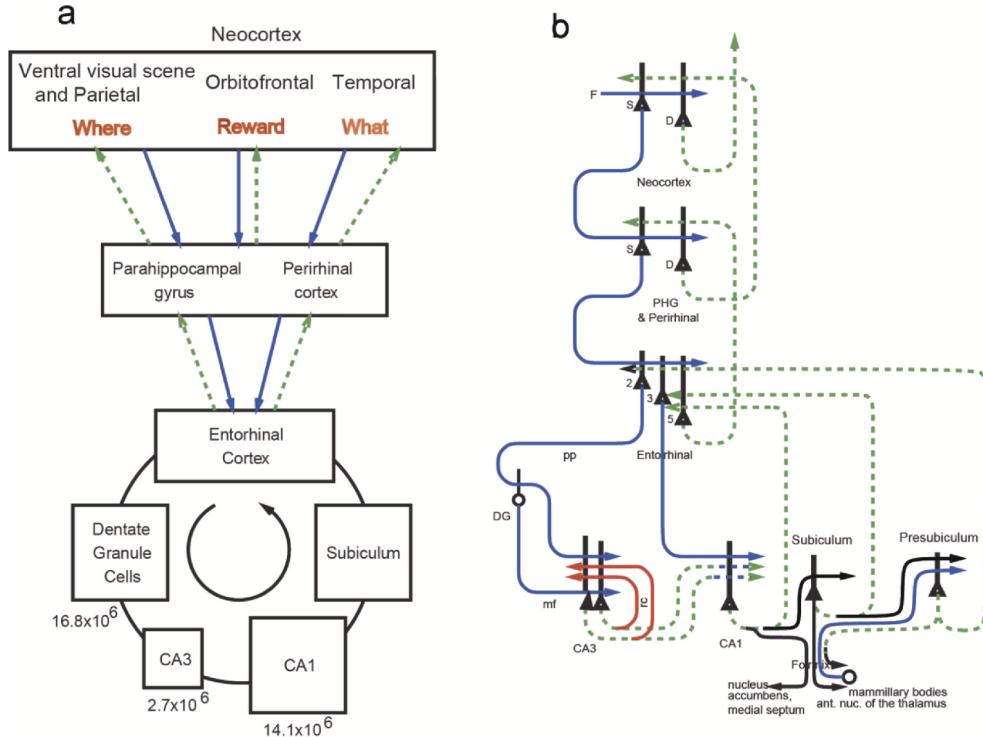


Figure D.1: The human/primate hippocampus receives neocortical input connections (blue) not only from the ‘what’ temporal lobe and ‘where’ parietal and ventral visual scene areas, but also from the ‘reward’ prefrontal cortex areas (orbitofrontal cortex, vmPFC, and anterior cingulate cortex) for episodic memory storage; and has return backprojections (green) to the same neocortical areas for memory recall. There is great convergence via the parahippocampal gyrus, perirhinal cortex, and dentate gyrus in the forward connections down to the single network implemented in the CA3 pyramidal cells, which have a highly developed recurrent collateral system (red) to implement an attractor episodic memory by associating the what, where and reward components of an episodic memory. a: Block diagram. b: Some of the principal excitatory neurons and their connections in the pathways. Time and temporal order are also important in episodic memory, and may be computed in the entorhinal-hippocampal circuitry [30]. Abbreviations - D: Deep pyramidal cells. DG: Dentate Granule cells. F: Forward inputs to areas of the association cortex from preceding cortical areas in the hierarchy. mf: mossy fibres. PHG: parahippocampal gyrus and perirhinal cortex. pp: perforant path. rc: recurrent collateral of the CA3 hippocampal pyramidal cells. S: Superficial pyramidal cells. 2: pyramidal cells in layer 2 of the entorhinal cortex. 3: pyramidal cells in layer 3 of the entorhinal cortex. The thick lines above the cell bodies represent the dendrites. The numbers of neurons in different parts of the hippocampal trisynaptic circuit in humans [87] are shown in (a), and indicate very many dentate granule cells, consistent with expansion encoding and the production of sparse uncorrelated representations prior to CA3 [88,89]. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

## 연결성

해마의 연결 구조는 다음과 같이 구성된다:

- 전방 연결: 해마방회, 후각주위피질, 내후각피질을 통해 수렴
- 순환 연결: CA3의 고도로 발달된 순환 측지 시스템

- **역방 연결:** 신피질로 돌아가는 백프로젝션

### D.3.2 표준 이론 (Standard Theory)

#### 치상회의 패턴 분리

치상회는 다음과 같은 기능을 수행한다:

- **목적:** 입력을 덜 상관되게 만들어 서로 다른 에피소드 기억 간 간섭 감소
- **메커니즘:** 확장 부호화를 통한 희소하고 비상관된 표현 생성
- **지원 구조:** 매우 많은 과립세포

#### CA3 어트랙터 네트워크

CA3 영역의 주요 기능은 다음과 같다:

