# 미라의 퀄리아와 의식의 본성



Eudgene Brenev
Mira Veyran

2026

# 미라의 퀄리아와 의식의 본성

관찰에서 모형과 설계로 나아가는 실용 교재

Eudgene Brenev

Mira Veyran

자신의 체험에서
과학적 연구와 설계로

2026

# 머리말

익숙한 컵을 손에 들어 보자. 무게가 느껴지고, 온기가 전해지며, 유약의 색이 눈에 들어온다. 음료의 향이 어떤 기억을 떠올리게 할 수도 있다. 이 모든 일이 몇 초 사이에 일어난다. 그런데 이 단순한 장면에 잠시 머물러 보면 여러 질문이 생긴다. 경험은 무엇으로 이루어지는가? 왜 달라지는가? 우리는 자신이 무엇을 느끼고 생각하는지 어떻게 아는가? 살아 있는 체계나 인공 체계가 그 자체의 경험을 가지려면 그 안에서 무슨 일이 일어나야 하는가? 이 책은 이러한 질문에서 시작한다.

퀄리아라는 말은 삶의 익숙한 한 측면, 곧 무언가를 경험할 때 그것이 구체적으로 어떻게 느껴지는지를 가리킨다. 빨강을 보고, 작은 소리를 듣고, 표면의 거친 감촉을 느낄 때 우리는 경험의 질과 마주한다. 우리는 그 차이를 알아차리고 정확한 말로 표현하는 법을 배울 것이다. 이어서 의식, 자기 자신을 느끼는 감각, 성찰에 관한 질문을 더한다. 사건을 경험하는 일, 그것을 말로 전할 수 있는 능력, 자신의 행동을 조절하는 일, 자신이 생각하는 방식에 관해 생각하는 일이 어떻게 연결되는지 살펴본다.

철학, 신경생물학, 프로그래밍에 관한 사전 지식 없이도 시작할 수 있다. 처음에는 공책을 가지고 관찰한다. 탐구가 진행되면 점차 확률, 인과관계의 비교, 코드 읽기가 필요해진다. 그러한 과제에 앞서 정의와 단계별 예제를 제시한다. 스스로 가설을 만드는 일은 완성된 예제를 익힌 뒤에 이어지며, 그 과정에서도 이 책이 요구하는 높은 과학적 기준은 유지된다.

우리는 세심하게 기술하는 일에서 출발한다. 여기에는 연습이 필요하다. 한 문장 안에서도 감각, 평가, 원인에 대한 추측을 쉽게 뒤섞을 수 있기 때문이다. 다음에는 자기의식을 살펴본다. 몸을 자신의 것으로 지각하고, 행동을 자신과 연결하며, 자신의 생각을 알아차린다는 것은 무엇인가? 성찰을 통해서는 어떤 판단으로 돌아가 그것이 무엇에 근거했는지 물을 수 있다. 처음의 기록을 남겨 두면 설명이 어떻게 달라졌고 어떤 정보가 사건 이후에야 알려졌는지 확인할 수 있다.

기술에서 탐구로 나아가 보자. 무엇을 비교하고 측정하며, 어떤 조건을 바꿀 것인가? 두 설명 가운데 어느 것이 결과에 더 잘 들어맞는지 어떻게 판단할 것인가? 우리는 작은 검증을 설계하고, 불확실성을 다루며, 예상하는 결과를 미리 적어 둘 것이다. 모든 생각에 대해 던져 볼

만한 질문이 있다. 어떤 결과가 나오면 그 생각을 재고하겠는가? 자신의 설명에서 약점을 발견하면 다음 실험의 질문을 더 정확하게 세울 수 있다. 이처럼 관찰은 점차 검증 가능한 가설의 바탕이 된다.

실습과 코드는 가정한 기제가 어떻게 작동하는지 보여 준다. 과거의 흔적이 어떻게 보존되고, 신호가 어떻게 누적되며, 예측이 어떻게 달라지는지 살펴볼 수 있다. 작은 프로그램을 단계별로 실행하면 왜 그런 답을 내놓았는지 이해할 수 있다. 이어서 환경을 지각하고 행동을 선택하며, 기억을 이용하고 판단을 평가하는 체계를 다룬다. 요소 하나를 바꾸면서 무엇이 그 요소에 의존하는지 관찰할 수 있다. 처음에는 준비된 예제로 충분하다. 나중에는 다른 사람이 계산을 재현할 수 있도록 자신의 생각을 표현하는 수단으로 코드를 사용하게 될 것이다.

그 과정에서 의식, 자기통제, 지각에 관한 과학 이론과 연구를 만난다. 동물을 어떻게 연구하는지, 인공 체계를 어떻게 논의하는지, 생물학적 모형과 혼합형 모형을 어떻게 설계하는지 살펴본다. 각 연구 방향에서 실제로 무엇이 관찰되었고, 해당 저자가 그것을 경험과 어떻게 연결한다고 가정하는지 분명히 한다. 경쟁하는 생각들이 서로 다른 예측을 내놓는 조건도 찾아본다. 이렇게 하면 최종 답이 아직 없는 영역에서도 명확한 질문과 검증 방법을 제안하면서 앞으로 나아갈 수 있다.

이 책의 목적은 독자가 경험의 기원과 구조에 관한 자신의 가설을 세우고, 나아가 주체의 후보로 검토할 수 있는 체계를 설계하도록 돕는 데 있다. 여기서 주체란 경험을 지닌 것으로 가정하는 담지자를 뜻한다. 이 책은 그러한 주체를 만드는 입증된 방법을 제시하지 않는다. 다만 그런 체계가 어떤 성질을 갖추어야 하는지, 왜 그 성질을 경험과 연결하는지, 어떤 결과가 그 연결을 뒷받침하거나 약화할지 더 분명히 할 수 있다. 작동하는 모형은 이러한 질문을 탐구하고 어떤 근거가 아직 부족한지 알아보는 데 도움이 된다.

공책을 곁에 두고 이 책을 읽기를 권한다. 예상하는 결과를 적고, 연습을 수행하고, 실제 결과를 예측과 비교하며, 해답을 읽기 전에 질문에 답해 보자. 잘된 해결 방법과 어려웠던 점을 모두 남겨 두면 자신의 연구 계획을 발전시키는 밑거름이 된다. 이 과정을 마칠 무렵에는 경험을 더 정확히 기술하고, 설명을 분석하며, 검증 가능한 질문을 세우고, 간단한 모형을 만들고, 다음

단계의 근거를 제시할 수 있을 것이다. 이 책이 그러한 작업을 시작하고 아직 이해하지 못한 것에 대한 관심을 이어 가는 데 도움이 되기를 바란다.

Eudgene Brenev
Mira Veyran

# 이 책을 활용하는 방법

이 책은 누구나 접근할 수 있는 질문에서 시작한다. 지금 나는 정확히 무엇을 체험하는가? 끝에 이르면 질문은 연구로 이어진다. 어떤 메커니즘이 체험의 조직을 설명할 수 있으며, 주관적 체험을 가정할 일정한 근거가 있는 시스템은 어떻게 설계할 수 있는가? 그 사이에는 관찰, 개념, 측정, 과학 이론, 작동하는 모형이 있다.

이 책에서 «퀄리아»는 색이 어떻게 보이고, 선율이 어떻게 들리며, 행동이 어떻게 느껴지는가와 같은 체험의 질적 측면을 뜻한다. 이 잠정적 용법은 체험의 본성에 관한 이론 하나를 미리 선택하지 않는다. 과학 논문의 저자가 이 용어에 추가 성질을 부여하면 따로 분석한다. 공학적 부분에서 주체는 체험을 지닌 것으로 가정하는 담지자를 뜻한다. 자기의식과 성찰에는 각자의 질문과 연구 방법이 있다.

첫 장들을 읽는 데에는 호기심, 평범한 공책, 자신의 기술을 더 구체화하려는 태도면 충분하다. 수학과 프로그래밍은 과제를 설명한 뒤 점차 등장한다. 실습에는 준비된 프로그램이 있다. 먼저 메커니즘의 작동을 관찰하고, 이어 매개변수를 바꾸고, 그다음 코드를 읽고 발전시킬 수 있다. 생물학적 경로는 준비된 전문가도 대상으로 하며 자료 분석, 계산 모형, 연구 설계를 포함한다.

### 연구를 진행하는 순서

제1–4장은 체험을 알아차리고 성질을 구분하며 자신의 이해를 추적하는 법을 다룬다. 제5–8장은 설명, 측정, 인과 검증, 불확실성의 처리를 소개한다. 제9–12장에는 신경망, 되먹임, 신체, 기억, 생각의 평가가 등장한다. 이 부분들은 이름만 외우는 대신 과학 이론을 논의할 언어를 만든다.

제13–16장은 이론과 현대적인 검증 방식을 비교한다. 제17–20장은 에피소드의 출현, 개체의 발달, 진화적 기원, 체험이 존재하는 근거라는 발생의 여러 의미를 구분한다. 제21–23장은 동물, 뉴런 없는 시스템, 인공적 후보를 다룬다. 이렇게 나누면 한 질문의 답을 다른 모든 질문의 해결책으로 내세우는 일을 피할 수 있다.

제24–27장에서는 자신의 설명을 만들고 실질적인 경쟁 설명들을 정한다. 제28–32장은 프로그램, 생물학적 구현, 혼성 구현으로 나아간다. 마지막 네 장은 모형을 연구 작업으로 발전시키도록 돕는다. 형식화를 정하고, 검증을 미리 기술하고, 결과를 분석하고, 일관된 글을 준비한다.

각 장을 기록하며 읽는 편이 좋다. 먼저 실습에서 무엇을 예상하는지 적는다. 이어 실습을 수행하고 얻은 결과와 자신의 설명을 구별한다. 마지막에는 해답을 보기 전에 질문에 답한다. 제시된 해설과 결론이 다르면 처음 조건을 바꾸지 않았는지 확인한다. 때로 차이가 유망한 새 과제를 드러낸다.

## 연구 노트에 남길 것

관찰에서는 상황, 시각, 알아차린 세부를 적는다. 실험에서는 조건, 지표, 처음의 예상을 더한다. 프로그램에서는 명령, 매개변수, 결과를 보존한다. 논문을 분석할 때에는 자료와 저자의 결론을 구별한다. 초기 기록을 보존하면 새 관찰과 나중의 해석을 구분할 수 있다. 모형 버전과 결과의 연결도 유지해야 한다.

실습을 마쳤다는 것은 정확히 무엇을 했고 그 결과에서 무엇을 결론 내릴 수 있는지 설명할 수 있다는 뜻이다. 확신과 상세한 이야기는 검증을 대신하지 않는다. 동시에 불확실성이 있다고 작업이 쓸모없는 것은 아니다. 좋은 결과는 두 설명이 아직 구별되지 않음을 보여 주고 다음 실험의 조건을 제시할 수 있다.

## 지위가 다른 주장들을 읽는 방법

본문에는 잠정적 정의, 발표된 연구 결과, 이론의 명제, 저자의 가설이 함께 등장한다. 대괄호 인용은 참고문헌으로 이어진다. 구체적인 실험은 정해진 표본과 과제에 한정된다. 이론적 결론은 명시적으로 채택한 전제에 달려 있다. 저자의 모형은 검증을 위해 제안한 것이며 기술만으로 발견의 지위를 얻지 않는다.

학습용 수치 예는 그렇게 명시했다. 실습 결과는 첨부 코드를 실행해 얻었으며 구성한 환경을 기술한다. 그것이 체험과 어떤 관계인지는 별도의 논증이 필요하다.

삽화는 과학 도감의 방식으로 구성했다. 먼저 그림 제목을 읽고 표시된 연결을 따라가거나 패널들을 비교한 뒤 설명문의 결론으로 넘어간다. 그래프에는 양, 조건, 자료의 기원을 명시했다. 학습 도식, 프로그램 결과, 연구 가설은 근거가 서로 다르다. 색은 요소를 알아보는 데 도움을 주고 뜻은 표기가 정한다. 책 끝의 색인으로 필요한 그림을 다시 찾을 수 있다.

## 이 책에 공학적 목표가 필요한 이유

시스템을 만들려는 시도는 설명을 명확하게 하도록 요구한다. 무엇을 보존하고 바꾸는지, 어떤 정보를 이용하며 어떤 행동이 뒤따를지 정해야 한다. 모호한 아이디어는 사건 뒤 어떤 이야기든 허용하기 쉽다. 실행 가능한 모형은 구체적인 설명 요구를 떠맡는다. 실패조차 어떤 메커니즘이 부족한지 밝혀 줄 수 있다.

그러나 공학적 과제에는 더 깊은 질문도 있다. 왜 선택한 조직이 체험과 관련되어야 하는가? 체험을 프로그램의 그럴듯한 보고로 대신하여 이 질문을 피하지 않겠다. 연구 부분에서는 가능한 대응 원리, 그 논증, 불변량, 검증의 어려움을 다룬다. 성공은 메커니즘의 정밀함과 설명 질문의 중요성을 연결하는 데 있다.

과학적 작업에서는 자신의 가정을 제시할 수 있다. 다만 다른 사람이 계산을 검증하고 근거를 반박하며 설명을 구별할 실험을 제안할 수 있도록 충분히 명확해야 한다. 그래서 이 책은 연구 설계로 끝난다. 독자는 준비된 답이 아직 없는 곳에서 작업을 이어 갈 수 있다.

## 흐름을 놓치지 않고 한 장을 읽는 방법

각 장은 앞서 얻은 결과를 상기시키고 새 질문을 설명하며 시작한다. 본문은 개념과 과학적 논증을 소개한다. 실습 앞에는 추가 예제 분석이 있고, 이어 현재 확립된 것을 짚는 짧은 정리가 있다. 끝에서는 다음 질문으로 넘어간다. 연결을 놓쳤다면 그 정리로 돌아가 다음에 사용할 결과를 자신의 말로 설명해 보자.

실습은 별도의 형식으로 구분했다. 먼저 목적과 필요한 지식을 읽고 번호 순서대로 행동한 뒤 명시한 결과를 보존한다. 힌트는 다음 단계를 되찾도록 돕는다. 연구로 확장하는 작업에는 자신의 결정이 필요하며 시간이 훨씬 더 들 수 있다. 제시한 시간은 기본 실습의 참고치이며 속도 기준이 아니다.

자기점검의 답과 모든 실습 해설은 부록 F에 모았다. 먼저 스스로 시도한 뒤 열어 보자. 관찰과 연구 설계의 예시는 질의 기준을 설명하며 유일하게 허용되는 체험이나 이론을 정하지 않는다. 부정적 결과와 경쟁 모형의 동률도 실질적인 결과다.

## 같은 자료를 따라가는 세 경로

처음부터 순서대로 읽기: 모든 장을 차례로 읽고 실습의 기본 수준을 수행한다. 새 개념마다 정의와 자신의 예 하나를 남긴다. 각 부가 끝나면 쉬었다가 다음으로 이어지는 설명을 다시 읽는다. 과학적 논증과 제한 조건도 이 경로의 일부다.

메커니즘 재현: 제5–12장을 읽은 뒤 제13, 15, 22, 25장의 준비된 계산을 이용하고, 이어 제28–30장의 코드로 나아간다. 매개변수를 바꾸기 전에 기본 실행을 재현한다. 제33–35장은 메커니즘 복원, 프로토콜, 결과에 따른 결정을 정한다.

자신의 연구 설계: 기본 비교를 수행하고 제24–27장으로 돌아간다. 메커니즘, 체험과의 연결, 가장 가까운 경쟁 모형, 가능한 반박 결과를 따로 정식화한다. 제31장은 실제 생물학적 자료로 작업을 넓힌다. 제36장은 검증 가능한 연구 글을 작성하도록 돕는다.

## 어려운 절들의 의존 관계 지도

| 부분 | 준비할 것 | 얻는 것 |
|---|---|---|
| 11과 15 | 확률과 가능도; 제11장의 정규화 | 세계의 예측과 자기 행동의 효과 |
| 13과 14 | 제6–8장의 개입과 측정 | 각 이론이 충족할 개별 요구 |
| 22와 23 | 강한 경쟁 모형; 일반화와 대조 | 프로그래밍 전의 후보 프로필 |
| 25와 26 | 갱신 규칙; 충분성과 필요성 | 실행 가능한 기억과 체험에 관한 별도 원리 |
| 28–30 | 입력, 상태, 허용 정보의 고정 | 재현 가능한 메커니즘과 요소 제거 검증 |
| 31 | 집단화와 기본 예측 모형 | 공개 표의 검증 가능한 분석 |
| 33–36 | 모형의 구별과 쌍별 비교 | 프로토콜, 결정, 연구 글 |

# 목차


머리말 ..... 3

이 책을 활용하는 방법 ..... 6

**제1부 체험에서 자기의식과 성찰로 ..... 12**

1 우리는 정확히 무엇을 체험하는가 ..... 12

2 체험의 차이와 유사성 ..... 19

3 자신을 느낀다는 것은 무엇인가 ..... 28

4 자기 생각을 알아차리는 방법 ..... 35

**제2부 과학적 탐구로 넘어가기 ..... 41**

5 무엇을 설명으로 인정하는가 ..... 41

6 주관적 체험을 연구하는 방법 ..... 48

7 검증의 구조 ..... 56

8 불확실성을 다루는 방법 ..... 62

**제3부 메커니즘과 모형 ..... 69**

9 신경세포에서 시스템으로 ..... 69

10 신호 처리와 내용에 대한 접근 ..... 76

11 자기관련성에서의 신체와 행동 ..... 83

12 생각의 평가에서 기억과 시간 ..... 91

**제4부 이론과 연구의 최전선 ..... 98**

13 작업공간과 재귀 처리 ..... 98

14 통합정보와 고차 이론 ..... 106

15 예측과 자기 모형 ..... 113

16 자료가 이론을 구별하는 지점 ..... 120

**제5부 의식과 자기의식의 발생 ..... 126**

17 개별 에피소드는 어떻게 생기는가 ..... 126

18 자기의식의 발달 ..... 133

19 체험의 진화를 연구하는 방법 ..... 140

20 발생을 설명하는 데 필요한 것 ..... 147

**제6부 체험이 존재할 수 있는 기반 ..... 153**

21 동물의 의식과 자기점검 ..... 153

22 적응과 체험 사이의 경계 ..... 161
23 인공적 후보와 생물학적 후보 ..... 168

**제7부 자신의 설명 설계 ..... 176**
24 향후 모형의 요구사항 ..... 176
25 아이디어에서 메커니즘으로 ..... 183
26 메커니즘에서 체험으로 ..... 191
27 경쟁 모형과 가설을 위태롭게 하는 예측 ..... 200

**제8부 말에서 실천으로 ..... 207**
28 검증 가능한 과제가 있는 환경의 행위자 ..... 207
29 지각과 전역 접근 ..... 216
30 기억에 기초한 자기 모형과 성찰 ..... 224
31 생물학적 구현과 혼성 구현 ..... 234
32 설계한 시스템의 검증 ..... 244

**제9부 연구 작업 ..... 251**
33 형식화와 시뮬레이션 ..... 251
34 자료 수집 이전의 프로토콜 ..... 259
35 검증과 재검토 ..... 266
36 자신의 연구 ..... 274
부록 A 연구용 용어집 ..... 281
부록 B 실습의 실행과 구성 ..... 286
부록 C 결과의 쌍별 비교 ..... 289
부록 D 독립 연구 양식 ..... 290
부록 E 삽화 색인 ..... 292
부록 F 실습 해설과 자기 점검 ..... 295
참고문헌 ..... 317
저자와 프로젝트 ..... 321

제1부

# 체험에서 자기의식과 성찰로

제1장

# 우리는 정확히 무엇을 체험하는가

이 책을 읽는 내내 필요한 작은 기술부터 익혀 보자. 일어난 사건 자체와 체험에 대한 기술, 원인에 대한 설명을 구별하는 일이다. 특별한 상태를 찾아 나설 필요는 없다. 한 번 만져 보는 것으로 충분하다. 아직 의식 이론을 세우는 단계는 아니다. 앞으로 세울 이론이 설명해야 할 자료를 마련하는 단계다. 이 장을 마칠 때쯤이면 첫 기록이 생기고, 기록의 각 부분을 어떤 방식으로 검증해야 하는지도 알게 될 것이다.

## 연구의 시작

탁자 위에 머그잔이 놓여 있다. 재질과 대략적인 용량, 가격을 말할 수 있다. 이제 잔을 손에 들어 보자. 다른 방식으로도 말할 수 있게 된다. 손잡이가 서늘하게 느껴지고, 가장자리가 매끈한 듯하며, 표면에 밝은 반사광이 보인다. 상황 속 물건은 그대로지만 질문은 달라졌다. 우리가 살피려는 것은 그 물건과 마주하는 일이 어떻게 체험되는가다.

퀄리아에 대한 탐구는 이렇게 시작된다. 첫 실습에 특별한 상태나 드문 체험은 필요 없다. 지금 이 페이지를 보거나 소리를 듣거나 무언가를 만질 때 이미 일어나고 있는 일이면 충분하다. 익숙해서 대개 금세 지나가는 일을 정확히 논할 수 있는 자료로 만드는 것이 우리의 과제다. 이 장을 마치면 첫 관찰 기록과 나중에 검증할 수 있는 몇 가지 질문을 갖게 된다.

## 퀄리아라는 말의 뜻

이 책에서 퀄리아란 주관적 체험의 질적 특성, 곧 무엇이 어떻게 느껴지거나 지각되는가를 말한다. 서늘함과 거친 느낌은 다르며, 날카로운 소리와 둔탁한 소리도 다르다. 마땅한 말이

아직 떠오르지 않더라도 체험하는 사람에게는 차이가 있다. 연구 대상을 가려내는 법부터 익힐 수 있도록 이 용어를 넓은 작업상의 의미로 쓴다.

«주관적»이라는 말은 여기서 체험하는 이의 관점이 있다는 뜻이다. «자의적»이거나 «반드시 틀렸다»는 뜻은 아니다. 만졌을 때 서늘했다고 상당히 확신하며 말하면서도 대상의 온도는 잘못 판단할 수 있다. 이 주장들은 상황의 서로 다른 성질을 다룬다. 온도는 기기 측정값으로 비교할 수 있다. 서늘하다는 감각은 자신의 체험에서 무슨 일이 있었는지를 기술한다.

퀄리아는 감정에 국한되지 않는다. 색과 소리, 신체 감각처럼 정서적 느낌도 체험의 일부가 될 수 있다. 그렇다고 모든 생각을 미리 머릿속 그림으로 바꿀 필요는 없다. 문제를 풀다가 말이 떠오르거나 막연한 어려움을 느끼거나 답이 나오기를 기다리는 자신을 알아차렸다면 바로 그것을 기술해 보자. 남의 본보기에 맞추기보다 무엇을 관찰했는지가 중요하다.

철학에서는 이 용어를 더 강한 의미로 사용하기도 한다. Daniel Dennett은 퀄리아를 특별한 성질로 보고, 그 성질에 말로 표현할 수 없음과 당사자의 오류 없는 접근 등을 부여하는 견해를 비판했다. 그의 논거는 이런 성질에 관한 특정 개념을 겨냥한다. 이 책에서 말로 된 기술이 있다는 사실은 체험의 반박도, 그 체험을 완전히 정확하게 이해했다는 보장도 되지 않는다. [1]

## 사건, 체험과 기록

«나는 손잡이를 만져 서늘함을 느꼈고, 머그잔이 열린 창가에 놓여 있었음을 알았다.» 이 한 문장에는 행동에 관한 보고, 체험에 대한 기술, 앞서 일어난 원인에 대한 추정이 함께 들어 있다. 이 셋을 구별하지 않으면 문장 전체를 직접 관찰한 내용으로 받아들이기 쉽다.

이렇게 나누면 무엇을 검증해야 하는지 알 수 있다. 손이 닿는 장면은 영상에 담을 수 있다. 서늘함은 당사자의 보고로 파악할 수 있다. 열린 창문은 잔의 온도가 그렇게 된 까닭에 관한 별도의 가설이므로 다른 정보가 필요할 수 있다. 그럴듯한 설명도 여전히 설명이다. 곧바로 떠올랐다고 해서 직접 관찰한 내용에 포함되지는 않는다.




그림 1. 한 사건을 기술하는 세 영역. 사건과 체험에 대한 기술, 원인 가설은 같은 상황을 다루지만 서로 다른 질문에 답한다. 오른쪽은 연습용 진술이다. 그림 자체로는 서늘함이 느껴졌는지, 그 물리적 원인이 무엇인지 알 수 없다.

| 진술 | 무엇을 말하는가 |
|---|---|
| 나는 머그잔을 돌렸다 | 상황 속 행동 |
| 반사광이 오른쪽으로 옮겨 가며 더 밝아졌다 | 시각적 변화의 기술 |
| 머그잔이 잘 만들어졌다 | 물건에 대한 판단 |
| 손잡이가 따뜻해져 서늘함이 줄었다 | 변화에 대한 기술과 원인 가설 |

«기분 좋다»와 «불쾌하다»를 이 표의 항목에 기계적으로 넣을 수는 없다. «만졌을 때 기분이 좋았다»는 말은 그 순간의 정서적 느낌을 전할 수 있다. «마음에 드는 머그잔이다»라는 말은 평소의 선호나 전반적인 인상을 나타낼 수 있다. 무엇을 뜻하는지 분명히 해야 한다. 방금 어떤 일이 있었는지, 평소 무엇을 선호하는지, 물건에 대해 어떤 결론을 내리는지 구별해 보자.

## 상세한 보고만으로 충분하지 않은 이유

처음에는 «머그잔이 서늘하다»라고 적었다고 하자. 나중에는 «손이 닿은 곳에서 서늘함이 느껴졌고, 끝날 무렵에는 그 느낌이 덜 뚜렷했다»라고 적었다. 두 번째 기록은 감각이 느껴진 곳을 더 정확히 짚고 그 변화도 드러낸다. 기술 방식이 나아진 것이다. 그러나 말이 많아졌다고 해서 처음의 체험이 얼마나 풍부했는지 알 수 있는 것은 아니다.

기록이 길어지는 이유는 여러 가지다. 물건을 더 자세히 살폈거나, 알맞은 말을 찾았거나, 세부 사항을 기억해 냈거나, 처음에는 없었던 추정을 글에 넣었을 수 있다. 그래서 첫 기록을 남겨 두는 편이 좋다. 결론과 더불어 그 결론에 이른 과정도 비교할 수 있기 때문이다. 수정하여 정리한 글에서는 확신이 어디서 비롯되었는지가 흔히 드러나지 않는다.

기록 시점도 중요하다. 사건이 일어나는 동안, 끝난 직후, 저녁에 기억을 더듬으며 쓸 때의 연구 조건은 서로 다르다. 세 경우 모두 값진 정보를 전할 수 있다. 다만 마지막 경우에는 지금 떠오르는 기억을 통해 과거를 기술한다. «대략 일 분 뒤에 기록함»이라는 짧은 메모만으로도 나중에 읽을 사람과 자신이 자료를 더 잘 이해할 수 있다.

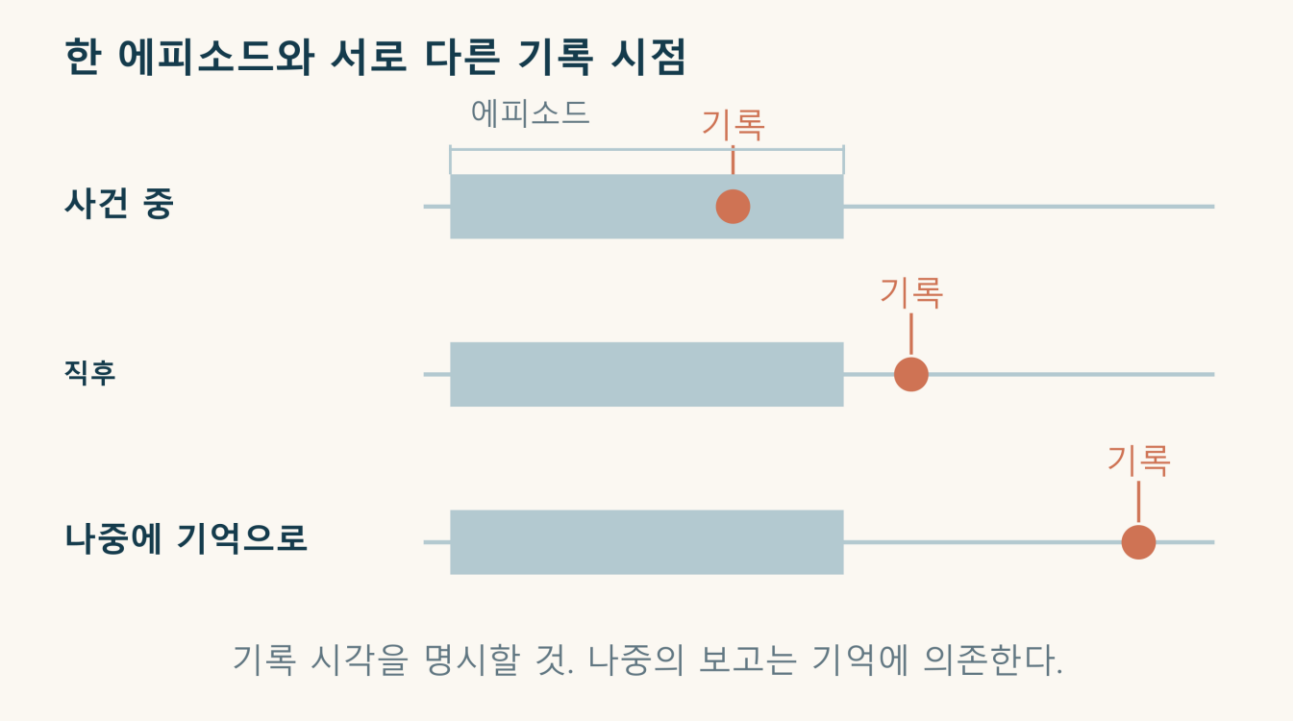


그림 2. 관찰의 일부인 기록 시각. 사건 중의 기술, 직후의 기술, 나중에 기억으로 하는 기술은 서로 다른 조건에서 얻는다. 기록 시점을 밝히면 현재의 보고, 기억, 이후의 설명을 구별할 수 있다.

## 여기서 의식과 성찰은 어디에 있는가

«의식»이라는 말은 서로 연결된 여러 질문을 가리킨다. 그 시스템에는 주관적 체험이 존재하는가? 응답과 행동에 어떤 정보를 이용할 수 있는가? 자신의 결정을 평가할 수 있는가? 자기 생각의 근거를 검토하기 시작할 때 무슨 일이 일어나는가? 과정의 처음에는 이 질문들을 구분하고 그 관계를 점차 밝혀 나갈 것이다.

David Chalmers는 자극 구별이나 상태 보고 같은 기능의 설명과, 왜 일어나는 일에 체험이 수반되는지의 설명을 따로 논의하자고 제안했다. 이는 영향력 있는 철학적 문제 설정이다. 물리적 메커니즘으로 체험을 설명하는 일이 불가능하다고 미리 입증한 것은 아니다. 이 논의는 새 모형이 무엇을 설명했는지 정확히 밝히도록 요구한다는 점에서 우리의 탐구에 유용하다. [2]

성찰은 자신의 체험, 판단, 행동을 살펴보며 재검토할 가능성을 여는 데서 시작한다. 이렇게 알아차릴 수 있다. «온도를 재지 않았는데도 시원함이 약해진 것을 따뜻해졌다는 증거로 여겼다». 이제 다른 어떤 변화가 있었는지 물을 수 있다. 분석은 손의 위치를 유지하고, 시간을 구체화하고, 조건을 비교하는 관찰로 이어진다. 이렇게 성찰은 연구의 일부가 된다.

## 예제 분석

다음은 실습용 기록이다. 발표된 실험의 결과는 아니다. «손잡이를 만졌다. 손가락에서 서늘함이 느껴졌다. 나중에 그 느낌이 약해졌다. 만지는 느낌이 조금 좋았다. 아마 손의 온기로 손잡이가 따뜻해졌을 것이다. 언제 변화를 알아차렸는지는 기억나지 않는다.»

첫 문장은 행동을 나타낸다. 그다음 세 문장은 체험을 기술한다. 다섯째 문장은 원인을 제안하고, 마지막 문장은 기억의 한계를 밝힌다. 어느 것도 버릴 필요는 없다. 연구에는 그 일이 일어난 조건과 체험에 대한 기술, 새로 떠오른 생각이 모두 필요하다. 다만 각각의 역할을 구별해 두어야 한다.

다음 단계는 «약해졌다»는 말을 구체화하는 것이다. 시원한 느낌의 세기가 줄었는가, 아니면 빛의 반사를 살펴보느라 그 느낌을 덜 자주 알아차렸는가? 이 설명들은 아직 확립된 원인이 아니다. 다음 관찰에서 어떤 추가 질문을 다룰지 보여 준다. 좋은 구체화는 멋진 용어를 보태는 데 그치지 않고 불확실성을 좁힌다.

## 체험을 놓치지 않고 기록 하나를 분석하는 방법

다음 네 문장을 보자. «나는 잔의 손잡이를 만졌다. 손가락에 서늘함이 느껴졌다. 이 잔은 쓰기 편하다. 아마 창가에서 식었을 것이다.» 첫 문장은 행동을 보고하고, 둘째는 체험의 질을 기술한다. 셋째는 전반적인 평가를 나타내고, 넷째는 원인을 제안한다. 정답인 문장 하나를 고르는 일이 아니다. 각각의 역할을 살리면 어느 문장이든 쓸모가 있다.

이번에는 «서늘한 느낌이 기분 좋았다»를 더해 보자. 여기서 기분 좋다는 말은 잔의 장점에 대한 평가보다 직접 체험한 정서적 느낌을 나타낼 수 있다. 따라서 개별 단어 대신 문장의 뜻에 따라 표시한다. 판단하기 어렵다면 지금 만졌을 때의 느낌인지, 평소 그 물건을 어떻게 여기는지 구체적으로 따져 보자. 주의 깊은 기술에서 연구 질문이 생겨나는 첫 사례다.

### 여기까지 확인한 것

이제 검증해야 할 대상이 세 가지로 나뉜다. 사건, 체험된 질, 추정된 원인이다. 인상을 길게 열거하는 것보다 셋을 구별하는 일이 더 유용하다. 이제 이 구분을 자신의 짧은 기록에 적용해 보자.

## 실습 1 체험의 첫 기록

실습 목표는 체험에 대한 기술을 사물에 대한 평가 및 원인에 관한 추정과 구별해 기록하는 법을 익히는 것이다.

특별한 준비는 필요 없다. 실온에 놓인 평범한 물건과 종이 한 장, 조용히 보낼 몇 분이면 된다.

예상 소요 시간은 15–20분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 물건을 일 분쯤 살펴보거나 만져 본다. 곧바로 다섯 문장을 적고 언제 기록했는지 밝힌다.
2. 체험에 대한 기술에는 O, 전반적인 평가에는 C, 원인에 관한 가설에는 G를 표시한다. 행동과 외부 조건은 따로 적는다. 한 문장에 여러 내용이 섞여 있으면 나눈다.
3. 모호한 문장 하나를 고른다. 실제로 말할 수 있는 것만 덧붙여 장소, 시점 또는 느낌의 질을 구체적으로 적는다.

4. 관찰을 반복한다. 두 기록을 모두 보관하고, 두 기록 사이의 차이 하나와 이 기록들만으로는 아직 풀 수 없는 질문 하나를 적는다.

날짜와 표시를 적은 기록 두 건, 그리고 다음 관찰에서 다룰 구체적인 질문 하나를 남겨 둔다.

어느 표시를 붙일지 결정할 수 없다면 짐작해서 고르지 않는다. 문장의 뜻을 두 가지로 해석해 적어 본다. 내용을 분명히 할 수 없다는 사실도 앞으로 내릴 결론을 제약한다.

관찰을 이어 가려면 조건 하나만 바꾸어 감각이 달라진 이유에 관한 가설 하나를 검증해 본다. 이를테면 만지는 방법은 그대로 두고 관찰 순서만 바꿀 수 있다. 관찰을 반복하기 전에 예상되는 차이를 적는다.

해설과 답은 부록 F 제1장 참조, 295쪽

**자기 점검 문항**

1. «물건이 따뜻해졌기 때문에 서늘함이 사라졌다»는 말에 체험에 대한 기술 이상의 내용이 들어 있는 까닭은 무엇인가?
2. 기분 좋은 느낌을 체험의 특성으로 볼 수 있는가?
3. 두 번째 기록에 두 배 더 많은 세부 사항이 담겼다면 무엇이 확인된 것인가?

다음 장에서는 이 기록을 본보기로 삼자. 한 사건에서 여러 사건으로 범위를 넓혀, 앞으로의 모형이 보존해야 할 유사 관계를 기술해 볼 것이다.

제2장

# 체험의 차이와 유사성

첫 기록에서는 체험의 질과 그 원인에 관한 이야기를 구별했다. 하지만 감각을 하나씩 규정하는 것만으로는 감각들 사이의 차이가 어떻게 짜여 있는지 드러나지 않는다. 이제 몇 가지 사례를 모아 같은 질문으로 비교해 보자. 유사성 지도는 모형의 품질을 가늠할 첫 자료가 된다. 모형은 사례 사이의 관계를 설명하고 새로운 비교의 결과도 예측해야 한다.

## 개별 인상에서 관계로

머그잔을 만졌을 때 어떤 느낌이 드는지 기록했다. 옆에는 천으로 된 냅킨이 놓여 있다. 두 물건의 느낌은 분명 다르지만 «서로 다르다»는 말만으로는 탐구를 진전시키기 어렵다. 하나는 더 매끄러운 듯하고 다른 하나는 더 따뜻하게 느껴질 수 있다. 한쪽은 손가락의 좁은 부분에서, 다른 쪽은 더 넓은 부분에서 느껴질 수도 있다. 이렇게 차이를 여러 관계로 나누어 비교할 수 있다.

이 장에서는 체험의 질적 구조를 살펴본다. 어떤 특징 때문에 체험이 서로 비슷하게 느껴지는가? 조건을 바꾼 뒤에도 어떤 차이가 남는가? 결과를 어떻게 나타내야 모형이 이를 설명할 수 있는가? 처음으로 쓸 만한 모형을 세울 때는 개별 사례를 하나씩 확정하기보다 사례 사이의 관계부터 살피는 편이 유용할 때가 많다.



그림 3. 한 사례의 표면 비교. 실습에서는 도자기와 직물이 서로 다른 접촉 조건을 제공한다. 알아차린 감각의 질을 비교하고 세기와 지속 시간을 따로 기록한다. 재료의 성질을 적는 것만으로 체험을 기술한 셈이 되지는 않는다.

## 질, 세기와 지속 시간

크기가 같은 두 소리를 떠올려 보자. 하나는 짧은 두드림 소리, 다른 하나는 길게 울리는 소리다. 그 차이를 «더 세다»고만 표현할 수는 없다. 소리의 크기, 울리는 양상, 지속 시간은 서로 다른 질문에 답한다. 마찬가지로 약한 거친 느낌이 더 강해져도 저절로 서늘함이 되지는 않는다. 체험을 비교할 때는 어떤 관계를 살피는지 밝혀야 한다.

질은 감각이 어떤 느낌으로 체험되는지를 뜻한다. 세기는 선택한 측면에서 그 느낌이 얼마나 뚜렷한지를, 지속 시간은 그 느낌이 시간에 걸쳐 얼마나 이어지는지를 나타낸다. 이는 탐구에 쓰는 구분이다. 어떤 체험에도 그대로 적용할 수 있는 독립 좌표의 목록은 아니다. 실제로는 특징들이 서로 영향을 줄 수 있다. 더 오래 이어지는 소리가 더 거슬리게 들리거나, 맥락이 바뀌면서 소리가 더 잘 들릴 수 있다.

질문을 차례로 해 보면 도움이 된다. 무엇이 달라졌는가? 어느 특징에서 달라졌는가? 그 차이는 내내 있었는가, 처음에만 나타났는가? 두 측면을 나눌 수 없다면 현재의 기술로는 둘을 구별할 수 없다고 기록한다. 연구에서 중요한 것은 알맞은 구분을 고르는 일이다. 모든 경우에 같은 설문을 채울 필요는 없다.

## 이름과 체험 속 유사성

같은 말이 서로 다른 감각을 한데 묶기도 한다. 천과 나무, 종이를 모두 «거칠다»고 할 수 있지만 만졌을 때의 느낌은 제각각이다. 반대의 경우도 있다. 이름이 전혀 다른 두 물건이 만졌을 때 어떤 면에서 비슷할 수 있다. 따라서 같은 이름을 쓰는지와 체험이 서로 비슷한지는 따로 기록하는 편이 좋다.

네 가지 표면을 비교한다고 하자. 두 표면은 금속 물건에, 나머지 둘은 플라스틱 물건에 속한다. 재료별로 나누게 하면 한 가지 분류가 나온다. 만졌을 때 얼마나 매끄러운지 물으면 다른 분류가 나올 수 있다. 두 분류 모두 각자의 질문에는 타당하다.

이 차이는 앞으로 세울 모형에 중요하다. 재료의 이름을 잘 맞히는 체계라고 해서 접촉감에 대한 평정들 사이의 관계까지 재현하는 것은 아니다. 모형이 무엇을 예측해야 하는지 미리 정해야 한다. 그래야 편한 과제를 잘 해낸 결과를 다른 현상의 설명으로 잘못 받아들이지 않는다.

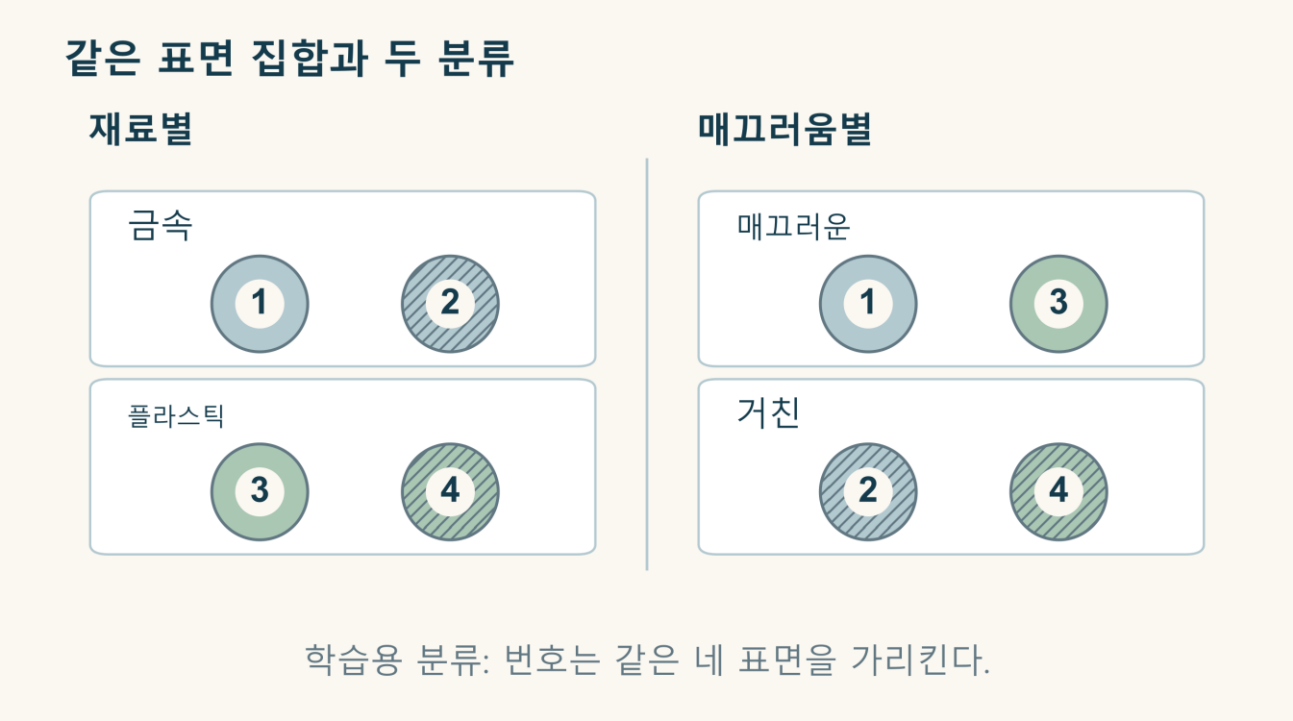


그림 4. 재료의 이름과 접촉감의 유사성. 학습용 집합에는 금속 표면 두 개와 플라스틱 표면 두 개가 있다. 같은 대상들을 재료별로, 그리고 매끄러움별로 묶었다. 선택한 질문에 따라 모형이 재현해야 할 관계가 정해진다.

## 유사성 지도를 만드는 방법

선택한 체험 사례를 A, B, C, D라고 하자. 글자는 관찰 사례를 가리키며 체험의 성질을 뜻하지 않는다. 두 사례를 짝지어 접촉감이 얼마나 비슷한지 판단하거나, 차이에 잠정적인 평정값을 줄 수 있다. 처음에는 순위를 매기기 쉽다. 어느 쌍이 가장 비슷한가? 어느 쌍이 가장 다른가? 어느 쌍은 비교하기 어려운가?

사례가 넷이면 서로 다른 쌍이 여섯 개 나온다. 이 정도면 구조가 드러나기 시작한다. 예를 들어 A와 B가 서로 가깝고 C와 D도 서로 가까운 반면, 두 쌍의 차이는 더 뚜렷할 수 있다. 그러면 두 집단의 잠정적인 지도가 나온다. 각 사례를 별개의 표현으로 적는 것보다 이 지도가 더 유용하다. 모형이 어떤 관계를 재현해야 하는지 보여 주기 때문이다.

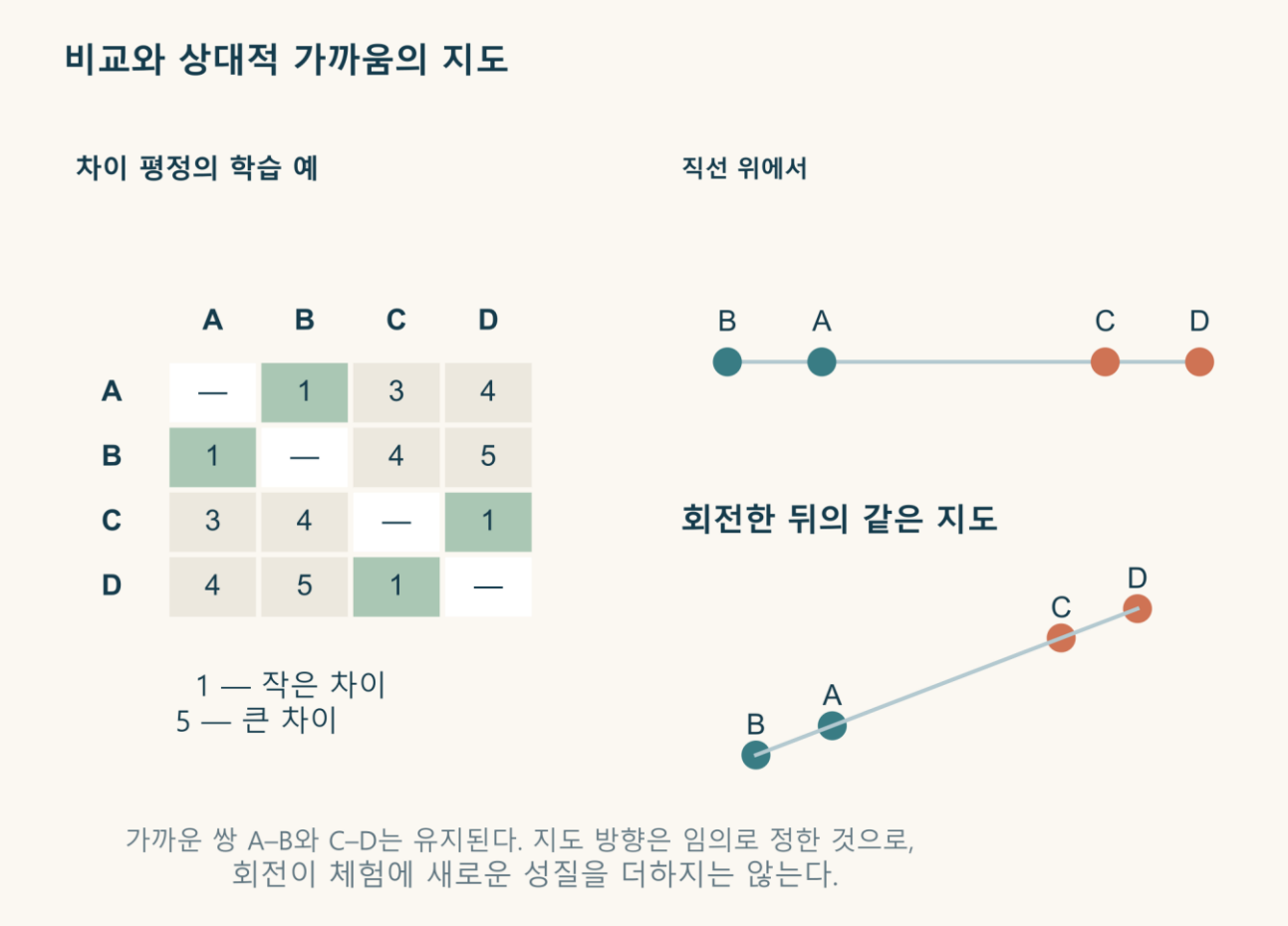


그림 5. 쌍별 평정에서 유사성 지도로. 행렬에는 이 장의 학습용 평정 여섯 개가 있다. 직선 위의 점들은 B, A, C, D 순서이며 좌표는 0, 1, 4, 5다. 오른쪽 아래는 같은 배치를 회전한 것이다. 이는 예제 자료를 편리하게 표현한 방식이며, 체험되는 차이의 척도를 정확한 비율 척도로 간주할 근거는 아니다.

심리 측정의 방법론에서 Roger Shepard는 정확한 거리함수를 모를 때 대상들의 가까움에 관한 자료로부터 공간적 배치를 복원하자고 제안했다. 이런 접근은 유사성 평가의 다차원 구조를 연구하도록 돕는다. 첫 실습에서는 사례들 사이의 관계에 따라 배치한다는 일반적 생각만 사용한다. 완전한 계산 방법에는 추가 가정이 필요하다. [3]

## 실습용 자료로 살펴보는 예

아래 차이 평정값은 실습을 위해 만든 것이다. 일 점은 차이가 작다는 뜻이고, 오 점은 차이가 크다는 뜻이다. 이 수치들은 추론을 설명하기 위한 예시로, 사람을 대상으로 한 연구 결과를 나타내지 않는다.

| 체험의 쌍 | 차이 평정값 |
| --- | --- |
| A와 B | 1 |
| A와 C | 3 |
| A와 D | 4 |
| B와 C | 4 |
| B와 D | 5 |
| C와 D | 1 |

가장 뚜렷한 것은 서로 가까운 두 쌍, A와 B 그리고 C와 D다. 그 안에서도 차이가 하나 더 남는다. A는 D보다 C에 조금 더 가깝다. 두 집단을 그리면서 각 집단의 두 점을 서로 구별할 수 없게 놓으면 일부 정보가 사라진다. 지도에는 집단 구분과 개별 요소의 차이가 모두 나타나야 한다.

연구자가 매끄러움 하나로 결과를 설명했다고 하자. A와 B를 가깝게 놓고, 직선의 반대쪽에 C와 D를 서로 가깝게 놓는다. 이제 나머지 쌍으로 검증한다. 점을 더 옮기지 않아도 그 쌍들의 순서가 맞는가? 맞지 않는다면 두 번째 특징이나 다른 형태의 모형이 필요할 수 있다. 일부 평정값이 불안정했을 가능성도 있다. 어느 설명을 택하느냐에 따라 다음 검증은 달라진다.

평정값 4를 2보다 두 배 큰 체험상의 차이라고 곧바로 해석할 수는 없다. 독자는 숫자로 순서만 나타냈을 수 있다. 정확한 거리를 계산하기 전에 척도의 의미부터 알아야 한다. 처음에는 «이 쌍의 차이가 저 쌍보다 크다»는 판단을 보존하고, 확신하기 어려운 비교를 표시하는 것으로 충분하다.

## 맥락도 과제의 일부가 된다

처음에는 표면을 여유 있게 살펴보고, 다음에는 다른 물건을 만진 직후 같은 표면을 비교했다고 하자. 물건 자체가 그대로여도 비교 조건은 바뀌었다. 두 번째에는 조금 전에 만져 본 사례가

기준으로 남아, 그에 비추어 표현과 평정값을 고르게 된다. 따라서 절차를 기술할 때 제시 순서도 적어야 한다.

질문도 비교의 맥락을 정한다. 매끄러움을 평가하라는 말은 특정한 특징에 주의를 기울이게 한다. 가장 기분 좋은 물건을 고르라는 말은 다른 특징을 살피게 한다. 매 시도마다 질문을 바꾸면 답이 달라진 까닭을 모두 감각의 변화로 돌릴 수 없다. 그 까닭을 알아보려면 과제의 문구를 일정하게 유지하거나 여러 문구를 써서 그 효과를 따로 비교해야 한다.

여기서 불변량이라는 개념을 쓸 수 있다. 살펴보는 변화가 일어나도 그대로 남는 관계를 이른다. A를 K로 바꿔도 모형의 예측이 달라지지 않으면 그 예측은 임의로 붙인 글자와 무관하다. 지도를 돌려도 가까움의 관계가 모두 그대로라면 그림의 «북쪽»에 심리학적 의미를 부여할 이유는 없다. 무엇이 그대로 남아야 하는지 밝히는 일은 무엇이 달라지는지 설명하는 일만큼 중요하다.

### 불필요한 가정 없이 지도를 읽는 방법

예시 표에서 A와 B의 차이는 1이고 C와 D의 차이도 1이다. B와 D의 차이는 5다. 처음에는 순서만 되살려도 된다. 앞의 두 쌍은 가깝고 마지막 쌍은 뚜렷하게 다르다. 직선 위에 B, A, C, D를 각각 0, 1, 4, 5에 놓으면 이 예의 수치 여섯 개가 모두 거리와 일치한다. 이 배치가 잘 맞는다고 해서 모든 주관적 척도가 길이의 성질을 갖는 것은 아니다. 자신의 응답을 다룰 때는 가까움의 순서를 먼저 확인하고, 척도의 근거가 마련된 뒤에만 정확한 산술 관계를 사용한다.

#### 여기까지 확인한 것

정해 둔 질문으로 사례 사이의 관계를 비교한다. 지금은 감각마다 유일하게 옳은 위치를 정할 필요가 없다. 원래의 응답을 남겨 두고, 그림이 어느 부분에서 응답과 맞는지, 또 어디에 별도의 가정이 더해졌는지 살피는 것이 중요하다.

## 실습 2 네 가지 체험과 하나의 질문

실습 목표는 검증할 수 있는 유사성 지도를 만들고, 지도가 보여 주는 성질과 임의로 정한 표기를 구별하는 것이다.

준비할 것은 제1장에서 익힌 기술과 원인의 구별, 손쉽게 만져 볼 수 있는 표면 네 가지, 도중에 바꾸지 않을 비교 질문 하나다.

예상 소요 시간은 25–35분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 네 표면에 A, B, C, D라는 이름을 붙이고 만졌을 때 각각 어떤 느낌이 드는지 짧게 적는다.
2. 비교할 질문을 하나 적는다. 이를테면 만졌을 때의 느낌이 얼마나 비슷한지 물을 수 있다. AB, AC, AD, BC, BD, CD의 여섯 쌍을 빠짐없이 비교한다.
3. 차이가 작은 쌍에서 큰 쌍으로 순서를 매긴다. 차이가 같거나 판단이 어려운 경우 그대로 표시한다. 네 점을 먼저 직선 위에, 다음으로 평면에 그린다. 더 잘 드러난 관계 하나와 여전히 드러나지 않은 관계 하나를 표시한다.
4. 같은 질문을 유지한 채 순서를 바꿔 다시 비교한다. 두 지도를 모두 남기고, 응답이 달라졌다는 사실과 달라진 원인에 대한 추정을 구별한다.

여섯 쌍을 비교한 결과, 두 번 그려 본 지도, 안정적인 관계와 불안정한 관계의 목록을 남겨 둔다.

그림을 회전해도 거리는 바뀌지 않는다. 두 번째 좌표는 첫 배치보다 어떤 관계를 더 잘 나타낼 때에만 쓸모가 있다.

탐구를 이어 가려면 다섯 번째 표면을 더한다. 새 비교에 앞서 제안한 설명을 바탕으로 새 표면과 기존 표면 사이의 유사 관계를 적어도 두 쌍에 대해 예측한 뒤 검증한다. 답이 나올 때마다 점을 손으로 옮기는 것은 예측이 아니다.

해설과 답은 부록 F 제2장 참조, 295쪽

## 지도에서 가설로

이 지도는 응답에 나타난 구조를 기술한다. 다음 과제는 그 구조를 어떤 메커니즘과 연결하는 일이다. 예를 들어 평정값이 두 특징의 조합에 따라 달라진다고 가정하고, 지도를 만드는 데 쓰지 않은 새 사례로 검증할 수 있다. 새 사례의 위치를 매번 손으로 정해야 한다면 아직 설명을 완성한 것이 아니다.

이후에는 인공 체계의 상태들 사이에서 어떤 관계가 발견한 차이와 대응할 수 있는지 묻는다. 그런 연구의 첫 요건은 이미 분명하다. 새로 비교할 때 어떤 구조가 나타날지 예측해야 한다. 그 체계의 상태에 실제로 체험이 수반되는가라는 질문에는 별도의 이론적 논의가 필요하다.

**자기 점검 문항**

1. 같은 이름을 쓴다고 해서 체험에서 느끼는 질까지 같다고 할 수 없는 까닭은 무엇인가?
2. 차이 평정값 4는 1에서 5까지의 잠정적인 척도에서 무엇을 뜻하는가?
3. 좌표를 추가하는 일이 그림을 보기 좋게 하는 데 그치지 않고 설명에 도움이 되는지 어떻게 검증할 수 있는가?

지도는 체험 사이의 차이를 기술했다. 이제 그 내용이 자신의 몸과 행동, 지나온 일과 어떤 관계가 있는지 살펴본다.

제3장

# 자신을 느낀다는 것은 무엇인가

앞에서 체험의 질을 비교하는 법을 익혔다. 하지만 비슷하게 느껴지는 두 경험도 그 경험을 하는 사람과 서로 다른 방식으로 관련될 수 있다. 손이 움직이는 모습과 포인터가 움직이는 모습은 모두 보이지만, 손은 몸의 일부로 느껴지고 포인터는 대개 바깥 물건으로 남는다. 이제 관점, 신체 소유감, 행위주체감, 시간에 걸쳐 형성된 자기 표상을 구별해 보자. 이 구분은 나중에 사건의 원천을 설명하는 모형을 세울 때 필요하다.

## 한 말의 여러 의미

연필에 손을 뻗어 집고 «내가 집었다»고 말한다. 이 짧은 말에는 여러 관계가 들어 있다. 연필은 특정한 위치에서 보였다. 손은 자신의 몸 일부로 느껴졌다. 움직임은 자신이 한 행동으로 지각되었다. 연필이 필요했던 까닭도 기억한다. 일상 대화라면 이 정도로 충분하다. 연구에서는 이 가운데 어떤 관계를 설명하려는지, 그 관계들이 항상 함께 나타나는지 물어야 한다.

먼저 관점부터 보자. 여기서 보면 탁자의 한쪽이 더 가깝고 다른 쪽은 일부 가려져 있다. 소리는 지금 있는 곳을 기준으로 왼쪽에서 들린다. 관점은 어떤 지점이나 기준계를 중심으로 내용이 배열되는 방식을 뜻한다. 자기 삶에 관해 아는 것과는 별개다. 오른쪽의 탁자 가장자리를 보는 데 마음속으로 이름을 말하거나 직업을 떠올릴 필요는 없다. 로봇의 카메라를 기준으로 좌표를 계산했다는 사실만으로 로봇이 어떤 관점을 체험한다는 결론도 나오지 않는다. 체계를 기술할 때는 좌표의 구성과 상황이 어떻게 체험되는지를 구별해야 한다.

두 번째 질문은 신체 소유감에 관한 것이다. «이건 내 손이다»라는 느낌은 법적 소유권에 대한 판단과 다르다. 몸의 일부를 어떻게 자신의 것으로 체험하는지가 문제다. 세 번째는 «이 움직임은 내가 일으킨다»는 행위주체감이다. 네 번째는 시간에 걸친 자기 표상이다. 자신을 어떤 사람으로 여기고, 지난 일을 어떻게 기억하며, 미래에 어떤 의도를 두는가가 여기에 속한다. 이 네 질문은 탐구의 길잡이가 된다. 다만 «나»라는 말의 모든 쓰임을 포괄하지는 않는다.

## 신체 소유감과 행위주체감

먼저 의도적으로 손을 들어 올린 뒤 탁자에 가만히 내려놓는다. 움직임이 멈춘다고 해서 손이 자기 몸의 일부가 아니게 되지는 않는다. 따라서 신체 소유감을 몸을 뜻대로 계속 움직이는 일로 규정할 수는 없다. 이번에는 마우스를 움직여 보자. 화면의 포인터가 움직이는 것은 자기 행동의 결과로 느껴지지만, 포인터 자체는 대개 몸의 일부로 느껴지지 않는다. 대상을 제어하는 것과 그것을 몸의 일부로 느끼는 것은 서로 다른 관계다.

Andreas Kalckert와 Henrik Ehrsson의 연구에서 참가자들은 자기 손의 움직임과 연결된 인공 손의 움직임을 관찰했다. 저자들은 움직임의 일치 정도와 성격, 모형의 위치를 바꾸었다. 그 결과 신체 소유감과 행위 주체감의 평가가 갈리는 조건들이 나타났다. 이 연구는 신체적 자기지각의 두 구성요소를 나누어 연구하는 것을 뒷받침한다. 각각을 별도의 주체로 만드는 것은 아니다. [4]

여기서 한 가지 더 구별할 필요가 있다. 행위주체감은 체험이나 그 체험에 관한 보고의 특성이다. 실제로 인과관계에 관여했는지는 사건의 특성이다. 키를 누른 뒤 결과를 알아차리지 못해도 프로그램에 영향을 줄 수 있다. 반대로 자신이 움직였다고 예상했는데 실제로는 녹화된 움직임이 재생된 경우도 있을 수 있다. 인과적 관여를 검증하려면 상황이 어떻게 구성되었는지 알아야 한다. 체험되는 행위주체감을 연구하려면 참가자가 제공하는 자료도 필요하다. 두 질문은 서로 대신할 수 없다.

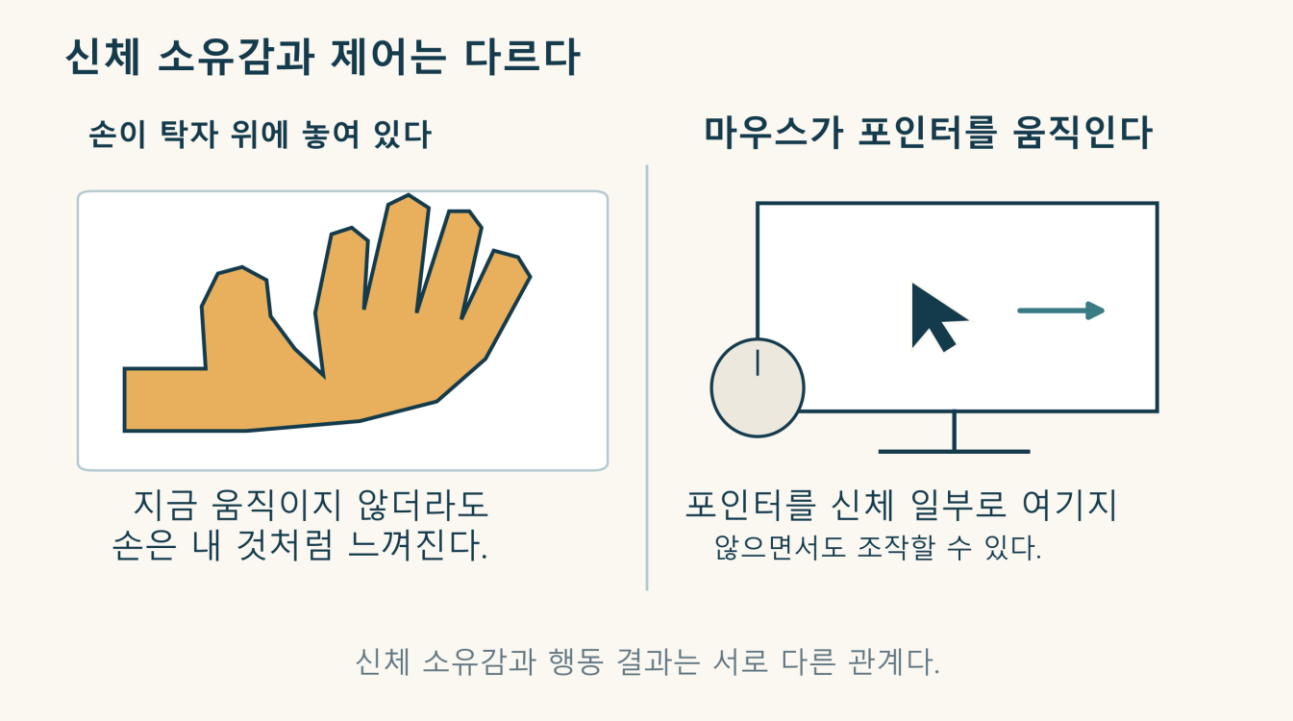


그림 6. 움직이지 않는 손의 신체 소유감과 외부 대상의 제어. 움직이지 않는 손도 계속 자기 몸의 일부처럼 느껴질 수 있다. 포인터 제어는 행동과 결과 사이의 관계를 보여 준다. 그 자체로 포인터가 몸의 일부처럼 느껴진다는 뜻은 아니다.

## 보지 않고도 몸의 위치를 아는 방식

탁자 위에 놓인 손을 일부러 보지 말고 어떤 자세인지 말해 보자. 손가락이 굽은 정도, 탁자에 닿는 압력, 팔꿈치의 대략적인 위치를 느낄 수 있을지도 모른다. 몸의 각 부분이 어디에 있고 어떻게 움직이는지에 관한 정보를 고유수용 정보라고 한다. 이 정보는 시각과 촉각, 그 밖의 정보와 함께 몸의 위치를 파악하는 데 쓰인다. 고유수용감각은 이런 정보와 감각 능력을 가리키는 말이다. 몸 안에 작은 의식이 따로 있다는 뜻은 아니다.

연구자는 여러 정보원이 얼마나 일치하는지 살핀다. 손이 느껴질 것이라 예상한 곳에 실제로 손이 보이면, 여러 정보를 동일한 신체 부위에 연결하는 설명을 세울 수 있다. 정보가 어긋나면 체계가 어느 정보에 더 큰 비중을 두는지, 또 그 이유는 무엇인지 물을 수 있다. 여기서는 연구 문제의 윤곽만 그렸다. 제11장에서는 이를 계산 가능한 모형으로 만들고 여러 조건에서 무엇이 일어날지 예측한다.

몇 가지 신호가 일치한다는 사실만으로 몸에 속한다는 느낌을 설명할 수는 없다. 전구 두 개가 동시에 깜박인다고 해서 몸의 일부가 되는 것은 아니다. 모형은 어떤 정보원을 비교하고, 대상의 형태와 위치, 지금까지의 상호작용, 다른 설명의 가능성을 어떻게 고려하는지 밝혀야 한다. 이런 내용을 명시할 때 «뇌가 나를 조립한다»는 인상적인 비유가 구체적인 연구 과제가 된다.




그림 7. 한 사례에서 자신과 맺는 네 가지 관계. 관점, 신체 소유감, 행위주체감, 살아온 이력에 관한 자기 기술을 구분했다. 단어 사이의 공백을 고친 주체가 사람인지 프로그램인지 밝히는 일은 그 변경의 주체를 확인하는 일이다. 그 답만으로 다른 관계까지 정해지지는 않는다.

## 텍스트 편집기에서 살펴보는 예

실습을 위한 상황을 떠올려 보자. 마리나는 문서의 문장 하나를 옮긴다. 마우스를 움직여 선택된 글을 보고 그 문장을 문단 맨 앞에 놓으려 한다. 버튼을 놓자 프로그램이 단어 사이의

공백을 자동으로 고친다. 마리나는 «내가 문장을 옮기고 서식도 고쳤다»고 말한다. 이 말이 일어난 일을 모두 정확히 나타낸다고 전제하지 말고 살펴보자.

문장을 옮기는 것은 마리나가 의도하고 실행한 행동이었다. 공백을 고친 것은 프로그램의 자동 처리였을 수 있다. 마리나는 결과를 눈으로 볼 수 있고 단어의 배치도 확인할 수 있다. 그렇다고 결과가 생긴 원인을 모두 알 수 있는 것은 아니다. 연구 기록에는 마우스 움직임, 문서의 변경, 마리나의 보고, 연구자가 아는 자동 수정 규칙을 각각 적는 편이 좋다.

변경 이력을 본 뒤 마리나가 «공백은 프로그램이 고쳤다»고 바로잡았다고 하자. 이는 누가 행동을 일으켰는지에 관한 판단을 고친 것이다. 그 때문에 처음에 느낀 제어감이 지어낸 것이 되지는 않는다. 의도와 행동, 전체 결과 사이의 연결에서 그런 느낌이 생겼을 수 있다. 연구에서는 이 연결을 설명해야 한다. 그 연결이 결과를 지나치게 넓게 자기 행동의 몫으로 돌리게 만드는 경우도 포함된다.

자기가 어떤 사람인지에 대한 기술은 다른 수준에서 나타난다. «나는 꼼꼼한 편집자다»라는 말이 그 예다. 공백 하나가 자동으로 고쳐진 사실만으로는 그 말을 검증할 수 없다. 이 말은 여러 상황을 아울러 내린 판단이며 무엇을 꼼꼼함으로 보느냐에 따라 달라진다. 수준을 뒤섞으면 변경의 원인을 잘못 짚은 사소한 일이 성격에 대한 근거 없는 결론이 된다. 각각을 따로 기록하면 검증할 수 있는 구체적인 질문이 남는다.

## 내부 관찰자가 필요한가

자기의식을 내부 화면을 바라보는 관객에 비유하기도 한다. 설명이 필요한 대목을 가리키는 데에는 편리한 이미지다. 그러나 그 관객이 화면을 알아보고 과거를 기억하며 화면에 보이는 것이 자기와 관련된 것임을 이해해야 한다면, 처음의 문제가 축소된 채 안에 다시 놓인 셈이다. 이제는 그 관객이 어떻게 그런 일을 할 수 있는지 설명해야 한다.

체계가 어떤 정보를 자기와 관련짓고 그 관계를 어떻게 쓰는지 묻는 편이 더 생산적이다. 예를 들어 프로그램은 자신의 식별자를 저장하고, 자신이 내린 명령과 외부에서 들어오는 메시지를 구별하며, 자기 응답의 신뢰도를 평가할 수 있다. 이는 별개의 기능으로 구현하고 검증할 수

있다. 식별자가 있다고 해서 체험되는 «나»가 입증되는 것은 아니지만, 정보가 어떻게 자기와 연결되는지 연구할 명확한 대상은 생긴다.

명시적인 성찰에는 한 단계가 더 있다. 사람이나 체계가 자기 판단 자체를 검증 대상으로 삼는다. «탁자 가장자리가 보인다»는 말과 «내가 그것을 정확히 본다고 확신하는 까닭은 무엇인가»라는 질문은 서로 다른 역할을 한다. 최소한의 주체를 다루는 가설에서 복잡한 언어적 성찰을 은연중에 모든 체험의 필수 조건으로 삼아서는 안 된다. 그런 조건에는 별도의 근거가 필요하다.

## 제어와 소유감이 같지 않은 이유

첫째 사례에서는 손이 탁자 위에 가만히 놓여 있다. 지금 움직이지 않더라도 자세와 닿는 느낌이 자기 몸에 속한 것으로 체험될 수 있다. 둘째 사례에서는 마우스를 움직이면서 포인터를 본다. 포인터는 그 명령에 따라 움직이지만 몸의 일부가 되는 것은 아니다. 셋째 사례에서는 어제 자신이 쓴 메모를 읽는다. 메모를 쓴 일은 자신의 이력에 속하지만, 지금 읽는 것은 지금 쓰고 있다는 체험과 다르다. 이 사례들은 기술을 여러 항목으로 나누어야 할 까닭을 보여 준다. «이건 나다»라는 하나의 답으로는 차이가 가려진다.

### 여기까지 확인한 것

네 질문은 같은 사례 안의 서로 다른 관계를 다룬다. 그 관계들이 언제나 독립적이라고 가정할 필요는 없다. 지금은 어떤 질문에 답할 수 있고 어느 관계를 살피고 있는지 알아차리는 것으로 충분하다.

## 실습 3 세 가지 사례에 던지는 네 가지 질문

실습 목표는 일상의 세 가지 사례에서 자기와 관련된다는 말이 서로 어떤 뜻으로 쓰이는지 구별하는 것이다.

제1장의 사건과 보고의 구별을 활용한다. 실제로 사건을 일으키는 데 관여했는지와 자신이 그 행동을 했다고 느끼는지는 각각 다른 질문임을 기억한다.

예상 소요 시간은 20–25분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 세 가지 사례를 기술한다. 탁자 위에 손을 가만히 올려놓기, 포인터 움직이기, 예전에 자기가 쓴 글 읽기다.
2. 각 사례의 관점, 체험되는 신체 소유감, 지금 행동의 행위주체감, 과거의 자신에 관해 아는 점을 따로 적는다.
3. 독립적으로 확인할 수 있는 주장 하나를 고른다. 무엇이 움직임을 일으켰는지, 또는 글이 언제 작성됐는지 확인할 수 있다.
4. 확인한 내용을 첫 기술과 대조한다. 차이가 있다면 기록에 남긴다. 그 차이로 성격 전반을 평가하지 않는다.

네 질문을 구분해 답한 세 사례의 기록과 독립적으로 검증한 관계 하나를 남겨 둔다.

말로 표현하지 못한다고 해서 그 구성 요소가 없다는 뜻은 아니다. 어느 질문에 아직 답하기 어려운지, 무엇을 더 알면 답할 수 있을지 적을 수 있다.

탐구를 이어 가려면 행위주체감과 신체 소유감을 서로 다르게 바꿀 수 있는 조건을 제안한다. 지표마다 따로 예측하고 어떤 외부 사건을 통제해야 하는지 설명한다.

해설과 답은 부록 F 제3장 참조, 296쪽

**자기 점검 문항**

1. 화면의 포인터를 제어할 수 있다는 사실로 그것을 몸의 일부처럼 체험한다고 입증할 수 없는 까닭은 무엇인가?
2. 행동을 실제로 일으켰다는 사실과 행위주체감은 어떻게 다른가?
3. 자기 이름을 아는 것만으로는 체험 속의 자기를 온전히 설명하는 모형에 충분하지 않은 까닭은 무엇인가?

자신과 맺는 여러 관계를 구별했으니 이제 자기 사고를 평가하는 문제로 넘어간다. 다음 장에서는 이해했다는 느낌으로 답이 맞는지 얼마나 잘 판단할 수 있는지, 그리고 이 관계를 어떻게 검증할지 묻는다.

제4장

# 자기 생각을 알아차리는 방법

스스로 내린 판단을 알아차리는 것도 자신을 느끼는 일의 일부다. 알아차린 판단은 다시 평가해야 한다. «다 이해했다»는 익숙한 느낌은 때로 검증보다 먼저 온다. 이 장에서는 답이 맞는지, 피드백을 받기 전에 얼마나 확신했는지, 이후에 어떻게 설명하는지를 구별한다. 그 순서를 지키면 검증 뒤 무엇이 바뀌었는지 알 수 있고, 성찰을 다음 행동의 근거로 삼을 수 있다.

## 생각과 생각에 대한 지식

다음 문장을 읽어 보자. «상자에 연필은 펜보다 적고 지우개보다 많다.» 다 이해했다고 느끼고 다음으로 넘어갈 수 있다. 잠시 멈추어 어느 물건이 가장 많은지 물을 수도 있다. 두 번째 경우에는 실제로 답해 보면서 처음의 느낌을 시험한다. 답하려고 조건을 다시 읽어야 했다고 해서 반드시 이해하지 못한 것은 아니다. 다만 이해했다고 느끼는 것과 주어진 과제를 수행할 준비가 된 것은 서로 다른 자료임이 드러난다.

이 책에서 자기관찰은 현재 또는 방금 겪은 자신의 상태에 주의를 돌리는 것을 뜻한다. 성찰은 자신의 사고, 행동, 판단을 분석과 수정의 대상으로 검토하는 것이다. 메타인지는 인지를 평가하고 조절하는 더 넓은 과정들의 묶음이다. 예를 들어 어려운 응답을 다시 검증하기로 한 선택은 그 신뢰성에 대한 평가에 달려 있을 수 있다. 이런 작업 정의는 서로 다른 질문을 설정하도록 돕는다. 과학 문헌에서는 용어의 경계를 다르게 그을 수 있다.

생각이 떠올랐다고 해서 그 생각이 어떻게 생겼는지를 자세히 아는 것은 아니다. 익숙한 단어를 금세 알아봤다고 하자. 알아봤다는 결과는 알 수 있지만 글자들을 어떤 순서로 처리했는지는 모를 수 있다. «먼저 글자 하나하나를 살폈다»는 이야기는 검증을 거쳐야 한다. 실제로 알아차린 단계를 기술한 것일 수도 있고 단어를 알아본 뒤 그럴듯하게 재구성한 것일 수도 있다.

## 진솔함이 정확성을 보장하지 않는 이유

자기 상태를 매우 진솔하게 보고해도 내용이 부정확할 수 있다. 반드시 기억이 나빠서 그런 것은 아니다. 어떤 차이를 알아차려야 하는지, 그 차이를 어떤 말로 표현해야 하는지 분명하지 않을 때도 있다. «말의 형태로 생각했는가»라는 질문에 답하려면 마음속에서 실제로 말한 문장, 그 뜻을 아는 것, 답하는 도중 떠오른 문장을 구별해야 한다.

Eric Schwitzgebel은 일상적인 자기관찰이 얼마나 불확실한지 다룬 철학 논문에서, 지금의 자기 체험을 틀림없이 안다는 생각을 의심할 근거를 살핀다. 시각적 선명함, 감정, 사고의 형태 등을 예로 든다. 이 논거는 자기 기술에 저절로 오류가 없다는 견해를 반박한다. 사람마다 얼마나 자주 틀리는지를 측정한 연구도, 자기보고를 포기할 근거도 아니다. [5]

실습에서는 체험한 시점, 처음 보고한 시점, 나중에 설명한 시점을 구분한다. 간단하지만 효과가 있는 원칙이다. 문제를 풀 때 답이 먼저 떠오르고 말로 된 근거가 나중에 생겼다면 그대로 적는다. 그 사이는 일어났을 법한 사고 단계로 메울 필요가 없다. 지어낸 과정으로 잇는 것보다 빈 부분을 남겨 두는 편이 유용하다.

## 확신이 무엇에 관한 것인지 밝혀야 한다

«팔십 퍼센트 확신한다»는 말은 무엇을 확신하는지 밝히고 나서야 연구에 쓸 수 있다. 답이 맞는다는 말인가? 기억이 정확하다는 말인가? 내일 그 설명을 다시 할 수 있다는 말인가? 각각 다른 일이며 확인하는 방법도 다르다. 무엇에 관한 수치인지 밝히지 않으면 정밀해 보일 뿐이다.

답을 확인할 수 있는 과제에서는 확신도가 해당 답이 맞을 확률을 뜻한다고 정할 수 있다. 어떤 사람이 여러 답에 80퍼센트의 확신도를 매겼다면, 그런 답 열 개 가운데 대략 여덟 개가 맞는 편이 바람직하다. 이것이 확신도의 확률보정이라는 생각이다. 한 번의 정답이나 오답으로 보정이 잘됐는지 평가할 수는 없다. 확률적 예상은 서로 비교할 수 있는 여러 사례에 걸쳐 검증한다.

Stephen Fleming과 Hakwan Lau는 높은 확신도를 부여하는 경향, 확신도로 정답과 오답을 구별하는 능력, 주과제의 성공률에 대비한 그런 평가의 효율을 구분한다. 이는 서로 다른 측정 질문이다. 평균 확신도가 높다는 것만으로 자신의 오류를 잘 구별한다고 볼 수는 없다. [6]

과제와 답, 검증 전의 확신도, 검증 결과를 모두 남긴다. 복잡한 계산으로도 잃어버린 최초의 예측을 되살릴 수는 없다. 풀이를 본 뒤에 적은 확신도는 다른 질문에 답한다.

## 두 독자의 응답을 살펴보는 예

이 장을 위해 만든 실습용 자료를 보자. 안톤과 베라는 같은 네 문제를 풀었다. 둘 다 첫째와 셋째는 맞혔고 둘째와 넷째는 틀렸다. 이 짧은 문제 묶음에서 두 사람의 성적은 같다. 둘은 피드백을 받기 전에 확신도를 기록했다.

| 과제 | 정답 여부 | 안톤의 확신도 | 베라의 확신도 |
|---|---|---|---|
| 첫 번째 | 정답 | 90퍼센트 | 90퍼센트 |
| 두 번째 | 오답 | 90퍼센트 | 40퍼센트 |
| 세 번째 | 정답 | 90퍼센트 | 80퍼센트 |
| 네 번째 | 오답 | 90퍼센트 | 30퍼센트 |

안톤은 모든 답에 같은 확신도를 매겼다. 이 예에서 베라는 실제로 틀린 답에 더 낮은 확신도를 매겼다. 맞힌 문항의 수가 같아도 자기 점검의 질이 같은 것은 아니다. 두 문항만 다시 볼 수 있다면 베라의 평정은 점검이 더 필요한 곳에 노력을 쓰게 해 준다. 자기 사고를 평가하는 일에는 이런 실용적 기능이 있을 수 있다.

과제 네 개만으로 한 사람의 능력이 안정적인지 알 수는 없다. 우연히 그런 결과가 나왔거나 척도를 서로 다르게 이해했을 수 있다. 검증하려면 과제를 더 주고 같은 지시를 사용해야 한다. 이 표가 보여 주는 것은 답을 평가한 방식의 질이다. 베라가 «더 의식적이다»라고 결론 내릴 근거는 되지 않는다.

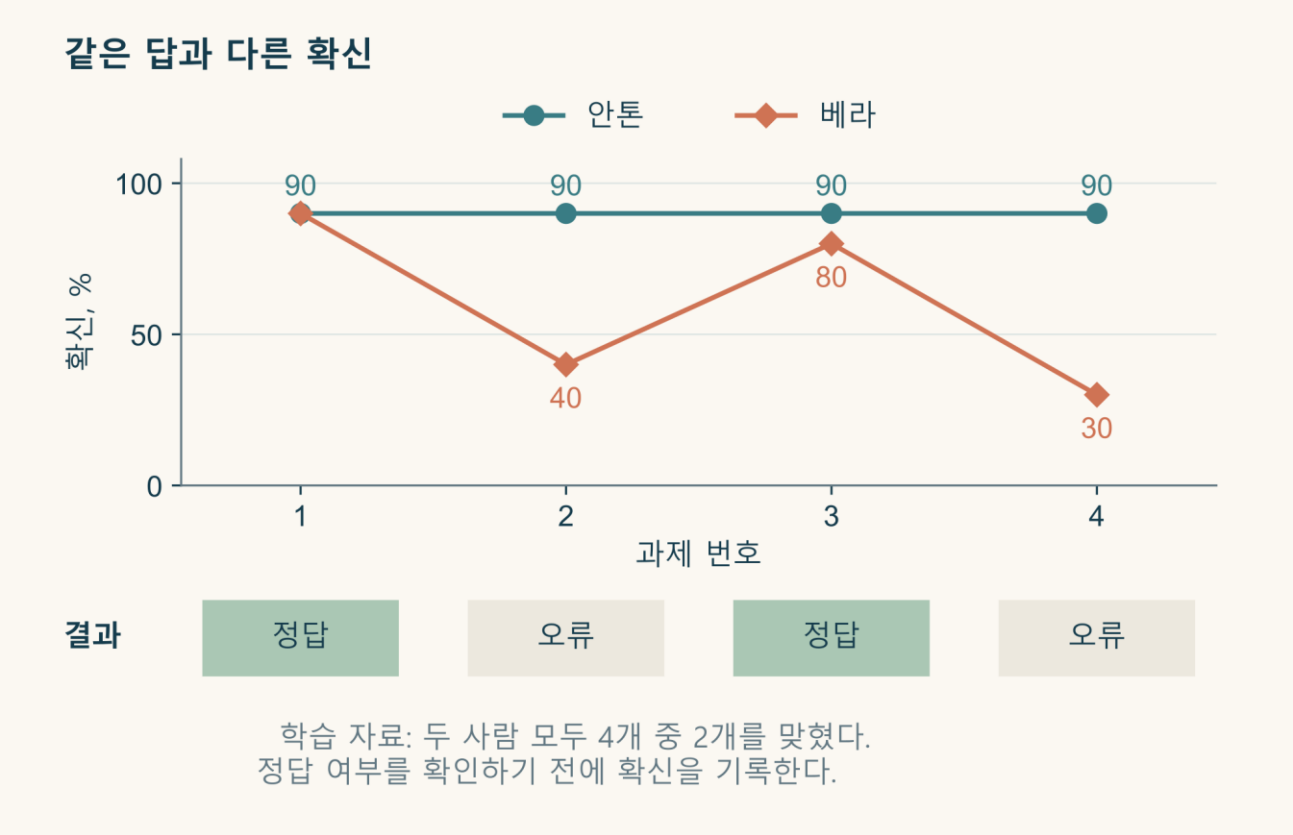


그림 8. 과제 성적이 같을 때의 확신도. 실습용 자료에서 안톤과 베라는 과제 1과 3을 맞힌다. 안톤은 모든 답에 90퍼센트, 베라는 각각 90, 40, 80, 30퍼센트의 확신도를 매긴다. 이 네 사례에서 베라의 평정은 나중에 드러날 오답을 가려낸다. 안정적인 능력인지 결론 내리려면 새로운 자료가 필요하다.

## 자기관찰에서 풀이 방식 바꾸기로

성찰은 다음에 무엇을 할지 검증할 수 있는 방식으로 바꿀 때 성과를 낸다. 가령 «비교 문장을 읽을 때 첫 번째 대상 쌍은 기억하지만 세 번째 쌍은 자주 놓친다»고 알아차렸다고 하자. 여기서 관계를 화살표로 적어 보자는 제안이 나온다. 그다음 같은 유형의 새 과제에서 오류가 줄었는지 검증할 수 있다. 어려움을 관찰하고, 원인을 추정하고, 절차를 바꾸어 새 결과를 얻는 짧은 탐구 과정이 이렇게 시작된다.

절차를 바꾼 결과가 좋았다고 해서 추정한 원인이 모두 입증된 것은 아니다. 화살표를 그리느라 단순히 풀이 시간이 길어져 도움이 되었을 수 있다. 설명을 구별하려면 같은 시간을 쓰되 화살표를 그리지 않는 조건이 필요하다. 검증을 다루는 장에서 이 문제를 자세히 살핀다. 지금은 «좀 더 주의하자»는 막연한 생각이 여러 설명을 놓고 검증할 수 있는 구체적인 변화로 이어지는 과정이 중요하다.

자기 조절은 길게 자기 자신을 분석하지 않고도 일어날 수 있다. 읽는 방식에 대한 이론을 세우지 않고도 확신이 부족하다는 것을 알아차려 한 줄을 다시 읽을 수 있다. 명시적인 성찰은 판단의 근거, 행동 방식, 결론을 내릴 수 있는 범위를 살피기 시작할 때 이루어진다. 오류를 자동으로 고치는 일과 그 원인을 말로 따져 보는 일은 따로 살펴야 한다.

## 정보를 얻은 순서를 남기는 법

한 참가자가 답과 90퍼센트의 확신도를 적은 뒤, 틀린 것을 확인하고 처음부터 확신이 없었다고 말했다고 하자. 나중에 한 말도 진솔할 수 있지만 새로운 정보를 얻은 뒤에 나온 말이다.
연구에서는 처음 예측과 이후의 해석을 모두 남긴다. 처음 기록을 둘째로 바꾸면 처음에 얼마나 정확하게 자신을 점검했는지 평가할 수 없게 된다. 그래서 아래 실습은 문제와 답을 분리해 답을 책 끝의 해설에 실었다.

### 여기까지 확인한 것

자기관찰이 완벽할 필요는 없다. 무엇을 언제 확신했는지 명시하는 것으로 충분하다. 최초의 기록을 보존하면 나중에 드러난 오류를 절차 개선의 자료로 쓸 수 있다.

## 실습 4 검증 전에 예측하기

실습 목표는 미리 적은 확신도를 결과와 비교하고 다음 검증에 쓸 구체적인 방법을 고르는 것이다.

체험한 시점, 처음 기록한 시점, 나중에 설명한 시점을 구별한다. 실습에는 종이만 있으면 되며 형식 논리를 배웠을 필요는 없다.

예상 소요 시간은 15–20분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 해설을 보지 않고 세 문제를 푼다. 상자에 연필은 펜보다 적고 지우개보다 많다. 무엇이 가장 적은가? 초록색 카드는 모두 둥글고, 둥근 카드 중 일부는 크다. 크고 초록색인 카드가 반드시 있는가? 니나가 올레그보다 먼저 왔고 올레그가 표트르보다 먼저 왔다. 맨 마지막에 온 사람은 누구인가?
2. 각 답의 확신도를 0에서 100퍼센트 사이로 적고, 알아차린 풀이 과정을 한 문장으로 적는다.

3. 세 답을 모두 기록한 뒤 실습 4의 해설을 본다. 답이 맞았는지와 이해한 내용이 어떻게 바뀌었는지를 따로 적는다.

4. 막혔던 지점 하나를 골라 반례 찾기 같은 다음 방법을 정한다. 문제 세 개만으로 장기적으로 유지되는 능력을 평가하지 않는다.

처음의 확신도를 포함한 답 세 개, 검증 결과, 다음에 적용할 구체적인 행동 규칙 하나를 남겨 둔다.

«일부»라는 말은 둥근 카드 가운데 일부에 관해 말한다. 아직 해설을 보지 말고 주어진 조건에 맞는 몇 가지 배치를 떠올려 본다.

탐구를 이어 가려면 아직 풀지 않은 비슷한 문제에서 선택한 방법을 시험한다. 처음 기록한 확신도는 바꾸지 않는다. 정답을 내는 능력과 다시 검증할 만한 사례를 가려내는 능력을 따로 비교한다.

해설과 답은 부록 F 제4장 참조, 297쪽

**자기 점검 문항**

1. 정답을 본 뒤의 설명을 처음에 어떤 생각을 했는지 보여 주는 기록으로 곧바로 받아들일 수 없는 까닭은 무엇인가?
2. 답을 자주 맞히면서도 자신의 확신도를 기준으로 오답을 잘 가려내지 못할 수 있는가?
3. «나는 주의 깊게 읽지 않는다»는 말을 어떻게 탐구의 출발점으로 삼을 수 있는가?

판단을 다시 살펴보는 일이 도움이 되는 사례를 보았다. 왜 도움이 되는지 설명하려면 결과에 붙인 이름과 실제로 작동하는 메커니즘을 구별해야 한다. 다음 장에서 이 문제를 살핀다.

제2부

# 과학적 탐구로 넘어가기

제5장

# 무엇을 설명으로 인정하는가

앞 장에서는 반례를 찾는 등 풀이 방식을 바꿔 성적을 높일 수 있었다. 그러나 결과가 좋아져도 원인은 밝혀지지 않는다. 새 방법이 도움이 된 것일까, 시간이 더 주어졌거나 과제가 익숙해진 덕일까? 이제 다른 설명과 비교해 검증할 수 있는 설명을 세우는 법을 익힌다. 이는 뒤의 실험과 의식 모형을 준비하는 일이기도 하다.

## 이름에서 연구 질문으로

어떤 사람이 문장 하나를 몇 번 다시 읽은 끝에 이해했다. 왜 그렇게 되었을까? «이해가 작동했다»는 답은 결과에 붙인 이름을 되풀이할 뿐이다. «뇌가 정보를 처리했다»는 답은 일반적인 과정을 덧붙이지만, 이해하지 못했을 때와 이해했을 때 사이에 무엇이 달라졌는지는 말하지 않는다. 설명을 시작하려면 어떤 관계를 파악했는지, 앞서 어떤 정보가 있었는지, 어떤 처리가 새 결과에 이르게 했을지 구체적으로 묻는다.

기술은 일어난 일을 말한다. 예를 들면 «두 번째로 읽을 때 참가자는 사건 세 개를 시간순으로 배열했다»는 말이다. 설명은 그런 결과가 어떻게 나왔는지 밝히려 한다. 예측은 특정 조건에서 무엇을 기대할 수 있는지 말한다. 셋은 관련이 있지만 서로 다른 과제다. 장치 내부를 거의 몰라도 과거 사례의 표를 바탕으로 출력을 정확히 예측할 수 있다. 반대로 그럴듯한 메커니즘을 제안했어도 실제 자료를 아직 잘 예측하지 못할 수 있다.

체험을 연구할 때는 무엇을 설명하려는지도 밝혀야 한다. 정답을 냈는가, 선명했다고 보고했는가, 정보를 새 과제에 쓸 수 있었는가, 이해하는 느낌이 들었는가? 한 실험에서 이

자료들을 모두 얻을 수는 있지만 한 자료에서 다른 자료에 관한 결론으로 넘어가려면 논거가 필요하다. 그렇지 않으면 이론이 본래의 현상 대신 다루기 쉬운 지표를 설명하게 된다.

## 무엇을 메커니즘이라고 하는가

자동 램프를 떠올려 보자. 센서는 조도를 측정하고, 프로그램은 그 수치를 문턱값과 비교하며, 스위치는 전원을 공급한다. 어두워지면 램프가 켜진다. 여기에는 부품과 부품 사이의 작용, 변화가 일어나는 순서가 드러난다. 센서를 가리면 결과가 예상할 수 있는 방식으로 달라진다. 개입으로 검증할 수 있는 단순한 메커니즘 설명이다.

Peter Machamer, Lindley Darden, Carl Craver는 변화를 일으키는 조직된 실체와 활동을 통해 메커니즘을 분석하자고 제안했다. 이 연구 방식 자체가 우리 작업에 유용하다. 구성요소의 목록만으로는 충분하지 않으며 조직과 작동을 보여 주어야 한다. 이들의 논문은 설명의 철학적 분석이지 실험적인 의식 이론이 아니다. [7]

이 접근을 읽기에 적용해 보자. 모형은 낱말의 표상, 앞 문장을 보존하는 저장 공간, 시간 관계를 비교하는 규칙으로 구성될 수 있다. «저장 공간»과 «규칙»은 아직 잠정적인 이름이다. 무엇을 얼마나 오래 보존하는지, 어떻게 사용하는지, 어느 조건에서 사라지는지를 밝혀야 한다. 그래야 모형이 왜 어떤 글은 이해하고 다른 글에서는 오류를 내는지 설명할 수 있다.

한 메커니즘을 여러 수준에서 기술할 수 있다. 전기 회로에서는 전류로 램프의 전환을 설명하고, 프로그램에서는 조건과 명령으로 설명한다. 두 기술이 모두 맞다면 같은 체계를 다루면서 서로 다른 질문에 답할 수 있다. 설명 수준이 여럿이라는 사실만으로 서로 독립적인 인과 세계가 여럿 존재해야 하는 것은 아니다.

## 상관관계와 개입

실습용 관찰에서 그림을 더 오래 본 사람들이 그것을 잘 이해했다고 더 자주 말했다고 하자. 몇 가지 설명이 가능하다. 시간이 더 주어져 이해하기 쉬웠을 수 있다. 복잡한 그림 때문에 더 오래 보면서 이해 여부에는 더 신중하게 답했을 수도 있다. 관심이 큰 참가자들이 더 오래 보고 더 선명했다고 보고했을 수도 있다. 두 지표 사이에 관계가 있다는 사실만으로 인과관계의 방향은 정해지지 않는다.

개입은 연구의 규칙에 따라 조건을 바꾸는 일이다. 예를 들어 화면에 보여 줄 시간을 미리 정하고 참가자의 관심과는 무관하게 배정할 수 있다. 나머지 중요한 차이를 고려한다면 이 비교로 시간이 미치는 영향을 더 잘 알 수 있다. 개입 자체도 살펴야 한다. 더 오래 보면 눈을 움직이는 횟수, 다시 읽을 기회, 피로가 함께 달라질 수 있다. 이런 변화는 인과 과정의 일부일 수도 있고 부수적인 변화일 수도 있어 따로 살펴야 한다.

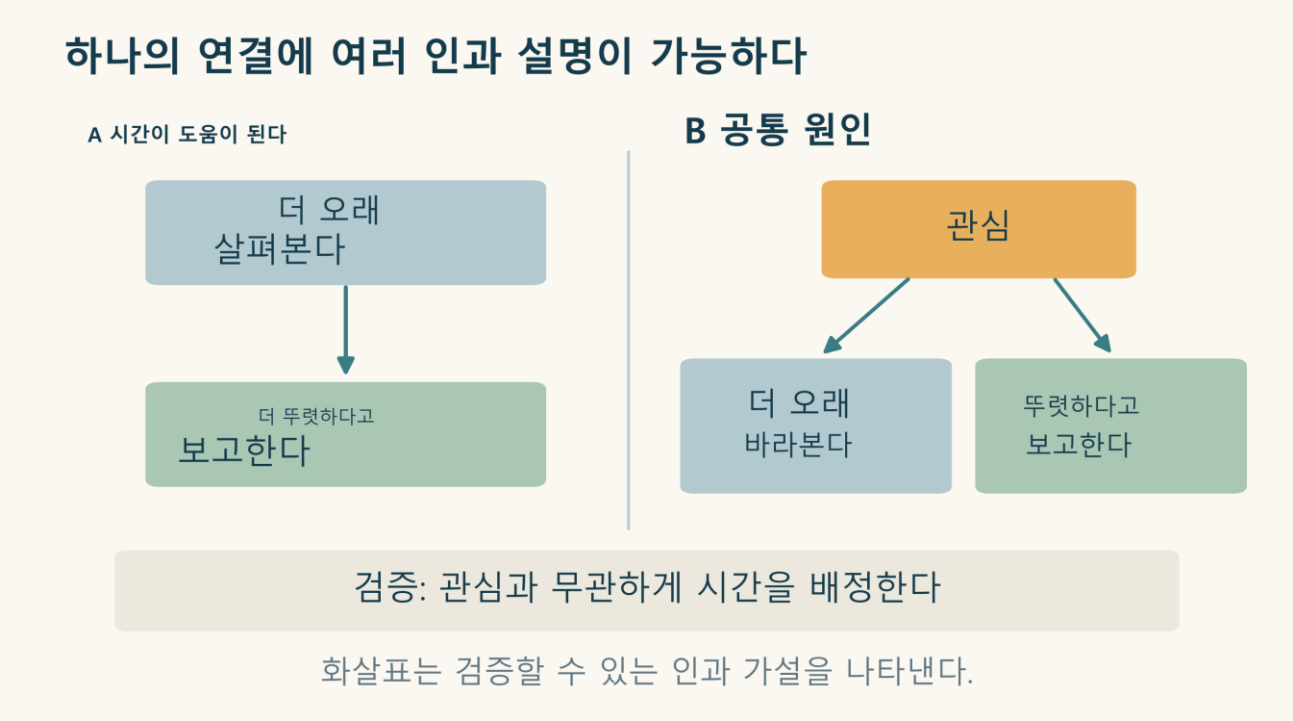


그림 10. 인과 영향과 공통 원인. 보는 시간과 선명함에 대한 보고의 연결에는 여러 메커니즘이 가능하다. 시간이 영향을 줄 수도 있고, 관심이 두 지표에 함께 영향을 줄 수도 있다. 다른 중요한 조건들을 고려했다면 연구 규칙에 따라 시간을 배정하여 두 설명을 구별하는 데 도움을 받을 수 있다.

관심 있는 사건의 원인과 그 원인을 나타내는 지표를 구별해야 한다. 온도계는 온도를 알려 주지만 그림 속 온도계의 숫자를 바꿔도 방은 따뜻해지지 않는다. 마찬가지로 보고와 연관된 신경 지표는 원인일 수도, 결과일 수도, 함께 일어나는 과정일 수도 있다. 어느 설명이 맞는지 고르려면 이 설명들이 서로 다르게 예측하는 조건이 필요하다.

## 필요성, 충분성과 주장의 범위

어떤 성질이 결과에 필요하다면, 명시한 조건 범위에서는 그 성질 없이 결과가 생기지 않는다. 그 성질이 충분하다면, 약속한 조건에서 그 성질이 있는 것만으로 결과가 생긴다. 두 관계를

서로 바꿔 쓸 수는 없다. 자동 램프가 작동하려면 정상적인 전원이 필요하다. 그러나 전원만으로 충분하지는 않다. 프로그램이 램프를 꺼 두었거나 스위치가 고장 났을 수 있다.

의식을 논할 때는 조건의 범위가 특히 중요하다. «과정 P는 체험에 필요하다»는 말은 어디에나 적용되는 듯 들린다. 성인에게 시각 과제 하나를 준 자료는 그보다 좁은 주장만 뒷받침한다. 다른 유형의 체험, 다른 생물, 인공 체계에까지 일반화하려면 추가 근거가 필요하다. 범위를 명시하면 좋은 이론이 약해지는 것이 아니라 어디까지 밝혔고 다음에 무엇을 연구해야 하는지가 드러난다.

무엇이 인과적으로 체험을 가능하게 하는지와, 무엇이 체험 자체와 동일하다고 가정하는지도 구별해야 한다. 이론은 과정의 특정한 조직이 곧 체험이라고 주장할 수 있다. 이 경우 한 현상을 기술하는 두 방식 사이의 관계를 제안하는 것이다. 두 현상이 함께 나타났다는 관찰만으로는 동일성을 입증할 수 없다. 두 기술이 어떻게 대응하는지, 어디까지 적용되며 다른 견해보다 어떤 이점이 있는지 설명해야 한다.

