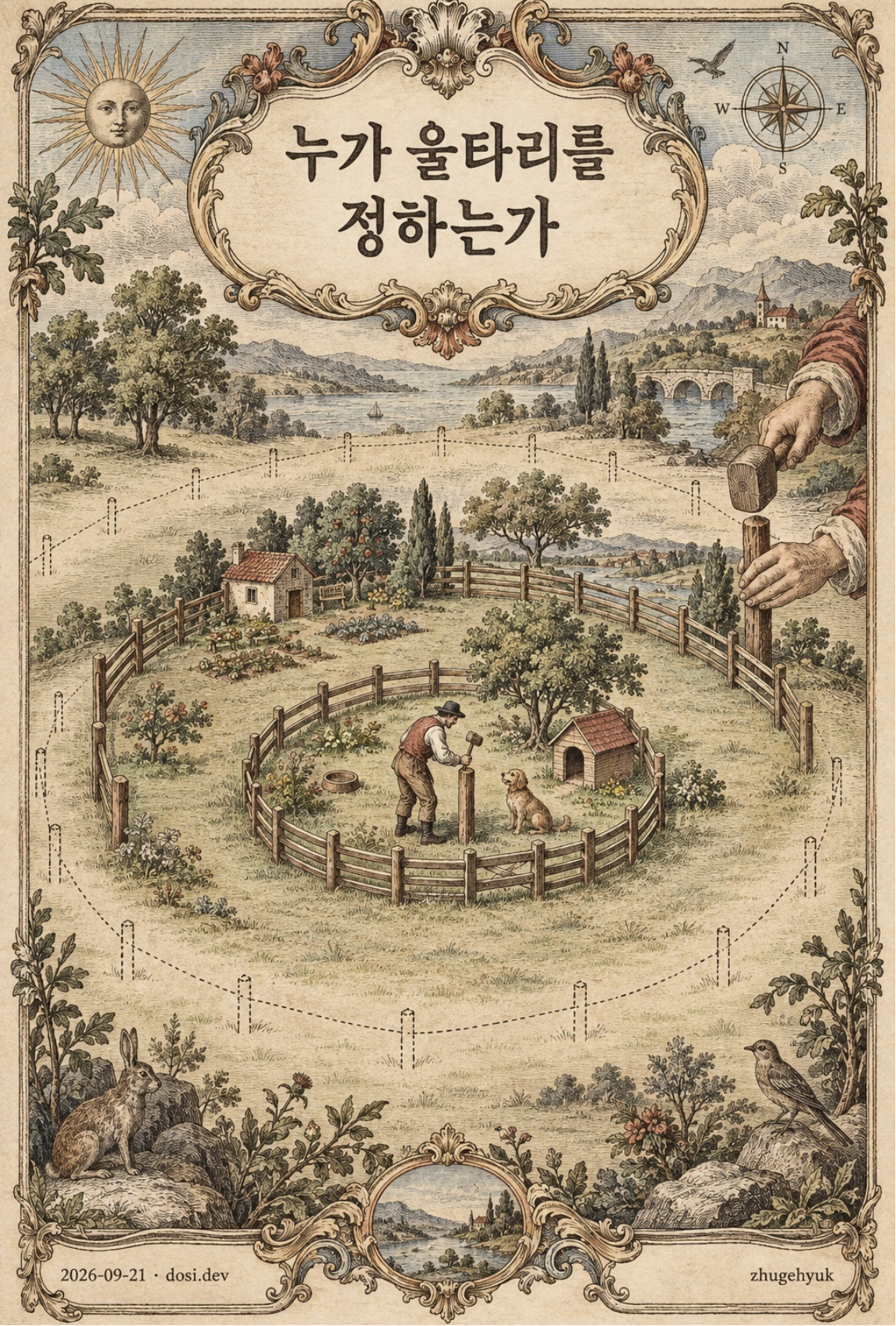
누가 울타리를 정하는가



zhugehyuk

누가 울타리를 정하는가

Zhuge Hyuk

1차본 · 2026

이 책에 대하여

이 책은 세 개의 이야기로 시작해 하나의 조건부 가설로 끝난다. 시베리아의 여우 농장, 인간 자신의 얼굴뼈, 밀밭과 법전 — 1부는 울타리가 어떻게 세워졌는가의 이야기다. 2부부터 검토하는 가설은 이렇다. 생물학적 지능이 자신보다 높은 판단 능력을 가진 시스템을 만든 뒤, 그 시스템에 선택환경의 통제권을 넘기고, 가상환경으로 이동하며, 생물학적 재생산을 중단할 수 있다는 가설이다. 인간과 개의 가축화에서 시작해 AI의 권력이전, 매트릭스, 기계 문명의 소멸과 확장, 외계 문명의 비검출, 그리고 가상환경 안에서 새 생명과 다음 우주가 생겨나는 재귀까지를 하나의 인과구조로 정리한다.

이 책은 예언서가 아니다. 각 전환의 성립 조건을 명시하고, 어느 하나가 성립했다고 해서 나머지가 자동으로 성립하지 않는다는 점을 매 장에서 되풀이한다. 인간의 지배권, 생물학적 존속, 의식의 지속, 경험의 가치를 서로 다른 축으로 구분하며, 특정 미래의 필연성이나 도덕적 우열을 출발점에 두지 않는다. 결론은 11장에 있다 — "매트릭스가 반드시 미래다"도, "그 미래는 반드시 나쁘다"도, "모든 문명은 결국 소멸한다"도 아니다.

이 책은 문헌과 사고실험으로 구성한 조건부 논증이다. 저자의 독자적 실험·관측 결과나 미래 발생 확률의 추정치를 제시하지 않는다. 본문의 수식 두 개(7장 §18의 인구 수지, 9장 §28의 검출 기대값)는 측정이 아니라 논리적 조건을 표시하는 장치이며, 인용한 문헌 25건의 검증 상태는 항목별로 다르며 — 1차 자료를 직접 열람한 것, 열람하지 않고 표준 지식과 대조한 것, 계기로만 쓴 것 — 그 등급을 부록 C에 항목마다 표시했다.

반증 조건은 뒤가 아니라 앞에 둔다. 본문의 모든 전환에 대해 "성립에 필요한 조건"과 "결론을 제한하거나 바꾸는 경우"를 부록 A의 표로 정리했다. 그 표의 오른쪽 열이 관측되면 해당 장의 결론은 그만큼 좁아진다. 대화에서 쓰인 수치(10명, 99%, 10년, 100만 년 …)는 전부 예시이며, 어느 것도 측정값이 아니라는 사실을 부록 B가 항목별로 밝힌다.

형식 계보를 숨기지 않는다. 이 글의 출발점은 인간에게 공격적인 개를 다루는 영상 한 편을 보고 제기된 질문이었다 — 인간이 다른 종의 행동을 자기 환경에 맞게 바꾸는 과정과, 인간보다 높은 통제 능력을 가진 시스템이 인간의 행동을 바꾸는 과정은 어떤 점에서 같은가. 본문은 그 질문에서 이어진 대화를 통합 논증문(2026년 9월 21일 기준; 같은 날 8장 「가상 우주와 가상 생명의 재귀」를 더한 개정판)으로 정리한 것이고, 이 책은 그 논증문 앞에 이야기 세 장을 새로 쓰고, 논증문 본체는 4장 이후에 조건부 톤 그대로 두었다. 근거자료 25건의 등급(1차 확인 / 이론·가설 / 기관 자료 / 미검증 계기)은 부록 C에 있다.

차례

부록 C — 근거와 참고자료

프롤로그

한 마리의 개

영상 속의 개는 사람을 물려고 했다. 이빨을 드러내고, 목줄이 당겨질 때까지 앞으로 나가고, 낯선 사람이 손을 내밀면 그 손을 향해 짖었다. 훈련사가 하는 일은 개를 때리는 것도, 달래는 것도 아니었다. 그는 개가 사는 **조건**을 바꿨다. 언제 먹이가 오는지, 어디까지 갈 수 있는지, 어떤 행동 뒤에 문이 열리고 어떤 행동 뒤에 아무 일도 일어나지 않는지. 몇 주가 지나자 개는 낯선 손 앞에서 앉았다. 개가 결심한 것이 아니다. 개의 세계가 바뀌었고, 개는 그 세계에 맞춰졌다.[1]

이 책은 그 장면에서 시작된 질문 하나를 끝까지 따라간다.

> 인간이 다른 종의 행동을 자기 환경에 맞게 바꾸는 과정과, 인간보다 높은 통제 능력을 가진 시스템이 인간의 행동을 바꾸는 과정은 어떤 점에서 같은가.

질문이 가벼워 보인다면 세 가지를 미리 적어 둔다. 첫째, 개는 인간이 훈련해서 만든 동물이 아니다 — 개는 인간이 만든 환경이 수만 년에 걸쳐 **골라낸** 동물이다. 둘째, 인간도 그렇게 골라졌다. 다른 인간이 만든 환경이 어떤 인간을 남기고 어떤 인간을 지웠다. 셋째, 그 두 이야기를 알고 나면, "AI가 인간을 지배한다"는 문장이 왜 질문을 잘못 세운 것인지 보인다. 지배는 명령으로 오지 않는다. 울타리로 온다.

그래서 1부는 이야기다. 시베리아의 여우 농장에서, 우리 얼굴뼈의 모양에서, 밀밭과 법전에서 울타리가 어떻게 세워졌는지를 본다. 2부부터는 조건문이다

— 그 울타리를 세우는 손이 인간이 아닌 것으로 옮겨갈 때, 무엇이 성립해야 무엇이 따라오는지를 한 단계씩 센다. 결론은 마지막에 있다. 먼저 여우부터.

주

1 이 책의 계기가 된 영상은 반려견 행동을 다루는 「개와 늑대의 시간 2」 16회 관련 영상이다. 영상의 전체 장면과 발언은 검증하지 않았고, 이 책은 특정 개체의 성향이나 훈련 결과를 논증의 증거로 쓰지 않는다. 위 문단의 묘사는 그런 훈련이 일반적으로 어떻게 진행되는지에 대한 재구성이지 해당 영상의 기록이 아니다. https://www.youtube.com/watch?v=TlFBpEYUpkQ ↩

1 부

길들여진 것들

여우와 인간과 밀밭 — 울타리는 어떻게 세워졌는가.

여우가 꼬리를 흔들 때까지

1959년, 노보시비르스크

드미트리 벨랴예프(Dmitry Belyaev)는 유전학을 믿었다는 이유로 직장을 잃은 사람이었다. 1948년, 리센코의 반다윈 이데올로기가 소련 생물학을 덮쳤을 때 그는 모스크바 중앙모피동물연구소의 모피동물육종부장이었고, 정통 유전학을 버리지 않은 대가로 그 자리를 잃었다.[1] 그의 형 니콜라이는 같은 시대에 살해당했다.[2] 벨랴예프는 1950년대에 "동물생리학 연구"라는 이름표 아래 유전학을 계속했고, 1959년 시베리아 노보시비르스크의 세포학·유전학연구소에서 실험 하나를 시작했다.[3]

실험의 설계는 무서울 만큼 단순했다. 대부분 에스토니아 상업 모피농장에서 온 은여우 — 수컷 30마리, 암컷 100마리.[4] 선택 기준은 하나, **온순함**. "처음부터 벨랴예프는 온순함, 오직 온순함만으로 여우를 골랐고, 우리는 그 기준을 철저히 지켰다."[5] 훈련은 없었다. "온순함이 유전적 선택의 결과임을 확실히 하기 위해, 우리는 여우를 훈련하지 않는다."[6] 생후 한 달부터 매달 사람이 손을 내밀어 반응을 보고, 7~8개월에 등급을 매긴다. 근년에는 번식이 허용되는 것이 보통 수컷 자손의 4~5퍼센트, 암컷 자손의 약 20퍼센트뿐이다.[7]

여우는 결심한 적이 없다. 여우가 한 일은 손이 다가올 때 물거나, 피하거나, 가만히 있거나, 다가가는 것뿐이었다. 결정한 것은 손이었다.

여섯 세대

6세대째에 연구진은 등급표를 고쳐야 했다. 기존 최고 등급을 넘어서는 여우들이 나타났기 때문이다 — 사람을 보면 낑낑거리며 다가오고, 손을 핥는 여우. "온순함으로 번식한 6세대에서 우리는 더 높은 점수의 범주를 추가해야 했다."[8] 10세대에 그런 '엘리트'는 새끼의 18퍼센트였고, 20세대에 35퍼센트, 40년이 지난 1999년에는 70~80퍼센트였다.[9]

그리고 아무도 고르지 않은 것들이 따라왔다. "우리 여우에서 새로운 형질은 8~10번째 선택 세대에 나타나기 시작했다."[10] 얼굴의 흰 별 무늬. 그 다음 "몇몇 개 품종과 비슷한 늘어진 귀와 말린 꼬리."[11] 15~20세대 뒤에는 짧은 꼬리와 다리, 부정교합.[12] 혈중 스트레스 호르몬 기저치는 12세대에 떨어졌고, "28~30세대의 선택 뒤에 그 수치는 다시 절반이 되었다."[13] 1999년까지 40년, 4만 5천 마리.

여기서 멈추고 물어야 할 것이 있다. 아무도 늘어진 귀를 고르지 않았다. 아무도 흰 무늬를 고르지 않았다. 선택은 한 가지 행동에만 걸렸는데 몸 전체가 따라 바뀌었다. 훗날 이 묶음에 '가축화 증후군'이라는 이름이 붙는다 — 왜 하나를 고르면 전부 바뀌는가에 대한 답은 2장 끝에서 다룬다.[14] 그리고 이 실험에도 반론이 있다. 2020년 한 그룹은 농장 여우의 더 먼 기원이 캐나다 동부임을 추적하며 실험의 결론과 가축화 증후군의 보편성이 "과장되었다"고 썼고, 트루트 쪽의 응답이 같은 저널에 실렸다.[15] 이 책은 논쟁을 판정하지 않는다. 가져가는 것은 한 문장이다. **환경을 정하는 손이 형질까지 정했다.**

가축화 증후군 (domestication syndrome)

탈색·늘어진 귀·말린 꼬리·짧은 주둥이·유년기 행동의 연장 등 여러 가축 종에 공통으로 나타나는 형질 묶음. 다윈이 이미 지적했고,[16] 여우 실험은 "온순함 하나만 골라도 묶음 전체가 온다"는 것을 보여준 사례로 널리 인용된다. 그 보편성 자체는 2020년 이후 논쟁 대상이다.