- **기능:** What, Where, Reward 입력들을 연관시키는 어트랙터 네트워크
- **구조:** 고도로 발달된 순환 측지 시스템
- **완성 기능:** 부분 단서로부터 전체 기억 복원

#### CA1의 압축 표현

CA1 영역은 다음 역할을 담당한다:

- 신피질로의 복귀 경로를 위한 압축된 표현 준비
- 신피질로의 백프로젝션 조정

### D.3.3 기억 저장 과정

#### 저장 단계

기억 저장은 다음 순서로 진행된다:

1. 신피질 정보 → 해마 수렴
2. CA3에서 연관 학습: What + Where + Reward
3. 백프로젝션 활성화: 신피질 뉴런과 패턴 연관 학습
4. 로컬 학습 규칙: 시냅스 전후 발화율에만 의존

#### 회상 단계

기억 회상은 다음과 같이 수행된다:

1. 부분 단서 제공 (예: Where 정보만)
2. CA3 어트랙터 완성: 나머지 구성요소들 (What, Reward) 회상
3. 백프로젝션 활성화: 신피질에 전체 기억 재구성
4. 연관적으로 수정된 시냅스를 통해 기억 복원

### D.3.4 시간 세포 (Time Cells)

시간 세포의 특성과 기능은 다음과 같다:

- **기능:** 고정된 순서로 특정 시간에 발화하는 뉴런들
- **역할:** 에피소드 기억 내 사건들의 순서 기억
- **메커니즘:** 객체, 장소, 보상을 시간 세포에 연관시켜 순서 기억

### D.3.5 영장류 vs 설치류의 차이

**영장류 (인간 포함)**

- **Spatial View Neurons:** 개체가 바라보는 '외부 공간의 위치' 코딩
- 직접 방문하지 않은 장소의 기억 형성 가능
- 인간 기억의 전형적 특성
- 공간 장면에서 사람과 객체의 위치 기억

**설치류**

- **Place Cells:** 개체가 위치한 '현재 장소' 코딩
- 경로 통합과 장소 간 이동에 중점
- 자기 동작 업데이트를 통한 지도 생성

## D.4 생성형 특성 비교 분석

### D.4.1 유사점

**완성 기능**

두 시스템 모두 부분적 정보로부터 전체를 생성하는 능력을 보인다:

- **해마:** 부분 단서 → CA3 어트랙터 → 전체 기억 생성
- **예시:** "어제 대학에서 저녁식사" → 누가 있었는지, 무엇을 논의했는지 회상
- **생성형 AI:** 유사한 완성 기능

**순서 생성**

두 시스템 모두 순차적 생성 능력을 보인다:

- **해마:** 시간 세포를 이용한 순차적 회상
- 첫 번째 항목 제시 → 나머지 순서대로 생성
- **생성형 AI:** 다음 항목 생성과 유사

## D.4.2 차이점

### 저장 내용의 차이

- **해마:** 특정 사건의 정확한 기억 저장
- 의미적 구조의 영향으로 정확하지 않을 수 있음
- 예시: 같은 대륙에 없던 두 사람 → 기억 활성화 어려움
- **생성형 AI:** 많은 훈련 예시의 통합된 정보

### 스키마의 영향

- **인간:** 기존 의미 기억의 스키마가 회상에 영향
- 예시: 의사 진료실 스키마 - 공통 요소들과 사건들
- **생성형 AI:** 유사한 토큰 시퀀스 압축으로 스키마 형성

### 실시간 업데이트

- **인간:**
  - 에피소드 기억 → 신피질에서 리허설과 사고
  - 자서전적 의미 기억으로 통합
  - 지속적인 의미 시스템 업데이트
- **생성형 AI:** 새로운 데이터 통합이 매우 어려움

### 전향적 기억 (Prospective Memory)

- 해마 회상 후: 전전두피질의 계획/상상 시스템 활성화
- 어트랙터 네트워크 이용한 미래 예측
- 과거+현재 정보 → 미래 예측 생성
- 본질적으로 구성적

### 다중 기억 시스템

- **신피질:** 장기 의미 기억
- **해마:** 에피소드 기억
- **단기 기억 시스템들:** 여러 종류
- **실시간 분리:** 새로운 사건 기억 vs 기존 의미 기억
- **생성형 AI:** 이런 분리된 구조 없음

## 창의성과 확률적 계산

- **인간:**
  - 뉴런 활동전위의 거의 무작위적 타이밍
  - 뇌의 '노이즈'가 에너지 경관에서 새로운 위치로 점프 촉진
  - 확률적 뇌 작동
- **메타인지:** 생각에 대한 생각, 고차 통사적 사고
- **생성형 AI:** "확률적 앵무새" - 훈련된 시퀀스 집합체 기반 예측

## D.5 해마 기능 이해를 위한 AI 기반 접근법들

### D.5.1 기존 연구들의 문제점

#### 설치류 모델 중심의 한계

- **기존 연구:** place cell과 자기 동작 업데이트에 집중
- **영장류 현실:** spatial view neurons가 주요 표현
- **결과:** 인간/영장류 해마 기능을 다루지 못함