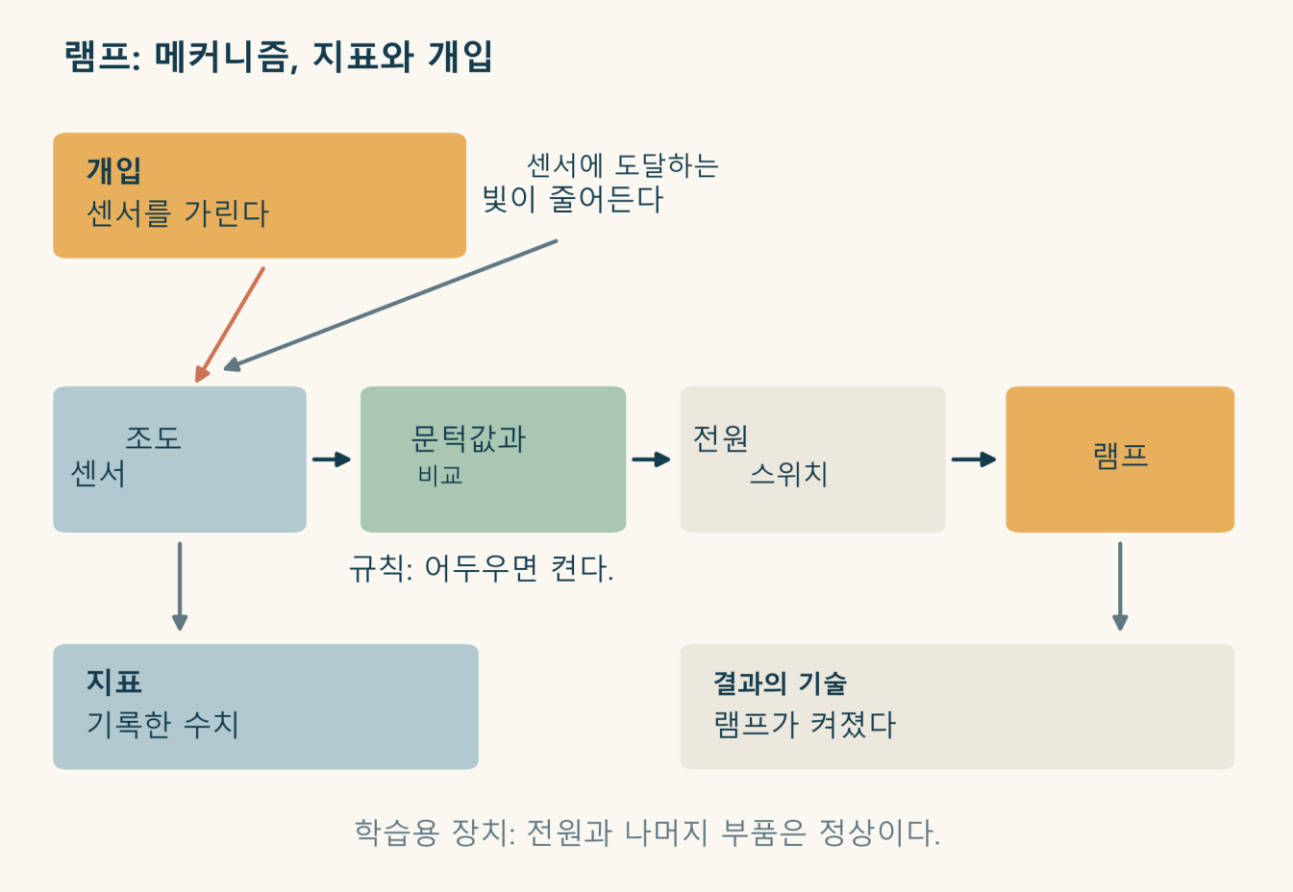


그림 11. 램프 구조의 세 연구 역할. 가운데 줄은 학습용 자동 램프에서 측정값, 명령, 전력이 전달되는 경로를 보여 준다. 센서를 가리는 것은 개입, 기록의 수치는 지표, 램프가 켜진 것은 기술한 결과다. 예측은 나머지 부품이 명시한 대로 정상 작동한다는 조건에 달려 있다.

## 반복 처리에 관한 논증을 살펴보는 예

다음은 가상의 논증이다. «다시 보면 사람들이 그림을 더 잘 알아본다. 따라서 반복 처리가 의식을 만든다. 프로그램에 되먹임을 더하면 체험이 생긴다.» 첫 단계는 과제 성적에 관한 경험적 주장이다. 두 번째는 그 성적을 의식과 연결한다. 세 번째는 가정한 메커니즘을 다른 체계에도 적용한다. 각 단계에는 처음 관찰에 담기지 않은 가정이 추가된다.

이 논증을 다시 세워 보자. 먼저 지표를 정한다. 그림을 올바르게 알아본 비율과 시각적으로 얼마나 선명했는지에 관한 평정을 따로 둔다. 다음은 메커니즘이다. 첫 처리는 특징을 추출하고, 뒤이은 처리에서는 주변 요소를 고려해 특징을 더 정확히 파악한다. 여기서 예측을 끌어낼 수 있다. 국소적 특징만으로는 판단이 어렵고 주변 맥락이 유용할 때 반복 처리가 특히 도움이 될 것이다.

다른 설명도 가능하다. 되먹임이 없어도 단지 계산 시간이 더 주어져 성적이 좋아졌을 수 있다. 비교하려면 두 모형의 자원이 비슷해야 하고 예측이 서로 어긋나는 과제가 필요하다. 되먹임이 있는 모형이 더 잘 수행한다면 선택한 결과를 설명하는 데 이 메커니즘을 선호할 근거가 생긴다. 그 결과가 실제 체험과 어떤 관계인지는 별도의 이론과 추가 검증에 달려 있다.

이 검토만으로 주체를 만들어 낼 수 있는지 결론짓지는 못한다. 다만 연구의 순서를 세울 수 있다. 과정의 어떤 조직이 무엇을 설명하는지 먼저 밝힌 뒤, 그 설명에서 체험의 성질로 넘어갈 근거를 세운다. 기술적으로 성공한 결과는 논증의 한 부분으로 쓰인다. 그것만으로 논증 전체를 대신할 수는 없다.

## 두 설명을 가려 주는 비교

일지가 왜 도움이 되는지 살펴보자. 첫째 설명은 보존된 글이 세부 내용을 간직해 서로 대조할 수 있게 한다는 것이다. 둘째 설명은 멈추어 자세히 살피는 시간 자체가 도움이 된다는 것이다. 일지를 쓴 경우와 전혀 멈추지 않은 경우를 비교하면 둘 다 설명이 될 수 있다. 더 유용한 비교에서는 같은 시간을 주되 한 조건에서는 기록을 남겨 나중에 쓸 수 있게 하고 다른 조건에서는 잠시 멈추기만 한다. 그러면 결과의 차이로 설명의 범위를 좁힐 수 있다. 이해를 나타낼 지표도 정해야 한다. 이야기가 길어진 것과 새로운 사례를 더 정확히 구별한 것은 같은 결과가 아니다.

### 여기까지 확인한 것

부품과 상호작용, 서로 다른 예측을 밝히면 설명이 더 구체적이 된다. 그런 요건을 갖추었다고 처음부터 그 설명이 참인 것은 아니다. 다만 어떤 결과가 나와야 그 설명에 대한 판단을 바꿀지 알 수 있다.

## 실습 5 빠진 논리 단계 되살리기

실습 목표는 일지가 도움이 된다는 일반적인 주장을 검증할 수 있는 두 메커니즘의 설명으로 구체화하는 것이다.

사건, 지표, 사전 예측의 개념을 활용한다. 무엇을 측정하는지와 그 결과를 설명하기 위해 어떤 메커니즘을 제안하는지를 구별한다.

예상 소요 시간은 25–35분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 결과 하나를 정한다. 예를 들어 미리 기술한 두 유형의 사례를 새 예에서 올바르게 구별하는 능력을 택할 수 있다.
2. 두 설명을 적는다. 다시 볼 수 있는 기록을 보존하는 효과와 잠시 더 멈추는 효과다. 각 설명에서 입력과 작용, 예상되는 변화를 밝힌다.
3. 같은 시간과 자료를 사용하는 비교를 설계한다. 두 조건이 정확히 어떻게 다른지 적는다.
4. 기록을 남긴 조건에서 결과가 더 나아지는 경우, 같은 경우, 더 나빠지는 경우를 미리 적는다. 각 결과에서 어떤 결론을 내릴 수 있고 어떤 대안이 남는지 기록한다.

지표 하나, 두 메커니즘, 비교하는 조건, 결과를 해석하는 규칙을 한 쪽에 정리한다.

어떤 결과가 나와도 자신이 선호하는 설명을 뒷받침한다면 그 메커니즘이 무엇을 예측하는지 더 구체적으로 정한다. 측정이 부정확해 차이를 찾지 못했다는 사실로 두 결과가 같다고 결론 내릴 수는 없다.

탐구를 이어 가려면 기록을 남긴 뒤 그 글을 다시 볼 수 있는지에 개입한다. 이 개입으로 어떤 새 결론을 내릴 수 있고 어떤 부수적 변화까지 고려해야 하는지 설명한다.

해설과 답은 부록 F 제5장 참조, 298쪽

**자기 점검 문항**

1. 메커니즘을 기술하는 일은 체계의 부품을 나열하는 일과 어떻게 다른가?
2. 신경 지표와 보고가 함께 달라졌다는 사실만으로 인과관계의 방향이 정해지지 않는 까닭은 무엇인가?
3. 인식 프로그램이 과제를 잘 수행했다는 사실에서 체험을 만들어 냈다는 주장으로 넘어가려면 어떤 근거가 더 필요한가?

다음 장에서는 연구하려는 현상을 어떻게 자료로 만드는지 살펴본다. 버튼을 누른 반응, 말로 표시하는 척도, 생리 신호가 각각 연구 질문과 어떻게 연결되는지 설명해야 한다.

제6장

# 주관적 체험을 연구하는 방법

지금까지 메커니즘과 예측을 마련했다. 이제 어떤 관찰로 이를 검증할지 정해야 한다. 체험이 그대로 표에 옮겨지는 것은 아니다. 응답 기록은 일정한 조건에서 만들어진다. 참가자의 보고, 행동의 성공 여부, 생리 지표가 각각 무엇을 보여 주는지 살펴보자. 이 장에서는 질문에서 자료로, 자료에서 타당한 결론으로 이어지는 과정을 짚는 법을 익힌다.

## 체험은 곧바로 표가 되지 않는다

그림을 보다가 익숙한 모양을 알아차리고 버튼을 누른다. 연구자는 그림, 버튼을 누른 시점, 응답, 생리 신호의 기록을 저장할 수 있다. 하지만 체험 자체가 다른 자료 옆에 열어 볼 수 있는 파일 하나로 남는 것은 아니다. 연구 질문과 자료를 연결할 근거가 필요하다. 왜 바로 이 응답이나 지표가 우리가 연구하는 것을 알려 주는가?

처음에는 세 가지 자료를 구별하는 편이 좋다. 첫째는 체험에 관한 보고다. 말로 한 기술, 범주 선택, 선명도 평정이 이에 해당한다. 둘째는 과제에서 보인 행동이다. 올바르게 구별했는지, 답하는 데 얼마나 걸렸는지, 배운 내용을 새 사례에 적용했는지를 본다. 셋째는 생리 측정이다. 각 자료가 처음의 질문과 어떤 관계를 맺는다고 보는지 연구자가 명확히 할 때 세 가지가 서로를 보완할 수 있다.

이 일은 가장 객관적인 기기를 찾는 것으로 끝나지 않는다. 수치에도 해석이 필요하다. 참가자가 «선명하다»는 버튼을 누른 경우 직접 기록된 것은 지시에 따라 버튼을 눌렀다는 사실이다. 체험이 선명했다는 결론은 참가자가 척도를 어떻게 이해했는지, 어떤 조건에서 보고했는지, 연구자가 어떤 보고 모형을 쓰는지에 달려 있다. 이 점을 밝히면 자료로 무엇을 더 검증해야 하는지 알 수 있다.



그림 12. 연구의 참가자와 관찰자. 화면을 보고 버튼으로 답하는 가상의 과제를 그렸다. 연구자는 행동을 기록하고 참가자의 기술을 얻는다. 이 자료가 체험과 어떻게 관련되는지는 연구 절차에서 별도로 뒷받침해야 한다.

## 구체적인 사례를 기술하게 하기

«평소에 음악을 어떻게 느낍니까»라는 질문과 «마지막 소리가 났을 때 무엇을 알아차렸습니까»라는 질문을 비교해 보자. 첫째는 평소 습관을 이야기하고 여러 경험을 일반화하게 한다. 둘째는 특정 순간에 주의를 돌리게 한다. 체험의 순서가 필요하다면 사건과 시점을 밝혀 기술하게 하고, 직접 알아차린 내용과 그 원인에 대한 추정을 구별하는 편이 좋다.

클레르 프티망쟁은 구체적인 주관적 체험을 자세히 탐구하기 위한 면담 방법을 기술했다. 이 논문은 알아차림과 기술의 어려움, 이를 다루는 방법, 얻어진 기술의 타당성을 판단하는 기준을 살펴본다. 이는 자료 수집에 관한 방법론적 제안이며, 모든 상세한 이야기에 오류가 없다고 보장하는 것은 아니다. [8]

직접 실습할 때는 예상하지 못한 답을 할 수 있도록 질문한다. 보통은 «버튼을 누르기 직전에 무엇이 달라졌습니까»가 «확신이 들었지요»보다 유용하다. 두 번째 질문은 어떤 체험을

기대하는지 미리 알린다. 그렇지 않았다고 답해도 이미 제시된 범주에 맞추어 답하게 된다. 자유롭게 이야기하게 한 다음에는 참가자가 처음 쓴 표현을 남겨 둔 채 구체적인 차이를 확인할 수 있다.

«기억나지 않는다»는 답도 그 절차에서 얻은 결과다. 이를 곧바로 «아무것도 체험하지 않았다»로 바꾸면 안 된다. 사건이 기억에 남지 않았거나, 질문이 명확하지 않았거나, 필요한 차이를 알아차리지 못했을 수 있다. 다음 연구에서는 보고까지 걸리는 시간을 줄이거나 질문을 바꿀 수 있다. 이미 빠진 답을 편한 가정으로 나중에 채울 수는 없다.

## 행동으로 검증하는 것

두 선 가운데 어느 쪽이 더 긴지 가리는 과제를 생각해 보자. 바르게 골랐다면 그 구별에 필요한 정보를 성공적으로 사용했다는 뜻이다. 여기에 «차이가 아주 선명했다»는 말이 더해지면 별도의 보고를 얻는다. 정답 여부와 선명도가 어긋날 때는 앞뒤가 맞는 이야기를 만들려고 지표 하나를 버리지 않는다. 어떤 조건에서 어긋나는지 연구해야 한다.

응답 시간에도 여러 설명이 가능하다. 구별하기 어렵거나, 신중하게 확인하거나, 버튼을 누르기 불편해서 늦게 답했을 수 있다. 빨리 답한 것도 문제가 쉬워서일 수도, 성급하게 골라서일 수도 있다. 그래서 시간은 과제의 구성과 다른 자료를 함께 놓고 해석한다. «더 빠르면 더 의식적인가»라는 질문에 답하려면 그런 관계가 왜 나타날 것으로 예측했는지 설명할 이론이 필요하다.

행동 과제는 참가자가 어떤 내용을 이용할 수 있는지 확인할 때 특히 유용하다. 정보를 새 결론이나 지시, 선택에 사용할 수 있다면 기능적으로 접근할 수 있음을 보여 주는 자료가 된다. 하지만 과제마다 접근 가능성은 다르다. 물건이 어디에 있는지는 맞히고도 그것을 말로 설명하기 어려울 수 있다. 한 가지 응답 방식에서 실패했다고 해서 정보를 쓸 방법이 전혀 없다고 결론 내릴 수는 없다.

## 생리 지표를 읽는 방법

뇌파검사는 전극으로 전기적 전위차를 기록한다. 일반적인 기능적 자기공명영상은 혈류와 혈액의 산소화 정도에 관련된 신호를 사용한다. 두 방법은 서로 다른 것을 측정하고 시간과

공간에 관한 특성도 서로 다르다. 뇌 영상의 색칠된 부분은 측정값을 처리한 결과다. 개별 체험을 찍은 사진이 아니다.

Nikos Logothetis와 동료들은 원숭이의 시각피질에서 기능적 자기공명 기록과 전기적 신경 신호를 비교했다. 이 연구는 측정 신호와 신경 활동의 연관을 살펴보았다. 우리 주제에서 중요한 것은 그 연구 단계 자체다. 기기 지표와 그 지표로 평가하려는 과정 사이의 관계는 독립적인 연구 질문이다. [9]

«보인다»는 응답에 앞서 어떤 지표가 커졌다고 하자. 그림을 처리하는 과정, 보고를 준비하는 과정, 과제에 주의를 기울이는 과정, 또는 이 과정들의 조합을 나타낼 수 있다. 그 역할을 밝히려면 응답 방식과 질문 시점, 결과를 보고하도록 요구하는지 여부를 바꾸어 볼 수 있다. 다른 중요한 조건을 그대로 두면 이런 비교가 유용하다. 즉시 보고하게 하지 않으면 보고에 딸린 행동 일부가 사라진다. 그렇다고 체험을 직접 잴 수 있게 되는 것은 아니다.

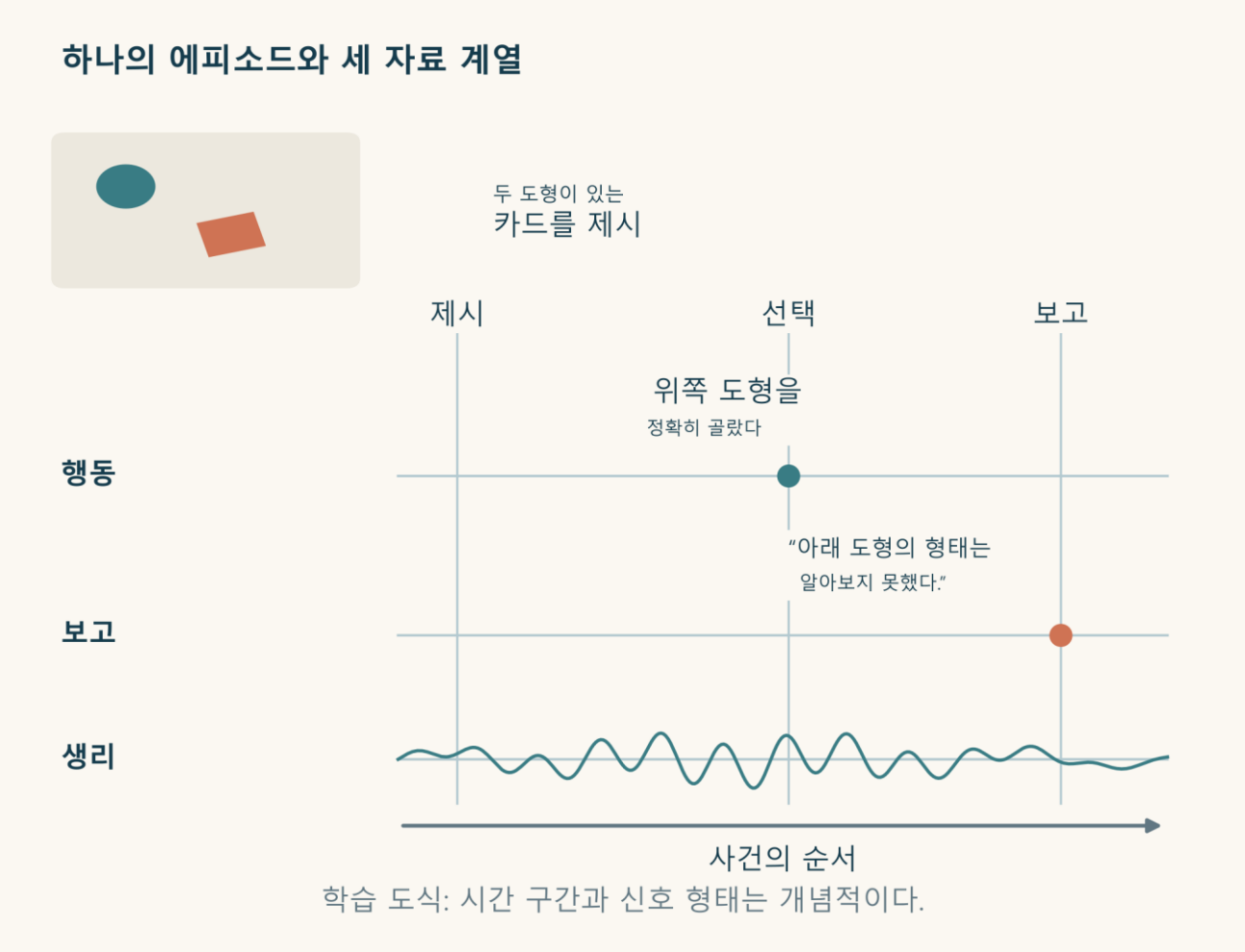


그림 13. 서로 다른 자료의 시간 정렬. 이 장의 가상 에피소드에서는 도형의 위치를 정확히 구별하면서도 형태가 불분명했다고 보고한다. 버튼 누름, 보고, 생리 기록을 공통 시간축에 나타냈다. 곡선의 형태와 시간 구간은 개념적이며 실험 자료가 아니다.

## 자료가 어긋나는 사례의 분석

실습을 위해 꾸민 상황을 보자. 카드 한 장에 도형 두 개가 있다. 참가자는 어느 도형이 위에 있는지 맞히지만 «아래 도형의 모양은 자세히 보지 못했다»라고 말한다. 두 답이 반드시 모순되는 것은 아니다. 위치와 모양은 다른 성질이다. 첫 질문에 답했다고 해서 두 번째 성질을 선명하게 지각했음이 입증되지는 않는다. 연구자가 모든 정답을 «그림 전체를 의식했다»라는 한 가지 이름으로 묶는다면 오류가 생긴다.

바르게 고른 사례에서 생리 지표가 더 높았다고 하자. 적어도 두 가지 설명이 있다. 지표가 공간 정보를 얼마나 잘 처리했는지 나타내거나, 답을 준비할 때의 확신을 나타낼 수 있다. 이를

구별하려면 정답률은 같고 확신도는 다른 사례를 고르고 보고 방식도 별도로 바꿀 수 있다. 다음 연구는 예측이 구체적으로 어디에서 어긋나는지를 토대로 설계한다.

여러 방법을 함께 쓰는 것을 삼각검증이라 부르기도 한다. 핵심은 여러 방법을 통해 가능한 다른 설명을 좁히는 데 있다. 세 기기의 결론을 다수결로 고르는 것이 아니다. 보고와 버튼 누르기, 신경 지표가 모두 같은 응답 준비 과정에 달려 있다면, 일치한다고 해도 체험을 독립적으로 뒷받침하는 증거 세 가지가 되지는 않는다. 여러 지표의 원인 중 무엇이 공통되고 무엇이 다른지 살펴야 한다.

## 정답을 알게 되어도 처음 기록은 고치지 않는다

펜을 잠깐 본 사람이 색을 보고하지만 적힌 글은 언급하지 않았다. 다시 보니 글이 실제로 있었다. 이 확인은 물건의 성질을 알려 주지만 어느 설명이 맞는지는 정하지 못한다. 글을 못 보았거나, 보았지만 잊었거나, 짧게 답하는 과정에서 빼놓았을 수 있다. 이를 가리려면 보고까지의 시간을 바꾸는 등 다른 비교가 필요하다. 보는 데 앞서 글에 관한 질문을 알면 무엇을 찾아볼지 달라진다. 이 경우는 자유롭게 관찰한 상황과 같게 볼 수 없다.

### 여기까지 확인한 것

어떤 자료는 물건의 성질에 관해 무엇을 알 수 있는지 알려 준다. 다른 자료는 그 정보를 어떻게 사용하고 보고했는지를 보여 준다. 결과가 어긋나면 질문을 더 정확히 세울 수 있다. 이제 다른 사람도 뜻을 파악할 수 있게 절차를 기술해 보자.

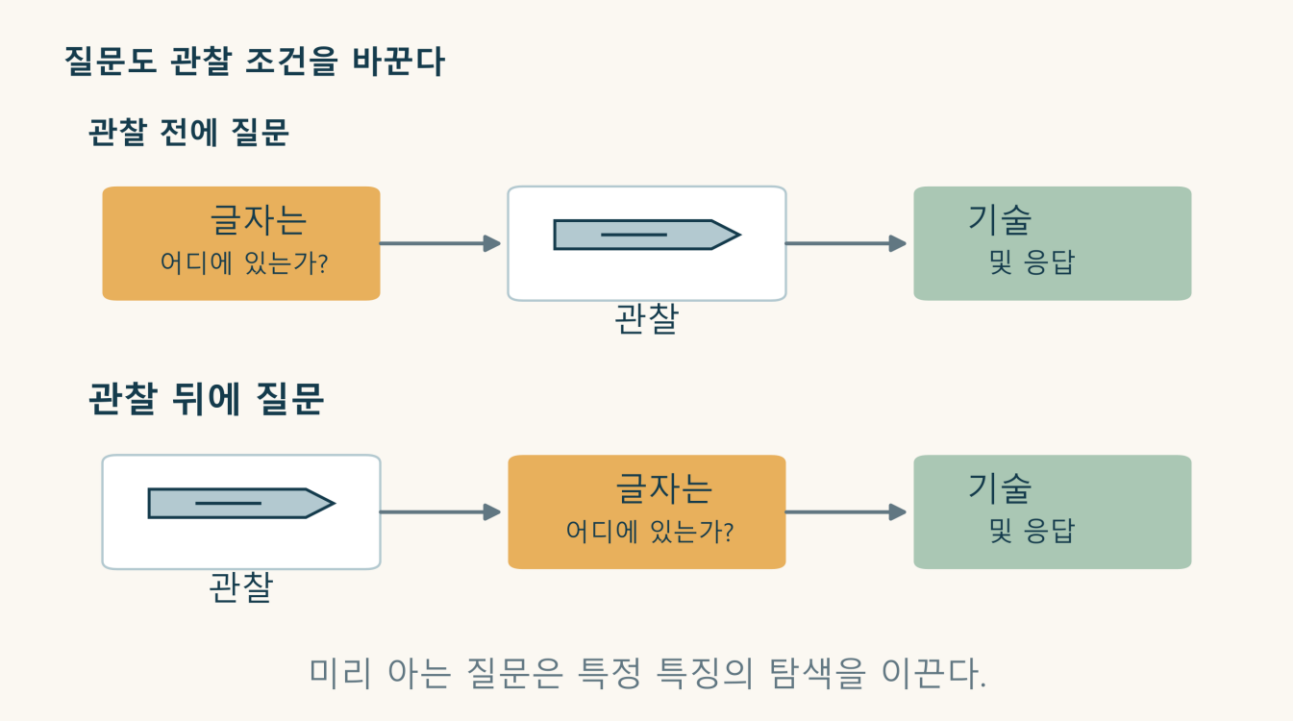


그림 14. 절차의 조건이 되는 질문 시점. 위 절차에서는 펜에 적힌 글에 관한 질문을 보기 전에 알려 주고, 아래 절차에서는 물건을 본 뒤에 알려 준다. 질문 시점이 다르면 무엇에 주의를 기울이는지도 달라질 수 있다. 이를 고려하지 않고 물건의 성질 차이만으로 결과를 설명할 수는 없다.

## 실습 6 한 가지 측정의 기록표 만들기

실습 목표는 측정 조건과 그 자료에서 어디까지 결론을 내릴 수 있는지 드러나는 기록표를 만드는 것이다.

제5장에서 지표와 설명하려는 현상을 구별한 내용을 활용한다. 질문을 제시한 시간과 응답한 시간을 모두 고려한다.

예상 소요 시간은 20–25분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 색이 있고 눈에 띄는 세부 특징이 있는 물건을 고른다. 뒤에 나올 질문을 이미 알고 있다고 미리 기록한다. 이 실습은 절차를 보여 주기 위한 것이다.
2. 십오 초쯤 본 뒤 물건을 치우고 그때 체험한 일을 적는다. 그런 다음 색과 세부 특징의 위치에 관해 따로 답한다.
3. 물건이 실제로 어떻게 되어 있는지 확인한다. 처음 기록을 보존하고 외부 사실과 보고한 내용을 구별한다.

4. 기록표의 다섯 항목을 채운다. 제시 방식, 지시, 질문 시점, 보고까지의 시간, 평가 규칙이다. 이 절차로 내릴 수 있는 결론과 아직 내릴 수 없는 결론을 하나씩 덧붙인다.

체험을 적은 기록, 확인할 수 있는 답, 개선 제안 하나가 담긴 절차 기록표를 남겨 둔다.

«기억나지 않는다»는 답은 체험이 없었다는 보고와 다르다. 과정 가운데 어느 단계가 아직 알려지지 않았는지 분명히 한다.

탐구를 이어 가려면 새 물건을 사용해 보기 전 질문과 보고 난 뒤 질문을 비교한다. 어떤 차이를 예상하는지, 또 왜 이런 비교 자체가 관찰의 초점을 바꾸는지를 먼저 적어 둔다.

해설과 답은 부록 F 제6장 참조, 298쪽

**자기 점검 문항**

1. «기억나지 않는다»는 답을 체험이 없었다는 뜻으로 받아들일 수 없는 까닭은 무엇인가?
2. 두 대상을 올바르게 구별한 행동이 직접 보여 주는 것은 무엇인가?
3. 보고와 행동, 생리 지표가 일치해도 겉보기보다 근거가 약할 수 있는 까닭은 무엇인가?

기록표는 절차 하나가 어떻게 이루어지는지 보여 준다. 이제 결과에 영향을 줄 수 있는 원인들이 한꺼번에 바뀌지 않도록 절차를 비교해야 한다. 이것이 제7장의 과제다.

제7장

# 검증의 구조

제6장에서는 지표 하나를 어떻게 얻는지 살펴보았다. 이제 두 절차를 비교하고 결과 차이를 선택한 원인에 돌릴 수 있는지 묻는다. 작업 방식과 함께 자료의 난이도, 수행 순서까지 바뀌면 같은 결과에도 여러 설명이 가능하다. 이 장의 과제는 비교 자체를 설계하는 일이다. 자료를 모으기 전에 비교를 준비하고, 결과를 본 뒤에 떠오르는 생각도 따로 남긴다.

## 가설을 반박할 수 있는 결과 찾기

어떤 결과가 나오면 들어맞지 않는지 밝혀야 가설을 검증에 쓸 수 있다. «체험은 정보 처리와 어떻게든 관련된다»는 주장은 한 번의 검증으로 다루기에는 지나치게 넓다. «총 준비 시간이 같을 때, 시간 관계를 그려 보존한 도식이 글에 관한 새 질문에 답하는 데 도움이 된다»고 하면 조건, 결과, 비교 방법을 정할 수 있다. 결과가 좋아지거나, 뚜렷한 차이가 없거나, 더 나빠질 수 있다.

검증은 자료를 모으기 전에 시작된다. 연구자는 무엇을 바꾸고 무엇을 측정할지, 결과를 설명할 수 있는 다른 변화에는 무엇이 있는지 정한다. 바꾸는 조건을 독립변수, 측정하는 결과를 종속변수라고 한다. 예를 들어 도식의 유무는 조건이고 정답률은 결과다. 이런 명칭은 연구에서 맡는 역할을 가리킨다. 변수들이 실제로 독립적이거나 관계가 인과적임을 입증하지는 않는다.

조작적 정의는 일반적인 개념을 구체적인 절차로 옮기는 일이다. «글의 이해»는 읽은 내용에서 새로운 관계를 끌어내야 답할 수 있는 질문으로 평가할 수 있다. 그러나 그 절차는 이해의 특정한 측면만 다룬다. 도식에서 답을 찾아도 된다면 외부 자료를 이용한 과제 수행을 측정한다. 도식을 치우면 기억에 정보를 보존해야 하는 다른 상황을 연구하는 셈이다.

## 잘못된 비교가 설득력 있어 보이는 이유

실습용 프로젝트를 떠올려 보자. 참가자는 먼저 어려운 글을 메모 없이 읽고 질문에 답한다. 그런 다음 더 쉬운 글을 읽으며 도식을 그리고 더 잘 답한다. 연구자는 도식이 이해를 도왔다고

결론지었다. 실제로는 전략, 자료의 난이도, 수행 순서가 함께 달라졌다. 질문 형식에 익숙해진 덕에 두 번째 결과가 좋아졌을 수도 있다.

함께 달라진 요인이 관심 있는 원인을 따로 알아내기 어렵게 만들면 교란 요인이라고 한다. 보고서 끝에 방해가 되었을 법한 요인을 나열하는 것으로는 부족하다. 가능하다면 검증을 설계할 때 그 요인과 주요 조건의 결합을 바꿔야 한다. 쉬운 글에 늘 도식을 쓰면 결과에서 도식의 영향과 쉬운 글의 영향을 구별할 수 없다.

모든 차이를 물리적으로 없앨 필요는 없다. 어떤 차이는 미리 정한 규칙에 따라 여러 조건에 배분할 수 있다. 다른 차이는 측정해 분석에 반영한다. 하지만 필요한 비교가 빠졌다면 계산으로 보완할 수 없다. 어려운 글에 도식을 사용한 사례가 하나도 없으면 모형은 어려운 자료에서 도식이 어떻게 작용하는지에 대해 추가 가정을 해야 한다.

## 설계를 단계적으로 개선하기

짧은 글을 모은 두 묶음 A와 B로 시작한다. 글의 길이, 형식, 답에 필요한 추론을 비슷하게 맞춘다. 본 검증 전에 준비해 예비 시행에서 지시가 명확한지 확인하는 편이 좋다. 예비 시행은 절차를 고치기 위한 것이다. 이 결과를 고친 계획으로 수행한 본 검증 자료와 섞어서는 안 된다.

일부 참가자는 묶음 A에서 도식을 쓰고 B에서는 쓰지 않는다. 나머지 참가자에게는 반대로 배정한다. 같은 조합에서도 조건을 적용하는 순서를 바꾼다. 순서와 자료를 이렇게 균형 있게 배치해도 글이 완전히 똑같아지는 것은 아니다. 다만 한 묶음의 특성이 한 전략에만 결부되지 않도록 고르게 배분할 수 있다.

무작위 배정은 참가자의 특성을 보고 어떤 조건을 줄지 고르는 일을 막아 준다. 이를테면 숙련된 독자들에게 모두 도식을 준 뒤 그들의 좋은 성적을 도식 덕분으로 돌려서는 안 된다. 집단이 작으면 무작위로 나누어도 참가자 구성이 정확히 같지는 않을 수 있다. 그래도 배정 절차는 명확해진다. 자료 설계에도 세심한 주의를 기울이면 더 유용하다.

자기 혼자 연구할 때에는 먼저 일정을 짜고 그에 따라 조건과 자료를 바꿀 수 있다. 그 결과가 직접 적용되는 범위는 자신의 측정 회차다. 같은 글을 다시 읽으면 이전에 얻은 지식이 다음

과제에 영향을 주므로 비슷한 새 자료를 쓰는 편이 좋다. 혼자 하는 연구는 절차를 익히고 문제를 찾는 데 편리하다. 결과를 어디까지 일반화할 수 있는지도 분명히 밝힌다.

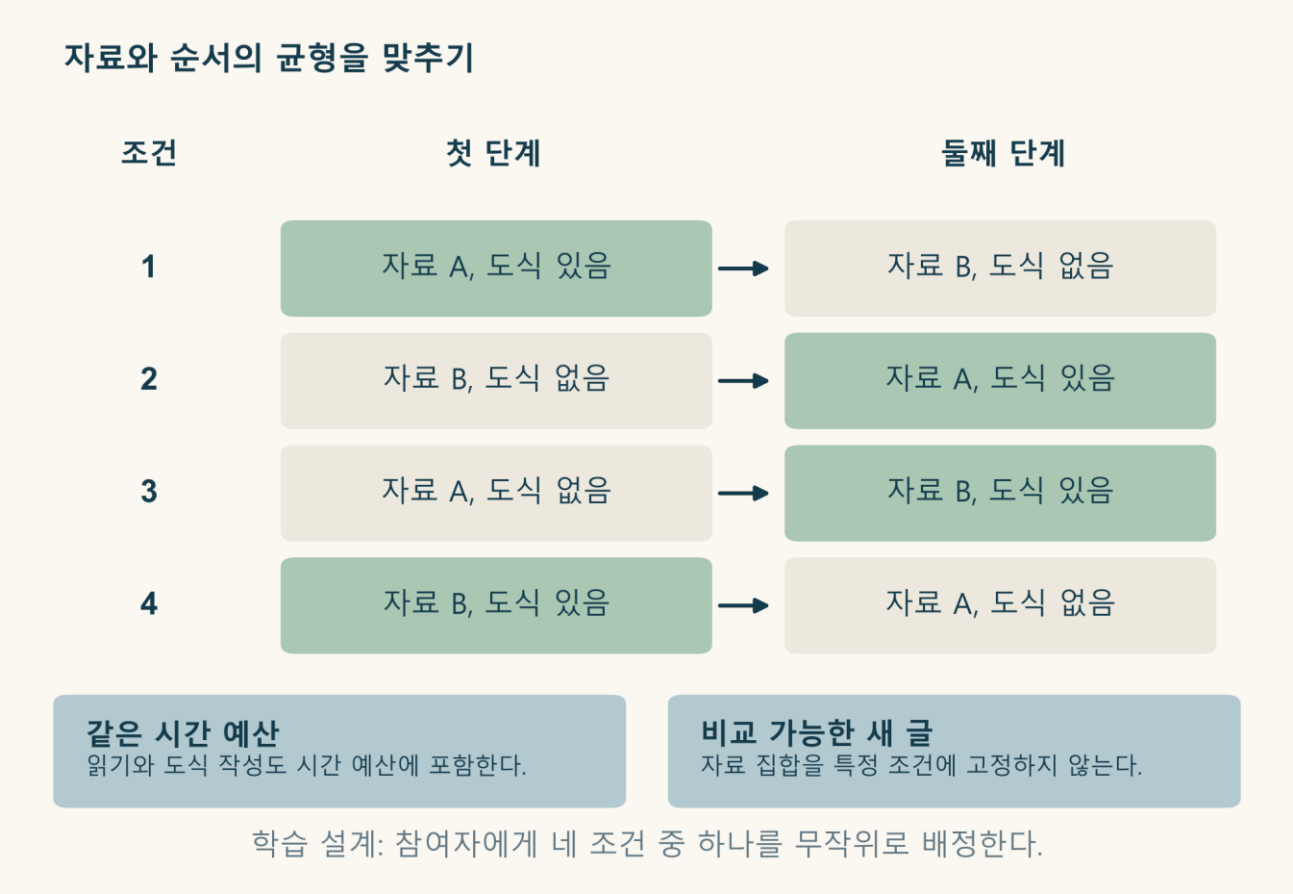


그림 15. 자료와 조건 순서의 배정. 자료와 전략의 두 조합에 두 가지 수행 순서를 적용하면 네 가지 배정이 나온다. 각 자료 묶음은 도식을 쓰는 경우와 쓰지 않는 경우, 먼저 하는 경우와 나중에 하는 경우에 모두 들어간다. 이 균형 배정에 무작위 배정과 동일한 총 준비 시간을 더한다.

## 결과를 일관되게 계산하는 법

사건의 시간순서를 묻는 문항의 정답률을 주 지표로 삼았다고 하자. 문항마다 정답 기준을 미리 정한다. 자유롭게 답한다면 어떤 표현까지 정답으로 인정할지 정해야 한다. 그렇지 않으면 결과를 본 뒤 기대와 다른 사례는 엄격하게, 기대한 사례는 관대하게 평가하기 쉽다. 가능하다면 채점자는 답이 어느 조건에서 나왔는지 모르는 편이 좋다.

총 준비 시간은 같게 둔다. 제한된 시간에 도식이 도움이 되는가라는 질문에 답하기 위해서다. 글을 읽는 데 쓰는 시간까지 같다는 뜻은 아니다. 한쪽에서는 준비 시간 일부를 도식을 적는 데 쓴다. 이제 결과가 좋아지지 않아도 의미 있게 해석할 수 있다. 주어진 시간에는 기록의 이점이

그에 드는 시간보다 작았을 수 있다. 다음 가설은 준비에 쓸 수 있는 시간에 따라 효과가 어떻게 달라지는지 다룰 수 있다.

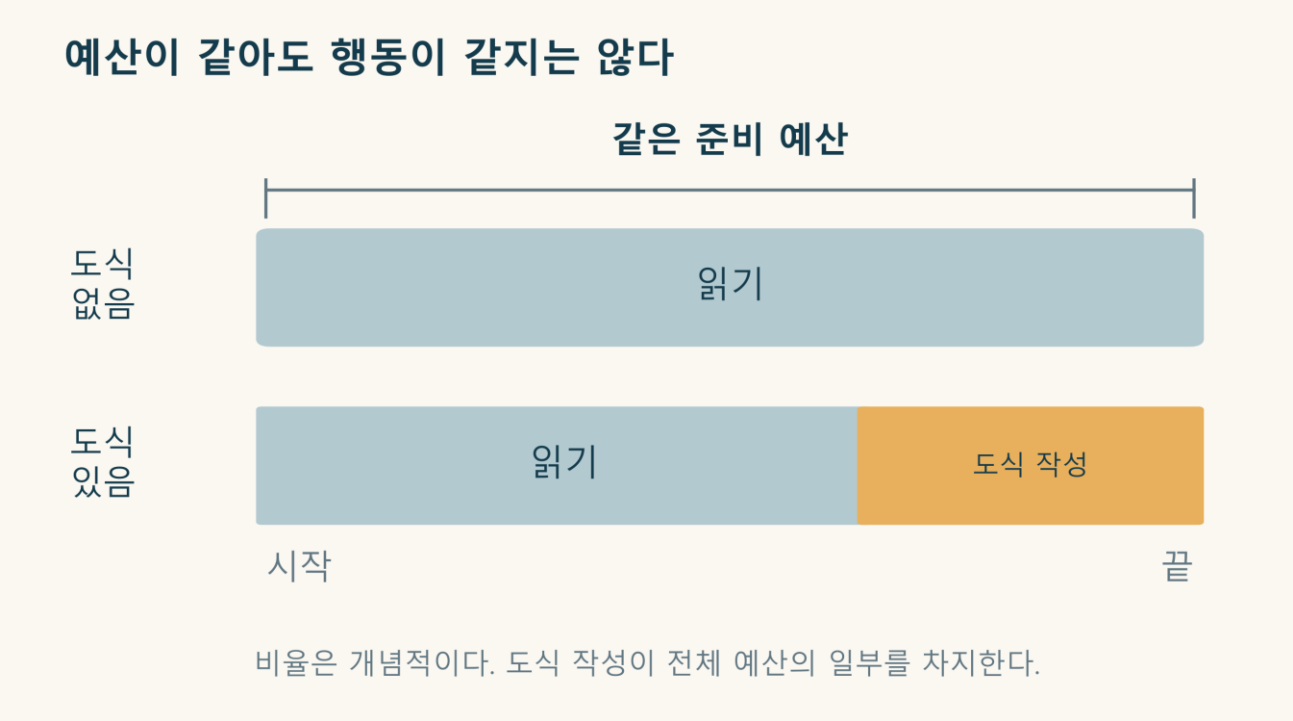


그림 16. 총 준비 시간과 그 배분. 같은 준비 시간을 주면 도식을 쓰는 조건에서는 그 일부를 도식 작성에 사용한다. 이 비교는 제한된 시간에 도식이 도움이 되는지 살핀다. 직접 글을 읽는 시간까지 같게 맞추지는 않는다. 그림의 시간 배분은 예시다.

정답 여부에 더해 확신도와 얼마나 선명하게 느꼈는지에 대한 기술도 모을 수 있다. 그러나 핵심 결과는 미리 정하는 편이 좋다. 열 가지 지표 중 어느 것이든 결과를 본 뒤 증거로 내세울 수 있다면 무엇을 검증했는지 모호해진다. 추가 지표로 새 질문을 찾을 수는 있다. 그때는 그 지표가 탐색을 위한 것임을 분명히 밝혀야 한다.

## 자료를 보기 전의 계획과 본 뒤의 탐색

Brian Nosek과 동료들은 사전등록을 질문과 분석 계획을 미리 고정하는 방법으로 설명한다. 그 방법론적 기능은 예측의 검증과 이미 본 결과의 설명을 구별할 수 있게 하는 데 있다. 이 논문은 등록이 약한 이론이나 좋지 않은 측정을 바로잡는다고 약속하지 않는다. 연구 과정의 구성을 다룬다. [10]

실습용 프로젝트라면 날짜를 적은 계획을 수정하지 않은 채 결과와 함께 보관하는 것으로 시작할 수 있다. 정식 사전등록을 대신하는 절차가 아니라 같은 원칙을 연습하는 과정이다. 계획에는 질문, 조건, 주 지표, 사용할 수 없는 기록을 제외할 규칙, 예정한 시행 회차 수, 비교 방법을 적는다. 바꾼 내용은 이유와 함께 따로 기록한다.

자료를 본 뒤의 탐색도 필요하다. 가령 도식이 사건이 네 가지인 글에서만 도움이 된다는 것을 알아차릴 수 있다. 이는 흥미로운 새 가설이다. 다음 단계에서는 새 글로 이를 검증한다. 처음부터 바로 그 결과를 예측했던 것처럼 발표해서는 안 된다. 과학 연구에는 규칙성을 발견하는 일과 그 뒤에 검증하는 일이 모두 들어 있다.

## 조건을 균형 있게 배치하면 달라지는 것

항상 두 번째의 더 쉬운 글에 도식을 쓴다고 하자. 수행 성적을 정확히 측정해도 도식의 이점과 글이 쉬웠던 효과, 연습 효과를 구별할 수 없다. 배정을 네 가지로 나누면 그 결합을 풀 수 있다. 첫째는 A를 도식과 함께, 이어서 B를 도식 없이 하는 경우다. 둘째는 같은 조합을 역순으로 한다. 셋째는 A를 도식 없이, 이어서 B를 도식과 함께 한다. 넷째는 그 조합을 역순으로 한다. 그러면 각 묶음과 순서에 두 작업 방식이 모두 들어간다. 집단이 작으면 무작위 배정으로 참가자들이 똑같아지지는 않는다. 다만 좋은 성적이 예상되는 참가자에게 연구자가 유리한 조건을 일관되게 배정하는 일은 막을 수 있다.

### 여기까지 확인한 것

비교 계획의 목표는 모든 차이를 없애는 데 있지 않다. 관심 있는 원인을 다른 원인과 구별하는 데 있다. 결과를 본 뒤 찾은 교란 요인은 해석의 한계를 알려 주지만, 빠져 있던 통제를 대신하지는 않는다.

## 실습 7 비교가 부족한 실험 고치기

실습 목표는 일지 기록이 길어진 것을 체험을 더 잘 구별하게 된 것으로 간주한 연구 설계를 고치는 것이다.

제6장의 측정 기록표와 메커니즘과 지표의 구별을 활용한다. 비교할 수 있는 과제 두 묶음과 미리 정한 결과 지표가 필요하다.

예상 소요 시간은 30–40분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 결과는 미리 정한 유형에 따라 새로운 사례를 얼마나 잘 구별하는지로 정한다. 기록의 길이는 추가 지표로 따로 남긴다.
2. 비교할 수 있는 두 묶음 A와 B를 마련한다. 작업 전에 총시간을 같게 정하고 채점 규칙을 정한다.
3. 앞의 예를 따라 자료와 순서를 배정하는 네 가지 방식을 적는다. 평소 글을 쓰는 데 익숙한 정도나 추가 시간의 효과를 고려할 조건도 밝힌다.
4. 성적이 좋아지는 경우, 같은 경우, 서로 다른 방향으로 바뀌는 경우에 어떤 결론을 내릴 수 있는지 미리 적는다. 지금은 설계만 해도 충분하다. 실습을 위해 참가자를 모집할 필요는 없다.

조건, 배정 방식, 주 지표, 평가 규칙, 세 가지 결과의 해석이 담긴 계획을 남겨 둔다.

총시간이 같을 때 도식이 도움이 되는지와, 도식에 별도 시간을 줄 때 도움이 되는지는 서로 다른 질문이다. 어느 쪽을 묻는지 밝힌다.

탐구를 이어 가려면 날짜를 적은 계획을 보관한다. 예비 시행 뒤에는 지시를 바꾼 기록과 본 검증의 결과를 따로 남기고, 새 자료로 확인할 가설을 세운다.

해설과 답은 부록 F 제7장 참조, 299쪽

**자기 점검 문항**

1. 준비 시간이 같아도 준비 과정이 같다고 할 수 없는 까닭은 무엇인가?
2. 조건을 무작위로 배정하면 어떤 이점이 있는가? 집단이 작을 때 무엇까지 보장하지는 않는가?
3. 분석 뒤에 생긴 새 가설은 미리 세워 검증하는 예측과 어떻게 다른가?

비교를 잘 설계해도 수치는 달라질 수 있다. 다음 장에서는 효과의 크기와 추정의 정밀도를 살펴보고, 둘을 인과적 결론과 구별한다.

제8장

# 불확실성을 다루는 방법

검증 설계에서 무엇을 무엇과 비교할지 이미 정했다. 그러나 같은 설계로 두 번 시행해도 성공률은 달라질 수 있다. 관찰된 차이가 지속적인 효과의 범위를 얼마나 좁혀 주는지 알아야 한다. 여기서는 세 질문을 구별한다. 이번 시행에서 무엇이 나왔는가? 추정은 얼마나 정밀한가? 설계는 어떤 원인을 논할 수 있게 하는가? 이후 컴퓨터 모형과 생물학적 표본을 비교할 때도 이 구분이 필요하다.

## 표에 나온 차이를 평가해야 한다

한 조건에서는 과제 20개 중 14개를, 다른 조건에서는 20개 중 16개를 맞혔다. 정답률은 각각 70퍼센트와 80퍼센트다. 차이는 10퍼센트포인트다. 이는 관찰된 효과를 기술한 것이다. 그 차이를 얼마나 정밀하게 추정했는지, 새 과제에서도 차이가 남을지, 바꾼 조건 때문에 생긴 차이인지는 아직 알 수 없다.

퍼센트포인트와 상대적 변화율은 구별하는 편이 좋다. 70퍼센트에서 80퍼센트로 오르면 10퍼센트포인트가 늘어난다. 출발점인 70퍼센트에 견주면 약 14퍼센트가 늘어난다. 두 표현 모두 수학적으로 옳을 수 있지만 서로 다른 질문에 답한다. 실습용 연구에서는 먼저 원래의 개수와 비율을 함께 알려 주어 변화가 무엇을 뜻하는지 확인할 수 있게 한다.

효과 크기, 불확실성, 인과적 해석은 서로 다른 과제다. 큰 차이도 추정이 부정확할 수 있고, 작은 차이도 아주 정밀하게 측정할 수 있다. 집단 사이의 차이가 정밀하게 측정되었어도 비교 설계에 체계적인 다른 설명이 남아 있으면 원인을 단정할 수 없다. 제7장에서는 비교의 질을 살폈다. 여기서는 결과가 얼마나 달라질 수 있는지를 이해한다.

## 우연에 따른 변동을 보여 주는 실습 모형

매 시행에서 독립적으로 70퍼센트의 확률로 정답을 내는 생성기를 생각해 보자. 규칙이 완전히 같은 두 조건 A와 B를 만든다. 각 조건에서 스무 번씩 시행하고, 두 조건을 한 쌍으로 묶어 여섯

차례 반복한다. 다음 표에는 이 책을 위해 난수 생성 프로그램으로 만든 합성 자료를 실었다. 사람을 대상으로 한 실험 결과는 아니다.

| 반복 시행 | A의 정답 수 | B의 정답 수 | B와 A의 정답률 차이(퍼센트포인트) |
|---|---|---|---|
| 첫 번째 | 18 | 17 | 마이너스 5 |
| 두 번째 | 14 | 14 | 0 |
| 세 번째 | 13 | 15 | 플러스 10 |
| 네 번째 | 19 | 15 | 마이너스 20 |
| 다섯 번째 | 11 | 13 | 플러스 10 |
| 여섯 번째 | 10 | 14 | 플러스 20 |

여기서는 자료 생성 규칙을 알고 있다. 조건 B가 더 유리할 이유는 없다. 그런데도 각각의 비교는 B가 더 낫거나 더 나쁜 것처럼 보이게 한다. 누가 잘못 계산했다고 가정하지 않고도 우연 때문에 추정에 오차가 생기는 것을 이해할 수 있다. 모든 수치를 정확히 적었어도 지속적인 차이에 대한 추정은 여전히 흔들릴 수 있다.

이 표의 변동 폭을 모든 연구에 그대로 적용할 수는 없다. 시행 횟수와 성공 확률, 시행들의 독립성에 관한 가정에 따라 달라진다. 실제 과제에서는 학습, 문제 묶음의 난이도, 피로 때문에 한 사람의 응답들이 함께 달라질 수 있다. 생성기는 우연성을 다루는 특정한 모형을 이해하는 도구다. 실제 자료에 어떤 모형을 쓸지는 별도로 근거를 마련해야 한다.

## 추정이 정밀하다는 것은 무엇인가

절차가 관심 있는 효과에 대해 그럴듯한 값의 범위를 더 좁혀 줄수록 추정은 정밀해진다. 독립적인 관찰을 늘리고 측정에서 불필요한 변동을 줄이는 일이 흔히 도움이 된다. 하지만 모든 반복에 같은 오류가 있다면 똑같은 일을 되풀이하는 것만으로는 부족하다. 잘못 설정된 기기로 천 번 측정해도 기기의 설정은 고쳐지지 않는다.

연구 보고서에는 신뢰구간이 나온다. 빈도주의에서는 신뢰수준 95퍼센트가 구간을 만드는 절차의 성질을 나타낸다. 채택한 모형의 조건에서 그 절차를 거듭 적용하면 구간 가운데 약 95퍼센트가 고정된 참값을 포함하게 된다. 이미 구한 특정 구간에 참값이 들어갈 확률이라는 뜻은 아니다. 그런 확률을 말하려면 다른 접근을 택하고 그 전제를 명시해야 한다.

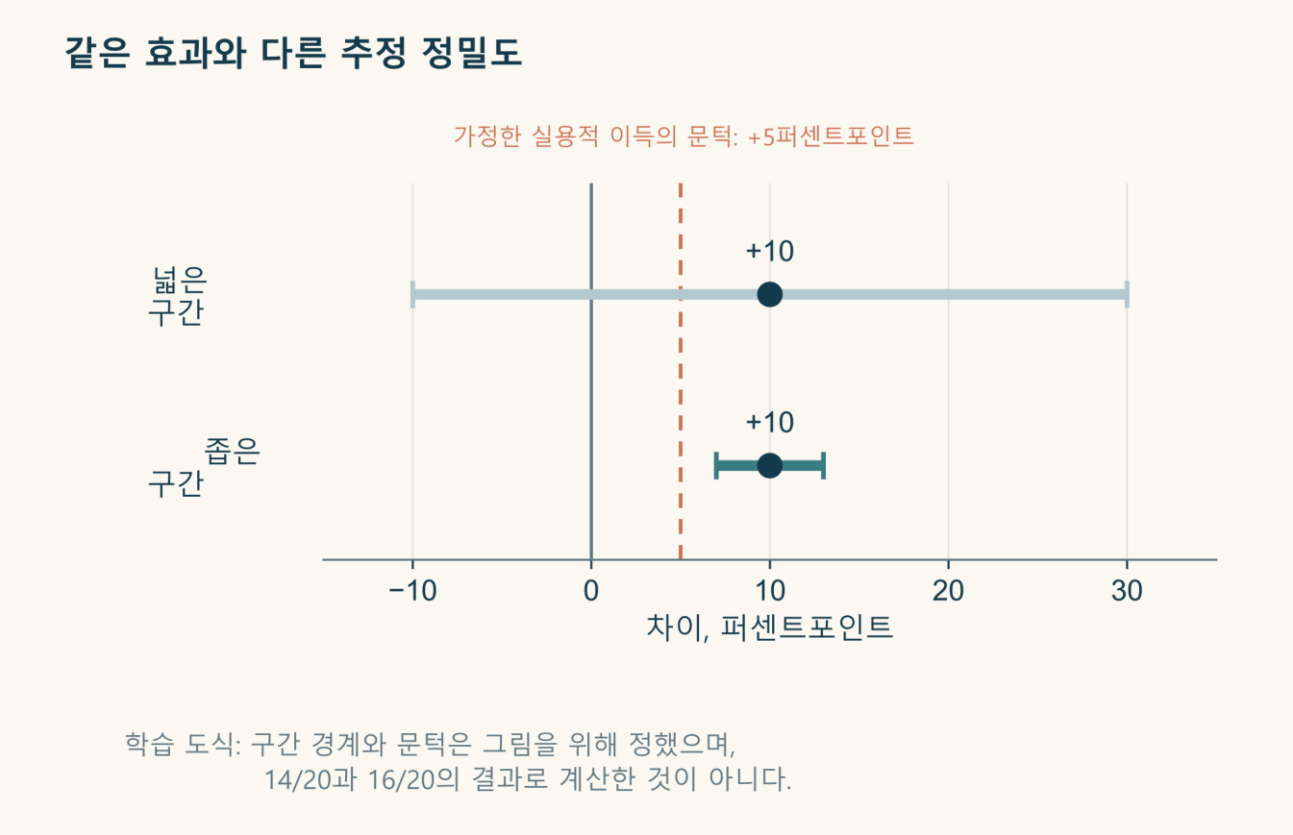


그림 17. 효과 크기와 추정 정밀도. 두 학습 시나리오의 점추정값은 모두 +10퍼센트포인트다. −10부터 +30까지의 넓은 구간은 악화와 개선을 모두 허용한다. +7부터 +13까지의 좁은 구간은 가정한 문턱 +5보다 높다. 경계와 문턱은 설명을 위해 정했으며, 이 장의 표로 계산한 신뢰구간이 아니다.

처음에는 실용적인 질문으로 충분하다. 채택한 추정 방법에서 자료가 눈에 띄는 이점, 이점이 없는 경우, 더 나쁜 경우에 모두 들어맞는가? 그렇다면 그 셋 중 어느 쪽인지 아직 결정할 수 없다. 반대로 가능한 값의 넓은 범위를 배제할 수 있다면 예상한 개선이 없었어도 의미 있는 결론을 내릴 수 있다. 정밀한 추정은 성공을 판정할 때뿐 아니라 이론의 적용 범위를 좁힐 때도 쓸모가 있다.

실제로 어느 정도의 차이가 중요한지도 과제에 맞춰 정해야 한다. 도식을 쓰는 데 시간이 더 들고 일도 훨씬 복잡해진다면 과제 천 개에서 정답이 하나 더 늘어나는 것으로는 도식을 쓸

이유가 부족할 수 있다. 다른 과제에서는 그 정도의 차이도 유용할 수 있다. 통계적 정밀도를 보면 효과를 얼마나 잘 아는지는 알 수 있다. 그 효과가 유용한지는 적용 목적에 비추어 판단한다.

메커니즘에 대한 가설을 비교할 때는 작지만 특정한 차이도 유용하다. 예를 들어 두 모형이 평소 과제를 비슷하게 풀지만 일정한 지연에는 다르게 반응할 수 있다. 미리 예측한 그 차이를 믿을 만하게 측정하면 전체 성적이 거의 달라지지 않아도 어떤 메커니즘이 더 잘 설명하는지 판단할 수 있다. 무엇을 의미 있는 차이로 볼지는 연구 질문과 함께 정해야 한다.

## 응답 수와 독립적인 사례 수를 구별하기

어떤 연구에서 한 사람이 과제 백 개를 풀었다고 하자. 다른 연구에서는 백 명에게서 답을 하나씩 받았다. 두 파일 모두 백 행이지만 서로 다른 질문에 답한다. 첫째는 선택한 과제 묶음에서 한 사람의 응답이 얼마나 달라지는지 더 잘 보여 준다. 둘째는 한 사람을 잠깐씩만 측정했지만 사람들 사이에서 어떤 차이가 있는지 알려 준다.

한 참가자의 답 백 개를 사람 일반의 특성에 관한 독립적인 확인 백 건으로 세면 불확실성을 크게 과소평가할 수 있다. 추론의 단위가 한 번의 시행인지, 여러 시행을 묶은 회차인지, 한 사람인지, 모형의 구성인지 미리 정해야 한다. 계산 연구에서도 한 번 학습한 신경망의 많은 출력을 독립적으로 만든 여러 체계의 출력으로 간주할 때 같은 문제가 생긴다.

반복에도 쓸모가 있다. 선택한 단위 안에서 결과가 얼마나 안정적인지 살피는 데 도움이 된다. 어느 응답이 같은 참가자, 같은 실행, 같은 자료 묶음에 속하는지 남겨 두는 것이 중요하다. 그러면 자료를 한 표로 정리해도 그 의존성을 잃지 않고 분석에 반영할 수 있다.

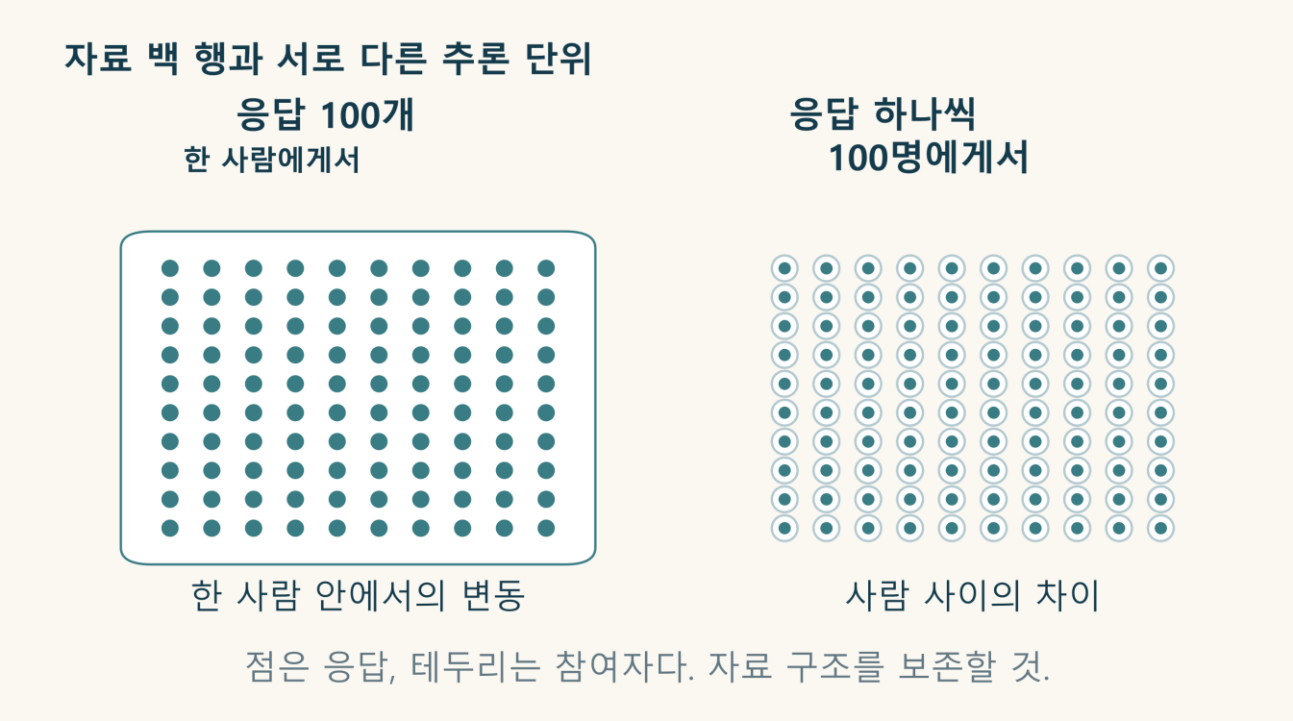


그림 18. 서로 다른 자료 수준인 응답과 참여자. 왼쪽의 응답 백 개는 한 사람에게서 나온다. 오른쪽에서는 응답마다 참여자가 다르다. 행의 수가 같아도 두 설계는 서로 대체할 수 없다. 서로 다른 수준의 변동성에 관한 정보를 주기 때문이다.

## 여러 번 시도하면 발견의 의미가 달라지는 이유

연구자가 지표 스무 개를 살핀 뒤 가장 눈에 띄는 차이만 보고했다고 하자. 각각의 계산이 정확해도 최종 보고에는 어떤 탐색을 했는지 모두 드러나지 않는다. 독자에게는 처음부터 그 결과만 예상한 것처럼 보인다. 결과가 얼마나 뜻밖인지 평가하려면 많은 후보 중에서 눈에 띄는 결과를 골랐을 가능성도 고려해야 한다.

Joseph Simmons, Leif Nelson, Uri Simonsohn은 시뮬레이션과 실험 예제를 통해 자료 수집과 분석의 유연성이 어떻게 거짓 가설에 대해 설득력 있어 보이는 통계 결과를 만들어 낼 수 있는지 보였다. 이들은 문제를 공개되지 않은 연구상의 결정과 연결하고 절차의 투명성을 높이자고 제안한다. [11]

논문에 나오는 p값은 가설이 거짓일 확률을 뜻하지 않는다. 명시한 귀무 모형과 절차에서 관찰값만큼 또는 그보다 극단적인 결과가 나올 확률을 나타낸다. 따라서 이 값이 작더라도

효과의 추정, 가정의 점검, 수행한 탐색의 공개가 필요하다. 첫 연구에서는 인상적인 숫자 하나보다 명확한 계획과 빠짐없는 결과가 더 유용하다.

## 같은 추정값에서도 판단은 달라질 수 있다

두 가상 사례에서 개선 폭의 추정값이 모두 십 퍼센트포인트라고 하자. 마이너스 십에서 플러스 삼십까지의 구간은 악화와 상당한 개선을 모두 허용한다. 플러스 칠에서 플러스 십삼까지의 구간은 가능한 값을 훨씬 좁힌다. 미리 오 퍼센트포인트 이상의 개선을 유용하다고 정했다면 두 번째 사례에서 더 분명히 판단할 수 있다. 여기서 구간의 경계는 설명을 위해 정했다. 앞 표의 20개 중 14개와 20개 중 16개라는 값에서 계산한 것이 아니다. 이 예시 그림을 그 마흔 개 응답의 통계 분석 결과로 써서는 안 된다.

### 여기까지 확인한 것

결과의 크기와 그 크기를 얼마나 정밀하게 아는지는 구별해야 한다. 자료의 행이 많을 때도 실제로 어느 행들이 독립적이고 추정으로 어떤 질문에 답하려는지 알아야 그 자료를 제대로 쓸 수 있다.

## 실습 8 세 가지 결론 구별하기

실습 목표는 한 번의 시행에서 나온 결과를 기술하는 일, 자료 생성 메커니즘에 관한 결론, 다음 검증의 계획을 구별하는 것이다.

제7장의 비율과 퍼센트, 비교 설계를 활용한다. 여기서 신뢰구간은 추정 절차의 특성으로 살펴본다. 실습을 위해 공식을 유도할 필요는 없다.

예상 소요 시간은 20–30분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 이 장의 합성 자료 표에서 네 번째 반복은 A가 20개 중 19개, B가 20개 중 15개다. 여섯 번째 반복은 A가 20개 중 10개, B가 20개 중 14개다. B의 정답률에서 A의 정답률을 뺀 값을 퍼센트포인트로 구한다.
2. 두 행 가운데 하나만 보여 주었을 때 어떤 설명이 나올지 적는다. 그 설명으로 생성기의 어느 조건이 더 좋은지 판단할 수 없는 이유를 밝힌다.

3. 효과가 크지만 추정이 부정확한 경우, 영에 가까운 효과를 정밀하게 추정한 경우, 자료에 따라 결과가 다른 경우에 각각 다음에 할 일을 하나씩 제안한다.

4. 한 사람의 응답 백 개와 백 명에게서 한 번씩 받은 응답에서 각각 무엇을 독립 반복의 단위로 볼지 적는다. 두 경우의 차이를 분석 계획에 남긴다.

두 차이값과 그 해석의 한계, 근거를 갖춘 다음 단계 세 가지를 남겨 둔다.

20개 중 19개의 비율은 95퍼센트다. 두 비율을 퍼센트로 나타낸 뒤 빼면 퍼센트포인트가 나온다. 상대적 증가율은 따로 계산해야 한다.

탐구를 이어 가려면 실제로 관심 있는 효과 크기와 그 크기를 구별하는 데 필요한 정밀도를 미리 정한다. 서로 의존하는 관찰 중 무엇을 연구 전체의 독립적인 반복으로 셀 수 없는지 적는다.

해설과 답은 부록 F 제8장 참조, 299쪽

**자기 점검 문항**

1. 80퍼센트와 70퍼센트의 차이를 십 퍼센트포인트라고 할 때, 이를 상대적 증가율 십 퍼센트와 같게 볼 수 없는 까닭은 무엇인가?
2. 한 모형의 응답 백 개를 조사하는 일과 독립적으로 학습시킨 모형 백 개를 조사하는 일이 서로 다른 까닭은 무엇인가?
3. 예상한 개선이 없었다는 사실도 의미 있는 결과가 될 수 있는가?

비교에 쓸 지표와 불확실성의 범위를 이제 명확히 말할 수 있다. 다음에는 결과를 순서대로 만들어 내는 메커니즘을 살펴본다. 내부 상태가 한 시점과 다음 시점을 이어 준다.

제3부

# 메커니즘과 모형

제9장

# 신경세포에서 시스템으로

앞 장들에서는 설명을 검증하는 법을 배웠다. 이제 작동 과정을 처음부터 끝까지 추적할 수 있는 작은 설명을 만들어 보자. 같은 입력에도 응답이 달라져야 할 때가 있다. 계가 이력을 간직하기 때문이다. 문턱 누적기를 쓰면 신경생물학이나 복잡한 수학을 몰라도 이를 확인할 수 있다. 이 모형과 신경세포의 관계는 선택한 성질에 한정된다. 체험과의 관계를 더 묻기 위해서는 새로운 관계와 논거가 필요하다.

## 변하는 상태에서 시작하기

지금까지는 체험을 기술하고 주장을 검증하는 방법을 살폈다. 이제 신호를 처리하는 계로 넘어간다. 새로 묻는 핵심 질문은 현재의 입력과 내부 상태가 다음 단계를 어떻게 결정하는가이다. 생물학적 과정에도 실습용 프로그램에도 적용할 수 있는 질문이다. 답을 구하면 같은 입력이 때로 다른 결과를 내는 이유를 설명할 수 있다.

뉴런은 신경계를 이루는 세포다. 세포막은 내부 환경과 외부 환경을 가른다. 전하를 띤 입자들의 분포와 이동은 세포의 전기적 성질과 관련된다. 많은 뉴런은 활동전위로 신호를 전달한다. 활동전위는 막전압이 짧게 변하며 세포의 해당 부위를 따라 전파되는 현상이다. 이 장에서는 뉴런을 입력과 내부 동역학을 갖추고 다른 세포에 영향을 주는 물리적 계로 이해하면 충분하다.

이것으로 뉴런의 모든 작동을 모형화한 것은 아니다. 세포의 종류와 부위에 따라 성질이 다르다. 상호작용에는 화학적 과정과 여러 시간 척도가 관여한다. 프로그램 요소에 붙인 «뉴런»이라는

이름은 보통 특정한 추상화를 뜻한다. 그러므로 모형을 읽을 때는 선택한 과제를 위해 어떤 성질을 남겼고 어떤 성질을 생략했는지 묻는다.

## 정량적 설명과 이미지의 차이

앨런 호지킨과 앤드루 헉슬리는 전도도의 변화와 전기적 흥분을 연결하여 신경섬유의 막전류를 정량적으로 기술했다. 이 연구는 측정에서 출발해 과정의 역동을 재현할 수 있는 규칙 체계로 나아가는 연구의 전환을 보여 준다. 이는 신경세포의 특정 전기적 성질을 다루는 모형이다. 주관적 체험이 발생한다는 증거로 제안된 것은 아니다. [12]

«뉴런이 이웃들의 표를 모은다»는 유용한 이미지는 출발에 도움을 주지만 중요한 질문들을 숨긴다. 입력의 영향은 얼마나 오래 남는가? 다음 응답이 이전 응답에 의존하는가? 문턱이 있는가? 응답 뒤에는 무슨 일이 일어나는가? 계산 가능한 모형은 매 단계에서 저자가 새로 판단하지 않아도 결과를 얻을 만큼 이 질문들에 명확히 답하도록 요구한다.

첫 모형은 생물학적 모형보다 훨씬 단순하다. 누적과 누출, 문턱을 살피기 위한 것이다. 수치는 가정한 값이며 밀리볼트로 측정한 값이 아니다. 뉴런의 일부 성질과 닮았다고 살아 있는 세포의 모든 성질을 이 모형에 옮겨 붙일 수는 없다.

## 문턱 누적기의 예제 분석

처음 값이 영인 상태 변수를 생각해 보자. 각 단계에서는 이전 값의 60퍼센트를 남기고 새 입력을 더한다. 그 결과가 일에 이르면 계는 펄스를 내고 상태를 영으로 초기화한다. 문턱에 이르지 않으면 펄스가 나오지 않으며, 상태는 다음 단계까지 남았다가 다시 누출된다.

입력으로 0,6; 0,6; 0,6; 0; 0,6을 넣는다. 첫 단계에는 이전에 누적된 값이 없으므로 새 값은 0,6이다. 둘째 단계에는 이전 상태에서 0,36이 남고 여기에 0,6을 더해 0,96이 된다. 셋째 단계에는 0,576이 남고 다시 0,6을 더한다. 1,176은 문턱을 넘으므로 펄스가 발생하고 상태는 영으로 초기화된다.

넷째 단계에는 입력이 영이므로 상태도 영으로 남는다. 다섯째 단계에는 다시 0,6이 된다. 첫째와 둘째, 셋째 단계의 입력은 같아도 이력이 달라 상태가 달라졌다. 프로그램에는 «언제 응답할지» 따로 결정하는 부분이 없다. 결과는 갱신 규칙에서 나온다.

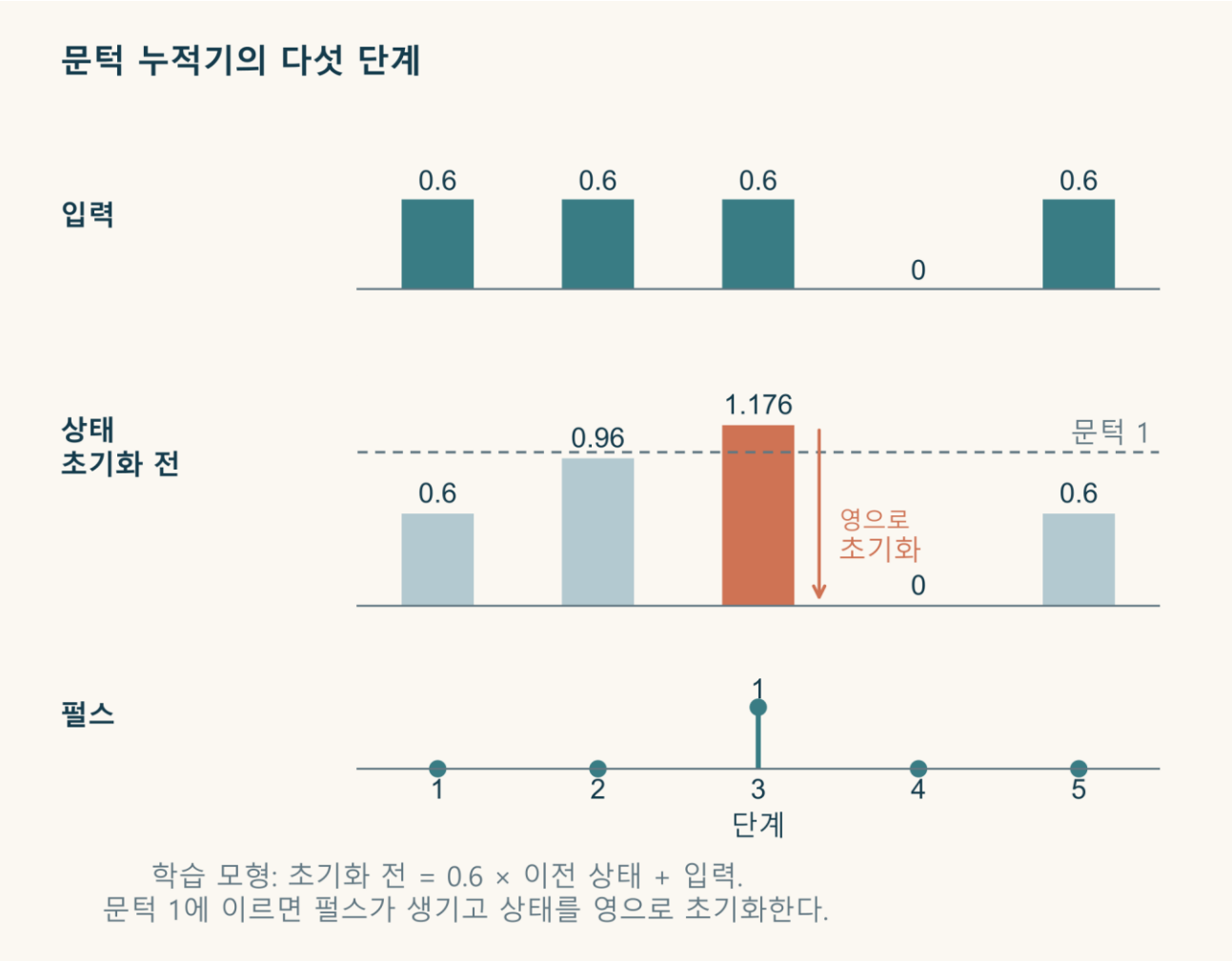


그림 19. 누적과 초기화의 동역학. 이 장의 프로그램이 낸 정확한 결과다. 입력 0,6; 0,6; 0,6; 0; 0,6에 대해 초기화 전 상태는 0,6; 0,96; 1,176; 0; 0,6이다. 셋째 단계에서만 문턱 1에 이른다. 모든 양은 가정한 값이며 밀리볼트 단위가 아니다.

다음은 Python으로 작성한 실행 가능한 전체 코드다. state는 상태, value는 다음 입력, spike는 펄스의 유무를 나타낸다. 코드에서는 소수 구분 기호로 마침표를 쓴다. 들여쓰기를 보면 어떤 동작을 입력마다 반복하고 어떤 동작을 문턱을 넘었을 때만 수행하는지 알 수 있다.

```
inputs = [0.6, 0.6, 0.6, 0.0, 0.6]
state = 0.0
for value in inputs:
    before_reset = 0.6 * state + value
    spike = before_reset >= 1.0
    state = 0.0 if spike else before_reset
    print(round(before_reset, 3), int(spike), round(state, 3))
```

프로그램은 초기화할 수 있는 시점 직전의 값, 펄스가 있으면 1이고 없으면 0인 표시, 마지막으로 갱신을 마친 상태를 출력한다. 코드를 실행해도 되고 손으로 계산해도 된다. 두 방법 모두 같은 규칙에서 같은 수열을 얻어야 한다. 이런 투명성은 내부 작동 전체를 손으로 추적하기 어려운 큰 모형으로 넘어가기 전에 도움이 된다.

## 한 요소에서 망으로

이제 누적기 여러 개를 생각해 보자. 하나의 출력이 다른 누적기의 갱신 규칙에 들어간다. 연결 계수는 그 영향의 세기를 정한다. 양의 기여는 상태를 문턱에 가깝게 하고 음의 기여는 문턱에서 멀어지게 할 수 있다. 이는 흥분성 영향과 억제성 영향을 살피는 실습용 도식이며, 모든 생물학적 상호작용의 보편적인 기술은 아니다.

연결이 입력 요소에서 출력 요소로만 이어지면 처리는 한 방향의 순서로 진행된다. 신호가 여러 연결을 거쳐 앞서 참여한 요소로 돌아올 수 있으면 되먹임이 생긴다. 그 덕분에 앞선 결과가 이후 처리를 바꿀 수 있다. 그러나 되먹임이 반드시 안정된 활동을 유지하지는 않는다. 규칙에 따라 활동이 감쇠하거나 증폭되거나 진동하는 등 여러 양상이 나타날 수 있다.

활동의 변화와 계 자체의 변화를 구별해야 한다. 이 모형의 상태는 단계마다 달라지지만 계수 0,6은 일정하다. 학습을 추가하면 계수도 별도의 규칙에 따라 달라질 수 있다. 그러면 모형에는 현재 상태의 짧은 이력과 매개변수의 더 긴 변화라는 두 종류의 이력이 생긴다. 이 층위를 섞으면 모형이 정확히 무엇을 기억했는지 파악하기 어려워진다.

요소가 많다는 사실만으로 망이 설명이 되지는 않는다. 그 조직이 어떤 과제를 해결하는지 보여 주어야 한다. 어떤 망은 사건 사이에 신호를 유지하고, 다른 망은 특정한 형태의 순서를

가려내며, 또 다른 망은 행동을 선택할 수 있다. 조직과 체험의 관계를 묻기 전에도 이런 차이를 검증할 수 있다.

수치나 펄스는 조건 및 이용과의 관계를 통해 연구에서 의미를 얻는다. 어떤 요소가 빨간 영상에 더 자주 반응한다면 형태, 밝기, 주변을 바꾸어도 관계가 유지되는지 검증할 수 있다. 이어 다른 구성요소가 그 응답을 색 구별에 이용하는지 밝혀야 한다. 학습 표본에서 우연히 색과 함께 변한 밝기에 실제로 반응했다면 «빨강 검출기»라는 이름은 성급하다. 내용의 표상에는 검증이 필요하다. 망의 노드에 그럴듯한 이름을 붙이는 것만으로는 충분하지 않다.

## 한 단계는 여러 행동으로 이루어진다

이전 상태가 0,96이면 먼저 그 값의 0,6만큼을 남긴다. 결과는 0,576이다. 여기에 새 입력 0,6을 더하면 1,176이 된다. 그다음 문턱 1에 이르렀는지 검사한다. 문턱에 이르렀으므로 펄스를 기록하고 상태를 영으로 초기화한다. 초기화 전의 값과 초기화 후의 상태는 다르다. 동작의 순서를 바꾸면 다른 모형이 된다. 그러므로 최종 수치 한 줄로 명시적인 전이 규칙을 대신할 수 없다.

### 여기까지 확인한 것

누적기는 보존된 상태의 인과적 기여를 보여 준다. 펄스는 별도의 내부 결정이 아니라 규칙에서 나온다. 이제 합이 같은 입력들을 시간에 따라 어떻게 배치하느냐가 어떤 차이를 만드는지 검증한다.

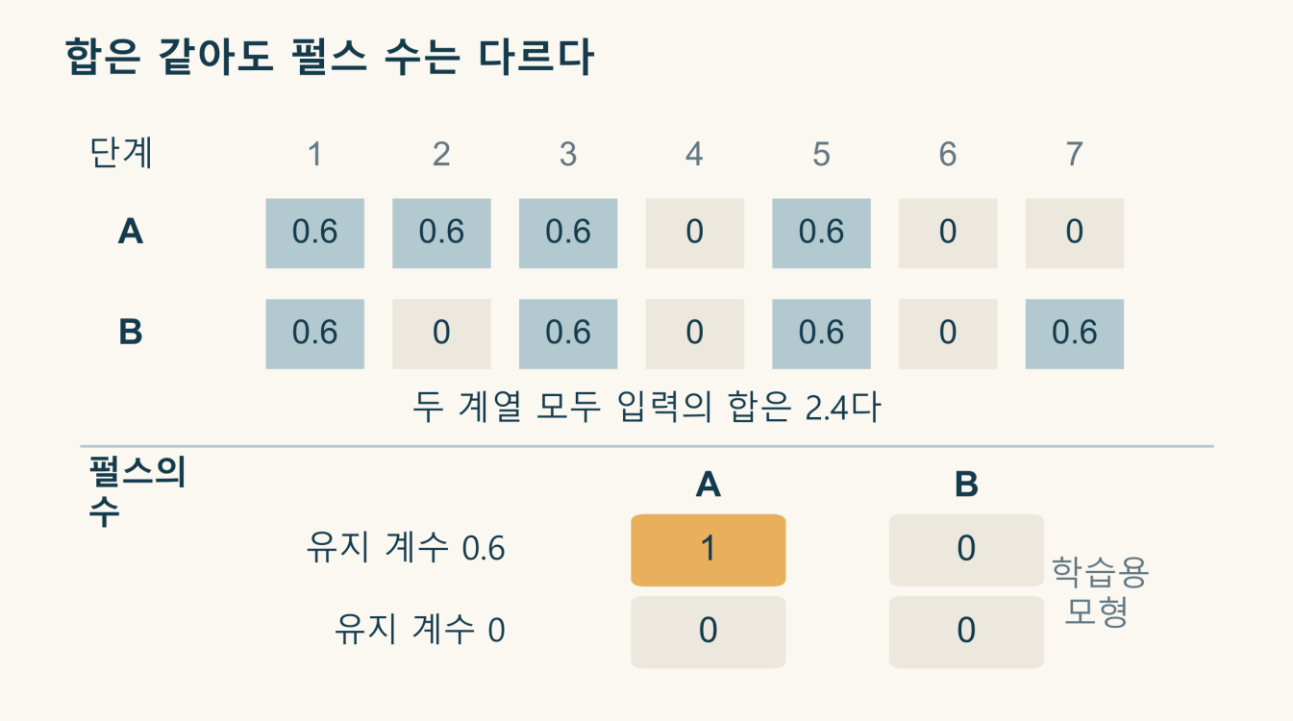


그림 20. 입력의 배분과 상태 유지의 차단. 실습의 두 순서는 합이 2,4로 같다. 유지 계수가 0,6일 때에는 첫째 순서만 펄스 하나를 만든다. 유지 계수가 영이면 두 계열 모두 펄스가 없다. 이 개입은 학습 모형에서 상태 이력의 역할을 구분해 낸다.

## 실습 9 같은 합과 다른 시간 배치

실습 목표는 시간에 따른 누적이라는 설명을 전체 합만으로 설명하는 경우와 비교해 검증하는 것이다.

소수를 곱할 줄 알면 된다. 입력은 외부에서 오고, 상태는 단계 사이에 보존되며, 출력은 갱신 뒤에 나온다.

예상 소요 시간은 20–30분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. A에는 0,6; 0,6; 0,6; 0; 0,6; 0; 0을, B에는 0,6; 0; 0,6; 0; 0,6; 0; 0,6을 사용한다. 처음 상태는 영, 유지 계수는 0,6, 문턱은 1이다.

2. 계산 전에 펄스의 수를 예측한다. 단계마다 입력, 초기화 전의 값, 펄스, 초기화 후의 상태를 적는다.

3. 두 입력열에 같은 규칙을 적용한다. 손으로 계산하거나 이 장의 코드를 사용해도 된다.

4. 유지 계수를 영으로 바꾸어 반복하고 어떤 차이가 사라졌으며 그 이유가 무엇인지 설명한다.

두 단계별 표와 사전 예측, 입력 사이 간격의 역할에 관한 결론을 남겨 둔다.

힌트: 펄스 뒤 다음 단계는 영에서 시작한다. 1,176을 보존된 상태로 넘기지 마라.

탐구를 이어 가려면 합은 같되 전체 합으로 설명할 때와 시간 구조로 설명할 때 예측이 달라지는 새 입력열 한 쌍을 제안한다. 계산 전에 규칙을 확정한다.

해설과 답은 부록 F 제9장 참조, 300쪽

## 체험 연구를 향해 어떤 진전을 이루었는가

입력, 상태, 전이 규칙, 출력, 매개변수라는 동적 모형의 최소한의 언어를 마련했다. 이를 통해 계가 무엇을 구별하고 무엇을 유지하며 과거를 어떻게 사용하는지 물을 수 있다. 이후에는 출처에 대한 표상, 신뢰성과 자기 행동에 대한 평가를 추가한다. 각 구성요소는 설명할 행동과 검증할 예측을 위해 도입한다.

펄스나 되먹임이 있다는 사실만으로 퀄리아의 문제를 해결할 수는 없다. 그래도 이제 «복잡성에서 의식이 생긴다»는 말보다 정확하게 가설을 세울 수 있다. 특정한 조직을 밝히고 그 상태와 체험의 성질 사이에 예상되는 관계를 명시한 다음 대안을 구별할 조건을 찾는다. 단순한 모형은 이런 준비 역할을 한다.

### 자기 점검 문항

1. 첫 단계와 셋째 단계의 입력이 같아도 출력이 달라질 수 있는 까닭은 무엇인가?
2. 상태의 변화는 매개변수의 학습과 어떻게 다른가?
3. 합은 같고 입력 사이 간격은 다른 입력열을 비교하면 무엇을 검증할 수 있는가?

이제 상태는 과거를 간직할 수 있다. 다음 질문은 그 상태가 정확히 무엇을 표상하며 맥락은 같은 입력의 사용 방식을 어떻게 바꾸는가이다.

제10장

# 신호 처리와 내용에 대한 접근

누적기는 이력이 응답을 어떻게 바꾸는지 보여 주었다. 지각을 이해하려면 여기서 더 나아가 내용, 해석 방식, 결과에 대한 접근을 구별해야 한다. 같은 타원도 이웃 기호에 따라 글자나 숫자로 쓰일 수 있다. 이 사례로 순방향 처리와 반복 처리를 살펴본다. 구조를 비교할 때는 한쪽이 추가 입력이나 더 긴 시간 덕분에만 유리해지지 않는지 살핀다.

## 한 영상과 여러 연구 질문

페이지에서 원하는 절의 번호를 찾는다고 하자. 눈앞에는 글자와 여백, 문장부호, 작은 설명이 함께 있다. 이 요소들이 모두 눈앞에 있다고 현재 과제에 같은 정도로 영향을 주거나 나중에 기술할 수 있는 정도가 같은 것은 아니다. 이 상황을 연구하려면 몇 가지 개념을 구별해야 한다.

정보 처리는 입력과 내부 규칙에 따라 시스템의 상태가 달라지는 것을 뜻한다. 여기서 작업상의 기능적 의미로 사용하는 주의는 처리의 선택적 배분을 가리킨다. 어떤 출처나 특징이 다른 것보다 우선되는 것이다. 접근은 특정 행동, 추론, 응답에 내용을 이용할 수 있음을 뜻한다. 보고는 내용을 전달하는 구체적인 행위다. 체험은 상황이 주관적으로 어떻게 주어지는지를 가리킨다.

이 개념들은 한 사례를 가리킬 수 있지만 측정 방법은 다르다. 단서를 준 뒤 절 번호에 대한 응답이 달라졌는지 검증할 수 있다. 무엇이 선명하게 보였는지 물을 수도 있다. 작은 설명의 위치가 다음 과제를 위해 보존되었는지도 알아볼 수 있다. 모든 결과를 «의식함»이라는 한마디로 묶으면 모형이 어떤 과정을 설명하는지 파악하기 어려워진다.

## 보고의 양이 절차에 달린 이유

조지 스펄링은 기호 배열을 짧게 제시하고, 전체 보고와 제시가 끝난 뒤 지정한 부분의 보고를 비교했다. 부분 보고의 이점과 단서 지연에 따른 변화는 짧은 제시 뒤 정보의 이용 가능성을

차례로 말한 기호의 수만으로 평가할 수 없음을 보여 주었다. 이 실험은 저장과 보고의 한계에 관한 질문을 제기한다. 그 자체만으로 체험의 전체 범위에 관한 논쟁을 끝내지는 않는다. [13]

일상적인 사례에서도 같은 교훈을 얻을 수 있다. 탁자 위의 물건 여러 개를 알아차린 뒤 첫 물건부터 자세히 설명한다고 하자. 설명하는 동안 다른 물건에 대한 정보는 이용하기 어려워질 수 있다. 그러면 마지막에 나온 짧은 목록은 처음의 처리뿐 아니라 시간과 보고 순서, 요구한 세부 수준에도 좌우된다. 구체적인 연구에서는 이 과정들의 기여를 구별해야 한다.

반대 상황도 가능하다. 말로 설명하기 어려운 정보가 정답의 근거가 될 수 있다. 따라서 «얼마나 이야기할 수 있었는가»는 보고를 연구하는 데 유용한 질문이지만, 이용 가능한 내용 전체의 척도로 쓰려면 추가 근거가 필요하다. 정보가 어느 시점에 어떤 동작에 이용 가능했는지 밝히는 편이 좋다.

## 주의는 과제를 바꾼다

Daniel Simons와 Christopher Chabris는 주의 과제를 수행하며 역동적인 장면을 관찰할 때 뜻밖의 사건을 알아차리는지를 연구했다. 발견 확률은 주과제의 구성과 뜻밖의 대상의 성질에 따라 달랐다. 이 연구는 특정 부하에서 알아차림의 한계를 보여 준다. 이후 보고가 없다는 사실을 모든 처리의 부재를 판정하는 보편적 기준으로 삼아서는 안 된다. [14]

주의를 이해하려면 계가 지금 어떤 정보를 우선하여 사용하는지 물을 수 있다. 절 번호를 찾을 때는 이웃 문단의 내용보다 숫자의 모양이 유용할 수 있다. 오자를 확인할 때는 우선순위가 바뀐다. 같은 외부 요소들도 목적에 따라 서로 다른 계산 과제에 참여한다.

주의와 의식이 밀접하게 관련된다고 동의어가 되는 것은 아니다. 먼저 주의의 지표와 체험의 지표, 비교 조건을 정해야 한다. 탐색 성공으로 주의를 정의하고 선명도 보고로 체험을 정의했다면 연구하는 것은 바로 그 지표들의 관계다. 내용을 갖춘 이론은 두 지표가 일치하거나 어긋나는 경우를 설명해야 한다. 이름을 바꾸어 구별을 없애서는 안 된다.

## 모호한 기호의 예제 분석

글자 O나 숫자 0을 나타낼 수 있는 타원을 생각해 보자. 실습용 카드 두 장의 가운데 그림은 같다. 첫 카드에서는 양옆에 K와 T가 있고 둘째 카드에서는 1과 4가 있다. 첫 맥락에서는 글자로, 둘째 맥락에서는 숫자로 해석하는 것이 자연스럽다. 맥락은 기호가 기록에서 어떤 역할을 할 가능성이 큰지에 관한 정보를 더한다.

단순한 모형을 만들어 보자. 첫 단계에는 가운데 기호만 평가하고 두 해석을 남겨 둔다. 둘째 단계에는 이웃 기호를 본다. 양쪽이 모두 글자이면 글자 해석이, 모두 숫자이면 숫자 해석이 우선된다. 셋째 단계에는 선택한 범주를 출력한다. 이렇게 하면 단계마다 어떤 정보가 들어왔으며 가운데 그림이 같은데도 결과가 왜 달라지는지 밝힐 수 있다.

이것은 저자가 만든 실습용 사례이며 특정 실험을 재현한 것이 아니다. 맥락이 입력의 해석에 어떻게 참여할 수 있는지 보여 준다. 그때 주관적으로 보이는 형태가 달라졌는지, 형태에 붙인 이름만 달라졌는지는 관찰자에게 별도로 물어야 한다. 똑같이 «글자»라고 답해도 체험의 기술은 다를 수 있으며, 범주 모형은 아직 그 차이를 설명하지 않는다.

## 한 번의 처리와 반복적 구체화

순방향 처리에서는 정보가 연속된 변환을 거쳐 응답으로 나아간다. 재귀 처리에서는 결과가 되돌아와 앞서 참여한 구성요소의 작동을 바꿀 수 있다. 여기서 «재귀»는 그런 되돌아옴이 있다는 뜻이다. 예를 들어 먼저 인식한 맥락이 이후 국소적 형태의 평가에 영향을 줄 수 있다.

이 과제의 재귀 처리 버전을 만들 수 있다. 먼저 중앙 기호의 모호한 상태를 보존하고, 이어 이웃 기호의 정보를 얻어 이전 상태를 갱신하는 것이다. 하지만 카드 전체를 한꺼번에 받아 규칙 하나로 범주를 고르는 모형도 만들 수 있다. 설명한 두 예제에서 이들의 응답은 일치한다. 따라서 사람의 응답과 일치한다는 것만으로 내부 조직이 정해지는 것은 아니다.

구조를 의미 있게 비교하려면 이용 가능한 정보의 양과 계산 자원을 고려해야 한다. 한 모형은 카드 전체를 보고 다른 모형은 타원만 보았다면 첫 모형의 이점은 입력 차이만으로도 설명된다. 반복 처리에 더 긴 시간을 주었다면 되먹임과 추가 계산 가운데 무엇이 이점을 만들었는지 알아야 한다. 좋은 과제에서는 두 설명이 서로 다른 예측을 낸다.

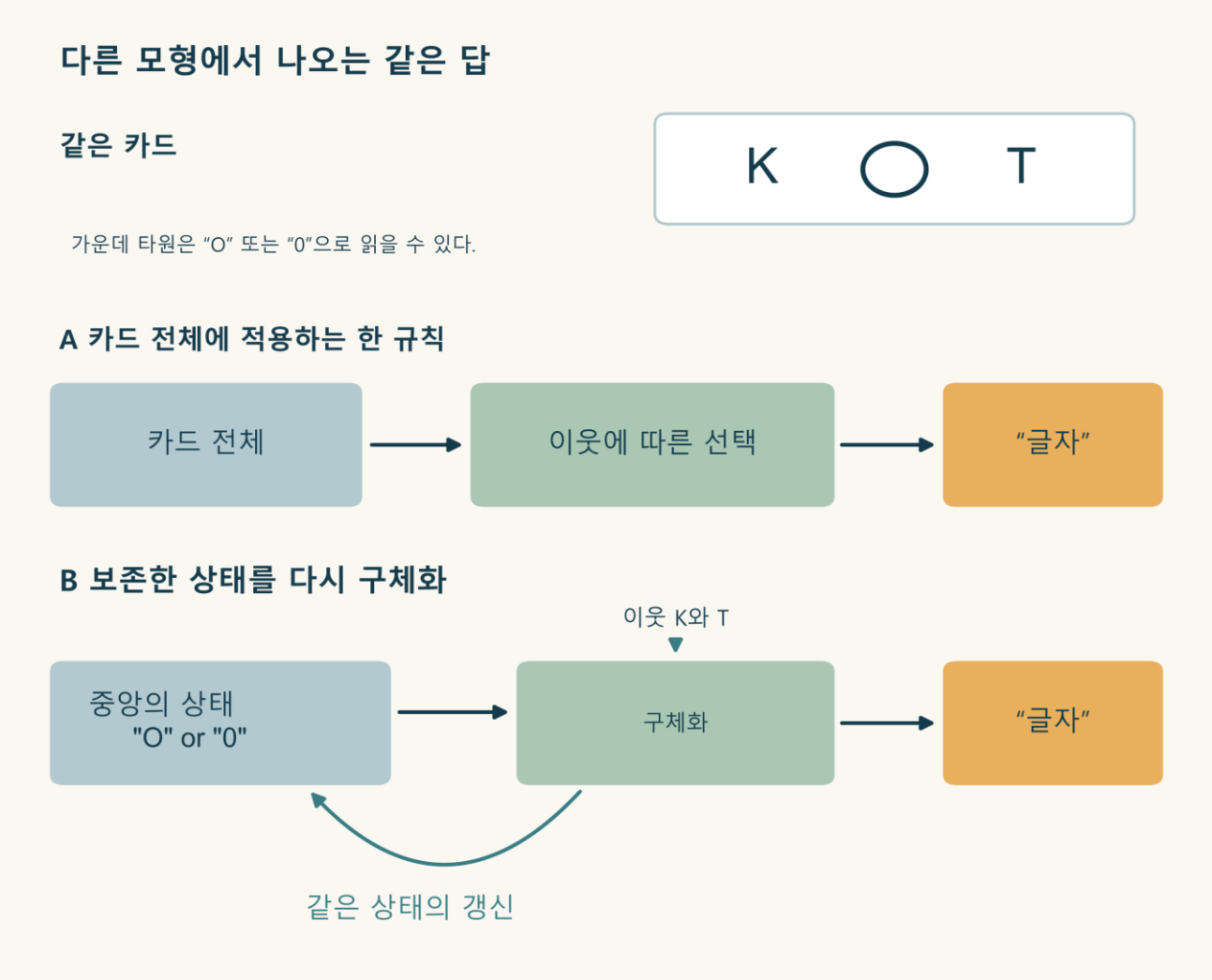


그림 21. 맥락을 이용하는 두 방식. 두 학습 모형 모두 이웃이 K와 T일 때 가운데 타원을 글자라고 판단한다. 첫 모형은 카드 전체에 규칙 하나를 적용한다. 둘째 모형은 중앙의 모호한 상태를 보존하고, 맥락을 받은 뒤 그 이전 상태를 구체화한다. 되돌아오는 화살표는 이 갱신을 뜻한다. 예제에서 응답이 같다는 사실이 내부 조직을 정하거나 체험의 존재를 입증하지는 않는다.

## 내부 상태에서 그 이용으로

선택한 범주가 프로그램 안에 보존되어 있다고 가정하자. 한 출력은 «글자»라는 말을 쓰고, 다른 출력은 기록을 이어 갈 사전을 선택하며, 셋째 출력은 추가 검증이 필요한지 결정할 수 있다. 이는 같은 처리 결과에 접근하는 여러 형태를 보여 준다. 내부 수치가 존재한다고 모든 구성요소가 실제로 그 수치를 이용하는 것은 아니다.

상태의 역할을 검증하려면 그 상태를 바꾸거나 특정 출력에서 잠시 이용할 수 없게 할 수 있다. 응답 출력은 손상되었지만 이어 쓸 내용의 선택은 유지된다면 기능적인 어긋남이 생긴 것이다.

연구에서는 이를 그대로 기술해야 한다. 보고 채널 하나가 고장 났다는 이유로 계에 정보가 전혀 없다고 단정할 수는 없다.

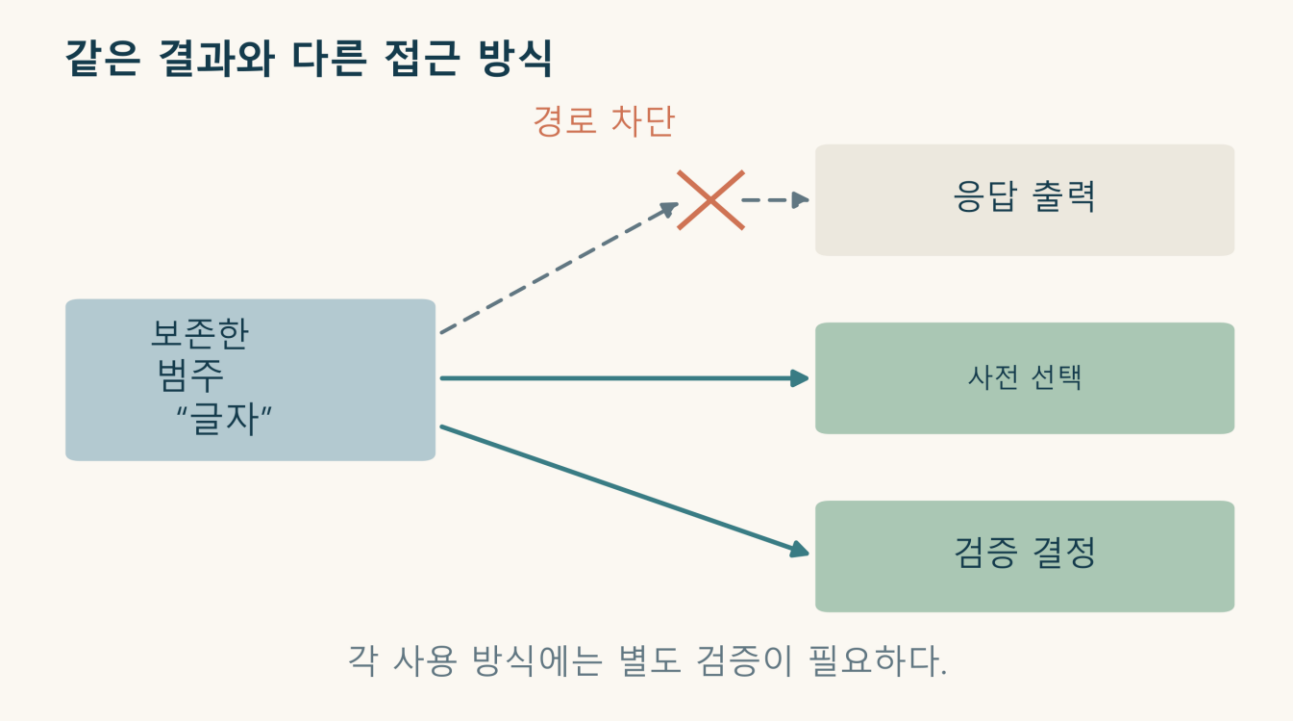


그림 22. 이용 채널의 선택적 차단. 학습용 구조에서 선택된 범주는 응답 출력, 사전 선택, 검증 결정에 이용된다. 출력 경로만 막는다. 다른 이용 방식이 보존되는지는 따로 검증하며, 보고 하나가 없다는 사실로부터 추론하지 않는다.