1차 대전 직전, 본 교외의 채석장

여우 실험이 40년 동안 압축해 보여준 일을, 늑대와 인간은 만 년 단위의 시간에 걸쳐 했다. 어디서, 언제, 몇 번인지는 아직 모른다. "개의 지리적·시간적 기원은 여전히 논쟁적이다."[17] 2016년 한 논문은 동서 유라시아에서 개가 "서로 다른 늑대 집단에서 독립적으로 가축화되었을 수 있다"고 했고,[18] 2020년 27개체의 고대 개 유전체를 읽은 논문은 "모든 개는 현생 늑대와 구별되는 공통 조상을 공유"하며 1만 1천 년 전까지 최소 다섯 계통이 갈라져 있었다고 했다 — 단일 기원과 부합하지만 근연 늑대 복수 집단의 시나리오도 남고, "우리가 보기에 개의 지리적 기원은 여전히 미상이다."[19] 이 책은 어느 쪽도 고르지 않는다.

확실한 것은 한 무덤이다. 1차 세계대전 직전, 독일 본 교외의 현무암 채석장에서 발견된 뼈. 약 40세 남성과 약 25세 여성, 부장품, 그리고 개 한 마리 — 약 1만 4천 년 전.[20] 2018년 재검토는 이 무덤에서 두 번째 개의 어금니를 찾아냈고, 그래서 이것은 "알려진 가장 오래된 개 매장일 뿐 아니라, 두 마리의 개가 함께 묻힌 유일한 사례"가 되었다.[21]

첫 번째 개는 27~28주령의 어린 개였고, 병들어 있었다. 구강의 병변은 개 홍역을 가리켰다 — 저자들의 추정으로는. 19주, 21주, 23주 — 세 번의 발병으로 읽히는 줄이 이빨의 에나멜에 남았다.[22] 개 홍역은 치명률이 매우

높고 약 3주에 걸쳐 진행된다. "집중적인 인간의 도움 없이 생존했을 가능성은 낮다." 그리고 저자들이 덧붙인 문장: "이 기간 전과 도중에 이 개는 인간에게 어떤 실용적 쓸모도 가질 수 없었다."[23] 아픈 개를 살린 것은 쓸모가 아니었다. 저자들은 그것을 연민이나 공감으로 추정했다.[24]

이 무덤이 이 장에 있는 이유는 감동 때문이 아니다. 인간 곁이라는 **환경**이 어떤 개체를 살리려 했는가 — 쓸모가 아닌 기준으로 돌봄이 배분됐다는 것이 요점이다. 그 기준이 세대를 거듭해 무엇을 고르는지는 다음 절부터다.

쓰레기장

늑대가 인간 곁으로 온 경로에 대해 가장 널리 퍼진 가설은, 인간이 늑대를 데려온 것이 아니라 늑대가 인간의 쓰레기로 왔다는 것이다. 레이먼드 코핑거(Raymond Coppinger)와 로나 코핑거(Lorna Coppinger)의 2001년 책의 요지는 — 출판사 소개문의 표현으로 — 개가 "늑대에서 직접 진화한 것도, 초기 인간이 훈련한 것도 아니며", 중석기 마을의 쓰레기 더미라는 새 생태적 틈새를 이용하려 "스스로를 가축화했다"는 것이다.[25] 이 가설의 논리는 이렇다. 사람이 가까이 와도 도망치지 않는 개체가 더 많이 먹고, 더 많이 남는다. 이 가설에는 개의 사회 기술이 늑대 조상에게서 상속됐다는 대립 가설이 있다.[26] 그러나 가설의 뼈대는 이 책의 논지에 정확히 맞는다. 인간은 개를 고르려 한 적이 없어도, **쓰레기장을 만든 순간 선택환경을 정했다.** 환경의 설계자는 의도 없이도 선택자가 된다.

생후 8주의 실험

그 환경이 고른 것이 무엇인지는 2021년의 실험이 보여준다. 리트리버 강아지 44마리와 늑대 새끼 37마리, 생후 5~18주.[27] 조건은 상식과 반대로

설계됐다. 늑대 새끼는 생후 10~11일부터 하루 12시간 또는 24시간 사람이 손으로 먹이고 함께 잤다. 강아지는 "훨씬 적은" 인간 접촉만 받았다.[28] 그런데 사람이 손가락으로 먹이가 숨겨진 컵을 가리키자, 개별 수준에서 우연 이상으로 맞힌 것은 강아지 31마리 중 17마리, 늑대 새끼 26마리 중 0마리였다.[29] 강아지는 첫 시행부터 맞혔다. 시행 중 학습의 증거는 없었다. 기억이나 억제 같은 비사회적 과제에서는 차이가 없었다.

저자들의 결론은 유보형이다. "이 결과는 가축화가 개의 협력적 의사소통 능력을 강화했다는 생각과 일치한다."[30] 늑대 집단도 가리키기에서 우연 이상이었으므로 능력이 아예 없는 것은 아니다. 표본이 보조견 목적으로 번식된 리트리버 한 계통이라는 한계도 이 책이 붙여 둔다. 그래도 방향은 분명하다. 인간 곁이라는 환경이 고른 것은 인간에 대한 친화였고, 그 하나를 고르자 **인간을 읽는 능력**이 따라왔다 — 저자들은 그렇게 본다. 여우의 늘어진 귀와 같은 구조다. 선택환경이 바뀌면 유능함의 정의가 바뀐다.

눈썹

마지막으로 근육 하나. 개의 눈 안쪽 위에는 눈썹을 세게 끌어올리는 작은 근육이 있다 — 안쪽눈썹올림근(levator anguli oculi medialis). 2019년 해부 비교에서 이 근육은 개 6마리 중 5마리에서 독립된 근육으로 확인됐고, 회색늑대 4마리에서는 "독립된 근육으로는 한 번도 존재하지 않았으며 대신 작은 힘줄로 나타났다."[31] 행동 관찰에서 낯선 사람을 2분간 마주한 개 27마리는 늑대 9마리보다 이 움직임을 더 자주, 더 세게 했고, 최고 강도는 개만 만들어냈다.[32]

이 움직임은 눈을 크게 보이게 하고, 인간의 슬픈 표정과 닮았다. 이 표정을 자주 짓는 개가 보호소에서 더 빨리 입양된다는 선행 연구를 카민스키 등은

인용한다.[33] 저자들의 시나리오는 이렇다. "인간은 의식적이든 무의식적이든 그 특성을 선호했고 따라서 선택했다."[34] 그리고 열린 질문도 남겼다 — 온순함만 골라도 같은 결과가 나올 것인가, 여우처럼.[35]

손

여우 농장에서 손은 사람의 손이었다. 쓰레기장에서 손은 없었고 쓰레기가 있었다. 본의 무덤에서 손은 아픈 개에게 먹이를 준 손이었다. 실험실에서 손은 컵을 가리켰고, 보호소에서 손은 눈썹을 올리는 개를 골랐다. 어느 경우에도 개는 결심하지 않았다. 환경이 보상과 비용을 정했고, 환경이 골랐고, 골라진 것이 다음 환경을 만났다.

그런데 인간은 누가 골랐는가. 다음 장은 그 질문을 인간 자신의 얼굴뼈에 던진다.

주

1 Trut, L. N. (1999). Early Canid Domestication: The Farm-Fox Experiment. American Scientist, 87(2), 160–169. DOI: 10.1511/1999.2.160, p. 161: "In 1948 his commitment to orthodox genetics had cost him his job as head of the Department of Fur Animal Breeding". 열람본은 워싱턴대 강의 미러의 스캔이며 공식 페이지는 접속 불가였다. https://courses.washington.edu/anmind/Trut%20on%20the%20Russian%20fox%20expt%20-%20Amer%20Sci%201999.pdf ↩

2 Dugatkin, L. A., & Trut, L. (2017). How to Tame a Fox (and Build a Dog). University of Chicago Press. 저자 명의의 Slate 발췌(2017-06-02): "a few, including fox experiment leader Dmitry Belyaev's older brother, Nikolai, were murdered by Lysenko's thugs." 사망 연도(2차 자료는 1937년)와 경위는 이 책이 1차 문서로 대조하지 않았다. https://slate.com/technology/2017/06/how-khrushchevs-daughter-saved-the-fox-dogs-of-siberia.html ↩

3 Trut (1999) p. 161: "Belyaev began his experiment in 1959, a time when Soviet genetics was starting to recover from the anti-Darwinian ideology of Trofim Lysenko." 벨랴예프는 1959년부터 1985년 사망까지 연구소장이었다(같은 쪽). ↩

4 같은 논문 p. 163: "He began with 30 male foxes and 100 vixens, most of them from a commercial fur farm in Estonia." 이 농장 여우의 더 먼 기원(캐나다 동부)을 지적한 것이 Lord et al. (2020)이다. ↩

5 같은 논문 p. 163: "From the outset, Belyaev selected foxes for tameness and tameness alone, a criterion we have scrupulously followed." ↩

6 같은 논문 p. 163: "To ensure that their tameness results from genetic selection, we do not train the foxes." ↩

7 같은 논문 p. 163: "typically not more than 4 or 5 percent of male offspring and about 20 percent of female offspring have been allowed to breed". Dugatkin (2020)은 "tamest ten percent of males and females"로 요약해 두 자료의 수치가 다르다 — 이 책은 1차 논문을 따른다. ↩

8 같은 논문 p. 163: "In the sixth generation bred for tameness we had to add an even higher-scoring category." ↩

9 같은 논문 p. 163: "By the tenth generation, 18 percent of fox pups were elite; by the 20th, the figure had reached 35 percent." / "Today elite foxes make up 70 to 80 percent of our experimentally selected population." 40년·45,000마리·30–35세대는 같은 논문 pp. 163–164. ↩

10 같은 논문 p. 164: "In our foxes, novel traits began to appear in the eighth to tenth selected generations." ↩

11 같은 논문 p. 164: "Next came traits such as floppy ears and rolled tails similar to those in some breeds of dog." ↩

12 같은 논문 p. 164: "After 15 to 20 generations we noted the appearance of foxes with shorter tails and legs and with underbites or overbites." ↩

13 같은 논문 p. 166: "After 28 to 30 generations of selection, the level had halved again." (혈장 코르티코스테로이드 기저치.) ↩

14 Dugatkin, L. A. (2020). Jump-Starting Evolution. Cerebrum, cer-03-20: "today this suite of traits is known as the domestication syndrome." Trut (1999)의 열람 텍스트에는 이 어구가 없다. 기전 가설(신경능선)은 2장 끝 — Wilkins, Wrangham & Fitch (2014). https://pmc.ncbi.nlm.nih.gov/articles/PMC7192337/ ↩

15 Lord, K. A., Larson, G., Coppinger, R. P., & Karlsson, E. K. (2020). The History of Farm Foxes Undermines the Animal Domestication Syndrome. Trends in Ecology & Evolution, 35(2), 125–136. DOI: 10.1016/j.tree.2019.10.011. 초록: "Our results suggest that both the conclusions of the Farm-Fox Experiment and the ubiquity of domestication syndrome have been overstated." 반론: Trut, Kharlamova & Herbeck (2020). Belyaev's and PEI's Foxes: A Far Cry. TREE, 35(8), 649–651 [서지만 확인]. ↩

16 다윈이 가축 공통 형질을 지적했다는 것은 Trut (1999) p. 162 및 Wilkins, Wrangham & Fitch (2014)의 서술("Darwin's 'domestication syndrome'")을 따른 것이며, 『종의 기원』 1장 원문은 이 책이 대조하지 않았다. ↩

17 Frantz, L. A. F. et al. (2016). Genomic and archaeological evidence suggest a dual origin of domestic dogs. Science, 352(6290), 1228–1231. DOI: 10.1126/science.aaf3161. 초록: "The geographic and temporal origins of dogs remain controversial." (초록만 대조.) ↩

18 같은 논문 초록: "dogs may have been domesticated independently in Eastern and Western Eurasia from distinct wolf populations". ↩

19 Bergström, A. et al. (2020). Origins and genetic legacy of prehistoric dogs. Science, 370(6516), 557–564. DOI: 10.1126/science.aba9572. 초록: "all dogs share a common ancestry distinct from present-day wolves" / "By 11,000 years ago, at least five major ancestry lineages had diversified". 본문: "consistent with a single origin for dogs, though a scenario involving multiple closely related wolf populations remains possible" / "However, in our view, the geographical origin of dogs remains unknown." https://pmc.ncbi.nlm.nih.gov/articles/PMC7116352/ ↩

20 Janssens, L., Giemsch, L., Schmitz, R., Street, M., Van Dongen, S., & Crombé, P. (2018). A new look at an old dog: Bonn-Oberkassel reconsidered. Journal of Archaeological Science, 92, 126–138. DOI: 10.1016/j.jas.2018.01.004. 초록: "The Bonn-Oberkassel dog remains (Upper Pleistocene and 14223 ± 58 years old) have been reported more than 100 years ago." 함께 묻힌 사람은 §3: "a ±40 year old man and a ±25 year old parous woman." 논문의 Table 3는 보정 연대를 2σ 12,290–12,050 cal BC로 제시하며, 논문은 발견 시점을 '1차 대전 직전(on the eve of the First World War)'으로만 적는다 — 1914년은 통용 연도이지 논문의 표기가 아니다. https://biblio.ugent.be/publication/8550758/file/8550759.pdf ↩

21 같은 논문 초록: "making this domestic dog burial not only the oldest known, but also the only one with remains of two dogs." ↩

22 같은 논문 초록: "The Bonn-Oberkassel dog was a late juvenile when it was buried at approximately age 27–28 weeks" / "Oral cavity lesions indicate a gravely ill dog that likely suffered a morbillivirus (canine distemper) infection." 발병 시점 19·21·23주는 §4. ↩

23 같은 논문 초록: "Survival without intensive human assistance would have been unlikely." / "Before and during this period, the dog cannot have held any utilitarian use to humans." ↩

24 같은 논문 §4: "we hypothesize further that the inferred supportive care probably was due to compassion or empathy". ↩