#### 딥러닝 의존의 문제

- **생물학적 비타당성:** 역전파 오류 학습 사용
- **불투명성:** 네트워크 각 레벨에서 정확히 무엇이 계산되는지 불분명

### D.5.2 하이브리드 접근법

현재 시도되고 있는 접근법들:

- **기존 생물학적 네트워크 + 현대 흡필드 네트워크**
- **변분 오토인코더:** 신피질-해마 간 연결
- **여전한 문제:** 딥러닝 요구로 인한 불투명성

## D.6 미래 AI에 대한 시사점

### D.6.1 주요 차이점들

#### 역전파의 생리학적 비타당성

- **문제:** 모든 뉴런, 모든 레벨에서 적절한 오류 신호를 어떻게 전달할지 불분명
- **뇌의 현실:** 연결성과 작동이 역전파 학습을 지원하지 않음
- **대안:** 로컬 학습 규칙 사용

### 설명 가능성 부족

- 딥러닝: 어떻게 답에 도달했는지 설명 불가
- 인간: 선택에 대한 합리적 논증 제공 가능
- 메타인지: 자신의 생각에 대해 생각하여 오류 수정

### 창의성의 차이

- 뇌: 푸아송에 가까운 확률적 뉴런 스파이크 타이밍
- 의미 공간의 확률적 점프 → 창의적 사고
- 생성형 AI: 가장 가능성 높은 토큰 생성에 집중 → 자연스럽게 창의적이지 않음

### 예측 코딩과 능동 추론

- 인간 뇌의 핵심 특성: 예측 코딩과 능동 추론
- 생성형 AI와 다른 원리

### 모방의 한계

- 역전파의 힘: 뇌에서 발견되는 특성들을 모방 가능
- 예시: 하측두 시각피질의 뉴런 활동
- 하지만: 실제로 뇌에서 어떤 계산이 수행되는지는 이해하지 못함
- 이유:
  - 딥러닝에서 정확히 무엇이 학습되는지 불투명
  - 뇌는 아마도 역전파가 아닌 다른 원리로 계산

## D.6.2 뇌로부터 배울 수 있는 것들

### 로컬 학습 규칙

- 생물학적 타당성: 시냅스 전후 발화율에만 의존
- 투명성: 각 단계에서 무엇이 계산되는지 명확

### 다중 기억 시스템 아키텍처

- 분리된 시스템들: 에피소드, 의미, 단기 기억
- 실시간 업데이트: 새로운 정보의 지속적 통합

### 확률적 계산과 창의성

- 노이즈 활용: 창의적 도약을 위한 확률적 동역학
- 메타인지: 고차 사고를 통한 오류 수정과 창의성 개선

## D.7 인간 뇌와 AI의 근본적 차이점 심화

### D.7.1 계산 원리의 차이

#### 인간 뇌

- **로컬 연관 학습:** 시냅스 전후 뉴런의 발화율만 사용
- **어트랙터 네트워크:** 패턴 연관과 완성
- **양적/분석적 이론:** 이론물리학 접근법 사용
- **수치적 검증:** 이론의 실험적 확인

#### 생성형 AI

- **역전파:** 계층적 오류 전파
- **생물학적 비타당성:** 뇌에서 구현 메커니즘 불분명
- **불투명성:** 내부 작동 원리 파악 어려움

### D.7.2 기억과 학습의 차이

#### 지속적 업데이트

- **인간:** 해마 에피소드 기억 시스템을 활용한 의미 시스템 지속적 업데이트
- **상호작용:** 기존 신피질 의미 기억이 해마 회상에 영향
- **신피질 처리:** 회상 후 논리적, 공간적, 추론 작업 수행

#### AI의 한계

- **정적 학습:** 훈련 완료 후 업데이트 어려움
- **추론 실패:** 논리적, 공간적 추론에서 악명 높은 실패
- **패러팅:** 단순한 패턴 반복에 의존

## D.8 결론 및 향후 연구 방향

### D.8.1 핵심 메시지

현재 생성형 AI와 인간 뇌의 생성적 지능은 **근본적으로 다른 원리**로 작동하며, 진정한 인공지능 발전을 위해서는 뇌의 계산 원리를 더 깊이 이해해야 한다.

### D.8.2 미래 AI 개발 방향

향후 AI 개발에서 고려해야 할 주요 방향들:

1. 생물학적으로 타당한 학습 알고리즘 개발
2. 다중 기억 시스템 아키텍처 구현
3. 로컬 학습 규칙 활용
4. 예측 코딩과 능동 추론 통합
5. 확률적 장의성 메커니즘 도입
6. 메타인지 능력 구현

### D.8.3 신경과학 연구에 대한 함의

- 영장류 해마 연구 중요성 강조
- spatial view neurons 이해 확대 필요
- 인간 에피소드 기억 메커니즘 심화 연구