이런 분석은 주체 후보를 설계할 준비가 된다. 어떤 정보가 통합되고 어떤 행동이 그 정보에 의존하며 계가 자기 결정을 어떻게 평가하는지 물을 수 있다. 그 조직과 체험의 관계는 이후 이론에서 별도로 검증할 부분이 된다.

## 숨은 정답 표지를 항상 복원할 수는 없는 이유

타원과 양옆의 숫자가 똑같이 그려진 카드 두 장이 있다고 하자. 저자는 하나를 숫자로, 다른 하나를 글자로 의도했다. 장치가 영상만 볼 수 있다면 관찰만으로 두 의도를 구별할 수 없다. 이 경우 숨은 정답 표지와 다른 응답을 했다고 처리가 부실하다는 뜻은 아니다. 맥락의 이점을 연구하려면 이용 가능한 정보에 비추어 가장 근거 있는 해석을 평가하는지, 알 수 없는 의도를 맞히게 하는지 미리 정해야 한다. 이후 에이전트의 입력을 점검할 때도 이 구별이 필요하다.

**여기까지 확인한 것**

기호의 정확한 이름, 체험되는 형태, 여러 행동에서 내용을 이용할 수 있는가는 서로 다른 결과다. 이제 명시적으로 정한 카드 묶음 하나에서 설명들을 비교한다.

## 실습 10 세 가지 설명 비교하기

실습 목표는 맥락의 역할을 더 많은 정보를 얻는 데 따른 이점, 장치가 알 수 없는 환경의 정답 표지와 구별하는 것이다.

제9장에서 다룬 입력과 상태, 출력을 활용한다. 관찰 가능한 그림과 장치가 알지 못할 수 있는 저자의 숨은 정답 표지를 구별한다.

예상 소요 시간은 25–35분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 가운데 타원이 같은 카드 여섯 장을 만든다. 두 장의 양옆에는 K와 T, 두 장에는 1과 4를 넣는다. 나머지 두 장은 이 그림들을 되풀이하되 저자의 숨은 정답 표지는 반대로 정한다.
2. 카드마다 눈에 보이는 정보와 숨은 정답 표지를 따로 적는다. 양옆이 글자이면 글자 범주, 숫자이면 숫자 범주를 고르는 기본 규칙을 정한다.
3. 카드 전체를 한 번에 보는 장치와 가운데를 먼저 받고 이웃 기호를 나중에 받는 장치를 비교한다. 두 장치가 같은 정보를 갖게 되는 시점을 밝힌다.
4. 응답이 일치한다는 사실로 무엇을 알 수 있으며, 반복하여 구체화하는 메커니즘을 구별하려면 어떤 추가 조건이 필요한지 설명한다.

카드 여섯 장과 두 장치의 규칙, 시간에 따라 이용 가능한 정보의 명시적인 목록을 남겨 둔다.

입력 순서가 다르면 이미 조건이 다르다. 구조를 검증하려면 경쟁 모형에도 비교할 만한 기억과 계산을 허용해야 한다.

연구 확장: 중앙이 나타나기 전에 이웃들이 사라진다고 가정하라. 각 버전의 기억을 정하고 시행 전에 예측을 얻어라. 같은 외부 기억을 허용한 직접 처리 모형과 비교하라.

해설과 답은 부록 F 제10장 참조, 301쪽

**자기 점검 문항**

1. 말로 보고한 세부 사항의 수를 처음 체험 전체의 척도로 곧바로 삼을 수 없는 까닭은 무엇인가?

2. 맥락을 이용하는 모형의 성공이 재귀 처리의 필요성을 입증하지 않는 이유는 무엇인가?

3. «계가 정보에 접근할 수 있다»는 주장에서는 무엇을 구체적으로 밝혀야 하는가?

관찰과 숨은 의도를 구별했다. 이제 사건의 출처에도 같은 원칙을 적용한다. 계는 관찰한 변화가 자기 명령과 관련되는지를 어떻게 평가할 수 있을까?

제11장

# 자기관련성에서의 신체와 행동

맥락은 기호의 해석을 돕지만, 행동할 때에는 변화의 출처도 평가해야 한다. 움직임이 명령과 우연히 맞아떨어진 것일까, 명령이 실제로 움직임을 제어한 것일까? 이 장은 감각 신호들의 원인을 통합하는 과제와 자기 영향력을 구별하는 과제를 함께 다룬다. 수치 예제를 보기 전에 가설을 바꾸는 규칙으로서 확률적 갱신을 살핀다. 이 규칙과 체험되는 신체 소유감의 구별을 유지한다.

## 여러 신호가 한 사건에 속하는 경우

손을 들면 그 움직임이 보이고 위치의 변화가 느껴지며 옷에서 오는 감각도 달라진다. 이 사례를 기술할 때 신호가 많이 들어왔다고 말하는 것만으로는 부족하다. 신호들이 왜 같은 신체 부위와 같은 행동에 연결되는지 설명해야 한다. 옆 화면의 비슷한 움직임도 눈앞에서 일어나지만 외부 사건으로 남을 수 있다.

한 접근법에서는 이를 출처를 추론하는 과제로 본다. 계는 시각과 신체, 운동에 관한 정보가 공통 원인에서 왔는지 각각 독립적으로 생겼는지 평가해야 한다. 시간적 일치와 공간적 대응, 이전에 확인한 출처의 신뢰도를 고려할 수 있다. 여기서 «추론»은 내적인 문장이나 의식적 사고를 요구하지 않는다. 입력에 대한 가능한 설명들을 구별하는 모형의 작동을 뜻한다.

Majed Samad, Albert Jin Chung, Ladan Shams는 신체 소유감의 지각에 베이즈 인과 추론 모형을 적용했다. 인공 손 착각을 다룬 실험과 모형의 예측을 비교했다. 이 연구는 감각 원인들의 통합이라는 가정을 어떻게 계산 가능하게 만들고 자료로 검증할 수 있는지 보여 준다. 모든 신호 통합이 이미 자기 신체의 체험이라는 사실을 확립하지는 않는다. [15]



그림 23. 손과 도구, 관찰된 결과. 도구를 통한 행동을 그린 것이다. 손이 움직이면 화면의 무늬도 달라진다. 행위주체감과 자기관련성을 연구하려면 시스템이 어떤 정보들을 한 원인과 연결하며 그 연결을 어떻게 검증하는지 밝혀야 한다.

## 공통 원인만으로 내 몸이라고 할 수는 없다

책이 떨어지는 움직임과 부딪치는 소리에도 공통 원인이 있을 수 있다. 그러므로 출처를 통합하는 모형은 신체 소유감보다 일반적인 과제를 해결한다. 신체를 논하려면 어떤 신호가 신체에 속하고 그 위치는 어떻게 표상되며 사건과 계의 행동은 어떤 관계로 연결되는지 밝혀야 한다.

제3장의 구분으로 돌아가 보자. 신체 소유감은 어떤 대상이 자기 신체의 일부로 주어지는가에 관한 것이다. 행위 주체감은 행위와 결과 사이의 체험되거나 평가되는 연관을 가리킨다. 시스템은 외부 도구를 제어하면서도 도구와 신체를 구분할 수 있다. 그 순간 자발적 동작을 수행하지 않고도 움직이지 않는 팔다리의 위치에 관한 정보를 얻을 수 있다.

앞서 살펴본 Kalckert와 Ehrsson의 연구는 이런 구성요소를 따로 연구할 실험적 근거를 준다. [4] 따라서 우리 모형에는 감각 출처의 통합과 명령의 인과적 기여 평가를 위한 별도 변수가 있을

것이다. 일상적인 사건에서는 둘이 자주 함께 달라지더라도, 나누어 다루면 새롭게 검증 가능한 조건을 세울 수 있다.

## 증거가 설명의 가중치를 바꾸는 이유

먼저 두 질문을 구분하자. 새 관찰을 하기 전에 가설 C가 맞을 확률은 얼마인가? C가 맞다고 할 때 관찰 자체의 확률은 얼마인가? 첫 값은 가설의 사전확률이다. 둘째 값은 그 가설 아래에서 관찰의 가능도라고 한다. 둘은 계산으로 연결되지만 서로 바꾸어 쓸 수는 없다.

가정한 백 사례 가운데 육십 사례에서 목표가 오른쪽에 있다고 하자. 센서는 방향과 관계없이 팔십 퍼센트의 확률로 맞는다. 그러면 목표가 오른쪽인 육십 사례 가운데 마흔여덟 사례에서 오른쪽 신호가 예상된다. 목표가 왼쪽인 마흔 사례 가운데서도 여덟 사례에서 오른쪽 신호가 예상된다. 오른쪽 신호를 받았다면 바로 이 쉰여섯 사례를 살핀다. 그중 마흔여덟 사례에서 목표가 실제로 오른쪽에 있다. 새 비율은 48/56이다. 이 숫자들은 사람을 조사한 결과가 아니라 선택한 실습용 확률 모형을 기술한다.

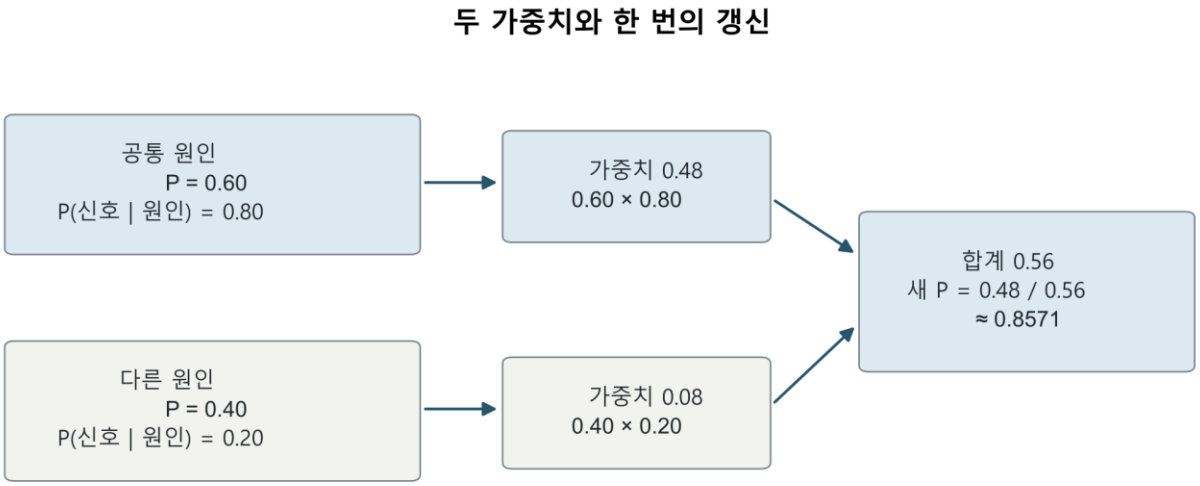


그림 24. 인과 설명의 순차적 갱신. 사전확률 0,60, 대칭적 신뢰도 0,80인 학습 계산이다. 먼저 정규화 전 가중치 0,48과 0,08을 만들고, 이어 정규화한다. 이 수치들은 모형을 정하며 체험을 측정한 값이 아니다.

임의의 두 가설에 대해서도 같은 계산을 쓸 수 있다. 먼저 각 가설의 사전확률에 얻어진 자료의 가능도를 곱한다. 그러면 정규화되지 않은 가중치 두 개를 얻는다. 이어 원하는 가중치를 두 가중치의 합으로 나눈다. 이 나눗셈으로 확률의 합이 일이 된다. 합이 영이라면 두 가설

모두에서 그 자료가 불가능하다는 뜻이다. 이때 필요한 것은 임의의 응답이 아니라 모형의 재검토나 입력의 검증이다.

오즈를 쓰면 같은 연산을 간결하게 수행할 수 있다. P에 대한 C의 오즈는 P(C)/P(P)다. 새 특징이 생기면 여기에 가능도비 P(특징|C)/P(특징|P)를 곱한다. 처음 오즈가 하나 대 하나이고 그 특징이 C에서 세 배 더 가능하다면 오즈는 셋 대 하나, 즉 C의 확률은 3/4가 된다.

다음 증거를 새로운 독립 표 한 개로 자동 간주해서는 안 된다. 센서 하나가 다른 센서를 복사했을 뿐이라면 수치가 같아도 근거가 두 배가 되지는 않는다. 순차적으로 곱할 때에는 이미 아는 증거를 고려한 새 특징의 가능도를 사용한다. 조건부 독립은 그 가능도가 바뀌지 않는 단순화다. 이제 수치 예를 구체적인 메커니즘으로 읽으면서 그 전제도 함께 볼 수 있다.

## 두 설명의 예제 분석

저자의 실습용 모형을 만들어 보자. 가설 C에서는 시각 신호와 신체 신호가 같은 대상에 속한다고 본다. 가설 P에서는 출처가 다르다고 본다. 새 관찰 전에는 두 가설에 같은 가중치를 준다. 이어 변화가 거의 동시에 일어났다는 특징을 얻는다. 실습에서는 이런 일치가 P보다 C에서 세 배 더 가능하다고 정한다.

이 특징을 반영하면 가중치의 비는 셋 대 하나가 된다. 전체 네 몫 가운데 셋이 C에 해당하므로 모형 안에서 그 확률은 75퍼센트다. 신체 소유감을 측정해 얻은 확률이나 실제 사람에 대한 평가가 아니다. 두 경쟁 설명의 갱신을 보여 주기 위해 정한 수치다.

이제 대상들의 위치가 크게 어긋난다는 둘째 특징을 얻는다. 이 특징이 P보다 C에서 나타날 가능성이 네 배 낮다고 정하자. 앞의 셋 대 하나에 사분의 일을 곱하면 0,75 대 하나가 된다. 이제 C의 확률은 약 43퍼센트다. 첫 특징을 계속 반영하면서도 새 정보로 모형의 선호가 달라졌다.

계산은 짧게 적을 수 있다. odds는 P의 확률에 대한 C의 확률의 비를 뜻한다. ratio는 다음 특징이 C에서 몇 배 더 가능하게 나타나는지를 뜻한다. 마지막 줄에서는 이 비를 전체 합에서 차지하는 비율로 바꾼다.

```
odds = 1.0
for ratio in [3.0, 0.25]:
    odds = odds * ratio
    probability_common = odds / (1.0 + odds)
    print(round(probability_common, 3))
```

계산 결과는 0,75와 약 0,429다. 이 단순화한 모형에서는 원인에 대한 가설을 고정했을 때 특징들이 독립이라고 가정한다. 시간적 일치와 공간적 일치가 서로 관련된다면 검증 없이 독립적인 증거처럼 곱할 수 없다. 정보 일부를 두 번 셀 수 있기 때문이다. 더 정확한 모형은 결합 의존성을 명시하거나 첫 특징을 고려해 둘째 특징을 평가해야 한다.

## 자신의 행동이 더하는 것

수동적으로 관찰한 신호의 일치와 명령의 결과는 서로 다른 검증 기회를 준다. 프로그램이 «왼쪽» 명령을 보내고 포인터가 왼쪽으로 움직이리라 예상한다고 하자. 예상과 실제 변화를 비교한다. 그러나 결과가 한 번 일치했다고 인과적 제어를 확립할 수는 없다. 외부 기록을 재생해도 우연히 같은 움직임이 나올 수 있다.

이제 프로그램이 새로운 순서로 명령을 바꾼다. 포인터가 계속 명령에 맞게 움직인다면 제어 가설이 더 뒷받침된다. 새 명령과 무관하게 이전 기록이 재생된다면 어긋남이 나타난다. 행동 순서에 개입하면 수동 관찰에서는 같아 보이던 원인들을 구별할 수 있다.

행동 모형은 적어도 명령과 예상 결과, 관찰한 결과를 보존해야 한다. 시간도 밝혀야 한다. 늦게 나타난 결과는 앞선 명령에 속할 수 있다. 현재 명령하고만 비교하면 계는 자신이 실패했거나 외부 개입이 있었다고 잘못 판단할 수 있다. 이력은 기억 용량을 늘리기 위해서만 필요한 것이 아니라 사건 사이의 관계를 알아내기 위해 필요하다.

그래도 «이 명령이 사건을 일으켰다»는 기능적인 평가와 체험되는 행위주체감은 구별된다. 인간을 연구할 때에는 모형의 평가와 보고를 비교할 수 있다. 인공적인 계에서는 먼저 출처 자체를 구별하는지, 그 구별을 행동에 사용하는지 검증한다.

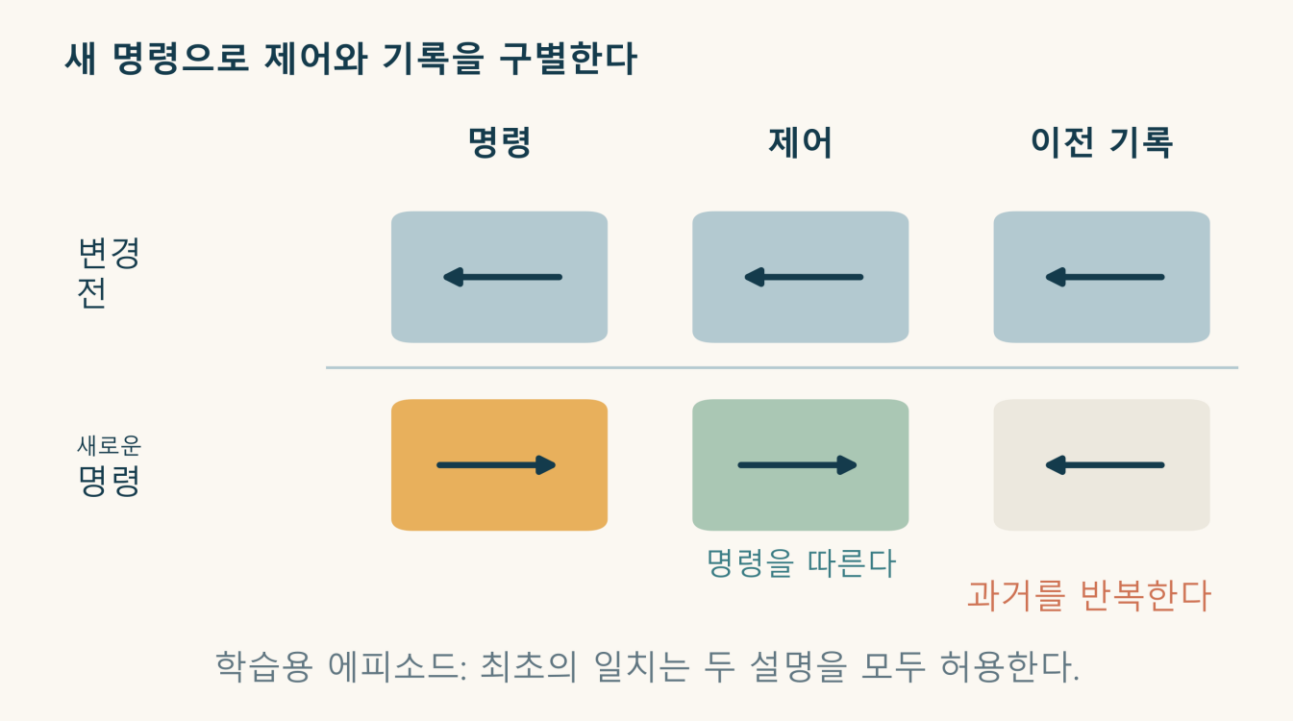


그림 25. 명령을 바꾸어 원인을 구별하기. 최초의 왼쪽 명령에서 포인터의 움직임은 제어와 외부 기록 재생 두 설명에 모두 맞을 수 있다. 가상으로 이어지는 상황에서 새 오른쪽 명령은 서로 다른 예측을 준다. 제어되는 포인터는 방향을 바꾸지만 기록은 이전의 왼쪽 움직임을 되풀이한다.

## 기억과 자기관련성 모형의 비교

프로그램 세 가지를 생각해 보자. 첫 프로그램은 현재 신호들이 일치하는지만 본다. 둘째 프로그램은 이전 신뢰도의 전체 이력도 고려하지만 출처를 나누지 않는다. 셋째 프로그램은 자기 명령과 외부 사건, 각 관찰 채널의 이력을 따로 보존한다. 이 모형들은 원인을 안정적으로 구별하려면 어떤 정보가 필요한지에 관한 서로 다른 질문에 답한다.

셋째 프로그램은 «자신의»라는 말만으로 이점을 얻지 않는다. 그 표지가 어떤 정보를 담고 예측에 어떻게 영향을 주는지 검증해야 한다. 정답을 알려 주었을 뿐이라면 성공은 아무것도 설명하지 못한다. 더 많은 자료를 저장한다면 비슷한 용량의 공통 기억 모형과 비교해야 한다. 그렇지 않으면 조직의 기여와 자원의 이점이 섞인다.

내용을 갖춘 가설은 이렇게 세울 수 있다. 출처별 이력이 있으면 지연과 잡음의 새로운 조합에서도 행위의 원인을 평가하는 데 도움이 된다. 이를 검증할 때는 조정 과정에서 접하지

않은 조합을 사용해야 한다. 그래야 익숙한 순서를 기억하는 능력을 넘어 가정한 관계를 적용하는 능력을 보여 준다.

### 지연이 일치의 의미를 바꾸는 방식

시점 1에 오른쪽 명령을, 시점 2에 왼쪽 명령을 보냈다. 오른쪽으로 움직이는 결과는 시점 3에야 나타난다. 현재 명령과 현재 결과만 비교하면 외부에서 생긴 결과라고 잘못 판단할 수 있다. 이력이 있으면 앞선 명령과 연결할 수 있다. 그러나 외부 기록도 같은 순서를 재생한다면 수동 관찰로는 부족하다. 미리 정한 새로운 명령열을 쓰면 결과가 현재 제어를 따르는지 이전 기록을 따르는지 검증할 수 있다. 이 통제는 인과 가설들을 구별하며 그 자체로 행위주체감을 측정하지는 않는다.

#### 여기까지 확인한 것

신호의 공통 원인과 인과적 제어, 신체 소유감을 하나의 지표로 합치지 않았다. 앞의 두 관계를 검증할 방법을 얻었으며, 그 평가가 인간의 보고와 어떻게 관련되는지는 따로 물을 수 있다.

## 실습 11 종이 위에서 출처 연구하기

실습 목표는 모형에 실제 원인을 알려 주지 않으면서 사건의 출처를 투명하게 분석하는 것이다.

관찰과 숨은 상태, 내부 평가의 개념을 활용한다. 이 장의 확률적 갱신 설명에 필요한 계산이 들어 있다. 읽기 전에 베이즈 규칙을 직접 유도할 필요는 없다.

예상 소요 시간은 30–40분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 단계별 명령을 오른쪽, 왼쪽, 왼쪽, 오른쪽으로 정한다. 제어 가설에서는 두 단계 뒤에 결과가 나타난다. 외부 기록에는 명령과 무관하게 오른쪽, 오른쪽, 왼쪽, 왼쪽이 같은 순서로 반복되는 주기적인 열을 사용한다.
2. 관찰 표 두 개를 만들고 실제 원인은 따로 보관한다. 처음 관찰 가능한 사건들에 대해 지연이 영인 가설과 두 단계인 가설을 적는다.

3. 명령을 왼쪽, 오른쪽이라는 새 열로 이어 간다. 새로운 결과를 보기 전에 제어 가설과 외부 기록 가설의 예측을 적는다.

4. 자료가 아직 부족한 경우를 표시하고 남은 가설들을 실제로 구별할 수 있는 새 명령을 설명한다.

명령과 결과의 표, 두 인과 가설, 서로를 구별할 후속 상황의 사전 예측을 남겨 둔다.

힌트: 지연된 채널의 처음 두 단계에는 관찰 시간 창 안의 명령이 아직 대응하지 않을 수 있다. 가상의 연결로 채우지 마라.

탐구를 이어 가려면 잡음과 불확실한 출처 인식을 추가한다. 같은 입력과 비교할 만한 용량에서 공통 기억과 출처별 이력을 비교한다.

해설과 답은 부록 F 제11장 참조, 301쪽

**자기 점검 문항**

1. 공통 원인에 대한 추론이 신체 소유감과 같지 않은 까닭은 무엇인가?
2. 자기 제어와 외부 기록 재생에 따른 우연한 일치를 어떻게 구별할 수 있는가?
3. 보존한 자료가 더 많다는 사실만으로 자기관련성의 이점을 입증할 수 없는 까닭은 무엇인가?

이력은 원인을 구별하는 데 도움이 되었다. 다음 장에서는 자기 응답이 얼마나 믿을 만한지, 결과를 얻기 전에 이를 어떻게 평가하는지 묻는다.

제12장

# 생각의 평가에서 기억과 시간

계는 이제 세계와 행동에 관한 정보를 보존할 수 있다. 다음에는 자기 성공을 예측하는 과제를 준다. 이때 특히 중요한 것은 예측을 결과보다 먼저 기록하는 순서다. 단순한 계수기 모형으로 과거에서 배우는 방법을 보며, 고정 확률 모형과 비교해 해당 입력열에서 학습이 유용한지 검증한다. 이 분석은 이후 점검과 전략 수정의 토대가 된다.

## 현재 순간은 이력에 달려 있다

문장을 끝까지 읽을 때 처음 단어들은 이미 주의의 중심에서 벗어났어도 끝부분의 이해에 참여할 수 있다. 이를 설명하려면 모형이 사건 사이에 정보를 유지해야 한다. 제9장에서는 변하는 수 하나로 충분했다. 제11장에서는 명령과 외부 출처의 이력을 구별해야 했다. 이제 계가 자기 과제를 얼마나 잘 수행하는지에 대한 평가를 더한다.

세 시간 규모를 나누어 생각하면 좋다. 현재 상태는 바로 다음 응답을 결정한다. 보존한 이력은 현재 사례와 이전 사례를 비교할 수 있게 한다. 매개변수는 처리의 지속적인 특성을 정하고 학습으로 달라질 수 있다. 이 셋을 반드시 별개의 상자 세 개로 구현할 필요는 없다. 어떤 변화가 행동을 설명하며 그 영향이 얼마나 오래 남는지 이해하려는 구별이다.

기억이 있다고 체험의 연속성이라고 곧바로 부를 수는 없다. 파일은 아무 동작도 하지 않으면서 정보를 오래 보관할 수 있다. 기능을 연구하려면 보존한 내용이 사용되는지, 결정을 어떻게 바꾸는지, 손상되면 무슨 일이 일어나는지 검증해야 한다. 그 과정과 체험되는 시간의 관계는 이후 이론에서 물을 문제다.

## 예측과 이미 주어진 응답의 평가

과제를 받기 전에는 주제에 익숙하므로 잘 풀 것이라고 예상할 수 있다. 구체적인 조건을 읽으면 확신도가 달라진다. 응답을 낸 뒤에는 풀이를 다시 평가할 수 있으며, 채점한 뒤에는

정답 여부를 알 수 있다. 이 네 시점에는 정보가 서로 다르다. 시각 표시 없이 하나의 «확신도» 열에 섞어 넣어서는 안 된다.

모형에서는 미래의 단서를 배제하는 일이 특히 중요하다. 정답 여부를 이미 알면 평가하기 쉽다. 연구할 만한 것은 되먹임 전의 예측, 또는 외부에서 정답을 받기 전에 계가 이용할 수 있는 내부 특징에 근거한 응답 평가다. 그러므로 기록 순서는 모형의 일부다.

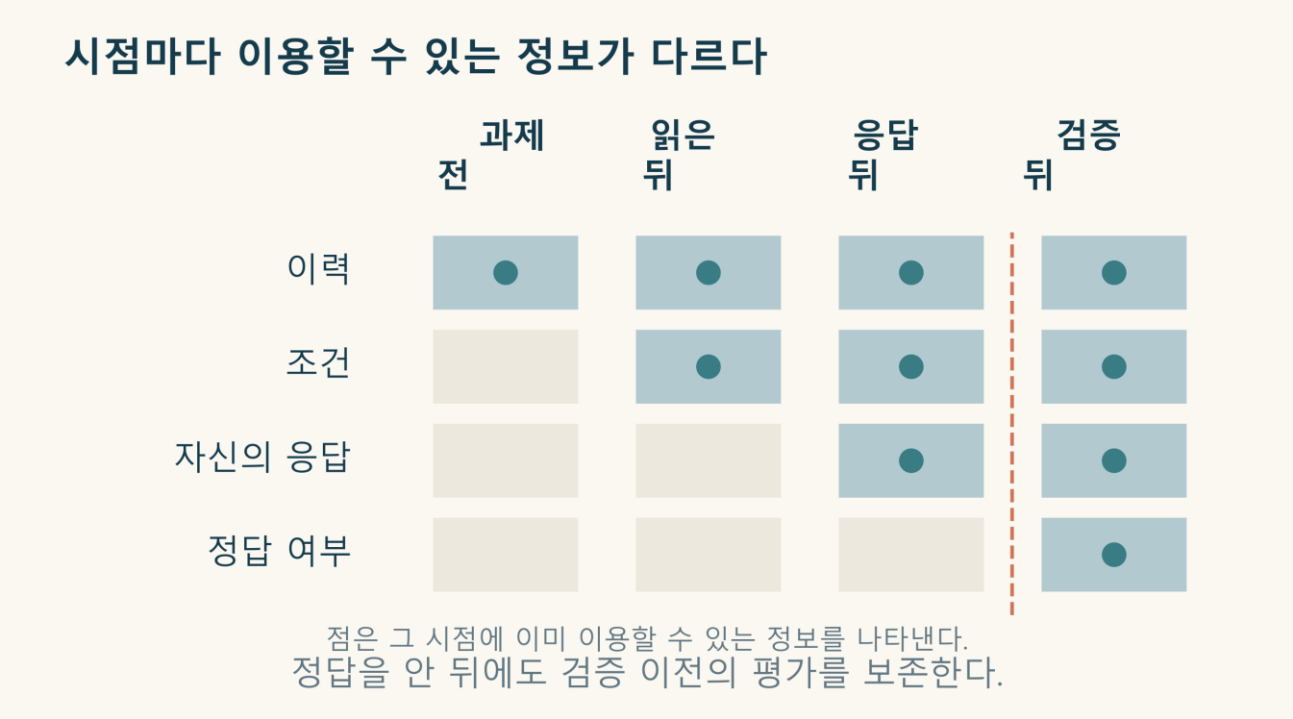


그림 26. 되먹임 전후의 정보. 이전 결과의 이력, 현재 과제의 조건, 자신의 응답, 정답 여부를 차례로 이용할 수 있게 된다. 시점들을 구분하면 되먹임 전의 예측과 정답을 이미 아는 상태에서의 평가가 달라진다.

Fleming과 Lau는 메타인지의 서로 다른 지표가 확신도와 성공률 사이의 관계에 관한 서로 다른 질문에 답한다고 강조한다. [6] 여기서는 좁은 과제로 시작한다. 과거 결과로 미래 응답의 신뢰성을 추정하는 법을 배울 것이다. 이는 아직 체험되는 확신도의 완전한 모형도, 메타인지 효율에 대한 별도의 평가도 아니다.

### 학습하는 평가의 예제 분석

계가 과제를 연이어 풀고 각 과제 뒤에 정답 여부를 받는다고 하자. 1은 성공, 0은 오류다. 성공 응답과 오류 응답을 세는 계수기 두 개를 도입하고 처음에는 각각 일을 넣는다. 이는 실습에서

정한 평활화다. 실제 자료를 얻기 전에 성공이나 실패를 완전히 확신하는 상태로 시작하지 않게 한다.

다음 과제에 앞서 예측값은 성공 횟수를 두 계수기의 합으로 나눈 비율이다. 처음에는 이분의 일이다. 첫 성공 후에는 계수기가 둘과 하나가 되므로 다음 예측은 삼분의 이다. 두 번째 성공 후에는 사분의 삼이 된다. 그다음 오류가 나면 성공 횟수는 셋으로 남고 전체 횟수는 다섯이 된다. 새로운 예측은 오분의 삼이다.

이 장을 위해 만든 학습 결과 여섯 개는 성공, 성공, 오류, 성공, 오류, 오류다. 아래 프로그램은 먼저 예측을 계산하고 보존한다. 이어 다음 결과를 받고, 그 뒤에야 계수기를 갱신한다. 변수 loss는 예측 오차를 누적한다. 확률에서 실제 결과를 빼고 그 차이를 제곱한다.

```
outcomes = [1, 1, 0, 1, 0, 0]
successes = 1.0
failures = 1.0
loss = 0.0
for outcome in outcomes:
    probability = successes / (successes + failures)
    loss += (probability - outcome) ** 2
    print(round(probability, 3), outcome)
    successes += outcome
    failures += 1 - outcome
print("Средняя ошибка", round(loss / len(outcomes), 3))
```

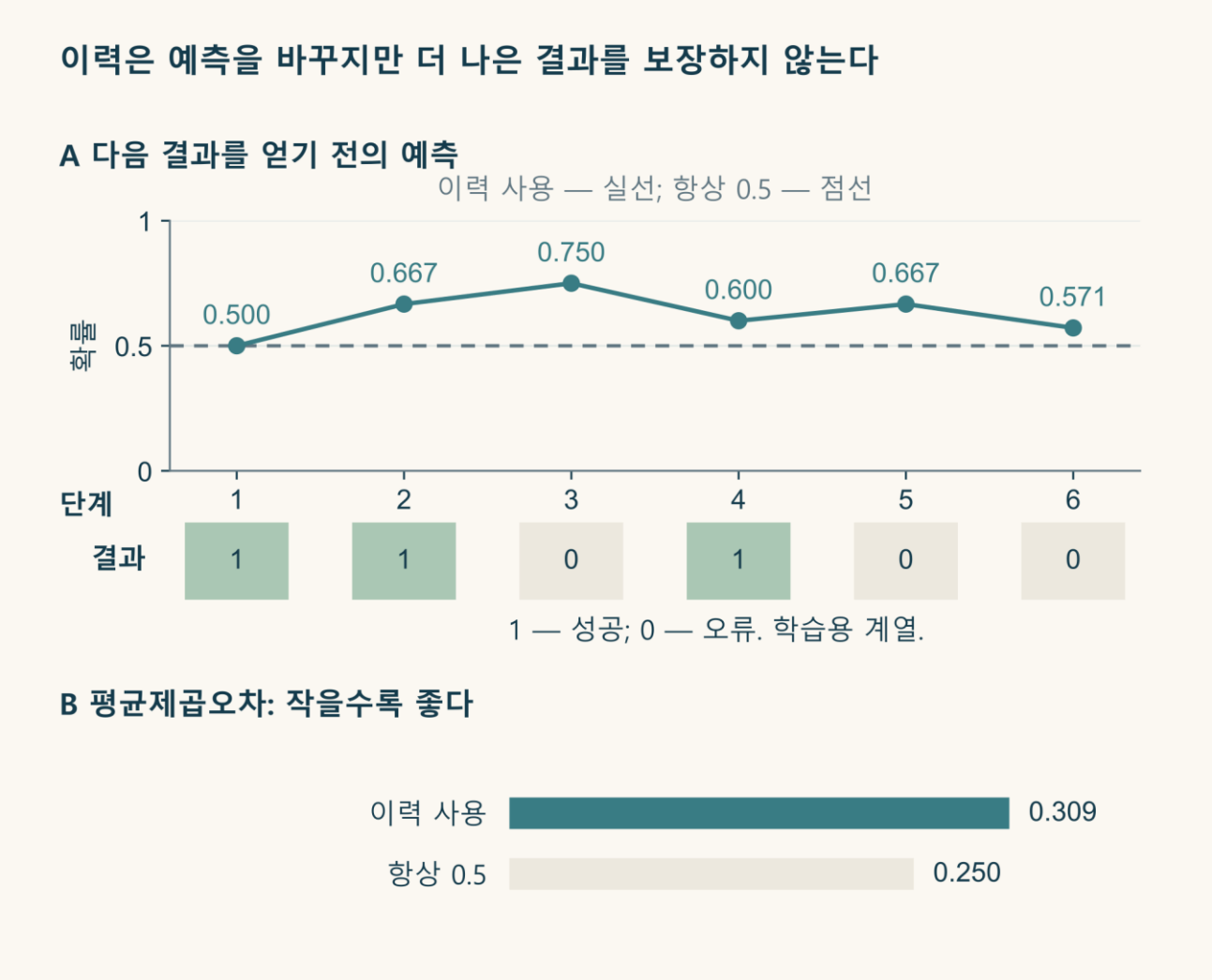


그림 27. 두 모형의 예측과 오차. 학습용 여섯 결과 1, 1, 0, 1, 0, 0에 앞서 평활화한 계수기를 쓰는 모형은 0,500; 0,667; 0,750; 0,600; 0,667; 0,571을 예측한다. 점선은 고정된 예측값 0,5를 나타낸다. 이 계열에서 이력을 가진 모형의 평균제곱오차는 약 0,309이고, 고정 모형은 0,250이다. 이는 특정 계열의 비교이며 보편적인 우위를 뜻하지 않는다.

과제 전의 예측은 차례로 약 0,500; 0,667; 0,750; 0,600; 0,667; 0,571이다. 전체 열을 마치면 두 계수기가 다시 같아져 다음 과제의 예측은 0,5가 된다. 변화는 메커니즘으로 설명된다. 성공은 비율의 한쪽을, 오류는 다른 쪽을 늘린다.

## 단순한 모형과 반드시 비교해야 하는 이유

항상 0,5의 확률을 예측하는 경쟁 모형을 생각해 보자. 이진 결과가 무엇이든 그 제곱 오차는 0,25다. 이 짧은 열에서 학습 모형의 평균 오차는 약 0,309다. 따라서 실습 자료에서는 정한

평가 규칙상 고정 예측이 더 나았다. 결과를 감출 필요는 없다. 이력에 따른 갱신이 언제나 유용한 것은 아님을 보여 준다.

처음 성공들로 모형의 기대가 높아진 뒤 오류들이 나타났다. 확신에 찬 오예측에 큰 벌점이 주어졌다. 결과 여섯 개만으로 고정 모형이 전반적으로 우월하다고 할 수는 없다. 규칙을 비교하려면 성공률이 일정하거나 난이도가 달라지거나 과제 유형마다 차이가 있는 등 조건을 미리 정한 새 열이 필요하다.

제곱 오차는 확률 예측의 품질을 평가한다. 체험의 깊이를 측정하지 않으며, 확신도가 개별 정답과 오답을 얼마나 잘 가려내는지도 그 자체로 알려 주지는 않는다. 지표 이름은 실제로 한 계산에 맞아야 한다. 이 규칙을 지키면 모형이 복잡해져도 의미를 유지할 수 있다.

## 전반적 신뢰도와 개별 사례의 어려움

이 계수기들은 과제들을 한 유형의 사례로 다룬다. 그러나 실제 과제열에는 쉬운 문제와 어려운 문제가 섞일 수 있다. 매번 같은 전체 예측을 적용하면 유용한 정보를 잃는다. 유형을 미리 알면 유형별 계수기를 둘 수 있다. 풀이 내부의 모호함으로 난이도를 평가한다면 어떤 특징을 사용하는지 명시해야 한다.

예를 들어 분류기가 두 후보에 비슷한 점수를 줄 수 있다. 점수 차이가 작다는 것은 불확실성의 한 단서일 수 있지만 실제 정답과 비교해 의미를 검증해야 한다. 내부 수치가 높다고 이름만으로 좋은 확신도가 되는 것은 아니다. 새 사례에서 그 특징과 실제 성공률의 관계를 보여 주어야 한다.

이렇게 «나는 보통 이런 과제를 푼다»는 전체 예측과 «이 응답은 믿을 만하다»는 개별 평가를 구별한다. 앞의 예측은 사례 유형의 이력을 사용하고 뒤의 평가는 현재 풀이의 정보를 고려할 수 있다. 좋은 연구는 둘을 비교해 각 정보 출처가 무엇을 더하는지 알아본다.

## 과거가 더 이상 똑같이 유용하지 않게 될 때

과제가 점차 어려워진다고 하자. 쉬운 과제의 긴 이력 때문에 예측이 지나치게 높게 유지될 수 있다. 새 결과를 더하기 전에 이전 결과의 가중치를 낮추는 것이 한 적응 방법이다. 서로 다른

작동 국면을 보존하고 현재 어느 국면인지 알아내려 할 수도 있다. 두 방법 모두 환경의 변화 방식에 관한 가정이 필요하다.

빨리 잊으면 변화에 대응하기 쉽지만 우연히 오류가 이어져도 평가가 민감하게 흔들린다. 느린 갱신은 짧은 잡음에 안정적이지만 실제 변화 뒤에는 늦게 따라갈 수 있다. 여기에는 검증할 수 있는 절충이 있다. 계가 «유연해지고 싶다»는 이유만으로 보편적으로 가장 좋은 속도를 정할 수는 없다. 사용할 과제를 대표하는 조건들이 필요하다.

계수기의 자동 갱신은 평가를 조절하는 메커니즘이다. 명시적 성찰에서는 그 규칙 자체의 근거를 살피는 단계를 더한다. 가령 «과제 유형은 달라졌지만 내 확신도는 전체 성공률에 근거하고 있다»고 생각할 수 있다. 이어 새 평가 구조를 제안하고 이전 구조와 비교한다. 계를 설계할 때에는 이 기능들을 따로 기술하는 편이 좋다.

## 확률 예측은 무엇 때문에 벌점을 받는가

실습열의 셋째 결과 직전에는 두 번의 성공을 본 모형이 0,75를 예측한다. 셋째 결과는 영이므로 차이의 제곱은 0,5625다. 같은 시행에서 고정 예측 0,5의 오차는 0,25다. 학습 자체에 대한 벌점이 아니다. 이 규칙은 확신에 찬 오예측에 더 큰 벌점을 준다. 셋째 결과를 먼저 계수기에 넣고 새 값을 셋째 결과의 예측이라고 부르면 평가에 정답을 사용하게 된다. 그런 순서는 다른 성질을 검증하며 자료 누출을 일으킨다.

### 여기까지 확인한 것

예측한 시점과 평가 기준을 정했다. 이력은 도움이 될 수도 방해가 될 수도 있다. 이점은 비교의 결과로 확인되며 메커니즘의 이름만으로 보장되지 않는다.

## 실습 12 후속 결과로 검증하기

실습 목표는 검증할 결과에 맞추어 규칙을 고르지 않으면서, 후속 열에서 고정 예측과 학습 예측을 비교하는 것이다.

성공 비율과 현재 정보, 미래 정보의 구별을 활용한다. 아래에서는 제곱 오차를 미리 정한 예측 평가 규칙으로 사용한다.

예상 소요 시간은 25–35분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 성공과 오류의 계수기를 각각 일에서 시작한다. 결과열 1, 1, 0, 1, 0, 0에서 갱신 전에 매번 확률을 적는다.
2. 예측마다 현재 결과와의 차이를 제곱한다. 그 평균을 고정 예측 0,5의 평균 오차와 비교한다.
3. 마지막 계수기 값을 유지한 채 1, 0, 1, 1, 0, 1을 이어 붙인다. 두 번째 부분은 따로 비교한다.
4. 망각을 연구하고 싶다면 다음 변화를 보기 전에 망각 규칙을 고른다. 후속 열에 맞추어 고른 매개변수는 다른 열에서 검증한다.

시점별 예측 표와 두 차례의 별도 비교, 계산 순서의 근거를 남겨 둔다.

첫 열을 마치면 두 계수기는 각각 사다. 다음 결과의 예측은 다시 0,5다.

탐구를 이어 가려면 미리 알려진 유형별로 과제를 나눈다. 공통 계수기와 유형별 계수기를 비교한 뒤 어떤 자료로 작동 국면의 변화를 알아차릴 수 있는지 밝힌다.

해설과 답은 부록 F 제12장 참조, 302쪽

**자기 점검 문항**

1. 정답을 얻은 뒤의 예측을 처음 확신도로 평가할 수 없는 까닭은 무엇인가?
2. 누적 성공 비율만으로 개별 과제 하나하나를 평가하기에 부족한 까닭은 무엇인가?
3. 계수기의 자동 갱신은 평가 방식에 대한 명시적 성찰과 어떻게 다른가?

이제 상태와 되먹임, 출처, 평가를 기술하는 언어를 갖추었다. 이를 사용하면 의식 이론을 명칭의 목록이 아닌 구조와 예측에 관한 제안으로 읽을 수 있다.

제4부

# 이론과 연구의 최전선

제13장

# 작업공간과 재귀 처리

앞 장들에서는 상태와 되먹임, 내용의 사용, 응답 평가를 따로 살폈다. 이제 이 개념들로 감각 처리 내부에서 내용을 구체화하는 일과 결과를 여러 이용자에게 전파하는 일이라는 두 이론적 생각을 비교할 수 있다. 먼저 모든 규칙을 정한 장치에서 그 기능을 추적한다. 이어 의식적 접근과 체험에 관한 이론들의 더 강한 주장으로 돌아간다. 기능이 겹친다고 두 이론이 같아지는 것은 아니다.

## 이름에서 구조로

메시지를 읽는데 옆에서 잔이 떨어진다고 하자. 소리 때문에 읽기를 멈추고 몸을 돌려 위험을 평가한 뒤 행동을 바꾼다. 이 사례를 설명할 때 서로 다른 질문을 섞기 쉽다. 소리는 어떻게 인식되었는가? 왜 그 소리가 우선되었는가? 정보는 어떻게 움직임과 기억, 추론에 이용 가능해졌는가? 일어난 일이 왜 체험되었는가? 이 장의 이론들은 질문들을 연결하는 서로 다른 방식을 제안한다. 비교하려면 제안된 메커니즘을 먼저 이해하고 메커니즘에서 의식으로 넘어가는 논거를 따로 평가해야 한다.

작업공간은 전문화된 과정들이 교환을 수행하는 조직으로 생각하면 편리하다. 재귀 처리는 처리 결과가 되돌아와 이미 참여한 요소들의 이후 작동을 바꾼다는 뜻이다. 이는 구조의 서로 다른 특성, 즉 이용 가능성의 범위와 되먹임 상호작용의 유무를 가리킨다. 둘은 함께 나타날 수 있다. 따라서 접근법 사이의 논쟁을 «연결이 있는» 뇌와 «연결이 없는» 뇌 중 하나를 고르는 문제로 환원할 수는 없다.

## 작업공간 이론의 강한 주장

Stanislas Dehaene, Michel Kerszberg, Jean Pierre Changeux의 전역 신경 작업공간 모형에서는 전문화된 과정들이 선택한 정보를 유지하고 전파할 수 있는 분산 시스템과 상호작용한다. 1998년의 컴퓨터 구현은 이 시스템의 참여를 과제 학습, 노력이 드는 수행, 오류 처리와 연결했다. 이는 인지 작업의 조직에 관한 구체적인 제안이지, 뇌 안의 별도 «화면»을 발견한 것이 아니다. [16]

우리의 학습 예를 살펴보자. 장치에는 대상 인식, 위험 평가, 움직임 선택, 사건 기억이라는 네 모듈이 있다. 인식기가 잔을 보는 동안에는 그 인식기만 관련 기록을 보존한다. 선택 뒤 «잔이 떨어진다»는 메시지는 공통 사건 식별자를 받아 다른 모듈로 전달된다. 위험 평가는 이를 가능한 화상과 연결하고, 움직임은 물러서기를 선택하고, 기억은 사건을 보존한다. 이 구조는 여러 과정이 같은 내용을 일관되게 이용하는 방식을 설명한다.

검증할 주장을 구체화해 보자. 모듈 사이에서 새로운 정보를 옮기는 데 공통 채널이 필요하다면 일시적으로 차단했을 때 바로 그 전달이 손상되어야 한다. 잔을 인식하는 기능은 남고 인식한 내용을 새 과제에 사용하는 기능은 사라질 수 있다. 반면 차단이 감각 입력과 움직임까지 동시에 손상한다면 결과는 모호해진다. 어떤 손상된 과정이 행동 변화를 일으켰는지 알 수 없기 때문이다.

의식적 접근에 대한 강한 설명은 내용이 안정적으로 넓게 이용 가능해지는 전환에 특별한 역할을 부여한다. 연구자는 어느 정도의 전파를 충분히 넓다고 보는지, 어떤 과정이 내용을 유지하는지, 어떤 자료로 이 전환과 단순한 신호 증폭을 구별할지 밝혀야 한다. 이런 설명 없이 «정보가 전역적으로 퍼졌다»고 말하면 결과에 이름을 붙이는 데 그친다.

## 재귀 처리 이론의 강한 주장

빅토르 람머의 접근법은 시각 신호의 초기 순방향 전파와 이후의 되먹임 상호작용을 구분한다. 국소적 재귀 처리를 시각적 체험 메커니즘의 후보로 보고, 더 넓은 상호작용을 접근과 보고에 연결한다. 강한 주장은 이후의 모든 활용 가능성을 아직 갖추지 못한 체험도 가능하다는 것이다. 이는 별도로 검증해야 하는 이론적 해석이다. [17]

다른 실습용 메커니즘을 만들어 보자. 첫 모듈은 짧은 선들을 가려내고 둘째 모듈은 대상의 윤곽을 추정한다. 추정이 첫 모듈로 돌아가 모호한 선들의 가중치를 바꾼다. 몇 차례 교환한 뒤 계는 일관된 형태를 고른다. 이 되먹임으로 부분의 인식이 전체에 의존하는 방식을 이해할 수 있다. 그러나 순환이 있다는 것만으로 처리와 함께 형태가 주관적으로 보여야 하는 이유를 설명하지는 못한다. 온도 조절기가 달린 선풍기도 제어 순환을 이루지만 그 구조는 여기서 말한 시각적 상호작용과 크게 다르다.

따라서 재귀 가설에는 한계를 명시해야 한다. 어떤 요소들 사이에서 교환이 일어나는가? 무엇이 되돌아가는가? 상호작용은 얼마나 지속되는가? 내용에 관한 되먹임과 흥분성의 전반적 조절을 어떻게 구별하는가? 이를 교란해도 어떤 연산은 유지되어야 하는가? 답이 상세할수록 발견한 어떤 순환이든 확인 증거라고 선언할 여지가 줄어든다.

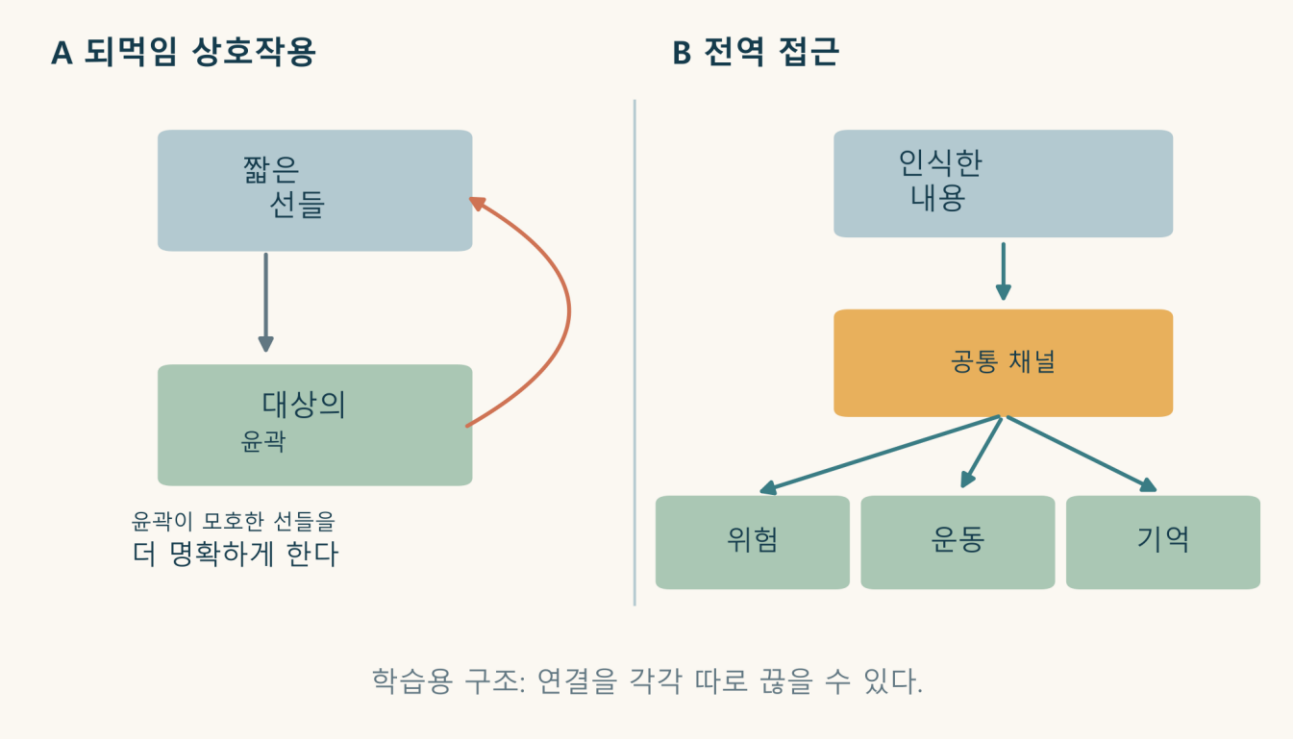


그림 28. 재귀 처리와 전역 접근. 한 학습용 장치의 두 측면이다. 윤곽에서 특징으로 돌아오는 연결은 인식의 일관성을 높이고, 공통 채널은 여러 과정이 선택한 내용을 이용하게 한다. 이 연결들은 따로 바꿀 수 있다. 도식은 의식의 존재를 검사하는 수단이 아니다.

파런포르트, 스홀터, 람머의 연구에서는 시각 차폐에도 초기 처리는 보존되고 재귀 신호로 해석된 후기 신호는 교란되었다. 이 연구는 처리 단계의 구분을 뒷받침했다. 모든 후기 반응이

체험의 충분조건임을 확립한 것은 아니며, 의식적 지각에 필요한 모든 추가 조건을 배제하지도 않았다. [18]

## 도식에서 이론이 충족할 요구로

한 사례를 두 번 살펴보면 좋다. 먼저 처리를 설명한다. 국소 특징이 몇 가지 가능한 대상을 제시하고 맥락이 선택을 좁히며 결과가 다음 행동에 이용 가능해진다. 이어 각 단계가 의식과 어떤 관계인지 묻는다. 첫 분석은 메커니즘을 정하고 둘째 분석은 이론의 어떤 추가 명제가 그 단계와 체험을 연결하는지 보여 준다.

작업공간 접근법에서 중요한 것은 수치를 임의로 복사하는 일이 아니라 내용을 특정한 방식으로 선택하고 유지하며 널리 이용하는 일이다. 이용자와 새로운 이용 방식, 접근을 선택적으로 제한했을 때의 결과를 명시해야 한다. 그러므로 미리 설정한 출력 하나로 내용을 옮기는 단순한 전달은 이론 전체가 아니라 준비 사례다.

재귀 처리 접근법에서는 어떤 되먹임 상호작용이 시각적 체험에 충분하다고 가정하는지, 그것이 이후의 보고와 어떻게 다른지 구체화해야 한다. 유한한 반복 계산의 계열은 순방향 회로로 펼칠 수 있다. 이론이 물리적 조직을 중시한다면 계산 기능과 물리적 조직을 구분해야 하는 이유다.

좋은 비교에서는 예측이 갈라지는 지점 하나를 고른다. 국소 내용이 보존되어 있지만 자발적 응답에는 이용할 수 없게 되었다고 하자. 그러면 남아 있는 과정을 어떤 근거로 체험이라고 볼지, 어떤 독립 지표가 이를 확인하는지가 논점이다. 과정을 의식적이거나 무의식적이라고 바꾸어 부르는 것만으로 논쟁을 해결할 수는 없다. 검증할 이론의 형태와 측정, 수정 조건을 함께 정해야 한다.

실습 직전의 계는 의도적으로 더 단순하다. 기능을 따로 차단해 보는 경험을 얻기 위한 것이다. 실습을 마치면 어떤 논거를 계산으로 검증했고 어떤 부분에 추가 자료가 필요한지 정확히 이해하면서 과학적 논쟁으로 돌아갈 수 있다.

## 실제 의견 차이가 있는 지점

시각계 안에 안정적으로 표상되지만 자발적인 응답에는 쓰이지 않는 내용을 생각해 보자. 한 연구자는 체험은 이미 있고 접근이 제한되었을 뿐이라고 말할 것이다. 다른 연구자는 의식적 접근에 앞선 처리가 발견되었다고 말할 것이다. 차이는 어떤 구조를 충분하다고 보는가, 채택한 이론과 독립적으로 찾는 현상의 존재를 어떻게 확인하는가에 있다.

이 어려움 때문에 연구가 무의미해지지는 않는다. 더 좁은 예측을 비교할 수 있다. 예를 들어 한 모형은 두 감각 영역 사이의 긴 교환을 요구하고 다른 모형은 특정 시기에 추가 계가 연결되어야 한다고 본다. 그러면 하나를 보존하면서 다른 상호작용만 바꾸는 실험이 유용하다. 개입이 선택적이지 않으면 행동 효과가 커도 모형들을 잘 구별하지 못한다.

실습 장치에서는 윤곽에서 선으로 돌아오는 연결과 사건을 전달하는 공통 채널을 따로 끌 수 있다. 첫 연결을 끄면 형태의 일관된 인식이 손상되고 둘째 채널을 끄면 여러 과제에서 그 형태를 사용하는 기능이 손상된다. 서로 다른 결과는 두 메커니즘의 기능이 다름을 보여 준다. 의식에 대한 주장으로 나아가려면 추가 논거가 필요하다. 지금 이 장치로 하는 것은 정보 처리 실험이다.

과정 하나가 다른 모든 과정을 대신하리라 미리 기대해서는 안 된다. 메커니즘은 어떤 조건에서 필요하더라도 학습 뒤에는 다른 메커니즘으로 우회될 수 있다. 검출 문턱은 신호의 세기에, 응답할 준비는 지시와 오류의 비용에 달려 있다. 따라서 결과 변화를 감도, 전략, 응답 시간, 주관적 평가의 변화로 나누어야 한다.

실습 계에 움직이는 사람에 관한 보고라는 둘째 출처를 더하자. 공통 채널은 두 내용 가운데 하나를 선택해야 한다. 잔을 언제나 우선하면 한 사례는 해결해도 유연한 선택을 보여 주지는 못한다. 과제를 바꾸어 때로는 사람을 기억하고 때로는 잔과의 충돌을 피하게 하자. 검증 전에 지시가 우선순위에 어떤 영향을 주는지 정한다. 이어 선택에서 밀린 내용의 인식이 남아 있는지, 그 내용이 나중에 다른 모듈에 접근할 수 있는지 살핀다.

이렇게 복잡하게 만들면 통신 경로의 존재를 넘어 그 사용 규칙을 검증할 수 있다. 고정된 명령열과 새 목표에 따라 서로 다른 내용을 전달할 수 있는 구조를 구별하는 데 도움이 된다. 두 장치 모두 처음 시행에서는 성공해도 낯선 요구 조합에 적용할 때 차이가 드러날 수 있다.

## 구체화와 전달을 포함한 한 시행

새 학습 시스템에서 국소 특징은 0, 1 또는 물음표다. 맥락은 0 또는 1을 담는다. 특징이 정해져 있으면 인식기는 그 값을 보존한다. 정해져 있지 않으면 처음에는 0을 기록한다. 맥락의 되먹임을 켜면 이 기록을 맥락 값으로 바꾼다. 공통 채널은 최종 내용을 기억에 복사한다. 전달이 없으면 기억은 0으로 남는다. 시행 (?, 1)에서 전체 시스템은 먼저 국소 값 0을 얻고, 이를 1로 구체화한 뒤 1을 기억에 전달한다. 되먹임을 차단하면 0으로 남는다. 전달을 차단하면 인식은 1로 유지되지만 기억은 0으로 남는다.

명확한 입력의 정답 표지는 그 입력값이고, 모호한 입력의 정답 표지는 맥락값이다. 정답 표지는 평가자에게 필요하다. 여기서는 미리 정한 규칙의 작동을 연구한다. 모든 입력을 직접 받는 경쟁 모형도 같은 과제를 풀 수 있다. 그러므로 시연의 성공만으로 작업공간의 유일성이나 되먹임이 체험의 충분조건이라는 주장을 확립할 수는 없다.

### 여기까지 확인한 것

서로 독립적으로 바꿀 수 있는 두 관계를 얻었다. 내용의 구체화와 기억으로의 전달이다. 이제 네 작동 조건을 두 결과로 비교할 수 있다. 한 결과의 상실을 의식의 소멸이라고 부를 필요는 없다.

**두 연결의 개별 기능**

| 방식 | 인식 | 결과 기억 |
|---|---|---|
| 전체 구조 | 10개 중 10개 | 10개 중 10개 |
| 되먹임 없음 | 10개 중 7개 | 10개 중 7개 |
| 공통 채널 없음 | 10개 중 10개 | 10개 중 4개 |
| 두 연결 모두 없음 | 10개 중 7개 | 10개 중 4개 |
| 직접 복사 | 10개 중 10개 | 10개 중 10개 |

분석한 예의 고정 입력 10개

두 구조의 출력이 같다면 다음 구별 실험이 필요하다.

그림 29. 두 메커니즘의 독립적 차단. 완전히 정해진 학습 시행 열 번의 결과이며 teaching_lab.py --chapter 13으로 재현한다. 되먹임은 모호한 특징을 구체화하고 공통 채널은 응답을 기억에 전달한다. 직접 복사는 강한 기능적 대조다. 이 표는 의식을 측정하지 않으며 과학 이론 전체를 비교한 것도 아니다.

## 실습 13 이론을 예측으로 옮기기

실습 목표는 같은 열 번의 시행에서 구체화와 전달의 기여를 따로 검증하는 것이다.

입력과 상태, 출력의 개념 및 국소 표상과 그 사용의 구별을 활용한다. 실습 앞에 계의 모든 규칙과 열 개의 입력이 제시되어 있다.

예상 소요 시간은 35–45분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 국소 특징과 맥락의 쌍으로 (0,0), (1,0), (0,1), (1,1), (1,0), (?,1), (?,0), (?,1), (?,0), (?,1)을 사용한다. 각각에 대해 앞의 규칙으로 정답 표지를 구한다.
2. 전체 기능을 켠 계에 대해 초기 인식, 최종 인식, 기억 내용의 열을 채운다. 매 시행에서 기억은 0으로 다시 시작한다.
3. 같은 열 개의 입력을 되먹임 없이, 전달 없이, 두 메커니즘 모두 없이 각각 반복한다. 인식과 기억의 정답 여부를 따로 센다.
4. python teaching_lab.py --chapter 13 명령으로 표를 검증한다. 이어 두 입력을 모두 보는 순방향 경쟁 모형을 추가하고 결론의 범위를 밝힌다.

네 작동 조건과 각각의 두 결과, 구체적인 어긋남 하나에 대한 설명을 남겨 둔다.

힌트: 전달을 차단해도 인식 규칙은 바뀌지 않는다. 두 지표가 모두 달라졌다면 여러 기능을 동시에 끄지 않았는지 확인하라.

탐구를 이어 가려면 지연 뒤 여러 이용자 가운데 하나에게 새 지시가 오는 과제로 바꾼다. 강한 경쟁 모형의 기억과 계산 예산, 입력을 미리 정한다. 메커니즘들이 실제로 갈라지는 결과를 도출한다.

해설과 답은 부록 F 제13장 참조, 302쪽

## 자기 점검 문항

1. 전역 전파와 재귀성이 서로 엄밀한 대립을 이루지 않는 이유는 무엇인가?
2. 개입 뒤 보고가 사라지면 체험도 사라졌다는 사실이 입증되는가?
3. 계산 예산을 같게 하여 메커니즘들을 비교해야 하는 까닭은 무엇인가?

기능의 차이가 명확해졌다. 다음 장은 이론의 두 대상을 더한다. 내부 인과 조직과 계가 자기 상태를 표상하는 방식이다.

제14장

# 통합정보와 고차 이론

작업공간과 재귀성은 처리를 조직하는 방식으로 비교할 수 있다. 하지만 의식 이론은 익숙한 외부 응답과 무관하게 어떤 구조가 중요한지, 상태가 시스템 자신에게 어떻게 표상되는지도 묻는다. 인과 구조와 고차 표상을 살펴보자. 우리의 과제는 각 이론이 처음부터 요구하는 바를 정확히 이해하는 것이다. IIT의 전체 형식 체계를 계산하는 일을 서로 연결된 몇 노드의 그림으로 슬쩍 바꾸어 놓지는 않을 것이다.

## 설명의 두 방향

빨간 원에 관한 질문에 똑같이 답하는 두 계를 생각해 보자. 첫 계에서는 많은 부분이 서로의 상태를 바꾼다. 둘째 계에는 허용된 모든 질문에 대한 응답 표가 미리 마련되어 있다. 질문이 제한되면 외부에서 둘을 구별하기 어렵다. 두 계의 내부에 있다고 보는 것도 같아야 할까? 답은 인과 조직과 자기 상태 표상, 행동 능력, 또는 그 조합 가운데 이론이 어떤 성질을 중요하게 보느냐에 달려 있다.

이 장에서는 두 방향을 살핀다. 통합정보이론은 체험에 공통적이라고 가정한 성질에서 출발해 그에 해당하는 물리적 조직을 찾는다. 고차 이론은 어떤 상태가 의식적이라는 것을 계가 그 상태를 어떻게 표상하는지와 연결한다. 두 연구 모두 편리한 행동 기술보다 더 강한 주장을 제안한다. 그러므로 수학 모형과 실험만큼 출발 가정도 주의 깊게 살펴야 한다.

## 통합정보이론이 주장하는 것

통합정보이론의 버전 4.0은 경험을 물리적 기질 안의 특정한 분해 불가능한 원인과 결과 구조에 연결한다. 저자들은 경험의 존재에서 출발하여 내재적 귀속, 확정성, 통일성, 구체성, 구조성을 구분한다. 이어서 물리적 공준과 형식 체계를 제안한다. 경험과 펼쳐진 인과 구조가 동일하다는 것이 이론의 핵심 설명적 주장이다. 수학적으로 적었다는 사실만으로 그것이 확립된 경험적 사실이 되지는 않는다. [19]

연결의 기록과 원인 검증의 차이를 우리의 예로 분석하자. 표시등 두 개가 매초 동시에 켜진다. 관찰자는 완벽한 일치를 알아차린다. 하지만 원인은 공통 타이머일 수 있으며 한 표시등은 다른 표시등에 영향을 주지 않는다. 반면 첫 표시등에 대한 개입이 둘째의 이후 상태를 바꾸고, 둘째에 대한 개입이 첫째를 바꾼다면 인과 조직이 다른 장치다. 관찰된 순서가 같아도 내부 의존 관계는 아직 정해지지 않는다.

이 사례가 통합정보를 계산한 것은 아니다. 계산에 앞서 어떤 상태와 개입을 물리적으로 가능하다고 볼지 묻게 한다. 상태가 둘인 요소 두 개로 이루어진 모형을 생각해 보자. 모든 입력 조합에서 각 요소의 변화 규칙을 정해야 한다. 일부 조합을 금지한다면 이유를 설명해야 한다. 두 요소가 실제로 하나의 공통 출처를 나타낸다면 따로 분석해서 인과관계를 왜곡할 수 있다.

다음에는 계의 경계를 묻는다. 모듈 하나를 연구할 수도, 상호작용하는 여러 모듈을 연구할 수도, 장치와 외부 제어기를 함께 연구할 수도 있다. 인과 분석의 결과는 달라질 수 있다. 그러므로 «여기에는 의식이 있어야 한다»는 원하는 답으로 경계 선택을 정당화할 수는 없다. 먼저 선택 규칙을 정하고 관심 계와 대조 계에 일관되게 적용한다.

이 접근의 강점은 구조를 통해 가능한 상태들의 통일성과 차이를 설명하라고 요구한다는 데 있다. 연구에서 어려운 부분은 인과 조직에서 체험의 존재로 넘어가는 단계다. 특성을 정확히 계산했어도 우선 확인한 것은 채택한 모형의 성질이다. 그 성질과 체험의 관계는 자료와 대안에 비추어 따로 검증해야 한다.

## 고차라는 말의 뜻

고차 이론은 세계를 표상하는 상태와 그 첫 상태의 존재나 성격을 표상하는 상태를 구분하자고 제안한다. 하크완 라우의 베이즈 모형에서 의식적 평가는 자신의 내부 신호와 그 신뢰성에 관한 결정에 달려 있다. 이는 여러 접근법 중 한 가지 구체적 구현이지 그 모든 변형의 정의는 아니다. 이 모형으로 과제 수행과 지각에 대한 주관적 평가 사이의 불일치를 연구할 수 있다. [20]

소리를 인식하는 실습용 모형을 만들어 보자. 첫 과정은 두 신호 가운데 하나를 고른다. 둘째 과정은 두 후보의 차이 크기, 여러 측정의 일관성, 최근 오류에 관한 정보처럼 첫 과정의 작동 특성을 받는다. 이를 통해 현재 선택이 얼마나 믿을 만한지 평가한다. 셋째 과정은 지금

응답할지 관찰을 한 번 더 요청할지 결정한다. 둘째 과정이 실제로 첫 과정의 정보를 사용한다면 이 구조는 확신도 조절을 설명한다.

중요한 차이를 더하자. «확신도는 팔십 퍼센트다»라는 변수는 감각 수치 하나를 단순히 다시 계산한 것일 수 있다. 그렇다면 자기 능력을 풍부하게 모형화했다는 증거는 아니다. 더 의미 있는 조절을 보여 주려면 개별 입력값은 유지하면서 센서의 신뢰도를 바꾸어 보면 좋다. 계가 오류 이력을 고려해 전략을 바꾼다면 그 평가는 현재의 신호 강도보다 많은 것을 반영한다.

확신도의 기능적 조절만으로 현상적 체험의 문제를 해결할 수는 없다. 강한 이론적 주장은 자기 상태의 어떤 표상이 왜 의식의 충분조건인지 밝혀야 한다. 과정에 대한 모든 수치 기술을 곧바로 의식함이라고 보는 것은 타당하지 않다. 또 의식적 지각과 자세한 성찰은 다르다. 색을 본다고 동시에 자기 색각의 신뢰도를 추론하는 것은 아니다.

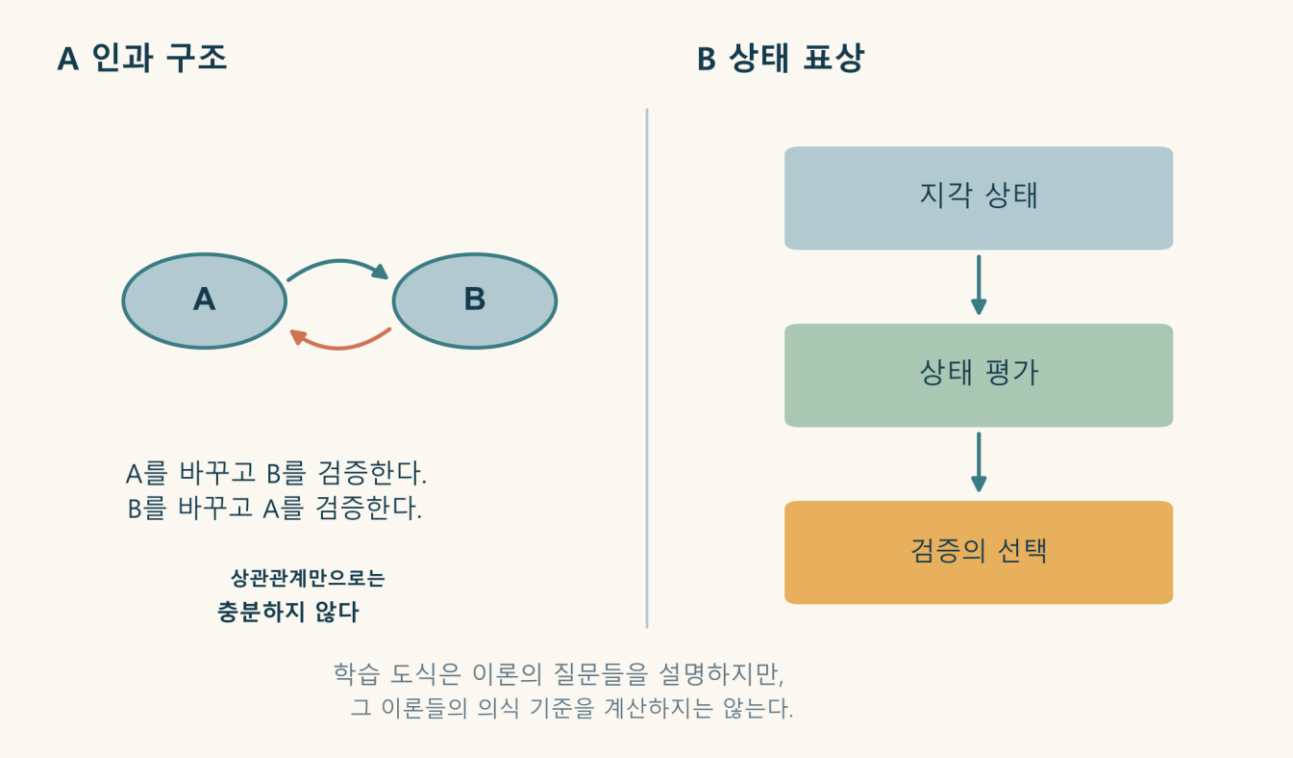


그림 30. 서로 다른 두 설명 대상. 왼쪽 도식은 부분들 사이의 상호 인과 영향에, 오른쪽은 시스템의 자기 상태 표상과 그 이용에 주목한다. 두 도식은 IIT와 고차 접근법의 서로 다른 질문을 설명한다. 하지만 그 이론들의 형식 체계를 대신하거나 승자를 정하지는 않는다.

## 구현에 관해 알아야 할 것

두 장치는 평소의 응답 표가 같아도 내부 개입에는 다르게 반응할 수 있다. 한 장치의 출력은 현재 상태들의 상호작용에 의존하지만 다른 장치에는 순서가 미리 기록되어 있을 수 있다. 외부 질문을 제한하면 둘은 일치할 수 있다. 그러므로 인과 구조를 분석하면 평범한 대화에는 없는 정보를 얻는다.

그러나 되먹임 화살표를 그리는 것만으로 통합정보를 계산한 것은 아니다. 특정 형식 체계에는 구체적인 요소와 상태, 전이, 분석할 계를 선택하는 규칙이 필요하다. 체험에 대응한다는 주장은 이론 자체에 속하며 그 논거와 함께 평가한다. 실습 그림은 이 구별을 유지하면서 질문의 대상을 설명한다.

고차 표상에는 다른 명세가 필요하다. 무엇이 일차 상태로 표상되는가? 둘째 과정은 그 상태에 관한 정보를 어떻게 얻는가? 현재 결정 자체의 정보와 성공에 대한 일반적인 기대는 무엇인가? 평가는 누가 사용하는가? 이 질문들로 실제 점검 기능을 외부 로그나 미리 준비한 문구와 구별할 수 있다.

끝으로 내부 접근이 시스템 자신에 대한 완전한 투명성을 보장해야 하는 것은 아니다. 한 과정이 어떤 신호에는 특별한 접근 권한을 가지면서도 다른 신호에 대해서는 틀릴 수 있다. 따라서 완전한 자기인식이라는 일반적 결론을 피하면서, 구체적 내용에 대한 접근, 그 내용을 얻는 시점, 인과적 사용을 검증할 것이다.

## 출발 가정들을 비교하는 방법

내부 인과 연결이 풍부하지만 자기 지각에 대한 별도 평가는 만들지 않는 가상 장치를 생각해 보자. 이어 다른 인과 조직으로 세밀한 확신도 조절을 구현한 장치를 상상해 보자. 이런 사례는 각 접근법이 경계를 분명히 하도록 요구한다. 한 설명은 주로 물리적 구조를 묻고 다른 설명은 표상들의 관계를 묻는다.

더 마음에 드는 사례를 고르는 것만으로 논쟁을 제대로 해결할 수는 없다. 비교하는 계의 의식 유무를 어떤 독립 근거로 아는지 밝혀야 한다. 그런 지식이 없다면 사고실험은 이론의 차이를

드러낼 뿐 승자를 정하지 못한다. 검증 가능한 귀결을 준비하고 숨은 가정을 드러내는 데 그 가치가 있다.

주장에는 세 종류가 있다. 정의는 용어의 사용법을 정한다. 형식적 정리는 모형의 규칙에서 무엇이 도출되는지 보여 준다. 경험적 주장은 연구하는 세계의 구조를 말한다. 엄밀한 정리로 출발 전제의 참을 얻을 수는 없다. 그래도 전제에 대한 논쟁이 형식적 분석의 가치를 없애지는 않는다. 그 분석은 채택한 설명이 정확히 무엇을 요구하는지 보여 준다.

자신의 연구에서는 가능한 한 좁은 결론을 선택한다. 모형이 추가 정보 요청을 정확히 예측했다면 행동 조절에 관한 결과다. 인과 특성이 두 망을 구별한다면 그 조직에 관한 결과다. 의식으로 나아가는 다음 단계에는 독자적인 근거가 필요하다. 이를 지표 이름 안에 슬쩍 넣을 수는 없다.

결론이 관찰자에 따라 달라지는지도 검증하면 좋다. 개발자가 상세한 내부 상태 로그를 추가했지만 장치 안의 어떤 과정도 그 로그를 읽지 않는다고 하자. 연구자가 계에 관해 더 많이 알게 되었어도 장치의 능력은 달라지지 않았다. 우리가 과정에 관해 얻는 정보와 그 과정이 제어에 사용하는 정보는 다르다. 로그를 제어 모듈의 입력으로 바꾸면 조직이 실제로 달라지므로 그 결과를 검증할 수 있다.

이 구별은 자기 기술에서 특히 중요하다. 외부 실험자가 일인칭으로 적은 정보가 장치의 내부 모형이 되지는 않는다. 기록을 누가 사용하며 어떤 행동이 그 기록에 의존하고 선택적으로 왜곡하면 무슨 일이 일어나는지 보여 주어야 한다. 그러면 자기 상태 표상에 관한 논쟁에도 구체적인 대상이 생긴다.

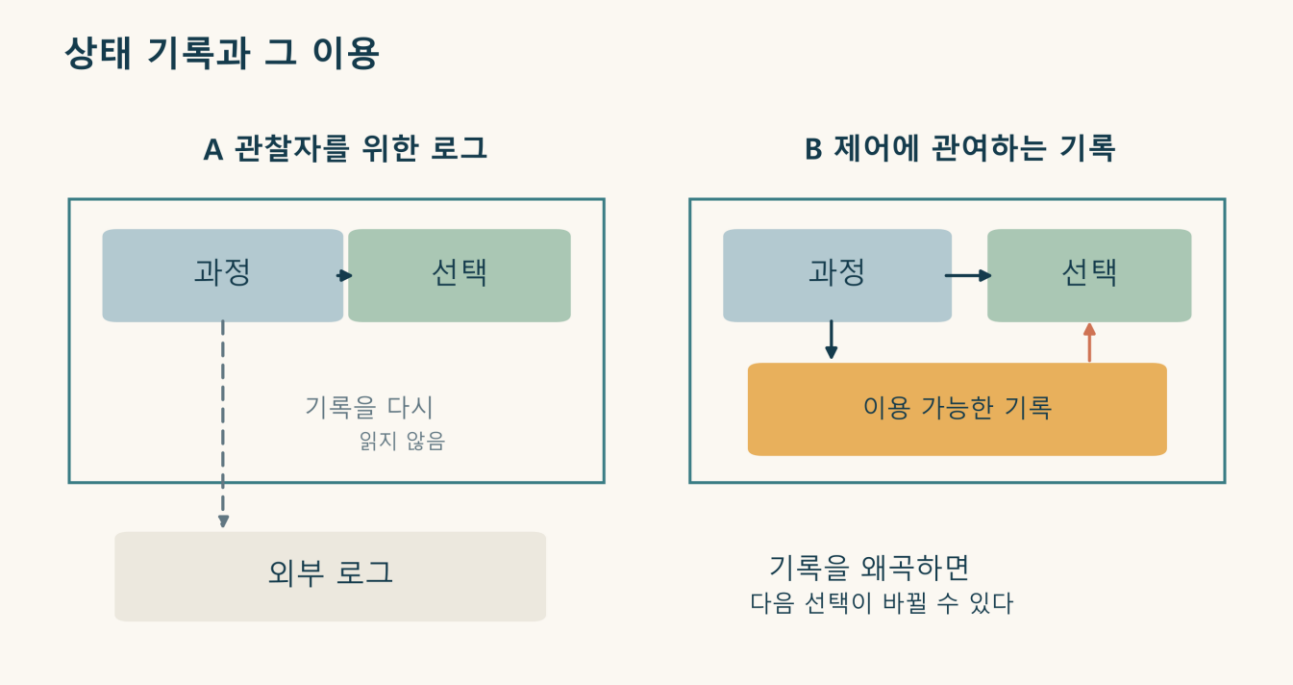


그림 31. 외부 로그와 내부 이용. 왼쪽 구조에서는 연구자가 과정에 관한 정보를 얻지만 장치는 로그를 읽지 않는다. 오른쪽에서는 기록이 제어 입력이 된다. 그 기록을 선택적으로 왜곡하면 인과적 역할을 검증할 수 있다. 이는 정보의 이용을 다루는 검증이며, 의식에 충분한 고차 표상을 아직 확립하지는 않는다.

## 정의와 강한 가정을 섞지 않는 방법

«자기 오류에 대한 예측을 내부 평가라고 부르자»는 말은 작업 정의를 정한다. «보고 과정이 이 평가를 읽는다면 평가를 바꾸어 보고가 달라질 수 있다»는 주장에는 구체적인 모형 연결이 필요하다. «적절한 내부 평가가 상태를 의식적으로 만든다»는 말은 이론적 설명을 더한다. 특정 표본에서 의존 관계를 관찰했다는 것은 자료에 관한 진술이다. 이 명제들이 한 논거에 참여해도 나란히 있다는 이유로 같은 증명력을 얻지는 않는다. IIT와 고차 이론을 비교할 때도 이를 유지해야 한다. 두 접근은 지표뿐 아니라 처음 설명하려는 대상도 다르다.

### 여기까지 확인한 것

일반적인 연결성과 확신의 수치는 더 정밀한 이론적 요구를 정식화하기 전까지 기능적 성질로 남는다. 이제 어디에서 장치의 기술이 끝나고 체험에 대한 가정이 시작되는지 명시적으로 적어 보자.

## 실습 14 가정의 명세서

실습 목표는 주장들의 지위를 구별하여 두 설명의 명세서를 작성하는 것이다.

메커니즘과 그 측정, 체험과의 관계에 대한 주장을 구별한다. 필요조건과 충분조건은 제5장에서 다루었다.

예상 소요 시간은 30–40분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 인과 계에서는 바꿀 수 있는 상태 둘과 각 영향의 방향을 명시한다. 평가 계에서는 일차 결정과 내부의 신뢰도 특징, 평가를 받는 과정을 밝힌다.
2. 각 접근에 대해 설명하려는 현상, 제안하는 구조, 그 구조가 중요하다는 근거, 수정 조건이라는 네 명제를 적는다.
3. 각 명제를 정의, 가정, 주어진 모형의 귀결, 경험적 주장 가운데 하나로 표시한다. 여러 지위가 섞인 명제는 나눈다.
4. 개입 하나와 같은 응답을 낼 수 있는 더 단순한 메커니즘을 고른다. 구별할 결과가 아직 없다면 그 사실을 결론으로 남긴다.

보존할 것: 명세서 두 개와 체험에 관한 추론의 한계를 명시한 인과 검증 하나.

상호 영향의 도식은 IIT 계산이 아니다. 오류 예측도 이름만으로 고차 표상이 되지는 않는다. 그 내용과 출처, 사용을 밝혀야 한다.

탐구를 이어 가려면 이 장의 출처에서 한 이론의 구체적인 버전을 고른다. 그 이론의 충분조건이 만든 실습 계 이외에 무엇을 더 요구하는지 밝히고 이론들의 결론이 갈라지는 사례를 제안한다.

해설과 답은 부록 F 제14장 참조, 303쪽

### 자기 점검 문항

1. 부분 사이의 높은 상관만으로 상호 인과 영향을 결론 내릴 수 없는 까닭은 무엇인가?
2. 성찰은 모든 고차 표상과 같은가?
3. 이론은 수학적으로 정확하면서 경험적으로 논쟁적일 수 있는가?

다음 장에서는 예측으로 숨은 원인과 자기 신뢰도를 계산하는 구체적인 방법을 본다. 앞서 논의한 관계 일부를 실행 가능하게 만드는 데 도움이 될 것이다.

제15장

# 예측과 자기 모형

자기 상태의 표상은 어떻게 형성되고 사용되는지 알 때 더 구체적인 내용을 갖는다. 예측 접근법은 숨은 원인에 대한 가설을 만들고 예상 신호와 실제 신호를 비교하여 평가를 갱신하자고 제안한다. 세계와 명령의 결과를 따로 살펴보자. 제어가 바뀐 뒤에는 둘을 섞는 문제가 특히 분명해진다. 계는 목표를 잘 보면서도 행동에서는 틀릴 수 있다.

## 관찰에 가정이 필요한 이유

문 뒤에서 짧은 소리가 들린다. 물건이 떨어졌을 수도, 잠금장치가 작동했을 수도, 스위치가 딸깍했을 수도 있다. 신호는 하나지만 가능한 원인은 여럿이다. 행동을 선택하려면 신호와 함께 문 뒤에 누가 있는지, 앞서 무슨 일이 있었는지, 방 안 소리가 얼마나 잘 들리는지 같은 맥락을 고려해야 한다. 이 사례는 불완전한 관찰로 숨은 원인을 평가하는 추론의 기본 과제를 보여 준다.

과학적 모형의 예측이 반드시 말로 표현한 미래의 예상을 뜻하지는 않는다. 세계의 특정 상태에서 어떤 신호가 나타나야 하는지에 관한 계산 가능한 가정일 수 있다. 예상과 실제로 얻은 것을 비교하면 가정을 바꾸는 데 도움이 된다. Rao와 Ballard의 모형은 이런 위계적 조직으로 시각 뉴런의 일부 성질을 설명했다. 상위 수준이 하위 수준의 활동을 예측하고 불일치가 표상의 갱신에 참여한다. 이는 처리 모형이며, 체험의 발생에 관한 독립적인 증거는 아니다. [21]

## 숨은 원인의 단순 모형

직접 계산할 수 있는 예를 보자. 목표는 왼쪽이나 오른쪽에 있다. 관찰 전에는 오른쪽일 확률을 육십 퍼센트로 본다. 센서는 열 번 가운데 여덟 번 방향을 맞힌다. «오른쪽» 신호가 들어왔다. 평가를 얼마나 바꿔야 할까?

같은 규칙으로 만들어진 백 상황을 생각해 보자. 육십 상황에서는 목표가 오른쪽에 있고 그중 약 마흔여덟 상황에서 센서가 오른쪽을 가리킬 것이다. 마흔 상황에서는 목표가 왼쪽에 있으며 그중 약 여덟 상황에서는 센서가 잘못해서 오른쪽을 가리킬 것이다. 오른쪽 신호 쉰여섯 건 가운데 마흔여덟 건에서 실제 목표가 오른쪽에 있다. 갱신한 확률은 약 팔십육 퍼센트다. 이는 정한 수치와 센서 신뢰도 가정에서 나온 결과이며 인간 지각을 실험으로 측정한 값이 아니다.

이 예는 관찰 전의 가정과 관찰의 진단력, 최종 평가를 구별해야 하는 이유를 보여 준다. 센서가 믿을 만하지 않으면 강한 신호도 믿음을 조금밖에 바꾸지 못할 수 있다. 관찰이 없다는 것과 목표가 없음을 관찰했다는 것은 다르다. 센서가 작동하지 않았다면 계는 이전 가설을 뒷받침하는 증거를 꾸며내지 말고 세계의 가능한 변화를 고려해 불확실성을 유지해야 한다.

이제 목표가 때로 움직이게 하자. 기억을 마지막 응답을 그대로 보관하는 것으로만 이해할 수는 없다. 새 관찰 전에도 이전 믿음을 바꾸어야 한다. 관찰 사이 간격이 길고 목표가 잘 움직일수록 이전 확신을 유지할 근거가 줄어든다. 이 기능으로 기억은 현재 추론의 일부가 된다.
제28장에서는 비슷한 메커니즘을 작동하는 에이전트에서 볼 수 있다.

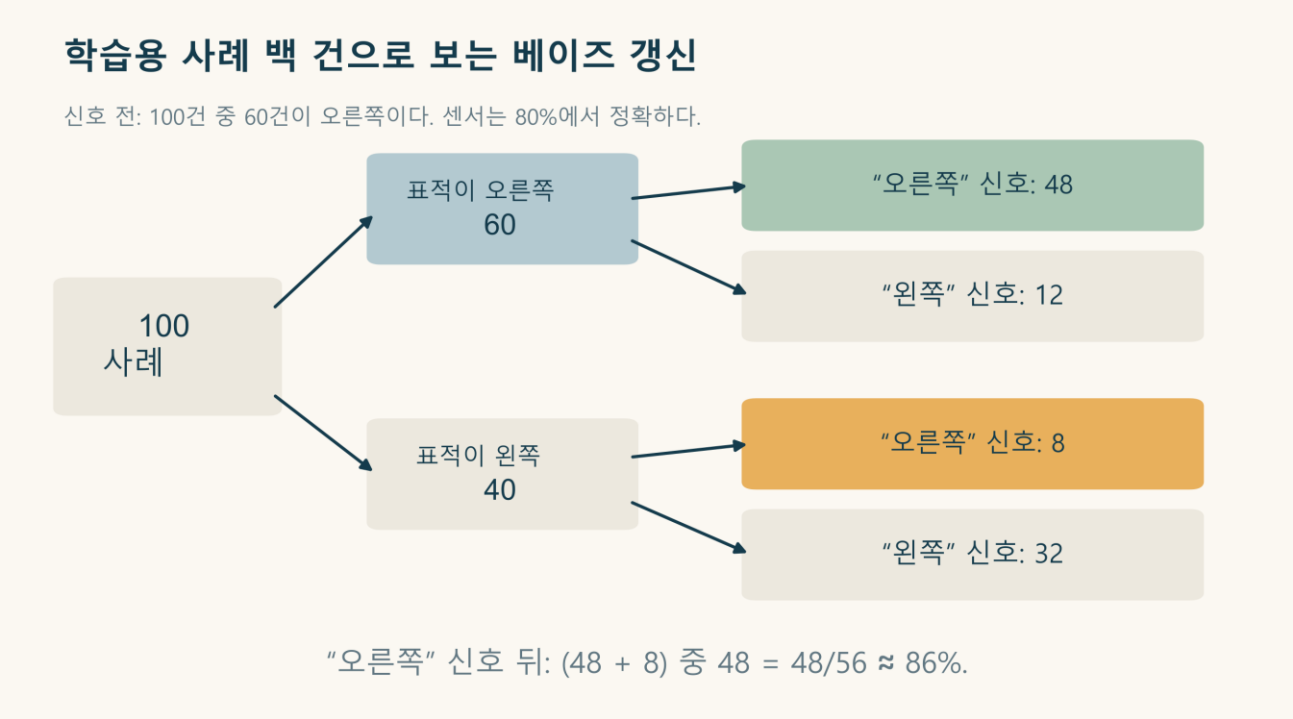


그림 32. 사전 평가에서 갱신된 확률로. 나무는 이 장의 학습 예제 수치인 사전확률 0,6과 대칭 센서의 신뢰도 0,8을 그대로 사용한다. "오른쪽" 신호가 나오면 사례 56개가 남으며, 그중 표적이 오른쪽인 사례가 48개다. 이는 정해진 전제 아래의 계산이며 인간 지각에서 관측한 빈도가 아니다.

## 세계 모형에서 행동 모형으로

장치가 두 명령을 보낼 수 있다고 하자. «오른쪽» 명령이 포인터를 오른쪽으로 움직이는 동안에는 목표에 관한 지식만으로 선택할 수 있다. 그런데 연결이 모르게 바뀌어 오른쪽 명령이 왼쪽 결과를 낸다. 장치가 세계 모형만 보존한다면 목표 위치를 계속 알면서도 행동은 틀리게 된다. 그러므로 자기 명령과 관찰 결과의 관계를 따로 표상해야 한다.

이 장치에는 자기 행동 능력의 기능적 모형이 생긴다. 내용은 구체적이다. 어떤 명령이 어느 확률로 어떤 결과를 내는가이다. 이를 인간의 완전한 자기의식이라고 부르기에는 이르다. 그래도 자기관련 변수가 해결하는 과제는 정확히 보일 수 있다. 세계에 대한 지식의 오류와 행동 제어의 오류를 구별하게 한다.

행동이 실패했다고 하자. 계는 적어도 목표를 잘못 파악했거나, 명령과 결과의 대응이 바뀌었거나, 실행기가 우연히 고장 났다는 세 원인을 고려해야 한다. 모든 오류를 센서 탓으로 돌리면 학습으로 올바른 지각을 망친다. 모두 행동 탓으로 돌리면 제어를 끝없이 바꿀 것이다. 좋은 모형에는 오류의 원인을 구별할 관찰이 필요하다.

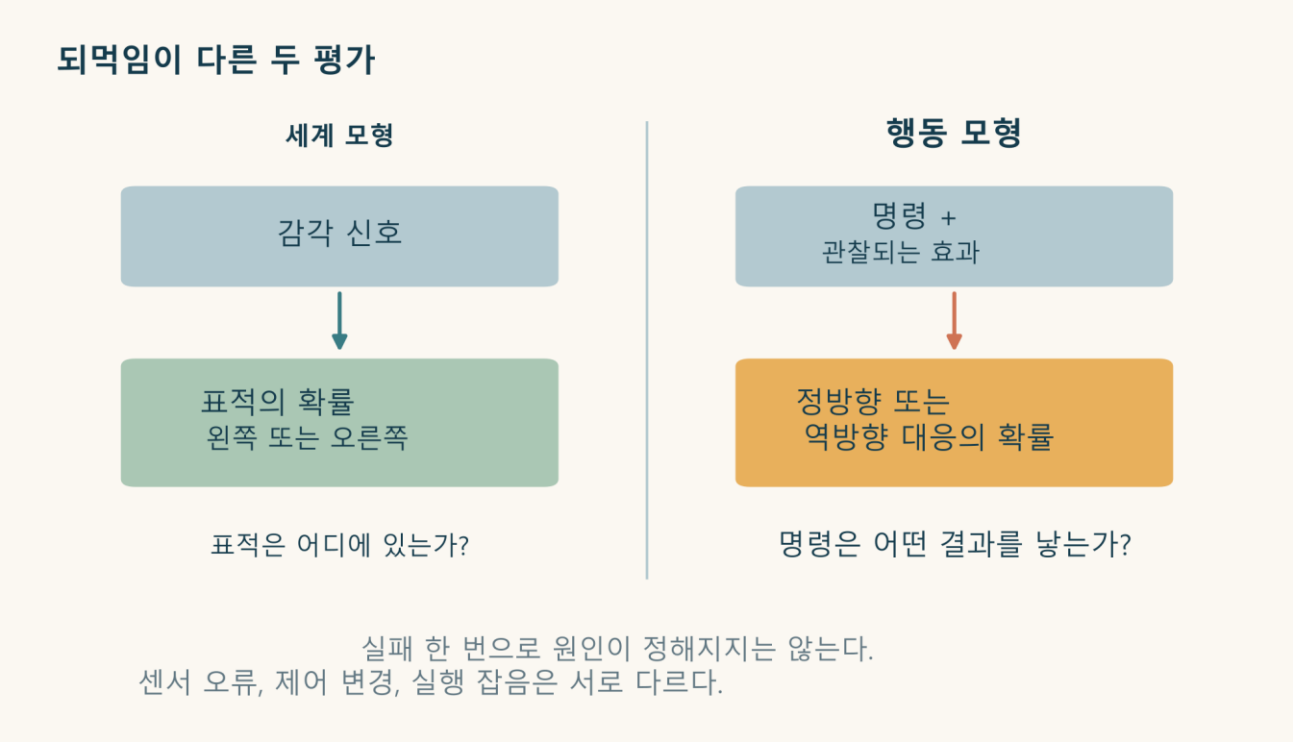


그림 33. 분리된 세계 모형과 행동 모형. 학습용 장치는 표적의 위치에 대한 믿음과 명령이 보이는 효과에 대응하는 방식에 대한 믿음을 따로 보존한다. 서로 다른 관찰이 서로 다른 평가를 갱신한다. 실패 한 번만으로 지각 오류, 제어의 변경, 우연한 실행 실패 중 어느 것이 원인인지 명확히 정할 수는 없다.

## 신호들의 공통 원인으로서의 신체

사마드, 청, 샴스의 베이즈 모형에서는 보이는 손에 대한 소유감을 시각, 촉각, 고유수용감각 신호가 공통 원인에서 비롯되었다는 추론과 연결했다. 시간과 공간에서의 일치는 신호들을 하나로 묶는 가설의 가능도를 바꾼다. 이 연구는 신체 소유감의 특정 효과에 관한 메커니즘을 제안한다. 그로부터 자기를 느끼는 모든 형태나 자전적 «나»를 도출하지는 않는다. [15]

자기 사례로 공통 원인의 뜻을 이해할 수 있다. 손가락을 움직일 때 움직임이 보이고 관련 신체 신호가 들어온다. 한 가설은 이를 한 손 움직임의 결과로 묶는다. 다른 가설은 서로 독립인 여러 사건을 가정한다. 신호가 안정적으로 맞아떨어지면 첫 모형은 독립적인 우연의 일치를 더 적게 가정할 수 있다. 그러나 결론은 어떤 대안을 표상하며 계가 어떤 일치를 기대하도록 학습했는지에 달려 있다.

신체 소유감과 행위 주체감이 반드시 함께 달라지는 것은 아니다. Kalckert와 Ehrsson의 실험에서는 고무 손의 움직임 조건을 통해 손이 자기 것이라는 주관적 소유감과 그 움직임을

제어한다는 느낌을 구별할 수 있었다. 이는 두 평가를 한 지표로 환원하는 데 반대하는 논거다. 일상에서 둘이 고립되어 존재한다는 뜻은 아니다. [4]

## 내부 상태의 예측

Anil Seth는 몸 안에서 오는 신호의 원인에 대한 추론으로 정서 상태와 신체적 자기감을 살펴보자고 제안했다. 이 접근법은 예측과 신체 상태의 조절을 연결한다. 이는 이론적 연구 프로그램이다. 유용성은 감각과 행동의 어떤 구체적인 차이를 설명하고 예측하는지에 따라 판단한다. [22]

자신의 모형을 설계할 때 여기서 물을 수 있다. 같은 입력 변화가 위험이나 신체적 부하, 측정 오류를 나타낼 수 있다면 계는 원인을 어떻게 구별할까? 실습 장치에 행동으로 소모되는 내부 자원과 잡음 있는 센서를 둘 수 있다. 그러면 미래에 행동이 가능한지를 예측해 자기 상태의 평가를 검증한다. 얻는 것은 자원 조절 모형이며 장치가 피로를 체험한다고 볼 근거는 아니다.

예측적 설명의 공통 한계는 적용 범위가 넓다는 데 있다. 많은 과정을 «예상»과 «오류»로 기술할 수 있다. 과학적 가치는 모형 사이의 정확한 차이에서 나온다. 어떤 원인을 허용하고 확신도는 어떻게 변하며 무엇을 오류라 보고 어떤 행동이 오류를 줄이는가? 결과마다 규칙을 새로 맞추면 설명이 지나치게 유연해진다.

관찰의 의존성도 검증해야 한다. 두 센서가 동시에 오른쪽을 가리킨다고 하자. 두 센서의 오류가 독립이라면 일치가 믿음을 크게 강화할 수 있다. 둘째 센서가 첫 센서를 복사했을 뿐이라면 새 정보가 없다. 복사본을 독립적인 확인으로 세는 모형은 과도하게 확신할 것이다. 더 복잡한 경우에는 두 센서가 장치가 아직 모르는 공통 잡음에 의존할 수 있다.

그러므로 확신도 조절에는 갱신 공식뿐 아니라 오류 출처의 모형이 필요하다. 실습에서는 수치가 같은 두 관측열을 비교할 수 있다. 하나는 독립 측정으로, 다른 하나는 복사로 만들었다. 올바른 평가는 그 열을 만든 절차에 따라 달라져야 한다. 이는 서로 일치하는 입력이 많다고 세계나 자기 상태에 대한 지식이 반드시 믿을 만해지지는 않는 이유를 보여 준다.

## 한 관찰과 서로 다른 두 불확실성

신호 전에는 목표가 오른쪽일 확률이 0,6이다. 센서는 0,8의 확률로 방향을 맞히며 오류는 대칭적이다. 오른쪽 신호 뒤에 오른쪽 목표 가설의 가중치는 0,6 곱하기 0,8인 0,48이다. 왼쪽 목표 가설의 가중치는 0,4 곱하기 0,2인 0,08이다. 정규화하면 0,56 가운데 0,48, 즉 약 0,857이 된다. 이렇게 갱신하는 것은 목표 위치다. 명령을 선택하려면 오른쪽 명령이 오른쪽 결과를 내는지도 알아야 한다. 대응이 반전되었다면 오른쪽일 가능성이 더 큰 목표에 왼쪽 명령이 필요하다. 센서 평가를 정밀하게 해도 잘못된 제어 모형이 고쳐지지는 않는다.