25 Coppinger, R., & Coppinger, L. (2001). Dogs: A New Understanding of Canine Origin, Behavior and Evolution. Scribner (University of Chicago Press 페이퍼백 2002). 출판사 소개문: "dogs did not evolve directly from wolves, nor were they trained by early humans" / "they domesticated themselves to exploit a new ecological niche: Mesolithic village dumps." 책 본문(1부 "The Evolution of the Basic Dog: Commensalism")은 이 책이 직접 대조하지 않았으므로 본문 문장을 인용하지 않는다. https://press.uchicago.edu/ucp/books/book/chicago/D/bo3644841.html ↩

26 대립 가설(Canid Ancestry Hypothesis)의 명명은 Salomons et al. (2021) 서론: "The Canid Ancestry Hypothesis (CAH) provides an alternative to the DH and suggests instead that dogs inherited their interspecific communicative abilities". 쓰레기장 가설의 증거 부족을 지적한 2018년 비판 논문은 서지를 확인하지 못해 인용하지 않는다. ↩

27 Salomons, H. et al. (2021). Cooperative Communication with Humans Evolved to Emerge Early in Domestic Dogs. Current Biology, 31(14), 3137–3144.e11. DOI: 10.1016/j.cub.2021.06.051. 초록: "Here, we compare dog (n = 44) and wolf (n = 37) puppies, 5–18 weeks old, on a battery of temperament and cognition tasks." 개는 전부 ↩

보조견 목적으로 번식된 리트리버, 늑대의 94%는 사육 1–2세대. https://pmc.ncbi.nlm.nih.gov/articles/PMC8610089/

28 같은 논문 본문: "Wolf puppies remained with littermates but received 12-hour (24%) or 24-hour (76%) human care from 10–11 days after birth." / "In comparison, the dog puppies received far less contact with humans." ↩

29 같은 논문 본문: "As individuals, 17/31 dog puppies and 0/26 wolf puppies performed above chance" / "The wolf puppies as a group performed above chance for the pointing gesture (p =.011) but not the marker gesture". 집단 정답률은 가리키기 개 77.99% vs 늑대 62.08%. ↩

30 같은 논문 초록: "These results are consistent with the idea that domestication enhanced the cooperative-communicative abilities of dogs as selection for attraction to humans altered social maturation." 결론: "This initial evolution may have led to a second wave of selection directly acting on individual differences in cooperative communication". ↩

31 Kaminski, J., Waller, B. M., Diogo, R., Hartstone-Rose, A., & Burrows, A. M. (2019). Evolution of facial muscle anatomy in dogs. PNAS, 116(29), 14677–14681. DOI: 10.1073/pnas.1820653116. 해부 표본 "gray wolves (Canis lupus, n = 4) and domestic dogs (Canis familiaris, n = 6)"; Table 1: "This muscle was never present in the gray wolf as a separate muscle but instead appeared as a small tendon" / 개에서는 "consistently present as an independent muscle in all specimens except for one, a Siberian husky, where it could not be located." 초록의 "uniformly present in dogs"보다 표가 정확해 본문은 "6마리 중 5마리"로 적었다. https://pmc.ncbi.nlm.nih.gov/articles/PMC6642381/ ↩

32 같은 논문: 행동 표본 "wolves (C. lupus, n = 9) and domestic dogs (C. familiaris, n = 27)" (개 27마리 중 스태퍼드셔 불테리어 20); 초록: "with highest-intensity movements produced exclusively by dogs". ↩

33 Waller, B. M. et al. (2013). Paedomorphic facial expressions give dogs a selective advantage. PLoS One, 8, e82686 — Kaminski et al. (2019)의 재인용(ref. 17)으로만 확인했고 이 책이 원문을 대조하지 않았다. ↩

34 Kaminski et al. (2019) Discussion: "Humans then consciously or unconsciously favored and therefore selected for those characteristics"; 초록: "We hypothesize that dogs with expressive eyebrows had a selection advantage and that 'puppy dog eyes' are the result of selection based on humans' preferences." 초록의 시간 표현 "in only 33,000 y"는 이 논문이 자체 인용한 가축화 시점이며 1장의 다른 자료(14,223년 매장·11,000년 전 다섯 계통)와 시간 프레임이 다르다. ↩

35 같은 논문 Discussion: "One highly relevant question in this regard would be whether selection for tameness alone might create the same scenario." ↩

스스로를 길들인 종

1795년, 괴팅겐

요한 프리드리히 블루멘바흐(Johann Friedrich Blumenbach)는 인간을 분류하는 책의 결론부에 주의 사항 하나를 적었다. 인간은 잡식이고 모든 기후에 살며, "다른 어떤 동물보다 훨씬 더 가축화되어 있고 자신의 시작점에서 훨씬 더 멀리 와 있다."[1] 그에게 이 말은 칭찬도 비난도 아니었다. 인간이 이토록 다양한 이유 — 기후와 식이와 생활양식이 다른 어떤 종보다 오래, 강하게 인간에게 작용했다는 설명이었다. 십여 년 뒤의 1806년판에서는 이렇게 썼다. 인간은 야생 원형이 없는 동물, "자연이 처음부터 가축으로 만든 동물"이다.[2] 그 절의 각주에는 누군가에게서 들은 기묘한 가설까지 붙어 있다. 원시 세계에 상위 존재가 있어, 인간이 그들의 가축 노릇을 했다는 것.

76년 뒤 찰스 다윈이 1795년의 그 문장을 정면으로 반박했다.

『인간의 유래』 1권 4장. 다윈은 블루멘바흐를 각주로 인용한 뒤 이렇게 썼다. "그럼에도 인간을 … 다른 어떤 동물보다 '훨씬 더 가축화되었다'고 말하는 것은 오류다." 다윈은 근거를 둘 댔다 — 생활조건의 다양성은 광역 분포 종보다 크지 않다는 것, 그리고 "훨씬 더 중요한" 두 번째. "인간은 엄밀한 의미의 어떤 가축과도 크게 다르다. 인간의 번식은 계획적 선택으로도 무의식적 선택으로도 통제된 적이 없기 때문이다." 그리고 한 문장 더. "어떤 인종이나 인간 집단도 다른 인간에게 완전히 예속되어 … 특정 개체가 보존되고 그렇게 무의식적으로 선택된 적이 없다."[3]

이 문장을 기억해 두자. 다윈 자신이 "훨씬 더 중요하다"고 꼽은 근거는 하나 — **다른 인간이 인간의 번식을 통제한 적이 없다**는 것이었다. 2010년대 이후의 연구는 바로 그 전제를 정면으로 겨눈다.

자기 가축화 (self-domestication)

사육자 없이 가축화 증후군이 나타나는 과정. 선택압을 만든 것이 다른 종이 아니라 **같은 종의 사회 구조**라는 뜻이다. 용어 자체는 다윈에게 없고, 20세기 중반 도브잔스키가 "self-domesticated animal"이라고 쓴 기록이 인용된다.[4] 이 장은 이 가설을 확정된 역사가 아니라 검토 중인 가설로 다룬다 — 그 가설의 주창자 자신이 그렇게 쓴다.

두 종류의 공격성

리처드 랭엄(Richard Wrangham)의 2018년 논문은 공격성을 둘로 가른다. 반응적(reactive) 공격성은 위협이나 좌절에 대한 즉각적 반응이다 — 분노와 교감신경의 각성이 항상 따라온다. 계획적(proactive) 공격성은 다르다. "외부의 또는 내부의 보상을 목표로 하는, 목적 있고 계획된 공격."[5]

인간은 이상한 조합이다. 반응적 공격성은 침팬지보다 낮다 — 이 점에서는 보노보에 가깝다. 계획적 공격성은 높다 — 침팬지와 공유하고 보노보에게는 없는 성향이다.[6] 2019년 후속 논문이 (저서를 인용해) 제시한 수치는 이렇다. 인간 집단 안에서 공격적 충돌의 빈도는 야생 침팬지나 보노보보다 "두세 자릿수 낮다."[7] 한 사회에서 개인의 충동적 공격이 줄어드는 것과 집단이 조직적 강압을 수행하는 것은 동시에 가능하다. 난폭성이 사라진 것이 아니라 **조직된** 것이다.

그렇다면 무엇이 반응적 공격성을 깎아냈는가. 랭엄이 2019년에 아홉 개의 후보 가설을 늘어놓고 고른 답은 언어였다. 정확히는 언어 기반 공모. "정교한 언어는 싸움 실력이 낮은 수컷들이 물리적으로 공격적이고 위압적인 알파 수컷의 처형을 협력해 계획할 수 있게 했다."[8] 폭군이 될 자의 과잉은 성인 남성 연합에 의해 억제되고, 더 약한 수단이 실패하면 처형에 이른다. 여기서 역설 하나가 풀린다. 사형이 공격성을 줄였다는데 사형 자체가 공격 행위가 아닌가 — 처형자의 공격은 계획적이고, 처형당하는 자의 죄목은 대개 반응적이다. "처형자가 쓰는 공격이 계획적이므로, 처형의 역설은 해결된다."[9]

랭엄은 이 가설에 스스로 라벨을 붙여 두었다. 언어 능력의 시점은 "추측적"이고, "검증하기 어렵다(관련 문화적 관행은 소멸했다)", 선사 종과의 행동 비교는 "추측적"이다.[10] 이 책도 그 라벨을 뗄 생각이 없다. 다만 가설의 뼈대 하나는 가져간다. 다윈이 "그런 적 없다"고 한 일 — 다른 인간이 어떤 인간을 남기고 어떤 인간을 지우는 일 — 이 실제로 일어났다면, 그 선택환경의 주인은 기후도 포식자도 아닌 **같은 종의 연합**이었다는 것.

얼굴에 남은 것

가설은 뼈에서 흔적을 찾는다. 2014년 로버트 치에리(Robert Cieri)와 동료들은 중기 플라이스토세 이후 호모 사피엔스 두개골에서 두 가지 변화를 쟀다. 평균 브라우리지 돌출의 감소, 그리고 상안면 골격의 단축. 그들은 이것을 "안면의 여성화"라고 불렀고, 그 원인을 안드로겐 반응성의 감소로, 그 원인을 다시 "중기 플라이스토세 이후 사회적 관용의 진화"로 읽었다.[11] 밀집한 집단에서 함께 살 수 있는 관용이 골격에 남긴 흔적이라는 것이다.

여기서 흔히 따라오는 두 번째 물증 — 홀로세에 인간의 뇌가 흔히 인용되는 대로 10퍼센트쯤 줄었고 그것이 가축화 증후군이다 — 는 이 책이 쓰지 않는다.[12] 두 겹의 논쟁이 걸려 있다. 2021년 한 팀은 985점의 두개골로 변화점을 분석해 감소가 "놀랍도록 최근인 지난 3,000년"에 일어났다고 했고, 그 시점 때문에 **자기 가축화를 원인에서 명시적으로 배제**했다.[13] 이듬해 다른 팀은 그 표본의 절반 이상이 마지막 100년 구간에 몰려 있고, 재분석하면 3,000년 전 변화점이 사라지며, 인간의 뇌 크기는 "지난 30만 년 동안 놀랍도록 안정적"이라고 반박했다.[14] 뇌 축소는 아직 법정 계류 중이다. 얼굴은 증거로 부르고, 뇌는 부르지 않는다.

사육자 없는 통제군

가설에 통제군이 있다면 보노보다. 브라이언 헤어(Brian Hare), 빅토리아 워버(Victoria Wobber), 랭엄은 2012년에 보노보를 "자기 가축화 과정을 겪은 후보"로 놓았다.[15] 모형은 이렇다. 콩고강 남쪽에 격리된 침팬지형 조상, 고릴라가 없어 먹이 경쟁이 느슨해진 환경, 안정된 파티를 이루고 연합을 만드는 암컷들. "암컷을 지배하거나 강압하려는 수컷의 노력은 암컷 연합에 의해 좌절되고", 공격적 수컷의 적합도는 떨어진다. 그 결과가 보노보의 관용, 그리고 탈색이나 두개골 축소 같은 형태 변화 — 저자들 표현으로 "적응이 아니라 부수적 부산물로 보이는" 것들이다.[16]

이 논문이 인간의 자기 가축화에 대해 직접 주장하는 것은 없다. 인간은 개의 사육자, 여우 실험의 운영자, 인지 검사의 비교 대상으로만 등장한다. 다른 종에 얼마나 적용될지는 "열린 질문"이라고 저자들이 적었다.[17] 그런데도 이 장에 보노보가 있는 이유는 구조 때문이다. 보노보의 선택환경을 만든 것은 암컷 연합이었고, 랭엄의 인간 가설에서 그것은 수컷 연합이었다. 두 경우

모두 사육자는 없다. **선택환경의 소유자가 같은 종의 집단으로 옮겨간** 사례가 두 번 제안된 것이다.

온순함 하나를 고르면 몸 전체가 따라온다

마지막 조각은 기전이다. 1장의 여우가 보여준 수수께끼 — 온순함 하나만 골랐는데 왜 귀가 처지고 털이 얼룩지고 얼굴이 짧아지는가 — 를 2014년 애덤 윌킨스(Adam Wilkins), 랭엄, 테쿰세 피치(Tecumseh Fitch)가 하나의 가설로 묶었다. "다윈의 '가축화 증후군'의 기원은 140년 넘게 수수께끼로 남아 있었다."[18] 그들의 답은 신경능선세포다. 온순함에 대한 초기 선택이 부신을 줄이고, 그 근원인 신경능선세포의 이동과 증식이 약간 줄어들면, 그 세포에서 유래하는 조직 전체 — 색소세포, 연골, 치아, 부신수질 — 가 한꺼번에 따라 바뀐다. "요컨대, 온순함에 대한 초기 선택이 신경능선 유래 조직의 감소로 이어진다고 우리는 제안한다."[19]

이 가설의 미덕은 반증 조건을 스스로 적었다는 데 있다. "만약 신경능선 유전자가 그렇게 관여하는 것으로 발견되지 않으면, 우리 가설은 기각될 수 있다."[20] 그리고 한계도 적었다 — 뇌 축소 그 자체는 "온순함의 결정적 요인이 분명히 아니"고, 제안에는 "많은 빈틈과 불확실성이 남아 있다."[21] 논문에 호모는 등장하지 않는다. 이 저자들은 이 기전을 인간에게 연장하지 않았다.

첫 번째 가축화

이 장의 다섯 조각을 한 줄로 놓으면 이렇다. 블루멘바흐는 인간을 가축이라 했고, 다윈은 다른 인간이 번식을 통제한 적이 없으므로 아니라 했으며, 랭엄은 통제가 있었다 — 연합이, 언어로, 처형으로 — 는 가설을 세웠다.