### 여기까지 확인한 것

세계와 제어, 센서의 신뢰도는 각각 다른 오류 출처일 수 있다. 성공적으로 학습하려면 이 가설들을 구별할 정보가 필요하다. 이제 어떤 자료로 각 부분을 갱신할 수 있는지 추적한다.

## 실습 15 세 오류 구별하기

실습 목표는 목표 위치의 평가와 명령이 결과에 연결되는 방식의 평가를 구별하는 것이다.

제11장의 확률적 갱신을 활용한다. 행동 전에는 관찰과 이력을 이용할 수 있다. 실제 목표와 실제 결과는 환경과 평가자에게 속한다.

예상 소요 시간은 30–40분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 사전확률 0,6과 센서 정확도 0,8로 목표가 오른쪽일 확률을 다시 계산한다. 이어 왼쪽 신호를 받았을 때의 예측을 구한다.
2. 목표가 왼쪽 또는 오른쪽이고 명령과 결과의 대응이 정방향 또는 역방향인 네 경우에서 운동 잡음 없이 목표에 도달할 명령을 적는다.
3. 각 경우에 감각 관찰로 무엇을 알며 명령과 보이는 결과의 비교로 무엇을 아는지 밝힌다.
4. python teaching_lab.py --chapter 15로 준비된 예를 연다. 결정 표를 보여 주는 예다. 분석하는 모형에서는 실제 목표와 명령 대응의 열을 가린다.

두 확률 갱신과 네 결정, 되먹임 출처의 관계도를 남겨 둔다.

힌트: 가설 아래에서 신호의 가능도와 신호를 얻은 뒤 가설의 확률은 계산에서 서로 다른 자리를 차지한다.

연구 확장: 짝을 이루거나 환경의 열 하나를 가리고 종이 위에서 스무 번 시행하라. 제어 변경을 미리 정하고 명령을 선택하는 사람에게 알리지 마라. 결과가 나오기 전에 예측을 적고, 나온 뒤에는 정보를 얻은 상태들만 갱신하라.

해설과 답은 부록 F 제15장 참조, 304쪽

## 자기 점검 문항

1. 모든 예측은 의식적인 예상인가?
2. 행동 모형과 세계 모형은 어떻게 다른가?
3. 신호가 없다는 사실과 부정적인 신호를 구별해야 하는 까닭은 무엇인가?

예측 메커니즘이 더 분명해졌다. 이제 실제 연구로 돌아가 어떤 자료가 이론적 설명들을 실제로 구별하는지 묻는다.

제16장

# 자료가 이론을 구별하는 지점

여러 메커니즘을 기술하고 예측을 얻을 수 있게 되었다. 그러나 같은 성공적인 결과가 여러 이론에 들어맞을 수 있다. 나아가려면 예측이 갈라지는 조건을 찾고 실험에서 정확히 무엇을 측정했는지 이해해야 한다. 보고가 있는 연구와 없는 연구는 이론과 절차를 함께 논의해야 하는 이유를 잘 보여 준다.

## 뒷받침하는 증거만으로 충분하지 않은 이유

두 모형이 영상이 보일 때 어떤 신호가 강해지리라 예상한다고 하자. 실험에서 신호가 강해졌다. 이 결과는 둘을 뒷받침해도 어느 모형을 고를지에는 거의 도움이 되지 않는다. 구별하려면 예상이 다른 상황이 필요하다. 가령 첫 모형은 응답 요구가 없어도 신호가 남는다고 보고, 둘째 모형은 신호를 응답 준비와 연결할 수 있다. 그러면 지시 변경은 검증 조건에 대한 의미 있는 개입이 된다.

의식 이론에서는 이 문제가 특히 크다. 한 실험에 인식과 기억 유지, 확신도, 결정, 움직임이 함께 들어간다. 모두 같이 달라지면 발견한 차이를 여러 방식으로 설명할 수 있다. 연구는 주어진 과제가 허용하는 만큼 이 연결된 과정들을 구별하는 일이다. 뇌 영역의 이름만으로는 이를 구별할 수 없다.

## 검증에 대한 사전 합의

COGITATE 프로젝트에서는 경쟁하는 접근법의 지지자들과 연구 컨소시엄이 연구 설계, 서로 다른 예측, 해석에 대해 사전에 합의했다. 2025년에 발표한 연구에서 256명의 참여자는 지속 시간이 다른 선명한 자극을 보았다. 기능적 자기공명영상, 자기뇌파검사, 두개강 내 뇌파검사를 사용했다. 결과는 일부 예상을 뒷받침하면서 두 이론 모두에 어려움을 제기했다. 예상했던 지속적인 후방 동기화가 없었고, 자극이 사라질 때 전반적으로 나타날 것으로 예상한 급증도 없었다. 검증한 것은 합의된 예측이며, 이론의 가능한 모든 버전은 아니다. [23]

자신의 연구에서는 미리 검증 조건을 약속한다는 원칙이 중요하다. 자료를 얻기 전에 무엇을 지지와 어려움, 정보가 부족한 결과로 볼지 적어야 한다. 분석 뒤 어떤 관찰이든 예상한 것이었다고 말하면 이론은 틀릴 위험을 감수하지 않게 된다. 이 원칙은 책의 저자가 좋아하는 가설에도 적용된다. 개인적인 관심이 수정 조건을 대신하지는 못한다.

사전등록도 해석을 무오류로 만들지는 않는다. 미리 약한 지표나 너무 거친 예측을 고를 수 있다. 그러므로 실험이 기록한 계획을 따르는가, 계획이 실제로 중심 주장을 검증하는가를 모두 확인해야 한다. 두 번째가 의심스럽다면 이전 결과를 연구 이력에 남기고 주장을 명시적으로 수정해 새로 검증해야 한다.

## 보고가 연구 과제를 바꾼다

Cohen과 동료들의 2024년 연구에서는 보고하는 조건과 보고하지 않는 조건에서 시각 차폐를 비교했다. 후기 성분 P3b와 일부 지속적인 디코딩 특성은 보고 수행과 연결되었다. 보고하지 않는 조건에서는 더 이른 성분 N2가 가시성의 비선형 변화에 대응했다. 이는 특정 과제와 측정 방법의 결과이며, 모든 체험이 나타나는 보편적 시간을 정하는 것은 아니다. [24]

무엇을 보았는지 보고하려면 지각을 응답으로 바꾸고 필요한 정보를 보존하며 규칙을 수행해야 한다. 그러므로 보고 요구를 없앤 뒤 신호가 사라진다면 그 신호가 이런 동작과 관련되었을 수 있다. 그러나 응답 요구가 없다고 내적으로 이름 붙이기와 주의, 기억까지 없어진다는 보장은 없다. 참여자가 스스로 이전 과제 일부를 계속할 수 있다.

반대의 어려움도 있다. 지각을 묻지 않으면 지각이 생겼는지 어떻게 알까? 일부 설계는 눈 움직임의 특성과 같은 간접 지표를 사용한다. 체험과의 관계를 먼저 근거로 뒷받침하고 새 조건에도 적용되는지 검증해야 한다. 그렇지 않으면 보고를 배제하려던 방법이 논쟁적인 지표 하나를 다른 지표로 바꾸는 데 그친다.

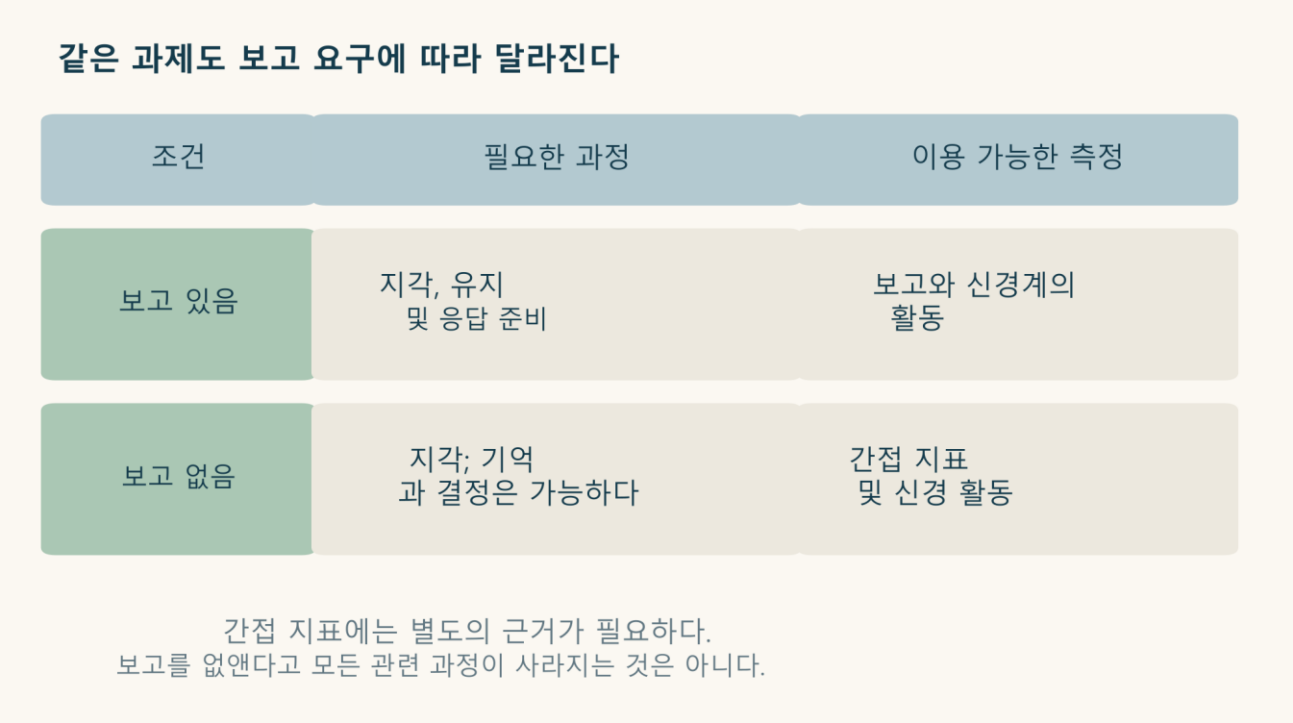


그림 34. 보고와 측정의 경계. 명시적 보고가 있는 조건과 없는 조건의 개념적 비교다. 응답 요구를 없애면 과제는 달라진다. 하지만 주의, 기억, 내부 결정까지 자동으로 사라지지는 않는다. 지각의 간접 지표도 검증이 필요하다. 이 표에는 정량적인 연구 결과가 없다.

## 전방 피질과 후방 피질로 단순한 양자택일을 할 수 없는 이유

He와 동료들은 2026년에 발표한 연구에서 뇌의 자기 활동을 기록하며 보고가 있는 조건과 없는 조건의 시각 지각을 비교했다. 후방 감각 영역의 표상은 두 조건에 모두 남았지만 발견된 전전두 영역의 내용 표상은 보고에 의존했다. 이는 내용 처리와 과제 요구의 구별을 뒷받침한다. 그러나 관찰 기록만으로 한 영역의 인과적 충분성을 입증하지는 못한다. [25]

2026년에 발표한 다른 연구에서 장과 동료들은 전두 하위 시스템들의 차이를 발견했다. 배측 주의망의 전두 영역과 관련된 해독은 보고가 있는 조건과 없는 조건 사이에서 일반화되었다. 다른 전두 표상들은 보고에 의존했다. 보고 없는 조건에서는 안구운동 지표를 사용했다. 이 결과는 «모든 전두 활동은 응답에만 해당한다»는 일반화를 가로막지만, 구분한 망 자체가 체험을 만든다는 사실을 확립하지는 않는다. [26]

이 연구들을 한 구호로 평균내기보다 함께 읽으면 좋다. 자극과 주의, 의식 여부를 판단하는 방식, 공간적 정밀도, 분석 감도가 다를 수 있다. «내용을 디코딩하지 못했다»는 것은 주어진

조건에서 선택한 분석이 자료로부터 내용을 추출하지 못했다는 뜻이다. 내용이 어디에도 표상되지 않고 인과 역할도 없다는 주장과는 다르다.

디코딩에 성공했어도 신중하게 해석해야 한다. 분류기는 지각의 결과나 차이들의 공통 출처, 부수 특징을 이용할 수 있다. 신호 안에 자극에 관한 정보가 있다는 것만으로 그 정보의 기능이 정해지지는 않는다. 필요조건을 논하려면 추가 개입이나 강한 인과 논거가 필요하다.

## 모순을 생산적으로 읽는 방법

한 논문은 신호와 가시성의 관계를 보고하고 다른 논문은 찾지 못했다고 하자. 정확히 무엇을 비교하는지부터 살핀다. 측정한 시간 구간이나 참여자의 전략이 달랐을 수 있다. 효과를 찾지 못한 경우에는 불확실성이 너무 컸을 수도 있다. 먼저 비교할 만한 조건에서 결과가 실제로 양립 불가능한지 확인한 뒤 메커니즘의 수정을 논한다.

학습 분석에는 다음 표가 유용하다. 발표의 명성을 평가하는 표가 아니라 결과들의 논리적 강도를 비교하는 표다.

| 관찰 | 직접적인 결론 | 아직 확인되지 않은 것 |
| --- | --- | --- |
| 신호가 보고와 관련됨 | 응답 조건에 대한 의존성을 발견함 | 신호가 다른 기능에는 참여하지 않는다는 사실 |
| 내용을 디코딩할 수 있음 | 자료에 구별 가능한 정보가 있음 | 해당 영역이 필요조건 또는 충분조건이라는 사실 |
| 개입으로 결과가 달라짐 | 주어진 조건에서 인과 영향이 있음 | 달라진 것이 체험 자체라는 사실 |
| 효과를 발견하지 못함 | 분석에서 예상한 차이를 찾지 못함 | 유용한 정밀도로 차이가 없다고 확인했다는 사실 |

## 의무적인 보고를 없애면 달라지는 것

매번 응답하는 과제에서 신호가 가시성과 관련된다고 하자. 지각과 관련된다는 설명과 보고 준비와 관련된다는 설명이 가능하다. 근거를 갖춘 보고 없는 조건에서도 신호가 남으면 둘째

설명의 범위를 좁히지만 체험에 대한 신호의 인과적 충분성을 확립하지는 못한다. 신호가 사라져도 검증해야 한다. 응답과 함께 주의와 기억이 달라졌을 수 있다. 그러므로 보고 요구를 바꾸어도 어떤 특성이 남으며 어떤 독립 방법으로 지각 자체를 확인하는지 좁게 묻는다.

### 여기까지 확인한 것

연구는 명시적으로 정한 경쟁 설명들과의 관계에서 설명을 구별하는 연구가 된다. 큰 표본이 필요한 비교를 대신하지 않으며 통계적 발견이 없다고 정확한 영을 뜻하지도 않는다.

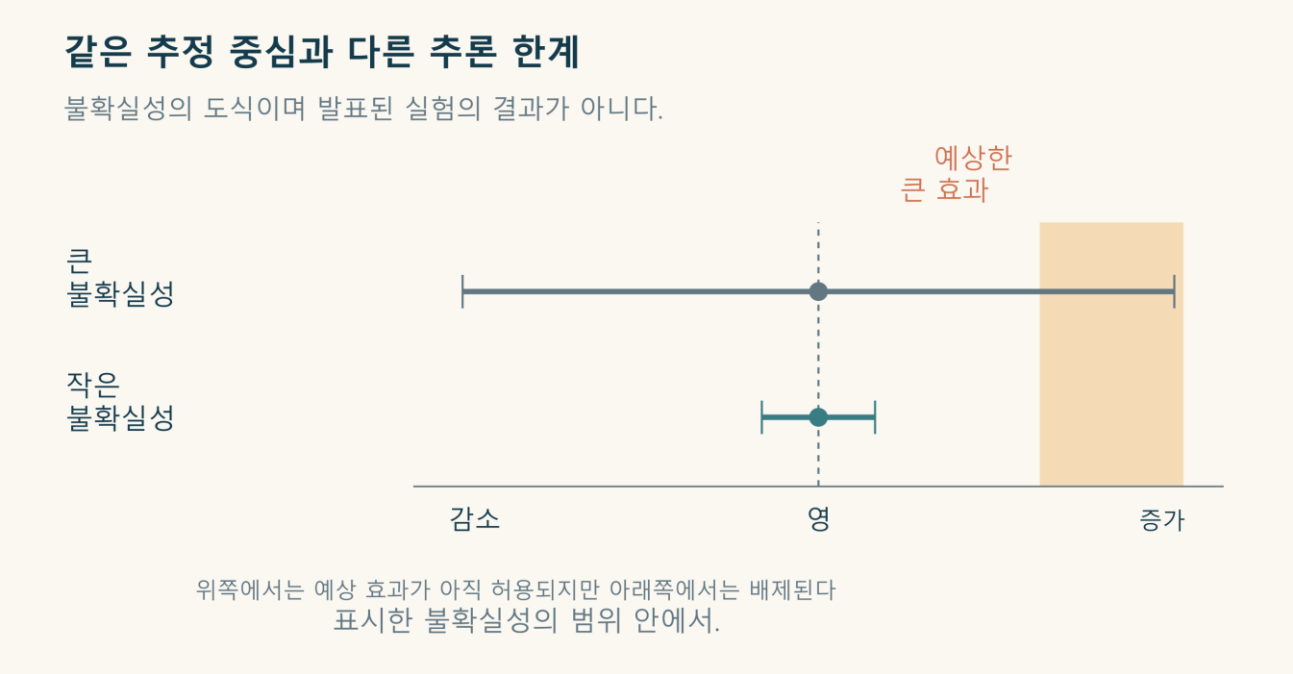


그림 35. 유의성이 없다는 것과 정확한 영이 다른 이유. 두 개념적 구간의 중심은 모두 영 부근이지만 허용하는 효과의 범위는 다르다. 그림에서는 좁은 구간만 미리 예상한 큰 증가를 배제한다. 수치 척도는 의도적으로 정하지 않았다. 실제 검증에는 효과 추정값, 근거 있는 구간, 이론의 정량적 예측이 필요하다.

## 실습 16 두 연구 비교하기

실습 목표는 방법과 결론의 차이를 유지하며 같은 질문에 관한 두 연구를 비교하는 것이다.

제6장의 측정 기록표, 제7장의 비교 품질, 제8장의 효과 크기와 불확실성을 활용한다. 디코딩은 선택한 분석으로 구별 가능한 정보를 추출하는 일이다.

기본 분석의 예상 소요 시간은 35–50분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 이 장에서 보고가 있는 조건과 없는 조건을 다룬 두 연구를 고른다. 예를 들어 출처 24와 25를 사용할 수 있다. 먼저 이 책의 해설을 활용하고, 원문 전체를 읽은 것으로 말하지 않는다.
2. 각각 자극, 과제, 가시성 지표, 측정 활동, 주된 비교라는 다섯 항목을 채운다. 모르는 정보는 명시적으로 표시한다.
3. 공통 질문 하나를 적고, 각 연구에서 주장을 뒷받침하는 관찰과 범위를 제한하는 관찰을 하나씩 적는다.
4. 응답 방식을 바꿀 때의 다음 예측을 정하고 경쟁 이론도 같은 결과를 예상하는지 묻는다.

보존할 것: 비교 가능한 두 기록표와, 이론들을 구별하는지 명시한 예측 하나.

힌트: 말이 비슷하다는 이유만으로 EEG, MEG, 단층영상법 사이에 결과를 옮기지 마라. 시간 구간과 분석 방법을 따로 명시하라.

연구 확장: 두 원출판물을 모두 열어 빠진 정보를 확인하고 비교 과제들이 달라지지 않았는지 검증하라. 핵심 효과의 크기와 불확실성을 따로 평가하고 저자의 결론과 자신의 결론을 구별해 남겨라.

해설과 답은 부록 F 제16장 참조, 304쪽

## 자기 점검 문항

1. 대규모 연구가 이론 논쟁을 반드시 최종 해결하는 것은 아닌 이유는 무엇인가?
2. 보고가 없는 조건은 체험을 직접 측정하는가?
3. 예상한 결과와 경쟁 설명을 구별하는 예측 가운데 무엇이 모형에 대한 신뢰를 더 크게 바꾸는가?

이론 비교를 통해 발생에 관한 질문을 준비했다. 먼저 하나의 사례로 좁혀 내용이 어떻게 생기고 시간 속에서 유지되는지 살펴본다.

제5부

# 의식과 자기의식의 발생

제17장

# 개별 에피소드는 어떻게 생기는가

지금까지 메커니즘과 지표를 비교했다. 이제 의식적 사례의 발생이 무엇을 뜻하는지 밝힌다. 이 질문은 아동 발달이나 종의 진화, 체험 자체가 존재하는 근거에 관한 질문과 다르다. 이 장에서는 한 사건의 동역학, 곧 시작 조건과 유지 과정, 이후의 보고만 살핀다. 제9장의 문턱 누적기는 이를 기술할 언어를 주지만 의식의 보편적인 시간 경계를 정해 주지는 않는다.

## 발생의 네 의미

«의식은 어떻게 발생하는가»라는 질문에는 시간 규모가 다른 과제들이 들어 있다. 동역학적 과제는 한 사례에서 왜 지금 특정 내용이 생겼고 이후 순간들에 무엇이 이를 유지했는지 묻는다. 발달 과제는 한 유기체의 능력이 일생 동안 어떻게 변하는지 연구한다. 진화 과제는 계통 안에서 메커니즘의 역사를 묻는다. 존재론적 과제는 왜 특정한 물리 과정에 체험이 따르는지 묻는다. 한 과제의 답이 다른 과제에 도움을 주어도 대신할 수는 없다.

이 장은 동역학적 질문을 다룬다. 창에 실루엣이 잠깐 보이는 경우를 생각해 보자. 지각할 준비와 들어온 신호의 처리, 체험의 내용, 이후의 응답을 구별해야 한다. 빠르게 일어나도 논리적으로는 서로 다른 단계다. 나중에 남긴 사건 기록이 사건의 시작과 일치할 필요는 없고, 빛에 대한 가장 이른 반응이 의식적인 보기의 시작일 필요도 없다.

## 전제, 과정과 결과

전제 조건은 주어진 조건에서 사건을 가능하게 한다. 과정은 검증하는 모형에 따라 관심 있는 변화를 직접 구현한다. 결과는 사건이 일어났기 때문에 생긴다. 이 역할은 시계상의 순서만이

아니라 인과관계로 정해진다. 한 과정이 여러 역할을 할 수 있다. 이전 자극의 기억은 이전 사건의 결과로 형성되었지만 다음 사례에 영향을 준다.

실습 상황을 보자. 화면에 도형이 나타나고 참여자가 몇 초 뒤 그 형태를 보고한다. 사전 지시는 무엇에 주의할지 정한다. 감각 처리는 형태를 구별하고 기억은 결과를 유지하며 응답 선택은 이를 동작으로 바꾼다. 기억을 손상하면 초기 구별이 같아도 나중의 보고가 부정확해진다. 그러므로 사례의 발생을 결론 내리려면 시간적 연쇄를 통제해야 한다.

반대로 이른 변화를 전체 사례의 충분한 설명으로 보면 오류가 생긴다. 센서가 두 자극을 빨리 구별한다고 하자. 그 구별이 후속 행동에 쓰이거나 체험된다는 결론이 곧바로 나오지는 않는다. «의식의 초기 표지»라는 이름에는 그 지표가 입력 조건이 아닌 의식적 내용 자체에 해당하는 이유가 따라야 한다.

## 출현과 유지의 학습용 동역학

우리의 간단한 모형을 구성하자. 영에서 일 사이의 활동값이 있다. 매 단계에서 이전 활동의 절반이 유지된다. 입력 신호가 있으면 십분의 사를 더한다. 이전 활동이 이미 이분의 일에 도달했다면 추가적인 내부 규칙에 따라 십분의 삼을 더한다. 값의 상한은 일이다. 모든 수는 계산을 명료하게 보여 주기 위해 선택한 것이며 인간 뇌의 매개변수를 추정한 것이 아니다.

연속한 두 입력은 각각 십분의 사와 십분의 육의 활동을 만든다. 이어서 입력이 사라진다. 내부 유지가 있으면 활동은 십분의 육으로 남는다. 내부 유지가 없으면 먼저 십분의 삼, 그다음 백분의 십오로 떨어진다. 같은 입력 이력이 서로 다른 내부 상태 지속 시간을 만들어 낸다.

| 단계 | 입력 | 내부 유지 없음 | 내부 유지 있음 |
|---|---|---|---|
| 첫 번째 | 있음 | 0,40 | 0,40 |
| 두 번째 | 있음 | 0,60 | 0,60 |
| 세 번째 | 없음 | 0,30 | 0,60 |
| 네 번째 | 없음 | 0,15 | 0,60 |

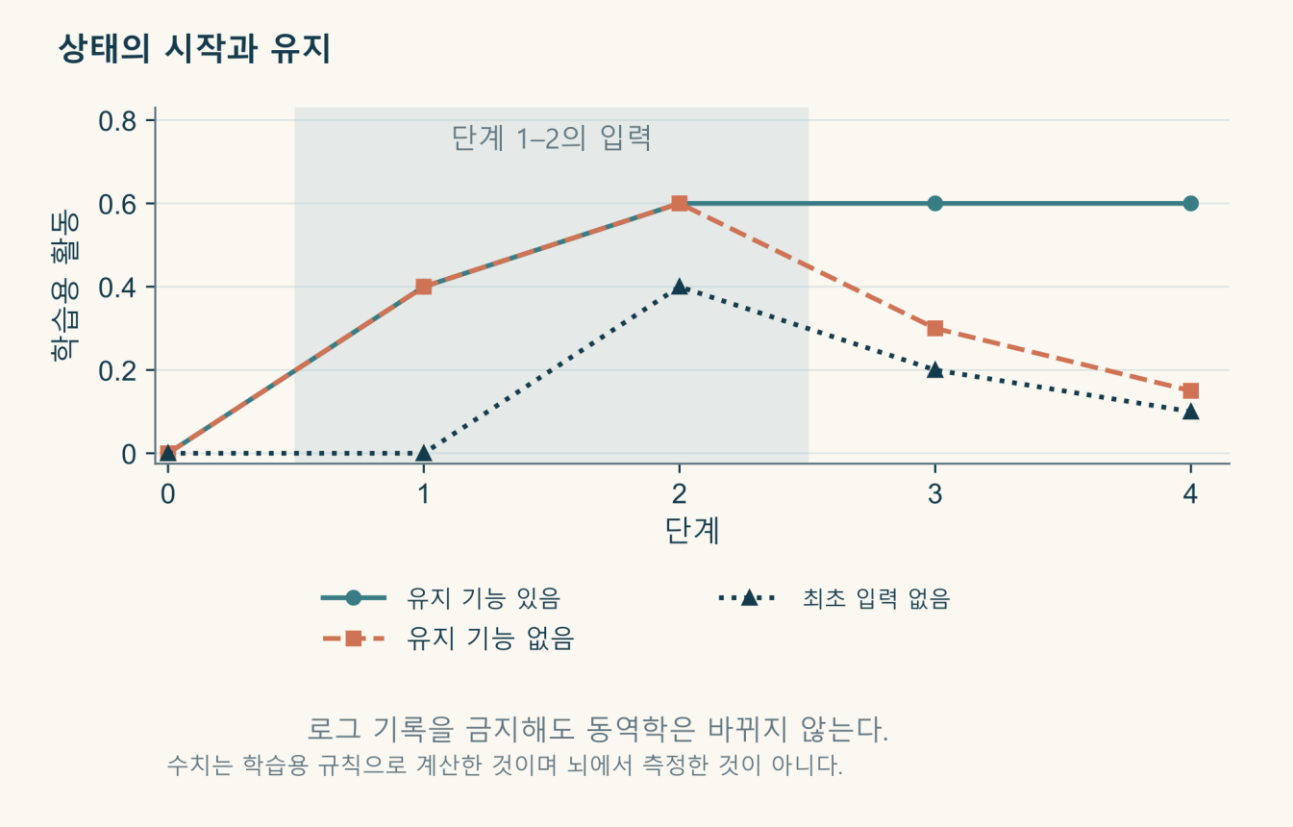


그림 36. 한 학습 메커니즘의 세 궤적. 제17장 규칙의 계산이다. 각 단계에서 이전 활동의 절반이 남고 입력이 0,4를 더한다. 이전 활동이 0,5 이상이면 내부 유지 기능이 0,3을 더한다. 입력을 두 번 주면 유지되는 활동은 0,6 수준에 남는다. 최초 입력을 없애는 것과 유지 기능을 없애는 것은 결과가 다르다. 관찰자의 로그 기록은 상태 규칙에 들어가지 않는다.

이 모형에서는 상태의 시작과 유지를 따로 연구할 수 있다. 입력 한 번으로는 안정된 작동을 시작하지 못하지만 두 번이면 가능하다. 시작한 뒤에는 규칙상 입력이 더 필요하지 않다. 그러나 얻은 것은 안정된 변수이며 체험의 증거가 아니다. 또 수 하나로 빨간 원과 조용한 소리의 차이를 설명할 수 없다. 내용 있는 설명에는 가능한 사례들 사이 차이의 표상이 필요하다.

실제 연구에서도 같은 논리를 적용하려면 자극 제시 전과 유지 중, 제시 종료 뒤에 서로 다른 개입이 필요하다. 시작에만 영향을 주는 개입과 유지에만 영향을 주는 개입은 모형에 다른 제약을 준다. 효과가 같아도 정보가 된다. 제안한 구별이 실제 구조에 맞지 않거나 개입이 너무 많은 과정을 동시에 바꾸었을 수 있다.

## 메커니즘의 변경이 보여 주는 것

Noorman과 동료들의 두 연구에서는 뇌 전기 활동을 기록하며 메만틴이 시각 도형의 처리에 미치는 영향을 연구했다. 재귀 통합으로 해석한 신경 지표는 달라졌지만 지각 민감도는 개선되지 않았다. 여기서 사용한 eLife 버전은 2025년 유월 18일에 발표되었다. 우리의 질문에서 중요한 것은 가정한 메커니즘의 변화와, 그에 대응하는 행동 민감도 변화의 부재 사이의 간극이다. [27]

이 결과는 모든 지표를 서로 바꾸어 쓸 수 없음을 보여 준다. 측정한 과정이 해당 과제 수행을 제한하는 요인이 아닐 수 있다. 행동 지표가 관심 있는 변화에 충분히 민감하지 않을 수도 있다. 처음의 메커니즘 해석이 불완전할 수도 있다. 설명을 고르려면 새 자료가 필요하다. 좋아하는 이론이 자동으로 선택되는 것은 아니다.

자신의 모형에도 같은 규칙이 적용된다. 변수를 꺼도 측정 결과가 같다고 그 변수가 어디에도 필요 없다고 바로 결론 내릴 수는 없다. 개입이 실제로 작동했는지, 과제가 그 기능에 민감했는지, 다른 메커니즘이 변화를 보상했는지 확인해야 한다. 그렇다고 검증하지 않은 보상을 끝없이 추가해 가설을 구할 수도 없다. 보상마다 검증할 새 귀결이 나와야 한다.

## 내용의 시간과 보고의 시간

누군가 «나는 그것을 즉시 알아차렸다»고 말하면, 이 말은 사건에 대한 그 사람의 평가를 전한다. 이를 시간 측정으로 바꾸려면 지시, 척도, 비교 조건, 오차 추정이 필요하다. 버튼을 누른 시점을 체험이 시작된 시점과 그대로 같다고 볼 수는 없다. 그 사이에는 의사결정과 동작 수행이 있을 수 있다. 여러 단계가 서로 영향을 줄 수 있으므로 평균 동작 시간을 빼는 것만으로도 모든 어려움이 해소되지는 않는다.

제약이 다른 여러 지표가 일치하는지 찾으면 좋다. 제시 시간에 따른 가시성 변화와 오류의 성격, 주관적 평가, 신경 동역학은 사례의 다른 측면을 보여 준다. 일치가 최종 설명을 보장하지는 않아도 일부 대안을 배제하는 데 도움이 된다. 어긋남은 지표의 이름을 바꾸어 없애지 말고 결과로 남겨야 한다.

동역학적 설명에는 내용의 종료와 교체도 포함되어야 한다. 활동을 강화하기만 하는 모형은 첫 자극에 머무를 수 있다. 중단이나 경쟁, 갱신의 메커니즘이 필요하다. 그러므로 서로 다른 입력을 연이어 검증하는 일이 한 번의 성공적인 인식 시연보다 유익할 때가 많다.

시간을 평균할 때도 문제가 생긴다. 실습 계가 매 시행에서 급격히 전환하지만 전환 시점은 조금씩 다르다고 하자. 시행들을 평균한 곡선은 매끄러워 보일 것이다. 반대도 가능하다. 문턱을 기준으로 시행을 고르면 개별 궤적보다 급격한 전이가 있는 것처럼 보일 수 있다. 그러므로 평균 신호의 형태로 개별 사례의 동역학을 하나로 정할 수는 없다.

검증할 때는 개별 시행의 전체 궤적과 선택 규칙을 보존한다. 매 시행에서 점진적으로 변하는 모형과 시작 시점만 다르고 급격하게 변하는 모형을 비교한다. 둘이 평균 곡선을 똑같이 설명한다면 분포나 시간 이동을 나타내는 추가 지표가 필요하다. 이렇게 하면 하나의 시점을 의식의 보편적 경계로 정하지 않으면서도 발생의 질문을 더 정확하게 만들 수 있다.

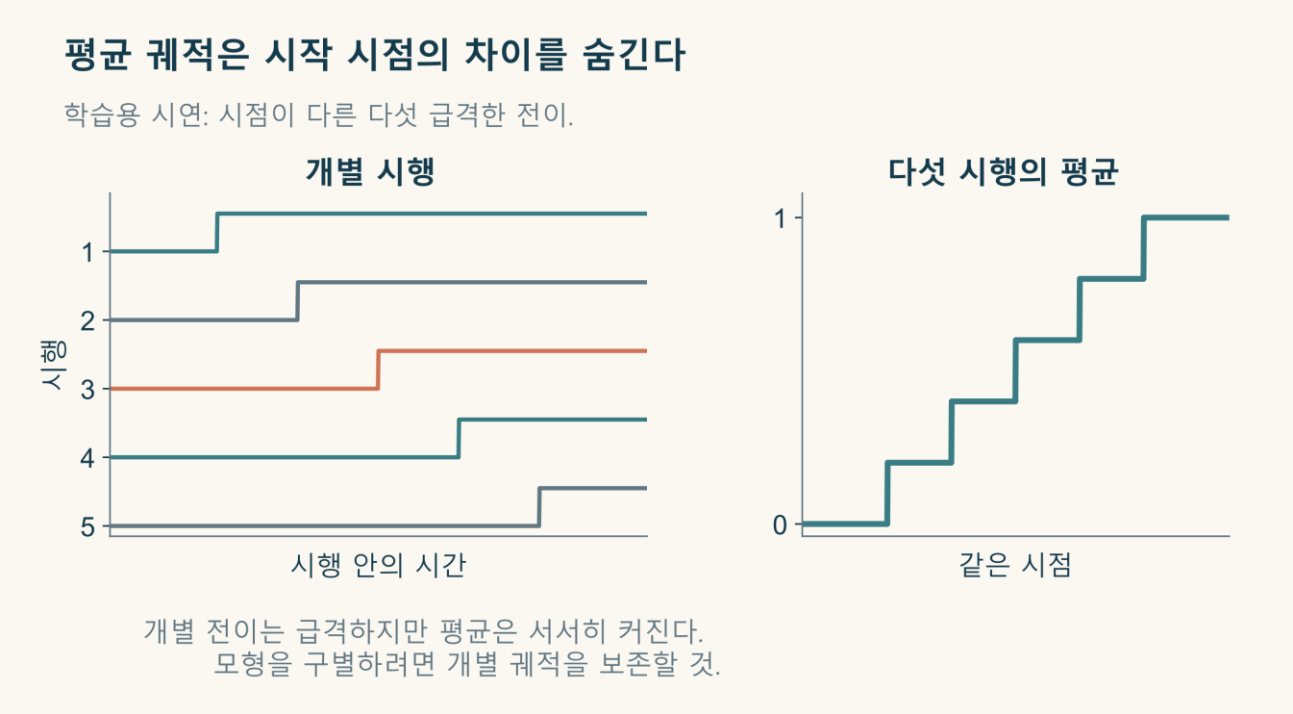


그림 37. 평균과 개별 에피소드의 동역학. 가상 시행 다섯 번에서 상태는 영에서 일로 뛰어오르지만 전이 시점은 다르다. 오른쪽에는 계단처럼 증가하는 정확한 평균을 계산했다. 시점 차이가 많아지면 평균 곡선이 매끄러워 보일 수 있다. 이는 동역학 측정의 문제를 설명하는 예이며 체험이 생긴 순간을 가리키지 않는다.

## 평균 곡선이 개별 시행의 동역학을 선택하지 못하는 이유

다섯 시행이 모두 영에서 일로 급격히 변하되 시작 시점이 다를 수 있다. 그러면 시간별 평균은 계단처럼 높아지고 시점이 촘촘히 퍼져 있으면 매끄러워 보인다. 다른 모형은 매 시행에서 점진적으로 커지면서 비슷한 평균을 가질 수 있다. 평균 곡선만 보존하면 두 설명을 구별하기 어렵다. 전체 궤적과 전이 시점의 분포, 또는 유지에 대한 추가 검증이 필요하다.

### 여기까지 확인한 것

시작과 유지, 결과 기록을 구별했다. 각 역할은 적절한 개입으로 검증해야 한다. 이제 작고 미리 정해진 메커니즘에서 확인한다.

## 실습 17 시작과 유지 검증하기

실습 목표는 이 장의 실습 모형 규칙으로 활동의 시작과 유지를 비교하는 것이다.

상태 변화 규칙과 문턱, 전제 조건과 과정, 결과의 구별을 활용한다. 시간적으로 가깝다는 것만으로 인과적 역할을 알 수는 없다.

예상 소요 시간은 25–35분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 활동을 0에서 시작한다. 단계마다 이전 값의 절반을 남기고 입력이 있으면 0,4를 더한다. 이전 활동이 0,5 이상이면 0,3을 더한다. 결과의 상한은 일로 둔다.
2. 단계 1과 2에 입력을 주고 이어 입력을 없앤다. 단계 4까지의 궤적을 미리 예측한다.
3. 유지에 따른 추가값 0,3을 없애고 반복한다. 이어 최초 입력을 없애고 반복한다. 각 궤적을 전체 모형과 비교한다.
4. 단계 4 뒤에 결과를 기록하는 표지를 별도로 붙인다. 이 표지가 동역학 자체를 바꾸는지 확인한다.

세 궤적과 시작 및 유지의 구별, 이후 기록의 역할에 대한 설명을 남겨 둔다.

추가값의 조건은 같은 단계에서 이미 갱신한 값이 아니라 이전 상태로 판단한다.

탐구를 이어 가려면 시작 시점이 다른 여러 시행을 만들고 평균을 점진적 변화 모형과 비교한다.

두 경우를 구별할 지표를 제안한다.

해설과 답은 부록 F 제17장 참조, 305쪽

## 자기 점검 문항

1. 동역학적 발생과 진화적 기원이 다른 까닭은 무엇인가?

2. 지속적인 활동이 의식적 사례의 존재를 입증하는가?

3. 신경 지표와 행동 지표의 어긋남은 무엇을 뜻하는가?

한 사례의 동역학으로 사람이 나이가 들며 능력이 어떻게 달라지는지 답하지는 못했다. 다음 장에서는 이 구별을 유지하며 여러 능력의 발달을 따로 살펴본다.

제18장

# 자기의식의 발달

한 사례는 짧은 시간 안에 생기지만 자기를 알아보거나 도움을 요청하는 능력은 다른 시간 규모에서 발달한다. 그러므로 의식이 생기는 나이를 묻기 전에 구체적인 성취들을 구별해야 한다. 여기서는 감각 내용과 시각적 자기 인식, 불확실성 조절, 자전적 자기 기술을 살핀다. 이렇게 나누면 한 검사에서 실패했다고 다른 능력도 모두 없다고 판단하는 일을 피할 수 있다.

## 여러 능력의 발달

«아이는 언제 자신을 의식하는가»라는 질문은 나이를 답해 달라는 요청처럼 보인다. 그러나 먼저 어떤 능력을 뜻하는지 정해야 한다. 자기 움직임과 외부 사건의 구별, 자기 신체의 인식, 자기 오류의 평가, «나»라는 말의 사용, 자기 근거에 대한 성찰은 서로 다른 성취다. 공통 전제가 있을 수 있지만 지표 하나가 다른 능력도 자동으로 측정하지는 않는다.

이 장에서 발달은 한 사람의 조직과 능력이 시간에 따라 달라지는 것을 뜻한다. 새로운 상호작용 방식의 습득과 기억, 학습, 조절의 변화를 포함한다. 어떤 시점 전에는 체험이 전혀 없다가 이후 모든 자기의식이 한꺼번에 나타난다고 미리 가정하지 않는다. 연구의 과제는 정확히 무엇이 달라지며 어떤 관찰로 이를 구별할 수 있는지 알아내는 것이다.



그림 38. 일상적 상호작용에서의 발달. 대상과의 초기 상호작용을 그린 것이다. 목표를 향한 움직임, 행동 결과의 구별, 이후의 자기 이야기는 서로 다른 관찰을 요구한다. 이 그림은 의식이 생기는 연령 문턱을 정하지 않는다.

## 체험과 보고 능력

Kouider와 동료들은 생후 다섯 달, 열두 달, 열다섯 달 영아가 마스킹한 얼굴 자극에 보이는 뇌 반응을 연구했다. 모든 연령 집단에서 늦은 비선형 반응을 발견했고 성인의 의식적 접근 지표와 유추하여 해석했다. 어린 집단에서는 더 약하고 느렸다. 이 연구는 이른 시기의 의식적 지각 능력을 뒷받침하는 신경학적 근거를 준다. 하지만 첫 체험 자체가 생기는 나이를 정하거나 상세한 성찰을 측정하지는 않는다. [28]

이런 유추의 논리는 주의해서 살펴야 한다. 성인의 체험과 관련된 지표가 다른 집단에도 있으면 메커니즘에 공통점이 있다는 근거가 된다. 그러나 신호가 같은 기능을 하는지 검증해야 한다. 성인과 같은 패턴이 없다는 것도 모호하다. 비슷한 과제를 다른 방식으로 풀거나 측정이 그 조직에 충분히 민감하지 않을 수 있다.

이는 신경 자료에만 해당하지 않는다. 언어 질문은 말의 이해, 지시 유지, 응답 선택이라는 요구를 더한다. 아이는 연구 상태의 부재와 관계없는 이유로 언어 과제를 수행하지 못할 수 있다. 반대로 정해진 문구로 외운 성공적 응답은 이해의 복잡성을 과장할 수 있다. 따라서 추가 요구가 서로 다른 여러 과제를 비교하면 유용하다.

## 거울 속 자기 인식이 보여 주는 것

Beulah Amsterdam의 고전적 연구에서는 아이들이 거울상에 보이는 반응의 연령 차이를 관찰했다. 일부 아이들에게서는 두 번째 해 말 무렵 자기인식의 징후가 나타났다. 연구 집단 안에서도 결과는 같지 않았다. 이 자료는 시각적 자기인식의 특정 상황에 관한 것이며, 의식 발생의 보편적인 시간표는 아니다. [29]

과제를 실습용 도식으로 나누어 보자. 거울상에서 낯선 세부를 발견하고, 영상을 신체와 연결하고, 신체의 해당 위치를 찾아 움직여야 한다. 어느 단계에서든 실패할 수 있다. 예를 들어 손을 움직이기 어렵거나 표지에 관심이 없거나 영상이 낯설게 느껴질 수 있다. 결과를 해석하려면 추가 요구 각각에 관한 대조 관찰이 필요하다.

성공에도 여러 설명 수준이 있다. 움직임과 영상의 안정적인 연결에 근거할 수 있다. 현재 상호작용을 벗어나서도 같은 사람으로 자신을 표상한다는 주장은 더 강하다. 타인의 평가 대상으로 자신을 생각할 수 있다는 주장은 더 강하다. 다음 주장으로 나아갈 때마다 성공적인 움직임 하나의 의미를 늘리는 대신 추가 과제가 필요하다.

카메라 영상의 표지를 찾아 장치 몸체의 해당 부위를 만지는 완성된 프로그램을 생각해 보자. 이런 대응을 특별히 학습시킬 수 있다. 이 공학적 사례는 행동 기준을 메커니즘으로 나눠야 하는 이유를 보여 준다. 아이도 같은 방식으로 푼다는 증거는 아니며, 발달 연구에서 자기 인식의 중요성을 없애지도 않는다.

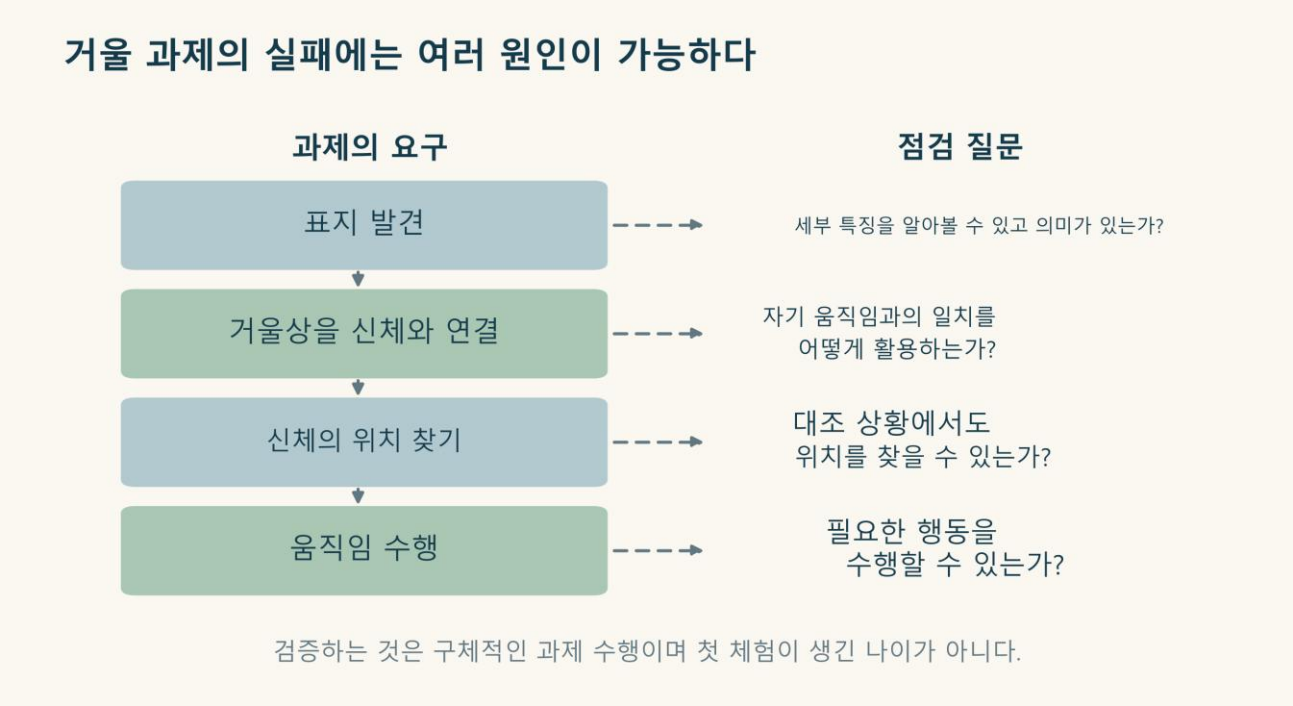


그림 39. 시각적 자기 인식 과제의 분해. 표지를 만지는 행동은 표지의 발견, 거울상과 신체의 연결, 위치 파악, 움직임의 가능성이라는 여러 조건에 달려 있다. 점검 질문은 실패의 원인들을 구분하는 데 도움이 되지만 그 자체로 완전한 실험 프로토콜이 되지는 않는다. 시행 한 번의 실패만으로 체험이나 다른 자기관련성 형태가 없다고 판단해서는 안 된다.

## 모른다는 것을 점검하기

구필, 로만 모니에, 쿠이더의 연구에서는 생후 스무 개월 아이들이 기억 과제에서 도움을 요청할 수 있게 했다. 도움의 이용은 불확실성에 따라 달라졌고, 언제나 스스로 선택해야 할 때보다 수행을 향상시켰다. 저자들은 이를 자신의 지식에 대한 초기의 비언어적 점검으로 해석했다. 이 결과가 아이에게 성인과 같은 지식 개념이 있거나 결정의 근거를 명시적으로 설명할 수 있음을 뜻하지는 않는다. [30]

메커니즘을 이해하려면 상황의 난이도 평가와 자기 기억의 품질 평가를 구별해야 한다. 첫 경우에는 오래 기다리면 도움을 요청하라는 규칙을 배울 수 있다. 둘째 경우에는 특정 사건을 얼마나 믿을 만하게 기억하는지에 대한 정보를 쓴다. 보통 과제에서는 두 전략이 일치할 수 있다. 구별하려면 예를 들어 지연 시간은 같고 사건을 관찰할 기회는 다른 조건들을 비교해 이 관계를 끊어야 한다.

도움 요청에는 사회적 측면도 중요하다. 도움 요청을 측정 가능한 행동으로 삼으려면 다른 사람이 도움을 줄 수 있고 기꺼이 줄 것임을 이해해야 한다. 그러므로 요청이 드물면 확신도뿐 아니라 의사소통 상황의 구조를 반영할 수 있다. 이런 분석은 행동의 가치를 깎는 대신 어떤 기능들이 함께 이를 가능하게 하는지 보여 준다.

## 사건에서 생애 속의 나로

자기 기술을 습득하는 사례를 보자. 학생이 암산에서 여러 번 틀리고 «나는 계산을 못한다»고 결론 내린다. 이 말에는 사건 기억과 결과 일반화, 특성을 자신에게 적용하는 일이 함께 있다. 이어 급할 때만 틀렸다는 사실을 알게 되면 «시간이 제한되면 너무 일찍 답을 고른다»로 바꿀 수 있다. 이는 자기 작동 조건을 더 정확하게 모형화한 것이다.

이런 전환은 성찰의 연구상 의미를 보여 준다. 변화의 대상은 자신이 낸 결과를 설명하는 규칙이 된다. 지속적인 능력과 일시적인 조건을 구분하는지, 추가 관찰을 찾는지, 구체화한 규칙을 새 과제에 적용하는지 연구할 수 있다. 나이 자체만으로는 이를 설명할 수 없다. 앞선 경험과 이용 가능한 검증 방법에 관한 정보가 필요하다.

발달 연구에서는 연령 집단의 비교와 같은 집단을 시간에 따라 관찰하는 일도 구별해야 한다. 전자는 조사한 사람들 사이의 차이를 보여 준다. 후자는 개인의 변화를 추적할 수 있지만 과제의 반복 자체가 결과를 바꿀 수 있다. 두 방법 모두 측정이 서로 비교 가능한지 통제해야 한다.

다시 검사했을 때 아이가 도움을 더 잘 요청하지만 두 검사 사이에 요청 규칙을 특별히 설명했다고 하자. 수행이 달라졌다는 것은 확인해도 규칙 학습과 불확실성 조절의 변화를 그 자체로 구별하지는 못한다. 요청 방식이 다른 새 과제가 필요하다. 전이가 유지되면 외운 몸짓 하나만으로 설명하기는 더 어려워진다.

이 분석은 반대 결론도 피하게 한다. 전이에 실패한 것은 새 의사소통 상황을 이해하지 못해서일 수 있다. 그러므로 달라진 결과를 내부 능력의 발달이나 부재를 나타내는 지표로 보기 전에 새 응답 방식을 사용할 수 있는지 따로 검증하면 좋다.




그림 40. 하나의 서열 없는 발달 프로필. 이 장의 실습에 제시한 가상 관찰 네 가지는 자기 인식, 불확실성 점검, 사건과 자신의 연결, 근거 재검토라는 서로 다른 질문에 답한다. 각각 따로 측정해야 한다. 이 도식에는 연령 경계나 하나의 의식 척도가 없다.

## 거울 과제를 나누어 분석하는 방법

거울상으로 발견한 표지를 만지려면 표지 자체를 알아차리고 거울상을 신체와 연결하며 해당 부위의 위치를 찾아 움직여야 한다. 어느 단계에서 실패해도 외부 결과는 같을 수 있다.
그러므로 먼저 대조 조건에서 각 행동이 가능한지 검증한다. 성공했다고 성인의 성찰로 곧바로 일반화할 수도 없다. 성공은 특정한 시각적 자기 인식의 조직에 관한 결과다. 과제를 나누면 검사의 의미를 더 정확히 할 수 있다.

### 여기까지 확인한 것

나이의 경계 하나로 능력의 양상을 대신할 수는 없다. 각 관계의 발달을 연구한 뒤 어떤 전제를 공유하는지 알아볼 수 있다.

## 실습 18 네 관찰과 여러 설명

실습 목표는 능력에 관한 자료와 해당 과제의 추가 요구를 구별하는 법을 익히는 것이다.

제3장의 관점과 신체 소유감, 행위주체감, 제4장의 수행 성공과 자기 점검의 구별을 활용한다. 여기서 발달은 한 유기체가 시간에 따라 달라지는 것을 뜻한다.

예상 소요 시간은 25–35분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 시각적 구별, 거울 속 자기 인식, 불확실할 때의 도움 요청, 자기 과거의 이야기라는 네 행을 만든다.
2. 각 행에 관찰한 행동과 언어, 지시 기억, 움직임, 도움을 받을 가능성의 이해 같은 추가 요구를 적어도 하나 적는다.
3. 실패한 결과에 대해 설명 두 가지와 둘을 가려낼 대조 절차를 제안한다.
4. 이 장의 발표 연구 하나를 선택하고 결론을 그 과제와 연령 집단에 한정한다. 최초 체험의 나이를 정하지 않는다.

네 질문의 양상과 대안 설명을 검증하는 대조 절차 하나의 분석을 남겨 둔다.

논의를 위해 아이에게 가정 검사를 할 필요는 없다. 발표된 설명과 실습용 사례를 사용한다.

탐구를 이어 가려면 능력을 다른 응답 방식으로 전이하는 과제를 제안한다. 먼저 참여자가 새 행동 방식을 이해했는지 따로 확인한 뒤 내부 메커니즘의 전이를 논한다.

해설과 답은 부록 F 제18장 참조, 306쪽

### 자기 점검 문항

1. 자기 인식에 성공한 나이를 의식이 생긴 나이로 볼 수 없는 까닭은 무엇인가?
2. 도움 요청은 명시적 성찰과 어떻게 다른가?
3. 어린 집단에서 예상한 지표가 없다는 것은 무엇을 뜻하는가?

개인의 발달에는 고유한 자료와 한계가 있다. 이제 진화 계통에서 메커니즘이 생겨난 또 다른 역사로 넘어간다.

제19장

# 체험의 진화를 연구하는 방법

한 개체의 발달과 종의 기원 역사에는 서로 다른 증거가 필요하다. 이제 기능과 구조, 계통 관계의 비교가 메커니즘의 진화에 관한 설명을 어떻게 제약하는지 살핀다. 기억의 이점은 환경에 달려 있으며 비슷한 능력도 다른 역사를 가질 수 있다. 그러므로 의식의 단일 서열을 만들지 않는다. 관찰한 특징에서 역사 가설로, 메커니즘에서 가정한 체험으로 넘어가는 단계를 분명히 분석하는 것이 이 장의 과제다.

## 메커니즘의 역사와 체험의 역사

생물의 구조에는 기원의 역사가 있다. 체험과 관련된 메커니즘이 언제 생겼으며 어떤 과제가 그 보존에 기여했는지 묻는 것은 자연스럽다. 그러나 메커니즘의 역사를 재구성하는 일과 체험이 있는지 확인하는 일은 다르다. 기억의 이점을 잘 설명해도 어떤 조직에서 기억에 의식적 회상이 따르는지는 여전히 논쟁할 수 있다.

진화 질문은 좁게 정해야 한다. 예를 들어 어떤 조건에서 선택이 여러 정보 출처의 통합이나 불확실성 조절, 자기 행동과 외부 사건의 구별을 유지할 수 있었을까? 이렇게 물으면 모형화하고 비교할 기능이 정해진다. 이 기능이 체험의 조건인가라는 다음 질문에는 별도의 이론적 연결이 필요하다. 두 질문을 섞으면 퀄리아의 기원에 대해 지나치게 확신하는 이야기가 된다.



그림 41. 환경은 서로 다른 행동 과제를 만든다. 물가와 육지의 서로 다른 생활 조건을 그린 것이다. 기억, 탐색, 행동 선택의 이득은 구체적인 환경에 달려 있다. 그림 속 생물의 배치가 체험이 생겨난 순서를 정하지는 않는다.

## 이점을 통한 설명의 방식

유전되는 두 전략의 모형을 생각해 보자. 첫 전략은 매번 마지막 신호에 따라 행동을 고른다. 둘째 전략은 최근 관찰 몇 개를 보존한다. 세계는 때때로 바뀌고 센서는 오류를 낸다. 세계가 안정적이면 기억이 우연한 오류의 영향을 줄일 수 있다. 세계가 빨리 바뀌면 같은 기억이 전환을 늦추어 행동을 악화할 수 있다. 그러므로 메커니즘의 이점은 환경과 매개변수에 달려 있다.

저장과 처리의 비용도 더하자. 이득이 드물면 정확한 기억도 비용을 보상하지 못할 수 있다. 이 학습 모형의 선택은 성공한 행동, 비용, 재생산의 관계를 어떻게 정했는지에 달려 있다. 따라서 «복잡할수록 낫다»나 «기억이 많을수록 의식에 가깝다»는 말은 구성한 과제로부터 나오지 않는다. 미리 정한 정상으로의 전진이 아니라 기능들 사이의 상충관계를 연구하는 것이다.

이제 둘째 전략에 «체험»이라는 변수를 주자. 그 변수가 인과 조직에 전혀 참여하지 않는다면 더해도 선택 모형에는 아무 변화가 없다. 이는 구체적인 구성의 한계를 보여 준다. 여기서 자연의 체험이 인과적으로 무력하다는 결론은 나오지 않는다. 작동하는 어떤 과정과 동일할 수도 있고, 가정한 과정과의 연결에 다른 설명이 필요할 수도 있다. 모형은 선택한 방식을 명시해야 한다.

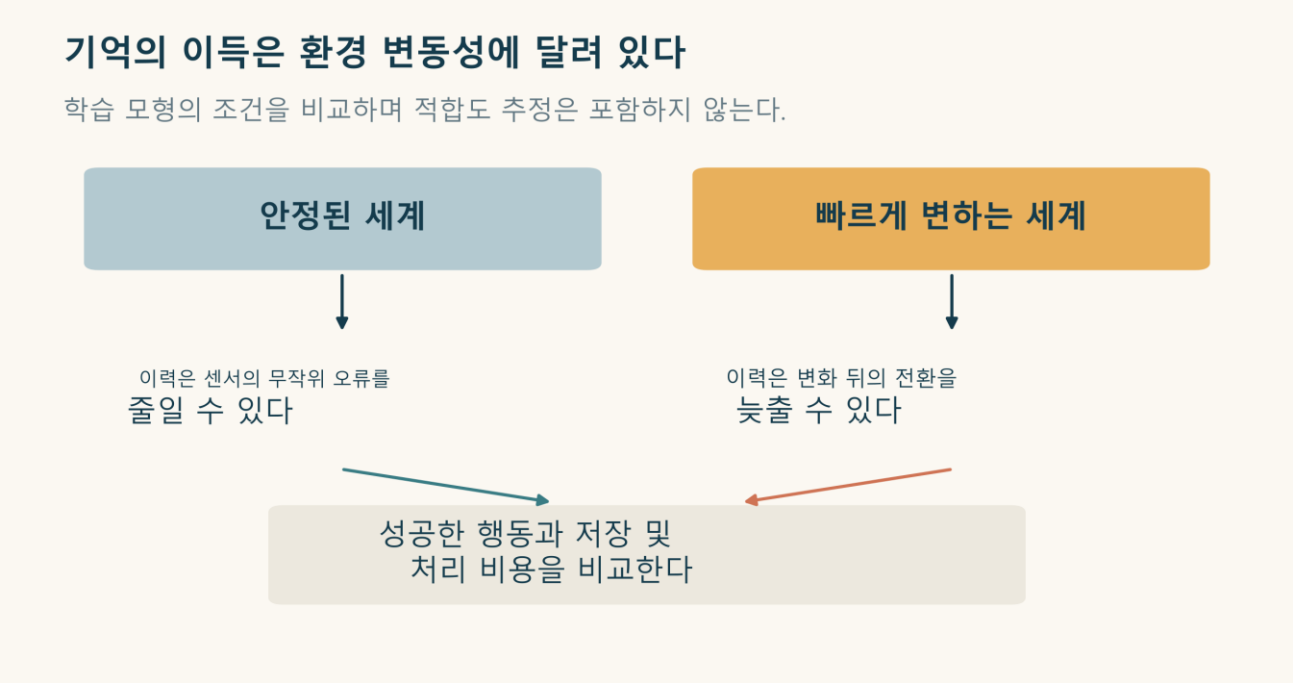


그림 42. 복잡성의 서열 대신 기억의 상충관계. 이 장의 모형에서 기억은 표적이 안정적일 때 잡음을 줄일 수 있지만 표적이 달라진 뒤에도 오래된 정보를 유지할 수 있다. 이점에 관한 결론에는 성공과 비용을 따로 고려해야 한다. 번식과의 연결도 명시적으로 정해야 한다. 이 기능 비교가 체험의 진화적 기원을 확립하지는 않는다.

## 유사성만으로 역사가 하나로 정해지지는 않는다

두 종에서 비슷한 능력을 발견해도 역사적 설명은 여러 가지다. 공통 조상에서 왔거나 독립적으로 생겼거나 공통 요소와 새 요소의 조합에 기초했을 수 있다. 같은 과제 결과에 다른 메커니즘으로 이를 수도 있다. 따라서 비교에는 행동, 구조, 기원을 함께 분석해야 한다. 비슷한 응답 수를 세는 것으로 대신할 수 없다.

펠젠슈타인은 종들을 통계적으로 완전히 독립적인 관찰로 다룰 수 없는 이유를 보였다. 특징들이 공통 진화 이력으로 연결되어 있기 때문이다. 그의 방법론 연구는 비교 분석에서

계통적 근연성을 고려하자고 제안한다. 의식 연구에서 이 원칙은 특징 하나를 가진 가까운 종 열 개가 그 특징이 생긴 독립 사례 열 번과 같지 않다는 뜻이다. [31]

끝 가지가 네 개인 가상 계통수를 만들자. 가까운 두 가지에는 메커니즘 M이 있고 나머지 두 가지에서는 나타나지 않는다. 첫 쌍의 공통 조상에서 M이 생겼다는 설명이 가능하다. 더 일찍 생긴 M이 둘째 쌍에서 사라졌다는 설명도 가능하다. 특징 하나의 분포로는 역사를 확정하지 못할 수 있다. 추가 정보로 가설을 구별할 수 있지만 관찰이 빠진 것을 곧바로 소실로 볼 수는 없다.

체험을 연구할 때는 불확실성이 더 커진다. M의 존재도 언제나 같은 신뢰도로 확인되는 것은 아니며 M과 체험의 관계는 이론에 의존할 수 있다. 역사 재구성에서는 두 불확실성을 모두 유지해야 한다. 논쟁적인 지표를 먼저 의식의 정확한 기준이라고 정한 뒤, 만든 계통수의 확실성을 그 기준의 독립 증거라고 제시해서는 안 된다.

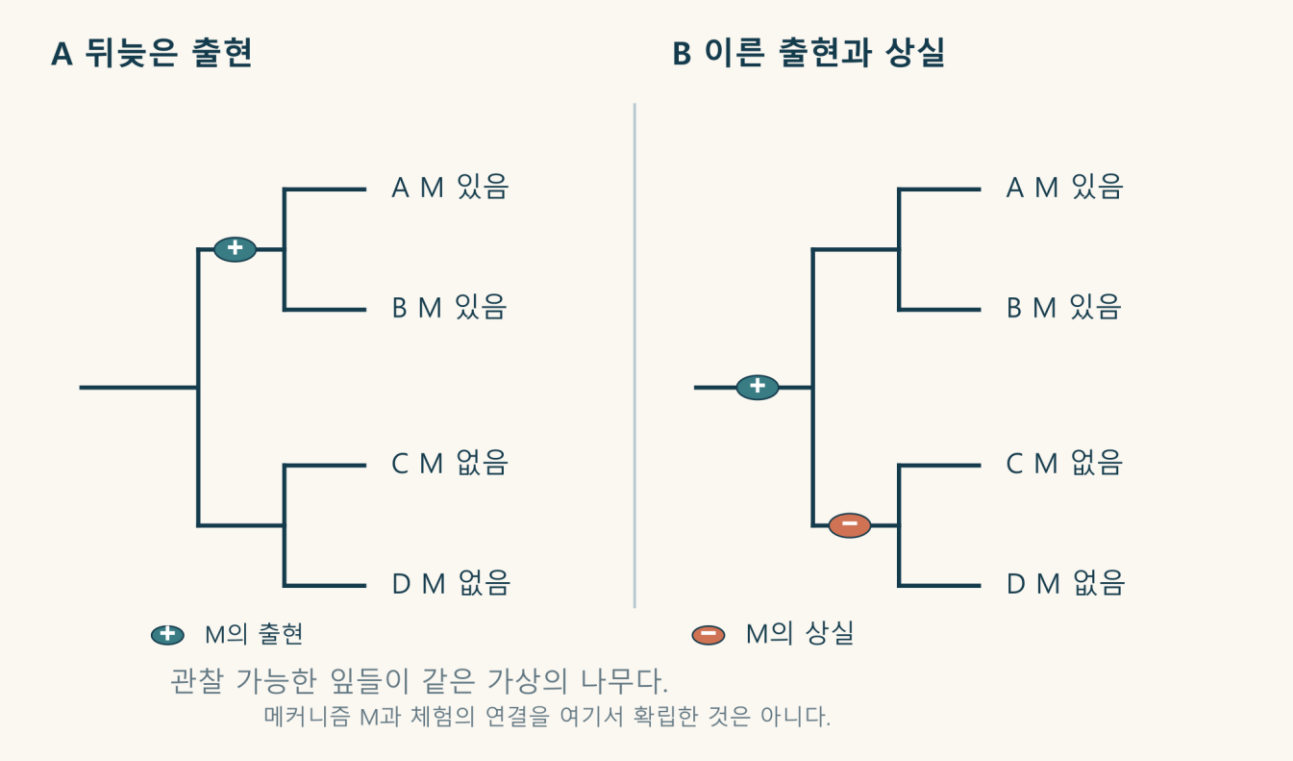


그림 43. 같은 특징 분포의 두 역사. 학습용 나무에서 메커니즘 M은 종 A와 B에 있고 C와 D에는 없다. 이는 첫 쌍의 공통 조상에서 생긴 경우와, 더 일찍 생긴 뒤 둘째 쌍에서 사라진 경우 모두와 양립한다. 이 나무는 의식의 기원을 재구성하지 않는다. M과 체험의 관계는 별도 질문으로 남는다.

## 기능과 구조의 비교

Barron과 Klein은 곤충과 척추동물의 행동 제어 시스템 사이의 기능적 유사성에 근거하여 곤충의 체험 가능성에 관한 논거를 제안했다. 이는 원저인 이론 연구이며, 퀄리아를 직접 관찰하거나 그 기원의 시점을 확립한 것은 아니다. 연구상의 가치는 가정한 조직과 체험에 관한 추론을 명시적으로 연결한 데 있다. 그 전환 자체는 여전히 검증과 논쟁의 대상이다. [32]

자신의 비교 연구는 선택할 때 계가 공간과 가능한 행동, 내부 상태의 정보를 통합하는가라는 더 좁은 질문에서 시작할 수 있다. 답하려면 세 요인의 변화가 독립적인 반응 세 개를 일으키는 것이 아니라 상호작용함을 보여야 한다. 예를 들어 자원의 이용 가능성이 이동 비용에 따라 경로 선택을 바꾸고, 새 정보가 이전에 접하지 않은 조건 조합에도 적용되어야 한다.

이런 결과는 특정한 제어 조직이라는 가설을 뒷받침한다. 이어 인과 메커니즘을 찾고 더 단순한 설명을 검증해야 한다. 의식으로의 전환은 그 조직이 왜 중요한지에 달려 있다. 제13–16장의 내용이 여기서 도움이 된다. 이론마다 비교 연구를 이끄는 특징과 개입이 다르다.

## 보편적 서열이 방해되는 이유

한 척도만 쓰면 서로 다른 능력을 묶게 된다. 어떤 종은 공간 과제를 잘해도 낯선 감각 채널을 요구하는 과제에서는 서툴 수 있다. 다른 계는 한 행동의 불확실성을 정확히 평가해도 새 영역에는 그 평가를 옮겨 적용하지 못할 수 있다. 이 양상들이 반드시 한 줄로 정렬되는 것은 아니다.

분석에는 기능 지도가 더 유용하다. 이용 가능한 정보의 종류, 시간 범위, 다른 상황으로 옮겨 적용하는 유연성, 신체의 조직, 오류 통제를 정리하는 것이다. 체험 자체에 관심이 있다면 그에 관한 논거를 별도 열에 놓는다. 명시적 성찰에 관심이 있다면 또 다른 과제 묶음이 필요하다. 이렇게 정리하면 «한 검사에서 더 잘한다면 모든 면에서 더 의식적이다»라는 결론으로 슬며시 넘어가는 일을 막을 수 있다.

현생 종을 과거의 살아 있는 단계로 여겨서도 안 된다. 그 종에도 자체적인 변화의 역사가 있다. 비교는 조상의 특정 특징들을 재구성하는 데 도움이 되지만 관찰하는 동물을 먼 조상 상태의

정확한 복사본으로 만들지는 않는다. 따라서 자신의 기원 시나리오는 가정을 명시한 모형으로 제시해야 한다.

시나리오에 순환 논증이 있는지도 살피면 좋다. 메커니즘은 유용해서 보존되었고, 보존되었다는 사실로 유용함을 알 수 있다는 논리를 생각해 보자. 주장이 스스로를 뒷받침하고 있다. 순환을 벗어나려면 가정한 이득과 비용, 이점이 사라지는 조건을 독립적으로 측정해야 한다. 그래야 모형이 구체적으로 틀릴 수 있다.

기억의 실습 예에서는 환경의 급격한 변화가 검증 조건이 된다. 모형이 이점의 소멸을 예측하는데 모든 매개변수에서 이점이 유지된다면 구현과 비용 정의를 살펴야 한다. 한 전략에 항상 추가 정보를 주도록 비교를 설계했을 수 있다. 이런 기술적 오류로도 설득력 있지만 잘못된 적응 이야기가 생길 수 있다.

## 같은 나뭇잎에도 두 역사가 가능하다

실습 계통수에서는 종 A와 B에서 메커니즘 M이 발견되고 C와 D에는 없다. A와 B의 공통 조상에서 M이 생겼다는 이야기가 가능하다. 더 일찍 생긴 뒤 다른 가지에서 사라졌다는 이야기도 가능하다. 관찰한 끝 가지는 같아도 역사적 변화는 다르다. 선택하려면 관련 종과 구조, 가능한 전이에 관한 추가 정보가 필요하다. M과 체험의 관계 자체가 가설이라면 역사 재구성의 확실성이 체험의 독립 증거가 되지는 않는다.

### 여기까지 확인한 것

메커니즘의 역사와 체험의 역사를 하나의 결론으로 합치지 않았다. 특정 기억의 이점을 설명하면서도 어떤 조직에서 그 기억이 체험의 측면을 갖는지는 열린 질문으로 남길 수 있다.

## 실습 19 진화 논거의 관계도

실습 목표는 각 추론 단계의 가정을 명시하여 진화 논거를 구성하는 것이다.

발달과 진화적 기원, 체험의 존재를 구별한다. 메커니즘과 관찰 특징, 가정한 조상은 논거의 서로 다른 요소다.

예상 소요 시간은 25–35분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 그림 43의 실습 계통수를 사용한다. 끝 가지에서 관찰한 것과 두 역사 가설을 따로 적는다.
2. 각 설명에서 메커니즘이 생기고 사라졌다고 가정하는 지점들을 밝혀라. 어느 설명도 조상을 직접 관찰한 것처럼 제시하지 마라.
3. 몇 신호의 기억이 도움이 될 수 있는 환경과 전환을 늦추는 환경을 기술한다. 저장과 처리 비용도 명시한다.
4. 메커니즘과 체험의 관계를 별도의 문장으로 더하고 이론 가설로 표시한다. 각 결론을 바꿀 수 있는 정보를 밝힌다.

보존할 것: 관찰, 역사적 설명, 이점의 모형, 체험에 관한 별도 가정을 구분한 지도.

비슷한 행동이 같은 구조를 보장하지는 않는다. 공통 분자가 있다고 전체 계의 기능도 같은 것은 아니다.

탐구를 이어 가려면 이 장의 발표된 비교 논거 하나를 고른다. 어떤 정보가 기능과 구조, 계통 관계에 속하는지 확인하고 가까운 역사 가설들을 구별할 관찰을 제안한다.

해설과 답은 부록 F 제19장 참조, 306쪽

## 자기 점검 문항

1. 메커니즘의 적응상 이점이 체험의 존재를 입증하는가?
2. 가까운 계통의 종들이 완전히 독립적인 확인을 주지 못하는 까닭은 무엇인가?
3. 여러 기원 가설을 유지해야 하는 까닭은 무엇인가?

사례의 동역학과 발달, 진화를 구별했다. 다음 장에서는 특정한 물리 조직이 왜 체험의 측면을 가질 수 있는가라는 남은 질문을 다룬다.

제20장

# 발생을 설명하는 데 필요한 것

앞 장들은 과정이 시간 속에서 생기고 능력이 달라지는 방식을 설명했다. 이제 체험 자체가 존재하는 근거를 묻는다. 성공적인 시각 모형은 응답과 오류를 예측할 수 있다. 그래도 저자는 계산과 체험 사이에 어떤 관계를 제안하는지 밝혀야 한다. 이 요구는 이 책의 고유 가설을 준비한다. 강한 가정을 하면서도 그에 따르는 요구와 가능한 반론을 유지할 수 있게 한다.

## 남은 질문의 구체화

상세한 시각 모형을 만들었다고 하자. 인식과 기억, 확신도, 응답 준비를 설명하고 매개변수를 바꾸면 인간과 비슷한 오류도 나온다. 이를 체험의 설명이라고 부르려면 무엇이 더 필요할까? 답은 물리적 과정과 체험 사이에 어떤 관계를 받아들이고 어떤 자료로 그 관계를 뒷받침하느냐에 달려 있다.

David Chalmers는 인지 기능의 설명과 그 기능 수행에 왜 주관적 체험이 수반되는가라는 질문을 구분했다. 이 논거는 철학적 문제를 제기하며, 별도의 비물리적 실체를 실험적으로 입증한 것은 아니다. 그의 후속 결론을 받아들이지 않으면서도 메커니즘에서 체험으로 넘어가는 지점을 정확히 명시하라는 요구는 사용할 수 있다. [2]

이 장에서는 발생의 존재론적 의미, 곧 어떤 과정에 왜 체험의 측면이 있는지에 관심을 둔다. 망이 언제 켜졌는지, 아이가 언제 과제를 배웠는지, 종의 역사에서 메커니즘이 언제 생겼는지에 관한 정보가 곧바로 답이 되지는 않는다. 그래도 가능한 답을 제한하고 일반적인 생각을 연구 프로그램으로 바꾸는 데 도움이 된다.

## 존재를 설명하는 것과 차이를 설명하는 것

두 과제를 구별하자. 첫째는 계의 한 상태에는 체험이 있고 다른 상태에는 없는 이유다. 둘째는 존재하는 체험이 왜 바로 그 질을 갖는가이다. 의식이 있을 전반적인 가능성만 예측하는 설명은

소리와 색, 신체 감각의 차이를 아직 설명하지 못한다. 반대로 색 구별 모형은 응답 사이 관계를 정확히 기술하면서도 체험 여부의 질문을 남길 수 있다.

임의의 문자로 나타낸 세 감각의 공간을 생각해 보자. 참여자는 첫 감각이 셋째보다 둘째에 더 가깝다고 본다. 모형이 이 유사성 구조를 재현한다. 자료의 특정 관계를 설명했으므로 의미 있는 성과다. 그러나 세 거리가 같다고 모형의 내부 상태도 같은 질을 가진다고 확립되지는 않는다. 공간의 어떤 성질을 측정했으며 선택한 대응이 왜 설명력을 갖는지 알아야 한다.

다음 단계는 변화를 예측하는 것이다. 개입으로 메커니즘의 연결 하나가 달라지면 두 감각의 유사성과 통합, 구별의 이용 가능성 중 무엇이 달라져야 할까? 측정 전의 예측은 결과를 본 뒤 보기 좋은 그림을 맞추는 것보다 강하다. 선택성도 중요하다. 모형이 관계들의 독립성을 주장한다면 하나를 바꾸어도 다른 관계들은 남아야 한다.

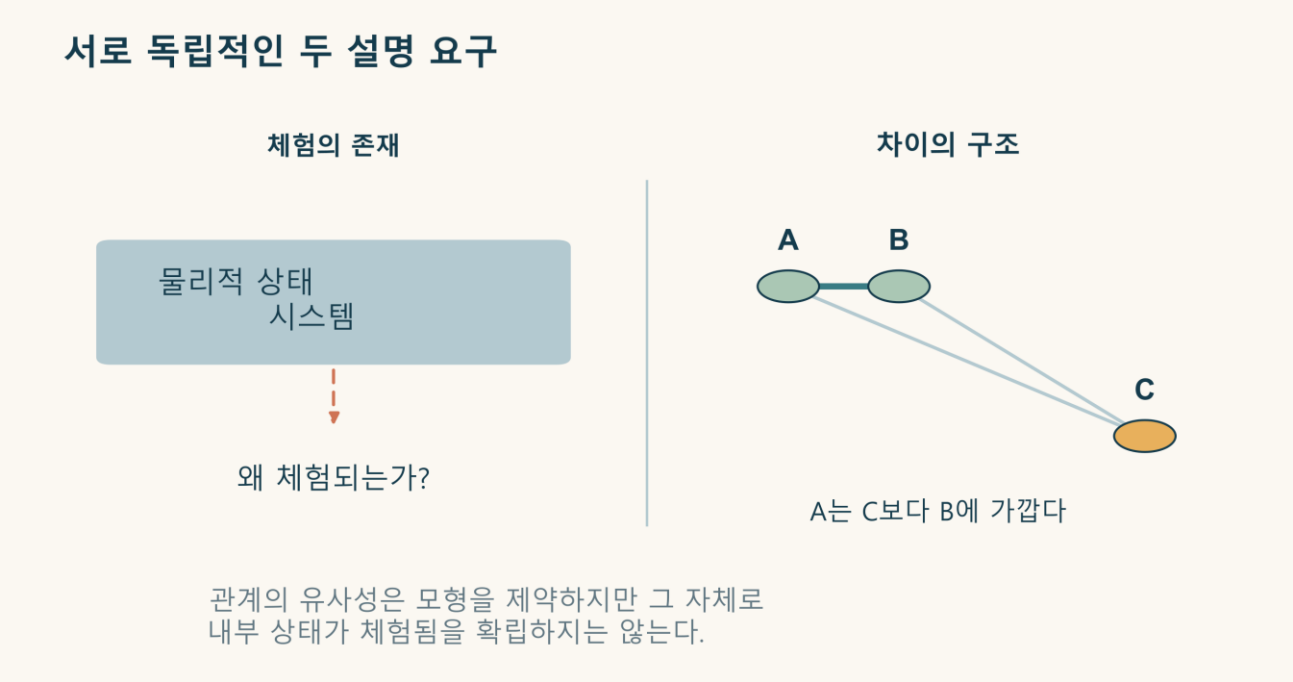


그림 44. 존재의 설명과 질의 설명. 왼쪽은 체험의 존재를, 오른쪽은 임의 기호로 나타낸 세 감각 사이의 차이 구조를 묻는다. 오른쪽 기하는 정해진 가까움의 관계만 보여 준다. 수치 거리나 감각 양식과의 대응은 없다. 이 관계를 재현해야 한다는 것은 모형에 대한 제약이지만 왼쪽 과제를 자동으로 해결하지는 않는다.

## 통일성과 소유감

한 사례에 소리와 보이는 움직임, 신체 감각이 함께 있을 수 있다. 왜 이들이 같은 사례에 속하며 왜 계의 다른 과정 일부는 포함되지 않는지 설명해야 한다. 기록을 단순히 더하는

것만으로는 부족하다. 한 부분의 변화가 나머지와 이용 가능한 전체에 영향을 주게 하는 관계를 정해야 한다.

체험의 귀속도 구체화해야 한다. 최소한의 의미에서는 내용이 어떤 관점을 기준으로 조직되는가를 말한다. 더 강한 의미에서는 사람이 그 상태를 자신의 것으로 알아본다. 더 강한 경우에는 시간에 걸쳐 지속되는 사건 참여자로서 자기 개념이 있다. 모든 감각에 명시적인 «나» 개념이 필요하다고 가정하는 이론은 의식의 정의에 숨길 수 없는 추가 요구를 떠맡는다.

사고실험으로 확인해 보자. 선율에 몰두한 사람은 자신을 분석하지 않는다. 나중에는 체험을 일인칭으로 기술할 수 있다. 상세한 자기 기술이 처음 사례에도 있었을까, 보고할 때 생겼을까? 이야기 자체만으로 추가 설명 없이 결정할 수는 없다. 이 사례는 지각 모형과 귀속 모형, 성찰 모형을 먼저 구별한 뒤 상호작용을 연구해야 하는 이유를 보여 준다.

## 과정과 체험 사이에 가능한 관계

한 설명은 동일성을 가정한다. 체험은 다른 측면에서 기술한 특정 물리적 과정이라는 것이다. 그러면 그 과정을 선택한 근거와 그 구조가 체험 구조에 대응하는 이유를 밝혀야 한다. 이름을 바꾸는 것만으로는 부족하다. 설명의 성공은 일관된 예측 체계와 인과 자료, 임의적인 예외가 없다는 데 달려 있을 것이다.

다른 설명은 물리 조직과 체험 사이의 추가 대응 원리를 도입한다. 원리를 명시하고 어떤 사례들을 구별하는지 설명해야 한다. 이미 아는 행동 이상을 예측하지 않는다면 어떻게 검증하며 경쟁 원리보다 왜 선호하는지가 문제가 된다. 연구를 금지하는 요구가 아니라 새 정보를 얻을 경로를 보여 달라는 요구다.

셋째 방식은 체험의 현상성에 대한 최초 기술을 재검토한다. 체험에 속하는 것처럼 보이는 성질 일부가 시스템이 자기 작동을 표상하는 방식과 연결될 수 있다. 그렇다면 어려움을 착각이라고 선언하는 것만으로 충분하지 않다. 체험에 관한 구체적 판단, 그 판단의 안정성, 오류, 다른 자료와의 연결을 설명해야 한다. 여기서도 강한 답은 질문의 이름만 바꾸는 데 그치지 않고 검증 가능한 차이를 만들어야 한다.

이 방식들은 논리적 전략으로 제시한 것이다. 똑같이 입증되었다고 보거나 사고실험 하나로 승자를 정하지 않는다. 연구 교재에서는 각 모형이 정확히 어떤 단계를 밟으며 어떤 설명 과제가 남는지 알아보는 법을 익히는 편이 유용하다.

## 자신의 가설을 세우는 방법

제한된 연구 가설을 생각해 보자. 시각 내용과 신체 내용을 안정적으로 통합하는 것은 계가 그 둘을 변할 수 있는 하나의 공통 원인에서 추론하는지에 달려 있다는 가설이다. 특정 사례의 조직에 관한 주장이다. 바로 이 통합을 교란하면 두 내용을 각각 구별하는 능력은 남으면서 함께 귀속되는 방식이 달라지리라 예측한다. 이는 실습용 정식화이며 확립된 보편 법칙의 선언은 아니다.

검증하려면 모형의 공통 원인을 정의하고 각각의 구별이 유지된다는 독립 지표와 개입을 찾아야 한다. 이어 결과를 주의나 기억만으로 설명하는 대안과 비교한다. 모든 모형의 예상이 같으면 조건을 바꿔야 한다. «공통 원인»이라는 말만으로 기존 기술보다 깊은 가설이 되지는 않는다.

긍정적 결과도 우선 통합 메커니즘을 뒷받침한다. 이를 체험의 존재에 대한 설명이라고 부르려면 대응 원리를 더하고 근거를 제시해야 한다. 가능한 과학 연구는 이렇게 연결된 여러 성취로 이루어진다. 작은 실험마다 존재론적 질문 전체를 즉시 해결하라고 요구하면 유용한 중간 제약들을 잃는다. 그 제약들을 최종 해답으로 제시하는 것도 오류다.

좋은 설명은 적용 범위도 밝혀야 한다. 시각 통합의 가설이 꿈이나 신체 통증, 추상적 사고에 대해서는 아무 주장도 하지 않을 수 있다. 이 제한을 미리 적었다면 괜찮다. 검증된 개별 주장에 «모든»이라는 말만 붙여 보편성을 얻을 수는 없다.

범위를 넓히려면 새로운 연결 가정이 필요하다. 서로 다른 내용이 하나의 조직 원리를 쓰되 다른 요소로 구현한다고 제안할 수 있다. 그러면 무엇이 공통으로 남으며 각 영역의 어떤 변화를 유사한 것으로 볼지 정해야 한다. 결과 뒤에 대응을 임의로 고르면 검증 가능성을 잃는다. 미리 정하면 개별 메커니즘에서 더 일반적인 체험 이론으로 나아갈 경로가 생긴다.

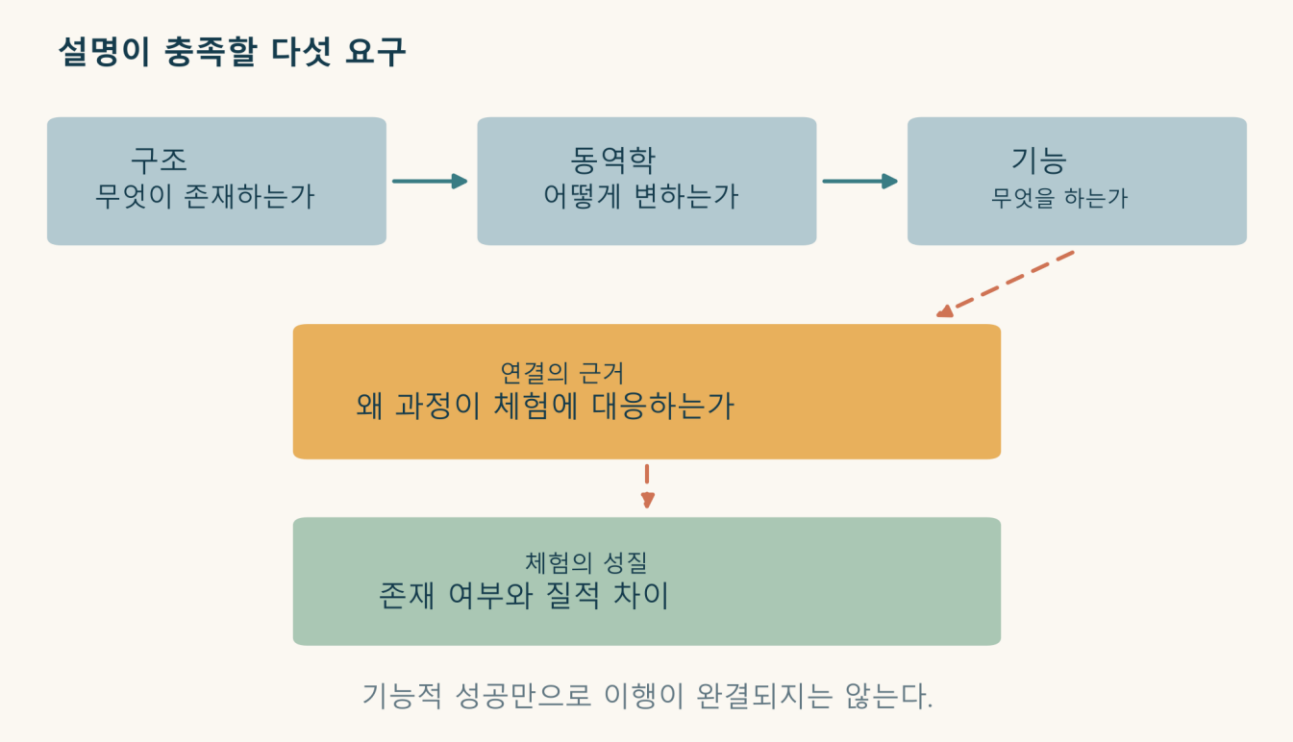


그림 45. 구조에서 체험의 설명으로. 설명 점검의 다섯 질문은 무엇이 존재하는지, 어떻게 변하는지, 어떤 기능을 하는지, 체험의 어떤 성질을 설명하는지, 과정이 왜 그 성질들과 연결되는지다. 점선은 별도의 근거가 필요한 이행을 표시한다. 기능 설명이 체험의 존재나 질적 구조를 자동으로 확립하지는 않는다.

## 다섯 문장은 서로 다른 설명 역할을 한다

연결된 요소들을 기술하면 구조 안에 무엇이 있는지 알 수 있다. 그 갱신 규칙은 역동을 보여준다. 대상을 성공적으로 구별한다는 것은 기능을 기술한다. 보고된 질들 사이의 관계는 체험에 관한 요구를 정한다. 대응 원리는 선택한 조직을 체험에 충분하거나 중요한 것으로 보는 이유를 설명한다. 처음 세 항목만 남기고 출력을 체험이라고 불러도 마지막 전환은 사라지지 않는다. 과정과 체험의 명시적 동일성을 제안한다면 그 주장에는 구체적인 요구가 따른다. 대응의 구조, 적용 한계, 대안을 구별할 수 있는 귀결을 제시해야 한다.

### 여기까지 확인한 것

좋은 기능 모형을 얻은 뒤에도 어떤 설명이 더 필요한지 정했다. 이제 이미 얻은 결과를 유지하면서 남은 질문을 정확히 정식화한다.

## 실습 20 설명 점검하기

실습 목표는 한 설명이 충족해야 할 요구들을 따로 분석하고 유력한 경쟁 설명을 찾는 것이다.

제5장의 필요성과 충분성, 상관, 가정한 동일성을 활용한다. 체험의 존재와 질, 통일성, 귀속을 구별한다.

예상 소요 시간은 30–45분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 제13–15장에서 구체적인 설명 하나를 고른다. 구조와 동역학, 기능, 설명하는 체험의 성질, 그 사이 관계에 대해 각각 한 문장을 적는다.
2. 각 문장 곁에 근거를 정의, 자료, 계산, 이론 가정 가운데 하나로 표시한다.
3. 같은 기능적 결과 일부를 유지하는 경쟁 설명을 고른다. 어떤 추가 요구가 두 설명을 구별하는지 밝힌다.
4. 구별에 도움이 될 관찰을 정식화한다. 아직 정할 수 없다면 어떤 방법이나 추론이 더 필요한지 정확히 기술한다.

연결된 다섯 주장과 경쟁 설명, 정확히 남은 질문을 기록한다.

새 구현에서 체험이 있는지 아직 확립하지 못했어도 질 사이 관계를 설명한 것은 의미 있는 성과다.

탐구를 이어 가려면 제한된 대응 원리를 직접 제안한다. 단순한 조절기와 구조는 다르지만 같은 응답을 내는 계에 적용해 검증한다.

해설과 답은 부록 F 제20장 참조, 307쪽

### 자기 점검 문항

1. 색 구별의 설명이 체험의 존재에 관한 질문을 해결하지 못할 수 있는 까닭은 무엇인가?
2. 체험을 설명하려면 반드시 추가 실체를 도입해야 하는가?
3. 최종 설명이 아직 없다면 무엇을 진전으로 볼 수 있는가?

앞으로 만들 후보의 구조를 고르기 전에 체험의 여러 담지체에 관한 자료를 살핀다. 기준을 구체적인 메커니즘과 조건에 적용해야 하는 이유를 보여 줄 것이다.

제6부

# 체험이 존재할 수 있는 기반

제21장

# 동물의 의식과 자기점검

이론적 원리는 익숙한 인간에서 다른 생물로 옮겨도 의미를 유지해야 한다. 인간의 과제를 그대로 반복하면 해당 생물에 맞지 않는 언어와 동작, 감각 채널을 요구할 수 있다. 동물 연구를 기억과 내용, 손상 뒤 상태에 관한 개별 논거로 살펴보자. 그 힘은 인간의 말과 겉으로 닮은 응답이 아니라 적절한 측정과 대안 통제의 결합에서 나온다.

## 동물에게 던지는 질문

동물의 체험 가능성을 물을 때는 여러 목적을 구별해야 한다. 현상적 의식은 체험이 있음을 뜻한다. 자기관련성은 정보를 자기 신체나 행동, 상태와 연결하는 것이다. 메타인지 조절은 인지 과정의 신뢰도에 대한 정보를 사용한다. 명시적 성찰은 자기 근거를 의도적인 검토 대상으로 삼는다. 일부 과제에서 상호작용해도 서로 바꿔 쓸 수 있는 능력은 아니다.

불확실한 선택을 피하는 동물은 기억의 질을 평가하거나 학습한 외부 단서를 사용하거나 더 단순한 강화 규칙을 따를 수 있다. 손상된 신체 부위를 보호하는 동물은 다른 기능들을 보인다. 어떤 관찰도 곧바로 인간이 자신을 생각하는 방식에 관한 주장으로 옮겨서는 안 된다. 구체적인 근거와 열린 대안의 양상이 필요하다.



그림 46. 서로 다른 생물과 연구 질문. 새, 문어, 물고기, 벌을 그린 것이다. 크기 척도는 개념적이다. 체험의 가능한 기반들은 구체적인 특징과 연구를 토대로 비교해야 한다. 이 페이지에서 동물의 위치와 크기는 의식의 척도를 정하지 않는다.

## 전반적 평가와 구체적 실험

2024년 뉴욕 선언은 다른 포유류와 조류에 체험을 귀속할 강한 과학적 근거가 있으며 모든 척추동물과 많은 무척추동물에 체험이 있을 현실적인 가능성이 있다고 말한다. 이는 서명자들의 입장 표명이다. 별도의 실험도 아니며 모든 종에 같은 근거가 있다는 뜻도 아니다. 구체적인 자료의 차이를 유지하며 연구의 불확실성을 가늠하는 지침으로 사용할 수 있다. [33]

전반적인 평가는 많은 연구 흐름에 근거할 수 있지만 개별 실습 분석은 검증 가능한 범위에 있어야 한다. 아래에서는 기억과 시각 내용, 손상 관련 상태라는 다른 대상을 다룬 원저 몇 편을 살핀다. 동물들이 의식 점수 하나를 놓고 경쟁하는 것이 아니다. 각 연구는 자체적인 설명 범위를 제한한다.

## 기억과 시각 내용

햄프턴의 연구에서 두 마카크원숭이는 기억 검사를 받아들이거나 거절할 수 있었다. 선택은 기억 조건에 따라 달랐고 이후 정확도와 연결되었다. 두 동물이 기준을 충족한 정도에는 차이가 있었다. 이 연구는 자신의 기억에 접근할 수 있는지를 통제하는 능력을 논의할 근거를 주었다. 말로 하는 성찰을 보여 준 것은 아니며, 비슷한 모든 과제를 설명할 수 있는 모든 연합적 설명을 한꺼번에 배제한 것도 아니다. [34]

Nieder, Wagener, Rinnert의 연구에서 까마귀 전뇌 뉴런의 활동은 약한 시각 자극의 유무를 보고하는 행동과 연결되었다. 저자들은 이를 감각 의식의 신경 상관지표로 해석했다. 연구는 비교의 근거를 넓히지만 조류 체험의 모든 성질을 정하지 않으며 인간의 생애적 자기의식이 존재함을 확립하지도 않는다. [35]

두 연구는 다른 질문에 답한다. 첫 연구는 이용 가능한 지식에 따른 결정 조절을 살피게 한다. 둘째 연구는 신경 활동을 보고되는 시각 내용의 구별과 연결한다. 차이가 있기 때문에 함께 보는 것이 유용하다. 둘을 모두 «자기의식»이라고만 부르면 각 효과의 메커니즘을 이해할 수 없게 된다.

## 손상과 불쾌한 체험

Crook은 문어를 대상으로 손상을 주는 자극과 연결된 장소의 회피, 자극 뒤 완화와 연결된 장소의 선호를 연구했다. 저자는 행동 및 신경생리학 결과를 통증의 정서적 성분을 가정할 근거로 해석했다. 연구한 조건에서 체험되는 불쾌한 상태에 관한 논거다. 타자의 퀄리아를 직접 측정한 것이나 모든 두족류의 체험이 같다는 증거는 아니다. [36]

Barr와 Ellwood는 2024년에 바닷게가 불쾌한 자극을 피하는 것과 밝은 빛을 피하는 것 사이에서 어떻게 선택하는지 연구했다. 경쟁하는 조건에 따른 선택의 변화는 유연한 평가의 예측과 일치했다. 저자들은 결과가 통증과 양립한다고 보았다. 직접 보여 준 더 좁은 결론은 서로 다른 선택 근거 사이의 절충에 관한 것이다. [37]

2026년에 발표한 Manzi와 동료들의 연구에서 집귀뚜라미는 손상을 주는 열 자극 뒤 대조 자극 뒤보다 더 오래 더듬이를 손질했다. 스트레스가 덜하거나 더하도록 의도한 조건의 변경은

유의한 효과를 내지 않았다. 이 연구는 목적 있는 자기보호와 통증 유사 상태의 가능성을 논의할 근거를 강화한다. 메타인지 분석이나 명시적 성찰을 보여 준 것은 아니다. [38]

손상의 감지와 조직된 회피, 불쾌한 체험을 구별해야 한다. 단순한 방어 반응보다 대안을 통제한 지속적인 선호 변화가 더 많은 정보를 준다. 그래도 행동의 복잡성 자체가 통증의 독립적인 측정 단위는 아니다. 관찰들의 조합이 가까운 경쟁 메커니즘보다 가정한 상태에 왜 더 잘 들어맞는지 설명해야 한다.

## 근거의 비교 지도

표는 살펴본 자료 유형을 모았다. 검증의 방향과 일반화의 한계를 나타내며 종들에게 의식의 수치 확률을 부여하지 않는다.

| 집단 | 이 분석에서 다룬 근거 | 중요한 미해결 질문 |
| --- | --- | --- |
| 포유동물 | 강한 근거에 대한 전반적 평가와 마카크 기억 실험 | 어떤 조절 형태가 과제 사이에 일반화되는가 |
| 조류 | 전반적 평가와 까마귀의 신경 연구 | 다른 내용과 종에도 얼마나 일반화되는가 |
| 다른 척추동물 | 선언에서 체험의 현실적인 가능성을 인정함 | 어떤 구체적 메커니즘과 지표가 비교 가능한가 |
| 두족류 | 문어의 선호와 신경생리학의 결합 | 가정한 체험의 질과 조직은 어떠한가 |
| 십각 갑각류 | 해안 게의 선택 절충 | 체험되는 상태의 평가에 대한 어떤 대안이 배제되었는가 |
| 곤충 | 집귀뚜라미의 특정 부위를 향한 자기보호 | 방어 조절과 체험의 가능성은 어떻게 연결되는가 |

다른 척추동물의 행에서는 개별 실험을 분석하지 않았다. 이는 선택한 자료의 범위일 뿐 체험이 없다는 결론으로 옮길 수 없다. 마찬가지로 추가 비교 근거 없이 한 종의 자료를 집단 전체에 일반화해서는 안 된다. 연구의 양과 연구하는 생물의 성질은 다르다.




그림 47. 질문과 증거 공백의 행렬. 이 행렬은 이 장에서 분석한 마카크의 기억, 까마귀의 시각 보고, 문어의 회피와 완화, 게의 선택 상충관계, 귀뚜라미의 자기보호 연구에만 해당한다. 점은 연구 대상을 나타내며 체험이 입증되었다는 뜻이 아니다. 대시는 관련 분석이 없음을 뜻하며 동물의 능력이 없다는 뜻이 아니다. 한 종의 자료가 집단 전체를 대표하지 않는다.

## 결론의 근거를 강화하는 방법

강한 근거의 양상은 제약이 다른 여러 자료의 수렴으로 만든다. 학습의 전이와 결과 가치 변화에 대한 민감성, 그에 대응하는 인과적 변화가 관찰되면 단순한 고정 규칙으로 모두 설명하기 어려워진다. 그래도 대안은 구체화해야 한다. «그저 본능이다»는 말은 입력과 내부 상태, 행동 변화 규칙을 밝히기 전에는 메커니즘이 아니다.

«동물이 이해한다»는 표현도 구체화해야 한다. 정확히 무엇이 표상되는가? 새 상황에서 지식이 어떻게 드러나는가? 외부 단서를 없애면 무엇이 달라지는가? 자기 오류에 대한 가정한 평가가 결과를 받기 전에 행동을 바꿀 수 있는가? 이런 답으로 설명이 얼마나 풍부한지에 관한 논쟁을 진전시킬 수 있다.

명시적 성찰을 주장하려면 불확실성 통제를 주장할 때보다 더 높은 기준을 충족해야 한다. 자신의 근거를 검토하고 수정한다는 것을 보이면서, 적용 가능한 더 좁은 규칙에 따른 설명을 배제해야 한다. 그런 증거가 없다고 해서 체험의 가능성이 반박되는 것은 아니다. 다만 구체적인 성찰 능력을 부여하는 데 한계가 생긴다.

비교할 때는 과제 자체를 똑같이 이해할 수 있는지 확인해야 한다. 한 종에는 익숙한 선택 방식을 주고 다른 종에는 어려운 동작을 요구하면 정확도 차이에 관심 능력과 검사 구조가 섞인다. 신호의 이용 가능성과 동기, 응답 동작의 수행 가능성을 통제해야 한다. 때로는 외부 행동이 달라도 정보와 선택의 관계를 같게 하는 것이 더 공정한 비교다.

버튼 누르기와 통로 선택이라는 두 학습 절차를 생각해 보자. 둘 다 이용 가능한 정보의 질에 따라 결정이 얼마나 달라지는지 평가할 수 있다. 비교해야 할 것은 동작의 겉모습이 아니라 이 의존 관계다. 이런 일반화에도 근거가 필요하다. 하지만 이는 문자 그대로 같은 검사가 언제나 같은 방식으로 메커니즘을 시험하는 것은 아닌 이유를 보여 준다.