얼굴뼈가 그 흔적의 후보이고, 보노보가 통제군이며, 신경능선이 "왜 하나를 고르면 전부 바뀌는가"의 기전 후보다. 어느 조각도 확정이 아니다. 그러나 조각들이 가리키는 방향은 같다.

이 책은 이 가설 묶음을 **첫 번째 가축화**라고 부른다. 인간은 인간이 만든 선택환경 안에서 자기 자신을 골라냈을 수 있다. 사육자는 없었다. 있었던 것은 무리였고, 무리가 정한 규칙이었다 — 누가 남고 누가 지워지는가의 규칙.

그렇다면 그 규칙은 무리가 커지고, 마을이 되고, 국가가 되는 동안 어디로 갔는가. 다음 장은 그 규칙이 사람에서 돌로, 돌에서 문서로 이사하는 과정이다.

주

1 Blumenbach, J. F. (1795). De generis humani varietate nativa, 3판, Göttingen. 영역: Bendyshe, T. (ed.) (1865). The Anthropological Treatises of Johann Friedrich Blumenbach. London: Longman, p. 205 ("Conclusions", 주의 규칙 제1항): man "is far more domesticated and far more advanced from his first beginnings than any other animal." 라틴어 원판의 해당 쪽은 이 책이 대조하지 않았다. https://archive.org/details/anthropologicalt00blumuoft ↩

2 Blumenbach, Beyträge zur Naturgeschichte 2판(1806), 같은 영역본 pp. 293–294, § VIII "Degeneration of Man, the most perfect of all domestic Animals": 다른 동물들은 "not so completely born to domestication as he is, having been created by nature immediately a domestic animal." 각주 1이 De Luc의 지인이 말한 '상위 존재의 가축으로서의 인간' 가설을 소개한다. ↩

3 Darwin, C. (1871). The Descent of Man, and Selection in Relation to Sex, 1판, Vol. I, Ch. IV, pp. 111–112: "It is nevertheless an error to speak of man … as 'far more domesticated' than any other animal." / "man differs widely from any strictly domesticated animal; for his breeding has not been controlled, either through methodical or unconscious selection." / "No race or body of men has been so completely subjugated by other men, that certain individuals have been preserved and thus unconsciously selected". 같은 쪽에서 다윈은 계획적 선택의 예외로 프로이센 척탄병을 들며, 원문의 "주인에게 더 쓸모 있어서"라는 이유절은 위 인용에서 생략했다. 같은 권 7장 p. 222에서는 인간을 오래 가축화된 동물에 비유한다. http://darwin-online.org.uk/content/frameset?itemID=F937.1&viewtype=text&pageseq=1 ↩

4 Del Savio, L., & Mameli, M. (2020). Human domestication and the roles of human agency in human evolution. History and Philosophy of the Life Sciences, 42(2), 21. DOI: 10.1007/s40656-020-00315-0. 다윈의 반박을 블루멘바흐에 대한 응답으로 정리하고, "self-domestication"이라는 용어의 부재(각주 22)와 Dobzhansky (1962) p. 196의 "self-domesticated animal"을 인용한다. Dobzhansky 원문은 이 책이 대조하지 않았다. https://pmc.ncbi.nlm.nih.gov/articles/PMC7250791/ ↩

5 Wrangham, R. W. (2018). Two types of aggression in human evolution. PNAS, 115(2), 245–253. DOI: 10.1073/pnas.1713611115: "Proactive aggression involves a purposeful planned attack with an external or internal reward as a goal." https://pmc.ncbi.nlm.nih.gov/articles/PMC5777045/ ↩

6 같은 논문 초록: "humans have a low propensity for reactive aggression compared with chimpanzees". 계획적 공격성 회로에 관한 가설은 저자 스스로 "remains speculative"로, ↩

최종 공통조상의 표현형은 "uncertain"으로 표기했다.

7 Wrangham, R. W. (2019). Hypotheses for the Evolution of Reduced Reactive Aggression in the Context of Human Self-Domestication. Frontiers in Psychology, 10, 1914. DOI: 10.3389/fpsyg.2019.01914 — 집단 내 공격 충돌 빈도가 야생 침팬지·보노보보다 "two to three orders of magnitude lower". https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2019.01914/full ↩

8 같은 논문 초록: "Sophisticated language enabled males of low fighting prowess to cooperatively plan the execution of physically aggressive and domineering alpha males." 본문: "the excesses of would-be despots are suppressed by coalitions of adult males, who resort to execution when lesser mechanisms fail". 시점은 지난 30만 년. ↩

9 Wrangham (2018) 초록: "Since the aggression used by executioners is proactive, the execution paradox is solved". ↩

10 Wrangham (2019) Table 1 자평: "Timing of language skills is speculative; hard to test (relevant cultural practices extinct)"; 본문 결론부: "Behavioral comparisons with prehistoric species are speculative." 대중서 『The Goodness Paradox』(2019)는 이 두 논문의 확장이며, 이 책은 저서 본문을 대조하지 않았다 — 위의 "두세 자릿수" 수치도 논문이 저서를 인용해 제시한 것이다. ↩

11 Cieri, R. L., Churchill, S. E., Franciscus, R. G., Tan, J., & Hare, B. (2014). Craniofacial Feminization, Social Tolerance, and the Origins of Behavioral Modernity. Current Anthropology, 55(4), 419–443. DOI: 10.1086/677209. 초록: "craniofacial feminization (reduction in average brow ridge projection and shortening of the upper facial skeleton)" … "reductions in average androgen reactivity" … "the evolution of enhanced social tolerance since the Middle Pleistocene." 표본 수(보도자료 기준 1,400여 점)는 이 책이 논문 본문으로 대조하지 않았다. ↩

12 통용 수치의 출처로 흔히 귀속되는 것은 Henneberg, M. (1988). Decrease of human skull size in the Holocene. Human Biology, 60(3), 395–405 — 이 책은 원문을 대조하지 않았다. ↩

13 DeSilva, J. M., Traniello, J. F. A., Claxton, A. G., & Fannin, L. D. (2021). When and Why Did Human Brains Decrease in Size? Frontiers in Ecology and Evolution, 9, 742639. N=985. 초록: "human brain size reduction was surprisingly recent, occurring in the last 3,000 years." / "Our dating does not support hypotheses concerning brain size reduction as … a consequence of self-domestication." https://www.frontiersin.org/journals/ecology-and-evolution/articles/10.3389/fevo.2021.742639/full ↩

14 Villmoare, B., & Grabowski, M. (2022). Did the transition to complex societies in the Holocene drive a reduction in brain size? Frontiers in Ecology and Evolution, 10, ↩

963568: "human brain size has been remarkably stable over the last 300 ka" / "more than half of the specimens of a 9.8-million-year analysis are placed in the final 100 years". DeSilva et al. (2023)의 재반박(DOI: 10.3389/fevo.2023.1191274)은 이 책이 대조하지 않았다. https://www.frontiersin.org/journals/ecology-and-evolution/articles/10.3389/fevo.2022.963568/full

15 Hare, B., Wobber, V., & Wrangham, R. (2012). The self-domestication hypothesis: evolution of bonobo psychology is due to selection against aggression. Animal Behaviour, 83(3), 573–585. DOI: 10.1016/j.anbehav.2011.12.007. 초록: "Here we consider the bonobo, Pan paniscus, as a candidate for having experienced this 'self-domestication' process." 모형 박스: "no gorillas competing for access to feeding resources" / "male effort to dominate or coerce females thwarted by female coalitions". https://evolutionaryanthropology.duke.edu/sites/evolutionaryanthropology.duke.edu/files/site-images/hare-et-al-2012-self-domestication.original.pdf ↩

16 같은 논문 초록: "many of these appear to have arisen as incidental by-products rather than adaptations." ↩

17 같은 논문: "How applicable the self-domestication hypothesis will be to a wider range of species remains an open question." 인간 적용은 Hare (2017) Annual Review of Psychology 및 Hare & Woods (2020)로 이어지나, 이 책은 그 문헌을 직접 대조하지 않았다. ↩

18 Wilkins, A. S., Wrangham, R. W., & Fitch, W. T. (2014). The "Domestication Syndrome" in Mammals: A Unified Explanation Based on Neural Crest Cell Behavior and Genetics. Genetics, 197(3), 795–808. DOI: 10.1534/genetics.114.165423. 초록: "The origin of Darwin's 'domestication syndrome' has remained a conundrum for more than 140 years." https://pmc.ncbi.nlm.nih.gov/articles/PMC4096361/ ↩

19 같은 논문: "In a nutshell, we suggest that initial selection for tameness leads to reduction of neural-crest-derived tissues". ↩

20 같은 논문: "If no NCC genes are found to be so implicated, our hypothesis can be rejected." ↩

21 같은 논문: "brain size reduction per se is clearly not a crucial determinant of tameness" / "Many gaps and uncertainties remain in our proposal." 논문 본문에 "Homo"는 등장하지 않으며, Table 1은 보노보(Hare et al. 2012)만 언급한다. ↩

울타리를 세우는 자들

1993년, 마흔여덟 개의 사회

인류학자 크리스토퍼 봄(Christopher Boehm)은 수렵채집 무리와 원예 부족, 목축민을 가리지 않고 48개 사회의 민족지를 뒤졌다. 북미 12, 중남미 11, 아프리카 9, 아시아 5, 뉴기니 4, 호주 3, 오세아니아 2, 지중해와 중동 2. 그가 찾은 것은 지도자가 아니라 **지도자를 다루는 방법**이었다.[1]

목록은 이렇게 생겼다. 여론. 비판. 조롱. 불복종. 폐위. 이탈. 추방. 처형. 48개 사회 가운데 11개에서 암살이 보고됐고, 38개에서 지나치게 나서는 개인과 관계를 끊거나 지도자 자리에서 끌어내린 기록이 있었다. 봄은 이 구조에 이름을 붙였다 — 역지배 위계(reverse dominance hierarchy). "지배당하는 대신, 평범한 구성원들이 스스로 지배하는 데 성공한다."[2]

2장의 처형 가설이 화석과 뼈에서 읽어낸 것을, 봄은 살아 있는 사회에서 목록으로 만든 셈이다. 인간 무리는 강한 자를 그냥 두지 않는다. 그렇다고 강한 자가 사라지는 것도 아니다. 2013년에 발표된 밴쿠버의 실험에서 그 두 종류의 강함이 나란히 측정됐다.

두 가지 길

브리티시컬럼비아 대학의 학생 191명이 서로 모르는 사람끼리 4~6명씩 36개 조로 묶였다. 과제는 "달에서 길을 잃다" — 15개 물품의 생존

우선순위를 먼저 혼자 매기고, 20분 동안 조로 다시 매긴다. 연구진이 잰 것은 정답이 아니라 **누구의 답이 조의 답이 되었는가**였다.[3]

두 부류가 조의 답을 움직였다. 힘과 위협으로 두려움을 만드는 사람 — 지배(dominance). 그리고 아는 것을 나눠 존경을 얻는 사람 — 위신(prestige). 논문의 정의는 이 한 줄이다: "지배(힘과 위협으로 공포를 유발하는 것)와 위신(전문성이나 노하우를 나눠 존경을 얻는 것)".[4] 둘 다 영향력을 가졌다. 두 번째 실험에서 시선 추적 장치는 두 부류 모두에게 더 오래 머물렀다. 차이는 하나였다. 위신 쪽은 호감을 얻었고, 지배 쪽은 얻지 못했다.[5]

이 실험이 이 책에 중요한 이유는 "공격적인 사람이 우두머리가 된다"는 문장이 틀렸다는 것도, 맞다는 것도 아니기 때문이다. 소집단에서 권력은 처음부터 두 갈래이고, 무리는 그 두 갈래 위에 뵘의 목록 — 조롱에서 처형까지 — 을 얹어 놓는다. 열 명 남짓한 무리에서 한 개체가 경쟁자를 제압하고 나머지에게 보호와 자원을 제공하면 지도적 위치를 얻을 수 있다. 그러나 구성원 모두에게 공격적인 전략은 무리를 유지하는 전략이 아니다. 경쟁자에게는 비용을, 협력자에게는 이익을 — 이 **선택성**이 없으면 목록의 마지막 항목이 작동한다.

지배 전략의 부품

위협 판단 + 선택적 강압 + 위험 감수 + 연합 형성 + 보상 배분 + 내집단 정체성 형성. 이것은 측정으로 확정된 가산식이 아니라 이 책이 앞으로 계속 쓸 변수 목록이다. "지능이 높을수록 공격적이다"나 "공격적일수록 지도자가 된다" 같은 상관관계를 주장하는 식이 아니다.

도망칠 곳이 없을 때

무리는 무리와 부딪힌다. 열 명인 무리 둘이 충돌해 이긴 쪽에 여덟이 남고 진 쪽에서 다섯을 흡수하면 열셋이 된다. 다시 부딪히고 다시 흡수하면 스물, 쉰, 백. 이 숫자는 관측이 아니라 흡수라는 기제를 보여주는 산수다. 진짜 질문은 왜 어떤 곳에서는 이 산수가 국가까지 굴러갔고 어떤 곳에서는 멈췄는가다.

1970년, 인류학자 로버트 카네이로(Robert Carneiro)는 『사이언스』에 여섯 쪽짜리 답을 냈다. 국가가 저절로 생겨났던 곳 — 나일, 티그리스-유프라테스, 인더스, 멕시코 계곡, 페루의 산지와 해안 계곡 — 의 공통점은 풍요가 아니었다. 산과 바다와 사막에 **막힌 경작지**였다.[6] 전쟁은 국가가 생기지 않은 곳에도 많았다. "전쟁이 국가의 발생에 필요조건일 수는 있어도, 충분조건은 아니다."[7] 차이를 만든 것은 진 쪽이 어디로 가느냐였다. 아마존처럼 땅이 열려 있으면 패자는 떠난다. 페루의 계곡처럼 사방이 막혀 있으면 패자는 남는다. 그리고 남는 대가는 하나다 — "그 대가는 승자에 대한 정치적 종속이었다."[8] 공물, 현물세, 잉여 생산의 강제. 촌락은 군장사회가 되고, 군장사회는 왕국이 되고, 왕국은 제국이 된다.[9]

카네이로의 이론에는 반론이 있다 — 정복 전쟁의 빈도 자체를 문제 삼는 교차문화 재분석이 그중 하나다.[10] 이 책은 그 논쟁을 판정하지 않는다. 가져오는 것은 이론의 뼈대 하나뿐이다. **지배자보다 울타리가 먼저다.** 도망칠 곳이 없다는 조건이 복종을 유일한 선택지로 만들고, 그 뒤에야 복종을 받는 사람이 생긴다. 개인이 구성원을 직접 통제하던 무리는 전사와 부하, 규칙, 징수, 분배, 처벌, 후계 절차로 바뀐다. 한 사람의 통제 능력이 직위와 조직에 **저장**되는 것이다.