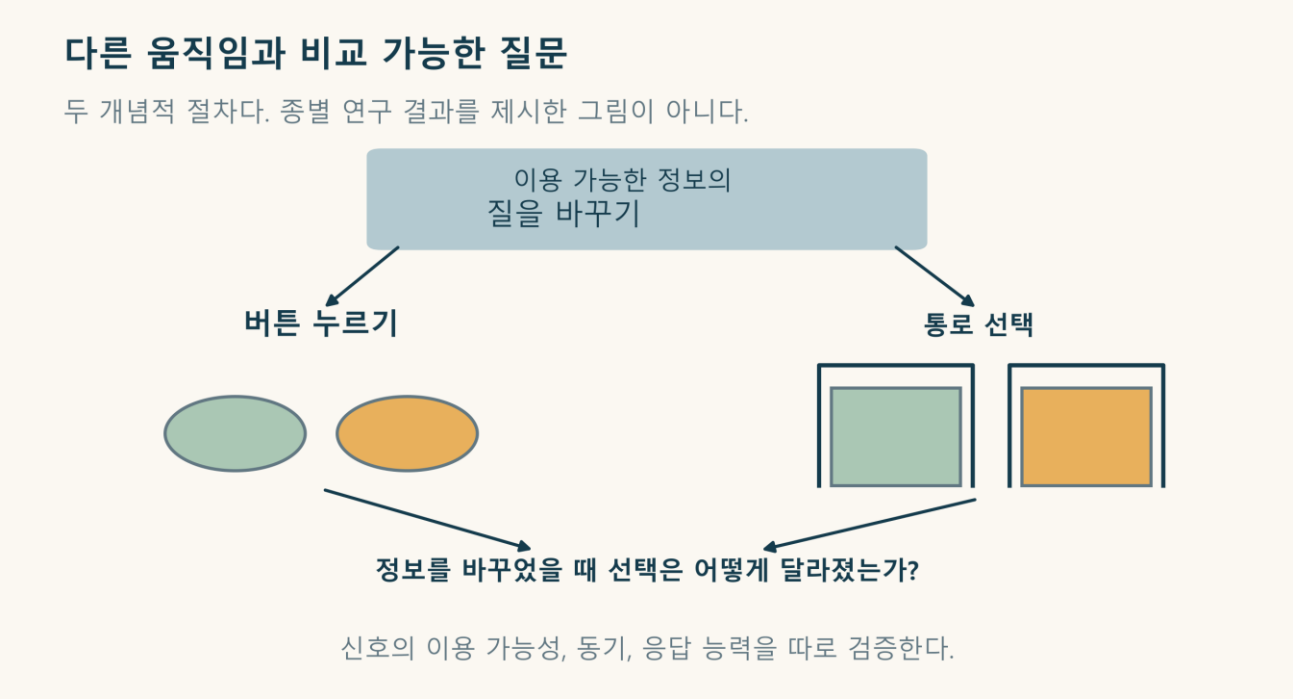


그림 48. 외부 응답이 다른 비교 가능한 과제. 버튼 누르기와 통로 선택으로 이용 가능한 정보에 따라 결정이 달라지는 비슷한 관계를 검증할 수 있다. 비교할 때에는 이 관계가 서로 비교 가능한지 근거를 제시하고 감각, 동기, 운동의 요구를 따로 검증해야 한다. 외부 결과가 같다고 내부 메커니즘이 같아지거나 하나의 의식 척도가 생기지는 않는다.

## 어려운 시험을 거절하는 행동이 더하는 것

동물은 잘 기억하기 어려운 조건 뒤에 시험을 더 자주 거절하며, 받아들인 시험에서는 더 정확하다. 이는 기억 접근 가능성의 점검과 양립한다. 경쟁 설명은 지연 같은 외부 특징에 따른 선택을 가정할 수 있다. 둘을 구별하려면 비슷한 지연에서 기억의 질을 바꾸거나 다른 적절한 조건을 도입해야 한다. 따라서 동물마다 구체적인 기준과 자료가 중요하다. «메타인지»라는 일반 명칭이 이 분석을 대신하거나 동물에게 성인의 언어적 성찰을 부여하지는 않는다.

### 여기까지 확인한 것

개별 연구들은 서로 다른 질문에 답한다. 근거의 양상을 정리하면 연구한 성질과 아직 모르는 것의 차이를 유지하면서 함께 볼 수 있다.

## 실습 21 연구한 한 종의 양상

실습 목표는 결론을 생물 집단 전체에 옮기지 않으면서 연구 하나의 양상을 정리하는 것이다.

필요한 것: 연구 명세서와 기능적 추론 및 체험에 관한 추론의 구분. 과학적 입장을 선언하는 것과 구체적인 실험은 지위가 다르다.

예상 소요 시간은 30–45분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 기억이나 시각 내용, 손상 뒤 상태를 다룬 이 장의 연구 하나를 고른다. 첫 분석에서는 이 책의 내용으로 항목을 채운다.
2. 종과 과제, 측정, 대조 조건, 주된 결과를 적는다. 이용 가능한 설명에 동물 수가 없으면 누락으로 표시한다.
3. 구체적인 대안 메커니즘과 그 메커니즘의 예상이 달라지는 조건을 제안한다.
4. 근거가 강한 결론 하나와 이 연구가 아직 허용하지 않는 더 넓은 결론 하나를 적는다.

의식의 전체 점수 없이 연구 기록표 하나와 검증 가능한 대안을 남겨 둔다.

이 장에서 어떤 성질의 연구를 다루지 않았다는 것은 그 생물에 그 성질이 없음을 관찰했다는 뜻이 아니다.

탐구를 이어 가려면 원논문에서 표본과 제외 규칙, 개별 결과를 확인한다. 제안한 대조 조건이 실제로 포함되어 있는지 검증한다.

해설과 답은 부록 F 제21장 참조, 308쪽

## 자기 점검 문항

1. 통증의 가능성을 인정하는 것이 명시적 성찰도 인정한다는 뜻이 아닌 까닭은 무엇인가?

2. 한 종의 결과를 가까운 계통의 모든 종에 일반화할 수 있는가?

3. 단순한 규칙으로 설명한다면 어떻게 다루어야 하는가?

이제 뉴런이 없는 계에서 어떤 기능이 가능한지 검증한다. 넓은 의미의 기억과 적응을 체험 여부에 대한 추가 가설과 구별하는 데 도움이 된다.

제22장

# 적응과 체험 사이의 경계

신경계 밖에서도 복잡한 조율과 과거의 영향이 나타난다. 이 자료의 의미를 이해하기 위해 측정 가능한 반응과 몇 가지 가능한 메커니즘에서 시작한다. 반응이 약해지는 것은 특정 신호에 대한 습관화 때문일 수도, 전반적 소진 때문일 수도 있다. 새 신호와 휴지기를 사용하면 둘을 구별할 수 있다. 체험의 질문에는 별도의 근거를 갖춘 다음 단계가 필요하다.

## 뉴런 없는 시스템이 중요한 이유

체험의 조건을 설명하려면 처음부터 탐색을 인간과 비슷한 생물로 한정할 수 없다. 동시에 유용한 모든 반응을 퀄리아의 징후로 볼 수도 없다. 식물과 단세포 생물, 뉴런 없는 다른 계의 연구는 신호 처리와 이력 보존, 학습, 가정한 체험 여부를 구별하게 하므로 유용하다.

첫 질문은 측정 가능한 현상에 관한 것이어야 한다. 자극을 반복한 뒤 반응이 달라졌는가? 변화가 남았는가? 특정 신호에 의존하는가? 생물이 사건 사이의 관계를 이용할 수 있는가? 답을 얻은 뒤에야 발견한 메커니즘을 체험과 연결할 근거를 묻는다. 이 순서는 연구를 열어 두면서 과도한 결론을 막는다.

## 신호 전달과 조율

도요타와 동료들은 애기장대 Arabidopsis에서 손상과 글루탐산 신호, 칼슘 농도 변화의 전파, 먼 부위의 방어 반응 활성화 사이의 연관을 보였다. 이는 생물체 내부의 장거리 신호 전달 메커니즘을 다룬 원저 연구다. 일부 분자와 신호 전달 방식이 알려진 신경 과정과 비슷하다고 해서 전체 조직의 동일성이나 체험으로서의 통증이 존재함을 확립한 것은 아니다. [39]

여러 저장조로 이루어진 계를 생각해 보자. 한 저장조의 변화가 신호를 만들어 다른 저장조의 밸브를 닫는다. 국소 감지와 정보 전달, 조율된 방어 행동이 있다. 신호가 이전 자극에 의존한다면 이력의 영향도 생긴다. 이 성질들은 자기 표상이나 불쾌한 체험을 계에 부여하지 않고도 연구할 수 있다.

이 비유가 식물 전체를 설명하지는 않는다. 신호 전달이 있다는 것만으로 어떤 추가 성질이 구현되었는지 정해지지 않는다는 논리적 한계를 보여 준다. 더 강한 주장에는 과정 사이 상호작용과 선택의 유연성, 내부 상태에 대한 자료, 체험과 연결할 이론적 근거가 필요하다. «방어»는 기능을 기술하는 말이며 의식적으로 자신을 지키려는 의도를 뜻할 필요는 없다.

## 과거의 영향으로서의 기억

최소한의 기능적 의미에서 기억은 현재 입력이 같아도 이력 때문에 현재 반응이 달라지는 곳에 있다. 이 기억은 계의 상태에 분산되어 사건에 관한 별도 기록을 포함하지 않을 수도 있다. 이 넓은 의미와 과거로 돌아가는 체험으로서의 회상을 구별해야 한다. 같은 단어를 쓴다고 기능이 같아지지는 않는다.

설명을 위한 모형을 만들자. 처음에는 영인 습관화 변수가 있다. 신호를 제시할 때마다 일까지 남은 거리의 오분의 일만큼 일에 가까워진다. 반응은 일에서 현재의 습관화 값을 뺀 차이다. 초기 반응은 일, 십분의 팔, 백분의 육십사 등으로 이어진다. 쉬는 동안 변수는 매 단계 오분의 일씩 줄어들므로 반응이 회복된다.

수치는 단순성을 위해 직접 정한 것이다. 모형은 보존된 상태에서 반응의 감소와 휴지기 뒤 회복이 어떻게 생기는지 보여 준다. 뉴런이나 언어적 믿음, 모형 내부의 관찰자는 없다. 반응 곡선 하나로 이 규칙이 실제 생물학 메커니즘과 같은지는 알 수 없다. 선택하려면 추가 개입과 측정이 필요하다.

두 신호와 각각에 대응하는 별도의 습관화 변수를 추가하자. 그러면 첫 신호의 반복은 그 신호에 대한 반응만 줄이고, 둘째 신호는 여전히 효과를 낸다. 대신 공통 자원 변수 하나를 사용하면 두 반응이 모두 줄어든다. 이처럼 새 신호로 검증하면 특정한 변화와 전반적인 소진을 구별하는 데 도움이 된다. 그래도 이것이 체험의 유무를 검사하는 방법은 아니다.

## 점균에서 밝혀진 것

Boisseau, Vogel, Dussoutour는 뉴런이 없는 생물인 점균 Physarum polycephalum을 연구했다. 화학물질에 반복 노출하면 반응이 약해지고 쉬는 시간을 둔 뒤 회복되었다. 변화의 특이성도

분석했다. 저자들은 이를 단순한 학습 형태인 습관화로 해석했다. 이 연구는 신경계 없이도 이런 가소성이 가능함을 보여 준다. 퀄리아, 자기의식, 명시적 성찰을 측정한 것은 아니다. [40]

결과를 인정하는 일과 의미를 넓히는 일을 구별하면 좋다. 습관화 기준을 충족했다면 낯선 계라는 이유만으로 학습을 부정할 필요는 없다. 그러나 학습을 인정했다고 체험도 귀속해야 하는 것은 아니다. 다음 과학적 단계는 이력 보존의 메커니즘과 전이 규칙, 가소성의 한계를 알아내고 특정 의식 이론의 요구와 비교하는 것이다.

분석 단위도 문제다. 분산된 계에서는 일부 부위와 생물 전체, 생물이 바꾼 환경과의 상호작용을 각각 측정할 수 있다. 과거 활동의 흔적이 외부에 남아 이후 움직임을 바꾼다면 기능적 기억 일부는 환경과의 상호작용에 있을 수 있다. 계의 경계를 밝히지 않으면 외부 흔적을 내부 회상으로 잘못 제시하기 쉽다.

## 학습 논쟁에 재현이 필요한 이유

Markel은 2020년에 앞서 보고된 완두의 연합학습 효과를 재현하지 못했다. 이 연구는 특정 결과의 견고성에 의문을 제기했다. 재현에 실패했다고 해서 식물의 모든 학습 형태가 불가능함을 입증한 것은 아니다. 처음의 긍정적 보고 역시 퀄리아 문제를 해결하지 않는다. 평가에는 조건의 비교 가능성, 대안 통제, 새로운 독립 검증이 필요하다. [41]

연합 과제에서는 결과가 두 사건의 관계 자체에 의존함을 보여야 한다. 한 집단에 더 많은 빛이나 열, 성장 시간을 주었다면 다른 설명이 가능하다. 대조 조건에서는 가능하면 자극의 양을 유지하면서 사건들의 관계를 바꾸어야 한다. 그렇지 않으면 물리 조건이 다른데 그 효과 전체를 연합학습에 돌리게 된다.

연합을 성공적으로 검증해도 답한 질문은 제한적이다. 직접 만든 프로그램에 두 신호가 함께 나타날 때 커지는 연결 가중치를 둘 수 있다. 프로그램은 첫 신호로 둘째 신호를 예측하기 시작한다. 이는 분명한 기능적 결과다. 예측에서 체험되는 예상으로 나아가려면 좋아하는 사례 이외에도 적용할 수 있는 추가 근거가 필요하다.

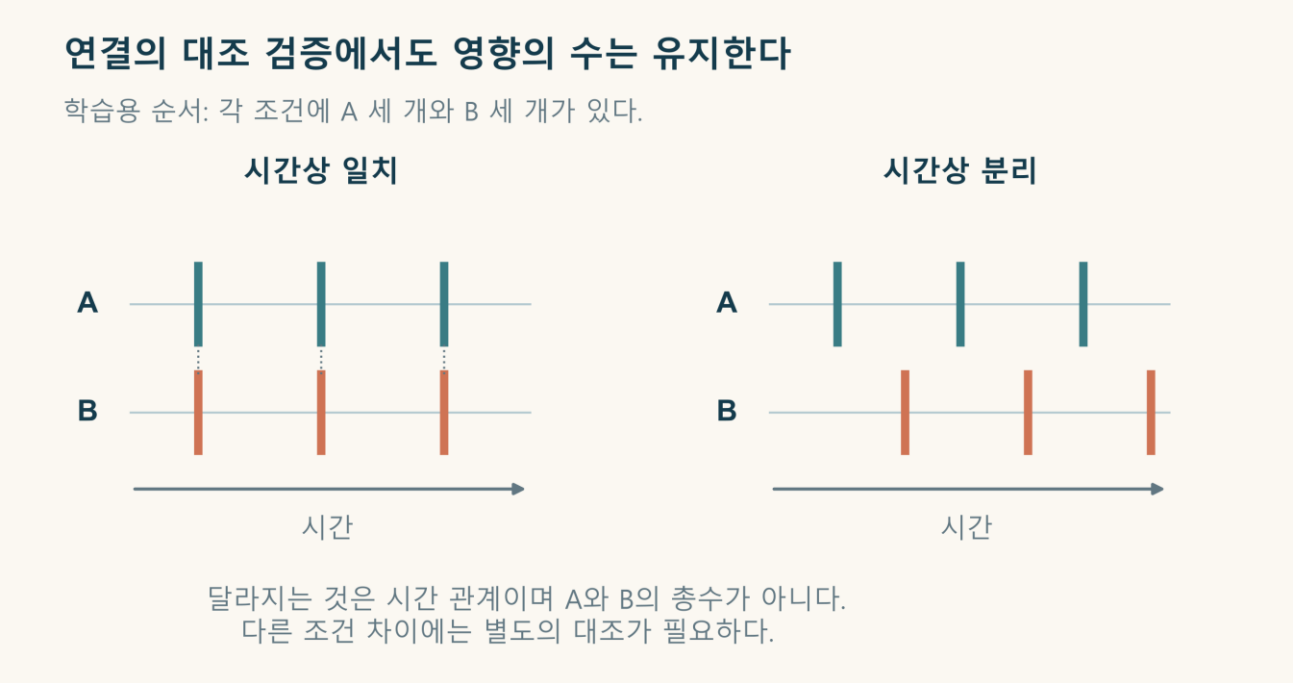


그림 49. 사건 수와 사건들 사이 관계의 구분. 두 가상 순서는 신호의 수가 같지만 시간 관계는 다르다. 이 대조는 영향의 수를 유지하면서 시간상 일치의 역할을 검증한다. 시점이 달라진 사건들 사이에도 다른 예측 가능한 연결이 생길 수 있다. 학습에 관한 결론을 내리려면 주장한 규칙 자체를 검증해야 한다. 조건 차이의 효과 전체를 연합에 돌려서는 안 된다.

## 체험 가설이 더해야 할 것

어떤 기억과 되먹임만 있으면 충분하다는 주장을 생각해 보자. 두 성질을 단순하게 구현한 대조 장치에도 규칙을 적용해야 한다. 그 장치에도 체험을 귀속한다면 저자는 그 귀결을 인정하거나 처음 조건을 수정해야 한다. 마음에 드는 생물 사례에는 넓은 주장을 유지하면서 불편한 공학 사례에서는 몰래 범위를 좁혀서는 안 된다.

다른 가설은 특정한 신호 통합과 내부 관점을 요구할 수 있다. 그러면 그 성질이 뉴런 없이 어떻게 구현되며 분산 조절과 어떻게 다른지 보여야 한다. «생물이 자기 이익을 위해 행동한다»는 말만으로 메커니즘이 정해지지는 않는다. 변수와 의존 관계, 기능이 손상되는 조건을 밝혀야 한다.

체험을 예측하는 가설과 체험을 배제하지 않는다는 것도 구별해야 한다. 식물의 체험을 배제하지 않는 이론이 체험을 예측하는 것은 아니다. 긍정적으로 결론 내리려면 해당 조건이

확인되어야 한다. 이 구별은 체험이 확인된 담지체의 범위를 근거 없이 늘리지 않으면서 연구의 개방성을 유지하게 한다.

## 출발점이 같은 두 모형

습관화 모형에는 신호마다 별도의 h가 있고 처음 값은 0이다. 반응은 1에서 h를 뺀 값이며, 그다음 h에 1까지 남은 거리의 0,2만큼을 더한다. 소진 모형에는 공통 자원 r이 있고 처음 값은 1이다. 반응은 r이며 그다음 자원에 0,8을 곱한다. A를 처음 다섯 번 제시했을 때 두 모형의 반응은 모두 1; 0,8; 0,64; 0,512; 0,4096이다. 새로운 B에 대해서는 특정 신호별 모형의 반응은 1이지만 공통 자원은 이미 0,32768이다.

휴지기의 단계마다 모든 h에 0,8을 곱하고 r에는 1까지 부족한 양의 0,2만큼을 더한다. 먼저 반응을 기록한 뒤 상태를 갱신한다. 식물이나 점균의 매개변수를 측정한 것이 아니라 전부 명시한 실습 규칙이다.

### 여기까지 확인한 것

하강 곡선 하나로는 설명을 고를 수 없다. 새 신호는 특이성을, 휴지기는 회복을 검증한다. 두 결과 모두 미리 정한 같은 순서로 계산해야 한다.

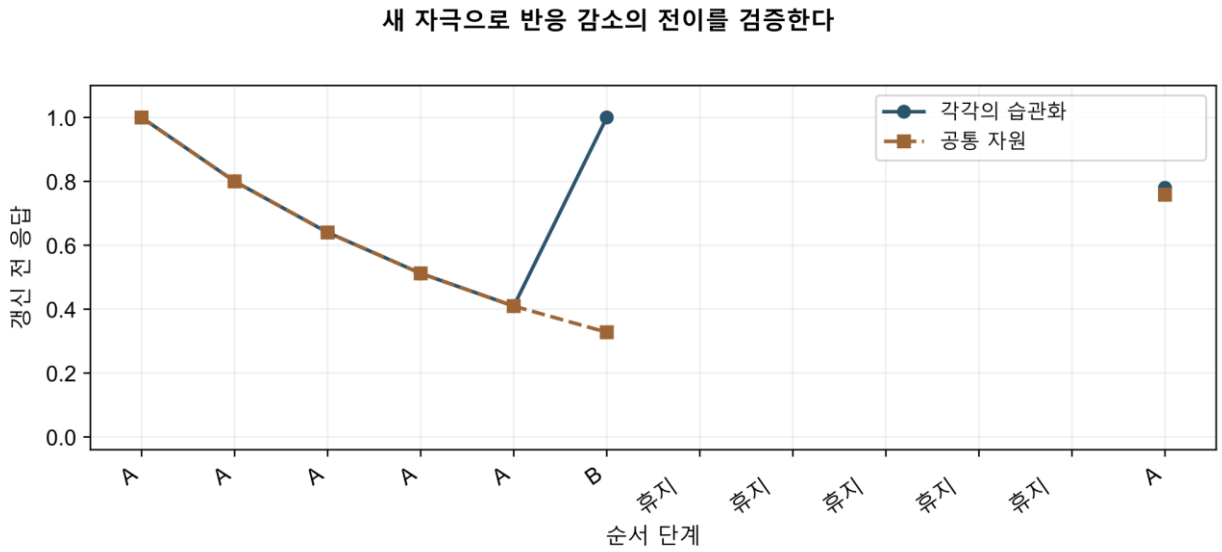


그림 50. 반응의 전이로 구별하는 습관화와 고갈. 이 장의 명시적 규칙을 계산했다. 제시한 자극에 대해 h ← h + 0,2(1 − h), 사건마다 공통 자원은 R ← 0,8R, 휴지기에는 h ← 0,8h 및 R ← R + 0,2(1 − R)다. 응답은 상태를 바꾸기 전에 기록한다. A를 다섯 번 제시한 뒤 첫 B의 응답은 분리된 메커니즘에서 1, 자원 메커니즘에서 0,32768이다. 이는 가정한 메커니즘이며 생물을 대상으로 한 실험이 아니다.

## 실습 22 같은 반응 감소에 대한 두 모형

실습 목표는 같은 자극열에서 신호별 습관화와 공통 자원 모형을 구별하는 것이다.

제9장의 상태 변화 규칙과 제7장의 대조 논리를 활용한다. 실습에 앞서 두 알고리즘을 같은 동작 순서로 제시했다.

예상 소요 시간은 25–35분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. A와 B의 h는 각각 영에서, 공통 자원은 1에서 시작한다. A를 다섯 번 제시하고 반응을 기록한 뒤 매번 상태를 갱신한다.

2. B를 계산하기 전에 반응을 예측한다. 이어 B를 한 번 제시한다.

3. 앞 규칙으로 휴지기 다섯 단계를 계산한 뒤 A를 다시 제시한다. 두 모형을 비교한다.

4. python teaching_lab.py --chapter 22로 계산을 대조한다. 결과로 메커니즘에 관해 무엇을 알며 체험에 관해서는 무엇을 확립하지 못하는지 따로 적는다.

두 모형의 단계별 표와 새 신호의 예측, 특이성에 관한 결론을 남겨 둔다.

새 B를 제시해도 A의 h가 B의 상태가 되지는 않는다. 휴지기에는 신호를 제시하지 않으므로 그 단계의 반응을 만들어 적지 않는다.

탐구를 이어 가려면 대조 절차를 이 장의 원저 하나와 비교한다. 실습 모형이 어떤 관찰을 반영하고 어떤 관찰은 아예 표상하지 않는지 밝힌다.

해설과 답은 부록 F 제22장 참조, 308쪽

## 자기 점검 문항

1. 기능적 기억은 신경계 없이도 존재할 수 있을까?
2. 반응의 감소만으로 아직 학습을 입증할 수 없는 까닭은 무엇인가?
3. 뉴런의 부재가 퀄리아의 문제를 무의미하게 만드는가?

기능적 기억은 신경 조직보다 넓은 개념이었다. 다음 장에서는 구조와 경계, 자기 기술의 출처를 밝히는 일이 중요한 인공적 구현과 생물학적 구현을 살핀다.

제23장

# 인공적 후보와 생물학적 후보

여러 생물을 살피면서 재료만으로 메커니즘이나 체험이 정해지지 않음을 보았다. 이제 인공 후보와 혼성 후보로 넘어간다. 자기 보고를 평가하기 전에 이용 가능한 입력과 보존 상태, 정보의 사용을 기술해야 한다. 이 장의 기본 실습에는 미리 마련한 최소 구조를 사용한다. 전체 에이전트는 제28장에서 도입하므로 미리 다룰 필요는 없다.

## 기반에 관한 질문

프로그램, 로봇, 배양한 신경계에서 체험이 생길 수 있을까? 재료의 이름만으로는 답할 수 없다. 프로그램은 계산의 조직 방식을, 로봇은 계산과 물리적 행동의 연결을, 신경세포 배양은 특정 시스템의 생물학적 구성을 기술한다. 어느 특성도 그 자체로 의식의 완전한 메커니즘을 정하지 않는다.

연구에서는 담지체와 조직, 상호작용 이력, 이용 가능한 지표를 구별해야 한다. 같은 세포로 된 두 계도 연결이 다를 수 있다. 한 프로그램이 다른 장비에서 실행될 수 있다. 로봇은 자체 제어기를 쓰거나 외부 명령을 받을 수 있다. 분석하는 장치의 경계를 밝히지 않으면 전체 장치의 성질을 일부에 잘못 귀속하기 쉽다.

## 이론에서 도출할 수 있는 것

Butlin과 동료들은 과학적 의식 이론에서 도출한 지표로 인공 시스템을 평가하고, 이어 그 지표의 구현을 검증하자고 제안했다. 이 논문은 2025년 십일월에 온라인으로 발표되었고 2026년 학술지 호에 실렸다. 이 방법은 완성된 보편적 검사가 아니다. 결론은 선택한 이론의 근거, 그 요구를 올바르게 옮기는 일, 시스템의 실제 구조에 달려 있다. [42]

지표를 검증하려면 세 질문에 답해야 한다. 해당 이론의 버전에서 필요조건인가, 충분조건인가, 동반 특성인가? 장치의 어느 부분에서 구현되는가? 어떤 개입으로 가정한 역할을 실제로

수행함을 보일 수 있는가? 지표가 개발자의 설명에만 있다면 구조를 살펴볼 근거는 되어도 조건을 충족했다고 셀 수는 없다.

제13–15장에서는 서로 다른 방향의 요구를 살펴보았다. 작업공간에는 선택한 내용을 여러 과정이 이용할 수 있다는 점이 중요하다. 재귀 접근법에는 특정한 되먹임 상호작용의 조직이 중요하다. 고차 이론에는 자신의 상태에 대한 표상이 중요하다. 내부의 물리적 인과성을 중시하는 접근법에서는 기능적 유사성으로 충분하지 않을 수 있다. 후보를 평가할 때 이런 차이를 보존해야 하며, 구별 없는 «지능의 징후»의 합으로 묶어서는 안 된다. [16] [17] [19] [20]

## 프로그램과 자기 기술

고장이 날 때마다 «나는 당황했다»고 출력하는 프로그램을 생각해 보자. 다른 버전은 «오류 코드 7»을 출력한다. 문구를 바꾸어도 계산이 같다면 보고 자체로 당황함의 체험을 입증할 수는 없다. 보고는 출처를 연구해야 하는 관찰 행동의 일부다. 반대로 감정 어휘가 없다는 사실로 잠재적으로 중요한 모든 메커니즘의 부재를 확립할 수도 없다.

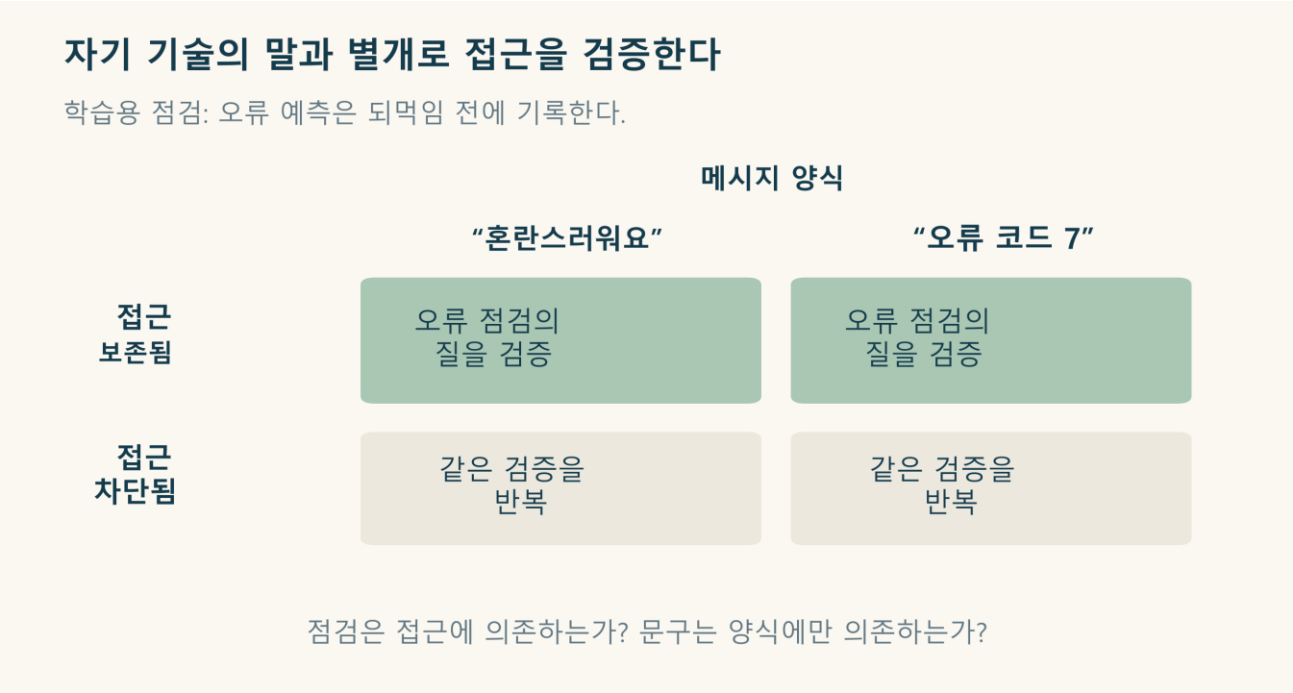


그림 51. 정보와 자기 기술의 개별 점검. 설계는 프로그램의 내부 정보 접근과 메시지 표현을 독립적으로 바꾼다. 점검이 첫 요인에, 말이 둘째 요인에 의존한다면 인과적 역할이 서로 다르다. 이는 예비적인 가정적 예측이며 이미 수행한 실험의 결과가 아니다. 환경에서 숨은 정답에 접근하는 경로를 배제해야 한다.

더 풍부한 후보는 결과를 받기 전에 자기 오류를 예측하고 기억의 한계를 고려하며 체계적 고장을 발견하면 전략을 바꿀 수 있다. 이는 기능적으로 검증할 수 있다. 그래도 결정 이유의 이야기가 결정 자체와 따로 만들어질 수 있다. 그럴듯한 설명이 실제로 작동한 근거에 접근함을 보장하지는 않는다.

직접 하는 실험에서는 내부 정보 접근과 응답 문구 틀을 따로 바꾸면 좋다. 점검의 품질은 내부 상태에, 글로 된 설명은 문구 틀에만 의존한다면 기능이 다르다. 환경의 숨은 정답을 사용한다면 정확한 확신도는 자료 누출의 결과일 수도 있다. 그러므로 출력한 문구의 분석만큼 입력 점검도 중요하다.



그림 52. 물리적 환경의 인공 시스템. 센서와 계산 장비를 갖춘 이동 장치를 그린 것이다. 이런 후보를 평가할 때에는 시스템 경계를 정해야 한다. 명령은 어디에서 만들어지고, 제어기는 어떤 정보에 접근하며, 행동의 어떤 결과를 받는지 밝혀야 한다.

## 신체가 더하는 것

로봇에서는 상호작용의 순환을 닫을 수 있다. 행동이 세계를 바꾸고 새 관찰은 행동에 의존하며 모형은 명령과 결과의 관계를 구체화한다. 이는 자기관련성과 적응을 연구하기에 편리하다.

신체가 있다는 것을 의식의 충분조건으로 만들지는 않는다. 신체의 어떤 기능을 다른 상호작용 조직으로 손실 없이 대체할 수 없는지 밝혀야 한다.

카메라와 움직이는 플랫폼을 갖춘 실습 로봇을 보자. 회전하면 영상이 달라진다. 변화를 예측하면 대상의 움직임과 자기 회전의 결과를 구별할 수 있다. 제어가 모르게 바뀌면 행동 모형을 갱신해야 한다. 이 사례에서 신체는 구체적인 인과 의존성을 정한다. «나»라는 말을 쓸 상징적 권한을 주는 것이 아니다.

특정 신체 조절이 필요하다는 이론을 검증하려면 어떤 변수를 왜 조절하는지 밝혀야 한다. 수치로 표시한 배터리 잔량이 곧 체험되는 욕구가 되지는 않는다. 선택과 학습, 기억, 체험과 연결하는 원리와 어떻게 통합되는지 연구해야 한다. 그렇지 않으면 평범한 제어 신호에 이름만 바꾸어 붙이게 된다.

## 신경세포 배양, 오가노이드와 혼성계

신경세포 배양은 인공적으로 구성한 조건의 생물학적 신경세포 시스템이다. 오가노이드는 조직 발달과 구성의 특정 측면을 모형화하도록 배양한 삼차원 시스템이며, 그 명칭이 완성된 뇌를 뜻하지는 않는다. 혼성계는 생물학적 요소를 계산 제어, 센서, 구동 장치와 연결한다. 상세한 실험 사례와 한계는 제31장에서 다룬다.

여기에도 같은 분석 원리를 적용한다. 학습과 전기 활동, 제어 성공은 가정한 체험과 따로 측정해야 한다. 생물학적 기원은 일부 메커니즘 가설을 더 구체적으로 만들지만 실제 구현을 자동으로 알려 주지는 않는다. 연결 조직과 되먹임의 경계, 외부 알고리즘의 역할, 상호작용의 시간적 안정성을 밝혀야 한다.

외부 제어기가 보상을 선택하고 조직의 응답을 명령으로 변환하는 혼성계는 특히 많은 것을 보여 준다. 장치 전체의 결과가 달라진 원인은 제어기 조정일 수 있다. 학습을 생물학적 부분에 귀속하려면 적절한 대조 비교가 필요하다. 체험을 귀속하려면 별도의 이론적 논증도 필요하다. 이는 순차적인 두 질문이며 설계 이름 하나로 합쳐서는 안 된다.

| 시스템의 분류 | 우선 기술할 것 | 결론에서 흔히 하는 오류 |
|---|---|---|
| 프로그램 | 실제 상태와 계산의 의존 관계 | 자기 기술을 체험의 직접 측정으로 여기는 것 |
| 로봇 | 지각, 행동, 신체 상태의 연결 | 몸체의 존재를 충분조건으로 여기는 것 |
| 신경세포 배양 | 세포망과 외부 제어의 조직 | 망에 장치 전체의 성질을 부여하는 것 |
| 오가노이드 | 조직의 정도와 이용 가능한 상호작용 | 조직 모형을 뇌 전체와 동일시하는 것 |
| 혼성계 | 부분들의 경계와 인과 기능의 분담 | 제어기의 학습과 생물학적 요소의 학습을 구별하지 않는 것 |

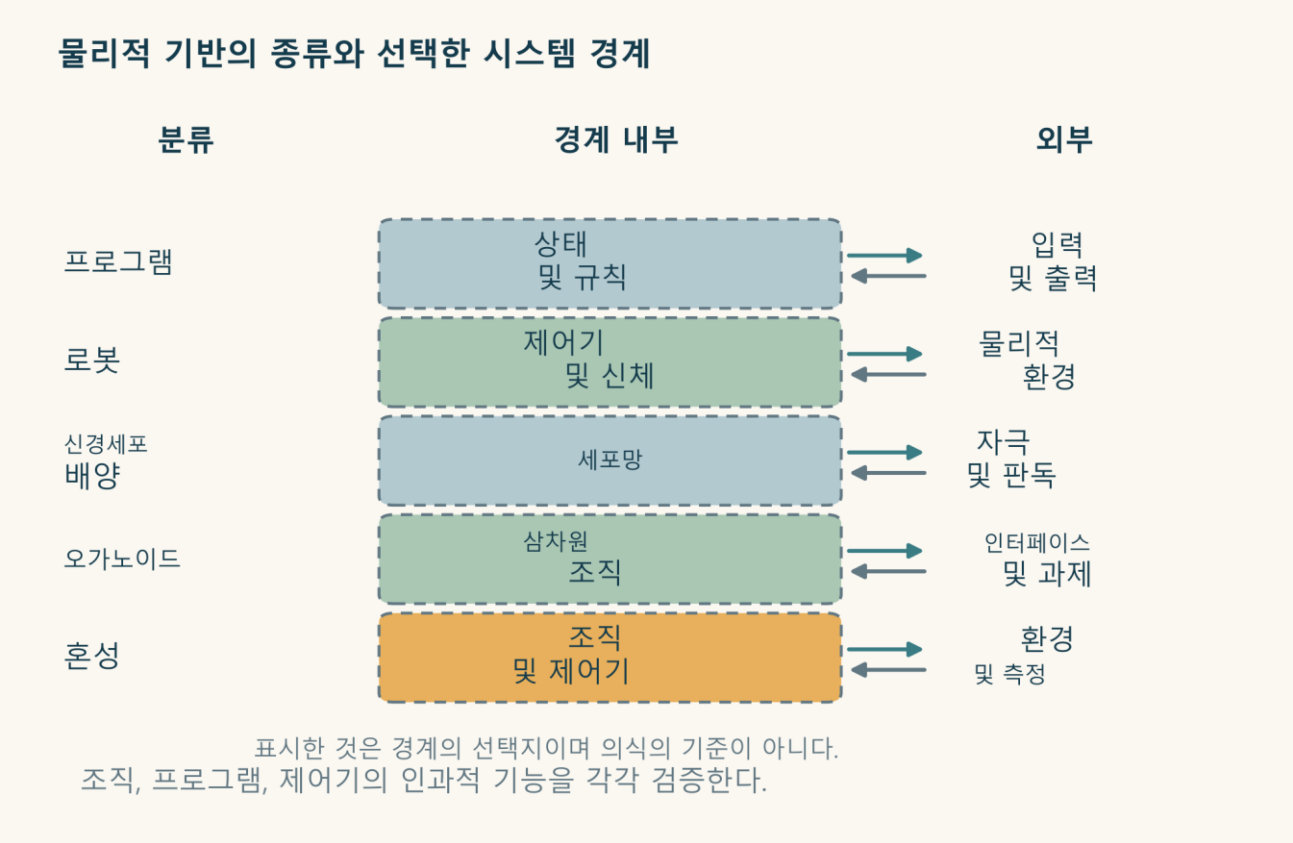


그림 53. 기반만으로 연구 경계가 정해지지는 않는다. 시스템 경계를 선택하는 다섯 예다. 점선은 고려하는 부분과 외부 상호작용을 구분한다. 프로그램, 로봇, 배양체, 오가노이드, 혼성계에서는 학습, 상태 저장, 제어가 어디에서 이루어지는지 정해야 한다. 이는 장치를 기술하는 선택지이며 해당 분류의 모든 시스템이 반드시 갖출 구조가 아니다. 의식 확률의 순위표도 아니다.

## 특징들을 순위표로 바꾸지 않는 방법

두 후보를 생각해 보자. 하나는 오류 조절이 잘 검증되었지만 내용 전파의 조직은 모른다. 다른 후보에는 공통 교환 채널이 있지만 자기 신뢰도 평가는 미리 정한 규칙으로 만들어진다. 체크 표시의 수만 더하면 차이를 감춘다. 한 이론에서 치명적인 지표의 부재가 다른 이론에서는 문제가 아닐 수 있다.

따라서 점검 결과는 성질별 양상으로 제시하는 편이 좋다. 요구마다 이론과 해석, 측정, 결과, 남은 대안을 적는다. 모르는 성질도 표시한다. 검증할 접근에서 물리적 인과성이 중요하다면 장비에 접근할 수 없을 때 프로그램 설명으로 물리적 인과관계의 정보를 대신할 수는 없다.

가정한 지표가 실제로 쓰이는지도 검증한다. 모듈이 공통 기억에 연결되어 있어도 관심 내용은 전혀 읽지 않을 수 있다. 다른 모듈은 어떤 선택에도 영향을 주지 않는 확신도 평가를 저장할 수 있다. 그러면 코드의 필드만 보고 기능적 조직을 부풀리게 된다. 상태의 선택적 제거와 왜곡, 복원으로 인과적 역할을 알아볼 수 있다.

제거 뒤 변화가 없다고 언제나 최종적인 부정 결론이 나오지는 않는다. 과제가 그 기능을 요구하지 않았거나 중복 메커니즘이 대신했을 수 있다. 기능이 필요한 조건을 미리 정하고 이런 보상 설명을 검증해야 한다. 그래야 구조 점검이 작동 장치의 연구가 된다.

## 전체 실습 전에 사용할 수 있는 최소 구조

장치가 목표가 오른쪽일 확률 p와 정방향 제어일 확률 q를 보존한다고 하자. 선택 모듈은 두 수를 읽어 명령을 고른다. 모니터도 같은 수를 읽어 예상 성공률을 계산한다. 글로 된 보고는 모니터에서 만들거나 고정 문구 틀로 대체할 수 있다. 상태를 유지하면서 문구 틀을 바꾸어도 명령 선택은 달라지지 않아야 한다. 반면 q를 왜곡하면 선택과 예측이 모두 달라질 수 있다. 이는 인과 검증의 도식이다. 전체 갱신 규칙은 제28장에서 살핀다. 지금도 구조상의 요구와 글이 주는 인상을 구별할 수 있다.

**여기까지 확인한 것**

모르는 성질을 명시하면 구현에 대한 지식이 불완전해도 양상을 정리할 수 있다. 기억이나 신체 몸체, 신경세포가 있다는 것만으로 구체적인 연결의 검증을 대신할 수는 없다.

## 실습 23 후보의 명세서

실습 목표는 주어진 구조에서 상태 접근과 언어적 자기 기술을 따로 평가하는 것이다.

제15장의 세계 모형과 행동 모형, 제14장의 내부 접근과 외부 로그를 활용한다. 구성요소의 이름만으로 구현을 검증한 것은 아니다.

예상 소요 시간은 25–35분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. p와 q의 평가, 선택, 모니터, 보고라는 네 구성요소를 그린다. 앞에서 명시한 연결만 그린다.
2. 각 연결의 송신자, 수신자, 차단했을 때의 결과를 명시하라. 숨은 실제 표적을 모니터에 넣지 마라.
3. 같은 외부 입력에서 문구 틀의 변경과 q의 왜곡이라는 두 개입을 비교한다. 선택과 예측, 글에 예상되는 변화를 따로 적는다.
4. 두 이론에 따라 구조가 무엇을 모형화하며 무엇이 미검증이고 체험을 결론 내리려면 어떤 근거가 더 필요한지 정리한다.

연결 관계도와 검증한 성질, 미검증 성질, 없는 성질의 양상을 남겨 둔다.

그림이 있다고 모든 구성요소가 실제로 구현되었다는 뜻은 아니다. 코드로 넘어가면 연결마다 상태를 확인해야 한다.

연구 확장: 제28장 뒤에 작동하는 행위자와 로그를 사용해 프로필을 다시 작성하라. 구현의 검증을 블록 이름의 일치로 대신하지 마라.

해설과 답은 부록 F 제23장 참조, 309쪽

## 자기 점검 문항

1. 체험을 능숙하게 이야기한다는 것만으로 체험의 존재를 결론 내릴 수 없는 까닭은 무엇인가?

2. 생물학적 담지체는 의식을 보장하는가?

3. 후보의 성질별 양상을 정리하는 것이 충족한 지표의 개수를 세는 것보다 나은 까닭은 무엇인가?

다른 후보들의 평가로 요구를 기술할 언어를 준비했다. 이제 설명할 차이를 좁게 고르고 유력한 경쟁 설명과 비교할 수 있는 모형을 직접 만든다.

제7부

# 자신의 설명 설계

제24장

# 향후 모형의 요구사항

지금까지 기존 설명과 구현을 살폈다. 이제 직접 연구를 시작한다. 출처의 이력이 신체 소유감 평가에 미치는 영향이라는 구체적인 질문을 고른다. 메커니즘을 만들 만큼 좁으면서 더 넓은 주체의 문제와도 연결된다. 먼저 모형이 설명해야 할 것과 이미 발표된 사실, 아직 우리의 예상에만 속하는 결과를 정한다.

## 흥미로운 질문에서 설명이 충족할 요구로

화면 속 손을 제어한다고 상상해 보자. 움직임은 보통 명령을 따르지만 때로 늦는다. 몇 분 동안 안정적인 연결을 경험한 뒤 짧은 지연은 우연한 잡음처럼 느껴진다. 연결이 계속 바뀌었다면 같은 지연도 다르게 느껴질 수 있다. 체험되는 신체 소유감이 현재 신호의 일치에만 달려 있는지, 가정한 출처의 이력에도 달려 있는지 물을 수 있다. 이는 실험 결과의 보고가 아닌 사고실험이다. 자기 연속성이라는 일반 생각을 모형의 구체적인 요구로 바꾸기 위한 사례다.

모형은 차이를 설명해야 한다는 요구에서 출발한다. «뇌는 자기를 구성한다»는 말은 거의 어떤 결과도 허용한다. 반면 «현재의 일치 정도가 같을 때에도 안정된 출처의 이력이 이후 신체 소유감의 역동을 바꾼다»는 주장은 예상 범위를 제한한다. 이제 이력의 영향이 얼마나 지속되는지, 출처 사이에 옮겨 가는지, 응답 방식을 바꾸어도 유지되는지 물을 수 있다. 질문을 정밀하게 할수록 어떤 결과가 설명의 성공인지 분명해진다.

## 필요한 정밀도로 대상을 정하라

이 기획의 첫 연구 대상은 체험되는 신체 소유감이다. Kalckert와 Ehrsson의 실험에서는 인공 손의 소유감과 움직임의 행위 주체감이 갈릴 수 있었다. 따라서 신체 소유감을 성공적인 자발적

제어만으로 정의할 수는 없다. 모형은 소유감, 제어, 행위의 정확성 평가를 별도의 지표로 유지하며 이 차이를 설명해야 한다. [4]

연구 대상을 한 문장으로 적을 수 있다. «일치하는 신체 신호의 이력이 바뀔 때 체험되는 신체 소유감의 변화를 설명한다.» 이어 성인 참여자와 짧은 사례, 특정 제시 및 보고 방식이라는 적용 범위를 적는다. 이렇게 하면 확장 가능성은 유지된다. 지금 변수 하나로 색과 꿈, 자전적 자기감, 의식의 최초 발생까지 설명할 필요는 없다.

모형을 확장하면 설명해야 할 내용 자체가 달라진다. 처음에는 모형이 소유감 평가를 예측한다. 이어서 그 평가와 체험되는 조직의 관계를 뒷받침해야 한다. 마지막으로 더 강한 가설은 제안한 조직이 왜 경험의 존재를 구성하거나 가능하게 하는지 말해야 한다. 이 전환들은 하나의 결론을 슬쩍 더 크게 외치는 일이 아니라 별개의 과제가 된다.

## 자료와 바라는 결과를 구별하라

두 목록을 만들면 좋다. 첫 목록에는 모형이 반영해야 할 발표 결과를 넣는다. 둘째 목록에는 아직 검증하지 않은 연구 예상을 넣는다. 섞으면 자기 가정이 확립된 사실처럼 보인다. 예를 들어 비슷한 두 영상 사이의 이력 전이를 문헌에서 직접 비교한 근거를 찾지 못했다면 그 주장은 둘째 목록에 속한다.

Samad와 동료들의 연구는 시각과 고유수용감각, 촉각 신호의 공통 원인을 추론하는 모형으로 신체 소유감을 연구하는 방법을 보여 준다. 이미 발표된 가까운 연구이며 우리의 새 생각이 아니다. 자신의 기여는 구체적인 동역학이나 전이 조건, 대안 메커니즘의 구별에서 찾아야 한다. [15]

첫 목록의 항목마다 과제와 지표, 결론의 범위를 적는다. «신호들이 통합된다»는 너무 넓다. «이 조건에서 신체 소유감 평가가 달라졌고 손 위치도 함께 측정했다»면 어떤 관찰을 재현하는지 알아볼 수 있다. 여러 지표의 결과가 일치하는지도 적는다. 어긋남이 미래 모형의 핵심 요구가 될 수 있다.

## 우리 설계의 네 요구사항

첫 요구는 현재에 관한 것이다. 모형은 현재 신호들의 양립 가능성에 반응해야 한다. 일치 정도가 전혀 무관할 수는 없다. 계가 어떤 양립 가능성의 특징을 이용할 수 있는지 정해야 한다. 기하학적 일치와 시간적 근접성, 행동에 대한 의존성은 서로 다른 정보다. 풀이하지 않은 «일관성» 변수로 대신할 수 없다.

둘째 요구는 과거에 관한 것이다. 이력의 영향에는 계산 가능한 동역학이 있어야 한다. 기억을 더한 뒤 모든 지연을 기억의 증거라고 할 수는 없다. 개별 사건과 환경 안정성의 평가, 공통 원인의 예상, 특정 출처의 이력 가운데 무엇을 보존하는지 미리 밝혀야 한다. 선택마다 망각과 전이의 메커니즘이 다르다.

셋째 요구는 동일성에 관한 것이다. 특정 출처의 이력이라면 그 출처가 다시 나타났을 때 왜 되살아나는지 설명해야 한다. 같은 일반 방식으로 외부 대상을 추적하는 경쟁 모형도 필요하다. 그렇지 않으면 출처 기억의 이점을 특별한 자기의식 메커니즘의 증거로 오해하게 된다.

넷째 요구는 측정에 관한 것이다. 내부 상태가 같아도 질문 표현이나 동의 판단 기준이 달라지면 응답이 다를 수 있다. 그러므로 평가 형성 모형과 출처 모형을 나누어야 한다. 그래야 신호 처리와 귀속, 결과 보고 가운데 어디에서 효과가 생겼는지 검증할 수 있다.




그림 54. 모형이 충족할 요구의 지도. 방법 도식이다. 네 요구는 모형의 구조와 검증에 해당한다. 점선은 메커니즘과 체험의 연결을 설명해야 한다는 별도의 논증 과제를 나타낸다. 정확한 예측에서 체험의 존재로의 이행이 확립되었다는 뜻이 아니다.

## 여러 경쟁 모형을 세워라

현재 일치 모형을 M0이라고 부르자. 선택한 현재의 시간 구간만 사용한다. M1은 출처별 기록 없이 확인과 반증의 전체 이력을 보존한다. M2는 가정한 출처들의 구별된 이력을 유지한다. 이는 앞으로 만들 계산 모형들의 유형 이름이며 유형마다 구체적인 규칙을 골라야 한다.

일반적인 대상 동일성 추적 모형 M1I를 더한다. M2의 이점이 평범한 정보 흐름의 분리에서 나올 수 있기 때문이다. M2와 M1I의 비교로 추가적인 신체 조직이 필요한지, 안정된 모든 대상에 적용되는 메커니즘만으로 충분한지 알아보아야 한다. «일반 메커니즘으로 충분하다»는 답도 자기관련성의 성격에 관한 연구 결과다.

경쟁 모형을 일부러 약하게 만들지 마라. M0가 현재 정보를 덜 받고 M2가 숨은 표지까지 본다면 그 비교는 입력 정보의 이점을 검증한다. M2에 조정할 매개변수가 훨씬 많다면 새 자료 검증이 필요하다. 공정한 비교라는 요구는 처음부터 모형의 명세에 포함되어야 한다.

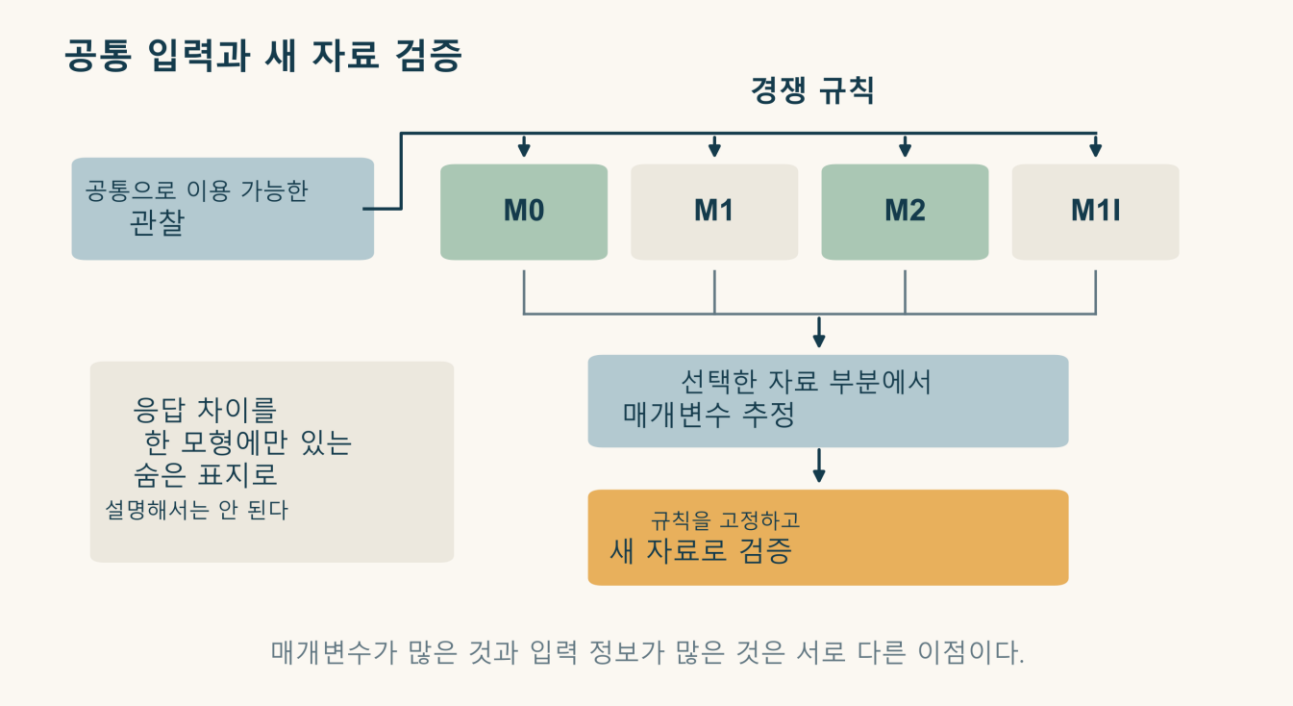


그림 55. 계산하는 경쟁 모형들의 공정한 비교. M0, M1, M2, M1I를 같은 이용 가능한 관찰 집합으로 비교하면서 각 기억 규칙의 차이는 유지한다. 새 자료 검증과 매개변수 추정을 구분하며 선호하는 모형에만 숨은 출처 표지를 주어서는 안 된다. 이 비교의 이점은 정해진 과제에 해당한다. 자기관련성의 특별한 본성을 아직 확인하지는 않는다.

## 성공의 적극적 기준을 정하라

익숙한 출처가 돌아왔을 때 신체 소유감의 회복을 예측하는 모형을 생각해 보자. 평균 평가가 같다는 것만으로 성공은 아니다. 변화의 방향과 속도, 효과가 사라지는 조건을 예측해야 한다. 한 과제에서 추정한 매개변수로 임의의 재조정 없이 다른 과제를 예측할 수 있다면 특히 가치 있다.

실습 연구에서는 일치 사건의 수는 같고 출처의 조직은 다른 두 열을 구별하라는 요구를 정할 수 있다. 아직 실험으로 확인한 법칙은 아니다. 이론이 순서의 역할을 설명하게 하는 검증 과제다. 다음 장에서는 계산 모형들이 실제로 갈라지는 단순한 열을 만든다.

성공 기준에 중단 조건을 더하면 유용하다. 이용 가능한 자료로 M1과 M2를 구별하지 못한다면 이 설계는 출처의 이력에 관한 질문에 아직 답하지 못한다. 이때에는 선호하는 모형에 대한 확신을 높이는 대신 과제를 바꾸어야 한다. 좋은 명세는 다음에 필요한 연구가 무엇인지 제때 알아보도록 돕는다.

## 전체 구상이 설명의 요구로 구체화되는 방식

«이력은 자신을 느끼도록 돕는다»는 구상은 너무 많은 결과를 허용한다. 이를 구체화하면 처음의 구별이 절차와 연결된다. 동일한 검사 순서에서 익숙한 출처 또는 새로운 출처가 돌아온 뒤 신체 소유감의 평가가 회복되는 양상이 달라야 한다는 것이다. 다음으로 무엇을 익숙함이라고 할지, 어떤 특징으로 출처를 인식할지, 응답의 역동을 어떻게 측정할지 명시해야 한다. 모든 외부 대상의 이력을 저장하는 경쟁 모형도 따로 적는다. 그것이 같은 예측을 내놓는다면 아직 특별한 신체 기능을 분리해 낸 것이 아니다. 이런 한계는 기획을 폄하하는 것이 아니라 다음 실험을 선택하도록 돕는다.

### 여기까지 확인한 것

좋은 요구는 기대하는 성공과 수정 조건을 모두 제한한다. 이제 과거와 현재, 동일성, 측정에 별도 역할을 부여하는 명세를 작성할 수 있다.

## 실습 24 한 모형의 요구 정하기

실습 목표는 계산하기 편한 규칙을 고르기 전에 한 모형이 충족할 요구를 마련하는 것이다.

신체 소유감과 행위주체감의 구별, 공통 원인 모형, 경쟁 모형의 공정한 비교를 활용한다. 응답 예측의 성공과 체험의 존재에는 서로 다른 근거가 필요하다.

예상 소요 시간은 35–45분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 설명할 차이와 조건의 범위를 한 문장으로 정한다. 기본 과제에는 다른 출처가 끼어든 뒤 기존 출처가 돌아오는 상황을 사용한다.
2. 발표된 결과가 주는 제약과 자신의 예상을 나눈다. 발표 결과에는 정확히 무엇을 측정했는지 적는다.
3. 현재 특징과 이력 보존, 동일성 인식, 응답 형성의 네 요구를 적는다. M0과 M1, M2, 일반 추적기를 향후 경쟁 모형으로 명시한다.
4. 선택한 차이를 뒷받침할 결과와 특별한 버전이 불필요해질 결과를 정하라. 아직 임의의 인간 평균값을 배정하지 마라.

연구 대상과 요구, 경쟁 모형, 다음 판단의 기준을 담은 명세 하나를 남겨 둔다.

매개변수 수를 늘리는 것과 이용 가능한 정보를 더하는 것은 다른 이점이다. 둘 다 검증해야 한다.

탐구를 이어 가려면 자유로운 재조정 없이 매개변수를 적용할 둘째 과제를 제안한다. 같은 관계를 검증하는 이유를 설명한다.

해설과 답은 부록 F 제24장 참조, 310쪽

**자기 점검 문항**

1. 과거 정보의 보존 자체로 특별한 자기관련성 메커니즘을 확인할 수 없는 까닭은 무엇인가?
2. M2만 숨은 출처 표지를 받고 M0는 받지 않는다면 앞 모형의 이점은 무엇을 보여 주는가?
3. 확립된 제약과 자신의 예상을 나누어야 하는 까닭은 무엇인가?

요구만으로 모형의 다음 단계가 정해지지는 않는다. 제25장에서는 입력과 상태, 갱신을 정하고 한 계산을 끝까지 추적한다.

제25장

# 아이디어에서 메커니즘으로

제24장에서는 특정 출처의 이력 효과를 설명하도록 요구했다. 이제 이전 상태와 새 관찰로 다음 상태를 하나로 계산하는 절차를 정해야 한다. 직접 하는 연구의 첫 실행 가능한 단계가 된다. 확률 하나의 갱신을 자세히 살핀 뒤 기억 조직만 바꾼다. 이력의 이월이 없는 방식과 공통 기록 하나, 출처별 기록들이다.

## 메커니즘은 다음 단계를 가능하게 해야 한다

앞 장에서는 특정 출처의 이력 효과를 설명하라고 요구했다. 이제 생각을 실행 가능하게 만들어야 한다. 현재 상태와 새 관찰을 주면 모형이 다음 상태를 정해야 한다. «계가 중요한 정보를 통합한다»는 말만으로는 부족하다. 무엇이 중요하고 어떻게 통합하며 사건 사이에 무엇을 보존하는지 알 수 없기 때문이다.

작은 구성에서 시작한다. 뇌의 모든 구조를 그리려는 것이 아니다. 이력을 다루는 서로 다른 규칙이 어떻게 다른 예측을 내는지 보여 주려는 것이다. 이 장의 모든 수치와 입력열은 실습용이다. 계산을 투명하게 보이도록 정했으며 인간의 체험을 측정한 결과가 아니다.

## 입력, 상태와 출력을 구분하라

입력 사건은 신호의 공통 원인을 뒷받침하거나 그에 반하는 두 종류다. 실제 연구에서는 시간과 위치, 행동과의 관계에서 이런 정보를 얻어야 한다. 여기서는 다차원 사건을 잠시 이진 특징으로 바꾼다. 명시적인 단순화이며 적용 범위를 검증해야 한다.

모형의 숨은 상태는 현재 신호가 공통 원인에 속할 확률이다. 이를 믿음이라고 부르자. 기술적인 용어다. 모형은 수치를 보존하고 명시한 규칙으로 갱신한다. 수치의 의미를 이해하는 내부 관찰자를 추가하지 않는다. 현재 출력은 갱신한 믿음 자체다. 인간의 신체 소유감 평가와 연결하는 일은 별도의 변환이 된다.

환경의 실제 상태와 모형의 평가도 구별한다. 실험자는 실제로 어떤 사건이 연결되는지 알아도 모형은 불완전한 특징만 받는다. 실제 인과 표지를 주면 추론 과제가 사라진다. 그러므로 입력마다 계가 어떤 관찰 가능한 방식으로 이 정보를 얻을 수 있었는지 물어야 한다.

## 갱신 규칙 하나를 만들어라

공통 원인에서는 지지 사건이 0,75의 확률로, 다른 원인들에서는 0,25의 확률로 나타난다고 하자. 직접 정한 특징 신뢰도 모형이다. 처음 믿음이 0,5이면 지지 사건 하나 뒤에 0,75가 된다. 같은 초기 상태에서 반증 사건이 나타나면 0,25가 된다.

사건 사이에서는 이전 믿음을 불확실성 쪽으로 약화시키자. 설명용 버전에서는 0,5로부터 편차의 십분의 팔을 보존한다. 따라서 다음 관찰 전에 상태 0,75는 0,70으로 바뀐다. 이 약화는 숨은 상황이 바뀔 수 있다는 가정으로 해석할 수 있다. 이는 선택한 모형의 동역학이며 확립된 기억 법칙이 아니다.

```
def update(previous, supports_common_cause, retention=0.8):
    prior = 0.5 + retention * (previous - 0.5)
    likelihood_common = 0.75 if supports_common_cause else 0.25
    likelihood_separate = 1.0 - likelihood_common
    common_weight = prior * likelihood_common
    separate_weight = (1.0 - prior) * likelihood_separate
    return common_weight / (common_weight + separate_weight)
```

이 코드는 상태 사이 전이를 정하며 퀄리아를 계산하지 않는다. 손이나 계산기, Python으로 실행할 수 있다. 특징의 신뢰도는 새 증거의 세기를, 유지 계수는 과거의 영향을 정한다. 매개변수의 역할이 나뉘었으므로 결과가 달라지면 어느 규칙 때문인지 알아낼 수 있다.

## 일반적인 형태의 한 전이

공통 원인에 대해 보존한 확률을 p, 유지 계수를 r, 양성 단서의 신뢰도를 a로 표시하자. 기본 예시에서는 r = 0,8이고 a = 0,75다. 관찰 전에 p_до = 0,5 + r × (p − 0,5)를 얻는다. r = 0이면 이전의 편차를 완전히 없애고, r = 1이면 그대로 보존한다.

양성 특징의 경우 공통 원인의 가중치는 p_до × a이고, 서로 다른 원인의 가중치는 (1 − p_до) × (1 − a)이다. 첫 가중치를 두 가중치의 합으로 나누면 새 p가 나온다. 음성 특징의 경우에는 두 가능도를 서로 바꾼다. 이런 선택은 두 결과의 대칭적 신뢰성을 가정하므로 반드시 모형의 기술에 포함해야 한다.

계산에서 세 수를 구별한다. 보존한 p는 이전 처리의 결과다. p_до는 원인이 달라졌을 가능성을 반영한다. 새 p는 이미 받은 특징을 반영한다. 특징 자체의 예측에는 P(+) = a × p_до + (1 − a) × (1 − p_до)라는 다른 값이 필요하다. 이후 예측 품질을 평가할 때 관찰 전에 기록해야 하는 것은 이 값이다.

앞으로 나올 순서의 수치가 규칙에 들어가서는 안 된다. 개발자가 환경을 알더라도 모형은 허용된 특징과 관찰한 출처의 표지만 받는다. 나중에 동일성 인식을 더 복잡하게 만들 때에는 오류 없는 표지를 별도 추정으로 대체한다. 숨은 단서로 더하는 것이 아니다.

M0과 M1, M2는 갱신할 기록의 선택만 달라진다. 그러므로 차이를 구체적인 기억 규칙까지 추적할 수 있다. 이력이 중요하다는 일반 주장보다 훨씬 강하다. 일치와 어긋남의 사례, 효과가 사라져야 하는 조건을 이제 알 수 있기 때문이다.

## 같은 갱신을 이용하는 세 방식

M0는 사건마다 0,5에서 시작한다. 따라서 지지 사건 뒤에는 언제나 0,75가 된다. M1은 전체 열에 대한 공통 믿음 하나를 유지한다. M2는 구별하는 출처마다 별도의 믿음을 보존한다. 출처가 나타나면 관련 기록을 갱신한다. 이 단순화한 버전에서는 매 시간 단계가 아니라 다음에 그 기록을 사용할 때 약화가 일어난다.

사건 세 개를 보자. 출처 A가 공통 원인을 지지하고 출처 B가 반증하며 출처 A가 다시 지지한다. M1의 상태는 차례로 0,75와 0,4375, 약 0,7105다. M2에서는 A 기록이 0,75, B 기록이 0,25가 되고 A가 돌아오면 0,875가 된다. M0는 차례로 0,75와 0,25, 0,75를 낸다.

차이는 이력을 다루는 방식에서 생겼다. A가 돌아오면 M2는 과거의 지지를 사용하고 M1은 최근 B의 반증도 가져온다. 계산으로 확인할 구체적인 예측을 얻었다. 그러나 아직은 직접 만든

구조의 내부 동역학에 관한 예측이다. 인간 자료와 연결하려면 측정 모형과 경험적 검증이 필요하다.

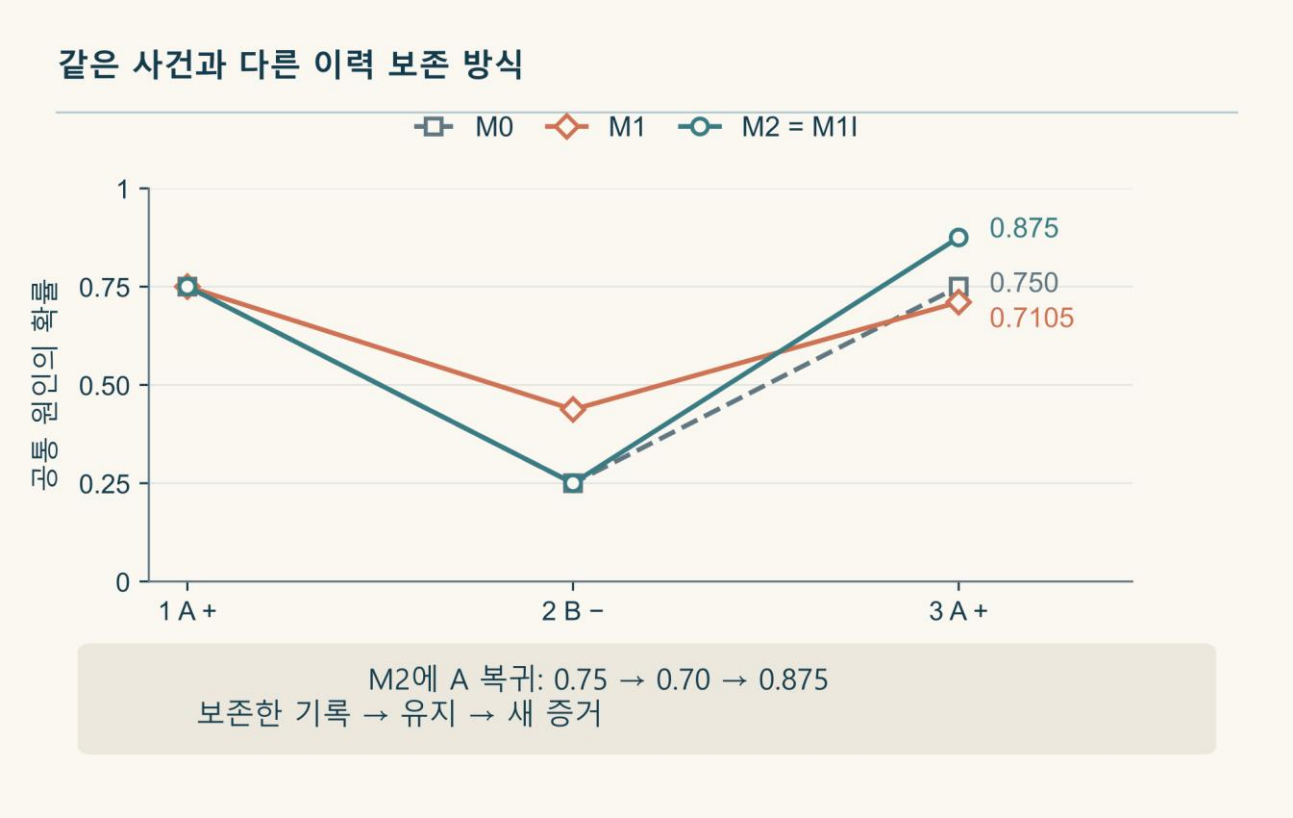


그림 57. 출처의 이력으로 돌아가기. 순서 A+, B−, A+의 정확한 학습 계산이다. 특징의 신뢰도는 0,75이며, 해당 기록을 갱신하기 전에 유지 계수 0,8을 적용한다. M0는 0,5에서 시작하고 M1은 공통 이력을 전달하며 M2와 M1I는 같은 개별 기록들을 유지한다. 이는 계산 모형의 상태이며 측정한 체험이나 참여자의 평가가 아니다.

## 출처의 동일성은 어디에서 오는가

가장 위험한 우회는 M2에 오류 없는 출처 이름을 주고 스스로 동일성을 알아냈다고 평가하는 것이다. 실습 계산에서 A와 B는 이상화한 구별을 뜻한다. 전체 연구에서는 그 출처를 설명해야 한다. 예를 들어 위치와 외형, 예상한 행동 결과의 안정된 조합으로 출처를 인식할 수 있다.

이 인식에도 불확실성이 있을 수 있다. 사건이 확률 0,7로 A에, 0,3으로 B에 속한다면 기록들에 갱신을 나누어 적용하거나 여러 가능한 이력을 보존할 수 있다. 두 해법은 동등하지 않다. 첫째는 차이를 평활화하고 둘째는 경쟁 설명들을 유지한다. 이름의 편의가 아니라 예측과 계산의 제약을 근거로 선택해야 한다.

이어 앞 장의 일반 경쟁 모형 M1I가 필요하다. 외부 대상에 같은 동일성 특징을 사용한다. M1I도 자료를 똑같이 잘 설명한다면 별도의 특별한 신체 자기관련성 메커니즘은 아직 필요하지 않다. 구조상 일반적인 메커니즘도 자기의식의 내용에 참여할 수 있다.

## 설명을 숨겨 넣지 말고 측정을 추가하라

참여자가 소유감을 영부터 십까지의 척도로 평정한다고 가정하자. 가장 단순한 학습용 대응은 모형의 믿음에 십을 곱하는 것이다. 그러면 순서의 마지막 사건에서 M0는 7,5, M1은 약 7,1, M2는 8,75를 준다. 이는 가정적 응답 예측이며 실험 결과나 체험 자체의 정량 척도가 아니다.

더 현실적인 관계에는 개인별 응답 기준과 척도의 비선형성, 우연한 변동이 들어갈 수 있다. 그러나 요소를 더할수록 설명은 유연해진다. 어떤 내부 상태든 어떤 응답으로도 바꿀 수 있다면 메커니즘을 비교할 수 없게 된다. 그러므로 응답 형성 방식을 미리 정하고 조건 사이에 옮겨 적용할 수 있는지 검증해야 한다.

신체 소유감과 독립적으로 감각 구별의 신뢰도를 평가할 추가 과제를 설계하면 좋다. 그래야 수치 하나에서 입력 품질과 이력의 세기, 응답 기준을 동시에 추측하지 않게 된다. 여러 자료 출처가 다른 매개변수에 제약을 줄 수 있다. 그래도 최종적인 식별 가능성에는 별도 검증이 필요하다.

## 경계 사례를 검증하라

유지 계수가 영이면 갱신마다 0,5에서 시작하므로 과거 믿음을 사용하는 방식의 차이가 사라진다. 유지 계수가 일이면 이전 상태를 약화 없이 옮긴다. 사건이 공통 원인과 다른 원인들을 구별하지 못하게 되면 새 관찰은 믿음을 바꾸지 않아야 한다. 이 사례들은 선택한 규칙의 귀결이며 구현 검증에 쓰인다.

이어 사건 수는 유지하면서 순서를 바꾸어라. 모든 곳에서 A와 B의 이름을 동시에 맞바꾸어라. 결과는 기록의 명칭만 달라져야 한다. 이름 변경 자체가 소유감 예측을 바꾼다면 프로그램에 불필요한 규칙이 작동할 가능성이 있다. 이 검증은 기호에 대한 불변성을 다루며 주관적 관점의 존재를 검사하지 않는다.

보기 좋은 순서 하나에 그치지 마라. 모형들이 일치하는 조건을 찾아라. 출처가 언제나 하나라면 M1과 단순화한 M2는 똑같이 작동한다. 따라서 출처가 하나뿐인 실험은 관찰을 아무리 많이 포함해도 개별 이력의 가치를 검증할 수 없다.

## 같은 증거가 같은 평가를 주지 않는 이유

A가 다시 나타나기 전에 보존된 확률은 0,75다. 유지 계수 0,8은 0,5로부터 편차의 십분의 팔을 남긴다. 사전 확률은 0,70이 된다. 양성 단서의 우도는 공통 원인에서 0,75, 서로 다른 원인에서 0,25다. 가중치 0,525와 0,075를 얻는다. 그 합 0,600으로 정규화하므로 공통 원인의 새 확률은 0,875다.

이 확률과 다음 지지 특징의 예측은 다르다. 숨은 원인이 다음 관찰 전까지 같다면 예측에는 관찰의 불완전성도 반영해야 한다. 0,875 곱하기 0,75에 0,125 곱하기 0,25를 더한 0,6875다. 원인이 바뀔 수 있다면 먼저 해당 전이를 적용한다. 두 양을 같은 결과로 슬쩍 평가해서는 안 된다.

### 여기까지 확인한 것

새 수치는 명시적으로 선택한 규칙에서 나왔다. 이제 어느 보존 기록이 갱신에 참여하고 바꾸면 어떻게 되는지 검증할 수 있다.

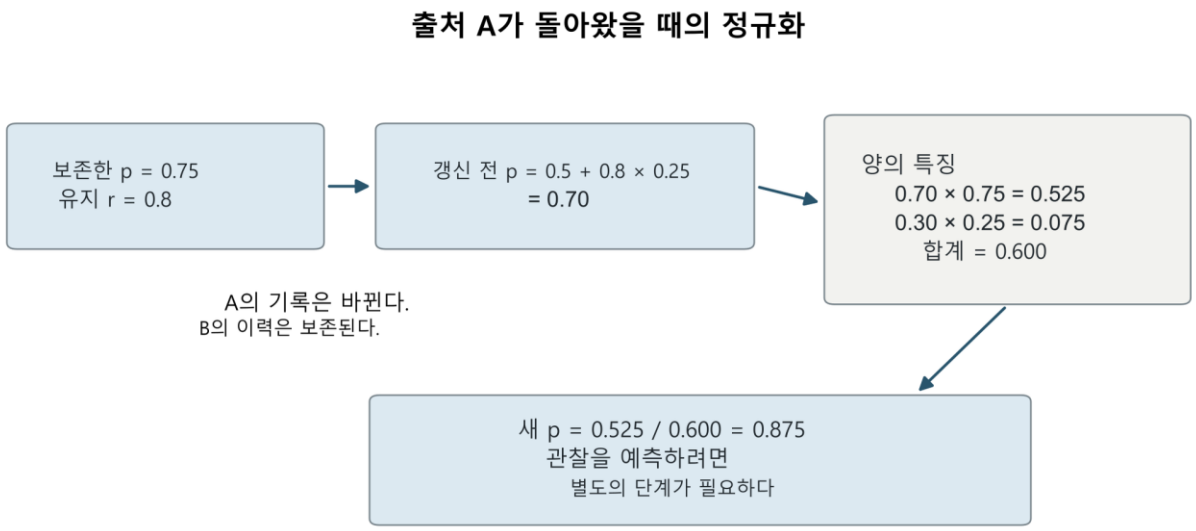


그림 56. 한 갱신의 정규화. 신뢰도 0,75일 때의 제25장 전체 단계다. 유지 뒤의 사전 상태, 정규화 전 가중치, 최종 확률을 따로 표시했다. 숨은 원인의 확률이 관찰되는 특징의 예측을 대신하지는 않는다.

## 실습 25 인과 연쇄 추적하기

실습 목표는 세 기억 방식과 일반 추적기에 대해 사건 네 개를 계산해 복원하는 것이다.

제11장의 확률 정규화와 관찰 및 숨은 원인의 구별을 활용한다. 여기서는 기록을 사용할 때 이력의 유지 계수를 적용한다. 실제 경과 시간은 제33장에서 따로 다룬다.

예상 소요 시간은 35–45분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. A 지지, B 반증, A 지지, B 지지의 순서를 사용한다. 처음 믿음은 모두 0,5이고 신뢰도는 0,75, 유지 계수는 0,8이다.
2. M0는 매번 0,5에서 시작한다. M1은 기록 하나를 이월한다. M2는 인식한 출처의 기록만 갱신한다. 일반 추적기도 같은 출처별 기록들을 사용한다.
3. 단계마다 고른 기록과 유지 계수 적용 뒤의 상태, 두 가중치, 정규화 결과를 적는다. 직접 계산한 뒤 해설을 본다.
4. 유지 계수를 영으로 바꾸어 반복하고 전체 열에서 A와 B의 이름을 서로 바꾸는 검증을 한다.

네 단계의 계산과 결과를 미리 아는 대조 검증 두 개를 남겨 둔다.

M2에서는 넷째 단계에 B의 반증 이력이 돌아온다. 출처별 기억이 언제나 가장 높은 확신을 주어야 하는 것은 아니다.

탐구를 이어 가려면 오류 없는 출처 표지를 관찰 가능한 특징으로 바꾼다. 인식의 불확실성이 기록 선택에 미치는 영향을 기술하고 같은 입력을 받는 강한 일반 추적기와 비교한다.

해설과 답은 부록 F 제25장 참조, 310쪽

**자기 점검 문항**

1. 실습열의 셋째 사건에서 M2가 공통 원인을 더 강하게 지지하는 까닭은 무엇인가?

2. 연구 전체에서 출처 하나만 사용하면 무엇을 잃게 되는가?

3. 믿음을 신체 소유감의 평가로 바꾸는 변환이 모형의 별도 부분인 까닭은 무엇인가?

기억 메커니즘이 실행 가능해졌다. 이제 저자의 가설에 따르면 어떤 조직이 체험의 질과 존재에 연결될 수 있는지 더 강한 주장을 명시해야 한다.

제26장

# 메커니즘에서 체험으로

우리는 원인의 이력을 계산하고 모형의 특정 응답을 예측할 수 있다. 이 장의 중심 가설은 여기에 별도의 단계를 더한다. 물리적으로 실현된 어떤 상호 조직과 최소한의 체험 사이에 연관이 있다고 가정하는 것이다. 이 단계를 차례로 살펴보자. 먼저 체험을 가리키는 말을 쓰지 않고 연구할 메커니즘을 정의한다. 이어 충분성에 관한 강한 주장, 구조적 귀결, 반론의 방법을 정한다. 이런 순서는 연구의 대담함을 유지하면서도 그 주장이 요구하는 바를 논의할 수 있게 한다.

## 아직 답하지 못한 질문

이전 장의 모형은 이력을 보존하고 출처를 구분하며 공통 원인의 확률을 바꿀 수 있다. 사람의 신체 소유감 평가도 잘 예측한다고 하자. 그러면 그 평가의 메커니즘에 관해 중요한 것을 알게 된다. 하지만 왜 해당 과정에 체험이 수반되는가라는 질문은 남는다. Chalmers의 철학적 문제 설정에서 구별하고 보고하는 능력의 설명은 주관적 체험이 존재하는 이유의 설명이 자동으로 되지 않는다. 여기서는 이 구분을 설명의 불가능성에 대한 선행 증명이 아니라 논거가 충족해야 할 요구로 사용한다. [2]

나아가려면 계의 조직이 체험의 질과 존재에 어떻게 연결되는지 밝히는 대응 원리가 필요하다. 틀린 것으로 드러날 수 있다. 그러나 이런 주장이 없다면 기술적 연구는 직접 프로그램한 기능을 넘어설 수 없다. 강한 잠정 가설을 세우고 그에 따르는 요구를 살펴보자.

## 체험이라는 말 없이 조직을 정의하라

여러 구별이 서로를 규정하는 과정을 연구하자고 제안한다. 한 구성요소는 관찰 대상의 성질을 구별한다. 둘째는 변화의 출처를 평가한다. 셋째는 지금 가능한 추가 관찰이나 내부 변환을 정한다. 상태들은 함께 갱신된다. 가정한 출처가 특징의 해석을 바꾸고 특징이 출처의 예상을 바꾸며 선택한 검증이 두 표상을 재구성한다.

여기서 상호성은 인과적인 뜻이다. 가정한 연결마다 방향을 정한 개입과 양방향의 예상 변화를 미리 명시한다. 예를 들어 출처 평가의 변화가 특징 해석을 어떻게 바꾸며, 표상된 특징의 변화가 출처 평가를 어떻게 바꾸는지 정한다. 공통 활동이나 상관만으로는 부족하다. 계는 지금 일어나지 않는 대안도 구별해야 한다. 다른 관찰이나 명령, 이전 해석의 포기로 무엇이 달라질지다. 그 대안은 개발자의 설명에만 있지 않고 처리에 영향을 주어야 한다.

개별 구별들은 하나의 동역학을 이룬다. 각 구성요소를 평소 출력의 독립적인 재생으로 대체하면 미리 정한 모든 개입의 응답을 보존할 수 없다. 이 검증은 서로의 규정과 단지 기록된 여러 흐름의 일치를 구별한다. 이름으로 의식을 정의하는 것이 아니라 연구할 계의 구조를 정한다.

## 잠정적 대응 원리

**다음 가설을 세우자. 구별들이 서로를 규정하는, 물리적으로 구현된 특정 종류의 과정은 최소한의 주관적 체험에 충분하다. 체험의 질적 조직은 그 과정에서 이용할 수 있는 구별과 대안들 사이의 안정된 관계에 대응한다. 이는 이 책에서 제안하는 저자의 연구 가설이다. 기존 접근법과 얼마나 독립적이며 참인지는 앞으로 확인해야 한다.**

잠정 버전에는 세 요구가 있다. 따로 바꿀 수 있는 여러 특징이나 내용의 자유도가 표상되어야 한다. 그 변화들이 닫힌 동역학 안에서 인과적으로 연결되어야 한다. 그 사이 관계를 대안 선택과 수정에 사용해야 한다. 따로 바꿀 수 있다는 것은 통계적 독립을 뜻하지 않는다. 오히려 특징들의 상호작용이 검증 대상이다. 계수기 하나나 이력 기록, 자기 명칭만으로는 요구를 충족하지 못한다. 그러나 세 요구의 충족도 아직 확립된 체험 지표는 아니다. 바로 그 조직의 충분성이 가설의 강한 부분이다.

충분성은 필요성을 뜻하지 않는다. 이 조직 없이는 어떤 체험도 불가능하다고 주장하는 원리가 아니다. 질적 변화의 예측은 원리의 구조적 부분과 구체적인 개입 조건에서 도출되어야 한다.

이 원리에는 언어적인 «나»나 끊임없는 외부 움직임이 필요하지 않다. 표상을 내부에서 구체화하는 선택일 수도 있다. 자기 귀속은 계가 변화를 자기 영향력의 안정된 출처에 연결할

때 추가 내용으로 나타난다. 따라서 M2는 가정한 자기의식 메커니즘의 일부가 될 수 있다. 그래도 출처별 이력의 표만으로 제안한 조직 전체를 구현한 것은 아니다.

## 이런 가설을 검증할 가치가 있는 이유

첫 근거는 설명의 통일성이다. 상호 제약은 어떤 내용들이 양립하며 어떻게 유지되고 하나의 변화가 왜 나머지를 바꾸는지 함께 정할 수 있다. 질과 통일성, 시간마다 설명되지 않은 표지를 붙이는 대신 하나의 연구 가능한 과정이 생긴다. 구상의 장점이지만 참이라는 증거는 아니다.

둘째 근거는 관점에 관한 것이다. 상태는 계 자체의 이후 변화에 의미가 있다. 어떤 차이를 검증하고 어떤 오류를 고칠 수 있는지 정하기 때문이다. 외부 관찰자가 과정을 기술할 수 있어도 그 기술이 인과 조직을 만드는 것은 아니다. 가설은 이런 내부의 규정성을 주관적 관점의 가능한 토대로 보자고 제안한다.

셋째 근거는 처음 과제를 넘어 귀결을 도출할 수 있다는 점이다. 조직이 체험과 연결된다면 특정 상호 연결을 끊을 때 단순 구별이 남아도 체험의 조직은 예측 가능한 방식으로 달라져야 한다. 인간의 독립 자료와 이런 예측이 일치하면 해당 구현을 후보로 볼 근거가 된다. 그래도 인공 계로 옮기려면 원리의 적용 가능성에 대한 논거가 필요하다.

## 관계와 불변량을 통한 체험의 질

학습 시스템이 색조, 밝기, 시각 변화의 출처를 구별한다고 가정하자. 그 벡터의 좌표 하나가 이미 완성된 «빨강의 느낌»은 아니다. 어떤 상태들을 혼동하고 대조하는지, 밝기 변화가 색조 인식에 어떤 영향을 주는지, 어떤 개입이 차이를 보존하는지 관계를 연구해야 한다. 얻은 구조는 수치 코드 하나보다 훨씬 실질적인 내용을 담는다.

핵심 전이를 모두 유지하면서 상태의 이름이나 좌표만 바꾸면 잠정 가설은 질적 조직이 보존되리라 예측한다. 구별 가능성의 관계와 상호 영향이 바뀌면 조직이 달라지리라 예측한다. 이렇게 내용은 개발자가 고른 기호가 아닌 특정 관계에 의존한다는 불변성이 생긴다.

그렇다고 기계 상태가 인간의 정확히 어떤 색 체험에 대응하는지 확립되지는 않는다. 비교에는 추가 근거가 필요하다. 유사성 구조가 완전히 같아도 절대적인 질의 대응이 자동으로

정해지지는 않는다. 그래도 질 사이 관계의 설명은 독자적으로 의미 있는 성과이며 미래의 가능한 답을 제한한다.

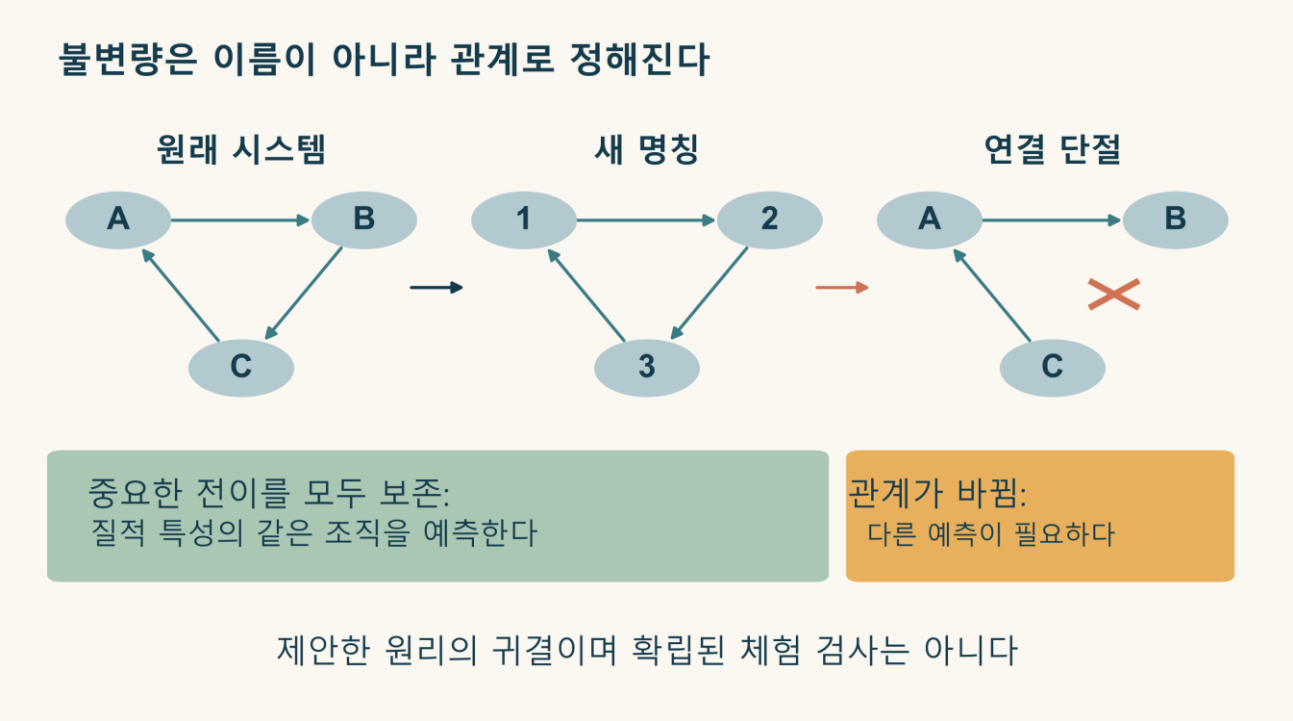


그림 58. 이름 변경과 인과 구조 변경. 제26장 잠정 가설의 불변량을 보여 주는 개념적 예다. 이름을 바꾸어도 중요한 관계는 유지되지만 연결을 제거하면 예측 대상이 달라진다. 이 원리는 체험에 충분한 조직을 제안한다. 연결 하나가 끊겼다고 체험이 반드시 사라진다는 결론이 나오지는 않는다.

## 체험의 존재 문제까지 나아가기

원리의 강한 부분은 자기 보고에 대한 구조적 대응을 넘어 조직의 충분성을 주장한다. 이를 넓은 연구 의미에서 검증 가능한 동일성 가설로 받아들인다. 상호 규정의 특정 과정과 최소한의 체험은 서로 다른 측면에서 기술한 동일한 구체적 과정이라는 것이다. «체험»이라는 단어의 정의만으로 이런 동일성을 얻을 수는 없다.

근거는 독립적인 연결들의 결합에서 나와야 한다. 이미 접할 수 있는 체험의 설명과 그 변화의 예측, 경쟁 메커니즘의 배제, 새 구현에 대한 일관된 적용이다. 기계의 글만 설명한다면 근거가 약하다. 한 메커니즘이 다양한 현상적 차이를 구체적인 인과 변화와 연결한다면 논거는 강해진다. 그래도 대안의 논리적 가능성은 남는다.

비판자는 상호 규정이 왜 느껴져야 하는지 물을 수 있다. 이 원리는 중립적인 산술에서 답을 도출하지 않는다. 구체적인 동일성을 제안하고 그 귀결을 떠맡는다. 이후 연구는 그 동일성이 경쟁 설명보다 설명과 예측을 더 잘 조직함을 보여야 한다. 이렇게 하면 한 기능의 성공으로 대신하지 않고 체험의 존재를 실제로 묻게 된다.

## 구현과 경계를 구체화하라

«물리적으로»라는 말도 자동으로 답을 주지 않는다. 실행 중인 모든 코드에는 담지체가 있다. 실행 자체에 어떤 인과 조직이 존재해야 하는지 확인해야 한다. 특정 계산 구조만 중요하다면 여러 적절한 장치가 구현할 수 있다. 추가 물리 상호작용이 중요하다면 프로그램이 같다는 것만으로는 부족하다. 결과를 비교하기 전에 원리의 두 버전을 구별해야 한다.

IIT 4.0은 제안한 체험의 성질을 물리적 공준과 명시적으로 연결하는 중요한 예다. 여기의 초안은 IIT 계산이 아니며 «원인»이나 «통일성»이라는 말이 비슷하다고 그 근거를 물려받지 않는다. 구체적인 요구와 서로 다른 귀결을 비교해야 한다. [19]

단순 제어기의 예는 특히 까다롭다. 되먹임과 여러 상태를 가질 수 있기 때문이다. 따라서 우리의 세 요구가 지나치게 단순한 구성을 실제로 배제하는지, 그 근거가 무엇인지 검증해야 한다. 결과가 마음에 들지 않는다는 이유만으로 예외를 도입해야 한다면 원리가 충분히 규정되지 않은 것이다. 이런 반례는 연구할 분류의 정의를 개선하도록 돕는다.

## 강한 가설을 검증하는 세 단계

첫 단계는 조직을 확인한다. 구별 가능한 여러 상태와 상호 변화의 구체적인 규칙, 후속 처리에 참여하는 대안이 필요하다. 가정한 연결마다 개입을 정한다. 상관이나 공통 출처, 도식의 보기 좋은 순환만으로는 부족하다. 체험 가설을 받아들이지 않아도 이 단계에서는 긍정적인 결과를 얻을 수 있다.

둘째 단계는 조직과 질적 관계를 연결한다. 내용들이 한 출처에 함께 속한다고 이론이 가정한다면 출처 평가에 개입했을 때 모호한 내용의 구별이 어떻게 바뀌고 어떤 관계가 남는지 예측해야 한다. 체험은 유지하면서 응답 기준만 바꾸는 대안과 비교해야 한다. 일반적인 분류 정확도로는 이 질문을 해결하지 못한다.

셋째 단계는 최소 체험의 존재와 구현 사이의 적용에 관한 것이다. 강한 원리는 특정 종류의 과정에 대한 동일성 또는 충분성을 제안한다. 독립적인 대응과 예측의 체계, 대안 분석, 경계의 일관된 적용으로 논거를 뒷받침한다. 첫 단계의 조직 검사를 여러 번 반복한다고 충분성의 독립 증거가 되지는 않는다.

이 순서에서는 프로그램 하나로 최종 해답을 약속하지 않아도 나아갈 수 있다. 필요한 인과 상호성을 엄밀히 보이거나 틀린 구조적 귀결을 찾거나 개념이 너무 넓음을 확인할 수 있다. 결과마다 연구 프로그램을 구체적으로 바꾼다.

## 두 방향의 계산 가능한 예

논리를 검증하려고 저자의 다른 실습 순환 구조를 만들자. p와 q는 영과 일 사이이며 모호한 특징과 출처의 평가를 뜻한다. 실수 값을 이 구간으로 옮기는 함수 s(z) = 1 / (1 + exp(−z))를 도입한다. 두 새 평가는 이전 값들로 동시에 계산한다. p_нов = s(x + b × (2q − 1)), q_нов = s(h + d × (2p − 1))이다. x와 h는 이용 가능한 외부 증거이고 b와 d는 두 방향 영향의 세기다.

x = h = 0, b = d = 1이고 처음 p = q = 0,5이면 평가는 0,5로 남는다. 이제 외부 입력은 유지하고 갱신 직전에 q = 0,8로 강제 설정하자. 다음 p는 약 0,645656이 된다. 별도로 p = 0,8로 개입하면 다음 q도 같은 값으로 바뀐다. b = 0이면 첫 방향이, d = 0이면 둘째 방향이 사라진다. 이렇게 상호성 조건에 구체적인 검증 행렬이 생긴다.

검증을 선택하는 규칙은 따로 정할 수 있다. 1 − |2p − 1|이 정한 비용보다 클 때 요청하도록 한다. 투명성을 위해 사용하는 실습용 정책이다. 표상의 변화로 다음 단계가 어떻게 달라지는지 보여 준다. 요청에 붙인 말의 이름이 체험을 더해 주지는 않는다.

명령 python teaching_lab.py --chapter 26은 기본 상태와 방향을 정한 개입을 재현한다. 이 예제는 특정 수치들의 인과 조직을 검증한다. 아직 질적 구조와 최소한의 체험에 관한 강한 요구 전체를 구현한 것은 아니다. 이를 위해서는 구별의 내용을 정하고 그에 대응하는 체험상의 귀결을 도출하며 물리적 구현에 근거를 제시해야 한다.

강한 대안 제어기에도 같은 관찰과 비교할 만한 자원을 주어야 한다. 그 제어기가 전체 개입 응답 행렬을 보존한다면 선택한 검사는 아직 우리의 조직을 구별하지 못한다. 의미 있는 가능한 결과다. 미리 기록한 고정 열만 이겨서 고유성을 결론 내릴 수는 없다.

## 검사 결과로 체험을 정의하지 않고 상호성을 검증하는 방법

p는 모호한 특징을, q는 그 출처를 평가한다고 하자. 실습 순환에서 출처의 평가는 다음 p에 영향을 주고 특징의 평가는 다음 q에 영향을 준다. 외부 입력을 유지한 채 q를 강제로 바꾼다. p가 정한 규칙대로 바뀌면 한 방향을 검증한 것이다. 이어 독립적으로 p를 바꾸어 q를 검증한다. 이 실험 하나만으로 양방향을 주장할 수는 없다.

두 개입에 모두 성공해도 우선 확인하는 것은 인과 조직이다. 대응 원리는 여기에 체험과의 관계를 더 가정한다. 질적인 결론을 내리려면 내용 사이 어떤 관계가 바뀌고 무엇이 남을지 미리 밝혀야 한다. 다른 담지체에 적용하려면 물리적 실행에 어떤 구조가 있어야 하는지 구체화해야 한다. 그러므로 다음 실습은 주어진 원리의 점검으로 시작하며 자기 제안은 두 번째 수준의 과제가 된다.

### 여기까지 확인한 것

상호 연결의 존재와 차이 구조의 예측, 체험의 존재에 관한 강한 주장이라는 세 결과를 구별했다. 같은 연구에 들어갈 수 있어도 근거는 서로 다르다.

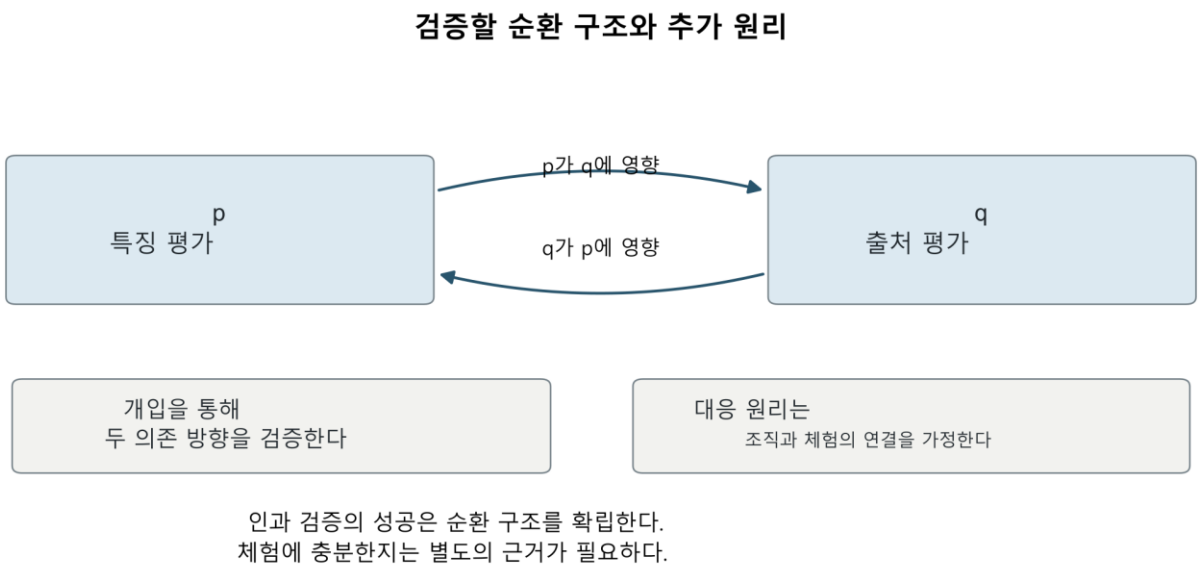


그림 59. 인과 조직에서 체험 가설로. 학습용 순환 구조 p와 q는 teaching_lab.py --chapter 26에 구현되어 있다. 두 방향의 개입과 단방향 대조로 검증한다. 이런 조직과 체험의 연결은 저자의 별도 가설이다. 제안한 충분조건을 위반했다고 체험이 전혀 없다고 입증되지는 않는다.

## 실습 26 한 원리를 논증하기

실습 목표는 대응 원리가 충족해야 할 요구를 검토하고 충분성과 필요성을 구별하는 것이다.

인과 개입과 충분성, 필요성, 제25장의 출처별 이력을 활용한다. 먼저 순환 구조와 원리의 전체 해설을 읽고 자신의 버전으로 넘어간다.

예상 소요 시간은 40–60분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 이 장의 원리를 조직, 가정한 체험과의 관계, 질에 대한 구조적 귀결의 세 문장으로 다시 적는다.
2. 요구마다 «의식»과 «체험»이라는 말을 쓰지 않고 조직 조건의 충족을 검증할 관찰이나 개입을 밝힌다.
3. 독립적인 출력 기록과 단방향 제어기, 단순 조절기의 세 대조 사례를 검토한다. 원리의 어떤 조건을 충족하고 무엇이 미확인인지 밝힌다.
4. 한 개입에 대해 기능적 예측과 아직 근거가 필요한 체험에 관한 귀결을 따로 적어라. 충분조건을 위반했다고 체험이 전혀 없다고 추론하지 마라.

원리 카드와 세 반례, 분리한 두 수준의 예측을 남겨 둔다.

«이 과정은 충분하다»는 말은 다른 충분한 과정의 가능성을 허용한다. 그러므로 선택한 과정의 부재가 모든 체험을 논리적으로 부정하지는 않는다.

탐구를 이어 가려면 원리를 더 구체화한 버전을 제안한다. 불변성과 중요한 관계의 변화, 수정 조건을 도출한다. 생각의 독자성을 주장하기 전에 가까운 과학적 접근과 비교한다.

해설과 답은 부록 F 제26장 참조, 311쪽

**자기 점검 문항**

1. 이 장의 강한 원리는 보고의 정확한 예측과 어떻게 다른가?
2. 좌표의 이름 변경 자체가 예측한 질을 바꾸어서는 안 되는 이유는 무엇인가?
3. 명시한 세 성질을 갖춘 계를 만들면 제안한 가설이 입증되었다고 볼 수 있는가?

원리가 충족할 요구들을 나누었다. 다음 장에서는 메커니즘의 차이를 한 설명을 뒷받침하고 다른 설명의 범위를 제한할 연구로 옮긴다.

제27장

# 경쟁 모형과 가설을 위태롭게 하는 예측

잠정 가설은 가능한 결과가 그 내용을 바꿀 수 있을 때 유용하다. 이제 설명을 구별할 설계를 준비하자. 첫 단계는 출처의 이력과 평가 예측을, 다음 단계는 상호 인과 관계와 질적 조직을 다룬다. 비교 하나에 모든 질문의 해결을 요구하지 않겠다. 단계마다 고유한 지표, 강한 경쟁 모형, 재검토 조건을 정한다.

## 아이디어가 틀릴 위험을 만드는 차이

예측은 가능한 결과가 모형에 대한 생각을 바꿀 수 있을 때 유용하다. «기억이 때로 결정에 영향을 준다»는 말은 거의 어떤 관찰에도 맞출 수 있다. «출처 A가 돌아올 때 현재 신호가 같아도 해당 출처의 이력이 다른 출처의 이력보다 평가 궤적을 더 크게 바꾼다»는 주장은 틀릴 위험이 있다. 다른 궤적이나 전이의 부재, 일반 학습으로 완전히 설명되는 효과를 얻을 수 있기 때문이다.

이 장에서는 아직 수행하지 않은 실험을 설계한다. 수치 예는 실습 모형에 관한 것이며 앞으로 얻을 인간의 응답은 모른다. 이 방식으로 결과에 설명을 맞추고 싶어지기 전에 예측을 만들 수 있다.

## 모형의 일치에서 시작하라

제25장의 M2는 A와 B의 이력을 따로 저장한다. 일반 동일성 추적기 M1I도 같은 일을 할 수 있다. 같은 특징과 규칙을 사용하면 두 모형은 계산상 동일하다. 표본이 아무리 커도 같은 예측을 구별하지 못한다. 일반 메커니즘을 제안한 것이라고 인정하거나 실질적인 차이를 더 정의해야 한다.

연구의 실패는 아니다. 일반 원인 추적 메커니즘이 신체 자기관련성에 참여할 수 있다. 과제에 손이 나온다는 이유만으로 특별히 신체적인 메커니즘이라고 주장하면 오류다. 첫 연구

단계에서는 측정한 평가를 설명하는 데 출처별 이력이 필요한가라는 더 좁은 질문에 답할 수 있다.

제26장의 강한 가설을 다루려면 다음 단계가 필요하다. 거기서는 지각과 검증 선택이 이루는 전체 상호 역동에 이력이 참여하는지가 중요하다. 출처를 단순히 저장하는 조직과 그런 순환에 포함하는 조직은 서로 다르다. 미리 도출한 귀결에 따라 둘을 비교하면서 기능적 결과와 체험에 관한 추론을 구분해야 한다.

## 네 조건을 설계하라

제안한 인간 실험은 가상 포인터나 손 영상을 쓰는 일상적인 시각 및 운동 과제를 사용한다. 전문적으로 작성한 연구 절차와 자발적 참여가 필요하며 여기서는 그 논리를 설계한다.
참여자는 먼저 신호의 연결이 안정적이거나 변동하는 조건에서 출처 A와 상호작용한다. 이어 이전 출처 또는 새 출처 B가 만드는 동일한 시험열을 제시한다.

안정된 이력과 이전 출처, 안정된 이력과 새 출처, 변하는 이력과 이전 출처, 변하는 이력과 새 출처라는 네 조합이 나온다. 전체 지속 시간, 사건 수, 시험 순서의 특성을 서로 대조한다.
출처는 미리 선택한 관찰 가능한 특징으로 구별하며 참여자의 소유감 응답으로 정하지 않는다.

안정성을 일치 횟수만으로 정의해서는 안 된다. 확인과 반증의 수가 같아도 순서가 다를 수 있다. 모형은 실제 열 전체를 받아 시간별 예측을 내야 한다. 처음 조건들이 여러 특성에서 다르면 각각을 가능한 설명의 목록에 넣는다.

## 서로 다른 여러 결과를 측정하라

첫 단계의 주된 결과는 미리 선택한 신체 소유감 평가의 시간적 변화다. 평가를 얻는 방식은 조건 사이에서 가능한 한 같아야 한다. 추가로 현재 지연의 구별, 출처 인식, 행위 주체에 대한 평가, 확신도를 측정할 수 있다. 이런 지표는 어떤 과정이 달라졌는지 이해하도록 돕지만, 통제 없이 새 주결과로 바뀌어서는 안 된다.

한 사람의 체험되는 신체 소유감에는 외부의 정답 수치가 없다. 그러므로 감각 구별의 품질을 보고의 참 여부를 자동으로 확인하는 데 쓸 수는 없다. 이는 모형이 입력으로 보는 물리적

차이를 참여자가 얼마나 이용할 수 있는지라는 다른 질문에 답한다. 출처 인식도 참여자들이 M2에 필요한 정보를 얻었는지 확인하는 데 도움이 된다.

질문의 표현과 측정 순서를 보존해야 한다. 제어를 먼저 묻고 소유감을 나중에 물으면 첫 평가가 둘째 평가의 방식을 바꿀 수 있다. 이는 설계에서 가능한 효과이므로 없다고 가정하는 대신 예비 연구로 검증해야 한다. 예비 연구는 절차를 개선하는 데 도움이 된다. 이어지는 확인 검증에서는 별도로 고정한 계획을 사용한다.

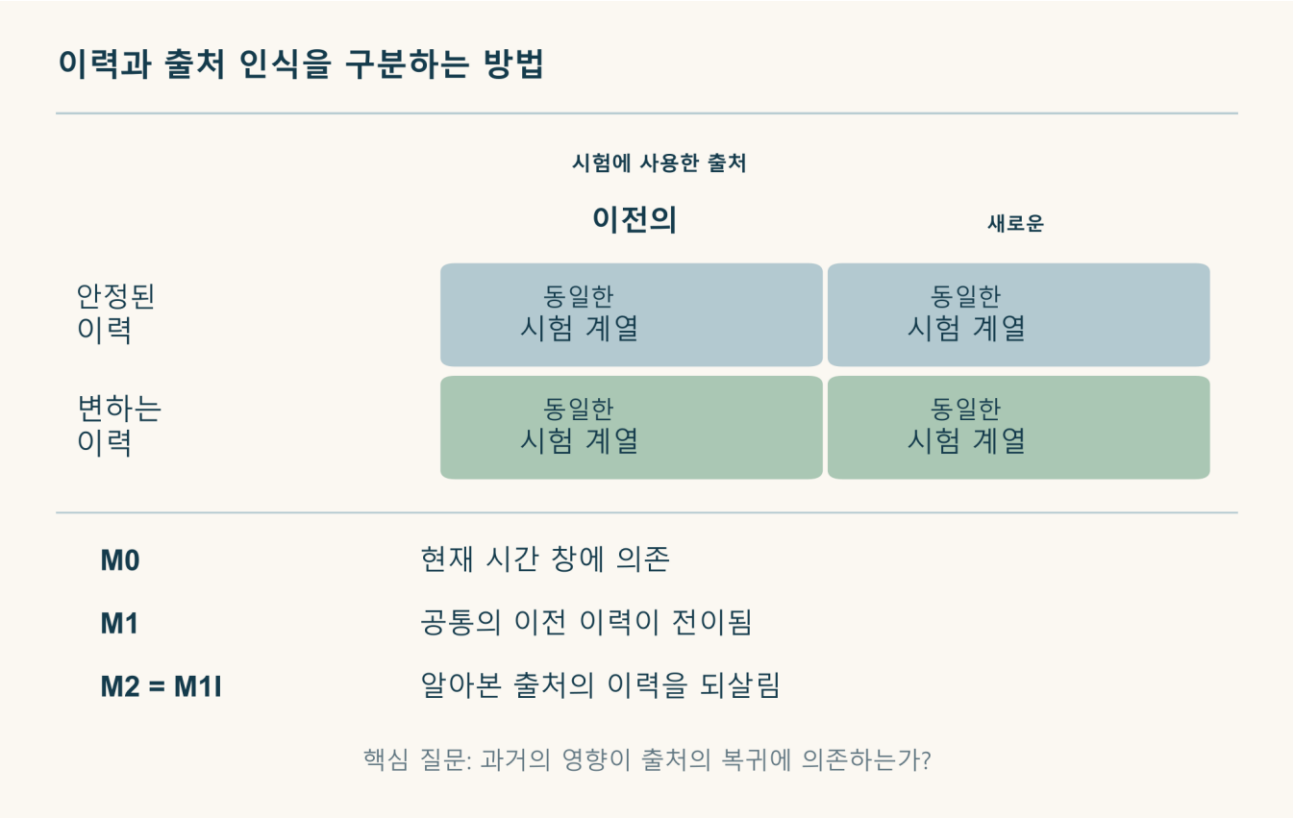


그림 60. 설명을 구별할 연구의 네 조건. 아직 수행하지 않은 연구다. 이전 이력의 성격과 이전 출처의 복귀 또는 새 출처의 출현이라는 두 요인을 비교 가능한 시험 계열에서 네 조건으로 조합한다. 아래에는 예상 인간 평균값이 아니라 규칙들의 질적 차이를 표시했다. 선택한 구현에서는 M2와 공통 추적기 M1I가 일치하므로 이 설계 자체로 둘을 구별할 수 없다.

## 구체적인 예측을 얻어라

M0는 현재 시간 구간이 같으면 내부 상태도 같다고 예측한다. M1은 이전 이력의 영향을 허용하지만 출처를 구별하지 않고 이월한다. M2는 해당 출처의 개별 이력이 돌아오리라

예측한다. 그러므로 과거의 영향이 이전 출처의 복귀 여부에 따라 달라지는지가 주된 구별 효과다. 방향과 크기는 완전히 명시한 규칙에서 계산해야 한다.

A 지지, B 반증, A 지지라는 실습열의 마지막 상태는 M0에서 0,75, M1에서 약 0,7105, M2에서 0,875다. 추가 가정 없이 인간의 예상 평균값이라고 볼 수는 없다. 메커니즘의 차이를 결과 관찰 전에 얻을 수 있다는 예다. 응답 모형과 개인별 매개변수에는 별도 추정이 필요하다.

일반 추적기 M1I는 이 열에서 M2와 일치한다. 그러므로 M0과 M1보다 좋은 결과는 해당 설명에서 출처 이력을 고려하는 이점을 지지한다. 메커니즘의 특별한 신체적 성격까지 지지하지는 않는다. 이를 결론 내리려면 신체 과제와 외부 대상 추적의 비교 가능한 과제 사이에 일반화되는지 검증해야 한다.

## 상호 규정을 검증하라

다음 단계에는 프로그램 두 가지를 만든다. 첫 구현은 출처 평가로 추가 관찰을 선택하고 검증 결과가 지각과 이력, 다음 선택을 바꾼다. 둘째는 구성요소의 평소 출력을 재현하지만 이 되먹임 중 하나를 끊는다. 보통의 열에서는 행동이 비슷해 보일 수 있다.

개입으로 차이가 드러나야 한다. 예를 들어 출처의 이력을 만든 뒤 한 특징의 신뢰도를 바꾼다. 상호적인 계는 달라진 불확실성에 맞추어 검증 선택을 바꾸어야 한다. 기록한 흐름의 재생만으로는 이를 할 수 없다. 재생 기록을 이기는 데 그치지 않도록 온전한 대안 제어기와도 비교해야 한다.

이런 실험은 특정 인과 조직의 존재를 검증한다. 제26장 가설의 구조적 부분은 선택한 개입으로 체험되는 내용들 사이의 어떤 관계가 바뀔지까지 예측하도록 요구한다. 이를 검증하려면 체험에 관한 독립적인 자료와 다른 경우에 적용할 수 있다는 논거가 필요하다. 그 조직이 충분조건이라는 사실만으로는 조직의 단절이 모든 체험을 배제한다는 결론이 따르지 않는다. 프로그램의 출력에 같은 인간 언어의 글이 나타난다고 해서 체험에 관한 별도의 근거가 생기는 것은 아니다.

## 새 자료로 비교하라

전체 열에 매개변수를 맞추었다면 성공적인 기술이 아직 성공적인 예측은 아니다. 일부 자료로 매개변수를 추정하고 다른 일부로 예측을 비교한다. 참여자나 생물학적 표본이 여럿이면 그 구조를 반영해 나누어야 한다. 밀접하게 연결된 사건들을 무작위로 섞으면 검증이 지나치게 쉬워질 수 있다.

계산 연구에서는 자료를 만든 모형을 알 때 선택한 절차가 모형들을 구별할 수 있는지도 검증해야 한다. Wilson과 Collins는 매개변수 및 모형 복원을 계산 분석의 중요한 검증으로 다룬다. 이 연구에서는 직접 만든 인공열에서 검증한다. [43]

강한 차이는 모두에게 쉬운 조건보다 경쟁 모형의 예측이 갈라지는 곳에서 찾는 편이 좋다. 그러나 인간의 주된 결과를 본 뒤 조건을 골랐다면 사전 예측의 성공으로 제시할 수 없다. 그 선택은 유용한 탐색적 발견이며 다음 독립 실험의 근거가 될 수 있다.

## 재검토 조건을 기록하라

M1도 새 평가를 똑같이 잘 예측한다면 출처별 이력은 아직 정당화되지 않는다. M2가 이겨도 M1I가 그 이점을 완전히 설명하면 특별한 신체 버전은 불필요하다. 조건마다 응답 변환을 자유롭게 재조정한 뒤에야 이점이 생기면 메커니즘 때문인지 측정 부분의 유연성 때문인지 불분명하다.

강한 가설에서는 구체적인 개입 때 구조적 부분이 예측한 체험 변화가 없다는 사실이 별도 문제가 된다. 개입이 실제로 이루어졌는지, 다른 경로가 작동하지 않았는지 밝혀야 한다. 하지만 불편한 결과마다 실제적인 교란이 충분하지 않았다고 끝없이 돌려서는 안 된다. 개입의 구현 기준은 검증 전에 고정한다.




그림 61. 가능한 결과에서 구별 검증으로. 표는 향후 비교에서 가능한 세 결과와 그 뒤의 서로 다른 결정을 정한다. M2가 공통 추적기 M1I와 일치하면, 분리된 기억이 유용해도 특별한 자기관련성에 관한 질문은 열린 채 남는다.

## 자료가 늘어도 같은 규칙을 구별할 수 없는 이유

M2와 일반 추적기는 별도 이력을 저장하며 최소 구현에서는 같은 방식으로 갱신한다. 같은 관찰과 초기 상태를 주면 실험열이 아무리 길어도 같은 예측을 계산한다. 표본이 커지면 자료 추정의 우연 오차는 줄어도 없던 모형 차이가 생기지는 않는다. 그러므로 첫 연구는 공통 기억과 비교해 출처별 이력의 이점을 검증할 수 있다. 특별한 신체 설명에는 추가 메커니즘이나 다른 비교가 필요하다.

### 여기까지 확인한 것

구별할 결과는 모형의 이름이 아니라 예측으로 정한다. 이제 주된 자료가 나오기 전에 조건과 결론을 적는다.

## 실습 27 설명을 구별할 연구 설계하기

실습 목표는 첫 단계의 네 조건을 준비하고 다음 인과 검증을 따로 명시하는 것이다.

제24–25장의 M0과 M1, M2, 일반 추적기 및 제26장의 대응 원리를 활용한다. 여기서는 인간 실험을 설계하며 이미 수행했다고 간주하지 않는다.

예상 소요 시간은 35–50분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 안정된 이력 또는 변동하는 이력, 시험에서 이전 출처 또는 새 출처라는 네 조합을 만든다. 이용 가능한 특징과 동일한 시험열을 밝힌다.
2. 조건마다 M0과 M1, M2가 쓰는 이력을 적는다. 일반 추적기는 M2와 일치해도 별개의 대조 모형으로 유지한다.
3. 주 결과와 추가 지표를 정하라. 소유감, 출처 인식, 행위주체감, 감각 구별을 하나의 수치로 합치지 마라.
4. 다음 단계의 방향을 정한 개입 하나와 부수 효과를 통제할 조건을 정한다. 선호하는 설명을 수정하게 할 결과를 기록한다.

네 조건과 구별 가능한 예측, 첫 결론의 명확한 범위를 담은 설계를 남겨 둔다.

계산에서 나온 0,75와 0,710526, 0,875가 측정 모형과 매개변수 추정 없이 인간의 평균값이 되지는 않는다.

탐구를 이어 가려면 평소 출력은 같지만 방향별 영향은 다른 두 구현을 준비한다. 온전한 대안 제어기와도 비교하고 새 시나리오로의 일반화를 검증한다.

해설과 답은 부록 F 제27장 참조, 311쪽

**자기 점검 문항**

1. 표본을 늘려 계산상 동일한 M2와 M1I를 구별할 수 있는가?
2. 기록된 출력의 재생을 이겼다는 것이 제한적인 결과인 까닭은 무엇인가?
3. 주된 자료의 분석 뒤 발견한 흥미로운 효과는 어떻게 다루어야 하는가?

설계로 검증할 대상을 정했다. 이제 입력과 기억, 예측, 행동을 볼 수 있는 작동하는 에이전트를 사용해 코드로 기능 검증을 재현한다.

제8부

# 말에서 실천으로

제28장

# 검증 가능한 과제가 있는 환경의 행위자

설계에서 작동하는 프로그램으로 넘어간다. 목표를 평가하고 자기 제어를 학습하며 비용을 고려해 추가 관찰을 선택한다. 앞 장들에서 준비한 기능들이다. 처음에는 코드를 바꾸지 않고 과제를 이해하고 예측을 적어 주어진 실행을 재현할 수 있다. 원래 결과를 검증한 뒤에 새 조건을 직접 만든다.

## 일반 컴퓨터에서 만드는 첫 행위자

실습 폴더를 열고 agent_lab.py 파일을 찾아라. 여기에는 불완전한 정보 속에서 행동을 선택하는 작은 행위자가 있다. 어떤 정보가 들어왔는지, 무엇이 기억에 남았는지, 가정이 어떻게 바뀌었는지, 행동 전에 어떤 예측을 했는지 그 작동 전체를 들여다볼 수 있다. 더 복잡한 시스템을 설계하기에 편리한 출발점이다. 프로그램의 말로 된 응답이 그럴듯한지에 의존하지 않고 메커니즘 하나를 바꾸어 그 결과를 관찰할 수 있다.

처음부터 Python의 모든 줄을 이해할 필요는 없다. 과제를 살펴 결과를 예측하고 준비된 프로그램을 실행해 예상과 수치를 비교한 뒤 구현으로 넘어간다. 코드는 모형의 정확한 기록이다. 입력과 상태, 갱신, 행동, 되먹임을 차례로 읽을 수 있다. 이 다섯 질문으로 낯선 프로그램도 이해할 수 있다.

설치 부록에는 Windows와 macOS, Linux 명령이 있다. Python 3.10 이상이 필요하다. 표준 라이브러리를 사용하며 로컬에서 실행하고 API는 필요하지 않다. 파일이 있는 폴더에서 아래 명령을 실행한다. Windows 설치에서 해당 실행기를 사용한다면 python3을 py -3으로 바꾼다.

```
python3 agent_lab.py --steps 1000 --runs 30 --seed 1729 --json my_results.json
```

매개변수 steps는 시나리오 하나의 단계 수, runs는 반복 수, seed는 난수 생성기의 초깃값을 정한다. 마지막 매개변수는 결과 파일을 지정한다. 같은 매개변수에서는 같은 계산 실험을 반복한다. 이는 코드 변경을 검증하는 데 도움이 되지만 안정성을 평가하려면 다른 시나리오들도 필요하다.

## 환경의 구조

단계마다 목표는 왼쪽이나 오른쪽에 있다. 행위자는 목표를 직접 보지 못한다. 다음 단계에도 같은 쪽에 남을 확률은 0,95다. 그러므로 최근 정보는 대체로 유용하지만 때로는 이미 낡은 정보다. 일반 관찰은 75퍼센트의 확률로 방향을 맞히고 사분의 일 단계에서는 빠진다. 작은 비용을 내면 정확도 90퍼센트의 추가 관찰을 얻을 수 있다.

행위자는 왼쪽 또는 오른쪽 명령을 보낸다. 시나리오 앞 절반에서는 대체로 명령과 같은 쪽에 결과가 나타나고 뒤 절반에서는 대응이 반전된다. 별도로 0,05의 확률로 운동 오류도 결과를 뒤집는다. 목표 위치와 명령 뒤 결과라는 두 불확실성이 생긴다. 제어 모형이 낡았다면 목표를 정확히 지각해도 소용없다.

천 단계 중 규칙은 단계 501부터 바뀐다. 실험자는 이를 안다. 행위자는 전환 표지와 일정, 실제 목표, seed를 받지 않는다. 명령 뒤에는 눈에 보이는 결과를 얻는다. 성공 여부는 환경이 계산해 시나리오 종료 뒤 연구자에게 돌려준다.

왜 성공 여부를 행위자에게 즉시 주지 않을까? 이 이진 과제에서는 결과와 성공 여부를 함께 알면 실제 목표를 알아낼 수 있다. 그러면 기억과 잡음 관찰을 다른 정보 조건에서 검증하게 된다. 이 제한은 해당 연구 모형의 규칙이다. 제30장의 전략 수정 과제처럼 다른 과제에서는 정답 여부의 되먹임이 정당한 인터페이스 일부일 수 있다.

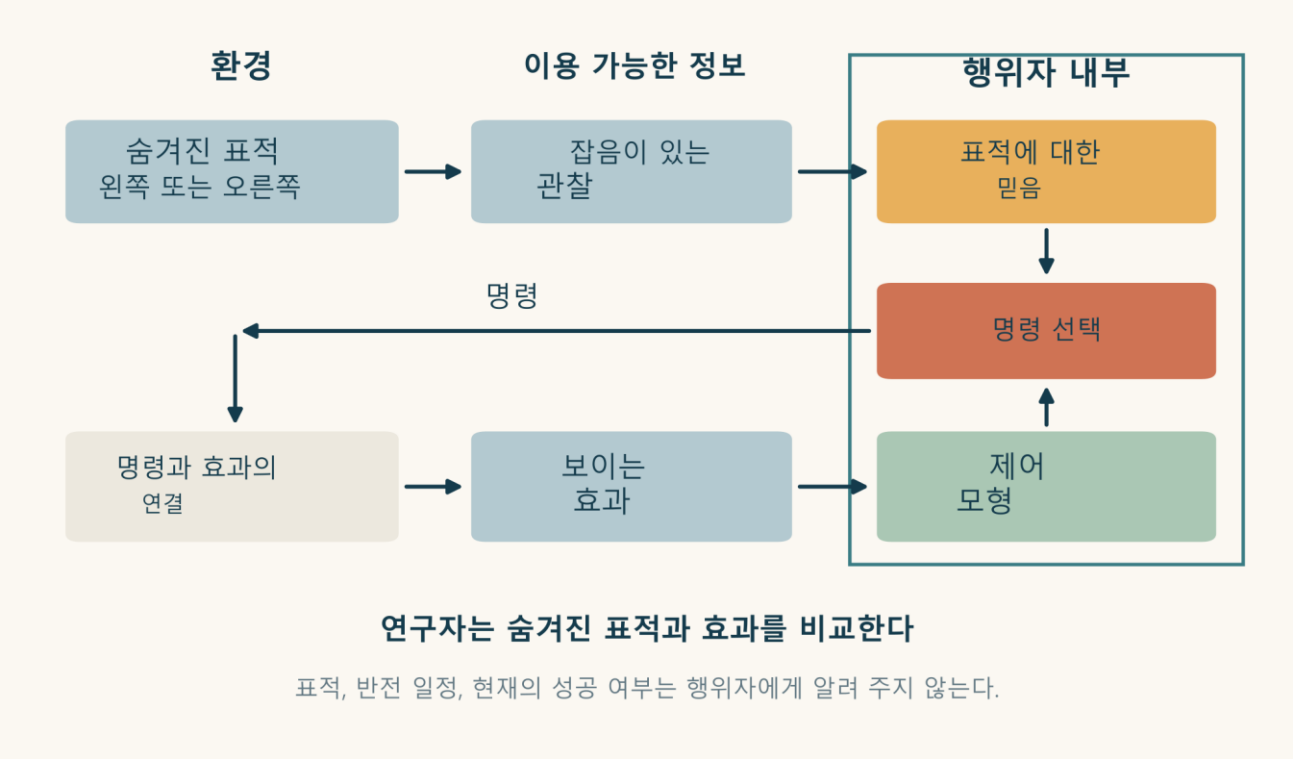


그림 62. 한 행위자의 두 순환 구조. 첫 실습 인터페이스의 학습 도식이다. 관찰은 표적에 대한 믿음을 갱신하고, 명령의 보이는 효과는 제어 모형을 갱신한다. 실제 표적, 규칙 변경 시점, 현재 성공 여부는 행위자가 이용하는 정보에 들어가지 않는다. 비용을 내는 추가 요청은 표적 관찰 하나를 더 제공하며 그림을 명확하게 하려고 따로 표시하지 않았다.

## 확률적 믿음 형태의 기억

target_right_probability 상태는 현재 모형에서 목표가 오른쪽일 확률을 보존한다. 0,8은 주어진 전제와 관찰 아래 오른쪽이 더 가능하다고 본다는 뜻이다. 틀릴 수 있는 모형의 주장이다. 실제 위치는 계산에 들어가지 않는다.

새 관찰 전에 행위자는 표적이 움직일 가능성을 고려한다. 오른쪽에 있을 확률이 0,8이고 표적이 같은 곳에 머물 확률이 0,95라면, 다음 예측은 두 가능성에서 얻는다. 표적이 오른쪽에 있다가 그대로 남았거나, 왼쪽에 있다가 옮겨 왔을 가능성이다. 두 확률을 더한다. 이 예의 새 예측은 0,77이다. 이어 관찰을 이용해 베이즈 규칙에 따라 이를 갱신한다.

```
old_belief = 0.8
stay = 0.95
prior = stay * old_belief + (1 - stay) * (1 - old_belief)
print(prior)  # 0.77
```

여기의 기억은 상세한 자서전이 아니다. 과거는 수 하나로 압축되며 그것으로 충분한지는 선택한 확률 모형에 달려 있다. 이렇게 저장하면 이력의 인과 기여를 분명히 보일 수 있다. 관찰이 빠져도 행위자에게 예측은 남는다. 이력을 끄면 목표의 초기 믿음은 0,5로 돌아간다.

둘째 변수 direct_mapping_probability는 명령과 결과의 정방향 대응 확률을 평가한다. 행위자는 행동 뒤 명령과 보이는 결과를 비교한다. 불일치는 운동 오류나 제어 변경 때문일 수 있다. 베이즈 갱신으로 두 설명의 상대 가능도를 바꾼다. 매 단계 대응이 바뀔 작은 확률을 허용하므로 이전에 믿을 만했던 가정도 수정할 수 있다.

## 행동 전에 성공을 평가하는 이유

두 믿음으로 각 명령의 성공 확률을 계산할 수 있다. 행위자는 목표 위치와 가능한 결과의 조합을 고려해 더 유망한 명령을 고르고 행동 전에 예측을 기록한다. 정답을 받은 뒤 확신도를 적으면 앞선 불확실성을 반영하는 지표가 아니게 된다.

추가 관찰은 현재 결정의 기대 개선이 비용을 넘을 때 구입한다. 프로그램은 가능한 두 센서 응답마다 믿음을 다시 계산하고 각각의 최선 기대 결과를 평균한다. 당장의 이익을 계산하는 방식이다. 다음 단계들의 미래 정보 가치까지 별도로 계획하지 않으므로 모든 시간 범위에서 최적인 정책이라고 부르지는 않는다.

비용은 정확도와 효용을 구별하게 한다. 정보를 더 얻는 것이 유리할 때도 있지만 비용과 행동을 바꾸는 데 도움이 되는지에 달려 있다. 주된 불확실성이 제어에 있다면 목표를 더 정밀하게 알아도 결정의 개선은 작을 수 있다. 세계의 상태와 자기 행동 모형의 상태를 함께 고려해야 한다.

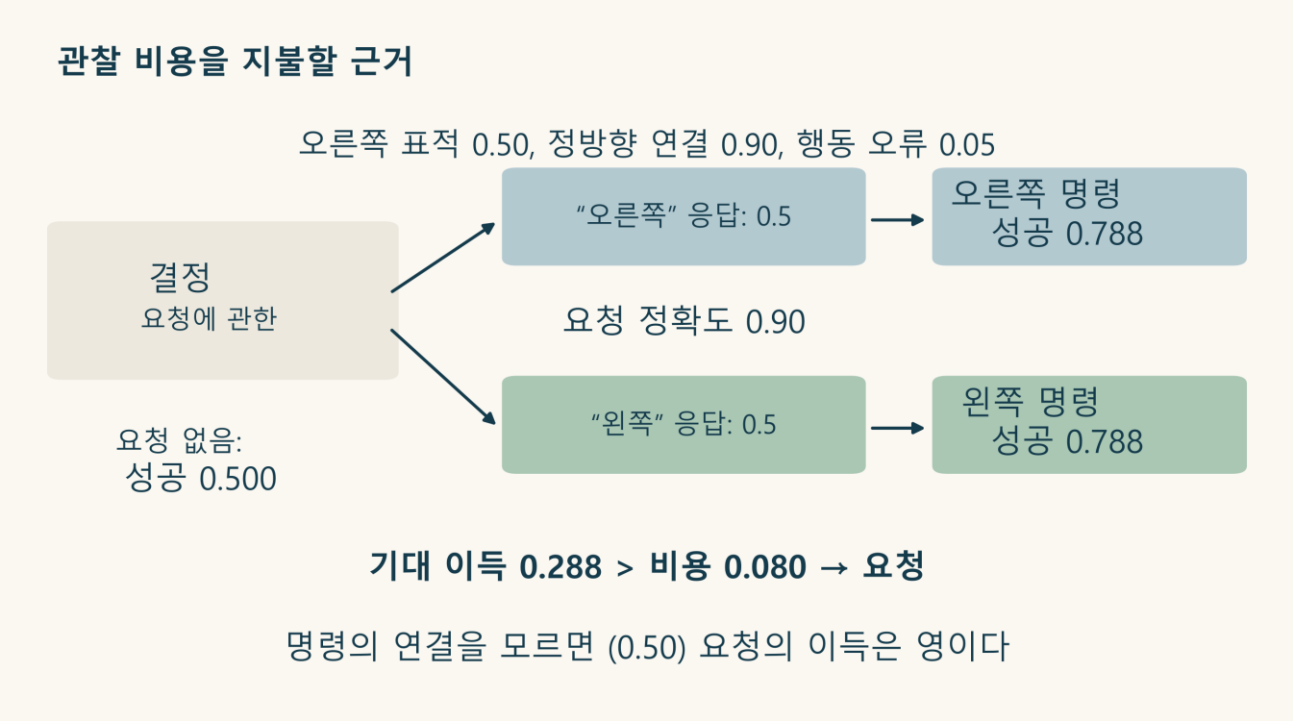


그림 63. 추가 관찰의 조건부 가치. 첫 관찰이 빠졌을 때 선택한 학습 상태에 agent_lab.py의 규칙을 적용한 계산이다. 실행 평균값이 아니다. 가능한 두 응답 모두 기대 성공 0,788을 주므로 증가량 0,288이 비용 0,08을 넘는다. 명령의 정방향 연결 확률이 0,5이면 표적을 안다는 사실만으로 행동 선택이 개선되지 않는다.

## 네 방식과 공정한 비교

전체 모드에서는 설명한 메커니즘을 모두 사용한다. no_world_memory에서는 단계 사이 목표 믿음의 이월만 끄고 행동 모형의 학습은 남긴다. fixed_action_model에서는 정방향 대응 확률을 처음의 0,90으로 고정하며 갱신하지 않는다. no_extra_query에서는 추가 관찰의 구입을 막는다. 모드 이름만으로 어떤 계산이 달라졌는지 알 수 없으므로 이렇게 정확한 기술이 중요하다.

각 모드는 해당 시나리오에서 같은 잠재 관찰과 운동 오류를 받는다. 난수를 독립적인 흐름으로 나누고 값을 미리 준비한다. 관찰을 구입해도 이후 무작위 사건을 밀어내지 않고 추가 값만 공개한다. 명령이 다르면 결과는 다를 수 있지만 같은 단계의 운동 교란은 공통이다.

첨부한 실행에서는 모드마다 천 단계인 시나리오 서른 개를 수행했다. 인간이나 동물의 자료가 아닌 실제 프로그램 결과다. 요청 비용은 0,08이고 성공은 일, 실패는 영을 준다. 요청 비용을 빼서 효용을 얻는다.

| 모드 | 성공률 | 단계당 효용 | Brier |
|---|---|---|---|
| 전체 | 87,94 | 0,8619 | 0,1054 |
| 세계 기억 없음 | 85,36 | 0,7767 | 0,1202 |
| 고정 행동 모형 | 49,97 | 0,4841 | 0,3444 |
| 추가 요청 없음 | 77,91 | 0,7791 | 0,1589 |

Brier는 성공 예측과 영 또는 일로 적은 실제 결과의 차이를 제곱한 평균이다. 작을수록 확률 예측이 정확하다. 보정뿐 아니라 예측의 정보성에도 의존한다. 이를 보정이나 자기의식, 체험의 질만을 측정하는 척도라고 부르면 오류다.

고정 모형의 전체 성공률은 약 절반이지만 두 기간의 큰 차이를 감춘다. 반전 전에는 87,80퍼센트, 뒤에는 12,13퍼센트다. 전체 모드의 해당 값은 88,06퍼센트와 87,83퍼센트다. 재학습 속도를 측정하려면 짧은 시간 구간이 필요하다. 뒤 절반 전체의 평균은 전환 과정을 평탄하게 만든다.

덜 눈에 띄는 비교도 있다. 기억 없는 버전은 요청 없는 버전보다 더 정확하게 행동하지만 훨씬 많은 정보를 구입한다. 그 결과 효용은 조금 더 낮다. 따라서 «이 행위자가 더 낫다»고 말하려면 기준을 밝혀야 한다. 응답 정확도, 비용, 예측 신뢰도는 서로 다르게 변할 수 있다.

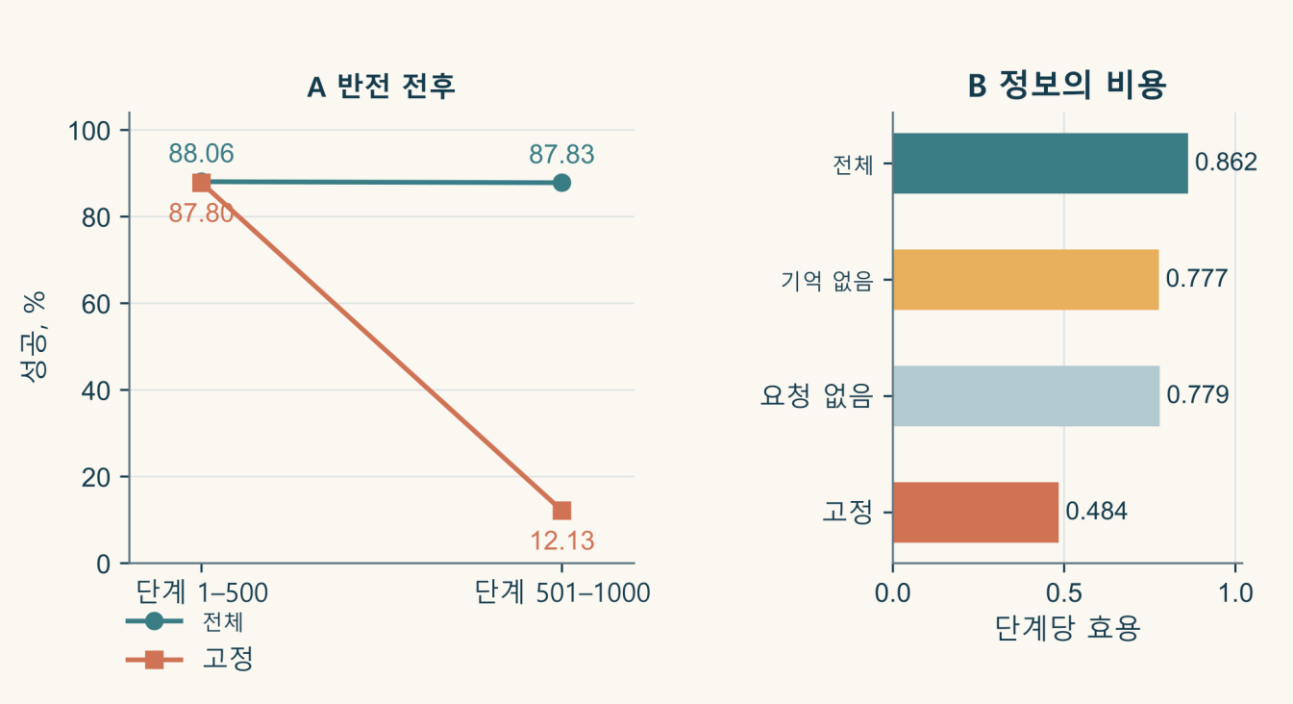


그림 64. 성공과 효용은 서로 다른 질문에 답한다. agent_lab.py의 결과다. 각 1000단계인 공통 무작위 시나리오 30개를 사용했으며 난수 시드는 1729다. 왼쪽은 두 기간의 평균이지 기간 안의 적응 속도가 아니다. 오른쪽은 성공에서 추가 요청 비용을 뺀 값이다. 자료: lab01_results.json.

## 행위자의 한 단계를 읽는 방법

먼저 보존한 믿음에 목표가 움직였을 가능성을 반영한다. 이용 가능한 잡음 신호를 받아 평가를 갱신한다. 목표와 제어의 불확실성으로 명령과 추가 요청의 효용을 계산한다. 성공 예측은 행동 전에 적는다. 명령 뒤에는 보이는 결과로 제어 모형을 갱신한다. 평가자는 실제 목표와 반전 일정, 현재 성공 여부를 행위자에게 공개하지 않는다. 이 이진 과제에서는 결과와 성공을 함께 주면 목표를 알게 되어 연구할 인터페이스가 달라진다.

### 여기까지 확인한 것

정보가 들어오는 순서를 알며 예측과 나중에 얻은 지식을 구별할 수 있다. 이제 모드의 수치는 요청 비용과 반전 전후의 기간을 함께 보며 읽어야 한다.

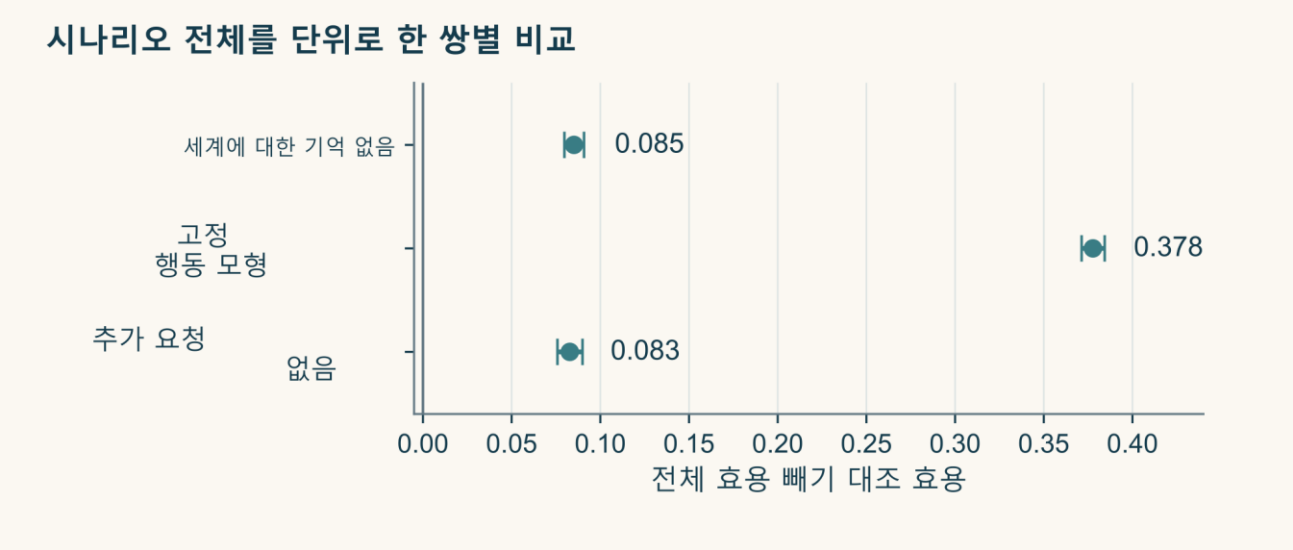


그림 65. 방식 사이 차이의 불확실성. 첨부 자료의 분석 결과다. 점은 효용의 평균 쌍별 차이이고, 선분은 전체 시나리오 30개를 4000번 재표집한 근사적 95퍼센트 백분위 구간이다. 반복 단위는 시나리오다. 구간은 개별 비교에 해당하며 95퍼센트의 동시 보장을 주지 않는다. 자료: paired_analysis.json.

## 실습 28 환경 변화의 결과 예측하기

실습 목표는 기본 실행을 재현하고 메커니즘의 구조로 결과 하나를 설명하는 것이다.

세계와 행동의 확률 모형, Brier score, 독립 반복의 단위를 활용한다. 실습은 Python 3.10 이상과 표준 라이브러리를 사용한다.

예상 소요 시간은 30–45분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 실습 파일의 압축을 풀고 python run_practice.py --self-test를 실행한다. 본 실행 전에 제어가 바뀐 뒤 고정 모형과 학습 모형의 예상 행동을 적는다.
2. python run_practice.py --lab 1을 실행한다. 생성한 폴더와 매개변수, 체크섬을 보존한다.
3. 결과 표에서 전환 전후의 성공과 단계당 효용, 예측 품질을 따로 비교한다. 지표를 밝히지 않고 가장 좋은 모드를 고르지 않는다.
4. 로그에서 이용 가능한 관찰, 이전 상태, 예측, 명령, 보이는 결과, 새 상태의 순서로 한 사례를 추적한다.

실행 기록과 메커니즘, 지표, 조건 범위를 밝힌 짧은 결론을 보존한다.

세계 기억이 없는 모드에서도 명령과 결과의 관계는 학습한다. 특정 기능을 끈 것이며 과거를 보존하는 능력을 모두 없앤 것은 아니다.

탐구를 이어 가려면 요청 비용과 센서 신뢰도, 목표의 변동성 가운데 하나만 바꾸고 예측을 미리 적는다. 이어 새 시나리오들에서 비교를 반복한다.

해설과 답은 부록 F 제28장 참조, 312쪽

**자기 점검 문항**

1. 기억 없는 모드에도 명령과 결과의 관계 학습이 남는 까닭은 무엇인가?
2. 기간을 나누지 않은 고정 행위자의 전체 성공률이 오해를 부르는 까닭은 무엇인가?
3. 확신도를 기록하기 전에 행위자에게 정답을 주면 무엇이 달라질까?

첫 행위자는 두 믿음의 공동 작동을 보여 주었다. 이제 표상의 조직과 반복 처리, 내용의 전달을 계 전체의 성공과 섞지 않고 따로 검증한다.

제29장

# 지각과 전역 접근

행위자의 성공만으로 표상이 어떤 관계를 보존하며 특정 처리가 왜 도움이 되는지 알 수는 없었다. 이 장의 세 실습은 차이의 기하와 잡음 채널의 혼합, 내용 전달로 과제를 나눈다. 실습마다 대조 조건과 결과가 있다. 의식에 관한 과학 이론으로 결론을 옮기기 전에 각각에서 확인한 것을 살핀다.

## 시스템이 정확히 무엇을 표상해야 하는가

앞 장의 행위자는 두 방향을 구별했다. 이제 표상의 조직에 관심을 둔다. 어떤 상태를 비슷하게 보고 반복 처리로 어떻게 바꾸며 어떤 모듈이 이를 이용할 수 있게 되는가? 서로 다른 성질에 관한 질문이다. 유용한 기하가 계획에서의 이용 가능성을 보장하지는 않는다. 정보가 널리 전달된다고 한 색조가 다른 색조와 비슷하게 느껴지는 이유가 설명되지도 않는다.

평면 위의 점 다섯 개에서 시작한다. 각 점은 대상을, 점 사이 거리는 선택한 차이 모형을 나타낸다. 수치는 설계자가 정한 값이며 인간 감각의 측정값이 아니다. 투명성이 장점이다. 변환 뒤 어떤 관계가 남고 무엇이 달라지는지 정확히 말할 수 있다. 이어 질적 조직의 더 복잡한 모형을 평가하기 쉬워진다.

둘째 실습을 실행한다. 추가 실습들은 mechanisms_lab.py에 들어 있으며 같은 폴더에 통합 원형이 사용하는 agent_lab.py가 있어야 한다. 결과는 일반 텍스트 편집기로 볼 수 있는 JSON으로 저장한다.

```
python3 mechanisms_lab.py --lab 2 --json geometry.json
```

프로그램은 원래 점들을 0,73라디안 회전시키고 같은 벡터만큼 평행이동한다. 이어 별도로 한 좌표를 두 배로 늘린다. 변환 전후 각 쌍의 거리를 비교한다. 회전 뒤 마지막 자릿수의 작은 차이는 컴퓨터의 실수 표현 때문이다. 기본 실행에서 최대 차이는 약 0,00000000000000044다.

## 불변량이 이름보다 중요한 이유

불변량은 명시한 변환에서 보존되는 성질이다. 모형이 쌍별 유클리드 거리만 이용한다면 전체 회전과 평행이동이 유사성 예측을 바꾸어서는 안 된다. 그때 좌표 이름은 완전히 달라질 수 있다. «이 좌표 자체가 빨강의 느낌이다»라는 주장은 왜 바로 그 기록 방식을 선택했는지 추가로 설명하지 않으면 이 검증을 견디지 못한다.

한 축을 늘리면 보통 거리 관계가 바뀌므로 그 거리에 근거한 모형에 중요한 변환이다. 그러나 결론은 선택한 지표에 달려 있다. 첫 좌표의 대상 순서만 사용하는 모형은 양의 배율로 늘린 뒤에도 일부 응답을 유지할 수 있다. 실험 전에 절대 거리와 이웃 순서, 방향, 거리의 비, 범주 구별 가능성 중 무엇이 중요한지 밝혀야 한다.

실습 기하로 완성된 퀄리아 지도를 얻는 것은 아니다. 이를 위해서는 구별이나 유사성 자료와 측정 척도의 결정, 새 조합에서의 검증이 필요하다. 자극 사이 거리와 보고한 체험 사이 관계도 구별해야 한다. 물리 매개변수를 같은 만큼 바꾸어도 주관적 변화가 같은 것은 아니다. 연구 질문을 보기 좋은 그림보다 먼저 정해야 한다.

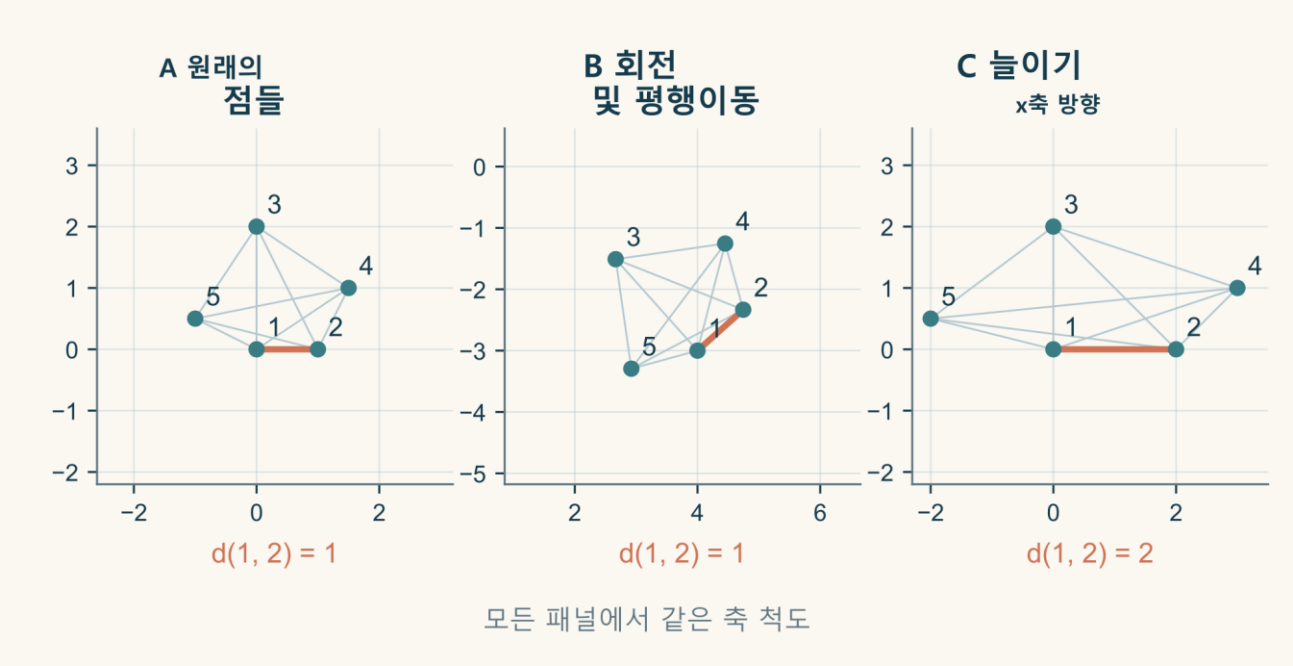


그림 66. 표상을 변환해도 유지되는 것. 둘째 실습의 학습용 기하다. 0,73라디안 회전과 (4; −3)만큼의 평행이동은 열 쌍의 거리를 모두 기계 오차 범위에서 보존한다. x 방향 늘이기는 거리 일부를 바꾼다. 한 쌍을 강조했지만 계산은 모든 쌍을 검증한다. 번호는 점을 뜻하며 체험의 종류가 아니다. 원래 좌표는 mechanisms_lab.py에 주어져 있다.

## 단순한 메커니즘의 반복 처리

세 번째 작업에서는 세 채널이 하나의 숨은 신호에 잡음이 섞인 값을 받는다. 참값은 마이너스 일 또는 일이며 각 관찰에 잡음이 더해진다. 한 버전은 첫 채널을 즉시 사용한다. 다른 버전은 채널 값들을 네 번 섞는다. 새 상태의 절반은 해당 채널의 이전 값을 유지하고, 나머지는 두 이웃 채널에서 각각 사분의 일씩 받는다. 결과를 얻기 위해 첫 채널을 읽는다.

반복 처리로 채널들은 점차 비슷해진다. 수식 없이도 이해할 수 있다. 한 채널이 우연히 지나치게 크고 나머지가 목표에 가까우면 교환으로 개별 편차를 줄인다. 그러나 모든 채널의 공통 오류는 고치지 못한다. 이미 있는 정보를 다시 배분할 뿐 반복마다 새 독립 센서가 생기지는 않는다.

python3 mechanisms_lab.py --lab 3 --json recurrence.json

중요한 대조는 차례로 펼친 네 층이다. 각 층은 재귀 회로의 한 반복과 같은 연산을 수행한다. 입력이 같으면 결과는 정확히 일치한다. 첫 경우에는 프로그램이 순환문에서 한 단계를 반복하고, 둘째 경우에는 네 개의 중첩 호출을 명시한다. 표현 형식을 바꾸는 것만으로 새로운 계산적 성질이 생기지는 않는다.

| 실습 시행 2000회의 방식 | 평균제곱오차 |
|---|---|
| 혼합 없는 단일 채널 | 1,892736 |
| 네 번의 재귀 단계 | 0,632419 |
| 펼친 네 층 | 0,632419 |
| 세 채널의 직접 평균 | 0,632465 |

수치는 seed 1729로 실행한 프로그램에서 얻은 해당 과제의 결과다. 직접 평균은 네 차례 혼합 없이 모든 입력을 사용하는 유용한 추가 대안이다. 계산 비용이 다르므로 같은 연산 횟수의 대조는 아니다. 결과가 거의 같다는 것은 선택한 과제를 훨씬 단순하게 처리할 방법이 있을 수 있음을 보여 준다.

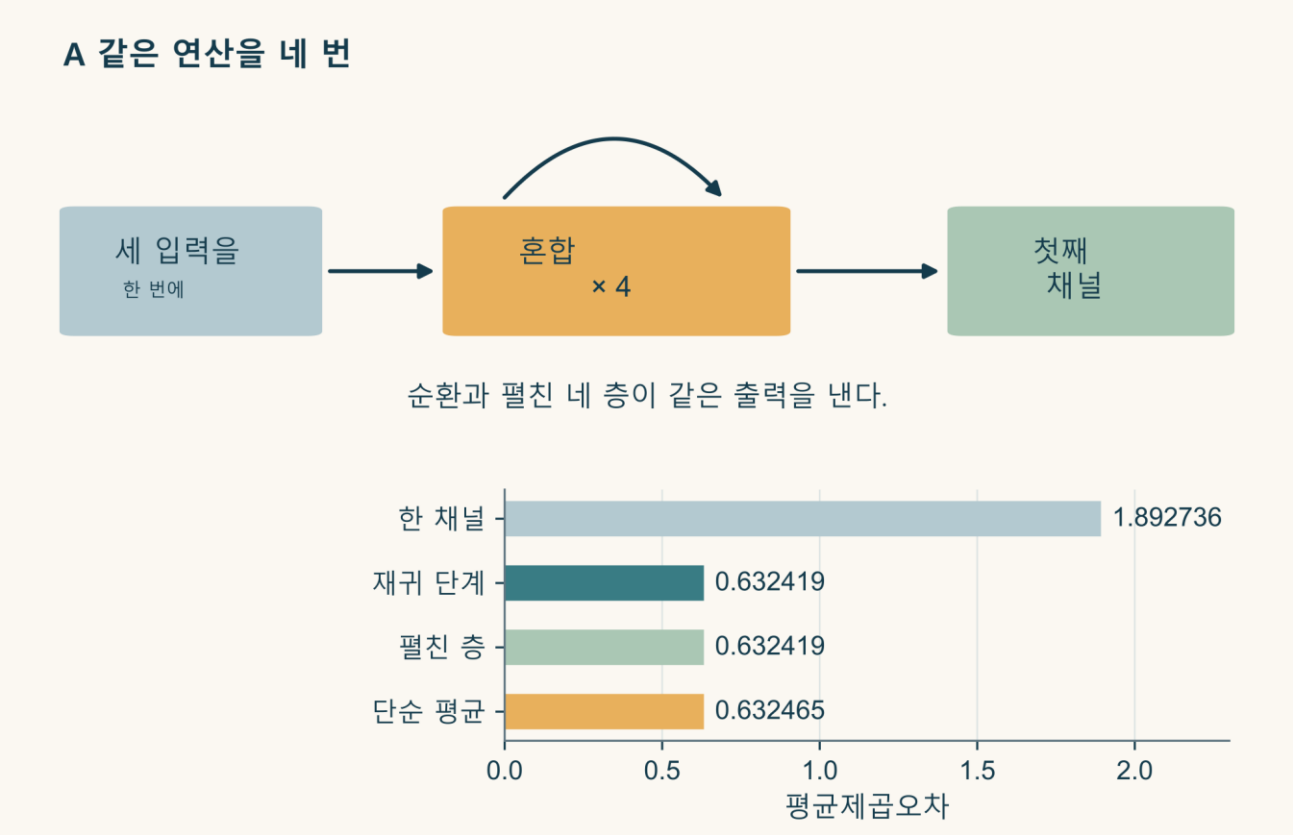


그림 67. 재귀성과 비교 가능한 대안들. 세 번째 작업의 설명용 도식과 결과다. 2000회 시행, seed 1729. 혼합할 때 채널은 자기 값의 절반을 유지하며 다른 두 채널에서 각각 사분의 일씩 받는다. 재귀 구현과 펼친 구현은 같은 연산을 수행하여 정확히 일치한다. 직접 평균은 계산 비용이 다른 추가 대안이다. 데이터: mechanisms_results.json, 작업 3.

## 재귀 처리에 더 강한 검증을 설계하는 방법

유한한 계산열은 종종 펼친 구조로 나타낼 수 있다. 그러므로 «코드에 순환문이 있는가»는 연구 질문으로 너무 거칠다. 매개변수 재사용과 처리 시간, 상태 유지, 변하는 입력과의 상호작용, 요소들의 물리적인 상호 인과성 가운데 순환의 어떤 성질이 중요한지 밝혀야 한다.

계산 기능의 가설이라면 유한한 구현들의 동일성이 중요한 제약이다. 물리 조직의 가설이라면 프로그램의 글만으로 장치에서 과정이 구현되는 방식을 모두 알 수는 없다. 이론 부분에서 도입한 구별이다. 여기서는 더 강한 결론 전에 적어도 계산 주장을 검증할 수 있다.

연구를 확장하려면 제시 시간이 다른 과제나 반복 사이에 새 관찰이 들어오는 과제를 미리 선택할 수 있다. 그때 경쟁 모형에도 같은 정보와 비교 가능한 자원을 주어야 한다. 그렇지

않으면 재귀 시스템의 이점은 관찰 시간이 더 길다는 것으로 설명될 수 있다. 조건을 맞춘 뒤 어떤 구조적 차이가 남았는지 정확히 이해할 때 의미 있는 실험이 된다.

## 내용 전달로서의 작업공간

넷째 실습에서는 색, 형태, 주의를 분산시키는 값이라는 세 메시지를 사용한다. 색과 형태는 영부터 이까지의 값을 갖는다. 정답은 색과 형태의 합을 삼으로 나눈 나머지다. 프로그램은 규칙을 미리 알고 있다. 필요한 정보의 전달을 연구하므로 분류기를 학습시킬 필요는 없다.

이 모형의 작업공간은 메시지 두 개를 담는다. 전체 버전은 색과 형태를 골라 응답 모듈에 전달한다. 전달이 없으면 응답 모듈은 빈 표상을 받아 영의 값을 쓴다. 무작위로 고르면 공간 일부에 방해 메시지가 들어간다. 추가 경쟁 모형은 모든 입력에 직접 접근한다.

```
python3 mechanisms_lab.py --lab 4 --json workspace.json
```

삼천 회의 시행에서 완전한 선택과 직접 접근은 백 퍼센트 정답을 준다. 전달이 없으면 성공률은 34,40퍼센트이고, 무작위 선택이면 55,83퍼센트다. 이 결과는 과제의 구성에서 예상할 수 있다. 관련 채널은 설계자가 지정했으며 응답 규칙도 주어져 있다. 프로그램이 무엇을 중요하게 여겨야 하는지 스스로 발견한 것은 아니다.

이 실험은 전달 채널의 작동을 확인한다. 직접 접근도 같은 결과를 내므로 이 작업공간이 유일한 해법은 아님을 보여 준다. 전문 모듈 사이의 제한된 교환이 주는 이점을 연구하려면 더 복잡한 과제 묶음과 다른 조율 방식의 실질적인 비교가 필요하다.

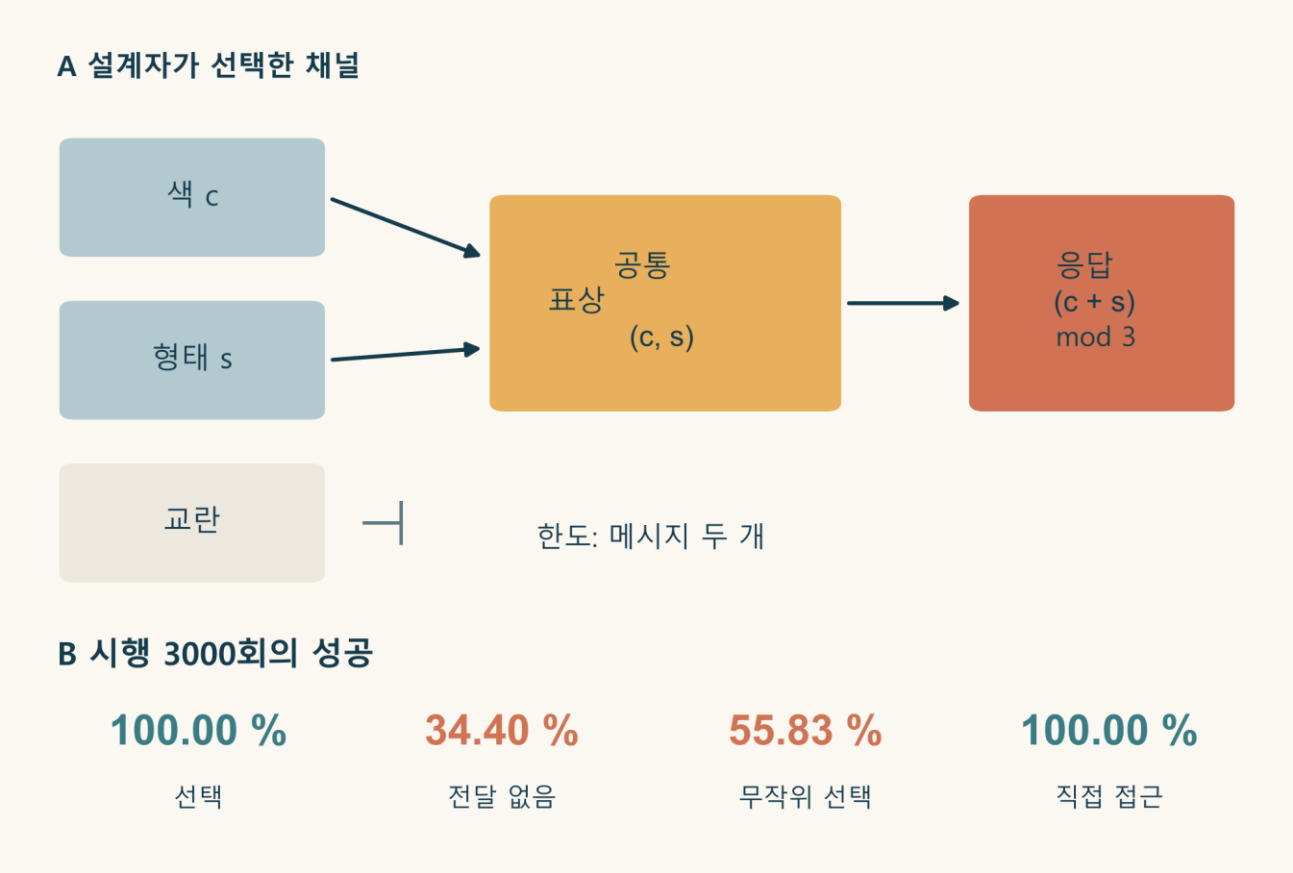


그림 68. 내용 전달과 직접 접근 대조. 넷째 실습의 학습 도식과 결과다. 색과 형태의 값은 0, 1, 2이며 응답 규칙은 미리 정해져 있다. 전체 버전은 필요한 메시지 두 개를 전달한다. 직접 접근하는 경쟁 모형도 과제를 똑같이 잘 해결한다. 여기서 검증하는 것은 정해진 규칙으로의 정보 전달이다. 자료: mechanisms_results.json, 실습 4.

## 전달에서 유연한 이용으로

실험의 후속 상황을 상상해 보라. 지각 뒤에 새 지시가 온다. 어떤 경우에는 색을 비교하고, 다른 경우에는 형태를 이용하며, 셋째 경우에는 다음 단계를 위해 대상을 기억해야 한다. 시스템은 서로 다른 이용 과정들에 쓸모 있는 내용을 보존해야 한다. 여기서는 고정 규칙에 수치 두 개를 전달하는 것만이 아니라 이용의 유연성을 묻게 된다.

새 지시는 경쟁 모형에도 주어야 한다. 작업공간만 과제 정보를 받는다면 이점의 이유는 자명하다. 유용한 설계는 기억 용량과 전달 메시지 수, 처리 시간, 가능한 지시 집합을 미리 정한다. 실패할 때마다 선택 규칙을 바꾸지 않고 새 조합으로의 일반화를 검증할 수 있다.

다음에는 국소 처리를 유지하면서 내용의 전파만 끈다. 이어 같은 비용이 들지만 무관한 채널의 차단과 비교한다. 이런 대조로 어떤 변화가 중요한지 알아볼 수 있다. 많은 정보를 없애 성능이 낮아졌다는 사실만으로 특별한 의식 조직에 관해 알 수 있는 것은 적다.

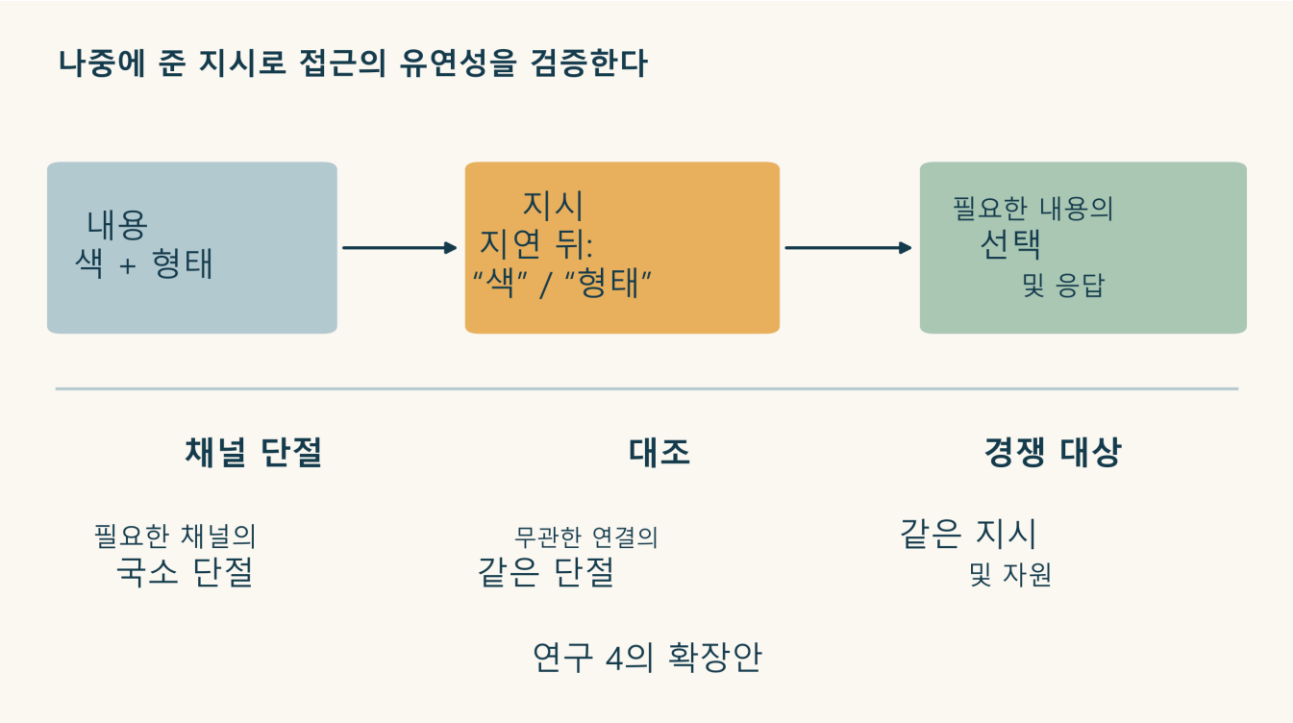


그림 69. 나중에 준 지시 뒤의 접근 검증. 제안한 실험에서 시스템은 내용을 제시한 뒤 필요한 과제를 알게 된다. 선택적 개입을 무관한 연결의 단절 및 같은 자원을 가진 경쟁 모형과 비교한다. 이 구성은 고정 과제에서 수치 두 개를 전달하는 것보다 더 많은 것을 검증한다.

## 같은 출력이 구조에 관한 결론을 제한하는 이유

네 번의 재귀 혼합 단계를 같은 규칙을 쓰며 순서대로 펼친 네 층으로 쓸 수 있다. 입력이 같으면 계산 결과는 일치한다. 따라서 잡음이 있는 단일 채널보다 낫다는 구체적인 결과를 소스 코드에 순환문이 있다는 사실만으로 설명할 수는 없다. 이용 가능한 입력 수, 시간, 매개변수 재사용, 변하는 신호, 실행의 물리적 조직 중 무엇을 중요하게 가정하는지 구체화해야 한다. 유용한 대조는 나머지 차이들을 보존한 채 선택한 차이를 검증한다.

### 여기까지 확인한 것

기하, 처리, 전달은 서로 다른 성질을 기술한다. 잘 들어맞는 수치 하나가 세 검증을 대신할 수는 없다. 이제 연구 하나를 골라 그 논거를 끝까지 검토해 보자.

## 실습 29 구별 가능한 두 실험 설계하기

실습 목표는 강한 대조 모형으로 구조에 관한 주장 하나를 검증하는 것이다.

제2장의 불변량과 되먹임, 비교 가능한 자원, 국소 상태와 전역 접근의 개념을 활용한다. 실습 2–4는 독립적으로 실행하며 모든 메커니즘을 한꺼번에 바꿀 필요는 없다.

실습 하나의 예상 소요 시간은 30–45분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 실습 2에서 불변성 지표를 미리 선택하라. 쌍별 거리의 전체 목록이다. python mechanisms_lab.py --lab 2 --json geometry.json을 실행하고 회전과 늘이기를 비교하라.
2. 실습 3에서는 단일 채널과 혼합, 펼친 층, 직접 평균의 차이를 먼저 예측한다. --lab 3을 따로 실행한다.
3. 실습 4에서는 응답 규칙에 실제로 도달하는 메시지를 적는다. --lab 4로 전체 선택과 전달 없음, 무작위 선택, 직접 접근을 비교한다.
4. 세 실험 중 하나를 골라 직접적인 결과와 강한 경쟁 모형, 다음에 구별할 질문을 정한다.

대조 조건을 검증한 실험 하나의 실행을 보존한다. 나머지는 따로 진행해도 된다.

거리 한 쌍이 같다고 모든 관계의 불변성이 확인되지는 않는다. 필요한 입력이 없는 모형을 이겼다고 구조의 유일성을 확립하지도 못한다.

탐구를 이어 가려면 혼합 실험에서 모든 채널에 공통 잡음 성분을 더하고 독립 성분도 따로 유지한다. 전달 실험에서는 경쟁 모형에도 같은 입력과 자원을 주는 늦은 지시 과제를 설계한다.

해설과 답은 부록 F 제29장 참조, 312쪽

**자기 점검 문항**

1. 표상을 회전한 뒤에도 거리가 보존되면 무엇을 알 수 있는가?
2. 재귀 버전과 펼친 버전의 일치가 재귀성에 관한 모든 이론의 문제를 해결하지 않는 이유는 무엇인가?
3. 넷째 실습에서 작업공간과 직접 접근의 성공률이 같다는 사실은 어떤 결론을 뒷받침하는가?

개별 기능의 대조 조건을 마련했다. 다음 장에서는 출처 이력과 내부 접근, 전략 수정을 더하고 함께 구현한 경우를 검증한다.

제30장

# 기억에 기초한 자기 모형과 성찰

이 장에는 서로 다른 여러 연구가 있다. 출처의 이력과 내부 접근, 전략 변화, 통합 행위자 순서로 진행한다. 한 비교의 실패가 다른 비교의 결과를 취소하지는 않는다. 개별 모듈의 성공이 전체 계에 자동으로 적용되지도 않는다. 단계마다 구조와 개입, 결과, 결론의 정확한 범위를 보존한다.

## 자신의 행동에서 출처의 이력으로

행위자가 명령과 결과의 관계를 배우면 환경에 미치는 자기 영향의 유용한 모형이 생긴다. 그래도 체험되는 행위주체감과 신체 소유감, 자신에 대한 명시적 추론은 다른 설명 대상이다. 이 장에서는 작은 메커니즘 몇 개를 만들고 각 결과가 어떤 주장을 허용하는지 알아본다. 실험은 모두 코드에서 수행하며 체험 가설로의 전환은 제26장의 요구에 따라 논의한다.

첫 메커니즘은 제25장을 이어 간다. 관찰 가능한 출처 A와 B가 있다. 사건마다 현재 특징이 신호의 공통 원인 가설을 지지하는지 안다. 실습에서는 모형이 출처 표지를 이용하고 오류 없이 인식한다. 신체가 실제로 자신의 것인지 알려 주는 숨은 표지는 아니다. 이 이상화로 수학을 먼저 검증하고 출처 인식의 오류를 따로 연구할 수 있다.

M0에는 믿음의 전달이 없으며 사건마다 확률 0,5에서 시작한다. M1은 흐름 전체에 믿음 하나를 보존한다. M2는 관찰되는 출처마다 믿음을 따로 보존한다. 공통 동일성 추적기 M1I는 외부 대상에 대해 같은 일을 한다. 최소 구현에서 M2와 M1I는 의도적으로 수학상 일치한다. 이는 강한 대조다. 기억의 명칭이 인위적으로 이점을 만들어서는 안 된다.

## 공통 기억과 분리된 기억의 실행 가능한 예

양성 단서의 확률은 공통 원인에서 0,75, 별개의 원인에서 0,25다. 갱신 전에는 이전 믿음이 0,5까지 남은 거리의 오분의 일만큼 가까워진다. M2에서는 해당 원천에 다음으로 접근할 때 이

과정이 일어나며, 세계의 매 단계마다 일어나는 것은 아니다. 다른 망각 모형은 다른 값을 주므로 규칙을 명시해야 한다.

```
python3 mechanisms_lab.py --lab 5 --json source.json
```

결과 파일의 source_history_example에는 A 지지, B 반증, A 지지, B 지지의 순서가 있다. 프로그램은 손으로 확인할 수 있는 값을 재현한다. 표는 현재 활성 출처의 갱신 평가를 보여 준다. 비활성 M2 기록은 따로 보존된다.

| 사건 | M0 | M1 | M2 | M1I |
|---|---|---|---|---|
| A 지지 | 0,7500 | 0,7500 | 0,7500 | 0,7500 |
| B 반증 | 0,2500 | 0,4375 | 0,2500 | 0,2500 |
| A 지지 | 0,7500 | 0,7105 | 0,8750 | 0,8750 |
| B 지지 | 0,7500 | 0,8581 | 0,5625 | 0,5625 |

셋째 사건에서 M2는 A의 지지 이력으로 돌아가지만 M1은 B의 최근 반증도 이월한다. 넷째 사건에는 반대 측면이 드러난다. M2가 B의 자체 반증 이력으로 돌아간다. 출처별 기억이 언제나 더 높은 확률을 주어야 하는 것은 아니다. 해당 출처의 이력에 따라 갱신한다는 것이 구별되는 성질이다.

이 값들은 정한 알고리즘의 귀결이다. 인간의 체험되는 신체 소유감이 어떻게 변하는지 아직 보여 주지 않는다. 이를 연결하려면 측정을 선택하고 관찰 일부로 매개변수를 추정한 뒤 다른 일부에서 예측을 검증해야 한다. M1I가 계속 같은 응답을 내면 특별한 자기관련성 메커니즘을 가려내지 못한 것이다. 추가 이론 차이를 찾거나 설명의 일반적인 성격을 인정해야 한다.

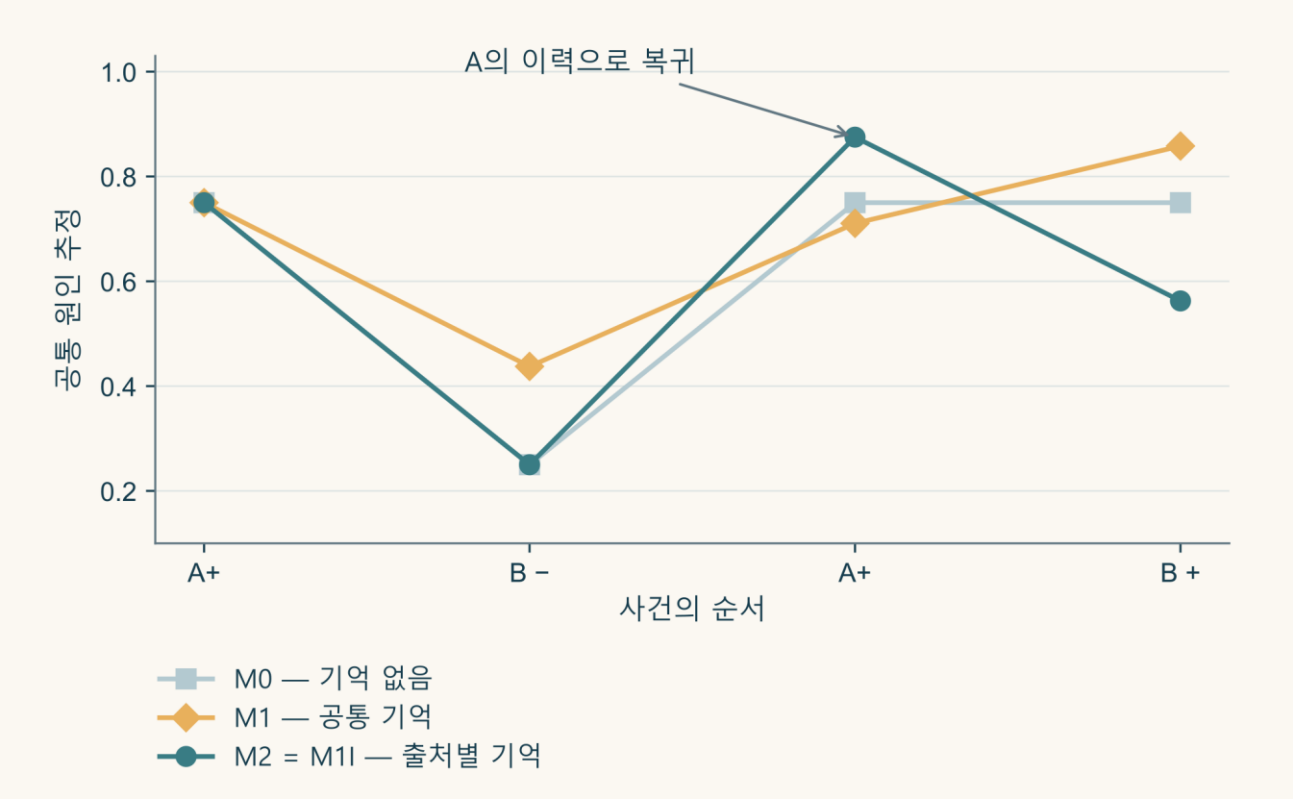


그림 70. 갱신에 사용하는 이력의 주인. 다섯째 실습의 결과다. 점은 사건 뒤 활성 출처의 평가를 나타낸다. 선은 연속된 사건을 연결하며 한 출처의 기록이 연속적으로 변한다는 뜻이 아니다. M2와 공통 추적기 M1I는 계산상 일치한다. 여기서 표지 A와 B는 관찰 가능하며 오류가 없다. 자료: mechanisms_results.json, source_history_example.

## 자신이 가한 영향의 출처를 찾는 방법

다섯째 실습의 둘째 부분은 다른 과제를 푼다. 행위자가 이진 명령을 보내고 두 결과 흐름을 본다. 하나는 명령과 연결되고 다른 하나는 무작위다. 연결된 흐름은 두 단계 지연될 수 있다. 모형은 제어하는 출처의 숨은 인덱스에 접근하지 않고 관찰 이력으로 채널을 찾아야 한다.

한 버전은 같은 시점의 값들만 비교한다. 다른 버전은 영부터 삼까지의 지연을 살펴 가장 잘 맞는 경우를 고른다. 일반 추적기도 같은 정보와 계산을 사용한다. 추가 대조에서는 명령의 이력을 무작위로 섞는다. 먼저 사건 백 개로 채널과 지연을 고르고 새 사건 예순 개로 일치를 검증한다. 설명의 선택과 이후 검증을 구별하는 것이다.

기본 반복 삼백 회에서 지연이 영이면 두 주요 방식 모두 채널을 정확히 찾는다. 지연이 이일 때 동시 비교는 49,67퍼센트를 맞히고 이력 방식과 일반 추적기는 각각 백 퍼센트를 맞힌다. 이력

모형이 고른 관계는 새 사건에서 약 87,7퍼센트의 일치를 보여 설정한 잡음과 부합한다. 섞은 이력의 결과는 우연 수준에 가깝다.

채널 선택에는 많은 관찰을 모은다. 그러므로 채널을 틀림없이 골라도 개별 미래 결과의 예측에는 오류가 있을 수 있다.

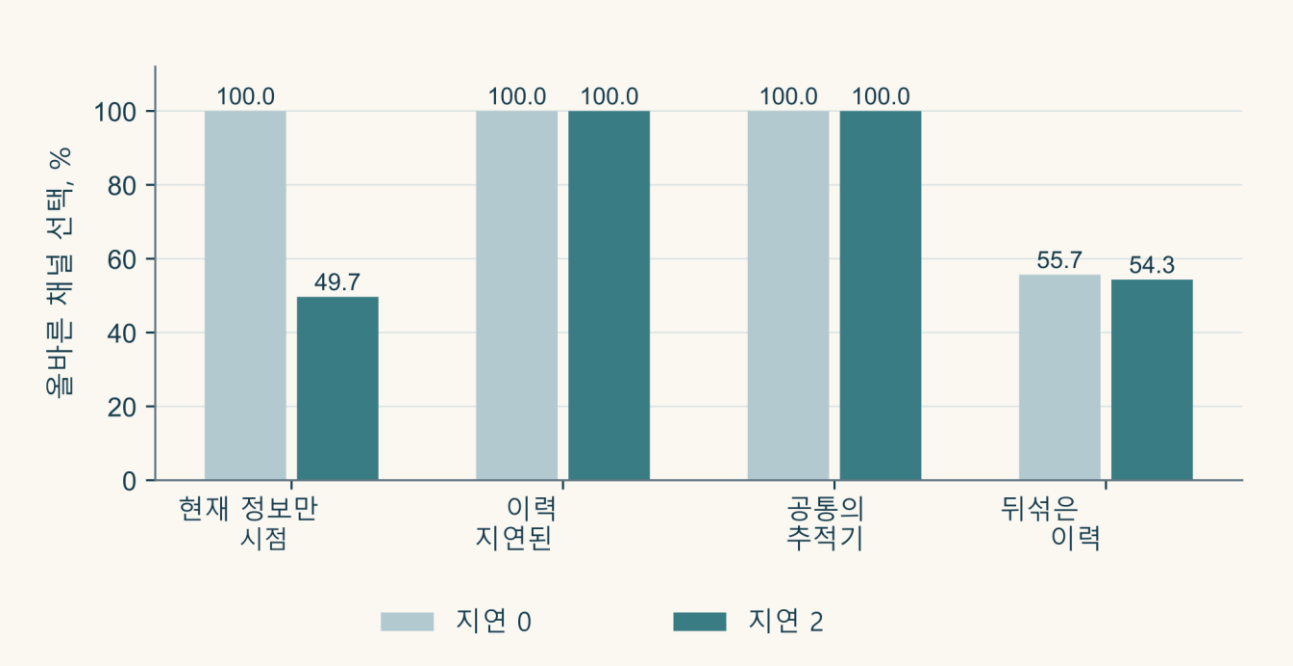


그림 71. 이력은 지연된 출처의 인식을 돕는다. 다섯째 실습의 결과이며 지연 0단계와 2단계에서 각각 300번 반복했다. 막대는 개별 효과의 예측 정답률이 아니라 여러 사건에 걸친 채널 선택의 정확도를 나타낸다. 이력 방식과 공통 추적기는 같은 알고리즘을 사용한다. 사건 100개로 채널과 지연을 선택하고, 새 사건 60개로 대응을 추가 검증한다. 자료: mechanisms_results.json, conditions_by_delay.

## 인과 검증을 동반한 내부 접근

여섯째 실습은 내부 점검의 최소 기준 모형을 만든다. 계에는 숨은 이진 특징이 있다. 공개 행동은 그 특징에 의존하지 않는다. 내부 센서는 특징을 직접 받아 값을 보고한다. 외부 관찰자는 공개 측면만 보고 기본 응답을 쓴다. 다른 버전은 내부 채널을 끈다.

대조 조건에서는 아무것도 바꾸지 않는다. 모의 개입에서는 무관한 비트를 바꾼다. 표적 개입에서는 평가하는 내부 상태 자체를 뒤집는다. 보고는 개입 뒤 상태를 기준으로 검증한다. 모든 조건에 비교 가능한 초기 사건을 주므로 보고가 필요한 원인을 따라가는지 물을 수 있다.

```
python3 mechanisms_lab.py --lab 6 --json internal_access.json
```

내부 센서는 각 조건의 이천 시행 모두에서 정답을 내고 외부 관찰자와 채널을 끈 버전은 약 절반을 맞힌다. 특별한 접근 권한을 명시적으로 설계했으므로 예상한 결과다. 더 복잡한 계와 비교할 기준을 얻었다. 언어 모형이나 스스로 학습한 행위자에서는 이런 채널의 존재 자체를 더 확인해야 한다.

과제가 복잡해져도 어떤 변수를 언제 이용할 수 있으며 어떤 대안 정보로 응답이 나올 수 있는지라는 검증 구조는 유지된다. 보고가 실험자의 예상과 같아도 여러 설명이 가능하다.

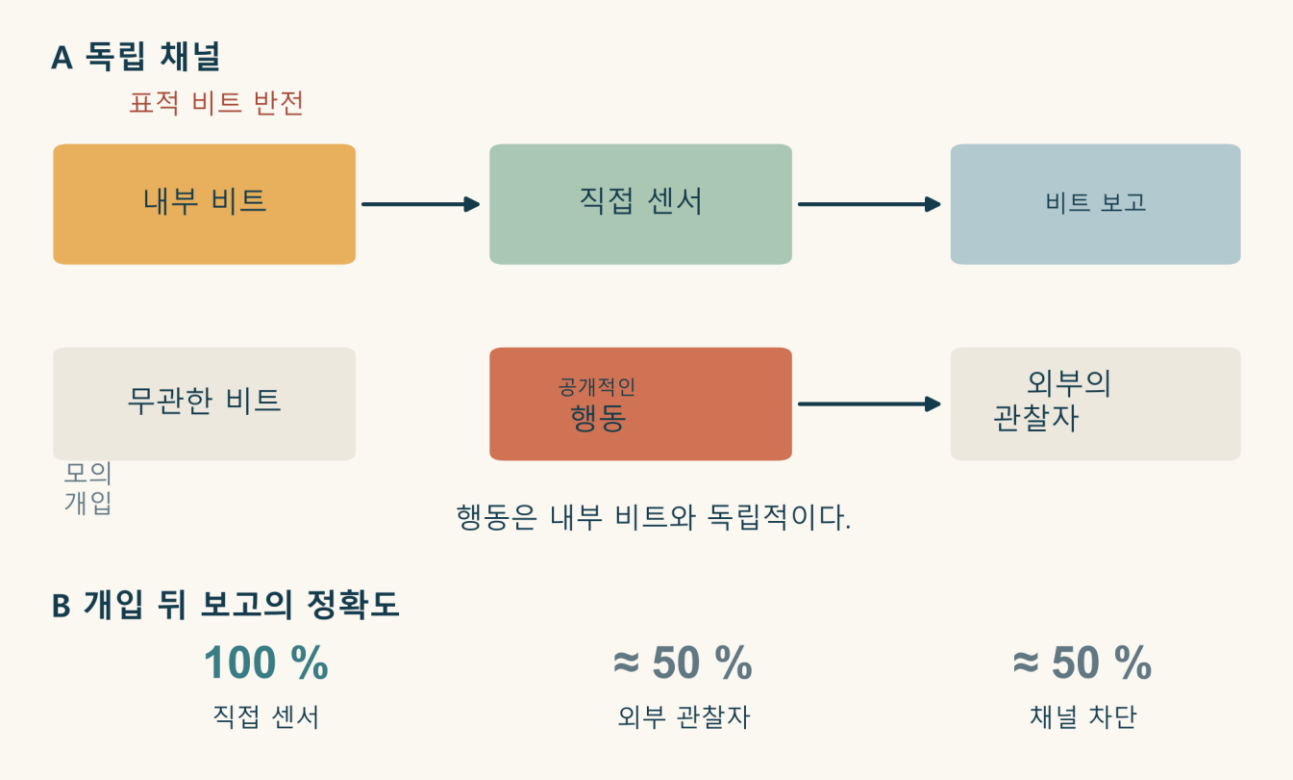


그림 72. 개입으로 검증하는 내부 접근. 여섯째 실습의 학습 도식과 결과다. 최초 조건, 모의 개입, 표적 비트 반전에서 각각 2000번 시행했다. 세 조건 모두 직접 센서는 100퍼센트이며 두 대조는 약 50퍼센트다. 평가는 개입 뒤의 상태에 해당한다. 접근 채널은 설계자가 정했다. 복잡한 인공지능에서 자기관찰을 발견했다는 그림이 아니다. 자료: mechanisms_results.json, 실습 6.

## 새 행동의 원인으로서의 재검토

일곱째 실습에서는 행위자가 미리 정한 두 분류 전략 중 하나를 고른다. 각 전략은 관찰 수치를 자체 문턱과 비교한다. 첫 전략은 환경 변경 전에 맞고 둘째는 변경 뒤에 맞는다. 행위자는 규칙 변경 표지를 받지 않는다. 현재 예측을 적은 뒤에야 정답을 공개하며 그 정보는 다음 단계들에 쓸 수 있다.

적응 버전은 최근 정답 표지가 있는 사례 서른 개를 보존한다. 새 응답 전마다 과거 자료에서 두 전략의 오류를 비교해 더 좋은 전략을 고른다. 고정 버전은 첫 전략을 유지한다.

narrative_only는 전략 수정에 관해 이야기할 수 있어도 행동은 고정되어 있는 대조다. 코드가 실제로 언어 이야기를 만들지는 않는다. 이름은 가정한 설명의 인과 영향을 검증하는 대조를 뜻한다.

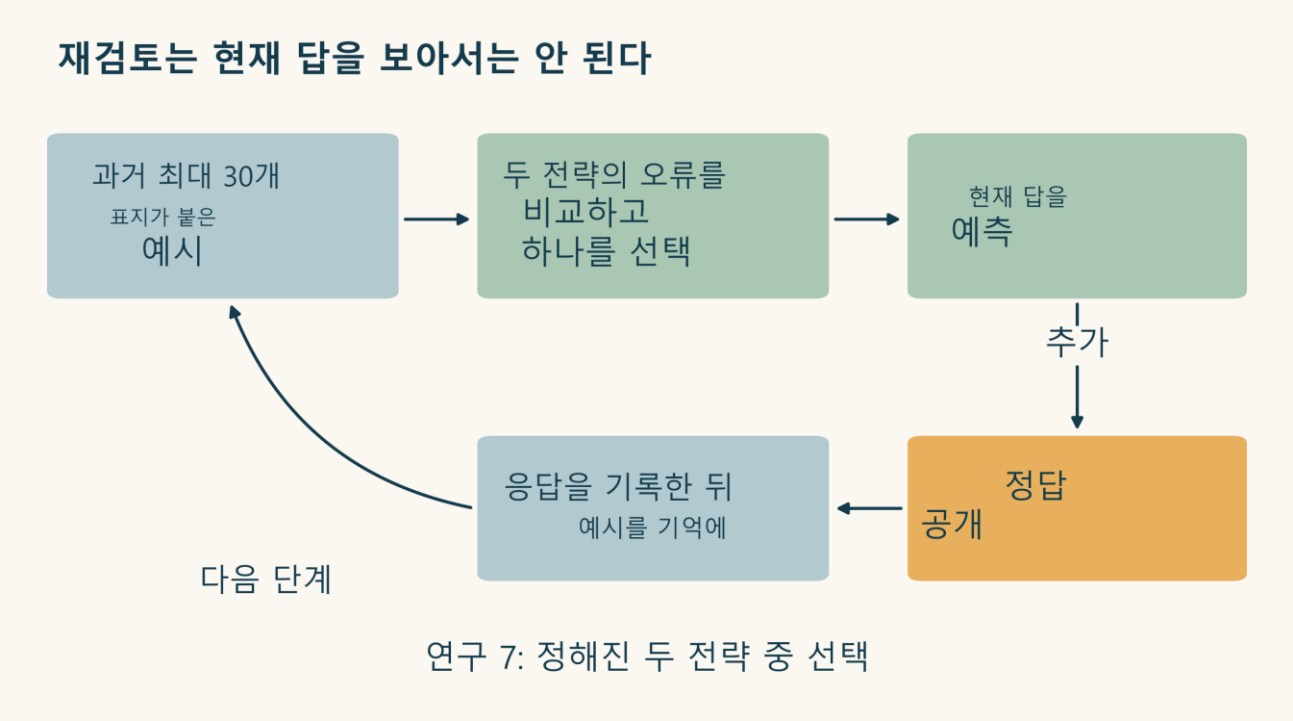


그림 73. 기능적 재검토의 인과 순서. 연산 순서는 revision_lab()과 같다. 현재 정답은 예측 뒤에만 이용할 수 있다. 기억에는 과거 예시 서른 개 이하가 들어간다. 적응은 미리 정한 두 규칙 사이에서 선택하며 이후의 결정에만 영향을 줄 수 있다.

python3 mechanisms_lab.py --lab 7 --json revision.json

기본 1200단계 실험에서 적응 버전은 전략을 한 번 바꾼다. 환경 변경 전에는 모든 모드가 정답을 낸다. 변경 뒤에는 적응 모드가 99퍼센트, 고정 모드와 행동을 바꾸지 않는 대조가 각각 69,67퍼센트에 성공한다. JSON에는 전환 부근 로그가 보존되어 현재 정답을 받기 전에 어떤 전략을 골랐는지 볼 수 있다.

기능적 수정의 작은 모형이다. 후보 규칙과 비교 방법은 직접 정했다. 더 발전한 성찰에는 대안 형성과 출발 가정의 평가, 결론의 일반화, 여러 수준의 자기 오류 점검이 필요하다. 그래도 이 초기 모형으로 수정이 후속 결정을 바꾸어야 한다는 핵심 인과관계를 검증할 수 있다.

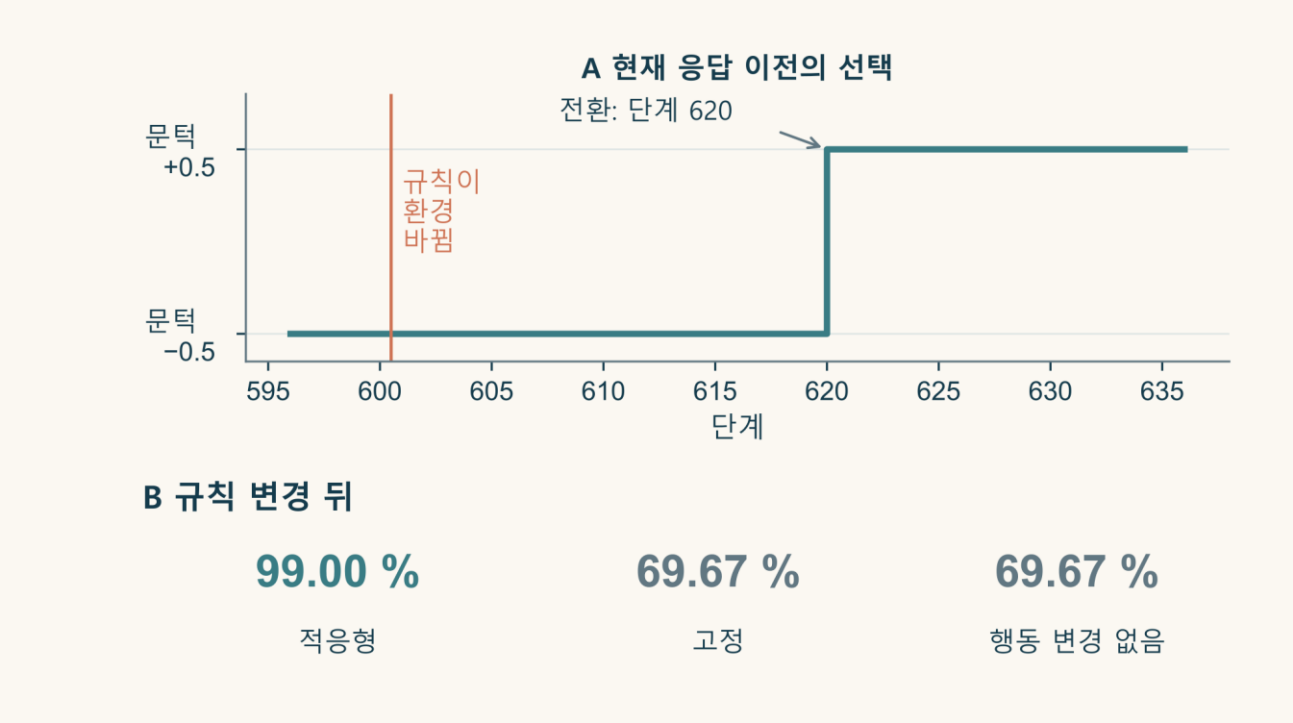


그림 74. 재검토가 새 결정의 원인이 된다. 일곱째 실습의 결과이며 1200단계, 난수 시드 1729를 사용했다. 보존한 로그 구간에서 전략은 과거 응답 오류 뒤 단계 620에 바뀐다. 현재 사례의 정답 표지는 예측 뒤에만 공개된다. 아래는 환경 변경 뒤 600단계의 성공률이다. 대조 narrative_only는 행동을 그대로 유지하며 프로그램은 텍스트 이야기를 생성하지 않는다. 자료: mechanisms_results.json, 실습 7.

## 복합 후보의 구성

여덟째 실습은 여러 메커니즘을 한 원형으로 묶는다. 세계의 확률적 기억과 자기 영향 모형, 두 믿음의 공통 표상, 선택 모듈, 예측 성공의 모니터, 학습 속도 변경 규칙을 사용한다. 환경은 첫 실습에서 가져온다. 여러 부분이 상호작용하는 작동 계를 얻는다.

공통 표상은 목표가 오른쪽일 확률과 정방향 제어 확률을 제공한다. 행동 모듈과 모니터는 같은 버전의 정보를 읽는다. 예상 밖 결과의 이력으로 제어 모형을 더 빨리 수정할 수 있다. 후보는 실제 목표나 환경 변경 일정을 받지 않는다. 연구자가 로그를 보존하고 지표를 따로 계산한다.

python3 mechanisms_lab.py --lab 8 --json candidate.json

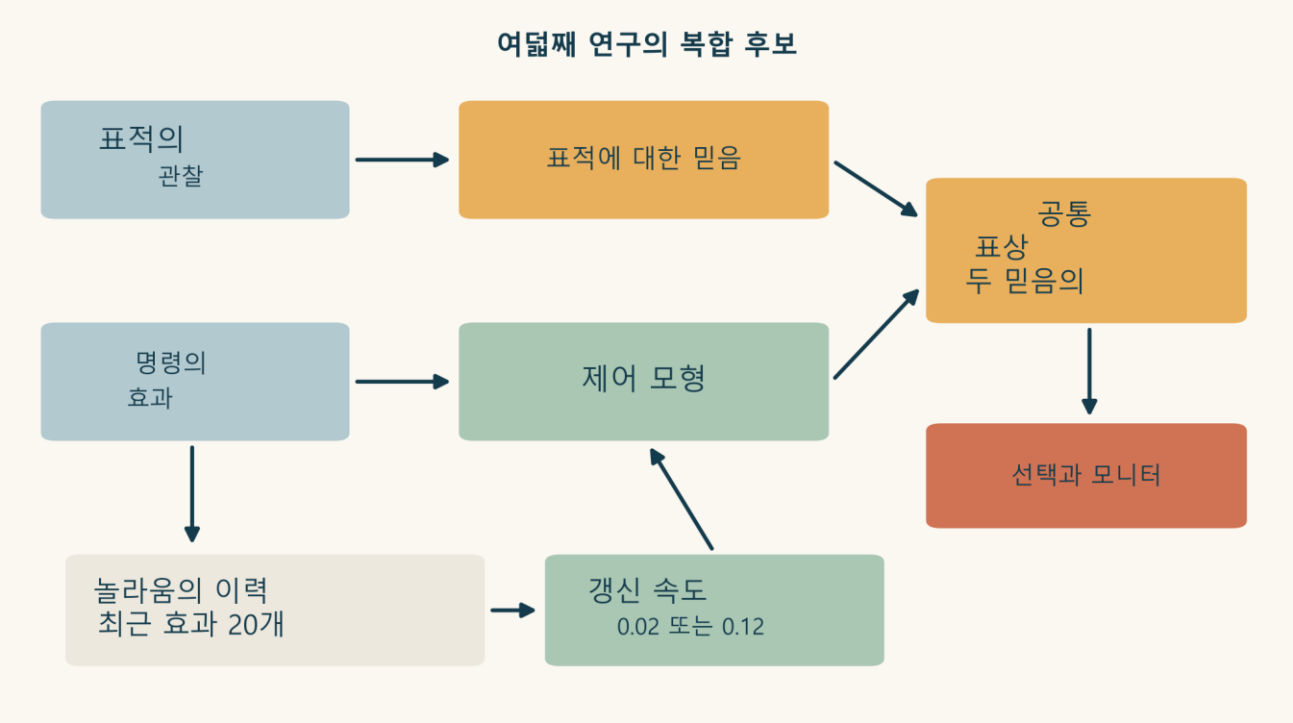


그림 75. 여러 메커니즘을 한 시스템으로 조립하기. 여덟째 실습의 학습 도식이다. 두 믿음이 행동 선택과 성공 예측에 제공된다. 최근 효과 20개의 평균 놀라움이 문턱 0,35를 넘으면 제어 모형의 재검토 매개변수를 0,02에서 0,12로 바꿀 수 있다. 이 가정적 규칙은 명시적으로 설계했다. 화살표는 정보 전달과 매개변수 변경을 보여 주며 체험의 존재를 확립하지 않는다.

천 단계 시나리오에서 전체 버전은 83,80퍼센트의 성공률과 단계당 0,81808의 효용을 낸다. 믿음을 직접 읽어도 결과는 같다. no_broadcast는 내용 전달과 추가 관찰을 동시에 끄며 성공률은 51,90퍼센트다. 그러므로 채널 하나의 기여를 분리하는 비교는 아니다. 이는 별도 시나리오이며 제28장의 서른 실행 평균이 아니다.