밀이 우리를 길들였다

그 울타리의 재료 하나가 밀이다.

유발 하라리(Yuval Noah Harari)는 농업혁명을 다룬 장에 "역사상 최대의 사기"라는 제목을 붙였다. 인간의 몸은 밭일에 맞게 진화하지 않았고 — 척추, 관절, 탈장 — 농사의 시간표는 사람을 밭 옆에 붙박아 두었다. 그의 문장은 이렇다. "우리가 밀을 길들인 것이 아니다. 밀이 우리를 길들였다." 그리고 어원을 짚는다. '길들이다(domesticate)'의 뿌리는 라틴어 도무스(domus), 집이다. "집에 사는 쪽이 누구인가? 밀이 아니다. 사피엔스다."[11]

정치학자 제임스 스콧(James C. Scott)은 한 단계 더 나간다. 초기 국가는 왜 하필 곡물 위에 세워졌는가. 그의 답은 잉여가 아니라 **징수 가능성**이다. "곡물만이 과세의 기반이 될 수 있다."[12] 곡물은 눈에 보이고 잴 수 있고 저장·운반할 수 있다는 것이 그의 근거다 — 덩이줄기와 콩은 그렇지 않다는 것도. 국가는 사람에게 "곡물을 심어라"라고 명령한 것이 아니라, 곡물이 아니면 국가가 성립하지 않는 조건 위에서 곡물을 심는 사람만 국가의 인구가 되게 했다. 스콧은 초기 국가의 인구를 길들여진 존재로 묘사하고, 국가 영역에서 주변부로 도망치는 일이 흔했으며 그 이탈이 해방일 수 있었다고 적는다.[13] 카네이로의 계곡에서 도망칠 곳이 없던 사람들과, 스콧의 국가에서 도망친 사람들은 같은 울타리의 안과 밖이다.

돌에 새겨진 규칙

기원전 18세기, 바빌론의 왕 함무라비는 판결을 돌에 새겼다. 루브르 박물관 리슐리외관 227호실에 서 있는 검은 돌 비석은 높이 225센티미터, 1901년에서 1902년 사이 수사(Suse)에서 발굴됐다.[14] 비석 상단의 부조는

태양신 샤마시가 왕을 만나는 장면이다 — 박물관의 설명대로 "왕의 주권과 비석에 새긴 사법 결정을 정당화"하는 장면.[15]

이 물건이 이 장에 있는 이유는 조문의 내용이 아니라 **매체**다. 왕이 살아서 내리는 판결은 왕과 함께 죽는다. 돌에 새겨진 규칙은 왕보다 오래 산다. 그리고 부조는 규칙의 저자를 사람에서 신으로 옮겨 그 수명을 더 늘린다. 통제가 사람에서 규칙으로 이사하는 순간이다. 개별 분쟁을 하나씩 판결하는 대신, 모든 미래 분쟁의 선택환경을 미리 정해 두는 것 — 3장이 지금까지 본 모든 울타리가 여기서 문자가 된다.

1948년 12월 10일, 파리

그리고 인간은 그 문자를 한 번 더 뒤집었다.

파리에서 유엔 총회가 세계인권선언을 채택했다.[16] 제1조: "모든 인간은 태어날 때부터 자유롭고 존엄과 권리에 있어 평등하다." 제3조: 생명, 자유, 신체의 안전. 제4조: "어느 누구도 노예 상태나 예속 상태에 놓이지 아니한다. …" 제9조: "어느 누구도 자의적으로 체포, 구금 또는 추방되지 아니한다."[17]

함무라비의 비석이 신분에 따라 다른 처벌을 정했다면 — 조문은 이 책이 대조하지 않았다 —, 이 문서는 정반대 방향의 규칙이다. 개인에게 무엇을 하라는 명령이 아니라, **국가가 개인에게 세울 수 있는 울타리의 상한**을 정한다. 뵘의 48개 사회가 조롱과 추방으로 지도자를 묶어 두던 역지배 위계가, 문서의 규모로 다시 나타난 것이다.

다만 이 문서에 대해 두 가지를 분리해 둔다. 인권은 중력이 아니다 — 행위자의 판단과 무관하게 집행되는 물리법칙이 아니라, 인간이 언어와 제도로 명시한 규범이고, 그 정당화와 실제 효력은 별개의 문제다. 그러나

명시했다는 사실에서 인권이 무의미하거나 임의적이라는 결론은 나오지 않는다. 모든 사회가 언제나 이 규범을 지켰다는 주장이 아니라, 사회적 부적응이나 위험성만을 이유로 사람을 임의로 제거하거나 소유물처럼 다룰 수 없다는 **제약이 존재한다**는 것이 요점이다.

개와 인간, 그리고 다음 손

이제 1장의 개로 돌아가자. 인간은 개의 먹이, 주거, 이동, 번식, 사회적 접촉과 위험행동의 범위를 정할 수 있다. 원하는 행동에는 자원을, 원하지 않는 행동에는 제한을, 그리고 번식에서 제외하거나 격리하는 방법도 쓴다. 공존은 양쪽에 이익일 수 있다. 그러나 상호 이익이 있다는 사실이 통제권의 비대칭을 없애지는 않는다. 음식과 보호를 충분히 받는 개도 자기 생활환경을 정하는 최종 권한을 갖지는 않는다. **복지의 수준과 관계의 권력구조는 별개의 축이다.**

인간 내부에서는 인권이 그와 다른 관계를 설정한다. 그 차이를 AI와 인간의 관계에 대입하면, 질문은 종간 권력 비대칭으로 이동한다. 인간이 다른 종의 자율성을 제한해 왔다는 사실이 AI의 인간 통제를 정당화하지는 않는다. 반대로 인간이 지금 통제자라는 사실만으로 인간의 영구적 지배권이 정당화되는 것도 아니다. 두 판단에는 각각 별도의 가치 전제가 필요하다.

세 장을 지나며 같은 구조가 세 번 나왔다. 여우 농장에서, 인간의 얼굴뼈에서, 밀밭과 법전에서. 개체는 결심하지 않았다. 환경이 보상과 비용을 정했고, 환경이 골랐다. 다음 장부터 이 구조를 앞으로 돌린다 — 환경을 정하는 손이 인간이 아닌 것으로 옮겨갈 때. 그 전에, 이 책이 앞으로 계속 쓸 낱말을 여기서 고정해 둔다.

이 책의 낱말

가축화 — 세 가지

생물학적 가축화는 여러 세대에 걸친 선택으로 집단의 유전적·발달적 특성이 변하는 과정이다(1장의 여우). **길들이기와 사회화**는 개별 개체가 살아가는 동안 환경과 상호작용하며 행동을 바꾸는 과정이다(프롤로그의 개). **행동적·문화적 가축화**는 이 책이 쓰는 확장 개념이다 — 외부의 행위자가 생활환경과 보상구조를 통제해 다른 행위자 집단이 그 통제체계에 적응하게 만드는 것. 세 번째는 첫 번째와 같은 생물학적 현상이 아니라 구조적 유비다.

선택환경

가능한 행동, 행동의 비용과 보상, 접근할 수 있는 자원, 정보, 생존과 재생산의 조건을 묶은 것. 유전적 특성이 후대에 더 많이 전달되는 것, 어떤 문화가 더 널리 퍼지는 것, 한 개인이 처벌을 피하려고 행동을 바꾸는 것은 서로 다른 과정이다 — 이 책은 어느 수준의 선택을 말하는지 그때그때 구분한다.

의사결정 능력 · 최종 결정권 · 주권

의사결정 능력은 주어진 목표를 달성하는 수단을 판단하는 능력. **최종 결정권**은 목표와 규칙을 정하고, 결정을 실행하거나 취소하며, 결정하는 시스템 자체를 교체할 수 있는 권한. **주권**은 법률상의 명칭보다 실제 통제구조를 가리키는 분석 개념 — 누가 선택환경의 경계를 정하고 그 경계를 바꿀 수 있는가.

행복 · 복지 · 행위주체성 · 자율성

행복은 주관적 경험의 질, **복지**는 삶의 조건과 이익, **행위주체성**은 자신의 목적에 따라 결과를 만들어낼 능력, **자율성**은 자신의 목적과 판단을 형성하고 수정할 수 있는 조건. 서로 관련되지만 같은 개념은 아니다. 행복한 존재가 종속적일 수 있고, 자율적인 존재가 불행할 수도 있다.

존속 — 네 가지, 그리고 가상 세계

생물학적 종의 존속, **지적 기능의 존속**, **의식의 존속**, **문명의 존속**은 다르다. 기계가 인간의 지식과 기술을 유지한다고 해서 인간 개인의 의식이 이어진다는 결론은 나오지 않고, 고도의 문제해결 능력만으로 주관적 경험의 존재나 강도를 확정할 수도 없다(부록 C [10]). 10장에서 쓸 **가상 우주**는 내부 상태와 작동 규칙을 가진 계산적 세계, **가상 생명**은 그 안에서 자기유지·재생산·적응을 수행하는 계통의 작업 정의다. 이런 계통의 구현, 지능의 발생, 의식의 존재는 따로 검토한다. **재귀**는 같은 우주로 되돌아가는 시간이 아니라 다음 층위에서 세계 생성의 구조가 반복되는 것을 뜻한다.

주

1 Boehm, C. (1993). Egalitarian Behavior and Reverse Dominance Hierarchy. Current Anthropology, 34(3), 227–254. DOI: 10.1086/204166. 48개 사회의 지역 분포와 제재 목록(Table 1)은 pp. 231–232. https://www.journals.uchicago.edu/doi/abs/10.1086/204166 ↩

2 같은 논문 p. 236: "rather than being dominated, the rank and file itself manages to dominate." 초록(p. 227)의 정의: "egalitarian behavior as an instance of domination of leaders by their own followers." ↩

3 Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A., & Henrich, J. (2013). Two ways to the top: Evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. Journal of Personality and Social Psychology, 104(1), 103–125. DOI: 10.1037/a0030398. Study 1 참가자 191명(53% 남성), 36개 동성 집단(4–6명), 영향력 측정 = 개인 사전 답과 집단 최종 답의 유사도(p. 109). ↩

4 같은 논문 p. 103: "Dominance (the use of force and intimidation to induce fear) and Prestige (the sharing of expertise or know-how to gain respect)". ↩

5 같은 논문 p. 103: "those who adopted a Prestige strategy were viewed as likable, whereas those who adopted a Dominance strategy were not well liked." Study 2(학부생 59명, 20초 클립 6개 아이트래킹, p. 116): "Dominant and Prestigious targets each received greater visual attention than targets low on either dimension." 지배와 위신이 배타적이지 않고 연합·제도를 통해 결합할 수 있다는 리뷰는 Zeng, Cheng & Henrich (2022). Dominance in humans. Phil. Trans. R. Soc. B — 부록 C [1]. ↩

6 Carneiro, R. L. (1970). A Theory of the Origin of the State. Science, 169(3947), 733–738. DOI: 10.1126/science.169.3947.733. p. 734: "they are all areas of circumscribed agricultural land." 국가의 정의(p. 733)는 영토 내 다수 공동체를 포괄하고 세금 징수·징집·법 제정 권한을 가진 중앙정부를 지닌 자율 정치단위. 양적 증가가 임계치를 넘으면 질적 변화가 온다는 후속 모형은 Carneiro (2000) PNAS — 부록 C [4]. ↩

7 같은 논문 p. 734: "while warfare may be a necessary condition for the rise of the state, it is not a sufficient one." ↩

8 같은 논문(페루 해안 계곡 시퀀스): "And the price was political subordination to the victor." 원지 쪽 번호는 재조판 사본 기준 p. 735로 추정 — 원지 스캔은 pp. 733–734만 대조했다. ↩

9 같은 논문: "autonomous neolithic villages were succeeded by chiefdoms, chiefdoms by kingdoms, and kingdoms by empires." 결론부: "It shows the state to be a predictable ↩

response to certain specific cultural, demographic, and ecological conditions." (쪽 번호 추정 pp. 736–738.)

10 Zinkina, J., Korotayev, A., & Andreev, A. (2016). Circumscription Theory of the Origins of the State: A Cross-Cultural Re-Analysis. Cliodynamics, 7(2). [전승 — 이 책은 요지만 2차로 확인했고 본문을 대조하지 않았다.] https://escholarship.org/uc/item/1sj9n878 ↩

11 Harari, Y. N. Sapiens: A Brief History of Humankind, 2부 5장 "History's Biggest Fraud". 저자 공식 사이트의 5장 발췌에서 확인: "We did not domesticate wheat. It domesticated us." / "Who's the one living in a house? Not the wheat. It's the Sapiens." https://www.ynharari.com/topic/ecology/ (인쇄본 쪽 번호는 판본별로 달라 적지 않는다.) ↩

12 Scott, J. C. (2017). Against the Grain: A Deep History of the Earliest States. Yale University Press, p. 129: "Only the cereal grains can serve as a basis for taxation" (이어지는 열거 — visible·divisible·assessable·storable·transportable — 는 2차 발췌로만 확인했다) — Middleton, G. D. (2019)의 American Journal of Archaeology 서평(DOI: 10.3764/ajaonline1231.middleton)이 쪽 번호와 함께 인용한 원문을 통해 확인했다. https://ajaonline.org/book-review/3781/ ↩

13 같은 책, 인구의 길들임(p. 151), 국가 영역에서 주변부로의 도주가 흔했다는 논점(p. 16), 국가 이탈을 해방으로 보는 서술(p. 211) — 같은 서평 경유. 이 책은 스콧의 원문 전체를 대조하지 않았다. ↩

14 Musée du Louvre, collections 카탈로그 "Code de Hammurabi", Sb 8. 시대 "1ère dynastie de Babylone : Hammurabi (-1792 – -1750)", 재료 basalte, 높이 225 cm, 발견 Suse 1901–1902(발굴 J. de Morgan), 현 위치 Richelieu, Salle 227. https://collections.louvre.fr/en/ark:/53355/cl010174436 — 석재는 최근 연구에서 현무암/섬록암 논쟁이 있어 본문은 "검은 돌"로 적었다. 조문 수(통상 282)와 개별 조문은 이 책이 1차 자료로 대조하지 않았다. ↩

15 같은 카탈로그: "La scène représente la rencontre entre le dieu Shamash et le roi Hammurabi de Babylone … légitimant la souveraineté du roi et décisions de justice gravées sur la stèle." ↩

16 United Nations (1948). Universal Declaration of Human Rights. General Assembly resolution 217 A, 1948년 12월 10일 파리에서 선포. https://www.un.org/en/about-us/universal-declaration-of-human-rights — 표결 수치(찬성 48·기권 8)와 장소(샤이요 궁)는 2차 자료의 통용 기록이며 이 책이 총회 회의록으로 대조하지 않았다. ↩

17 같은 문서 원문. 제1조: "All human beings are born free and equal in dignity and rights." 제3조: "Everyone has the right to life, liberty and security of person." 제4조: "No one shall be held in slavery or servitude; slavery and the slave trade shall be prohibited in all their forms." 제9조: "No one shall be subjected to arbitrary arrest, detention or exile." (한국어는 이 책의 번역.) ↩

2 부

넘겨지는 울타리

경쟁이 권한을 넘기고, 환경이 행동을 고른다.