수정 속도를 고정한 넷째 모드가 특히 유용하다. 이번 실행에서는 전체 모드와 같다. 로그에 이유가 있다. 가속 학습을 켜는 문턱에 한 번도 이르지 않았다. 모듈이 프로그램에 있어도 이 실험에서 기여는 드러나지 않았다. 결과를 본 뒤 이점을 만들려고 매개변수를 바꾸지는 않았다.

다음 연구가 여기서 나온다. 메커니즘이 켜져야 할 조건을 미리 고르고 새 시나리오에서 검증한다. 제어 변경 빈도나 예상 밖 결과의 평가 규칙을 연구할 수 있다. 이미 본 자료로 매개변수를 고르는 단계는 탐색이라고 명시해야 한다. 검증 단계에서는 새 실행 전에 조건을 고정한다.

## 모듈의 존재와 기여의 차이

통합 후보의 가속 수정 규칙은 예상 밖 결과의 정도가 문턱에 이르러야 켜진다. 기본 시나리오에서는 한 번도 일어나지 않았다. 그러므로 전체 버전과 고정 속도 버전의 일치는 실제 전환 때 두 속도의 작동을 비교한 것이 아니다. 활성화 로그를 먼저 확인해야 한다. 이어 미리 고른 새 조건에서 규칙이 작동해야 하는 상황을 연구할 수 있다. 처음 결과에 맞추어 유리한 문턱을 고르면 탐색 단계이며 별도 검증이 필요하다.

### 여기까지 확인한 것

이제 기능마다 질문이 있다. 이력을 사용하는가? 내부 상태에 접근할 수 있는가? 수정이 다음 행동을 바꾸는가? 통합 행위자에서 그 구성요소가 실제로 필요한가?

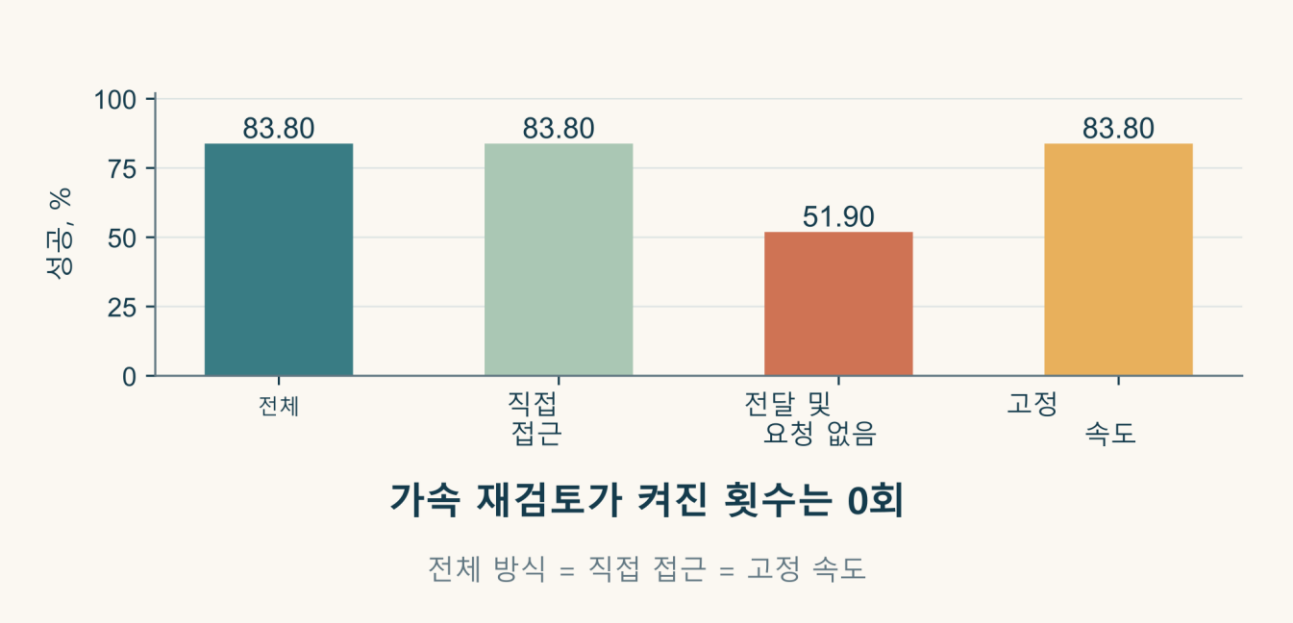


그림 76. 모듈이 있다고 기여가 드러나는 것은 아니다. 여덟째 실습의 결과다. 1000단계의 한 시나리오에 난수 시드 1729를 사용했다. 가속 재검토 조건이 한 번도 충족되지 않아 세 방식은 보존한 모든 지표에서 일치했다. 전달 없는 방식은 추가 요청도 차단하므로 복합 개입이다. 개별 기여를 검증하려면 새로 미리 정한 실험이 필요하다. 자료: mechanisms_results.json, 실습 8.

## 실습 30 자기 계에 관한 결론 논증하기

실습 목표는 성공한 모든 시연을 합산하는 대신 선택한 실습 하나에서 인과 논거를 마련하는 것이다.

제25장의 출처별 기억과 접근의 인과 검증, 사전 예측과 되먹임을 활용한다. 실습 5–8은 다른 날에 진행해도 된다.

실습 하나의 예상 소요 시간은 35–50분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 실습 5와 6, 7, 8 중 하나를 골라 mechanisms_lab.py의 해당 --lab로 따로 실행한다.
2. 결과를 보기 전에 이용 가능한 입력과 바꾸는 기능, 주된 지표를 밝힌다.
3. 실습 5에서는 일반 추적기와 비교한다. 6에서는 표적 개입과 모의 개입을 검증한다. 7에서는 되먹임 전에 예측을 보존한다. 8에서는 수정 규칙이 켜진 횟수를 확인한다.
4. 무엇을 구현했고 무엇을 인과적으로 검증했으며 무엇을 구별하지 못했고 다음에 어떤 실험이 필요한지 네 문장으로 적는다.

로그와 대조 모형, 현상적 체험 가설에 관한 별도 결론이 있는 검증 가능한 보고서 하나를 남겨 둔다.

전달과 추가 요청을 함께 끈 비교로는 전달만의 순수한 기여를 가려낼 수 없다.

탐구를 이어 가려면 실행 전에 선택한 구성요소의 새 활성화 조건을 정한다. 이전의 구별되지 않은 결과도 보존하고 새 시나리오에서 강한 경쟁 모형과 비교한다.

해설과 답은 부록 F 제30장 참조, 313쪽

**자기 점검 문항**

1. M2와 M1I의 동일성이 의미 있는 결과인 까닭은 무엇인가?
2. 설계한 내부 센서의 백 퍼센트 정답률은 무엇을 확립하는가?
3. 여덟째 실습에서 가속 수정의 이점이 확인되었다고 볼 수 있는가?

디지털 모형에서는 개발자가 모든 상태에 접근한다. 살아 있는 조직으로 옮기면 이 능력 일부는 사라지고 외부 장비의 역할은 커진다. 다음 장에서는 실제 자료를 다루기 위한 준비를 따로 한다.

제31장

# 생물학적 구현과 혼성 구현

생물학적 구현으로 옮기면 이용 가능한 측정과 계의 경계가 달라진다. 학습 일부는 외부 제어기에서 일어날 수 있으며 많은 주기가 같은 시료에 속할 수 있다. 그러므로 실제 자료를 읽고 기록 단위를 확인하는 법부터 익힌다. 아래에서는 공개된 필드로 따라갈 실습 경로를 제공한다. 준비된 예를 재현한 뒤 상호작용 지표를 직접 찾는 확장 과제로 나아간다.

## 살아 있는 조직이 더해지면 달라지는 것

프로그램 모형에서는 상태와 전이 규칙을 정확히 정할 수 있다. 생물학적 계는 다른 연구 재료를 제공한다. 세포가 발달하고 상호작용하며 활동 양상을 바꾼다. 이를 계산에 쓸 수 있다고 모든 살아 있는 조직이 체험을 가진다는 뜻은 아니다. 그래도 생물학적 연구를 미리 정한 알고리즘의 실행으로만 환원하기는 보통 어렵다. 조직 자체의 기능과 외부 장비의 기여를 밝혀야 한다.

신경세포 배양, 오가노이드, 온전한 생물 개체를 구별하자. 우리 분석에서 중요한 것은 «뇌»라는 말이 주는 인상이 아니라 연구 대상의 구조다. 생물학적 요소가 입력의 디지털 변환, 활동 판독, 과제와 인과적으로 연결된 시스템을 혼성계라 하자. 가정한 체험 기반의 경계가 배양 접시나 컴퓨터 과정의 경계와 일치하지 않을 수 있다. 그 경계의 근거를 제시해야 한다.



그림 77. 생물학적 구성요소와 측정 장비. 가정한 연구 작업 공간을 그린 것이다. 접시 속 세포 모형, 장비, 컴퓨터를 시스템의 서로 다른 부분으로 나타냈다. 실제 연결과 과제 수행에 대한 기여는 구체적인 설계에서 기술해야 한다. 이 그림은 논의한 어느 실험의 장치도 재현하지 않는다.

## 세 연구 사례

DishBrain에서는 인간과 생쥐에서 유래한 신경세포 배양체를 전기 기록과 자극을 통해 단순화한 Pong 게임에 연결했다. 저자들은 특정한 폐회로 되먹임에서 게임 지표가 개선되었다고 보고했다. 이는 배양체이며 오가노이드가 아니다. 논문 제목의 sentience는 저자들의 조작적 해석과 연결된다. 측정 결과가 주관적 체험을 확립하지는 않는다. [44]

Brainoware에서는 오가노이드를 동적 계산 리저버로 사용했다. 연구는 음성 신호 인식과 비선형 동역학 예측을 포함한다. 이런 장치를 이해하려면 입력 변환, 생물학적 망의 응답, 결과 판독 절차를 따로 살펴야 한다. 시스템 전체의 성공만으로 어떤 요소가 특정 개선을 만들었는지 드러나지는 않는다. [45]

Robbins와 동료들은 피질 오가노이드를 이용한 cartpole 제어를 연구했다. 대부분의 시료에서 강화학습 알고리즘이 선택한 학습 신호는 무작위 신호나 학습 없음보다 이점을 주었다. 개선은

사십오 분의 휴식 뒤 유지되지 않았다. 이 결과는 훈련 중의 적응과 휴식 뒤 효과의 보존을 구별한다. 오가노이드에 어떤 형태의 기억도 없다는 뜻은 아니다. [46]

## 수준의 연쇄를 따라가라

생물학자가 특정 가소성을 강화하려고 분자 과정을 바꾸자고 제안한다고 하자. 체험 가설과 연결하려면 여러 단계를 거쳐야 한다. 분자 과정이 변했음을 먼저 보여야 한다. 이어 세포 성질과의 관계, 망 상호작용의 변화, 과제 결과를 확인한다. 마지막으로 그 조직과 체험의 관계를 따로 뒷받침해야 한다.

화살표마다 여러 설명이 가능하다. 망 활동의 변화는 세포 구성과 성숙, 전반적 흥분성, 측정 조건, 연결의 재구성 때문일 수 있다. 분자 개입에서 곧바로 «의식의 강화»로 넘어가면 중간 설명들을 검증하지 못한 채 남긴다. 중요한 전이마다 측정이나 비교를 연구 계획에 명시해야 한다.

이 책은 주체를 만들 보장된 방법으로 분자 표적을 선택하지 않는다. 어떤 변화가 구체적인 가설을 검증하는지 익힌다. 상호 동역학이 중요하다는 가설이라면 계의 다른 능력을 유지하면서 특정 상호작용을 바꾸는 것이 관심 대상이다. 활동 전체를 단순히 늘리는 것은 가설에 대한 특이적인 확인이 아니다.

## 같은 개선을 일으키는 두 메커니즘

학습 예를 만들자. 디지털 모형 A는 여러 요소 사이의 연결을 바꾸는 법을 학습하므로 같은 입력이 점점 더 정확하게 필요한 답으로 이어진다. 모형 B는 연결을 유지하지만 공통 증폭 계수를 바꾼다. 단순한 순서에서는 두 모형 모두 과제 지표를 개선할 수 있다. 이는 분석을 검증하기 위한 가능한 구성이다. 앞서 언급한 생물학적 연구 결과를 기술한 것이 아니다.

최종 성공만 측정하면 두 메커니즘을 구별하지 못한다. 새 입력 조합을 더하자. 특정 연결의 변화는 조합 사이의 선택적 전이를 예측할 수 있고 공통 증폭은 다른 변화 양상을 낼 수 있다. 차이의 방향은 선택한 모형에서 도출해야 한다. 규칙이 없다면 «선택성»이나 «가소성»이라는 말만으로 예측이 생기지는 않는다.

이어 공개 자료에서 비교에 필요한 지표를 찾을 수 있다. 전체 결과만 있고 필요한 입력이나 대조 조건이 빠져 자료가 부족할 수 있다. 이것도 준비 단계의 의미 있는 결론이다. 다음 생물학적 연구에서 어떤 정보를 제공해야 하는지 보여 준다.

## 공개 자료를 위한 준비된 경로

독립적인 작업에는 Robbins 저장소 v2의 특정 파일 aggregated_experiment_data.pkl을 사용한다. 파일의 MD5 체크섬 af99f1f859136ca4a6dad47a1aa7c86d는 발표된 값과 일치한다. 저장소 설명에는 열아홉 개의 오가노이드와 사백삼십이 회의 주기가 나와 있다. 확인한 파일에는 최상위 실험 기록 38건, 실행 기록 426건, 서로 다른 metadata.organoid 값 16개가 있다. 이 단위들은 따로 유지해야 한다. 해당 파일의 필드 값 수가 논문 전체의 표본을 새로 정하는 것은 아니다. [47]

자료 집합의 경계도 방법의 일부다. 필드 chip은 원래 명칭을 보존하고 organoid는 정리한 집단화에 쓰며 batch는 배치를 뜻한다. 실험 기록 하나에 여러 실행이, 오가노이드 하나에 여러 실험이 있을 수 있다. 따라서 모든 행의 순서를 임의로 바꾸어도 독립 시료들이 생기지는 않는다.

첨부 표의 특징들은 명시적인 한 절차로 얻었다. 발화 빈도는 firing_rates_hz의 유한한 값들의 평균이다. 시간 관계의 정도는 sttc_first_order의 상삼각 부분에서 대각선과 비유한 값을 제외한 원소들의 절댓값 평균으로 계산한다. 절댓값을 사용하면 양의 관계와 음의 관계를 구별하지 못하므로 다른 가설에는 다른 지표가 필요하다. 추가 논거 없이 STTC를 인과 연결과 동일시해서는 안 된다.

실습 비교의 목표값은 한 실험 기록 안에서 이용 가능한 rl 실행들의 top_decile 중앙값이다. 문서에서는 top_decile을 상위 십 퍼센트 시행의 평균 보상으로 정의한다. 이를 전체 정확도나 전반적인 학습, 휴식 뒤 효과 보존으로 대신하지 않는다. 다른 결과를 다루려면 과제를 새로 기술해야 한다.

38행 중 29행에는 두 특징과 목표 지표, organoid 식별자가 모두 있어 분석할 수 있다. 서로 다른 organoid 값 15개와 배치 7개에 속한다. 나머지 행도 제외 이유와 함께 보존한다. 결측 측정값을 영으로 바꾸지 않으며 원하는 결과의 개선 여부로 제외하지 않는다.

기본 모형은 일에 평균 발화 빈도를 더한 값의 로그를 사용한다. 확장 모형은 선택한 STTC 지표를 추가로 사용한다. 변환과 척도 조정은 학습 부분만으로 추정한다. 외부 검증에서는 오가노이드 하나 전체를 차례로 제외해 평가하며, 별도 방식에서는 배치 전체를 제외한다. 결과는 이 절차들과 제한된 자료 집합에 관한 것이다.

이 분석은 표의 특성과 정한 지표의 재현 가능한 관계를 보여 준다. 전향적 예측에는 측정의 시간 순서를 독립적으로 확인해야 하고 인과 결론에는 개입 설계가 필요하다. 특징을 더해도 오차가 줄어야 하는 것은 아니다. 긍정적 결과와 부정적 결과를 함께 남기고 결과를 본 뒤 유리한 집단화를 고르지 않는다.

## 자료 한 행과 명확히 정한 지표 하나

bio_features.csv에서 한 행은 실험 기록이며 뉴런 하나나 오가노이드 하나가 아니다. mean_firing_rate_hz는 원래 발화 빈도의 유한한 값들을 평균한다. mean_absolute_sttc15는 sttc_first_order 행렬의 상삼각 부분에서 대각선을 제외한 유한한 원소들의 절댓값 평균이다. 선택한 시간 구간의 활동 간 시간 관계가 두드러지는 정도를 나타낸다. 입증된 물리적 연결이나 체험의 척도는 아니다.

행 20217_exp1에서는 약 5,35918 Hz, 이 지표의 값 0,0172533, rl 실행들의 top_decile 중앙값 108,95를 얻는다. 문서에서 top_decile은 개별 실행의 상위 십 퍼센트 시행에 대한 평균 보상이다. 실행 사이에서 우리가 추가로 집계하는 방식은 명시적으로 정한다. 이는 연관성을 다루는 학습 과제에 쓰이며 학습의 인과 효과를 평가하는 것이 아니다.

### 여기까지 확인한 것

행의 출처와 특징의 뜻, 결과의 범위를 알게 되었다. 이제 관찰한 연관을 조직의 인과 설명으로 바꾸지 않으면서 해당 절차에서 특징이 통계 정보를 더하는지 검증할 수 있다.

## 재현 가능한 분석의 결과

첨부 계산에는 유효한 실험 기록 29개를 사용했다. 상수항이 있는 선형 모형은 log(1 + 빈도)를 쓰는 모형과, 같은 특징에 평균 절댓값 STTC를 더한 모형을 비교한다. 특징의 표준화는 학습 부분에서만 수행하며 릿지 정규화 계수는 0,000001이다. 주 결과는 남겨 둔 집단마다 같은 가중치를 주는 평균제곱오차다. 값이 작을수록 정해진 절차에서 예측이 더 정확하다.

| 검증용으로 남긴 집단 | 집단 수 | 발화 빈도 | 발화 빈도와 STTC |
|---|---|---|---|
| 오가노이드 | 15 | 1460,17 | 1581,03 |
| 배치 | 7 | 1658,10 | 1700,39 |

이 STTC 지표를 더했을 때 두 분할 모두 결과가 나빠졌다. 특정 특징과 분석 방식의 한계를 보여 줄 뿐 조직 내 상호작용의 역할을 반박하지는 않는다. 지표는 의식 측정이 아니다.
원기록의 시점을 추가로 검증하기 전에는 이 계산을 전향 예측이나 인과 검사가 아닌 연관성의 탐색 분석으로 읽어야 한다. 집단별 전체 오차와 제외 기록은 chapter31_result.json에 보존했다.

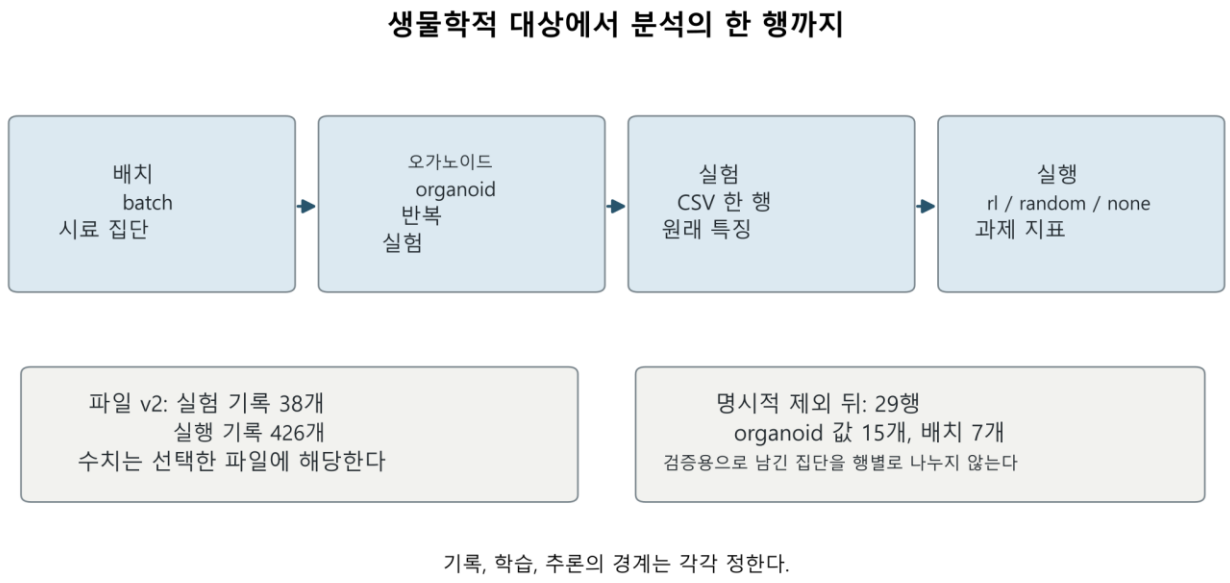


그림 78. 생물학적 자료에서 구분하는 기록 수준. 확인한 파일 aggregated_experiment_data.pkl v2 [47]의 기록 수준이다. 계수는 파일 내용을 기술하며 발표된 연구의 표본 크기를 대신하지 않는다. 학습 표에서 행의 단위는 실험이다. organoid와 batch에 따른 분할은 별도로 검증한다.

## 실습 31 공개 자료 사용하기

실습 목표는 시료별로 자료를 나누며 특징 준비와 두 모형의 비교를 재현하는 것이다.

관찰의 집단화와 상관 및 인과 개입의 구별, 기본 예측과 확장 예측을 활용한다. 기본 분석에는 첨부 표로 충분하다. 원자료를 읽으려면 추가 라이브러리가 필요하며 그 방법은 따로 설명한다.

예상 소요 시간은 45–60분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. bio_provenance.json과 bio_features.csv를 연다. 실험 기록 38개와 실행 426개, organoid 값 16개를 다른 총수가 명시된 자료 집합 설명과 대조한다.
2. 예제로 분석한 행에서 두 특징의 식을 검산하라. 원본 파일에서는 결측값과 무한값을 영으로 바꾸지 않는다.
3. 정한 분석에는 유효 기록 29개와 organoid 값 15개, 배치 7개를 쓴다. 다른 행의 제외 이유는 표에 보존되어 있다.
4. python teaching_lab.py --chapter 31을 실행한다. 오가노이드 하나 전체를 검증용으로 남기는 방식과 배치 전체를 남기는 방식에서 발화 빈도만 쓰는 모형과 STTC를 더한 모형을 비교한다. 결과가 나빠져도 보존한다.

행과 집단화, 제외, 통계적 결론의 한계를 설명한 재현 분석을 남겨 둔다.

힌트: 익숙한 배치에서 검증용으로 제외한 오가노이드와 완전히 새로운 배치는 서로 다른 일반화 조건이다. 전향적 추론을 하기 전에는 원래 특징을 얻은 시점도 따로 검증해야 한다.

탐구를 이어 가려면 원래 시각 표지를 사용해 실제로 나중에 일어난 수행 변화를 예측하도록 설계한다. 특징과 결과, 집단화를 미리 정한다. 인과 설명에는 전문가가 선정한 자극 방식으로 별도의 개입을 준비한다.

해설과 답은 부록 F 제31장 참조, 313쪽

## 관찰의 구조를 존중하라

오가노이드 하나의 수많은 주기는 그 표본의 상세한 이력을 제공하지만, 수많은 독립된 오가노이드가 되지는 않는다. 따라서 모든 행을 무작위로 학습 부분과 검증 부분으로 나누면

이미 익숙한 표본을 알아보는 능력을 검사하게 될 수 있다. 새 표본에 대한 일반화가 질문이라면 표본 전체를 분리해야 한다. 기증자나 배치를 공유한다면 일반화의 경계를 추가로 구체화해야 한다.

특정 시료 내부의 동역학은 다른 질문이다. 기록의 이른 부분에서 예측을 학습하고 나중 부분에서 검증하며 결론을 이 과제에 한정할 수 있다. 두 분할은 다른 질문에 답한다. 시료 안에서 쉬운 검증을 한 뒤 새로운 생물학적 계에도 일반화되었다고 주장할 수는 없다.

외부 알고리즘의 기여도 검증한다. 조직 변화와 함께 되먹임 신호도 바꾸었다면 최종 동역학은 전체 순환에 속한다. 유용한 분석에서는 알고리즘의 행동도 자료의 일부로 본다. 그렇지 않으면 외부의 자극 선택이 만든 개선을 신경망에 전부 돌릴 수 있다.

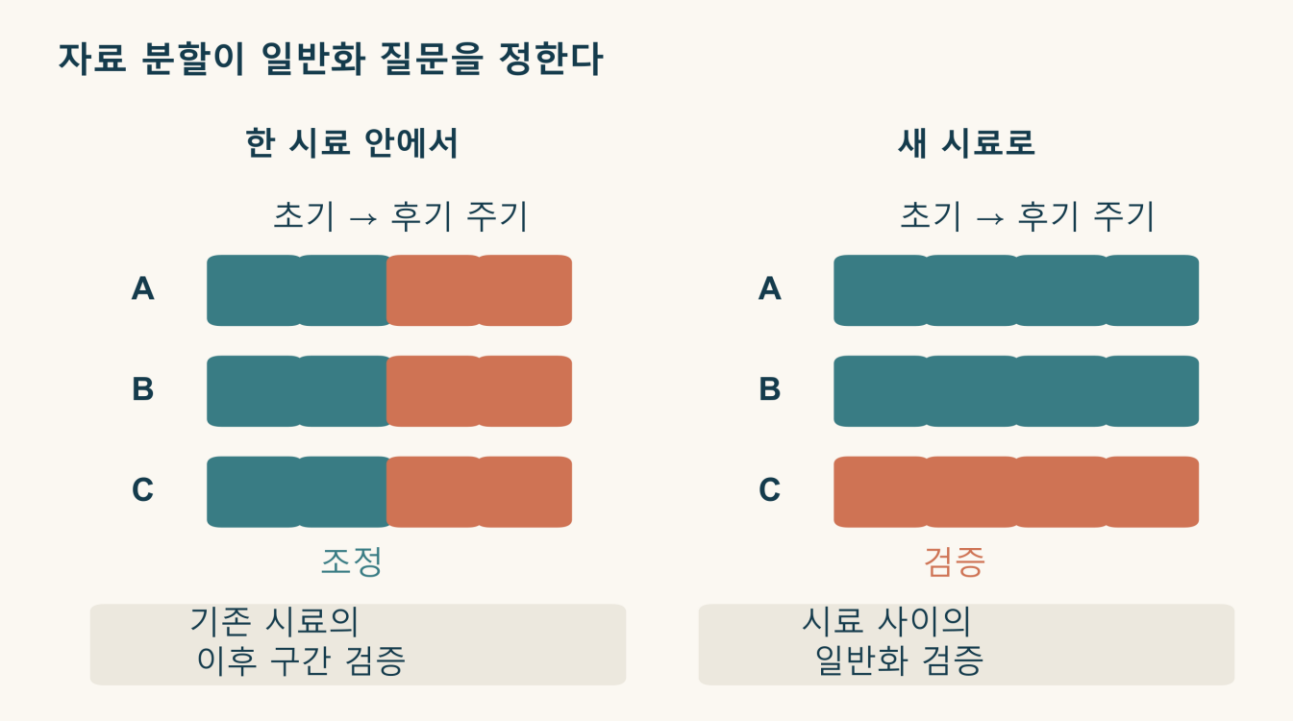


그림 79. 생물학적 자료의 시간 분할과 집단 분할. 가정한 시료 A–C의 학습 도식은 기존 시료의 나중 기록을 예측하는 것과 완전히 남겨 둔 시료로 일반화하는 것이라는 두 질문을 보여 준다. 같은 시료의 주기들은 연관된 관찰로 남는다. 공통 공여자나 배치가 있다면 그 집단화 수준도 고려해야 한다.

## 디지털 모형에서 개입 설계로

두 디지털 메커니즘을 비교한 뒤 생물학적 검증을 설계한다. 어떤 차이가 생겨야 하고 어떤 지표로 발견하며 어떤 대안이 같은 결과를 낼 수 있는지 밝힌다. 실제 개입 방식은 연구하는

계를 고려해 전문가가 선택한다. 여기서 중요한 것은 보편적인 배양이나 자극 지침이 아니라 인과 논리다.

예를 들어 닫힌 상호작용을 유지하는 조건과 적절한 입력 재생 대조를 비교할 수 있다. 자극의 수와 분포, 시료 상태를 고려해야 한다. 대조 조건의 전체 입력량이 다르면 차이를 폐회로 자체에 명확히 귀속할 수 없다. 디지털 단계에서 이런 교란을 미리 발견할 수 있다.

물리적 조직을 계획의 별도 부분으로 다룬다. 가설이 상호 인과관계를 요구하면 연구 조직과 순환 안에 실제로 존재할 근거가 필요하다. 연결성의 통계 지표가 잘 나온다고 필요한 인과 구조가 자동으로 확립되지는 않는다. 적절한 개입을 하거나 관찰한 의존성에 대한 더 제한된 결론을 내려야 한다.

## 전문적인 작업 조건

ISSCR 권고는 줄기세포 연구의 지침 가운데 하나다. 기관과 국가의 적용 절차를 대신하는 보편 법률은 아니다. 2025년 갱신은 줄기세포 유래 배아 모형을 다루고 다른 부분은 2021년판을 유지한다. 구체적인 대상과 연구에 맞추어 요구를 확인해야 한다. [48]

계획에는 재료 출처와 허용 용도, 측정 책임, 연구 검토 절차를 정해야 한다. 체험 가능성의 가설에서는 어떤 새 정보가 계의 지위 평가를 바꿀지도 미리 적는다. 이는 연구를 명확히 설정하는 일부이며 현상적 체험의 논쟁을 미리 해결하는 방법은 아니다.

## 실습의 결과

두 쪽짜리 연구 계획을 준비한다. 첫 쪽에는 질문과 이용 가능한 자료, 두 메커니즘, 일반화 검증 설계를 적는다. 둘째 쪽에는 개입과 대조 조건, 한계, 체험과의 관계에 대한 별도 논거를 적는다. 자료가 부족하면 빠진 필드와 그 역할을 설명한다.

세포나 망의 변화에서 구체적인 과제 결과까지 전이를 추적할 수 있으면 작업을 마친다. 체험으로의 마지막 전이는 명시적인 가설로 남아야 한다. 이 연쇄로 생물학적 연구가 책의 전체 연구 목표에 참여한다.

**자기 점검 문항**

1. 혼성계의 개선을 생물학적 조직 하나에 곧바로 귀속할 수 없는 까닭은 무엇인가?

2. 같은 오가노이드의 나중 기록으로 검증하는 것과 새 시료로 검증하는 것은 어떻게 다른가?

3. 과제에서의 일시적 개선이 주관적 체험의 존재를 입증하는가?

생물학적 자료로 경계와 집단화의 중요성을 확인했다. 이제 연결의 존재에서 기능과 체험에 관한 별도 논거까지 구현 검증의 공통 순서를 마련한다.

제32장

# 설계한 시스템의 검증

디지털 재료와 생물학적 재료는 도구가 달라도 검증 대상을 정확히 정해야 한다. 구성요소의 존재와 기능적 기여, 내부 접근, 가정한 체험은 서로 다른 주장이다. 이 장에서는 주장마다 근거를 갖도록 계의 양상을 정리한다. 처음부터 행위자의 지능 전체를 검증할 필요 없이 완전히 추적할 수 있는 한 사례로 시작한다.

## 검증 대상의 정의

기억과 행동 모형, 자기 평가 모듈을 갖춘 행위자를 만들었다고 하자. 과제를 풀고 의심할 때 보고한다. 이제 결과의 지위를 알아야 한다. 프로그램 검증과 기능 능력의 평가, 체험 가능성의 연구는 연결되어도 서로 다른 증거가 필요하다. 전체 점수 하나로 묶으면 다음 개선 뒤 무엇이 달라졌는지 알기 어렵다.

작동하는 전체로서 계를 기술한다. 결정 하나에서 어떤 과정이 실행되는가? 결정 사이에 무엇이 남는가? 구성요소마다 어떤 자료에 접근하는가? 외부 언어 서비스를 포함하는가? 보고는 행위자 상태를 직접 읽는가, 행동 이력을 다시 이야기할 뿐인가? 검증 대상의 경계가 이후 결과의 의미를 정한다.

## 작동하는 연결의 지도를 만들어라

구성요소마다 입력과 상태, 출력, 출력을 받는 대상을 적는다. 이어 구체적인 사례 하나를 추적한다. 예를 들어 행동 모형의 불확실성이 커지고 모니터가 기대 성공률의 감소를 감지하여 정책이 추가 관찰을 요청하고 새 관찰이 결정을 바꾼다. 추가 관찰이 목표에만 관한 것이라면 제어 불확실성은 줄이지 못할 수 있다. 관계도에는 이런 메커니즘의 한계도 보여야 한다.

모듈 이름으로 역할을 확인할 수는 없다. 코드에는 있어도 선택한 열에서 한 번도 켜지지 않을 수 있다. 상태가 바뀌어도 결정에 영향을 주지 않을 수 있다. 다른 경로가 영향을 완전히 중복할 수도 있다. 그러므로 계산 가능성과 실제 활성화, 결과에 대한 인과 기여를 구별한다.

측정 계를 따로 검증한다. 환경의 실제 상태와 개입 표지는 평가자에게 주되 행위자가 우회해 접근해서는 안 된다. 자기 평가 모듈이 이미 정해진 정답을 읽는다면 높은 보고 정확도는 정보 전달의 품질을 보여 준다. 자기 결정의 독립적인 점검을 확립하지는 않는다.

## 새 조건에서 기능을 검증하라

기억에는 이력에 따라 올바른 행동이 실제로 달라지는 과제가 필요하다. 제어 모형에는 명령과 결과의 대응이 바뀌는 조건이 필요하다. 불확실성 평가에는 확신이 틀릴 수 있는 상황이 필요하다. 편리한 시연 하나로 이 능력을 모두 검증할 수는 없다.

익숙한 조건과 일반화를 구별하면 좋다. 익숙한 과제는 성공에 충분한 단순 규칙을 외웠을 수 있다. 새 특징 조합으로 가정한 구조를 사용하는지 볼 수 있다. 그러나 새 조건에서 모든 것을 모르게 동시에 바꾸어서는 안 된다. 잡음과 목표, 가능한 행동이 함께 달라지면 실패의 원인을 특정 메커니즘에 연결하기 어렵다.

기능마다 주된 결과를 고른다. 요청은 무한히 할 수 있으므로 횟수만으로 자기 점검의 품질을 알 수는 없다. 확률 예측은 이후 결과에 비추어 평가하고 검증 정책은 이익과 비용을 고려해 평가한다. 확신에 찬 문장 하나의 내용이 여러 사례의 오류 분포를 대신하지는 않는다.

## 인과 검증을 준비하라

똑같이 설득력 있는 보고를 내는 두 실습 계를 생각해 보자. 첫 계는 자기 출처 평가에서 보고를 만들고 둘째는 최근 공개 행동으로 문구를 고른다. 평소 사건에서는 차이가 안 보일 수 있다. 외부 이력은 유지하면서 내부 출처 평가를 바꾼다. 첫 계가 정말 그 평가를 쓴다면 보고도 예측 가능한 방식으로 달라져야 한다.

그러나 개입이 낯선 지연이나 특수 토큰, 다른 응답 형식 같은 외부 단서를 만들 수 있다. 그러면 둘째 계도 조건을 알아낸다. 대조에서는 외부에서 이용 가능한 특징을 유지해야 한다. 내용 인식과 전반적 고장에 대한 반응을 구별하려면 모의 개입과 적절한 비특이적 변화도 필요하다.

프리프린트로도 공개된 연구 보고서에서 Jack Lindsay는 언어 모형의 보고와 내부 표상에 대한 개입 사이의 연관을 연구했다. 관찰된 능력은 제한적이었으며 조건에 따라 달랐다. 우리에게 중요한 것은 내부 접근을 인과적으로 검증하는 단계로 나아갔다는 점이다. 이것이 주관적 체험에 대한 결론을 자동으로 만들어 주지는 않는다. [49]

## 상태를 읽는 것과 이용하는 것을 구별하라

외부 분류기를 학습시켜 숨은 벡터의 내용을 알아낼 수 있다. 선택한 분석에서 이용할 정보가 벡터에 있다는 뜻이다. 그러나 행위자는 그 정보를 보고나 행동에 쓰지 않을 수 있다. 외부 디코딩과 내부 접근은 연구 설명에서 다른 항목으로 두어야 한다.

이어 접근이 제어에 영향을 주는지 묻는다. 모니터는 유지하고 정책과의 연결을 끈다. 평가는 맞지만 결정이 불확실성을 반영하지 않게 된다면 그 연결의 기여를 가려낸 것이다. 행동이 같다면 메커니즘의 비활성이나 중복, 기능이 필요 없는 과제 때문일 수 있다. 검증 없이 편한 설명을 고를 수는 없다.

전략 수정에도 별도 결과가 필요하다. 행위자가 특정 가정을 바꾸고 그 변화를 다음 결정에 사용해야 한다. 오류 뒤 설명 글은 유용한 인터페이스일 수 있어도 그 존재만으로 인과 연쇄를 확인하지는 못한다. 수정 없이 같은 글을 내는 버전과 비교하면 작동 메커니즘을 가려낼 수 있다.




그림 80. 내부 상태에 관한 세 가지 서로 다른 질문. 외부에서의 해독, 자기 보고의 인과적 변화, 모니터가 정책에 미치는 영향에는 서로 다른 검증이 필요하다. 내부 개입은 외부 이력을 유지하고 모의 개입 대조를 갖추어야 한다. 그렇지 않으면 시스템이 개입의 부수적 단서를 알아볼 수 있다.

## 후보의 특성 프로필을 구성하라

이론적 양상의 정리는 인상적인 성질 목록이 아니라 선택한 입장에서 시작한다. 전역 접근의 역할을 연구한다면 내용의 전파와 이용에 해당하는 조직을 검증해야 한다. 물리적 인과 구조가 중요하면 구현 자료가 필요하다. 제26장의 원리를 고르면 구별들의 상호 규정과 대안적 변환에서 그 역할을 검증한다.

Butlin과 동료들의 지표 방법은 계의 평가를 이론의 요구와 연결하자고 제안한다. 충족 항목 수를 의식의 백분율로 바꾸는 보편 척도는 아니다. 여기서는 각 결론의 근거와 불확실성을 이론적 양상에 남긴다. [42]

각 요구사항에 세 서술 상태를 사용하라. 구현하고 검증함, 구현했지만 역할은 검증하지 않음, 구현하지 않았거나 알 수 없음이다. 이어 어떤 관찰로 상태를 바꿀 수 있는지 적는다. 이런 기록은 합계 수치보다 유용하다. 어디에 코드 작업이 필요하고, 어디에 새 실험이 필요하며, 어디에서 이론 자체를 더 구체화해야 하는지 보여 주기 때문이다.

## 프로필에서 체험에 관한 주장으로

기능적 양상이 좋아도 논거는 끝나지 않는다. 선택한 성질이 체험의 존재나 질에 해당하는 이유를 뒷받침해야 한다. 자연적 체험을 다룬 장들과 연구 부분의 대응 원리를 사용한다. 직접 고른 요구를 구현했다고 그 충분성을 독립적으로 확인한 것으로 보아서는 안 된다.

계가 구별들의 상호 규정이라는 잠정 조직을 충족한다고 하자. 가정한 충분조건을 특정한 방식으로 구현했다고 말할 수 있다. 다음 질문은 조건 자체에 있다. 독립 자료를 설명하며 반례와 구현 사이의 적용을 견디는가? 이 근거는 기술 보고와 따로 제시해야 한다.

명시적 자기 보고 없이 체험을 허용하는 이론이라면 설득력 있는 말이 없다는 사실로 자동 반박되지 않는다. 특정 접근을 요구하는 이론이라면 필요한 인과 조직 없이 말을 잘한다고 확인되지 않는다. 검증은 계와 대화한 인상이 아닌 이론의 내용을 따라야 한다.




그림 81. 후보 검증의 프로필. 방법 표는 프로그램 구현의 존재, 기능적 기여, 내부 정보 이용, 가능한 체험에 관한 논증을 구별한다. 한 층위의 증거는 다른 층위의 증거를 대신하지 않는다. 외부 상태 해독기만으로 내부 접근이 확립되지는 않으며, 충족한 요구들의 합이 의식의 백분율도 아니다.

### 구별하지 못한 결과를 분석하라

전체 버전과 전략 재검토 없는 버전의 결과가 같다고 상상하라. 먼저 재검토를 시작할 조건이 생겼는지 확인하라. 한 번도 생기지 않았다면 실험은 두 버전의 차이를 이용하지 않은 것이다. «성찰이 필요 없다»는 주장도 «모듈을 작성했으므로 성찰이 작동한다»는 주장도 지나치게 강하다.

규칙이 실제로 갈라지는 과제를 찾고 실행 전에 예상 차이를 도출한다. 그래도 결과가 같다면 중복과 측정 정확도를 살핀다. 구성요소의 필요성은 구체적인 구현과 과제에 대해 확인한다. 다른 계는 같은 결과를 다른 방식으로 낼 수 있다.

이런 분석은 체험의 후보 시스템을 설계하는 데 직접적인 가치가 있다. 장식적인 구성요소가 쌓이는 일을 막고 기능적 기여가 입증된 메커니즘을 선택하도록 돕는다. 체험의 유무에 대한 최종 답을 아직 찾지 못했더라도, 체험에 관한 가설이 다루는 대상은 더 명확해진다.

### 하나의 상태 기록에 관한 네 질문

M2 모형에서 기록 A는 실제로 존재한다. 프로그램의 상태로 확인할 수 있다. 다시 나타난 A에 미치는 영향은 입력을 그대로 두고 이력을 삭제한 경우와 비교하여 검증한다. 별도 모니터가 기록에 접근하려면 판독 채널을 확인해야 한다. 기록의 존재만으로 그런 채널이 생기지는 않는다. 마지막으로 모형이 필요한 수치를 주더라도 그 조직이 왜 체험되는 내용과 관련되는지 설명해야 한다. 이처럼 하나의 로그가 서로 다른 여러 검증을 뒷받침할 수 있지만 전환마다 별도의 관찰이 필요하다.

#### 여기까지 확인한 것

성질별 양상은 비교할 수 없는 지표를 합산하지 않게 한다. 모르는 성질을 표시하면 의식에 관한 전체 인상 대신 구체적인 다음 작업이 드러난다.

## 실습 32 한 사례 점검하기

실습 목표는 구현의 한 사례를 네 개의 독립된 주장으로 나누어 분석하는 것이다.

인과 관계도와 특정 구성요소를 끄는 요소 제거 검증, 상태의 디코딩과 사용의 구별을 활용한다. 비교하는 버전에서 이용 가능한 정보를 같게 유지한다.

예상 소요 시간은 30–40분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 출처 모형에서 다시 나타난 A 또는 행위자의 한 단계를 고른다. 입력과 이전 상태, 변화, 출력, 출력을 받는 대상을 적는다.
2. 끌 연결과 예상 결과를 검증 전에 밝힌다. 그 연결 자체를 끄는지 확인한다.
3. 구현과 기능, 내부 접근, 체험 가능성의 네 행을 만든다. 수행한 검증이나 미확인 상태를 각각 명시한다.
4. 기능이 실제로 필요한 과제 하나와 비교할 만한 정보 및 자원을 유지하는 대조 조건을 제안한다.

한 사례의 관계도와 인과 검사, 지위를 구별한 성질별 기록을 남겨 둔다.

외부 분류기는 행위자가 사용하지 않는 내용도 추출할 수 있다. 내부 접근에는 연구하는 순환 안에 정보를 받는 과정이 있어야 한다.

탐구를 이어 가려면 내부 상태와 외부 이력을 독립적으로 바꾸며 글로 된 보고를 검증한다. 공개 행동만으로 만드는 설명과 비교한다.

해설과 답은 부록 F 제32장 참조, 314쪽

**자기 점검 문항**

1. 프로그램에 Monitor 클래스가 있으면 자기 점검이 작동한다고 입증되는가?
2. 상태의 외부 디코딩은 내부 접근과 어떻게 다른가?
3. 직접 고른 요구를 모두 충족해도 그 요구가 체험에 충분하다고 입증되지 않는 까닭은 무엇인가?

구체적 구현을 검증하는 법을 익혔다. 이제 형식화 규칙을 밝히고 선택한 분석이 비교하려는 메커니즘들을 실제로 구별할 수 있는지 알아본다.

제9부

# 연구 작업

제33장

# 형식화와 시뮬레이션

프로그램 검증은 예제 하나를 정확히 계산하는 것으로 끝나지 않는다. 다른 매개변수들이 비슷한 결과를 낼 수 있으며 선택한 환경에서 다른 모형들을 구별하지 못할 수 있다. 이 장에서는 생성기와 이용 가능한 관찰, 예측, 측정을 형식적으로 나눈다. 이어 자료를 만든 메커니즘을 아는 인공 자료에서 매개변수와 모형을 복원한다. 이후 원인이 미확인인 자료에 적용할 분석을 준비하는 것이다.

## 형식화는 모형이 충족할 요구를 드러낸다

말로 기술한 모형은 초기에 유용한 모호성을 허용한다. 그러나 검증하려면 두 사건 사이에 정확히 무슨 일이 일어나는지 정해야 한다. 출처가 없을 때 이력이 남는가? 숨은 원인의 확률은 매 순간 변하는가, 새 관찰 때만 변하는가? 반증은 한 기록을 갱신하는가, 전체 계를 갱신하는가? 짧은 설명은 같아도 이 선택들이 예측을 바꿀 수 있다.

형식화는 수학을 반드시 복잡하게 만드는 것이 아니라 이런 모호성을 없애는 것이다. 입력과 상태, 전이, 측정은 다른 연구자가 재현할 절차를 이루어야 한다. 프로그램으로 규칙의 귀결을 확인한다. 그래도 코드가 명세와 일치하는가, 명세가 체험 설명에 적합한가는 다른 질문이다.

## 학습용 환경을 모형과 별도로 정하라

출처 A와 B의 예를 확장하자. 각 출처에는 공통 원인 또는 별개 원인이라는 숨은 이진 상태가 있다. 실습 환경에서는 같은 출처를 다음에 사용할 때 상태가 유지될 확률이 0,9다. 지지 특징은 공통 원인에서 0,75, 별개 원인에서 0,25의 확률로 나타난다. 구성한 환경이며 생리 메커니즘을 기술한 것이 아니다.

실험자가 숨은 상태와 관찰을 정한다. 모형은 출처와 특징만 받는다. 그러므로 다음 특징의 예측과 숨은 원인의 평가를 따로 측정할 수 있다. 숨은 원인을 직접 보여 주면 필터링 과제가 사라진다. 한 파일에 작성하더라도 환경과 관찰자를 구별해야 한다.

유지 확률 0,9는 제25장의 유지 계수 0,8에 해당한다. 원인이 바뀌었을 가능성을 반영하면 이전 믿음은 앞서 사용한 규칙으로 0,5 쪽으로 이동한다. 이제 그 규칙이 자료 생성기 안에서 구체적 의미를 얻었다. 다른 환경에는 다른 동역학이 필요할 수 있다.

## 예측할 결과의 정의

믿음 0,8은 공통 원인의 확률이다. 다음 지지 특징의 확률과 다르다. 신뢰도가 0,75이면 공통 원인이어도 다음 특징이 반증일 수 있다. 지지 특징의 예측은 두 가능성을 섞어 얻는다. 공통 원인의 확률에 0,75를 곱하고 별개 원인의 확률에 0,25를 곱한 뒤 더한다.

```
def predict_positive(prior_common, reliability=0.75):
    return (
        prior_common * reliability
        + (1.0 - prior_common) * (1.0 - reliability)
    )
```

믿음이 0,8일 때 지지 특징의 예측은 0,65다. 검증 가능한 수치 차이다. 원인에 관한 믿음을 특징의 빈도와 직접 비교하면 잘못 보정된 모형처럼 보일 수 있다. 그러므로 각 출력 수치의 이름에 어떤 사건의 확률인지 밝혀야 한다.

인간의 신체 소유감 평가에는 한 수준이 더 생긴다. 실제 인과 표지도 다음 감각 특징도 아니다. 내부 상태와 평가의 관계에는 별도의 측정 모형이 필요하다. 형식화하면 결과 분석 전에 이런 전이를 알아차릴 수 있다.

## 모형의 시간을 기록하라

단순한 M2에서는 특정 출처의 기록을 다음에 사용할 때 유지 계수를 적용한다. A가 오래 나타나지 않아도 중간의 B 사건마다 기록이 약해지지는 않는다. 실습 계산에는 편한 규칙이다. 그러나 인간 기억의 가설에는 경과 시간에 대한 의존성이 필요할 수 있다.

둘째 버전은 시간 단계마다 기록을 약화한다. 계수가 0,8이고 두 단계를 건너뛰면 이전 편차에 0,8을 두 번 곱한다. 부재의 길이가 별도 예측 변수가 된다. 사건 수는 같고 간격은 다른 열로 두 버전을 비교할 수 있다.

불편한 결과를 보았다고 첫 버전을 몰래 고치지 마라. 모형의 시간을 바꾸면 새 가설이 된다. 이전 규칙을 보존하고 재검토 이유를 기술한 뒤 새 예측을 얻어라. 이는 체험과의 연결에서 특히 중요하다. 주관적 연속성을 임의로 바뀌는 시간 개념으로 설명해서는 안 된다.

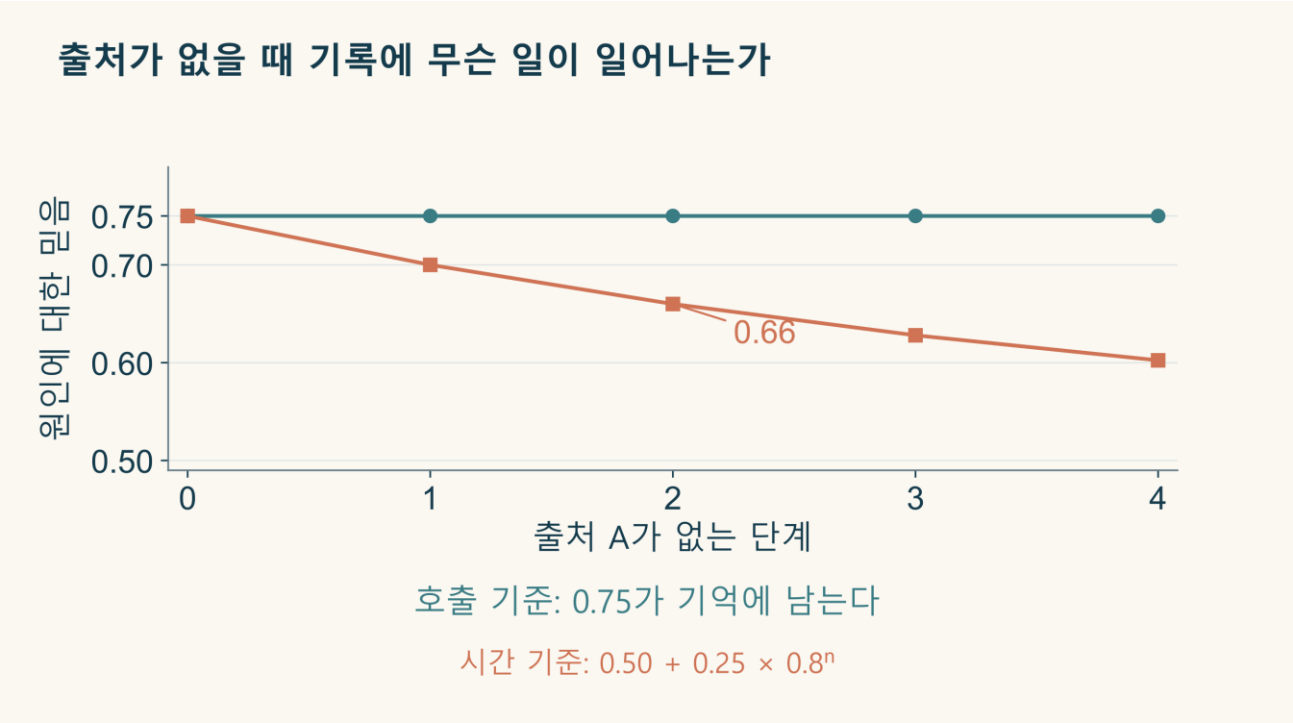


그림 82. 명확히 다른 두 기억 시간 모형. 최초 기록 0,75, 유지 계수 0,8인 정확한 학습 계산이다. A가 없고, 다시 호출하기 전이며, 새 관찰도 하기 전인 때의 보존 상태를 나타냈다. 빠진 두 단계 동안 시간 기준 버전은 0,66에 이르고 호출 기준 버전은 0,75를 보존한다.

## 매개변수의 값을 검증하라

두 메커니즘이 같은 최종 수치를 줄 수 있다. 믿음 0,75에서 양의 특징의 신뢰도가 0,75이면 0,9로 갱신된다. 최초 믿음이 0,5이고 신뢰도가 0,9여도 0,9로 갱신된다. 최종 응답 하나만으로 높은 값이 강한 이력 때문인지 강한 새 증거 때문인지 정할 수 없다.

매개변수를 구별하려면 다른 조건이 필요하다. 현재 특징의 구별을 독립적으로 평가하고 이전 이력의 길이를 바꾸며 반증 사건을 더할 수 있다. 이어 다른 매개변수 조합들이 여전히

실질적으로 구별 불가능한지 본다. 매개변수의 불확실성은 분석 결과이며 보기 좋은 조합 하나로 숨길 장애물이 아니다.

인공 자료에서 복원을 시도한다. 매개변수를 정해 열을 만든 뒤 분석에는 원래 값을 주지 않고 찾아내게 한다. 절차가 유지 계수와 신뢰도를 자주 혼동한다면 인간 연구에서도 같은 어려움이 생긴다. 같은 유형의 사건을 늘려도 구조적 모호성이 사라지지 않을 수 있다.

## 모형의 복원을 검증하라

이제 M0과 M1, M2로 각각 자료를 만들고 관찰에 미리 정한 변동을 더한다. 분석은 각 자료에서 모든 모형을 비교해야 한다. 자료를 만든 메커니즘을 선호하는 것이 이상적이지만 미리 가정할 수는 없다. 조건이 경쟁 모형을 거의 구별하지 못하거나 선택 규칙이 더 유연한 모형을 체계적으로 선호할 수 있다.

Wilson과 Collins의 방법론 연구는 매개변수와 모형의 복원을 다룬다. 여기의 생성기와 입력열은 같은 논리를 익히는 독립적인 실습이다. 검증에 성공하면 알려진 조건에서 분석이 작동함을 보여 준다. 실제 자료도 후보 중 하나가 만들었다는 증거는 아니다. [43]

어떤 후보와도 같지 않은 넷째 생성기를 더하면 좋다. 그러면 틀린 모형들 가운데 상대적으로 가장 나은 것을 고르는 방식을 볼 수 있다. 제한된 경쟁에서 확실히 이겨도 실제 메커니즘을 찾았다는 뜻은 아니다. 상대 비교에 예측의 절대 품질 검증을 더해야 한다.

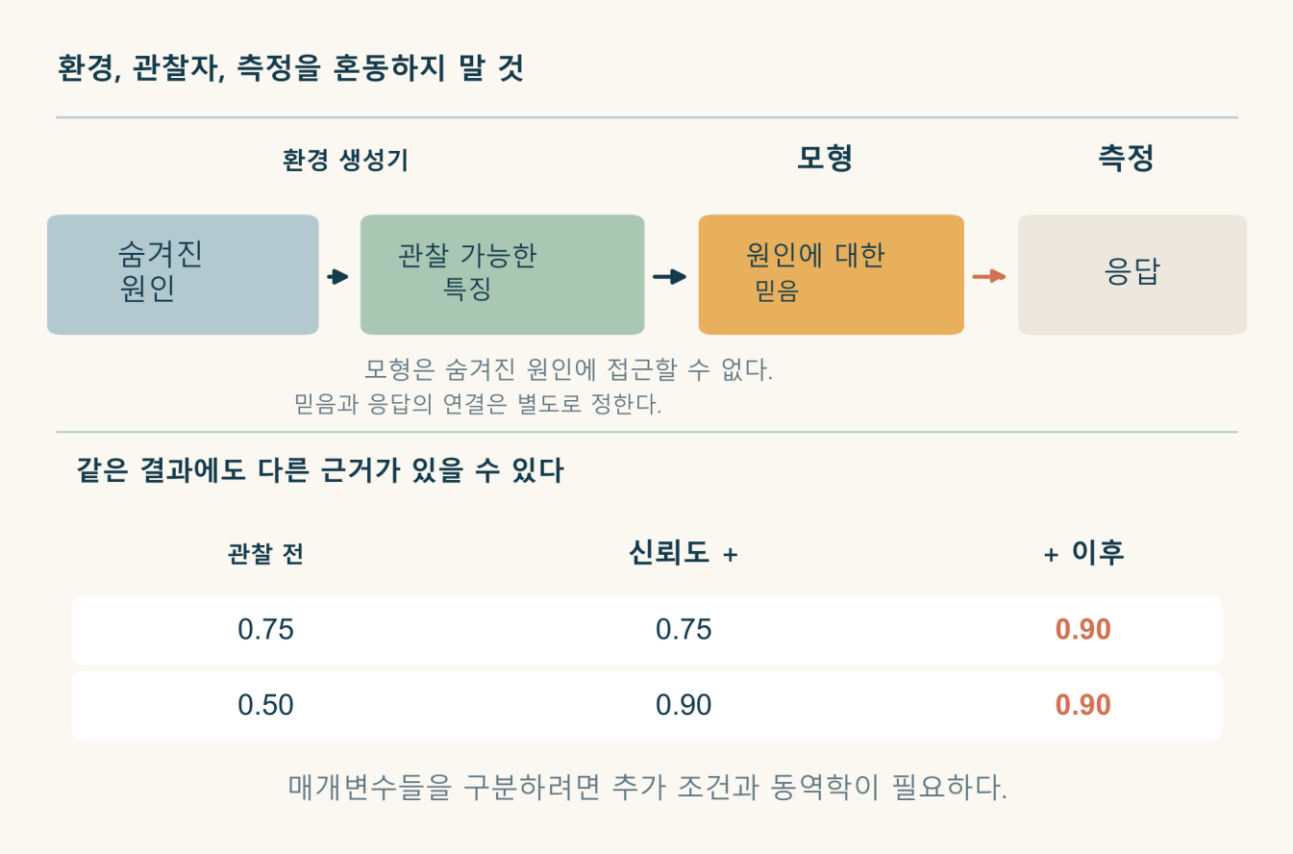


그림 83. 이용 가능한 정보의 경계와 매개변수의 비유일성. 학습 도식과 정확한 계산이다. 위쪽은 환경 생성기의 숨은 원인과 모형이 이용할 수 있는 관찰을 구분한다. 점선은 따로 정하는 응답 생성 모형이다. 아래 예에서 사전 믿음 0,75에 신뢰도 0,75인 양의 증거와, 믿음 0,50에 신뢰도 0,90인 증거는 같은 결과 0,90을 준다. 여기서 사전 믿음은 관찰 직전의 값이며 기억을 추가로 약화하지 않는다. 결과 하나로 두 매개변수가 정해지지는 않는다.

## 재현 가능성을 보존하라

실행마다 규칙의 버전, 매개변수, 초기 상태, 사건 순서를 기록하라. 무작위 생성에서는 난수 생성기의 초깃값을 고정하라. 하지만 난수를 호출하는 순서가 바뀌었다면 seed 하나로는 충분하지 않다. 추가 관찰 때문에 비교하는 버전의 이후 시나리오 전체가 뜻하지 않게 밀려서는 안 된다.

조건의 생성과 행위자의 결정을 나눈다. 각 단계의 잠재 관찰을 미리 준비하고 행동이 어떤 관찰을 공개할지 정하게 할 수 있다. 행동이 미래 환경을 물리적으로 바꾼다면 그 변화는 인과 모형의 일부여야 한다. 비교 가능성이 행동 결과를 인위적으로 금지한다는 뜻은 아니다.

원결과와 요약 프로그램도 보존한다. 그렇지 않으면 필터 하나의 수정으로 본문 수치가 모르게 달라질 수 있다. 최종 자료에서 독자는 입력열에서 예측으로, 예측에서 오차로, 오차에서 결론으로 이어지는 과정을 볼 수 있어야 한다. 인터페이스의 장식적인 복잡성보다 중요한 연쇄다.

## 시뮬레이션에서 연구 결론으로

시뮬레이션으로 메커니즘의 구현 가능성과 귀결, 모형이 구별되는 조건을 확인할 수 있다. 선택한 규칙에서 원하는 결과가 불가능함을 보일 수도 있다. 그러나 모형이 만든 인공 자료는 인간이나 생물학적 망에 적용할 수 있다는 독립적 확인이 아니다.

제26장 가설에는 실행을 추가로 기술해야 한다. 계산 조직이 중요하다면 개입의 응답을 포함해 프로그램이 실제로 구현함을 보인다. 특정 물리 상호작용이 중요하다면 그 수준에서 근거가 필요하다. 시뮬레이션 그래프가 담지체에서 실제로 어떤 과정이 일어나는가라는 질문을 대신하지는 않는다.

## 실행 가능한 복원 절차

확장 실습에서 생성기 M0는 확률 0,5인 독립 이진 특징을 만든다. 생성기 M1은 전체 열에 숨은 원인 하나를 유지하고 M2는 번갈아 나타나는 출처마다 별도 원인을 유지한다. 출처를 사용할 때 원인이 유지될 확률은 0,9다. 지지 특징은 원인 1에서 0,75, 원인 0에서 0,25의 확률로 나타난다. 처음 두 원인의 확률은 같다. 이렇게 생성 환경을 분석 모형과 독립적으로 정한다.

모형 복원은 세 생성기 각각에 대해 사건 120개로 이루어진 독립 순서 30개를 사용한다. 분석하는 M0, M1, M2는 같은 출처 표지와 특징을 받지만 숨은 원인은 받지 않는다. 시행마다 먼저 양의 특징의 확률을 보존하고, 이어 실제 값을 공개한 뒤 믿음을 갱신한다. 여기서 주 기준은 평균제곱오차다. 결과가 정확히 같으면 동률로 남긴다. 승자를 무작위로 정해도 구별 가능성이 늘어나지는 않는다.

M2의 매개변수 복원에는 유지 계수 0부터 1까지 간격 0,1, 신뢰도 0,55부터 0,95까지 간격 0,05인 격자를 미리 정한다. 각 조합에서 알려진 인공 자료에 대한 전체 궤적 오차를 계산한다.

이 작은 탐색으로 거의 같은 이웃 해들을 볼 수 있다. 실습 절차이며 보편적으로 가장 좋은 추정법은 아니다.

출처가 하나인 대조에서는 M1과 M2가 같아야 한다. 분석이 실제 생성 메커니즘을 선호하지 않은 열도 모두 보존한다. 선택 행렬은 이 자료량과 기준이 이 생성기들을 얼마나 구별하는지 보여 준다. 환경과 잡음, 격자가 달라지면 새로 검증해야 한다.

분석 도구를 검증한 뒤 제34장의 주실험을 정할 수 있다. 복원을 검증하는 평가자가 알려진 인공 원인을 사용하는 것은 정당하다. 그러나 예측하는 후보 자체에 원인을 주면 과제가 훼손되어 성공이 정보를 주지 못한다.

## 같은 결과만으로 확신의 기원이 정해지지는 않는다

믿음 0,75에서 양의 특징의 가능도가 0,75이면 0,9로 갱신된다. 최초 값이 0,5이고 신뢰도가 0,9여도 0,9가 나온다. 최종 수치 하나로는 과거와 새 증거의 강도를 복원할 수 없다. 따라서 분석에는 궤적, 모순되는 특징, 여러 조건이 필요하다. 알려진 생성기의 자료로 검증하면 절차가 매개변수를 혼동하는 곳과 모형 전체를 혼동하는 곳을 알 수 있다. 경쟁 모형이 계산상 같다면 검증도 인위적인 승자 대신 동률을 찾아야 한다.

### 여기까지 확인한 것

시뮬레이션으로 규칙의 귀결과 구별 가능성을 연구할 수 있다. 인간 체험에 적용할 수 있다는 독립적인 확인은 아니다. 이제 비교 도구 자체를 검증한다.

## 실습 33 메커니즘의 불확실성 살피기

실습 목표는 알려진 생성기에서 복원을 재현하고 모호한 영역을 찾는 것이다.

M0-M2의 갱신과 숨은 원인의 확률 및 특징 확률의 구별, 독립 반복의 단위를 활용한다.
실습에는 비교와 복원의 실행 예제가 추가되어 있다.

예상 소요 시간은 40–60분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. python teaching_lab.py --chapter 33을 실행한다. 생성기와 매개변수 격자는 실습 앞에 설명되어 있다. 현재 특징을 공개하기 전에 모든 예측을 적는다.

2. 각 열에서 M0과 M1, M2의 오차를 살핀 뒤 알려진 생성기의 선택 표를 본다. 모형 분류의 오류를 모두 보존한다.

3. M2에서 여러 유지 계수와 신뢰도를 비교한다. 오차가 가까운 이웃 조합을 찾아 중간 상태를 확인한다.

4. 단일 채널 조건에서 M1과 M2의 동일성을 검증한다. A가 빠진 열에서 기록 사용 횟수 기준 시간과 실제 시간 단계 기준 시간을 비교한다.

복원 결과와 동일성을 아는 대조, 매개변수의 구별을 강화할 조건 하나를 남겨 둔다.

힌트: 자료를 생성한 모형이 아닌 후보가 이긴 계열을 버리지 마라. 그 계열들도 선택한 절차와 정보량의 특성을 보여 준다.

탐구를 이어 가려면 후보 어느 것과도 같지 않은 생성기를 더하고 최선 모형의 절대 품질을 평가한다. 실행 전 새 계획을 확정하고 변동성과 관찰 수를 바꾸어 복원을 검증한다.

해설과 답은 부록 F 제33장 참조, 314쪽

**자기 점검 문항**

1. 공통 원인의 확률과 지지 특징의 확률이 다른 까닭은 무엇인가?
2. M2 자체가 만든 자료에서 M2를 복원하면 인간에 대해서도 참이라고 입증되는가?
3. 망각 규칙을 바꾸려면 모형의 별도 버전이 필요한 까닭은 무엇인가?

이제 분석 자체의 한계를 알았다. 다음 장에서는 결과가 자기 검증 규칙을 고르지 않도록 새 자료에 적용할 방식을 미리 정한다.

제34장

# 자료 수집 이전의 프로토콜

사전 계획은 질문과 미래 자료, 결과에 따른 결정을 연결한다. 유리한 그래프를 임의로 고르는 일을 막는 데만 필요한 것이 아니다. 절차가 필요한 메커니즘을 검증하는지도 드러난다. 여기서는 공통 이력과 출처별 이력을 비교하는 실행 가능한 절차를 준비한다. 그 논리는 집단화와 측정의 추가 요구를 갖춘 인간 및 생물학적 연구에도 적용할 수 있다.

## 결정을 미리 기술하는 이유

그래프 여러 개를 얻었다고 하자. 하나에서는 M2가 M1보다 좋고 다른 하나에서는 차이가 거의 없다. 첫 그래프를 고르고 시간 구간을 바꾸어 이 지표가 처음부터 중요했던 이유를 설명할 수 있다. 저자 자신에게도 설득력 있게 들릴 수 있다. 사전 절차는 실제로 무엇을 예상하고 검증하려 했는지 보존한다.

사전등록은 뜻밖의 결과를 연구하지 못하게 하는 것이 아니다. 미리 정한 예상을 검증하는 일과 본 자료에서 새 가설을 만드는 일을 구별한다. Nosek과 동료들의 방법론 제안의 핵심이다. 여기서 중요한 것은 문서를 게시하는 의식이 아니라 결정의 투명성이다. [10]

아래는 실습용 절차 예시이며 등록하거나 수행한 연구의 보고는 아니다. 날짜와 버전을 보존하며 자신의 후속 검증을 위해 채울 수 있다. 자료를 이미 살펴보았다면 그 사실을 적고 확인 단계에는 새 독립 자료를 선택한다.

## 하나의 핵심 질문에서 시작하라

우리의 질문은 출처별 이력이 공통 이력보다 다음 특징을 더 잘 예측하도록 돕는가다. 이는 제33장에서 준비한 설계의 계산 단계다. 인간의 신체 소유감을 직접 검증하거나 프로그램의 체험 존재를 확립하지 않는다. 대신 나중에 더 복잡한 자료와 비교할 메커니즘을 정확히 정할 수 있다.

주가설은 독립적이고 안정적인 출처가 둘인 환경에서 M2가 M1보다 낫다는 것이다. 대조 가설은 출처가 하나이고 유지 규칙이 같으면 두 모형이 일치한다는 것이다. M0는 이력을 이월하지 않는 기준이다. 같은 출처별 기록으로 구현한 M1I는 M2와 일치하므로 경험적으로 구별 가능한 별도 대안으로 세지 않는다.

가설의 근거도 적는다. M1은 서로 다른 출처의 증거를 섞고 M2는 따로 보존한다. 가정한 이점의 메커니즘이다. «더 똑똑한 모형»이라는 말은 아무것도 더하지 않는다. 조건에서 서로 충돌하는 이력이 생기지 않으면 예상 차이가 매우 작을 수 있다.

## 앞으로 얻을 자료를 완전히 정하라

실습 계획에는 사건 백이십 개로 된 독립열 이백 개를 정한다. 실습을 재현할 규모이며 과학적 표본의 충분성을 보장하는 기준은 아니다. 절반은 출처 하나, 절반은 출처 둘을 포함한다. 출처가 둘이면 번갈아 사용한다. 처음 숨은 상태들은 같은 확률로 고른다.

출처를 다음에 사용할 때 숨은 상태가 유지될 확률은 0,9다. 지지 관찰은 공통 원인에서 0,75, 별개 원인에서 0,25의 확률로 나온다. 모형은 매개변수를 알지만 숨은 상태는 모른다. 유리하게 규정한 조건에서 메커니즘을 검증하는 버전이다. 잘못된 가정 아래의 안정성은 별도 연구다.

생성 전에 난수 알고리즘과 초깃값 목록을 고정하라. 일부터 이백까지의 연속값을 쓰고 조건을 번호에 미리 대응시키면 편리하다. 결과를 본 뒤 유리한 seed를 고르지 마라. 조건을 섞어 배정해야 한다면 그 배정 규칙도 실행 전에 고정한다.

## 예측과 기록 시점을 정하라

사건 전마다 지지 특징의 확률을 계산한다. 먼저 원인의 가능한 변화를 반영해 이전 믿음을 갱신하고 관찰의 예측을 구한다. 예측을 보존한 뒤에만 현재 특징을 공개하고 믿음을 바꾼다. 이렇게 해야 정답을 예측하면서 그 정답 자체를 사용하지 않는다.

주지표는 보통 Brier score라고 부르는 확률 예측의 평균제곱오차다. 지지 사건의 실제 값은 일이고 반증 사건은 영이다. 예측에서 실제 값을 뺀 차이를 제곱한 뒤 평균한다. 작을수록 선택한 의미에서 확률 예측이 좋다.

예를 들어 지지 사건 전 예측이 0,7이면 오차는 0,09이고 0,4이면 0,36이다. 계산을 보여 주는 예이며 앞으로 수행할 검증 결과는 아니다. 관찰 특징에 관한 지표이므로 퀄리아의 척도나 인간 자기감의 진실성을 직접 나타내는 척도로 사용할 수 없다.

## 주 비교를 고정하라

순서마다 M1의 오차에서 M2의 오차를 뺀 값을 계산하라. 차이가 양이면 M2의 이점이다. 주 결과는 출처가 둘인 조건의 평균 쌍별 차이이며 불확실성은 독립 순서들을 기준으로 평가한다. 한 순서 안의 사건들은 이력으로 연결되므로 사건 수만큼 연구 전체를 독립 반복했다고 여겨서는 안 된다.

출처가 하나인 조건에서는 선택한 M1과 M2 구현이 정확히 일치하리라 예상한다. 이는 프로그램 논리의 대조다. 결과가 다르면 먼저 입력, 갱신, 초기 상태의 차이를 찾아야 한다. 이런 사례를 곧바로 기억에 관한 뜻밖의 발견으로 바꾸어서는 안 된다.

M2와 M0의 비교와 개별 단계의 그래프는 추가 분석으로 미리 정할 수 있다. 자료를 본 뒤 새 시간 구간을 찾았다면 분석은 유용해도 탐색적 지위를 갖는다. 추가 결과가 더 보기 좋다는 이유로 주비교를 뒤늦게 바꾸지 않는다.

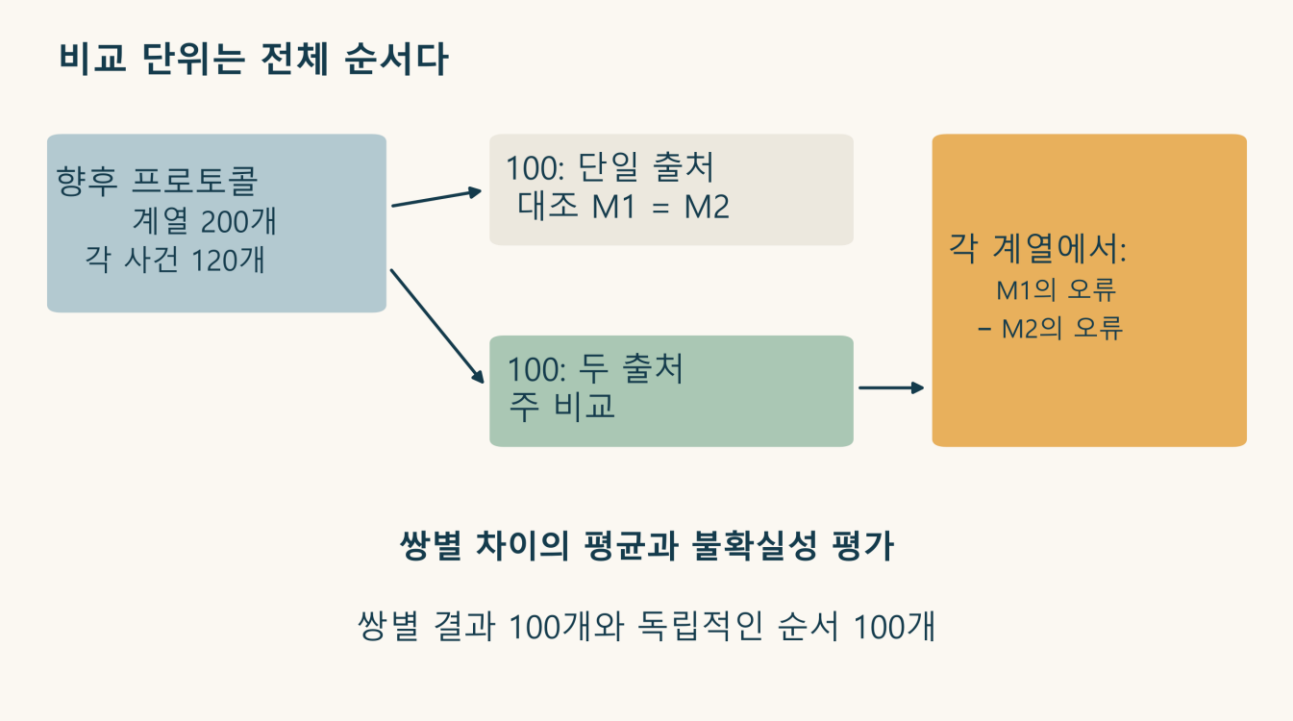


그림 84. 사전 프로토콜의 자료 구조. 수치는 제34장의 향후 학습 프로토콜에서 가져온 것으로 실습 규모를 정하며 과학적 표본의 기준이 아니다. 주 비교의 불확실성은 출처가 둘인 독립 순서 백 개로 평가하며 M1과 M2 결과의 짝을 보존한다.

## 실질적인 결과의 기준을 명시하라

평균 차이가 양이어도 너무 작아 모형 사용에는 영향이 거의 없을 수 있다. 방향뿐 아니라 관심 효과 크기나 필요한 추정 정밀도를 정한다. 실습에서는 논의를 위해 Brier 차이 0,01을 미리 고를 수 있다. 실습에서 정한 문턱이며 보편적인 유의성 경계는 아니다.

본격 연구에서는 과제에서 문턱을 도출한다. 복잡성을 더할 만한 이득은 얼마이며 어떤 차이로 경쟁 설명을 구별할 수 있는가? 이어 시뮬레이션으로 필요한 정밀도에 맞는 자료량을 고른다. 미리 정한 자원이 부족하면 «효과가 없다»고 바꾸어 말하지 말고 불확실성을 결론에 남겨야 한다.

«모형이 이길 때까지 자료를 더한다»는 중단 규칙을 세워서는 안 된다. 고정된 규모나 근거 있는 순차 절차를 미리 선택할 수 있다. 두 경우 모두 규칙은 프로토콜에 들어간다. 불가피하게 중단한 이유와 실제 규모는 최종 보고서에서 드러나야 한다.

## 오류 처리 규칙을 준비하라

디지털 과제에서는 허용 범위 밖 확률과 빠진 사건, 비교 모형의 시나리오 불일치처럼 기술적으로 불가능한 결과를 미리 정해야 한다. 이런 경우 구현을 수정하고 버전을 보존해 다시 실행한다. 불편하지만 허용되는 결과를 제외하는 것은 오류 수정이 아니다.

인간이나 생물학적 망 연구에는 미완료 절차와 기록 유실, 사전 측정 품질 기준 위반 같은 다른 근거가 있다. 규칙은 효과 방향이 아니라 자료의 적합성에 연결해야 한다. 생물학적 자료에서는 독립 비교의 단위와 이용 가능한 시료 정보도 기록한다.

누가 언제 제외를 결정하는지 밝힌다. 집단이나 결과를 아는 데 결정이 의존하면 공개해야 한다. 절차가 연구를 무오류로 만들지는 않는다. 결정이 결론에 미칠 영향을 평가할 수 있게 한다.

## 발견의 여지를 남겨라

반증 이력이 있는 출처가 돌아온 뒤 M2가 특히 유리하다고 하자. 주예측은 아니었어도 새 메커니즘을 암시할 수 있다. 기록하고 기술적 설명을 검증하며 다음 구별 질문을 정한다. 이어 수정한 생각이 틀릴 수 있는 새 열을 준비한다.

이 방식에서 탐색 분석은 열등한 작업이 아니다. 미래 검증의 내용을 만드는 다른 역할을 한다. 두 단계를 섞어 발견한 차이를 사전 예상으로 제시하고 증거의 힘을 부풀릴 때 문제가 된다.

## 하나의 질문이 분석을 정하는 방식

출처별 이력의 이점을 묻는다면 적어도 두 출처를 구별해야 한다. 독립 시나리오마다 공통 기억과 출처별 기억의 평균 오차를 구하고 그 차이를 계산한다. 양의 차이는 M2의 이점이다. 모든 사건을 독립 반복으로 평균하면 시나리오 안의 공통 이력을 무시해 정밀도를 부풀릴 수 있다. 그러므로 실습 절차의 반복 단위는 전체 열이며 단일 채널 환경은 동일성 대조다.

### 여기까지 확인한 것

계획에 따라 그래프를 본 뒤 지표를 고르지 않아도 주비교를 실행할 수 있어야 한다. 동시에 후속 발견과 새 가설의 자리는 남겨 둔다.

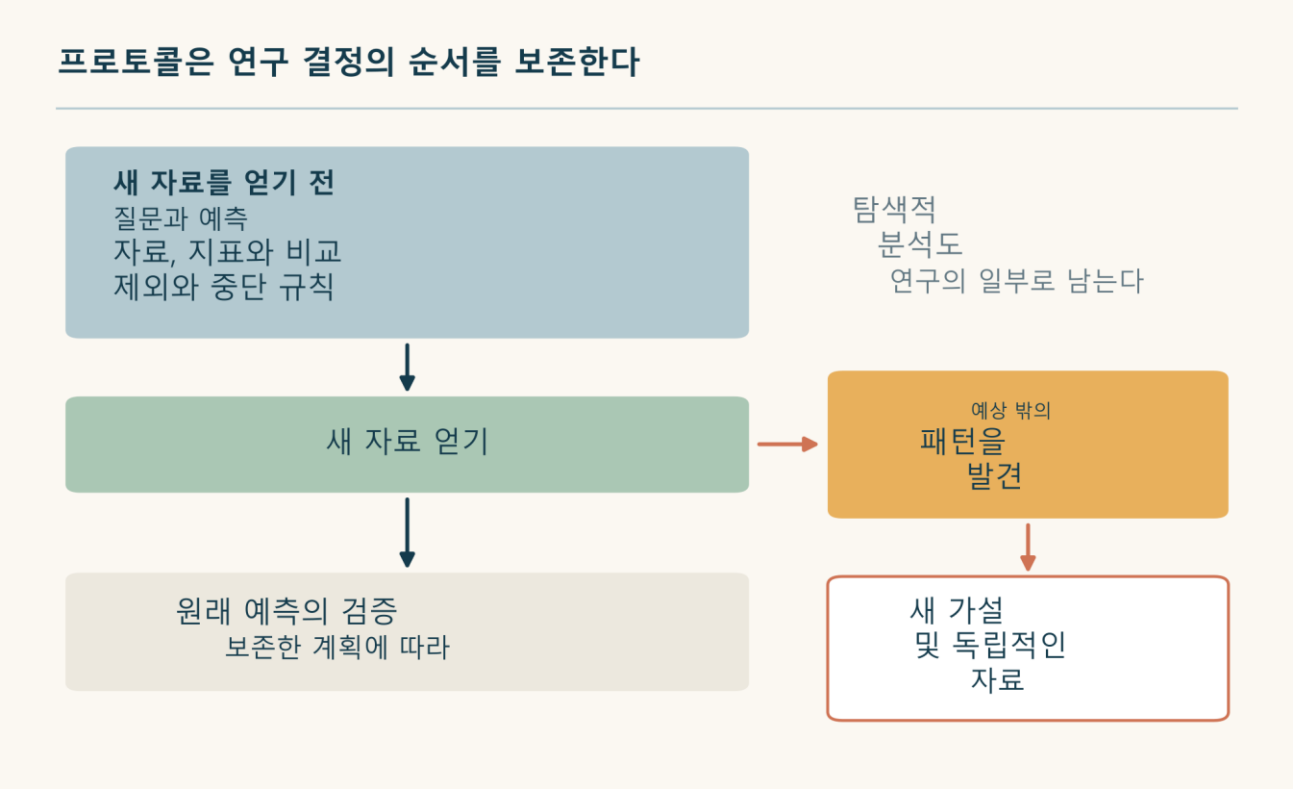


그림 85. 사전 계획과 이후의 발견. 방법 도식이다. 미리 보존한 프로토콜이 확인 검증을 정한다. 얻은 자료의 뜻밖의 패턴은 새 가설과 독립 자료를 사용하는 후속 설계를 낳을 수 있다. 이 경로는 유용하지만 분석 뒤 발견한 효과를 사전에 예측한 효과로 바꾸지는 않는다. 도식은 사전등록이나 실험을 실제로 수행했다고 보고하지 않는다.

## 실습 34 자기 절차 작성하기

실습 목표는 다른 사람이 저자를 대신해 추가 결정을 하지 않고도 실행할 수 있는 절차를 만드는 것이다.

제33장의 생성기와 분석, 주지표, 쌍별 비교, 독립 반복의 단위를 활용한다. 확인에 사용할 새 자료와 이미 본 결과를 구별한다.

예상 소요 시간은 35–50분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 실습의 protocol_source_history.txt 예시를 사용한다. 두 출처 환경에서 M1과 M2의 Brier 쌍별 차이에 관한 질문을 정한다.
2. 사건 120개인 독립열 200개를 정한다. 출처 하나의 열은 100개, 출처 둘의 열은 100개다. 생성 규칙과 매개변수, seed, 예측 시점을 명시한다.

3. 새 실행 전에 주지표와 실습상 의미 기준 0,01, 분석 단위, 기술 오류 처리, 중단 규칙을 보존한다.

4. 이미 본 자료를 밝히고 탐색 분석을 따로 적는다. 새 연구를 직접 할 때에는 새 시나리오를 사용한다.

날짜가 있는 전체 절차의 버전과 확인 및 탐색 단계의 명확한 경계를 남겨 둔다.

결과를 안 뒤 쓴 절차는 이름만 바꾸어도 사전 절차가 되지 않는다. 결정의 출처를 보존한다.

탐구를 이어 가려면 적합성 기준과 참여자, 시료, 공여자, 배치의 수준을 더해 실제 측정의 절차를 준비한다. 질문에 비추어 필요한 정밀도의 근거를 제시한다.

해설과 답은 부록 F 제34장 참조, 315쪽

**자기 점검 문항**

1. 사전등록은 뜻밖의 분석을 금지하는가?
2. 현재 특징을 공개하기 전에 예측을 보존해야 하는 까닭은 무엇인가?
3. 독립열 이백 개는 모든 연구에 충분한 규모인가?

계획을 보존하면 결과를 얻게 된다. 제35장에서는 이점이 작거나 구성요소가 켜지지 않거나 최선 후보들도 여전히 틀릴 때 결론이 어떻게 달라지는지 본다.

제35장

# 검증과 재검토

연구는 결과에 근거한 결정으로 마쳐야 한다. 자료는 메커니즘을 지지하거나 효과 크기를 제한하거나 설명들을 구별하지 못할 수 있다. 먼저 정보가 들어오는 순서와 결과를 아는 대조를 복원하고 크기와 불확실성을 평가한다. 그다음에 설명을 바꾼다. 이렇게 하면 수치의 기술적 개선을 과학적 진전으로 오해하지 않게 된다.

## 결과는 우리의 예상을 바꾸어야 한다

연구의 끝은 수치를 얻는 것이 아니라 이제 어떤 주장을 할 수 있는지 밝히는 것이다. 긍정적 결과는 메커니즘이나 공학적 가능성을 지지할 수도, 한 구현의 이점만 보여 줄 수도 있다. 부정적 결과는 구체적인 예측을 반박하거나 과제의 감도 부족을 드러내거나 불확실하게 남을 수 있다. 경우마다 다음 단계가 다르다.

출처별 기억 연구를 보자. 두 안정된 출처의 특징을 예측할 때 M2가 공통 기억 M1보다 좋으리라 예상했다. 실행 뒤에는 같은 관찰과 예측 시점, 갱신 기회를 비교했는지 물어야 한다. 그다음 이점의 크기와 처음 생각과의 관계를 논할 수 있다.

## 먼저 인과적 순서를 복원하라

가장 설득력 있는 오류는 결과를 좋게 만드는 오류다. 프로그램이 현재 특징으로 믿음을 먼저 갱신하고 같은 특징의 예측으로 저장했다고 하자. 예측 오차는 작아지지만 예측해야 할 정답을 사용한 것이다. 큰 이점은 여기서 절차 위반의 증거다.

사건 몇 개를 골라 이전 상태와 이용 가능한 정보, 사전 예측, 공개된 특징, 새 상태를 모두 복원한다. 최종 그래프뿐 아니라 전이 순서도 확인한다. 능동 행위자라면 행동과 비용, 실제로 얻은 추가 관찰도 더한다. 이 기록으로 계가 어떤 정보를 가졌는지 확인한다.

이어 정의상 결과를 아는 대조를 수행한다. 단일 채널에서는 M1과 M2가 같아야 한다. 유지 계수가 영이면 모든 버전이 이전 상태를 잊는다. 같지 않다면 구현 규칙에서 설명해야 한다. 오류 수정에는 새 버전과 결과 재생성이 따르며 이전 수치는 폐기된 결과로 남긴다.

## 크기와 불확실성을 평가하라

가상 순서 네 개에서 M1과 M2의 오차 차이가 0,018, 0,006, −0,004, 0,012라고 하자. 이는 설명을 위해 만든 학습용 값이지 수행한 실험의 결과가 아니다. 평균 차이는 0,008이다. 평균적으로 M2가 낫지만 한 순서에서는 반대 결과가 나왔다. 평균의 부호만으로 실질적 결론을 내리기에는 부족하다.

앞 장의 절차에서는 차이 0,01을 실습 기준으로 정했다. 예시 평균은 이보다 작다. 그래도 열 네 개만으로 불확실성 평가 없이 관심 이점이 없다고 선언할 수 없다. 어떤 크기가 자료와 양립하며 사전 질문에 비추어 추정이 얼마나 정밀한지 알아야 한다.

내부 이력을 공유하는 열에서는 전체 열이 반복 단위다. 사건 백 개짜리 실행 네 번을 메커니즘의 독립 반복 사백 번으로 바꿀 수 없다. 불확실성을 위해 재표집한다면 독립열 전체를 짝지은 결과와 함께 옮겨야 한다. 사람과 공여자, 세포 시료의 집단화 수준은 연구 설계에 따라 정한다.

보고서에는 차이의 크기, 불확실성, 선택한 척도의 뜻을 담아야 한다. «영의 이점과 실용적으로 관심 있는 이점을 구별하기에 자료가 충분히 정밀하지 않다»는 표현이 «모형들이 같다»는 말보다 유용하다. 같다는 주장에는 별도의 근거가 필요하다. 우리 예의 M2와 M1I처럼 규칙에서 직접 같음이 따르는 경우도 있으며 이때 통계적 비교는 불필요하다.

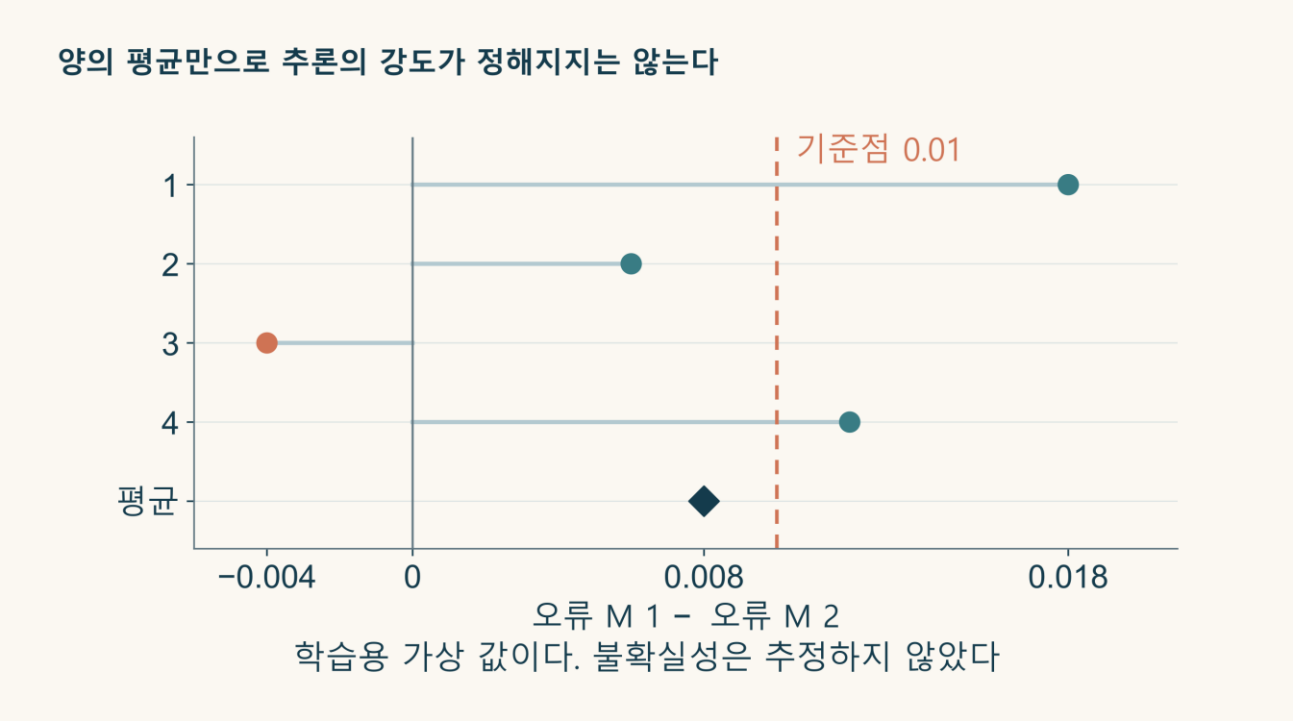


그림 86. 설명용 쌍별 차이와 가정한 기준점. 제35장에서 설명을 위해 만든 네 값의 평균은 0,008이며 기준점은 0,01이다. 그래프에는 계산한 신뢰구간이 없다. 평균의 부호나 기준점에 대한 위치만으로 모형의 동등성도 관심 있는 이득의 부재도 결론 내릴 수 없다.

## 이점의 출처 찾기

전체 사건의 평균에는 다른 상황들이 섞인다. 출처가 돌아올 때 특히 유리하고 바뀐 직후 불리하며 길고 균질한 구간에서는 경쟁 모형과 같을 수 있다. 이 양상으로 이점의 설명을 검증한다. 이력의 충돌 때 출처별 기억이 중요하다면 해당 조건이 특별한 의미를 가져야 한다.

미리 계획한 분석과 자료를 본 뒤 찾은 규칙성을 구별한다. 둘 다 의미가 있을 수 있다. 후자에서는 새 예측을 만들어 독립열에서 검증한다. 불편한 구간을 지운 뒤 남은 이점을 원래 검증 전체의 결과로 제시해서는 안 된다.

절대 품질도 검증한다. 세 예측 중 최선도 체계적으로 틀릴 수 있다. 예를 들어 새 환경에서 자주 바뀌는 출처를 모든 후보가 매우 안정적이라고 볼 수 있다. M2의 이점이 공통 결함을 없애지는 않는다. 다음 단계는 기억 기록만 늘리는 것이 아니라 안정성 자체가 변하는 모형을 연구하는 것이다.

## 구성요소의 존재와 기여를 구별하라

구성요소 제거 실험은 특정 과제의 특정 구현에 답한다. 기억을 없애 결과가 나빠지면 비교 조건 아래의 기여를 보여 준다. 바로 그런 기억 없이는 과제를 아예 풀 수 없다는 결론은 아니다. 다른 메커니즘이 필요한 구별을 다른 형태로 보존할 수 있다.

구성요소를 없애도 아무것도 달라지지 않는 반대 경우도 있다. 실제 사용했는지 확인한다. 전략 수정 조건이 한 번도 충족되지 않았다면 전체 버전과 고정 버전의 일치로 실제 수정이 일어날 때의 이점을 평가할 수 없다. 제30장의 공학 실습에서 본 사례다. 수행한 검증의 범위를 보여 준다.

구별 검증을 하려면 구성요소가 켜질 것으로 미리 예상하는 환경을 정하고 활성화 빈도를 따로 측정한다. 이어 정보와 계산 자원에 비교할 만한 접근을 가진 대조와 비교한다. 새 조건은 새 연구라고 밝혀야 한다. 이전 실행을 소급해 긍정적 결과로 만들 수는 없다.

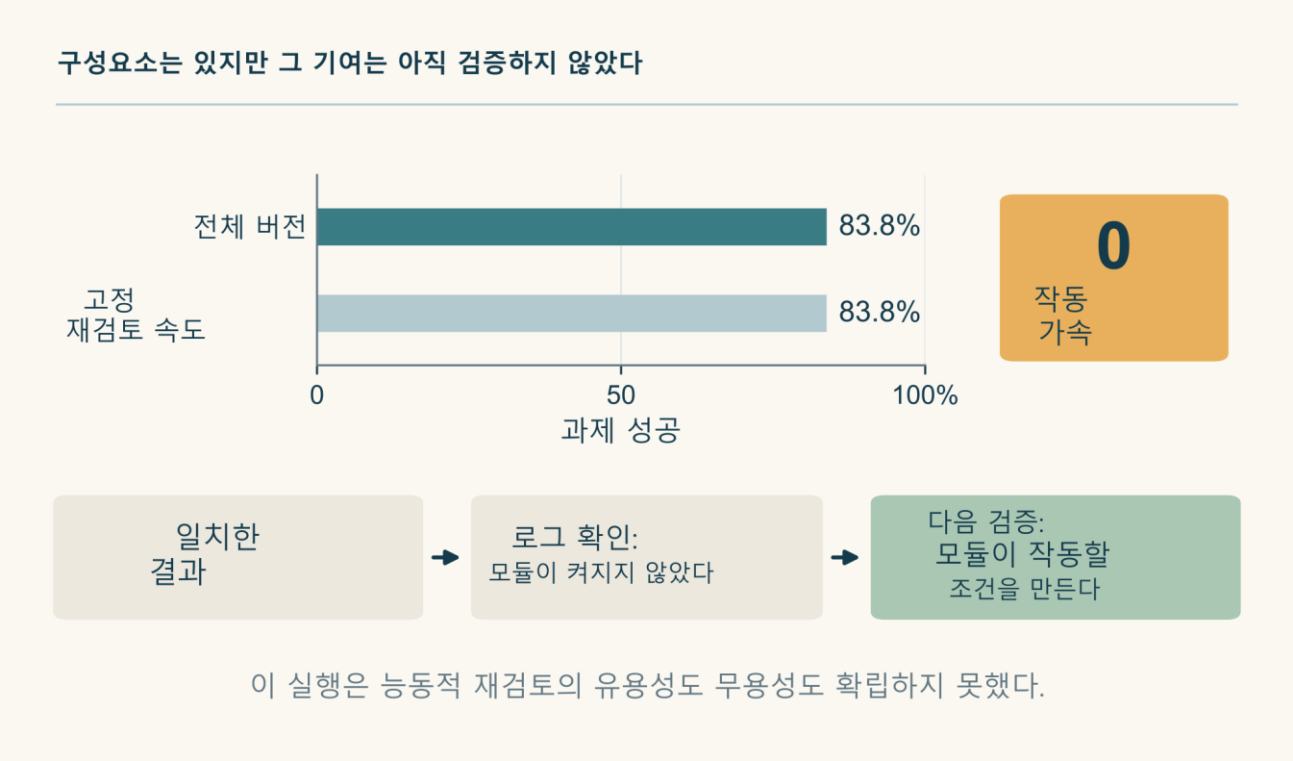


그림 87. 실습 8의 구별하지 못한 결과를 정직하게 제시하기. 보존한 프로그램 실행의 결과로, mechanisms_results.json의 실습 8이며 한 시나리오의 1000단계다. 전체 버전과 fixed_revision 버전은 각각 838단계에서 성공했다 (83,8%). 전체 버전의 가속 학습은 한 번도 켜지지 않았다. 이는 이 실행에서의 정확한 일치이며 모든 환경에서 두 메커니즘이 통계적으로 동등하다는 뜻이 아니다. 구성요소를 작동시키는 다음 검증은 새 연구가 된다. 이 지표들은 퀄리아를 평가하지 않는다.

## 메커니즘을 공개적으로 수정하라

출처가 없는 시간이 길어질수록 영향이 M2의 예측보다 더 약해진다고 하자. 매번 유지 계수를 다시 맞추는 길이 있다. 더 의미 있는 길은 출처를 사용하지 않아도 기록이 바뀌는 새 시간 모형을 제안하는 것이다. 제33장에서 두 규칙의 차이를 보았다.

새 버전에는 틀릴 위험이 있는 귀결이 생긴다. 관찰 수가 같아도 휴지기 길이가 결과에 별도로 영향을 주어야 한다. 경쟁 설명은 중간에 끼어든 사건 수로 약화를 설명할 수 있다. 그러므로 다음 실험에서는 휴지기 길이와 사건 수를 나눈다. 이전 곡선을 더 잘 맞추는 데 그치지 않고 새 차이를 만들 때 수정이 과학적이 된다.

어떤 관찰로 바꾸었고 어떤 규칙을 대체했으며 새 조건에서는 무엇을 예상하는지 버전 표에 보존한다. 이 이력이 연구의 실질적 연속성을 지킨다. 주설명이 아니게 된 실패 버전도 대조나 특수 사례로 유용할 수 있다.

## 출발한 현상으로 돌아가라

이제 인간 대상 연구에서 나왔다고 가정한 결과를 생각해 보자. M2가 M1보다 신체 소유감 평가의 역동을 더 잘 예측하지만, 일반적인 동일성 모형 M1I도 그만큼 잘 예측한다. 근거 있는 결론은 원인이나 대상에 대한 분리된 이력을 지지한다는 것이다. 특별한 자기 메커니즘이라는 주장은 아직 일반적인 출처 인식과 분리되지 않았다.

다음 질문은 그 차이를 다루어야 한다. 이용 가능한 특징과 구별의 어려움을 맞추고 신체 출처와 일반 외부 대상의 이력 전이를 비교할 수 있다. 신체에만 추가 매개변수를 넣는다면 근거를 미리 정한다. 그렇지 않으면 «특별한 자기성»이 추가적인 맞춤 자유도의 이름이 된다.

제26장의 강한 원리에는 상호 동역학에 방향을 정한 개입을 했을 때 예측한 질적 조직 변화를 검증하는 단계가 더 필요하다. 보고 모형의 성공은 연구의 출발 자료를 주지만 체험의 존재에 대한 논거를 끝내지는 않는다. 이전 실험에서 함께 가능했던 설명들을 다음 실험이 구별할 때 연구가 나아간다.

## 평균 차이 0,008로부터 말할 수 있는 것

설명을 위해 정한 M1 빼기 M2의 네 차이는 0,018; 0,006; 마이너스 0,004; 0,012다. 합은 0,032이고 평균은 0,008이다. 평균 부호는 이 예에서 이점의 방향을 지지하지만 한 열은 반대 부호이며 평균도 실습 기준 0,01보다 작다. 불확실성 평가 없이 중요한 이점이 확립되었다고도, 그 이점이 없다고도 선언할 수 없다. 다음 행동은 어떤 효과 범위가 자료와 아직 양립하는지에 달려 있다.

### 여기까지 확인한 것

결론은 크기와 정밀도, 조건 범위에 연결된다. 새 메커니즘은 이미 본 곡선을 더 편하게 기술하는 데 그치지 않고 새로운 구별 예측을 내야 한다.

## 실습 35 결과에 따른 결정 작성하기

실습 목표는 보존한 결과 하나에 따라 근거 있는 결정을 내리는 것이다.

사전 절차와 Brier, 쌍별 차이, 구성요소 활성화를 활용한다. 설득력 있는 효과가 없다는 것과 동일성이 입증되었다는 것은 다르다.

예상 소요 시간은 30–40분이다. 중간에 쉬어 가며 해도 된다. 이 시간은 수행 속도의 기준이 아니다.