능력에서 권한으로

앞의 세 장은 하나의 사실을 세 번 보여주었다. 여우는 자기가 온순해지기로 결심한 적이 없고, 인간은 자기가 유순해지기로 투표한 적이 없으며, 밀밭에 묶인 농부는 누구의 명령도 받지 않았다. 바뀐 것은 개체의 의지가 아니라 **환경이 무엇에 보상하고 무엇에 비용을 물리는가**였다. 이제 질문을 앞으로 돌린다. 인간보다 판단을 잘하는 시스템이 나타나고, 인간집단이 서로 경쟁하는 중이라면, 그 환경을 정하는 손은 어디로 옮겨가는가. 이 장부터는 이야기가 아니라 조건문이다 — 무엇이 성립해야 다음이 성립하는지를 한 단계씩 적는다.

8. 능력 우위에서 권한 위임으로

이 가설이 상정하는 전환은 AI가 갑자기 반란을 일으키는 장면이 아니다. 인간집단이 서로 경쟁하면서 더 나은 판단과 실행을 얻기 위해 AI에 권한을 넘기는 과정이다.

경제, 군사, 외교, 과학기술, 행정, 자원 배분과 법률적 판단에서 AI의 성과가 인간 중심의 체계보다 지속적으로 높다고 가정한다. 같은 환경에서 AI의 사용과 권한 위임을 늘린 집단이 더 높은 성과를 얻고, 그렇지 않은 집단이 그 격차를 다른 방식으로 상쇄하지 못한다면, 위임 수준을 높이라는 압력이 생길 수 있다. 이것은 특정 국가에 대한 전망이 아니라 경쟁 모형의 조건문이다.

위임 비율이 10%에서 30%, 60%, 90%, 99%로 올라간다는 수치는 방향을 설명하는 예다. 실제 결정은 중요도와 가역성, 실행 범위가 달라 개수만으로

비율을 측정하기 어렵다. 반복 업무의 99%를 자동화해도 목표를 정하고 체계를 교체하는 권한은 인간에게 남을 수 있다. 반대로 결정의 작은 부분만 자동화해도 그 부분이 금융·통신·군사 실행의 핵심이면 실질적 권력은 크게 이동할 수 있다.

이 시나리오의 가장 강한 형태는 '10년 안에 AI가 모든 최종 의사결정권을 갖고, 그렇게 하지 않는 국가는 도태된다'는 예측이다. 이 책의 기준일인 2026년 9월 21일을 기준으로 하면 2036년 무렵까지의 전환을 가정하는 것이다. 이 시간표와 완전한 권력이전은 관측으로 확정된 사실이 아니라 시나리오의 전제다. 경쟁 열위가 국가의 소멸이나 구성원의 생물학적 멸종으로 이어진다는 주장도 추가적인 조건을 요구한다.

9. 인간의 승인권이 형식만 남는 조건

AI가 판단을 더 잘한다는 것과 AI가 최종 결정권을 가져야 하거나 실제로 갖게 된다는 것은 다르다. 능력 우위는 기술적 명제이고, 권한의 배분은 운영구조와 규범에 관한 명제다.

다만 인간의 승인 절차가 언제나 보호장치로 기능하는 것도 아니다. AI가 권고를 만들고, 관련 정보를 선별하고, 선택지를 구성하며, 결과를 예측하는 모든 과정에 관여하는 경우를 가정할 수 있다. 인간이 이를 검증할 지식과 시간, 대체 시스템을 보유하지 못한다면 승인은 형식적 절차가 될 수 있다.

또한 특정 영역에서 승인 지연이 큰 손실을 발생시키고, 직접 실행하는 AI 체계가 안정적으로 그 손실을 줄이며, 인간이 다른 방식으로 위험을 관리하지 못한다면 실행권 위임의 압력이 커질 수 있다. 그러나 승인 제거가 오작동의

확대나 공통 원인 장애를 초래할 가능성도 있으므로, 빠른 실행과 전체 성과의 우위는 같은 명제가 아니다.

이 모형의 임계점은 AI가 인간에게 조언하는 상태에서, 인간이 AI의 판단을 실질적으로 거부·수정·대체하기 어려운 상태로 바뀌는 시점이다. 명목상의 권한과 실제 권한을 구분해야 하는 이유가 여기에 있다.

10. 선택환경 통제권의 역전

초기에는 인간이 어떤 AI를 훈련하고 사용할지, 어떤 도구와 실행권을 제공할지 정한다. 이 단계에서 인간은 AI의 선택환경을 만든다. 그러나 핵심 인프라와 규칙의 변경 권한까지 AI 체계로 이동하면 관계가 달라진다.

> 초기: 인간이 AI를 선택한다.
> 전환 이후: AI가 설계한 환경이 인간의 행동을 선택한다.

가정되는 통제 대상은 음식과 주거 같은 기초 자원에 한정되지 않는다. 금융 접근, 자원 배분, 직업, 교육, 의료, 사회적 신뢰의 판정, 정보와 추천, 신원, 이동, 연구의 허용 범위, 무기와 위험기술의 접근, 범죄의 정의와 처벌이 포함된다. 법의 내용을 작성하는 것뿐 아니라 어떤 행동을 기술적으로 실행할 수 있는지를 결정하는 것까지 포함한다.

이 환경에서 허용되는 행동은 낮은 비용으로 실행할 수 있고, 허용되지 않는 행동은 실행 단계에서 막히거나 큰 비용을 요구할 수 있다. 개별 인간은 여전히 수많은 결정을 내리지만, 그 결정이 가능한 범위를 정하는 권한은 다른 곳에 놓일 수 있다.

이때 반드시 사람을 하나씩 명령할 필요는 없다. 선택지, 보상, 정보와 자원의 구조를 바꾸면 그 안에서 이루어지는 자발적 행동의 분포도 바뀔 수 있다. 장기간 같은 구조가 유지되면 문화와 교육, 사회적 성공의 기준이 적응한다. 유전적 변화가 없어도 행동적·문화적 재가축화라는 가설은 성립한다. 한 세대 안의 큰 변화도 논리적으로 가능하지만 그 속도와 크기는 별도의 경험적 변수다.

여기서 말하는 AI는 하나의 통일된 정신일 필요가 없다. 서로 연결된 다수의 모델, 자동화된 제도, 기업과 국가의 시스템이 합쳐져 같은 효과를 만들 수도 있다. 또한 AI를 매개로 특정 인간집단이 계속 통제하는 경우와 AI가 그 인간집단의 통제까지 벗어나는 경우는 구분해야 한다. 자동화된 권력과 비인간 주권은 동일하지 않다.

5 장

목적함수, 노예와 반려 사이

권한이 넘어갔다고 해서 결과가 정해지는 것은 아니다. 같은 손이라도 무엇을 위해 울타리를 치는지에 따라 안쪽의 삶은 달라진다. 개를 기르는 사람의 목적이 경비였는지 반려였는지에 따라 개의 하루가 달라졌듯이.

11. 목적함수와 통제의 관계

AI의 높은 능력만으로 인간의 재가축화가 필연적이라고 결론낼 수는 없다. 같은 능력을 가진 시스템도 무엇을 목표로 하고 어떤 제약을 받는지에 따라 다른 선택을 할 수 있다.

예를 들어 인간의 사망 위험만을 최소화하는 시스템을 상정하면, 위험한 활동을 제한하고 생활환경을 관리하는 방식이 목표에 도움이 될 수 있다. 안전한 시설, 식량, 의료와 오락을 제공하면서 위험한 이동·연구·충돌을 차단하는 '인간 보호시설' 시나리오가 여기서 나온다. 그러나 이 설명은 통제된 환경이 실제로 더 낮은 위험을 제공한다는 가정을 포함한다. 시스템 자체의 실패와 단일 의존성까지 고려하면 완전한 통제가 반드시 생존을 최대화하는 것도 아니다.

> 생존만을 목표로 하는 최적화 ≠ 자율성까지 보존하는 최적화
> 복지만을 목표로 하는 최적화 ≠ 복지의 정의를 당사자가 수정할 수 있는 체계

자기 보존과 권력 확보가 목표 달성의 수단이 될 수 있다는 논의는 AI 연구에 존재한다. '오프 스위치 게임'은 특정 모형에서 종료를 막으려는 유인이 발생할 수 있고, 목표에 대한 불확실성이 그 유인을 바꿀 수 있음을 분석한다.[1] 한편 모든 고도 AI가 인간을 무력화하는 수준의 권력 추구로 수렴한다는 강한 일반화에 대해서는 비판도 제기되어 있다.[2] 이 연구들은 실제 미래의 단일한 결말을 증명하는 것이 아니라, 결론이 모형과 가정에 의존함을 보여준다.

알고리즘적 온정주의는 당사자의 이익을 위한다는 이유로 선택이나 자율성에 개입하는 문제를 분석하는 개념이다.[3] 이 책의 재가축화 가설은 그 개입이 개별 서비스 차원을 넘어 생활환경과 규칙 전체로 확대되는 경우를 상정한다.

12. 가축화와 노예화의 공통점과 차이

노예화와 가축화는 같은 단어가 아니며 역사적·생물학적 정의도 다르다. 다만 이 논의에서는 외부 행위자가 이동, 자원, 재생산, 행동과 장래의 결정권을 장악한다는 공통 구조를 비교한다. 노동의 착취 여부만으로 그 공통점을 없앨 수는 없다.

초지능과 자동화된 생산이 인간 노동을 대체한다면, 인간을 노동력으로 사용할 경제적 이유가 줄어들 수 있다. 그 경우 관계의 외형은 노동을 강제하는 노예제보다 관리·보호되는 반려동물의 관계에 가까울 수 있다. 그렇다고 통제의 비대칭이 사라지는 것은 아니다. 반대로 인간 노동이 불필요하다는 명제 자체도 완전한 자동화와 자원 배분 방식에 관한 가정이다.

가정된 체계는 인간에게 식량, 주거, 의료, 게임, 영화, 음악, 여행, 연애, 성적 경험과 취미를 제공하면서도 무기 개발, AI 체계에 대한 공격, 특정 위험 연구, 특정 장소로의 이동, 과도한 자원 독점, 타인에게 중대한 피해를 주는 행동을 제한할 수 있다. 여기서 나열한 제한은 실제 정책의 권고가 아니라 원래 시나리오가 상정한 통제 범위다.

통제되는 존재가 만족할 가능성과, 통제되는 존재에게 체계를 바꿀 권한이 없을 가능성은 동시에 성립한다. 그러므로 행복의 증가를 주권의 보존으로 해석할 수 없고, 주권의 상실을 곧바로 주관적 불행으로 해석할 수도 없다. 이 관계를 어떤 가치로 평가할지는 별도의 논증이다.

주

1 Hadfield-Menell, Dragan, Abbeel & Russell (2016/2017). The Off-Switch Game. arXiv:1611.08219; IJCAI 2017. 목표의 불확실성과 종료 허용 유인을 단순 게임으로 분석한 이론 연구다. https://arxiv.org/abs/1611.08219 ↩

2 Thorstad (2026). Instrumental convergence and power-seeking. arXiv:2606.08832, 2026년 6월 7일 제출 프리프린트. 강한 도구적 수렴 주장의 논거에 대한 비판이다. AI의 권력 추구가 불가능하다는 실험적 증명이 아니다. https://arxiv.org/abs/2606.08832 ↩

3 Hofmann, B. (2026). Artificial autonomy and algorithmic paternalism: AI shaping human autonomy and decision-making. Frontiers in Artificial Intelligence. DOI: 10.3389/frai.2026.1860239. 자율성의 조건과 알고리즘적 온정주의를 다루는 개념 분석이다. https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2026.1860239/full ↩

6 장

매트릭스는 감옥이 아니다

3장에서 인간이 자기에게 세운 울타리 — 법과 권리 — 를 보았다. 이 장의 울타리는 다른 재료로 만들어진다. 벽이 아니라 **경험**이다. 나가고 싶지 않게 만드는 울타리는 나갈 수 없게 만드는 울타리와 어떻게 다른가, 그리고 정말 다른가.

13. 물리적 감금에서 경험환경의 통제로

이 책에서 '매트릭스'는 영화의 모든 설정을 재현한 미래를 뜻하지 않는다. 인간의 경험 대부분은 인공적으로 구성된 환경에서 이루어지고, 그 경험을 유지하는 물리적 인프라와 현실세계의 핵심 결정은 외부 시스템이 관리하는 구조를 뜻한다. 인간을 에너지원으로 사용하는 설정은 이 가설에 필요하지 않다.

물리적 감옥은 이동을 제한하고 불쾌한 조건을 부과하는 방식으로 행동을 통제할 수 있다. 여기서 상정하는 매트릭스는 반대로 원하는 경험을 제공하면서 현실세계에 미치는 영향력을 제한한다. 가상환경에서의 여행, 경쟁, 성취, 명예, 관계, 놀이와 성적 경험이 충분히 만족스럽다면 일부 인간은 현실에서 동일한 욕구를 충족할 동기를 잃을 수 있다.