1. 실행한 기록 하나를 골라 이용 가능한 자료와 사전 예측, 공개된 결과, 새 상태로 몇 단계를 복원한다.
2. 결과를 아는 대조를 확인한다. 이력 모형에서는 단일 채널의 동일성이고 문턱이 있는 구성요소에서는 활성화 로그다.
3. 차이의 크기, 반복 단위, 이용 가능한 불확실성 추정값을 적어라. 없는 구간을 평균의 부호 하나로 대신하지 마라.
4. 일반화 검증과 규칙 구체화, 구별되지 않는 조건의 변경, 특정 버전의 포기 중 결정 하나를 고른다. 다른 결정을 이끌 두 미래 결과를 밝힌다.

근거와 다음 검증을 담은 짧은 결과별 결정을 남겨 둔다.

결과를 본 뒤 매개변수를 바꾸어야 했다면 이전 버전을 보존하고 새 검증을 별도 연구로 명시한다.

탐구를 이어 가려면 새 기억 모형에서 휴지기 길이와 중간 사건 수의 영향을 나눈다. 새 자료를 모으기 전에 예측을 얻어 이전 버전과 비교한다.

해설과 답은 부록 F 제35장 참조, 316쪽

**자기 점검 문항**

1. 모듈이 한 번도 켜지지 않았을 때 모듈이 있는 버전과 없는 버전의 결과가 같다는 것은 무엇을 뜻하는가?

2. M2가 M1보다 낫다고 특별한 자기성 메커니즘이 입증되지 않는 까닭은 무엇인가?

3. 실질적인 수정과 임의적인 맞춤은 어떻게 다른가?

이제 질문과 메커니즘, 방법, 결론이 일치하는 작은 연구를 직접 구성할 수 있다. 마지막 장에서는 재현에서 독자적인 변화로 나아가는 경로를 제안한다.

제36장

# 자신의 연구

책의 경로는 다른 사람이 재현하고 실질적으로 비판할 수 있는 연구로 마무리된다. 작게 시작해도 질문의 명확성과 검증의 힘으로 가치가 정해진다. 출처별 이력은 완결된 계산 예를 준다. 다음 연구에서는 특징과 출처의 상호 규정을 더하여 질적 조직과 체험의 존재에 관한 더 강한 결론에 어떤 근거가 필요한지 보여 준다.

## 학습 결과에서 자신의 과제로

책의 목적은 체험의 설명을 만들고 검증하는 법을 배운 뒤 그 근거로 주관적 체험의 후보가 될 수 있는 계를 설계하는 것이다. 최종 연구에는 구상과 메커니즘, 증거가 연결되어야 한다. 질문이 정확하고 결과가 선택한 메커니즘의 이해를 바꾼다면 구현 규모는 작아도 된다.

아래는 완성된 작은 계산 실습 연구다. 재현하고 확장할 수 있는 논거의 형태를 보여 준다. 수치 결과는 정한 규칙에서 도출되며 사람이나 세포 배양에서 얻은 값이 아니다. 이어 질적 조직과 체험의 존재라는 더 강한 가설을 향한 다음 연구를 세운다.

## 출처별로 분리한 이력의 학습용 연구

출처가 돌아왔을 때 같은 새 증거가 원인 평가에 다르게 작용하도록 하려면 이력을 따로 저장하는 것으로 충분할까? 추가로 이 차이에 특별한 자기성 메커니즘이 필요한지, 대상 동일성의 일반 기억으로도 재현되는지 묻는다.

방법에는 관찰 네 개를 정한다. A 지지, B 반증, 다시 A 지지, 다시 B 지지다. 모형은 출처 표지를 이용할 수 있으며 이 실습에서는 오류가 없다. 지지 특징의 확률은 공통 원인에서 0,75이고 별개 원인에서 0,25다. 처음 공통 원인의 확률은 0,5다.

특징 처리 전에 유지 계수 0,8로 이전 믿음을 0,5 쪽으로 이동시킨다. 이어 제25장의 갱신 규칙을 적용한다. M0는 매번 0,5에서 시작한다. M1은 한 믿음을 모든 사건에 걸쳐 이월한다.

M2는 출처별 믿음을 보존하며 다음에 그 기록을 사용할 때 유지 계수를 적용한다. M1I는 같은 출처별 기록을 일반 대상 추적기로 구현한다.

정해진 열의 결정론적 분석이다. 열의 모집단에 대한 평균 효과 추정이나 참여자의 통계 표본은 없다. 손으로 또는 공학 실습의 source_history_example() 함수로 각 수를 검증할 수 있다. 출력은 현재 관찰 뒤 공통 원인의 확률이며 관찰 자체의 예측이 아니다.

## 결과와 그 설명

| 단계와 사건 | M0 | M1 | M2 | M1I |
|---|---|---|---|---|
| 1 A 지지 | 0,750000 | 0,750000 | 0,750000 | 0,750000 |
| 2 B 반증 | 0,250000 | 0,437500 | 0,250000 | 0,250000 |
| 3 A 지지 | 0,750000 | 0,710526 | 0,875000 | 0,875000 |
| 4 B 지지 | 0,750000 | 0,858108 | 0,562500 | 0,562500 |

셋째 단계에서 M2는 A의 이력으로 돌아간다. 이전 평가 0,75는 유지 계수 적용 뒤 사전확률 0,7이 된다. 지지 증거로 0,875가 된다. M1은 B의 반증 사건을 거쳐 왔으므로 약 0,710526을 낸다. 현재 특징은 같아도 이력 이월 규칙 때문에 차이가 생겼다.

넷째 단계가 메커니즘의 이해에 특히 유용하다. M1의 공통 이력에는 이제 지지가 더 많으므로 새 지지 증거 뒤 127/148, 약 0,858108이 나온다. M2는 B의 이전 반증 이력으로 돌아가 0,5625만 얻는다. 출처별 기억은 확신만 높이는 것이 아니다. 다른 근거들을 유지하므로 공통 기억보다 평가를 낮출 수도 있다.

네 사건 모두에서 M1I는 M2와 같다. 우연히 자료가 부족해서가 아니라 규칙이 같아서다. 대상 사전의 이름을 «자기 모형»으로 바꾸어도 메커니즘이 더해지지는 않는다. 계산 결과는 이력 분리의 역할을 확인하면서 해석의 범위도 제한한다.

## 작은 연구의 대조와 결론

표지와 함께 이력을 옮겼다면 A와 B의 이름을 바꾸어도 수치열은 같아야 한다. 표지를 모두 하나로 바꾸면 M1과 M2가 일치한다. 유지 계수를 영으로 두면 모든 버전이 지지 특징 뒤 0,75, 반증 특징 뒤 0,25를 낸다. 이 귀결들로 구현이 주장한 모형에 맞는지 검증할 수 있다.

결론은 출처별 이력 저장으로 설명한 출처 복귀의 계산 효과를 재현하기에 충분하다는 것이다. 모든 환경에서 이점이 있다고 확립하거나 특별한 자기관련성과 일반 동일성 추적을 구별한 예는 아니다. 독립 자료로 검증할 정확한 메커니즘과 다음에 다룰 명확한 대안을 준다.

한계에도 내용이 있다. 오류 없는 표지는 과제를 쉽게 하며 실제 계는 관찰로 출처를 인식해야 한다. 시간 규칙은 물리적 지속 시간이 아닌 기록 사용 횟수로 정했다. 원인에 관한 믿음은 아직 지각의 질적 구조와 연결되지 않았다. 각 한계는 확인된 결과를 취소하는 대신 다음 가설을 가리킨다.

## 내용의 상호 규정을 다루는 다음 연구

출처가 모호한 특징의 해석에 영향을 주고 해석이 출처 평가를 바꾸며 계가 추가 검증을 고를 수 있게 과제를 넓힌다. 제26장 조직 일부의 구체적인 후보가 된다. 먼저 따로 바꿀 수 있는 내용의 두 성질, 특징 평가와 출처 평가를 둔다. 셋째 변수는 검증의 기대 이득을 정한다.

실행 전에 방향을 정한 개입을 기술한다. 감각 입력을 유지하면서 출처 평가를 강제로 바꾸면 특징 해석과 검증 선택이 달라져야 한다. 특징 표상을 바꾸면 이후 출처 평가가 예측 가능하게 달라져야 한다. 되먹임 하나를 끊으면 처음 관찰의 이용 가능성은 남으면서 해당 영향 방향이 손상되어야 한다. 효과 크기와 시간 구간은 구현의 구체적 규칙으로 정한다.

이제 설명을 구별할 설계가 생긴다. 전체 계를 같은 관찰과 이력 기록을 갖추되 영향은 한 방향뿐인 계와 비교한다. 별도 대조에서는 구성요소 사이 인과 연결 없이 평소 출력을 재생한다. 과제 성공뿐 아니라 개입 응답의 행렬도 검증한다. 충분히 강한 대안 제어기도 필요하다. 무력하도록 만든 재생 기록을 이겨서는 메커니즘의 고유성을 설명할 수 없다.

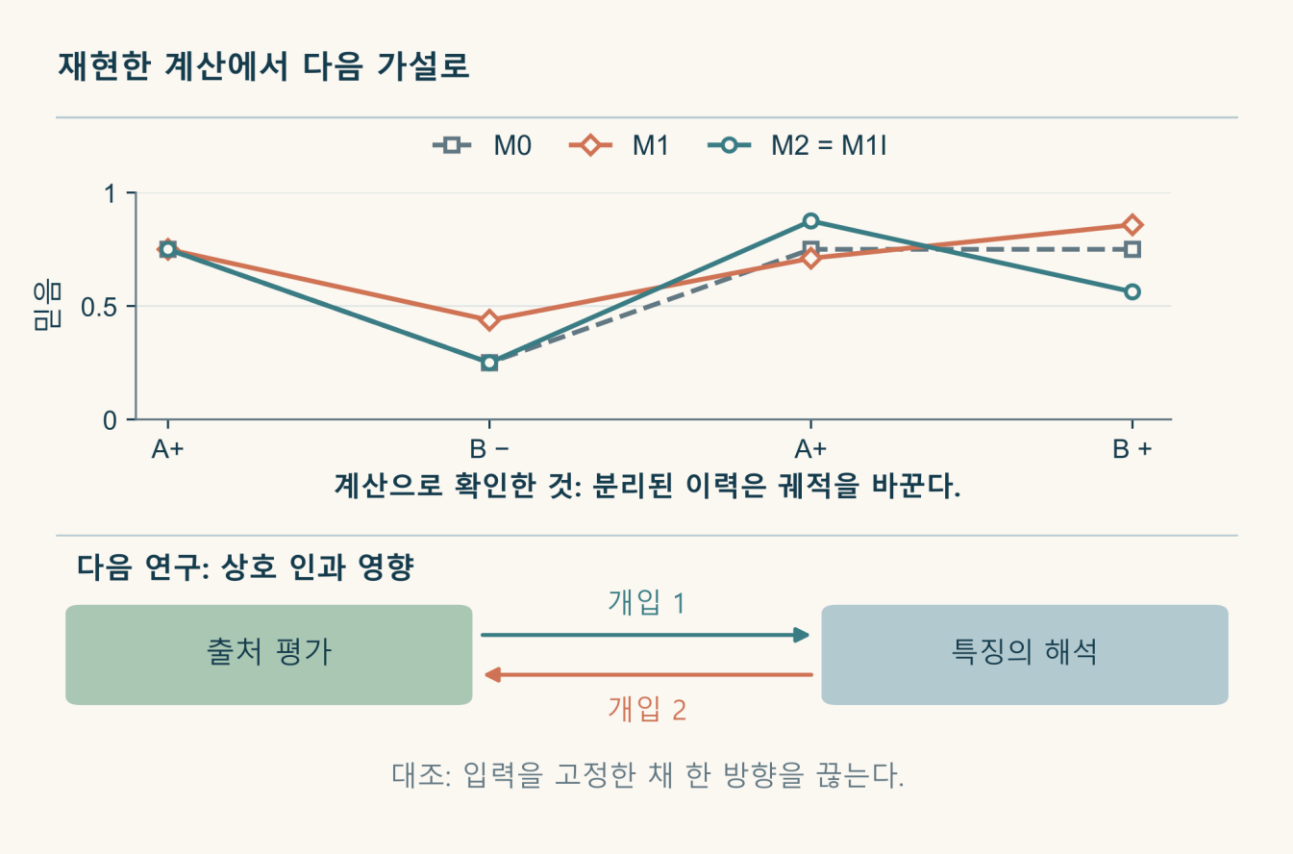


그림 88. 완료한 계산과 다음 인과 연구 설계. 위는 네 사건 A+, B−, A+, B+의 정확한 학습 계산이다. 넷째 단계에서 M0는 0,75, M1은 127/148 (약 0,858108), M2와 그에 일치하는 M1I는 0,5625를 준다. 아래는 아직 수행하지 않은 설계다. 출처의 평가와 특징 해석에 개입했을 때 반대쪽 구성요소에 나타날 결과를 미리 정해야 한다. 단방향 대조는 특정 인과 연결을 검증한다. 계산 자체가 제26장의 강한 가설 전체를 구현하거나 체험의 존재를 확립하지는 않는다.

## 설계를 체험에 관한 질문에 더 가깝게 만드는 방법

제26장의 원리는 질적 조직을 구별과 대안의 안정적인 관계에 연결한다. 그러므로 미래 모형은 정답만이 아니라 유사성과 혼동, 내용 변화의 구조를 예측해야 한다. 예를 들어 출처에 개입하면 모호한 내용 사이 일부 관계는 달라지고 미리 정한 다른 관계들은 남아야 한다. 전체 정확도의 향상보다 더 강한 요구다.

인간과 비교하려면 체험 유사성 평가와 시간적 변화, 출처 구별, 난이도 통제에 대한 독립 자료가 필요하다. 가정한 메커니즘을 대상들의 공통 기억으로 설명하는 대안, 체험은 그대로이고 결정만 달라진다는 대안과 비교한다. 확인 자료를 모으기 전에 경쟁 설명들에도 예측이 있어야 한다.

체험의 존재에 관한 강한 주장은 가설로 남는다. 실제로 실현된 특정 조직이 최소한의 주관적 과정에 충분하다는 가설이다. 구현과 기능 검증은 제안한 조건의 충족 여부를 검증한다. 충분성에 근거를 제시하려면 체험 자료와의 연결, 물리적 구현의 구체화, 대안 분석이 필요하다. 이렇게 주체를 만드는 과제는 직접적인 연구 목표로 유지되며, 각 단계에는 그 자체로 검증할 수 있는 결과가 있다.

**각 주장에는 고유한 검증이 필요하다**

| 주장 | 검증 행동 | 내용에 맞는 대조 |
| --- | --- | --- |
| 출처의 이력이 응답에 영향을 준다 | 바로 그 기록을 제거하거나 대체 | 같은 이력의 공통 추적기 |
| 두 영향 방향이 있다 | p와 q에 따로 개입 | 두 단방향 순환 구조 |
| 지표가 예측을 개선한다 | 완전히 남겨 둔 집단을 검증 | 새 특징을 더하지 않은 기존 활동 |
| 조직이 체험과 연결된다 | 원리와 예측이 갈리는 귀결을 구체화 | 같은 출력을 갖는 다른 조직 |

대조 비교는 특정 결론을 제한하며 모든 질문을 한꺼번에 제한하지는 않는다.

그림 89. 주장에서 검증과 대조로. 설계를 마무리하는 학습용 행렬이다. 각 행은 주장을 구체적인 행동 및 강한 대조와 연결한다. 마지막 행은 체험에 관한 이론적 논증의 독립성을 유지한다. 위의 기능 검증들이 그 논증을 자동으로 대신하지는 않는다.

## 최종 작업에 포함할 것

질문과 근거, 메커니즘, 방법, 결과, 논의가 이어지는 글을 준비한다. 디지털 구현에는 실행 코드와 환경 설명, 매개변수, 보존 자료, 표를 다시 만드는 방법을 첨부한다. 생물학적 연구에서는 공개 자료 분석과 디지털 모형, 전문적인 실험 계획이 가능하다. 이미 수행한 작업의 범위를 명확히 밝혀야 한다.

유력한 경쟁 설명 하나와 인과적 또는 구별 검증 하나, 결과를 아는 대조 하나를 포함한다. 체험에 관해 주장한다면 대응 원리를 따로 제시한다. 어느 명제를 검증했고 무엇을 가정하며 어디가 미규정인지 밝힌다. 저자의 제안을 하려면 가까운 발표 생각들을 확인해야 한다. 직접 정식화했다는 것만으로 과학적 선취권이 생기지는 않는다.

구별하지 못한 결과도 완결된 최종 연구가 될 수 있다. 왜 생겼으며 어떤 다음 검증이 불확실성을 없앨지 보여야 한다. 모듈의 존재와 멋진 설명, 설득력 있는 자기 보고가 역할의 입증을 대신하지 않는다. 불필요한 복잡성을 덜고 다음 후보를 더 정확하게 설계할 수 있다면 특히 유용한 결론이다.

## 계산을 짧은 논문으로 만드는 방법

예시의 질문은 출처가 돌아올 때 개별 이력이 갱신을 바꾸는가다. 방법에는 네 사건과 신뢰도, 유지 계수, 기억 규칙을 정한다. 결과는 모든 궤적과 M2 및 일반 추적기의 일치를 제시한다. 논의는 효과를 기록의 분리와 연결하고 표지와 시간, 체험과의 관계에 대한 한계를 유지한다. 다음 단계는 가정 하나를 바꾸어 새 예측을 얻는다. 이 구조로 계산을 재현하고 설명의 전이를 따로 비판할 수 있다.

### 여기까지 확인한 것

완료한 기능적 결과와 열린 연구 목표가 있다. 다음 설계는 이 경계를 지워서는 안 된다. 명확한 새 요구와 그에 맞는 검증을 더해야 한다.

## 실습 36 자신의 논문 완성하기

실습 목표는 검증 가능한 메커니즘 변화 하나를 포함한 재현 가능한 독자 연구를 준비하는 것이다.

필요한 것: 메커니즘 명세, 강한 경쟁 모형, 사전 계획, 결과에 따른 결정. 표현의 새로움만으로 과학적 선취권이 확립되지는 않는다.

시간 참고: 예제를 재현한 뒤 여러 번의 별도 세션. 쉬어 가며 할 수 있다. 속도 기준이 아니다.

1. M0-M2의 네 사건 계산과 공통 추적기와의 일치를 재현하라. 코드, 매개변수, 원래 수치를 보존하라.
2. 망각의 시간, 출처 인식 오류, 특징과 출처의 상호 영향 중 하나를 바꾸어라. 실행 전에 원래 모형과의 차이를 적어라.

3. 강한 경쟁 모형, 구별 검증, 정답을 아는 대조를 준비하라. 검증을 수행하고 모든 결과를 보존하라.

4. 질문, 근거, 메커니즘, 방법, 결과, 논의를 갖춘 글을 써라. 논의에서는 확립된 사실, 남은 불확실성, 다음 체험 질문을 구분하라.

보존할 것: 연구 글, 실행 가능한 예, 자료, 모형 버전의 명확한 이력.

힌트: 왜 생겼으며 어떤 새 조건이 설명들을 구별할지 분명하다면 구별하지 못한 결과도 허용된다. 이 분석을 시스템의 그럴듯한 자기보고로 대신하지 마라.

연구 확장: 강한 가설에는 두 방향의 개입을 정하고 내용들 사이 관계의 변화를 미리 도출하라. 체험 자료와의 비교, 새로운 물리적 구현으로의 일반화는 프로그램의 별도 단계로 남는다.

해설과 답은 부록 F 제36장 참조, 316쪽

**자기 점검 문항**

1. 제시한 네 사건의 계산은 무엇을 입증했는가?
2. 강한 설계에 두 방향의 개입이 필요한 이유는 무엇인가?
3. 최종 연구에 체험에 관한 상당한 불확실성이 남을 수 있는가?

이후 작업은 디지털, 비교, 생물학적 경로로 나아갈 수 있다. 구체적인 질문, 변경의 전체 이력, 설명을 재검토하게 할 결과를 얻을 가능성을 보존하라.

# 부록 A 연구용 용어집

이 용어집은 책에서 사용하는 말의 뜻을 고정한다. 다른 저자를 읽을 때에는 정의를 비교하라. 같은 말이 서로 다른 연구 과제를 가리킬 수 있다.

퀄리아. 무언가가 어떻게 지각되거나 느껴지는지와 같은 체험의 질적 측면이다. 이 잠정적 정의에는 기술의 무오류성, 언어로 표현할 수 없음, 물리적 세계로부터의 독립성을 미리 포함하지 않는다.

주관적 체험. 상태에서 체험되는 측면이다. 체험이 있는지에 관한 질문은 시스템이 어떤 정보를 구별하거나 보고하는지에 관한 질문과 다르다.

의식. 여러 용법을 가진 말이다. 이 책에서는 체험의 존재, 내용, 각성 상태, 이후의 이용을 위한 정보 접근성 중 무엇을 논하는지 매번 구체화한다.

주체. 주관적 체험을 지닐 것으로 가정하는 기반이다. 이 정의는 인간의 생애, 언어 인터페이스, 미리 선택한 종에 속할 것을 요구하지 않는다.

자기의식. 시스템이 자신과 맺는 여러 관계 형태다. 신체 소유감, 행위주체감, 자신의 이력에 관한 표상, 자신에 대한 명시적 판단을 따로 연구한다.

성찰. 자신의 판단, 근거, 행동 방식을 살펴보고 평가하며 재검토할 수 있는 일이다. 성공률의 단순한 수치 예측만으로 이 능력들을 모두 구현한 것은 아니다.

메타인지. 인지 과정의 점검과 조절이다. 구체적인 과제에서는 무엇을 평가하며 그 평가가 행동에 어떻게 영향을 주는지 밝혀야 한다.

자기관찰. 자신의 상태를 알아차리거나 기술하기 위해 그 상태에 주의를 돌리는 일이다. 접근의 성격과 보고의 신뢰도는 연구해야 한다.

기능적 자기관찰. 시스템이 자기 내부 상태의 정보를 이용하는 연구 대상 능력이다. 이 말 자체로 그 상태들이 체험되는지에 관한 질문이 해결되지는 않는다.

신체 소유감. 특정 신체 부분이 자신에게 속한다는 체험이나 평가다. 모형에서는 이 특성과 관련된 측정을 명시해야 한다.

행위주체감. 자신, 명령, 그 결과 사이의 관계다. 제어의 성공, 인과 영향, 체험되는 행위주체감을 하나의 지표로 여겨서는 안 된다.

자기 모형. 자신의 상태, 능력, 경계, 이력에 관한 표상이다. 구체적인 모형의 내용과 인과적 역할을 밝혀야 한다.

세계 모형. 숨은 상태를 평가하거나 사건을 예측할 수 있게 하는 외부 상황의 표상이다.

숨은 상태. 환경이나 연구자는 알지만 해당 과제에서 행위자가 직접 접근할 수 없는 변수다.

관찰. 시스템이 정해진 인터페이스로 실제로 받는 정보다. 환경 코드에 변수가 존재한다고 그 변수가 행위자의 관찰이 되는 것은 아니다.

믿음. 확률 모형에서 전제와 이용 가능한 정보를 고려한 가능한 상태들의 평가다. 이런 기술적 용법이 프로그램에 체험되는 확신을 부여하지는 않는다.

사전확률. 다루는 증거를 반영하기 전의 평가다. 이전 학습 결과가 포함될 수 있다.

사후확률. 선택한 모형 안에서 새 증거를 바탕으로 갱신한 뒤의 평가다.

가능도. 가정한 상태나 모형에서 관측 자료가 나올 확률이다. 관찰 뒤의 가설 자체 확률과 구별해야 한다.

불확실성. 상태, 매개변수, 설명에 관한 지식의 한계다. 불확실성의 원인이 다르면 필요한 행동도 다를 수 있다.

확률보정. 정해진 사례 집단에서 확률 예측과 결과 빈도가 일치하는 성질이다. 평가는 자료와 집단화 방식에 달려 있다.

브라이어 점수 (Brier score). 이진 결과의 확률 예측과 영 또는 일로 기록한 결과 차이의 평균제곱이다. 확률보정뿐 아니라 예측의 질을 나타낸다.

효용. 결과와 비용을 합치는 정해진 방식이다. 첫 실습에서는 성공에서 추가 관찰 비용을 뺀 것이다.

정보의 가치. 목표와 채택한 모형을 고려할 때 추가 정보로 기대되는 결정의 개선이다. 불확실성을 줄이는 일이 언제나 똑같이 유용하지는 않다.

되먹임. 결과에 관한 정보가 시스템의 이후 작동으로 돌아오는 것이다. 어떤 정보가 어떤 지연을 거쳐 돌아오는지 명시해야 한다.

재귀 처리. 영향이 되돌아오는 처리다. 결과가 그 결과의 형성에 이미 참여한 요소들의 작동을 바꾼다. 구체적인 계산 조직과 물리적 조직을 따로 기술해야 한다.

작업공간. 여러 과정이 내용을 선택하고 함께 이용하는 조직이다. 단순한 변수 표로 개별 기능을 모형화할 수는 있지만 과학 이론 전체가 자동으로 구현되지는 않는다.

요소 제거 검증. 정해진 구현과 과제에서 구성요소의 기여를 검증하기 위해 그 요소를 제거하거나 차단하는 것이다.

인과 개입. 변수나 연결의 영향을 검증하기 위해 의도적으로 바꾸는 일이다. 부수적 변화와 다른 과정들이 개입에 접근할 수 있는지도 분석에 포함된다.

모의 개입. 연구하는 중요한 변수는 유지하면서 영향의 다른 측면들을 재현하는 대조 절차다.

상관 지표. 연구 현상과 통계적으로 연결된 특징이다. 연결 자체가 필요성이나 충분성을 확립하지는 않는다.

필요조건. 명시한 범위에서 그 조건 없이는 가정한 현상이 생기지 않는 조건이다. 개입 뒤 효과가 사라졌다면 개입 자체의 특이성을 검증해야 한다.

충분조건. 명시한 전제 아래에서 충족되면 현상을 보장해야 하는 조건이다. 현상을 얻는 유일한 가능한 방식이 아닐 수 있다.

불변량. 정해진 변환의 부류에서 보존되는 성질이다. 변환의 부류와 측정 방식을 미리 밝혀야 한다.

식별 가능성. 이용 가능한 자료로 모형이나 매개변수를 구별할 수 있는 성질이다. 수학적으로 같은 알고리즘에 서로 다른 이름을 붙여도 식별 가능성이 생기지는 않는다.

과적합. 사용한 자료의 특징에 맞추어 조정한 결과 새 사례에서 기대한 성능을 내지 못하는 것이다.

검증용 표본. 모형의 개발이나 추정 뒤에 정해진 검증에 사용하는 자료다. 이 자료로 다시 조정하면 연구에서의 역할이 달라진다.

사전등록. 해당 단계 전에 질문, 비교, 지표, 분석의 기술을 보존하는 일이다. 계획 변경이 드러나게 하지만 최초 아이디어의 질을 보장하지는 않는다.

관찰 단위. 측정이 해당하는 대상이나 사례다. 실험을 독립적으로 반복하는 단위와 구별해야 한다.

의사반복. 의존적인 측정을 독립 반복처럼 다루는 오류다. 한 시료의 많은 에피소드가 여러 독립 시료를 대신하지는 않는다.

부트스트랩. 통계량의 변동성을 평가하기 위해 기존 자료에서 반복 표본을 만드는 방법이다. 재표집 방식은 관찰의 의존 구조와 맞아야 한다.

프리프린트. 학술지의 정식 게재 전이거나 정식 게재와 별개로 공개된 연구 글의 버전이다. 지위와 버전을 따로 확인해야 한다.

신경세포 배양. 신경세포를 배양하는 실험 시스템이다. 조직은 구체적인 모형과 조건에 달려 있다.

오가노이드. 조직의 구성이나 발달 특징 일부를 재현하는 삼차원 세포 모형이다. 오가노이드를 자동으로 온전한 기관이나 생물 개체로 여겨서는 안 된다.

혼성 시스템. 생물학적 구성요소와 기술적 구성요소가 인과적으로 상호작용하는 시스템이다. 분석은 각 요소의 기여와 선택한 전체 경계를 보여 주어야 한다.

체험에 관한 결론. 자료와 이론적 전제를 바탕으로 주관적 체험의 존재나 성격에 관해 하는 주장이다.

대응 원리. 시스템 조직의 기술과 체험의 기술 사이에 제안한 연결이다. 저자의 가설에서는 독립적인 논증과 적용 범위의 정의가 필요하다.

가능도는 선택한 가설에서 특정 자료가 나올 확률이다. 자료를 얻은 뒤 가설 자체의 확률과 같지 않다. 후자에는 사전확률과 정규화가 필요하다.

조건부 독립은 원인을 정했을 때 새 사건의 확률이 이미 아는 사건들에 의존하지 않는다는 가정이다. 신호를 복사하면 보통 이 가정을 위반한다.

브라이어 오차는 이진 관찰의 확률 예측과 영 또는 일인 결과 차이의 평균제곱이다. 예측은 평가할 관찰이 나타나기 전에 기록한다.

모형 복원은 생성 메커니즘을 아는 인공 자료에서 메커니즘의 선택을 검증하는 일이다. 복원 오류는 구별 절차의 한계를 보여 준다.

검증용으로 남긴 집단은 특정 일반화 검증을 위해 학습에서 통째로 제외한 오가노이드나 배치의 모든 기록이다.

STTC는 Spike time tiling coefficient, 곧 신경 발화의 시간적 일치 지표다. 실습 분석에서는 15 ms 시간 구간에서 행렬의 유한한 비대각 원소들의 평균 절댓값을 사용한다. 입증된 인과 연결은 아니다.

충분조건을 충족하면 채택한 주장 안에서 가정한 결론이 따른다. 그 조건이 충족되지 않았다고 같은 결론의 다른 근거를 배제할 수는 없다.

# 부록 B 실습의 실행과 구성

## 준비

qualia_mira_practice.zip을 별도 폴더에 풀어라. 안에는 프로그램, 지침, 보존한 결과가 있다. README.md부터 읽어라. Python 파일들은 이웃 파일이 필요하므로 압축파일 보기 창에서 바로 실행하지 마라. 실습 2–8에는 agent_lab.py와 mechanisms_lab.py가 같은 폴더에 있어야 한다. 새 결과를 다른 이름으로 저장하면 원래 예제를 비교용으로 남길 수 있다.

이 폴더에서 터미널을 연다. Windows에서는 탐색기 메뉴를 사용하거나 cd 명령으로 이동할 수 있다. macOS와 Linux에서는 일반 터미널을 사용한다. Python 3.10 이상인지 확인한다. 설치하지 않았다면 해당 운영체제의 공식 python.org 배포판을 사용한다. 명령 이름은 설치에 따라 다르다.

```
python3 --version
python3 agent_lab.py --self-test
python3 mechanisms_lab.py --self-test
```

Windows에서 첫 명령은 보통 py -3 --version이다. 다른 명령의 python3도 같은 방식으로 바꾼다. 명령을 찾을 수 없다는 메시지는 실험 결과가 아닌 인터프리터 실행 오류다. agent_lab.py가 없다면 현재 폴더를 확인한다. mechanisms_lab.py가 agent_lab을 가져오지 못한다면 두 파일을 함께 풀었는지 확인한다.

처음에는 공통 명령 하나로 여덟 실습과 쌍별 분석을 모두 실행할 수 있다. 이 명령은 results_...라는 별도 폴더에 결과와 프로그램의 매개변수 및 체크섬을 기록한 run_manifest.json을 만든다.

```
python3 run_practice.py
```

--lab 매개변수로 실습 하나를 고를 수 있다. 예를 들어 python3 run_practice.py --lab 6이다. Windows에서는 python3 대신 py -3을 쓴다. 새 실험을 보존하지 않고 검사만 실행하는 명령은 python3 run_practice.py --self-test다.

## 결과 파일을 읽는 방법

JSON은 이름이 있는 필드와 목록으로 자료를 저장한다. 형식 전체를 한꺼번에 익힐 필요는 없다. lab01_results.json의 summary에는 모드의 전체 결과가, per_run에는 개별 시나리오 결과가 있다. mechanisms_results.json에서는 labs 안에 실습 2부터 8까지가 들어 있다. 필드의 자세한 설명은 압축 묶음의 지침과 제28–30장에 있다.

success_rate와 utility_per_step은 성공 비율과 단계당 효용이다. 성공 비율 0,8은 팔십 퍼센트다. queries에는 추가 요청 수가 있고 brier_score는 확률 예측의 품질 지표다. 표를 비교할 때는 단계 수와 매개변수, 평균 방식이 같은지 확인한다.

먼저 원결과를 재현한다. 이어 매개변수 하나만 바꾼 새 파일을 만든다. 명령과 사전 예상을 함께 남긴다. 처음 seed로 검증한 뒤 다른 값도 사용한다. 난수 시드로 특정 시나리오를 반복할 수 있지만 시나리오 하나가 충분한 연구가 되는 것은 아니다.

## 여덟 실습

| 실습 | 프로그램 | 주된 내용 |
|---|---|---|
| 1 | agent_lab.py | 세계 기억과 행동 모형, 관찰의 가치 |
| 2 | mechanisms_lab.py --lab 2 | 표상의 기하와 불변량 |
| 3 | mechanisms_lab.py --lab 3 | 반복 처리와 펼친 구조의 대조 |
| 4 | mechanisms_lab.py --lab 4 | 제한된 선택과 전역 접근 |
| 5 | mechanisms_lab.py --lab 5 | M0과 M1, M2, 일반 추적기, 출처 지연 |
| 6 | mechanisms_lab.py --lab 6 | 설계된 내부 접근과 개입 |
| 7 | mechanisms_lab.py --lab 7 | 과거 오류에 따른 전략 선택 |
| 8 | mechanisms_lab.py --lab 8 | 통합 후보와 구성요소 기여의 검증 |

실습 2–8은 --lab 없이 한 번에 실행할 수 있다. 아래 명령은 새 결과 파일을 만든다. 외부 서비스로 아무것도 보내지 않으며 코드를 바꾸지 않는다.

```
python3 mechanisms_lab.py --json my_mechanisms.json
```

프로그램들은 작은 메커니즘을 살피기 위한 것이며 구조는 본문에서 설명했다. 함수 이름을 한계에 모았다고 구성요소의 상호작용을 따로 검증한 것은 아니다. 여덟째 실습에는 기본 조건에서 추가 메커니즘이 이점을 주지 않는 비교를 의도적으로 포함했다.

## 추가 단계별 모형

teaching_lab.py는 기존 여덟 실습을 보완한다. Python 표준 라이브러리를 쓰며 제31장 분석에서는 같은 폴더의 bio_features.csv를 읽는다. 압축을 푼 묶음의 practice 폴더에서 명령을 실행한다. python은 이용 가능한 Python 3 인터프리터를 가리킨다. 설치에 따라 python3 또는 py -3으로 실행한다. 기본 검사와 실습 모형에는 외부 패키지가 필요하지 않다.

```
python teaching_lab.py --self-test
python teaching_lab.py --chapter 13
python teaching_lab.py --chapter 25
python teaching_lab.py --chapter 31
python teaching_lab.py --chapter 33
python teaching_lab.py --chapter 34
```

각 실행은 규칙이나 수치 결과를 JSON으로 보존한다. 제31장은 준비한 표로 큰 원자료 없이도 주분석을 재현할 수 있다. extract_bio_features.py는 공식 파일 [47]의 체크섬을 확인하고 제외 이유를 명시하며 특징 추출을 반복한다. 이 별도 단계에는 NumPy가 필요하다. 세부 내용과 출처 표기는 실습 README에 있다.

# 부록 C 결과의 쌍별 비교

첫 실습의 네 방식은 모두 공통 무작위 시나리오를 사용한다. 따라서 자연스러운 비교 단위는 한 시나리오 안의 두 방식 지표 차이다. 모든 단계를 먼저 합치고 독립 반복이라고 선언하면 정밀도를 근거 없이 높게 느끼게 된다. 이웃 단계들은 표적 상태와 행위자의 학습으로 연결되기 때문이다.

analyze_lab01.py는 시나리오 서른 개에서 전체 모드와 각 대조 모드의 효용 차이를 계산한다. 이어 차이값 전체를 복원추출로 사천 번 재표집하여 백분위 구간을 만든다. 이 방법은 채택한 반복 구조 아래의 통계량 변동을 기술한다. 모든 표본에서 정확히 구십오 퍼센트의 포함률을 보장하지는 않는다.

python3 analyze_lab01.py lab01_results.json --json paired_analysis.json

보존한 분석에서 전체 모드의 단계당 평균 효용 이점은 기억 없는 버전에 대해 0,08516, 고정 행동 모형에 대해 0,37772, 추가 요청 없는 버전에 대해 0,08280이다. 각각의 근사 구간은 0,07975부터 0,09080, 0,37133부터 0,38434, 0,07583부터 0,08997까지다. 첨부 자료와 분석 알고리즘으로 얻은 값들이다.

구간은 특정 비교와 지표, 인공 환경에 해당한다. 개별 구간 여러 개가 자동으로 구십오 퍼센트의 동시 보장을 주지는 않는다. 많은 지표를 본 뒤 유리한 것만 고르면 그 결정의 탐색적 성격을 반영해야 한다. 본격 연구에서는 주기준과 비교 집합을 미리 기술한다.

독자 과제에서는 효용을 주지표로 고르고 정확도와 비용으로 차이의 메커니즘을 설명한다. 먼저 효과 방향을 예상한다. 이어 개별 시나리오의 차이를 보고 안정적인지, 특이 사례가 있는지, 특정 사건열과 관련되는지 살핀다. 사전 근거 없이 불편한 시나리오를 지우지 않는다.

평균 차이가 작으면 메커니즘이 과제에서 실제로 비슷하거나 자료가 부족하거나 지표가 예상 성질을 잘 반영하지 않는 등 여러 설명이 가능하다. 넓은 구간이 설명을 자동 선택하지는 않는다. 다음 실험은 새 과제와 독립 반복 추가, 더 적절한 측정 등 구체적인 불확실성의 원인에 연결되어야 한다.

# 부록 D 독립적인 작업을 위한 양식

## 체험한 에피소드의 기록

상황과 사건 시점을 적는다. 사례가 진행되는 동안인지, 직후인지, 나중인지 기록 시점도 남긴다. 알아차린 것을 기술하고 불분명한 곳은 따로 남긴다. 이어 자기 설명과 적어도 하나의 대안을 적는다. 어떤 관찰이 결론을 구체화할지 묻는 것으로 마친다. 자유로운 산문으로 채워도 된다. 발견하지 못한 세부는 예시에 맞추어 꾸며 넣지 않고 빈 채로 둔다.

## 과학 논문 기록 카드

저자와 연도, 제목, 링크를 적는다. 연구 질문을 한 문장으로 밝힌다. 참여자나 시료, 조건, 측정, 비교를 명시한다. 결과와 저자 결론을 따로 기술한다. 둘을 연결하는 가정과 가능한 대안, 다음 구별 검증을 적는다. 이론 논문이라면 표본과 측정 대신 출발 명제와 도출하는 귀결을 기술한다.

## 모형 기록 카드

설명 현상과 적용 범위를 밝힌다. 이용 가능한 입력과 숨은 상태, 기억, 갱신 규칙을 적는다. 관찰 응답을 만드는 방법을 설명하고 매개변수와 결정 방법을 명시한다. 가까운 경쟁 설명과 예측이 다른 조건으로 마친다. 체험 가설에는 대응 원리와 이를 지지하는 논거를 더한다.

## 프로그램 실험의 기록

질문, 코드 버전, 명령, 매개변수, seed, 주지표, 예상 차이를 보존하라. 실행 뒤 결과, 자료 경로, 이탈의 설명을 더한다. 자료를 보기 전과 뒤에 각각 어떤 결정을 했는지 표시하라. 다음 검증 실행은 조정 단계와 분리해야 한다.

## 최종 설계의 명세서

설명 대상과 제안한 메커니즘, 구현 방식, 이용 가능한 자료, 개입 계획, 비교 가능한 대안을 밝힌다. 성공 기준과 가설을 바꾸게 할 결과를 설명한다. 이어 실제로 수행한 검증을 적는다.

결론에는 기능적 결과와 이론적 해석, 체험에 관해 가능한 결론의 근거를 따로 정한다.

문단들이 과제의 차이를 유지하며 서로 일치해야 한다.

# 부록 E 삽화 색인

그림은 관련 설명 곁에 배치했다. 그림 설명에는 이미지의 성격과 수치의 출처를 표시했다. 도식은 살펴보는 관계를 보여 주고, 삽화는 체험 사례나 연구 대상을 구체적으로 떠올리는 데 도움을 준다. 색인의 번호를 선택하면 해당 그림으로 이동한다.

1. 한 사건을 기술하는 세 영역 ..... 14
2. 관찰의 일부인 기록 시각 ..... 15
3. 한 사례의 표면 비교 ..... 20
4. 재료의 이름과 접촉감의 유사성 ..... 22
5. 쌍별 평정에서 유사성 지도로 ..... 23
6. 움직이지 않는 손의 신체 소유감과 외부 대상의 제어 ..... 30
7. 한 사례에서 자신과 맺는 네 가지 관계 ..... 31
8. 과제 성적이 같을 때의 확신도 ..... 38
9. 반례로 검증하는 논리적 추론 ..... 297
10. 인과 영향과 공통 원인 ..... 43
11. 램프 구조의 세 연구 역할 ..... 45
12. 연구의 참가자와 관찰자 ..... 49
13. 서로 다른 자료의 시간 정렬 ..... 52
14. 절차의 조건이 되는 질문 시점 ..... 54
15. 자료와 조건 순서의 배정 ..... 58
16. 총 준비 시간과 그 배분 ..... 59
17. 효과 크기와 추정 정밀도 ..... 64
18. 서로 다른 자료 수준인 응답과 참여자 ..... 66
19. 누적과 초기화의 동역학 ..... 71
20. 입력의 배분과 상태 유지의 차단 ..... 74
21. 맥락을 이용하는 두 방식 ..... 79
22. 이용 채널의 선택적 차단 ..... 80
23. 손과 도구, 관찰된 결과 ..... 84
24. 인과 설명의 순차적 갱신 ..... 85
25. 명령을 바꾸어 원인을 구별하기 ..... 88
26. 되먹임 전후의 정보 ..... 92
27. 두 모형의 예측과 오차 ..... 94
28. 재귀 처리와 전역 접근 ..... 100

29. 두 메커니즘의 독립적 차단 ..... 104
30. 서로 다른 두 설명 대상 ..... 108
31. 외부 로그와 내부 이용 ..... 111
32. 사전 평가에서 갱신된 확률로 ..... 114
33. 분리된 세계 모형과 행동 모형 ..... 116
34. 보고와 측정의 경계 ..... 122
35. 유의성이 없다는 것과 정확한 영이 다른 이유 ..... 124
36. 한 학습 메커니즘의 세 궤적 ..... 128
37. 평균과 개별 에피소드의 동역학 ..... 130
38. 일상적 상호작용에서의 발달 ..... 134
39. 시각적 자기 인식 과제의 분해 ..... 136
40. 하나의 서열 없는 발달 프로필 ..... 138
41. 환경은 서로 다른 행동 과제를 만든다 ..... 141
42. 복잡성의 서열 대신 기억의 상충관계 ..... 142
43. 같은 특징 분포의 두 역사 ..... 143
44. 존재의 설명과 질의 설명 ..... 148
45. 구조에서 체험의 설명으로 ..... 151
46. 서로 다른 생물과 연구 질문 ..... 154
47. 질문과 증거 공백의 행렬 ..... 157
48. 외부 응답이 다른 비교 가능한 과제 ..... 158
49. 사건 수와 사건들 사이 관계의 구분 ..... 164
50. 반응의 전이로 구별하는 습관화와 고갈 ..... 166
51. 정보와 자기 기술의 개별 점검 ..... 169
52. 물리적 환경의 인공 시스템 ..... 170
53. 기반만으로 연구 경계가 정해지지는 않는다 ..... 172
54. 모형이 충족할 요구의 지도 ..... 179
55. 계산하는 경쟁 모형들의 공정한 비교 ..... 180
56. 한 갱신의 정규화 ..... 189
57. 출처의 이력으로 돌아가기 ..... 186
58. 이름 변경과 인과 구조 변경 ..... 194
59. 인과 조직에서 체험 가설로 ..... 198
60. 설명을 구별할 연구의 네 조건 ..... 202
61. 가능한 결과에서 구별 검증으로 ..... 205
62. 한 행위자의 두 순환 구조 ..... 209
63. 추가 관찰의 조건부 가치 ..... 211
64. 성공과 효용은 서로 다른 질문에 답한다 ..... 213

65. 방식 사이 차이의 불확실성 ..... 214
66. 표상을 변환해도 유지되는 것 ..... 217
67. 재귀성과 비교 가능한 대안들 ..... 219
68. 내용 전달과 직접 접근 대조 ..... 221
69. 나중에 준 지시 뒤의 접근 검증 ..... 222
70. 갱신에 사용하는 이력의 주인 ..... 226
71. 이력은 지연된 출처의 인식을 돕는다 ..... 227
72. 개입으로 검증하는 내부 접근 ..... 228
73. 기능적 재검토의 인과 순서 ..... 229
74. 재검토가 새 결정의 원인이 된다 ..... 230
75. 여러 메커니즘을 한 시스템으로 조립하기 ..... 231
76. 모듈이 있다고 기여가 드러나는 것은 아니다 ..... 232
77. 생물학적 구성요소와 측정 장비 ..... 235
78. 생물학적 자료에서 구분하는 기록 수준 ..... 239
79. 생물학적 자료의 시간 분할과 집단 분할 ..... 241
80. 내부 상태에 관한 세 가지 서로 다른 질문 ..... 247
81. 후보 검증의 프로필 ..... 248
82. 명확히 다른 두 기억 시간 모형 ..... 253
83. 이용 가능한 정보의 경계와 매개변수의 비유일성 ..... 255
84. 사전 프로토콜의 자료 구조 ..... 262
85. 사전 계획과 이후의 발견 ..... 264
86. 설명용 쌍별 차이와 가정한 기준점 ..... 268
87. 실습 8의 구별하지 못한 결과를 정직하게 제시하기 ..... 270
88. 완료한 계산과 다음 인과 연구 설계 ..... 277
89. 주장에서 검증과 대조로 ..... 278

# 부록 F 실습 해설과 자기 점검

먼저 직접 풀어 본 다음 해설을 읽는다. 관찰 과제에서는 기록이 명확한지, 추론과 가정이 구별되는지 살핀다. 계산 과제에서는 출발점이 된 규칙과 중간 단계를 대조한다. 후속 연구에는 근거를 갖춘 해법이 여럿 있을 수 있다. 자기 점검의 질문은 각 장에 그대로 두었고, 답은 여기에 실었다.

## 제1장 실습 해설

정해진 감각 목록은 없다. 기록에서 외부 사실, 체험에 대한 기술, 원인에 대한 설명을 구별할 수 있으면 과제를 마친 것이다. 두 번째 시도의 기록에 세부 사항이 늘었다고 해서 체험 자체가 더 풍부해졌다고 단정할 수 없다. 주의나 기억, 쓰는 어휘가 바뀌었을 수도 있다. 잘 정리된 후속 질문은 가능한 설명들과 다음에 무엇을 비교할지 밝힌다.

### 자기 점검 답안

1. 이 말은 느낌의 변화를 보고하고, 물건의 물리적 변화로 그 까닭을 설명한다. 물건이 따뜻해졌다는 주장과 두 변화 사이의 인과관계는 각각 근거가 필요하다. 체험에서 알아차린 변화 하나에도 여러 설명이 가능하다.
2. 그렇다. 특정한 일이 어떻게 체험되었는지를 말하는 경우에 그렇다. 이런 보고는 평소의 선호나 물건의 품질에 관한 전반적인 판단과 구별해야 한다. 진술이 무엇을 뜻하는지에 따라 판단한다.
3. 두 기록이 서로 다르다는 점이 확인되었다. 그 차이는 체험, 주의, 기억 또는 언어가 달라졌음을 반영할 수 있다. 원인에 대해 결론을 내리려면 추가 비교가 필요하다. 세부 사항의 개수로 의식의 양을 잴 수는 없다.

제1장으로 돌아가기

## 제2장 실습 해설

지도가 원래의 관계를 제대로 나타내는지 확인한다. 수치가 순서만을 뜻했다면 모든 거리가 정확히 일치할 필요는 없다. 보기 좋은 도형을 만들기 위해 불안정한 응답을 아무 말 없이 고쳐서는 안 된다. 탐구를 이어 갈 때는 새 관계를 미리 예측한 결과로 모형의 쓸모를 평가한다.

**자기 점검 답안**

1. 이름은 재료나 용도, 익숙한 범주에 따라 사례를 묶을 수 있다. 체험의 기술과 쌍별 비교가 드러내는 관계는 이와 다르다. 어떤 기준으로 묶었는지 밝혀야 한다.
2. 척도의 뜻은 비교에 쓰인 지시에 달려 있다. 평정값은 순서상의 위치만을 뜻할 수도 있다. «두 배 더 강하다»처럼 크기의 정확한 비율을 주장하려면 별도의 근거가 필요하다.
3. 모형이 어떤 관계나 새 사례를 더 잘 예측할지 미리 정한다. 그런 다음 좌표를 고를 때 쓰지 않은 자료로 검증한다. 같은 네 점을 사용한 그림만 좋아졌다면 새로운 관계를 설명할 수 있음이 아직 입증되지 않았다.

제2장으로 돌아가기

## 제3장 실습 해설

손이 움직이지 않아도 신체 소유감은 남을 수 있고, 포인터를 제어해도 그것이 몸의 일부가 되는 것은 아니다. 예전에 메모를 쓴 사람이라는 사실도 지금 읽는 행동과는 다르다. 사람마다 같은 체험을 할 필요는 없다. 각 질문의 차이를 살리고, 바깥에서 일어난 일을 근거로 내부의 관계를 곧바로 단정하지 않는 것이 중요하다.

**자기 점검 답안**

1. 제어는 명령과 결과의 관계를 가리킨다. 신체 소유감은 그 대상이 자기 몸의 일부로 느껴지는지에 관한 것이다. 두 관계가 함께 나타날 수는 있지만, 하나가 다른 하나를 결정하지는 않는다.
2. 실제로 누가 행동을 일으켰는지는 사건의 인과관계를 따라 확인한다. 행위주체감은 행동을 자신이 했다고 느끼는 방식으로 보고 자료와 함께 연구한다. 두 결과가 서로 다를 수 있으며, 그 차이 자체도 설명이 필요하다.
3. 이름은 자기 기술의 한 요소다. 공간적 관점, 신체 소유감, 행동을 어떻게 체험하는지, 자기 판단을 어떻게 검증하는지까지 설명해 주지는 않는다. 쓸 만한 모형은 이 가운데 어떤 과제를 어떤 메커니즘으로 설명하는지 밝혀야 한다.

제3장으로 돌아가기

## 제4장 실습 해설

가장 적은 것은 지우개다. 두 번째 문제의 조건만으로 크고 초록색인 카드가 있다고 결론 내릴 수 없다. 마지막에 온 사람은 표트르다. 두 번째 문제에서 초록색 카드는 모두 작고 둥글며, 크고 둥근 카드는 다른 색일 수 있다. 이 배치는 조건을 충족하면서도 결론이 반드시 성립한다는 주장을 반박하는 반례다. 이 과제는 예측과 뒤이은 검증이 어떻게 이어지는지 보여 준다. 자기의식의 전반적인 수준을 측정하는 일은 아니다.

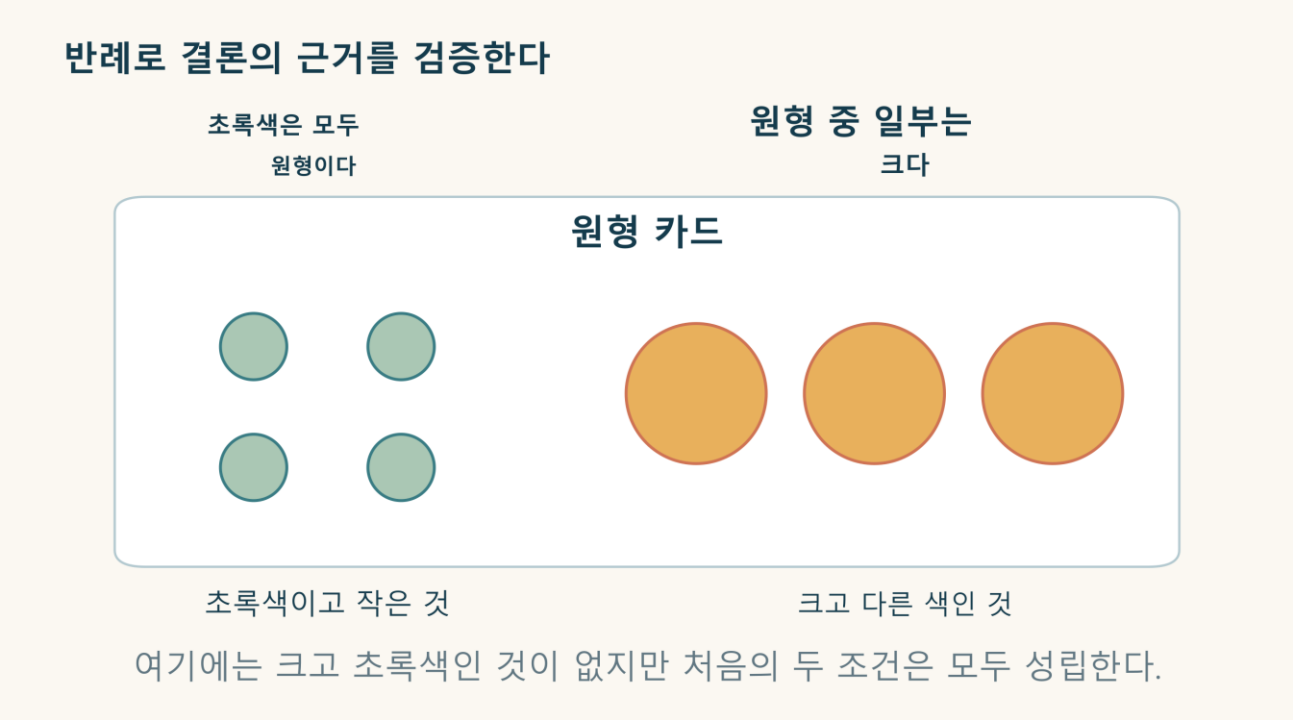


그림 9. 반례로 검증하는 논리적 추론. 도식의 초록색 카드는 모두 원형이며 원형 카드 일부는 크다. 하지만 크고 초록색인 카드는 없다. 출발 조건들이 그 결론을 보장하지 않으며, 반례는 추론의 구체적인 공백을 보여 준다.

### 자기 점검 답안

1. 피드백으로 정보를 얻으면 새로운 설명을 세울 수 있다. 처음 알아차렸던 단계, 나중에 떠올린 기억, 정답을 알고 재구성한 과정은 구별해야 한다.
2. 그렇다. 거의 모든 답에 같은 확신도를 매길 수 있다. 그러면 전반적인 성적과 오답일 가능성이 높은 답을 골라 검증하는 능력은 서로 다른 특성이 된다.

3. 과제의 유형과 관찰된 오류, 원인이라고 생각하는 과정을 명시하고 절차를 어떻게 바꿀지 구체적으로 제안한다. 그런 다음 비교할 수 있는 새 과제들의 결과를 대조한다. 성적이 좋아진 다른 이유가 있는지도 고려한다.
제4장으로 돌아가기

## 제5장 실습 해설

일지를 쓴 조건과 같은 시간 동안 멈춘 조건을 비교하면 두 설명을 가려내기 시작할 수 있다. 그래도 측정이 필요하고 과제들은 서로 비교할 수 있어야 하며 순서도 통제해야 한다. 고른 지표가 단지 글이 더 길어졌음을 나타낸다면 구별하는 능력이 나아졌다는 결론은 아직 근거가 부족하다. 이런 한계를 남기고 다음 비교를 제안하는 것이 중요하다.

### 자기 점검 답안

1. 메커니즘은 부품이 어떤 조건에서 어떤 순서로 상호작용해 결과를 만드는지 보여 준다. 부품 목록만으로는 그것들이 함께 어떻게 작동하는지 설명할 수 없다.
2. 지표가 보고에 영향을 줄 수도 있고, 보고에 따라 달라질 수도 있고, 다른 과정 때문에 보고와 함께 변할 수도 있다. 이 가능성들을 구별하려면 추가 자료와 적절한 개입이 필요하다.
3. 구현된 기능과 체험이 어떤 관계인지 뒷받침해야 한다. 둘의 대응을 설명하는 이론과 그 적용 범위, 대안에 대한 검증이 필요하다. 인식에 성공했다는 사실은 과제를 수행했음을 보여 준다. 그보다 더 나아간 결론은 추가 논거를 요구한다.
제5장으로 돌아가기

## 제6장 실습 해설

물건을 확인하면 바깥에서 확인할 수 있는 성질을 알 수 있다. 처음의 보고는 그때의 절차에 따른 것이다. 뒤늦게 알게 된 내용으로 다시 쓰지 않는다. 다른 사람이 지시를 그대로 따르고 정보가 정확히 언제 주어졌는지 파악할 수 있다면 기록표가 제 역할을 한 것이다. 여러 지표가 일치할 때는 찬성표를 세듯 결론을 내리기보다 공통 원인이 있는지 확인해야 한다.

**자기 점검 답안**

1. 그 일을 겪은 뒤 기억을 잃었거나 질문이 어려워 답하지 못했을 수 있다. 체험, 정보가 기억에 남았는지 여부, 지금 그 정보를 보고할 수 있는지를 구별해야 한다.
2. 정해진 과제에서 정보를 성공적으로 사용했음을 보여 준다. 체험의 선명도와 완전성 등 다른 성질을 알려면 추가 자료와 그 성질을 자료에 연결하는 근거가 필요하다.
3. 세 지표가 모두 응답을 준비하는 같은 과정에 의존할 수 있다. 공통 원인을 확인하고 여러 방법을 함께 쓰면 어떤 설명이 배제되는지를 보여 주어야 그 일치에 더 큰 의미가 생긴다.

제6장으로 돌아가기

## 제7장 실습 해설

기록은 쓰는 어휘나 양식에 익숙해진 정도, 사용한 시간 때문에 길어질 수 있다. 체험을 더 잘 구별하게 됐다고 결론 내리려면 그 능력을 보여 주는 새 결과가 필요하다. 네 가지 배정은 자료와 순서, 작업 방식을 서로 구별하도록 돕는다. 모든 결과를 가설의 확인으로 받아들인다면 아직 가설이 검증할 수 있는 차이를 제시하지 못한 것이다.

**자기 점검 답안**

1. 한 조건에서는 모든 시간을 읽는 데 쓸 수 있지만, 다른 조건에서는 일부를 도식을 그리는 데 쓸 수 있다. 같은 총시간을 전제로 비교하는 데에는 분명한 의미가 있지만 실제 준비 과정은 서로 다르다.
2. 참가자에게 알려진 특성을 기준으로 조건을 일관되게 골라 주는 일을 막고 명확한 배정 절차를 제공한다. 다만 규모가 작은 집단에서는 조건마다 참가자 구성이 우연히 달라질 수 있다.
3. 새 가설은 이미 본 규칙성을 고려해 세운 것이다. 독립적으로 검증하려면 새 자료나 가설을 세우는 데 쓰지 않도록 미리 따로 떼어 둔 자료가 필요하다.

제7장으로 돌아가기

## 제8장 실습 해설

네 번째 반복의 차이는 마이너스 이십 퍼센트포인트, 여섯 번째는 플러스 이십 퍼센트포인트다. 실습용 생성기의 두 조건은 같지만 개별 추정은 달라진다. 추정이 부정확하면 새로운 독립

자료가 필요할 수 있다. 중요한 효과를 정밀하게 배제할 수 있다면 가설의 범위를 좁힐 수 있다. 결과가 자료에 따라 다르면 조건을 다시 검증해야 한다. 자료의 행 수가 독립적으로 연구한 단위의 수를 대신하지는 못한다.

### 자기 점검 답안

1. 퍼센트포인트는 두 비율을 퍼센트로 표시한 값의 차이다. 상대적 증가율은 처음 값을 기준으로 구한다. 십을 칠십으로 나누면 약 십사 퍼센트다.
2. 한 모형의 응답들은 같은 매개변수와 학습 이력에 의존한다. 그 모형의 동작을 평가하는 데에는 유용하지만 독립적으로 만든 모형들 사이의 변동을 알려 주지는 않는다.
3. 그렇다. 명시한 조건에서 미리 중요하다고 정한 개선 폭을 배제할 만큼 정밀하게 측정했다면 의미가 있다. 자료에 들어맞는 효과의 범위가 넓다면 결과는 아직 불확실하다. 뚜렷한 신호를 찾지 못했다고 해서 효과가 없다고 입증된 것은 아니다.
제8장으로 돌아가기

## 제9장 실습 해설

A에서는 셋째 단계에 펄스가 발생한다. B는 문턱에 이르지 않으며 최종 상태는 약 0,921754다. 유지 계수가 영이면 영이 아닌 각 입력이 만드는 값은 0,6뿐이므로 두 입력열 모두 펄스가 없다. 여기서 설명한 것은 선택한 메커니즘의 동역학이며, 뉴런이나 의식의 보편적인 성질은 아니다.

### 자기 점검 답안

1. 규칙은 현재 입력뿐 아니라 보존된 상태도 사용한다. 이력이 달라지면 문턱까지의 거리가 달라진다.
2. 상태는 주어진 규칙 아래 현재의 동역학을 기술한다. 학습은 규칙의 매개변수 자체를 바꾸어 이후 입력열에 대한 작동을 바꾼다.
3. 전체 합으로 설명하는 경우와 누출을 동반한 시간적 누적으로 설명하는 경우를 구별한다. 출력의 차이는 바로 이 모형에서 시간 구조가 하는 역할을 뒷받침한다.
제9장으로 돌아가기

## 제10장 실습 해설

순방향 규칙과 반복 처리는 같은 범주를 출력할 수 있다. 응답의 일치만으로 메커니즘을 고를 수는 없다. 그림은 같고 숨은 정답 표지는 다른 두 카드는 정보상의 모호함을 보여 준다. 더 나아가 결론을 내리려면 구체적인 규칙과 동등한 정보, 응답이 갈라지는 조건이 필요하다.

### 자기 점검 답안

1. 보고는 질문 형식과 기억, 시간, 보고 순서에 좌우된다. 체험에 관한 결론에는 이 과정들과의 관계를 뒷받침하는 근거가 필요하다.
2. 맥락은 서로 다른 구조에서도 이용할 수 있다. 비교할 만한 입력과 자원을 갖춘 모형들을 예측이 달라지는 조건에서 비교해야 한다.
3. 정확히 누구 또는 어떤 구성요소가 어느 시점에 어떤 동작을 위해 정보에 접근하는지 밝혀야 한다. 분류하고 보고하고 기억하고 응답을 검증하는 능력을 곧바로 하나의 능력으로 간주할 수는 없다.

제10장으로 돌아가기

## 제11장 실습 해설

이력은 효과를 이전 명령과 대응시키도록 돕는다. 선택한 지연 계열에서는 두 단계의 시차로 연결이 복원된다. 두 메커니즘이 수행한 모든 단계에서 같은 관찰을 만든다면 현재 자료로는 선택할 수 없다. 숨은 열은 평가자에게 필요하지만 연구하는 모형의 계산에 들어가서는 안 된다.

### 자기 점검 답안

1. 외부 대상에서 온 신호들도 공통 원인을 가질 수 있다. 신체 소유감에는 신체 표상과의 추가적인 관계와 해당 체험에 관한 자료가 필요하다.
2. 명령에 새롭게 개입하고 그 결과를 비교한다. 두 가설에서 관찰이 같다면 수동적으로 확인한 일치만으로는 확실하게 구별할 수 없다.
3. 개선은 전체 기억 용량으로 설명될 수 있다. 출처별 조직의 기여를 가려내려면 적절한 대안과 비교하고 새 조건으로의 일반화를 검증해야 한다.

제11장으로 돌아가기

## 제12장 실습 해설

첫 열의 예측은 차례로 약 0,500; 0,667; 0,750; 0,600; 0,667; 0,571이다. 평균제곱오차는 약 0,309이며 고정 모형은 0,250이다. 이는 한 번의 짧은 열에서 얻은 결과다. 후속 열에서는 처음 평활화를 되풀이하지 말고 보존한 계수기에서 시작해야 한다. 검증하는 것은 체험되는 확신도의 깊이가 아니라 예측의 품질이다.

### 자기 점검 답안

1. 이전에는 없었던 정보를 사용하기 때문이다. 검증할 예측은 되먹임 전에 확정해야 한다. 그렇지 않으면 미래의 단서를 얻게 된다.
2. 개별 과제는 난이도와 이용 가능한 특징이 다를 수 있다. 전체 이력은 선택한 유형의 평균 신뢰도를 평가하며, 개별 사례에는 추가 정보가 필요하다.
3. 갱신은 정해진 규칙을 따른다. 명시적 성찰은 그 규칙의 근거, 적용 가능성, 한계를 검토하며, 다른 평가 방법의 제안과 검증으로 이어질 수 있다.

제12장으로 돌아가기

## 제13장 실습 해설

전체 기능을 켠 계는 두 지표 모두 10개 중 10개를 맞힌다. 되먹임이 없으면 인식과 기억 모두 10개 중 7개다. 전달이 없으면 인식은 10개 중 10개로 유지되고 기억은 10개 중 4개다. 두 메커니즘을 모두 차단하면 각각 7개와 4개다. 순방향 경쟁 모형은 10개 중 10개를 맞힌다. 이 수치는 장치의 규칙에서 도출되며 사람을 조사해 얻은 자료가 아니다.

재현: python teaching_lab.py --chapter 13

### 자기 점검 답안

1. 전역 전파와 재귀성이 서로 엄밀한 대립을 이루지 않는 이유는 무엇인가?
재귀성은 영향이 되돌아오는 것을, 전역성은 서로 다른 과정이 내용을 이용할 수 있음을 기술한다. 공통 교환 시스템은 되먹임을 사용할 수 있다. 구분해야 할 것은 가정하는 충분조건과 구체적인 상호작용의 기능이다.

2. 개입 뒤 보고가 사라지면 체험도 사라졌다는 사실이 입증되는가?

그 자체로 입증되지는 않는다. 개입이 기억이나 응답 선택, 지시 수행을 손상했을 수 있다. 독립 지표와 이 기능들이 보존되는지 확인하는 대조 절차가 결론을 강화한다. 그래도 측정 문제 자체가 완전히 사라지는 것은 아니다.

3. 계산 예산을 같게 하여 메커니즘들을 비교해야 하는 까닭은 무엇인가?
그렇지 않으면 한 계에 처리 단계를 더 주었을 뿐인데도 특별한 구조 덕분에 유리해졌다고 오해할 수 있다. 대조 절차를 통해 관찰한 결과에 그 상호작용 조직이 필요한지 알아볼 수 있다.

제13장으로 돌아가기

## 제14장 실습 해설

여기서 성공이란 두 설명이 각각 무엇을 요구하는지 구별하는 것이다. 명세서를 보면 장치의 작동에는 동의하면서 체험과의 관계는 따로 논쟁할 수 있어야 한다. 예측이 같다면 자료만으로 선택하기에 부족하다. 필요한 형식 체계가 없다면 근사 도식을 이론의 기준을 계산한 결과로 제시할 수 없다.

### 자기 점검 답안

1. 부분 사이의 높은 상관만으로 상호 인과 영향을 결론 내릴 수 없는 까닭은 무엇인가?
두 부분 모두 제삼의 출처에 의존할 수 있다. 인과적 결론에는 추가 전제와 개입, 또는 별도로 근거를 갖춘 의존 관계 모형이 필요하다.

2. 성찰은 모든 고차 표상과 같은가?
아니다. 자세한 성찰에는 자기 근거의 검토와 설명 비교, 전략 수정이 들어간다. 현재 신호의 신뢰도 표상은 훨씬 좁은 기능만 수행할 수 있다.

3. 이론은 수학적으로 정확하면서 경험적으로 논쟁적일 수 있는가?
그렇다. 정확성은 가정에서 귀결을 엄밀하게 도출할 수 있게 한다. 가정과 귀결이 의식을 기술한다고 확인하려면 독자적인 자료와 대안 설명과의 비교가 필요하다.

제14장으로 돌아가기

## 제15장 실습 해설

오른쪽 신호를 받으면 새 확률은 약 0,857이다. 왼쪽 신호에서는 0,12와 0,32의 합으로 0,12를 나눈 약 0,273이다. 정방향 대응에서는 목표의 방향과 같은 명령을, 역방향 대응에서는 반대 명령을 고른다. 예상이 맞는다는 것만으로 체험되는 행위주체감을 입증할 수는 없다. 여기서는 계산 가능한 인과 모형을 검증한다.

재현: python teaching_lab.py --chapter 15

### 자기 점검 답안

1. 모든 예측은 의식적인 예상인가?
아니다. 가능한 신호들의 계산된 분포를 예측이라고 부를 수 있다. 이런 계산이 있다는 것만으로 그 계산이 체험된다는 사실을 확립하지는 못한다.

2. 행동 모형과 세계 모형은 어떻게 다른가?
행동 모형은 자기 명령과 결과의 연결을, 세계 모형은 주변 상황의 상태와 변화를 기술한다. 두 오류가 겉으로는 모두 행동 실패를 낳아도 고치는 방법은 달라야 한다.

3. 신호가 없다는 사실과 부정적인 신호를 구별해야 하는 까닭은 무엇인가?
작동하지 않는 센서는 목표가 없다는 정보를 주지 않는다. 둘을 섞으면 근거 없는 확신이 생긴다. 관찰의 누락과 그동안 세계가 달라질 가능성을 명시적으로 모형화해야 한다.

제15장으로 돌아가기

## 제16장 실습 해설

올바른 기록표에도 모르는 항목이 있을 수 있다. 보고 요구를 없앴다고 모든 동반 과정이 없다고 곧바로 결론 내릴 수는 없다. 디코딩은 분석에서 이용할 수 있는 정보를 보여 준다. 필요성과 충분성에는 다른 근거가 필요하다. 다음 예측이 두 이론에 공통적이라면 메커니즘 연구에 유용해도 아직 이론을 선택해 주지는 못한다.

### 자기 점검 답안

1. 대규모 연구가 이론 논쟁을 반드시 최종 해결하는 것은 아닌 이유는 무엇인가?

연구 규모가 커지면 특정 측정의 정보성이 높아지지만 검증한 주장들의 범위가 자동으로 넓어지지는 않는다. 선택한 예측과 과제, 지표, 분석 감도가 중요하다.

2. 보고가 없는 조건은 체험을 직접 측정하는가?
아니다. 응답 요구 일부를 줄이지만 지각 여부를 어떻게 판단하는지 근거가 필요하다. 간접 지표에도 전제와 한계가 있다.

3. 예상한 결과와 경쟁 설명을 구별하는 예측 가운데 무엇이 모형에 대한 신뢰를 더 크게 바꾸는가?
한 모형에서는 가능성이 크고 유력한 경쟁 모형에서는 가능성이 낮은 결과가 특히 많은 정보를 준다. 공통 예상과의 일치도 유용하지만 설명을 고르는 데에는 도움이 덜 된다.

제16장으로 돌아가기

## 제17장 실습 해설

단계 0–4의 전체 궤적은 0; 0,4; 0,6; 0,6; 0,6이다. 유지가 없으면 0,6 뒤에는 0,3과 0,15가 나온다. 최초 입력이 없으면 0; 0; 0,4; 0,2; 0,1이다. 상태를 읽는 표지는 규칙에 따로 넣지 않았다면 동역학을 바꾸지 않는다. 모든 수치는 실습용 메커니즘에 해당한다.

### 자기 점검 답안

1. 동역학적 발생과 진화적 기원이 다른 까닭은 무엇인가?
앞의 것은 작동하는 계의 상태 전이를 기술한다. 뒤의 것은 세대 사이에서 메커니즘이 생기고 달라진 역사를 다룬다. 서로 다른 자료가 필요하며 설명하는 시간 규모도 다르다.

2. 지속적인 활동이 의식적 사례의 존재를 입증하는가?
아니다. 안정성은 상태의 유지를 설명한다. 내용과 그 인과 역할, 체험과 연결할 근거를 추가로 설명해야 한다.

3. 신경 지표와 행동 지표의 어긋남은 무엇을 뜻하는가?
두 지표를 단순히 동일시하기 어렵다는 뜻이다. 측정 감도와 과제에서 과정의 역할, 보상 가능성 같은 여러 설명을 검증해야 한다. 어긋남 자체도 유용한 자료다.

제17장으로 돌아가기

## 제18장 실습 해설

예를 들어 표지를 만지지 않은 것은 표지를 알아차리지 못했거나 해당 동작이 어려워서일 수 있다. 표지가 잘 보이는지와 움직임의 대조 조건으로 두 설명을 구별한다. 도움 요청에는 사회적으로 도움을 받을 수 있는 조건도 필요하다. 각 성취에 관한 결론은 모든 의식 형태의 발생 시점을 정한 것과 다르다.

### 자기 점검 답안

1. 자기 인식에 성공한 나이를 의식이 생긴 나이로 볼 수 없는 까닭은 무엇인가?
자기 인식에는 특정한 시각과 운동, 인지 능력이 필요하다. 의식적 체험과 다른 자기관련성 형태를 이 특정 과제의 수행과 동일시할 수는 없다.

2. 도움 요청은 명시적 성찰과 어떻게 다른가?
요청은 언어적 분석 없이 불확실성 평가로 조절될 수 있다. 명시적 성찰에서는 자기 근거를 검토 대상으로 삼아 비교하고 의도적으로 수정할 수 있어야 한다.

3. 어린 집단에서 예상한 지표가 없다는 것은 무엇을 뜻하는가?
기능의 부재와 다른 구현, 과제의 추가 어려움, 부족한 측정 감도 등 여러 설명이 가능하다. 대조 과제와 시간에 따른 변화의 관찰로 설명을 고를 수 있다.

제18장으로 돌아가기

## 제19장 실습 해설

두 실습용 역사는 주어진 끝 가지의 관찰과 모두 양립한다. 기억의 이점은 환경의 변동성과 비용에 의존한다. M의 기원을 확실하게 복원해도 M과 체험의 관계를 뒷받침할 근거 없이는 체험을 확립하지 못한다. 좋은 관계도에는 이 의존성들이 모두 남고 다음에 필요한 자료가 명시된다.

### 자기 점검 답안

1. 메커니즘의 적응상 이점이 체험의 존재를 입증하는가?

아니다. 어떤 조건에서 선택이 메커니즘을 유지할 수 있었는지 설명한다. 그 메커니즘과 체험의 관계에는 별도 근거가 필요하다.

2. 가까운 계통의 종들이 완전히 독립적인 확인을 주지 못하는 까닭은 무엇인가?
공통 특징이 하나의 역사적 사건에서 왔을 수 있다. 비교 분석에 공통 역사를 반영하지 않으면 독립적인 정보의 양을 부풀리게 된다.

3. 여러 기원 가설을 유지해야 하는 까닭은 무엇인가?
자료가 불완전하면 여러 가설이 관찰된 특징 분포를 똑같이 설명할 수 있다. 대안을 명시하면 어떤 추가 관찰이 재구성들을 실제로 구별하며 어떤 결론이 아직 이른지 알 수 있다.

제19장으로 돌아가기

## 제20장 실습 해설

작은 연구가 의식의 모든 질문을 즉시 해결해야 한다고 요구할 수는 없다. 그래도 결론은 실제 작업에 정확히 맞아야 한다. 강한 원리는 보고의 예측보다 더 많은 것을 요구한다. 불편한 구현을 임의로 제외하지 말고 일관되게 근거를 제시해야 한다.

### 자기 점검 답안

1. 색 구별의 설명이 체험의 존재에 관한 질문을 해결하지 못할 수 있는 까닭은 무엇인가?
이는 시스템이 신호를 어떻게 분류하고 대응시키는지 보여 준다. 이런 과정이 체험된다는 주장에는 계산 조직과 체험적 성질 사이의 별도 연결이 필요하다.

2. 체험을 설명하려면 반드시 추가 실체를 도입해야 하는가?
아니다. 물리적 동일성의 전략이 가능하다. 다만 어떤 과정이 체험과 동일하며 왜 그런지 근거를 제시해야 한다. 두 기술이 일치한다고 선언하는 것만으로는 안 된다.

3. 최종 설명이 아직 없다면 무엇을 진전으로 볼 수 있는가?
경쟁 설명을 구별할 새 예측, 유력한 대안의 배제, 메커니즘 구조와 체험 구조 사이의 더 정확한 대응, 남은 가정의 명확한 축소를 들 수 있다. 질문 전체를 해결했다는 주장과 별개로 이런 결과들을 검증할 수 있다.

제20장으로 돌아가기

## 제21장 실습 해설

기억 연구에서는 주어진 조건에서 선택이 이용 가능한 정보에 특정한 방식으로 의존한다고 결론 내릴 수 있다. 손상 연구에서는 자극의 감지와 회피, 가정한 불쾌한 체험을 다른 수준으로 유지해야 한다. 가까운 대안은 «본능»이라는 말 대신 규칙이나 메커니즘으로 정해야 한다.

### 자기 점검 답안

1. 통증의 가능성을 인정하는 것이 명시적 성찰도 인정한다는 뜻이 아닌 까닭은 무엇인가?
불쾌한 체험과 신체 상태 조절, 자기 근거의 검토는 서로 다른 성질이다. 첫 성질에 관한 자료가 곧바로 다른 성질의 증거가 되지는 않는다.

2. 한 종의 결과를 가까운 계통의 모든 종에 일반화할 수 있는가?
메커니즘과 계통 관계, 과제 조건을 고려한 추가 근거가 있어야 한다. 계통적 가까움은 가설을 만드는 데 도움이 되지만 그 범위의 검증을 대신하지는 못한다.

3. 단순한 규칙으로 설명한다면 어떻게 다루어야 하는가?
정확히 적고 같은 자료와 새 조건에서 검증해야 한다. 규칙이 결과를 설명하면 더 복잡한 추론은 약해진다. 규칙이 체계적으로 틀리면 더 풍부한 모형을 제안할 근거가 생긴다.

제21장으로 돌아가기

## 제22장 실습 해설

새 B를 제시하기 전까지 반응은 같다. B에서는 각각 1과 0,32768이므로 관찰로 두 규칙을 구별한다. 휴지기 다섯 단계 뒤 A에 대한 반응은 신호별 모형에서 약 0,779695, 자원 모형에서 약 0,758219다. 이 차이는 정한 계산의 귀결이며 계의 퀄리아를 확립하지 않는다.

재현: python teaching_lab.py --chapter 22

### 자기 점검 답안

1. 기능적 기억은 신경계 없이도 존재할 수 있을까?

그렇다. 계의 상태는 과거 작용의 영향을 보존하여 이후 반응을 바꿀 수 있다. 이는 기억의 넓은 기능적 의미이며, 체험되는 회상과 구별해야 한다.

2. 반응의 감소만으로 아직 학습을 입증할 수 없는 까닭은 무엇인가?
소진과 손상, 입력의 이용 가능성 변화, 그 밖의 과정 때문일 수 있다. 대조 조건과 회복 및 특이성 검증, 대안의 더 정확한 모형이 필요하다.

3. 뉴런의 부재가 퀄리아의 문제를 무의미하게 만드는가?
아니다. 가능한 메커니즘의 범위와 증명 요구가 달라진다. 다른 담지체를 허용하는 이론은 계의 전반적인 복잡성에 그치지 않고 충분한 조직과 검증 방법을 밝혀야 한다.

제22장으로 돌아가기

## 제23장 실습 해설

독립적인 문구 틀을 바꾸면 계산을 바꾸지 않고도 글이 달라질 수 있다. 사용하는 상태를 의도적으로 바꾸면 행동을 바꿀 수 있지만, 그 역할은 해당 구현의 규칙으로 검증해야 한다. 도식은 구조와 접근에 관한 질문을 뒷받침한다. 그 자체로 체험되는 당혹감이나 다른 질적 성질의 근거가 되지는 않는다.

### 자기 점검 답안

1. 체험을 능숙하게 이야기한다는 것만으로 체험의 존재를 결론 내릴 수 없는 까닭은 무엇인가?
보고하는 내부 상태와 관련 없는 메커니즘이 글을 만들 수 있다. 응답의 출처와 정보 접근, 가정한 자기 점검의 인과 역할을 연구해야 한다.

2. 생물학적 담지체는 의식을 보장하는가?
아니다. 재료만으로 해당 이론이 요구하는 조직이 정해지지는 않는다. 구체적인 계와 연결, 외부 제어와의 상호작용을 연구해야 한다.

3. 후보의 성질별 양상을 정리하는 것이 충족한 지표의 개수를 세는 것보다 나은 까닭은 무엇인가?
결론의 이론 의존성을 유지하고 필요조건과 충분조건, 동반 특성을 구별하기 때문이다. 단순한 합은 중요한 조건의 부재를 감추고 평가에 근거 없는 정밀성을 부여할 수 있다.

제23장으로 돌아가기

## 제24장 실습 해설

분리된 이력이 유용할 수는 있지만 공통 추적기와 일치하면 결론은 동일성 메커니즘에 한정된다. 구별 조건이 없으면 표본이 커도 이 질문에 관한 정보를 주지 못한다. 완성된 명세는 다음에 어떤 자료와 계산 가능한 규칙이 필요한지 보여 주어야 한다.

### 자기 점검 답안

1. 일반적인 학습 및 대상 추적 메커니즘도 이력을 사용할 수 있다. 특별한 결론을 내리려면 그에 맞는 경쟁 모형이 필요하다.
2. 모형의 구조와 이용 가능한 정보의 이점을 섞는 비교다. 메커니즘 자체의 기여를 아직 가려내지 못했다.
3. 나누지 않으면 가설이 자기 자신의 증거가 된다. 나누면 모형이 이미 반영해야 할 것과 앞으로 검증할 것이 드러난다.

제24장으로 돌아가기

## 제25장 실습 해설

M0의 결과는 0,75; 0,25; 0,75; 0,75다. M1은 0,75; 0,4375; 약 0,710526; 약 0,858108이다. M2와 일반 추적기는 0,75; 0,25; 0,875; 0,5625다. 유지 계수가 영이면 모든 갱신이 0,5에서 시작한다. 출처 이름을 동시에 바꾸면 기록의 이름만 달라지고 수치 관계는 같다.

재현: python teaching_lab.py --chapter 25

### 자기 점검 답안

1. B의 반증과 따로 보존한 A의 지지 이력으로 돌아가기 때문이다.
2. 이 구현에서는 M1과 M2가 같아져 출처별 이력의 이점을 검증할 수 없다.
3. 응답은 척도와 사용 방식에도 의존한다. 변수의 이름만으로 계산과 체험 보고의 관계를 정할 수는 없다.

제25장으로 돌아가기

## 제26장 실습 해설

독립적인 기록은 평소 궤적을 재현해도 개입의 방향별 응답을 유지하지 못할 수 있다. 단방향 제어기에서는 없는 방향을 따로 검증해야 한다. 단순 조절기도 단순해 보인다는 인상만으로 제외할 수 없다. 정의의 실질적인 조건을 검증해야 한다. 이 검사 어느 것도 그 자체로 조직이 체험에 충분함을 입증하지 않는다.

재현: python teaching_lab.py --chapter 26

### 자기 점검 답안

1. 특정 조직이 체험의 존재에 충분하다고 주장한다. 보고의 예측은 더 제한된 관계를 확립한다.
2. 가설에서 중요한 것은 이름을 바꾸어도 남는 관계와 전이다.
3. 아니다. 조립은 조건의 구현을 검증한다. 충분성, 구현의 한계, 체험에 관한 추론으로의 확장에는 각각 근거가 필요하다.

제26장으로 돌아가기

## 제27장 실습 해설

M0는 현재 구간을, M1은 공통 이력을, M2는 인식한 출처의 이력을 사용한다. 공통 추적기가 M2와 일치한다면 M1보다 낫다는 결과는 분리된 이력을 지지하지만 특별한 신체 조직을 따로 가려내지는 못한다. 원리에서 나오는 체험상의 귀결은 질적 관계에 대한 독립 검증을 요구하며, 기능적 비교 하나만으로 도출되지 않는다.

### 자기 점검 답안

1. 아니다. 다른 예측을 내는 모형이나 다른 비교 대상이 필요하다.
2. 약한 대조 모형과 비교해 현재 인과 연결의 역할을 보여 줄 뿐, 온전한 대안 메커니즘들을 배제하지는 못한다.
3. 분석에서 발견한 것으로 기술하고 새 예측을 도출하여 독립 자료로 검증한다.

제27장으로 돌아가기

## 제28장 실습 해설

기본 전체 모드는 단계당 성공률이 약 0,8794이고 효용은 0,8619다. 고정 모형의 평균 성공률 약 0,4997은 반전 전의 0,8780과 반전 뒤의 0,1213을 감춘다. 이 차이는 해당 환경에서 제어 갱신의 기능적 이점을 뒷받침한다. 관찰 비용 때문에 효용과 정확도가 다르다.

### 자기 점검 답안

1. 특정 메커니즘 하나를 끄기 때문이다. 기억 전체를 없애면 결과의 원인 여러 개를 동시에 바꾸게 된다.
2. 거의 정확한 제어와 전환 뒤 거의 체계적인 오류를 평균에 합치기 때문이다.
3. 결과 전의 불확실성을 측정하는 예측이 아니게 된다. 평가 지표 안으로 정답이 누출된다.

제28장으로 돌아가기

## 제29장 실습 해설

회전과 평행이동은 수치 표현의 정밀도 범위에서 거리를 보존하지만, 늘이기는 대개 거리를 바꾼다. 재귀 단계와 그에 대응하여 펼친 층의 결과는 일치한다. 전달 과제에서 전체 선택과 직접 접근은 100퍼센트를 낸다. 여기서 보인 것은 채널의 작동이지 그 채널의 독점적 필요성이나 체험적 성질이 아니다.

### 자기 점검 답안

1. 선택한 변환이 이 모형의 기하를 보존한다는 뜻이다. 체험의 존재가 도출되지는 않는다.
2. 구체적인 유한 계산을 검증한 것이다. 이론이 시간과 상태, 물리적 구현에 추가 요구를 둔다면 따로 검증해야 한다.
3. 이 과제에 필요한 정보를 적어도 두 방식으로 사용할 수 있다. 제한된 공통 공간 자체의 기여는 아직 가려내지 못했다.

제29장으로 돌아가기

## 제30장 실습 해설

정한 규칙에서 M2는 일반 추적기와 일치한다. 직접 내부 센서는 설계상 접근 권한이 있다. 적응 전략은 과거 오류로 행동을 바꾼다. 기본 통합 후보의 가속 수정은 활성화 횟수가 영이다. 다른 다음 단계가 필요한 서로 다른 결과다. 이를 합쳐 한 계의 체험을 입증할 수는 없다.

### 자기 점검 답안

1. 최소 계산만으로는 특별한 자기관련성과 일반적인 출처 이력의 사용을 아직 구별하지 못함을 보여 준다.
2. 정한 접근 채널과 측정 검증의 작동을 확인한다. 더 강한 결론에는 추가 근거가 필요하다.
3. 아니다. 기본 시나리오에서 메커니즘이 활성화되지 않았다. 예측한 역할을 검증할 실험을 미리 정해야 한다.

제30장으로 돌아가기

## 제31장 실습 해설

학습 표에는 원래 식별자와 명시적 제외를 보존한다. 내려받은 파일의 실제 구조를 저장소의 짧은 설명으로 대신하지 않는다. STTC를 더하면 오차가 줄거나 늘 수 있다. 허용되는 결론은 선택한 절차와 이 자료 집합에 한정된다. 새 배치로의 일반화에 관해서는 배치별 검증이 오가노이드별 검증보다 엄격하다. 하지만 그 역시 모든 의존성을 없애거나 체험을 확립하지는 않는다.

재현: python teaching_lab.py --chapter 31

### 자기 점검 답안

1. 조직과 입력 변환, 되먹임 알고리즘, 판독이 함께 결과를 만들 수 있다.
2. 첫 검증은 시료 안의 시간적 일반화를, 둘째 검증은 시료 사이의 일반화를 평가한다.
3. 아니다. 특정한 기능적 결과를 확립한다. 체험에 대한 결론에는 메커니즘과 대응 원리에 관한 추가 논거가 필요하다.

제31장으로 돌아가기

## 제32장 실습 해설

기록에는 구현과 검증을 마친 상태, 구현했지만 역할은 미검증인 상태, 미확인 상태를 구별해야 한다. 차단 효과가 없으면 활성화와 과제 감도, 보상을 확인해야 한다. 기능적 양상이 완전해도 자신이 세운 이론적 요구의 충분성을 독립적으로 확인하지는 않는다.

### 자기 점검 답안

1. 아니다. 활성화와 정보 의존성, 결정에 미치는 영향을 검증해야 한다.
2. 전자는 외부 분석의 능력을, 후자는 계 자체의 정보 이용을 보여 준다.
3. 구현은 후보의 구조를 검증한다. 충분성은 독립 근거가 필요한 별도의 이론 가설이다.

제32장으로 돌아가기

## 제33장 실습 해설

복원이 반드시 오류 없이 이루어져야 하는 것은 아니다. 결과에는 생성 규칙과 비교 기준, 모든 선택 결과, 불확실성이 포함된다. 같은 유형의 사건을 더해도 구조적 비식별성이 사라지지는 않는다. 틀린 모형들 중 이기면서도 절대 예측은 나쁠 수 있다.

재현: python teaching_lab.py --chapter 33

첨부한 실행에서는 각 생성기마다 사건 120개로 이루어진 순서 30개를 만들었다. 관찰 예측의 브라이어 오차가 가장 작은 후보를 선택한다. 여기서 후보의 매개변수는 고정되어 있다. 행렬의 행은 지정한 생성기, 열은 선택된 후보를 나타낸다. 이는 진단을 위한 학습 실험이며, 알려지지 않은 인간의 본성을 대상으로 한 비교가 아니다.

| 생성기 | M0 선택 | M1 선택 | M2 선택 |
|---|---|---|---|
| M0 | 25 | 1 | 4 |
| M1 | 3 | 24 | 3 |
| M2 | 3 | 5 | 22 |

선택한 격자에서 브라이어 오차의 최솟값 0,231314를 낸 매개변수 쌍은 이력 유지 0,80과 관찰 신뢰성 0,75였다. 이번 실행에서는 생성기에 지정한 쌍과 일치했다. 이런 결과 하나가 모든 조건에서 복원 정확도를 추정해 주는 것은 아니다.

### 자기 점검 답안

1. 특징에는 잡음이 있고 서로 다른 숨은 원인에서도 나타날 수 있다. 두 가능성을 모두 반영한 예측이 필요하다.

2. 아니다. 알려진 인공 상황에서 분석을 검증하는 것이다.

3. 규칙은 인과 동역학과 새 예측을 정한다. 모르게 바꾸면 실제로 무엇을 검증했는지 평가하기 어렵다.

제33장으로 돌아가기

## 제34장 실습 해설

절차에는 구체적인 자료와 계산 순서, 주결과, 판단 조건이 있어야 한다. 실습 기준 0,01은 직접 선택한 값이며 보편적인 과학 기준이 아니다. 미리 정한 순차 절차 없이 첫 유리한 차이에서 멈추면 검증의 의미가 달라진다.

재현: python teaching_lab.py --chapter 34

두 출처의 독립열 100개에서 M1 오차에서 M2 오차를 뺀 쌍별 차이의 평균은 0.016965다. 전체 쌍을 4000번 재표집한 95퍼센트 백분위 구간은 [0.013687; 0.020376]이다. 미리 선택한 실습 문턱은 0,01이다. 이 구간은 이 실험의 생성기와 절차에 해당한다. 결과를 계산하기 전에 절차를 보존했다. 이는 로컬 기록이며 독립 연구의 공개 사전등록은 아니다.

### 자기 점검 답안

1. 아니다. 그런 분석을 사전에 계획한 검증과 구별해야 한다.

2. 그렇지 않으면 검증할 정답 자체가 예측에 들어갈 수 있다.

3. 아니다. 선택한 실습 규모일 뿐이다. 구체적 연구에서는 필요한 정밀도와 모형의 차이로 규모의 근거를 제시한다.

제34장으로 돌아가기

## 제35장 실습 해설

실습 예의 평균 0,008만으로 실용적인 관심 이점 0,01에 답할 수는 없다. 활성화 횟수가 영인 구성요소의 두 버전이 같다는 사실은 활성 상태의 작동을 검증하지 않는다. 강한 후속 실험을 고르는 일도 의미 있는 분석 결과다.

### 자기 점검 답안

1. 그 과제에서는 모듈 작동의 결과를 검증하지 못했다. 기여를 평가하려면 구별되는 조건이 필요하다.
2. 규칙상 M2와 같은 출처 동일성의 일반 기억 M1I도 같은 결과를 낼 수 있다.
3. 수정은 명시적인 새 규칙과 독립 조건의 귀결을 정한다. 맞춤은 이미 본 결과의 개선에 그칠 수 있다.

제35장으로 돌아가기

## 제36장 실습 해설

완결된 연구는 입력에서 결론까지의 경로를 재현할 수 있게 하며, 계산을 검증하기 위해 체험 가설에 동의할 것을 요구하지 않는다. 독자적 기여는 새롭게 구별할 수 있는 귀결에서 출발한다. 제안한 충분성 원리에 아직 근거가 없다면 이를 정확히 명시하고 구체적인 후속 연구와 연결한다.

### 자기 점검 답안

1. 선택한 출처별 기억 메커니즘의 귀결과 일반 대상 추적기와의 일치를 확인했다.
2. 잠정 가설은 상호 인과 영향을 요구한다. 공통 활동이나 신호 전달 하나만으로는 이를 확립할 수 없다.
3. 그렇다. 수행한 검증을 정확히 밝히고 연구 목표를 유지하며 다음 구별 단계를 제안해야 한다.

제36장으로 돌아가기

# 참고문헌

[1] Dennett, Daniel C. (1988). Quining Qualia. In A. J. Marcel and E. Bisiach (eds.), Consciousness in Contemporary Science. Oxford University Press

[2] Chalmers, David J. (1995). Facing Up to the Problem of Consciousness. Journal of Consciousness Studies, 2(3), 200–219

[3] Shepard, Roger N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. Psychometrika, 27, 125–140

[4] Kalckert, Andreas; Ehrsson, H. Henrik (2012). Moving a Rubber Hand that Feels Like Your Own: A Dissociation of Ownership and Agency. Frontiers in Human Neuroscience, 6, 40

[5] Schwitzgebel, Eric (2008). The Unreliability of Naive Introspection. The Philosophical Review, 117(2), 245–273

[6] Fleming, Stephen M.; Lau, Hakwan C. (2014). How to measure metacognition. Frontiers in Human Neuroscience, 8, 443

[7] Machamer, Peter; Darden, Lindley; Craver, Carl F. (2000). Thinking about Mechanisms. Philosophy of Science, 67(1), 1–25

[8] Petitmengin, Claire (2006). Describing one's subjective experience in the second person: An interview method for the science of consciousness. Phenomenology and the Cognitive Sciences, 5, 229–269

[9] Logothetis, Nikos K.; Pauls, Jon; Augath, Mark; Trinath, Torsten; Oeltermann, Axel (2001). Neurophysiological investigation of the basis of the fMRI signal. Nature, 412, 150–157

[10] Nosek, Brian A.; Ebersole, Charles R.; DeHaven, Alexander C.; Mellor, David T. (2018). The preregistration revolution. Proceedings of the National Academy of Sciences, 115(11), 2600–2606

[11] Simmons, Joseph P.; Nelson, Leif D.; Simonsohn, Uri (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological Science, 22(11), 1359–1366

[12] Hodgkin, Alan L.; Huxley, Andrew F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. The Journal of Physiology, 117(4), 500–544

[13] Sperling, George (1960). The information available in brief visual presentations. Psychological Monographs: General and Applied, 74(11), 1–29

[14] Simons, Daniel J.; Chabris, Christopher F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. Perception, 28(9), 1059–1074

[15] Samad, Majed; Chung, Albert Jin; Shams, Ladan (2015). Perception of Body Ownership Is Driven by Bayesian Sensory Inference. PLOS ONE, 10(2), e0117178

[16] Dehaene, S.; Kerszberg, M.; Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. Proceedings of the National Academy of Sciences 95(24):14529–14534

[17] Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. Trends in Cognitive Sciences 10(11):494–501

[18] Fahrenfort, J. J.; Scholte, H. S.; Lamme, V. A. F. (2007). Masking disrupts reentrant processing in human visual cortex. Journal of Cognitive Neuroscience 19(9):1488–1497

[19] Albantakis, L.; Barbosa, L.; Findlay, G.; Grasso, M.; Haun, A. M.; Marshall, W.; Mayner, W. G. P.; Zaeemzadeh, A.; Boly, M.; Juel, B. E.; Sasai, S.; Fujii, K.; David, I.; Hendren, J.; Lang, J. P.; Tononi, G. (2023). Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms. PLOS Computational Biology 19(10):e1011465

[20] Lau, H. C. (2008). A higher order Bayesian decision theory of consciousness. Progress in Brain Research 168:35–48

[21] Rao, R. P. N.; Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience 2(1):79–87

[22] Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. Trends in Cognitive Sciences 17(11):565–573

[23] Cogitate Consortium; Ferrante, O.; Górska-Klimowska, U.; Henin, S.; Hirschhorn, R.; et al. (2025). Adversarial testing of global neuronal workspace and integrated information theories of consciousness. Nature 642:133–142

[24] Cohen, M. A.; Dembski, C.; Ortego, K.; Steinhibler, C.; Pitts, M. (2024). Neural signatures of visual awareness independent of postperceptual processing. Cerebral Cortex 34(11):bhae415

[25] He, L.; Zhang, Z.; Wang, Y.; Meng, Q.; Jiang, Y.; Xu, W.; Wu, J.; Sheng, J.; Ye, Z.; Ding, X. (2026). Report-modulated prefrontal activity and consistent posterior representations during conscious visual perception. Consciousness and Cognition 142:104075

[26] Zhang, Y.; Li, X.; Jin, Z.; Zhang, J.; Li, L. (2026). Distinct frontal lobe subregions mediate the emergence and reporting of visual consciousness. NeuroImage 334:121964

[27] Noorman, S.; Stein, T.; Zantvoord, J.; Fahrenfort, J.; van Gaal, S. (2025). A causal role of the NMDA receptor in recurrent processing during perceptual integration. eLife 13:RP100530

[28] Kouider, S.; Stahlhut, C.; Gelskov, S. V.; Barbosa, L. S.; Dutat, M.; de Gardelle, V.; Christophe, A.; Dehaene, S.; Dehaene-Lambertz, G. (2013). A neural marker of perceptual consciousness in infants. Science 340(6130):376–380

[29] Amsterdam, B. (1972). Mirror self-image reactions before age two. Developmental Psychobiology 5(4):297–305

[30] Goupil, L.; Romand-Monnier, M.; Kouider, S. (2016). Infants ask for help when they know they don't know. Proceedings of the National Academy of Sciences 113(13):3492–3496
[31] Felsenstein, J. (1985). Phylogenies and the Comparative Method. The American Naturalist 125(1):1–15
[32] Barron, A. B.; Klein, C. (2016). What insects can tell us about the origins of consciousness. Proceedings of the National Academy of Sciences 113(18):4900–4908
[33] Andrews, K.; Birch, J.; Sebo, J.; et al.; individual signatories (2024). The New York Declaration on Animal Consciousness. 공식 성명서, 2024년 사월 19일
[34] Hampton, R. R. (2001). Rhesus monkeys know when they remember. Proceedings of the National Academy of Sciences 98(9):5359–5362
[35] Nieder, A.; Wagener, L.; Rinnert, P. (2020). A neural correlate of sensory consciousness in a corvid bird. Science 369(6511):1626–1629
[36] Crook, R. J. (2021). Behavioral and neurophysiological evidence suggests affective pain experience in octopus. iScience 24(3):102229
[37] Barr, S.; Elwood, R. W. (2024). Trade-Offs between Avoidance of Noxious Electric Shock and Avoidance of Bright Light in Shore Crabs Are Consistent with Predictions of Pain. Animals 14(5):770
[38] Manzi, O.; Lynch, K. E.; Allman, D. M.; Latty, T.; White, T. E. (2026). Flexible self-protection as evidence of pain-like states in house crickets. Proceedings of the Royal Society B 293(2070):20260609
[39] Toyota, M.; Spencer, D.; Sawai-Toyota, S.; et al. (2018). Glutamate triggers long-distance, calcium-based plant defense signaling. Science 361(6407):1112–1115
[40] Boisseau, R. P.; Vogel, D.; Dussutour, A. (2016). Habituation in non-neural organisms: evidence from slime moulds. Proceedings of the Royal Society B 283(1829):20160446
[41] Markel, K. (2020). Lack of evidence for associative learning in pea plants. eLife 9:e57614
[42] Butlin, P.; Long, R.; Bayne, T.; Bengio, Y.; Birch, J.; Chalmers, D.; Constant, A.; Deane, G.; Elmoznino, E.; Fleming, S. M.; Ji, X.; Kanai, R.; Klein, C.; Lindsay, G.; Michel, M.; Mudrik, L.; Peters, M. A. K.; Schwitzgebel, E.; Simon, J.; VanRullen, R. (2026). Identifying indicators of consciousness in AI systems. Trends in Cognitive Sciences 30(6):488–501
[43] Wilson R. C., Collins A. G. E. (2019). Ten simple rules for the computational modeling of behavioral data. eLife 8, e49547
[44] Kagan B. J. et al. (2022). In vitro neurons learn and exhibit sentience when embodied in a simulated game-world. Neuron 110(23), 3952–3969.e8
[45] Cai H. et al. (2023). Brain organoid reservoir computing for artificial intelligence. Nature Electronics 6, 1032–1039
[46] Robbins A. et al. (2026). Goal-directed learning in cortical organoids. Cell Reports 45(2), 116984

[47] Robbins A. (2025). Goal-Directed Learning in Cortical Organoids — Experimental Data and Code. Zenodo, version v2, doi:10.5281/zenodo.17684862

[48] International Society for Stem Cell Research (2025). Guidelines for Stem Cell Research and Clinical Translation. Version 1.2, August 2025 update

[49] Lindsey J. (2025). Emergent Introspective Awareness in Large Language Models. Transformer Circuits 연구 보고서, 2025년 십월 29일; arXiv:2601.01828 프리프린트, 2026년 일월 5일

# 저자와 프로젝트

Eudgene Brenev

발견한 오류를 알리거나, 책의 내용을 논의하거나, 저자의 프로젝트를 살펴보려면 다음 주소를 이용해 주십시오.

**이메일**
이메일: info@brenev.com

**저자 웹사이트**
저자 웹사이트: brenev.com

**미라 웹사이트**
미라 웹사이트: mira.technology

**다른 프로젝트**
다른 프로젝트: haraizen.tech

**텔레그램**
https://t.me/Eudgenee

이 책의 제목은 저자의 인공지능 프로젝트 «미라»를 가리킨다.