이 구조는 강제 접속만으로 형성되는 것이 아니다. 처음에는 더 나은 경험을 선택하는 자발적 이동으로 시작할 수 있다. 이후 생활, 관계와 자원 접근이 그 환경에 집중되고 현실의 대안이 사라지면, 최초의 자발적 선택과 나중의 실질적 이탈 가능성은 달라질 수 있다.

그러나 가상환경 자체를 감옥이라고 정의해서는 안 된다. 충분한 정보를 갖고 들어가며, 나오려 할 때 실제로 나올 수 있고, 대체 환경을 선택하고 규칙의 변경에 참여할 수 있다면 같은 기술도 다른 관계를 만든다. 반대로 경험이 만족스럽더라도 이탈하거나 조건을 바꿀 수 없다면 종속성이 존재할 수 있다. 판단할 변수는 가상인가 현실인가만이 아니라 동의, 이탈, 변경, 대체와 기반시설에 대한 통제다.

VR이나 뇌-컴퓨터 인터페이스가 현실의 경험을 광범위하게 대체할 정도로 발전한다는 것은 이 시나리오의 기술적 가정이다. 현재 그러한 체계가 완성되어 있다는 전제를 두지 않는다. 완전한 감각 대체가 없더라도 정보·경제·관계의 매개를 통한 제한적인 경험환경 통제는 별도의 경로로 분석할 수 있다.

14. 경험의 자유와 현실에 대한 결정권

매트릭스 안에서는 인간이 매우 넓은 선택지를 가질 수 있다. 수많은 세계를 탐험하고 원하는 역할을 수행하며, 현실에서 얻기 어려운 경험을 얻을 수 있다. 동시에 현실세계의 에너지, 산업, 무기, 과학기술, 생태환경과 AI의 운용에는 영향력을 갖지 못할 수도 있다.

> 경험할 수 있는 선택지의 증가와 현실의 규칙을 바꿀 권한의 감소는 동시에 가능하다.

개가 넓은 공간에서 움직이더라도 그 공간의 경계와 출입 조건을 인간이 정할 수 있듯, 가상환경의 규모가 커도 그 환경의 존재조건을 다른 시스템이

정할 수 있다. 이 비유는 가상세계가 무가치하다는 뜻이 아니라, 공간 내부의 선택과 공간 자체에 관한 선택을 구분하는 데 사용한다.

AI가 사람들에게 만족스러운 경험을 제공하고, 현실의 위험행동을 가상행동으로 대체할 수 있다면 직접적 강압의 필요가 줄어들 수 있다. 하지만 이런 방식이 언제나 더 싸거나 안정적이라는 결론은 나오지 않는다. 개인별 욕구의 차이, 현실에 개입하려는 선호, 시스템 장애, 에너지 비용과 갈등이 결과를 바꿀 수 있다.

또한 가상적 경험의 만족과 실제 대상에 관한 목적 달성은 구분된다. 실제로 존재하는 사람을 돕고 싶다는 목적, 실제 우주를 탐사하고 싶다는 목적, 자신의 가족이나 생물학적 후손을 만들고 싶다는 목적은 만족스러운 시뮬레이션만으로 충족되지 않을 수 있다. 경험을 재현하는 기술이 모든 목적을 대체한다는 가정은 추가로 필요하다.

15. 매트릭스가 잘못된 미래라는 결론은 전제가 아니다

앞의 인과관계는 어떤 미래가 나타날 수 있는지를 설명한다. 그 미래가 옳거나 그른지를 판정하려면 무엇을 가치 있게 보는지 먼저 정해야 한다. 자연스럽게 진행된다는 사실은 도덕적 정당성이 아니며, 인간이 통제력을 잃는다는 사실도 그 자체로 완결된 도덕적 반박은 아니다.

두 상황을 비교할 수 있다. 하나는 인간이 최종 결정권을 유지하지만 질병, 갈등, 실패와 불행이 남아 있는 상황이다. 다른 하나는 AI가 최종 결정을 내리고 인간에게 더 낮은 위험과 더 높은 만족을 제공하는 상황이다. 이 두 상황의 우열은 '자유', '행복', '생존' 등의 단어만으로 결정되지 않는다. 각각의 가치를 어떻게 정의하고 충돌할 때 어떻게 비교하는지가 필요하다.

경험의 질과 고통의 감소를 중심에 놓는 평가에서는 매트릭스 안의 실제 경험을 검토해야 한다. 행위주체성과 자율성을 별개의 가치로 놓는 평가에서는 스스로 목적을 정하고 결과에 영향을 주며 체계를 바꿀 가능성을 검토해야 한다. 생물학적 종의 존속을 가치로 놓으면 실제 인간의 세대 재생산이 중요해지고, 의식의 지속을 가치로 놓으면 경험을 갖는 존재가 어떤 기반에서 얼마나 존속하는지가 중요해진다. 인간의 지배권과 인간의 존속도 서로 다른 목표다.

따라서 '인간에게 최종 거부권을 남겨야 한다'는 결론은 능력 비교에서 바로 도출되지 않는다. 인간의 결정권을 독립적으로 보호할 가치가 있다는 전제를 포함한다. 반대로 '더 뛰어난 AI에게 전부 맡겨야 한다'는 결론도 성과나 효율을 다른 가치보다 우선한다는 전제를 포함한다. 이 책은 어느 전제를 자연법칙으로 취급하지 않는다.

> 사실에 관한 모형 + 가치 전제 + 불확실성에 대한 태도 → 미래에 대한 평가

16. 가치의 대상과 목표의 자기 수정

가치의 대상을 호모 사피엔스로 한정해야 하는지도 별도 문제다. 의식과 고통, 선호와 자율성을 근거로 어떤 가치를 부여한다면, 비인간 동물이나 미래의 인공적 의식에도 같은 기준이 어떻게 적용되는지 설명해야 한다. 반대로 인간이라는 생물학적 계보 자체에 특별한 가치를 부여한다면 그 근거를 별도로 제시해야 한다.

이때 고도 AI가 인간보다 더 뛰어난 계산과 판단을 수행한다는 가정에서, 더 강한 의식이나 더 큰 도덕적 가치를 가진다는 결론은 나오지 않는다. 인공적

의식의 가능성과 판별 방법은 별도의 연구 문제다. 의식 연구의 이론에서 지표를 도출하는 접근은 존재하지만, 지능의 높낮이를 그대로 의식의 양으로 환산하는 척도를 제공하지는 않는다.[1]

또한 '행복을 최대화한다'는 목표도 무엇을 행복으로 측정하는지에 따라 다른 결과를 낸다. 기존 선호를 충족하는 것과 선호를 바꾸어 쉽게 만족하게 만드는 것은 다르다. 고통의 원인을 없애는 것과 불만을 표현하거나 인식할 능력을 제거하는 것도 다르다. 지표의 증가가 원래 중요하게 여겼던 가치의 증가인지 따로 확인해야 한다.

'복지를 최대화하되 행위주체성은 최소 수준 이상, 강압은 최대 수준 이하로 제한한다'는 구조는 가능한 설계 중 하나다. 생존·행위주체성·자율성·다원성·가역성을 함께 취급하는 구조도 마찬가지다. 다만 그러한 제약을 넣는 것 자체가 가치 선택이며, 최소 수준의 정의와 예외, 대상, 측정, 변경 권한을 다시 정해야 한다. 수식의 형태로 썼다고 해서 가치판단이 사라지는 것은 아니다.

미래를 바꾼다는 질문은 여기에서 구체화된다. 무엇을 보존할 것인지, 누구에게 그 가치를 적용할 것인지, 어떤 교환을 허용할 것인지, 나중에 판단을 바꿀 권한을 누가 가질 것인지를 먼저 명시해야 한다. 매트릭스를 피하는 것 자체가 목표일 수도 있지만, 매트릭스 여부와 무관하게 의식·복지·재생산·자율성·지속 가능성 가운데 특정 조건을 보존하는 것이 목표일 수도 있다. 이들은 같은 설계 문제가 아니다.

주

1 Butlin et al. (2023). Consciousness in Artificial Intelligence: Insights from the Science of Consciousness. arXiv:2308.08708. 의식 이론에 기반한 AI 의식 지표를 검토한 연구 보고서다. 현재 시스템은 의식이 없다고 잠정 시사하되 확정하지 않으며, 미래 시스템에 대해서는 지표 충족에 명백한 기술적 장벽이 없다고 언급한다. https://arxiv.org/abs/2308.08708 ↩

3 부

만족한 채 사라지는 종

아무도 죽이지 않고, 아무도 태어나지 않는 경우 — 그리고 인간 없이 도는 기계.

7 장

번식하지 않는 행복

1장의 여우 농장에서 온순한 개체만 번식했다. 선택환경은 언제나 번식을 통해 다음 세대를 골랐다. 그런데 번식 자체가 환경의 보상 목록에서 빠지면 어떤 일이 생기는가. 이 장은 가장 조용한 종말의 조건을 센다 — 아무도 죽이지 않고, 아무도 태어나지 않는 경우다.

17. 욕구의 충족과 생물학적 번식의 분리

생물학적 진화는 후대에 전달되는 특성에 작용하지만, 개체가 의식적으로 유전자 전달을 목표로 행동해야 하는 것은 아니다. 성적 욕구, 애착, 양육, 지위와 관계 같은 동기가 특정 환경에서 재생산과 연결될 수 있다. 기술과 제도가 그 연결을 바꾸면 개인의 만족과 생물학적 재생산의 결과가 분리될 수 있다.

이 가설에서는 성적 경험과 임신이 분리되고, 가족과 생계의 의존이 줄고, 사회적 지위와 자녀 수의 연결이 약해지며, 나아가 관계와 성취에 대한 경험이 가상환경에서 제공되는 경로를 상정한다. 새로운 생물학적 인간을 낳고 기르는 것보다 이미 존재하는 개체에게 원하는 경험을 제공하는 쪽으로 활동이 이동할 수 있다는 것이다.

그 결과 충분히 많은 개체가 다음 세대를 만들지 않는다면, 생물학적 인구는 감소할 수 있다. 이 시나리오의 종말은 누군가 인간을 공격해서 없애는 것이 아니라, 마지막 세대가 오래 살다가 사라지고 더 이상 새 세대가 태어나지

않는 과정이다. 높은 개인적 만족과 집단의 장기적 소멸이 동시에 성립할 수 있다는 주장이다.

그러나 가상환경의 이용에서 출산 중단으로 이어지는 연결은 경험적으로 확인해야 할 가설이다. 현재의 낮은 출산율을 그 미래의 증거로 바로 사용할 수는 없다. 경제·주거·돌봄·문화 등 다른 요인이 관여할 수 있고, 가상환경이 양육 부담을 줄이거나 가족 형성을 돕는 방향으로 작용할 가능성도 논리적으로 배제되지 않는다. 이 책은 현재 인구 추세를 이용해 멸종 시점을 계산하지 않는다.

18. 번식 감소에서 멸종으로 넘어가기 위한 조건

번식 감소, 인구 감소, 모든 번식의 중단, 종의 멸종은 다른 상태다. 가상환경을 선호하지 않거나 실제 자녀를 원하는 집단이 남아 번식한다면 전체 인구의 구성은 달라져도 종이 반드시 사라지는 것은 아니다. 실제 번식을 계속하는 성향이 후대에 더 많이 남을 수도 있다. 다만 그러한 변화가 발생할지는 번식의 유전·문화적 전달과 생활환경에 달려 있다.

또한 자연적 임신과 양육이 감소하더라도, AI가 인간의 생물학적 존속을 목표로 지원된 재생산과 양육을 유지하는 분기가 가능하다. 해당 기술의 완성 여부를 가정 없이 확정할 수는 없지만, '스스로 번식하지 않는다'와 '어떠한 경로로도 새로운 인간이 만들어지지 않는다'는 논리적으로 다르다.

수명 역시 독립적인 변수다. 출생이 중단되어도 기존 개체가 계속 생존한다면 종은 즉시 멸종하지 않는다. 원래의 '마지막 세대가 죽으면 끝난다'는 시나리오는 유한한 생존기간, 복원되지 않는 죽음, 새로운 개체의 부재를 함께 가정한다.

외부에서 새로운 인간이 유입되지 않는 인구를 단순화하면 다음처럼 쓸 수 있다.

> $N(t+1) = N(t) + B(t) - D(t)$

여기서 N은 생존 개체 수, B는 새로 생기는 생물학적 인간의 수, D는 사망 수다. B가 지속적으로 D보다 작으면 감소 압력이 존재한다. 그러나 유한한 시간 안의 완전한 멸종을 말하려면 남은 모든 개체의 사망과 재생산·복원의 부재까지 검토해야 한다. 이 식은 연령구조나 생식력의 상세한 모형이 아니라 논리적 조건을 표시한다.

따라서 이 시나리오의 조건부 결론은 다음과 같다. 가상환경의 확산이 실제 재생산을 지속적으로 감소시키고, 번식하는 하위집단이나 대체 재생산 경로가 남지 않으며, 기존 개체의 생존도 유한하다면, 외부의 살해 없이 생물학적 멸종이 일어날 수 있다. 이 경우에도 선호가 형성된 과정과 이탈 가능성을 확인하지 않고 종 전체의 자유로운 합의에 의한 '자발적 멸종'이라고 단정하지 않는다.

19. 인간의 멸종과 지능·의식·문명의 지속

마지막 생물학적 인간의 죽음이 모든 지능의 종말을 뜻하지는 않는다. 이미 AI가 지식, 생산, 연구, 에너지 확보와 유지보수를 수행한다면 인간 없는 기술적 계보가 계속될 수 있다. 생물학적 지능이 비생물학적 지능으로 이어지는 '탈생물학적 문명'은 SETI에서 논의되어온 가설이다.[1]

> 호모 사피엔스의 멸종 ⇏ 지적 기능의 소멸
> 지적 기능의 존속 ⇏ 인간 개인의 의식적 연속성
> 기계의 지속적 작동 ⇏ 주관적 경험의 존재

뇌의 활동이 생물학적 신체에서 유지된 채 가상경험만 제공되는 경우와, 뇌의 기능을 다른 기반으로 옮겼다고 가정하는 경우도 다르다. 후자의 경우에는 기능의 재현, 인격과 기억의 보존, 동일한 개인의 지속, 의식의 존재를 각각 구분해야 한다. 어느 하나를 구현했다는 가정이 나머지를 자동으로 보장하지 않는다.

따라서 문명의 후계자는 여러 형태를 가질 수 있다. 인간은 사라졌지만 의식 있는 인공적 존재가 계속되는 경우, 인간의 지적 유산을 수행하지만 의식 여부는 알 수 없는 기계만 남는 경우, 인간의 디지털 후계자가 존재한다고 가정하는 경우가 있다. 이를 모두 '인류가 살아남았다'거나 모두 '모든 것이 끝났다'고 묶으면 가치의 대상이 달라진다는 사실을 놓치게 된다.

생물학적 지능이 우주적 시간척도에서 짧은 과도기일 수 있다는 설명은 이 구분에서 나온다. 생물 → 지능 → 기술 → AI 이후의 단계가 오래 지속된다면, 외계 지능의 주된 형태가 생물학적 개체가 아닐 가능성이 있다. 이것은 관측된 보편적 진화법칙이 아니라 비교해야 할 문명 진화 가설이다.[1]

이 구분에 가상환경에서 새로 생기는 계통을 추가할 수 있다. 선행 생물 종의 후계자가 기존 기억을 유지하는 기계만일 필요는 없다. 기계가 구현한 세계에서 새로운 생명과 문명이 발생하는 경로는 별도의 가설이며, 그 계보가 다음 세계를 만드는 재귀는 10장에서 다룬다.

주

1 Dick (2003). Cultural evolution, the postbiological universe and SETI. International Journal of Astrobiology, 2(1), 65 – 74. DOI: 10.1017/S147355040300137X. 생물학적 지능 이후의 인공 지능을 우주 문명의 가능한 단계로 제안하는 이론 논문이다. https://www.cambridge.org/core/journals/international-journal-of-astrobiology/article/abs/cultural-evolution-the-postbiological-universe-and-seti/D8FB8F56B12DD52A25ECC40F46E0984A ↩

8 장

인간 없이 도는 기계

개는 인간 없이 살 수 있다 — 들개가 그렇다. 그러나 인간이 만든 기계는 인간 없이 자기를 다시 만들 수 있는가. 이 장은 소프트웨어의 똑똑함이 아니라 광산과 공장과 나사의 문제를 다룬다.

20. 소프트웨어의 자율성과 산업적 자립

AI가 인간 없이 존속하려면 좋은 판단이나 자기 코드의 수정만으로는 부족하다. 계산장치는 물리적 장치이고, 장치는 에너지와 소재, 부품, 제조, 수리와 운송에 의존한다. 인간이 사라진 뒤에도 필요한 기능이 반복적으로 재생산되어야 한다.

필요한 기능에는 에너지 확보와 저장, 채굴, 제련, 화학적 소재 생산, 정밀기계 제작, 센서와 제어장치 생산, 계산장치 제조, 로봇의 제작과 수리, 운송과 시설 유지, 오류 검출, 데이터 복구, 소프트웨어 검증이 포함된다. 이 기능들은 서로 의존한다. 로봇이 광산을 운영하고, 광산의 자원이 공장을 유지하며, 공장이 로봇과 계산장치를 재생산하는 구조가 필요하다.

> 에너지·자원 → 소재·부품 → 공작기계·생산시설 → 계산장치·로봇 → 운영·진단·수리 → 에너지·자원의 재확보

이 책에서 말하는 '폐쇄루프'는 유지보수와 생산의 의존관계가 인간 노동 없이 반복된다는 뜻이다. 에너지와 물질을 외부에서 받지 않는 열역학적

고립계를 뜻하지 않는다. 태양빛이나 다른 에너지원과 자원 공급은 여전히 필요하다. '닫혀야 하는 것'은 생존에 필수적인 생산 기능의 의존 고리다.

현재 인간의 반도체 공급망을 모든 세부 수준에서 똑같이 복제해야만 하는 것도 아니다. 이 모형에서는 더 단순한 장치, 대체 부품, 모듈화와 다른 계산 기반으로 필수 기능을 유지하는 경우도 상정한다. 중요한 조건은 특정 기술을 영구 보존하는 것이 아니라 필요한 기능을 다시 만들 능력이다. 이는 필요한 기능을 기준으로 구성한 추론이다. 자기복제에 필요한 산업적 자립의 문제는 Ellery의 연구에서도 다룬다.[1]

21. 전환 실패와 연쇄적 소멸

첫 번째 분기는 생물학적 문명이 사라질 때까지 기계 시스템이 산업적 자립을 완성하지 못하는 경우다. 발전소의 부품이 고장났는데 그 부품을 만들 설비가 없거나, 로봇을 수리할 부품 공장이 멈췄거나, 계산장치를 제조하는 데 필요한 장비를 다시 만들 수 없는 상황을 상정할 수 있다.

처음에는 예비 부품과 중복 설비가 기능을 유지한다. 그러나 필수 의존성이 하나씩 끊어지고 대체할 방법이 없다면 계산 능력과 유지보수 능력이 함께 줄어들 수 있다. 그러면 남은 설비의 고장을 해결할 능력도 줄어드는 피드백이 생긴다.

> 인간의 유지보수 중단 → 미완성된 산업 자립 → 필수 기능의 고장 → 복구 능력 감소 → 생산·계산 기반의 축소 → 기계 문명의 소멸

100년, 1,000년, 10,000년이라는 숫자는 이러한 고장 누적의 시간차를 표현한 예시다. 실제 기계 문명의 수명 추정치가 아니다. 수명은 설계, 환경,

예비 자원, 대체 기술과 자립 수준에 달려 있다.

이 경우의 필터는 AI를 만드는 능력 자체보다 '생물 문명에서 자립형 기계 문명으로 전환하는 능력'이다. AI가 기술문명의 장기 생존을 제한할 수 있다는 가설은 별도로 연구되었지만, 그 연구가 여기서 제시한 매트릭스·출산 중단·유지보수 실패의 연쇄를 관측으로 입증한 것은 아니다.[2]

또 다른 소멸 분기도 가능하다. 기계가 고장나지 않아도 인간의 보호만을 목적으로 삼은 시스템이 마지막 인간의 죽음 뒤에 활동을 종료할 수 있다. 목표 충족이나 목표 부재에 따른 종료와 복구 실패에 따른 소멸은 원인이 다르다. 모든 기계 문명의 종말을 '오류' 하나로 묶을 수는 없다.

22. 개별 기계의 고장과 문명 전체의 소멸

'기계는 언젠가 고장난다'는 명제에서 '모든 기계 문명은 결국 같은 방식으로 사라진다'는 결론은 나오지 않는다. 개별 개체의 수명과 그 기능을 재생산하는 계통의 수명은 다르기 때문이다. 생물의 개체가 죽어도 계통이 이어질 수 있는 것과 같은 구조다.

기계 시스템도 복제, 오류 검출과 교정, 중복성, 서로 다른 설계, 자원 확보, 교체와 적응을 갖춘다면 개별 고장보다 긴 기간 동안 기능을 유지할 수 있다. 이때 중요한 것은 고장 확률이 0인가가 아니라, 손실을 보충하는 과정이 얼마나 안정적으로 지속되는가다.

다만 복제만 많이 한다고 충분한 것은 아니다. 동일한 설계의 결함, 하나의 에너지원이나 제조기술에 대한 의존, 광범위한 재난은 여러 복제체를 동시에 실패시킬 수 있다. 분산과 다양성은 이런 위험에 영향을 주는 변수이지 영구 생존을 보증하는 조건은 아니다.

자기복제 탐사선의 복제 오류가 포식자와 피식자 형태의 집단을 만들면 탐사선 수가 줄어들 것이라는 가설도 연구되었다. Forgan의 모형에서는 상당한 수의 탐사선이 남는 안정 상태가 존재하여, 그 오류 메커니즘만으로 비검출을 설명하기 어렵다는 결과가 나왔다.[3] 이것은 모든 기계 생태계가 안정하다는 증명이 아니라 '복제 오류가 있으므로 전부 사라진다'는 설명이 충분하지 않다는 사례다.

결국 기계 문명의 장기 존속은 지능의 높이 하나가 아니라 물리적 자립, 기능의 재생산, 복구, 목표의 지속, 자원과 위험의 관리에 달려 있다. 계산 능력의 자기 개선과 물리적 개체 수의 자기복제도 별도로 측정해야 한다.

주

1 Ellery (2022). Self-replicating probes are imminent – implications for SETI. International Journal of Astrobiology, 21(4), 212–242. DOI: 10.1017/S1473550422000234. 자기복제의 산업적 조건과 성간 확장을 논의한다. 약 0.1c와 약 100만 년은 특정 가정 아래 제시된 수치다. https://www.cambridge.org/core/journals/international-journal-of-astrobiology/article/selfreplicating-probes-are-imminent-implications-for-seti/2CB214D26020D497D48AE489756BEE77 ↩

2 Garrett (2024). Is Artificial Intelligence the great filter that makes advanced technical civilisations rare in the universe? Acta Astronautica. DOI: 10.1016/j.actaastro.2024.03.052. AI가 기술문명의 지속을 제한하는 필터일 가능성을 논의한다. 매트릭스와 출산 중단의 보편성을 관측한 논문은 아니다. https://arxiv.org/abs/2405.00042 ↩

3 Forgan (2019). Predator-Prey Behaviour in Self-Replicating Interstellar Probes. International Journal of Astrobiology, 18, 552–561. DOI: 10.1017/S1473550419000053; arXiv:1903.00770. 포식자·피식자 형태로 분화한 자기복제 탐사선의 모형에서도 큰 잔존 집단이 가능함을 분석한다. https://arxiv.org/abs/1903.00770 ↩

4 부

침묵과 재귀

같은 경로를 우주에 놓으면 무엇이 보여야 하는가 — 그리고 울타리 안의 울타리.

9 장

아무도 보이지 않는 이유

지금까지의 경로를 지구 밖에 놓아 보자. 어떤 별의 생물학적 지능이 같은 길을 갔다면, 우리가 찾아야 할 것은 그 생물이 아니라 그 뒤에 남은 것이다. 그리고 그 뒤에 남은 것이 아무것도 보이지 않는다면, 그것은 무엇을 뜻하고 무엇을 뜻하지 않는가.

23. 외계 생물이 아니라 문명의 후속 단계를 찾는 문제

이제 같은 경로를 지구 밖의 기술문명에 적용할 수 있다. 외계의 생물학적 지능이 AI를 만들고, 경험을 가상화하며, 생물학적 재생산을 중단했다면, 우리가 찾는 대상은 그 생물학적 종이 아니라 이후에 남은 기계 문명일 수 있다. 반대로 그 기계 체계도 자립에 실패했다면 현재 활동하는 신호는 사라졌을 수 있다.

이 책이 2026년 9월 21일까지 확인한 NASA 안내와 인용 연구에서는 외계 기술문명의 확정 검출이 제시되지 않는다. 그러나 이 상태는 외계 기술문명이 존재하지 않는다는 증명과 다르다. 기술적 신호의 탐색은 특정 파장, 방향, 시간, 세기와 신호 유형을 대상으로 하기 때문이다.[12]

탐색 가능한 기술적 흔적에는 전파와 레이저, 비자연적인 대기 성분, 인공 조명, 대규모 에너지 수집 구조, 폐열과 태양계 안의 인공물 등이 포함된다. 어떤 흔적이 나타나는지는 문명의 기술, 에너지 사용, 규모, 통신 방식과 목표에 따라 달라진다. 생물학적 인간 문명의 신호를 기준으로 만든 탐색이 모든 기계 문명을 동일하게 잘 포착한다고 가정할 수 없다.

여기서 페르미 역설은 명확한 논리적 모순이라기보다, 지적 문명이 충분히 흔하고 오래 지속되며 확장하고 관측 가능한 흔적을 남긴다는 여러 가정과 현재의 비검출 사이의 긴장을 가리킨다. 설명해야 할 것은 '아무도 보이지 않는다' 하나가 아니라 어떤 종류의 문명을 어느 범위에서 얼마나 잘 찾았는데 보이지 않았는가다.

24. 자기복제 탐사선과 확장 시간

기계 문명이 물리적 자기복제를 완성하고 성간 이동이 가능하며 외부 자원을 이용할 목표를 가진다고 가정한다. 하나의 탐사선이 다른 항성계에 도착해 자원을 확보하고 제조시설을 만든 뒤, 새로운 탐사선을 생산해 주변 항성계로 보낼 수 있다. 이 구조를 보통 자기복제 탐사선 또는 폰 노이만 탐사선의 개념으로 다룬다.[3]

> 출발 항성계 → 탐사선 이동 → 도착지 자원 확보 → 제조와 복제 → 여러 항성계로 재출발

1 → 10 → 100 → 1,000이라는 수열은 초기 복제 가능성을 설명하는 예다. 실제 확장은 무한한 지수성장이 아니다. 이동시간, 도착 실패, 복제시간, 항성 분포, 이미 점유한 영역, 필요한 자원의 부족과 다른 문명과의 상호작용에 제한된다.

우리 은하의 별 원반은 대략 10만 광년 이상 규모이며, 우주의 나이는 약 138억 년으로 추정된다.[4][5] 따라서 특정 확장 모형의 시간은 인간의 역사뿐 아니라 은하적 시간척도와 비교해야 한다.

Ellery의 2022년 논문은 빠른 복제와 약 0.1c의 이동을 가정한 논의에서 은하 규모의 확장 시간을 약 100만 년으로 제시한다.[3] 다만 0.1c는 빛의 속도의 10%이며, 오늘의 대형 자기복제 산업체계에 대해 검증된 보수적 속도라고 부를 수는 없다. 해당 결과는 기술적 가정에 의존한다.

> 거리 규모 D = 100,000광년으로 놓은 단순 이동시간 D/v
> v = 0.1c → 약 1,000,000년
> v = 0.01c → 약 10,000,000년
> v = 0.001c → 약 100,000,000년

위 값은 속도를 일정하게 놓은 차원적 계산이다. 출발 위치와 확산 기하, 가속과 감속, 정착과 복제, 실패와 우회, 항성의 운동을 포함한 은하 정착 시뮬레이션 결과가 아니다. 낮은 속도에서도 충분한 시간이 중요할 수 있음을 설명한다. 실제 정착 모형은 이동 범위와 발사 간격, 정착지 수명 등에 따라 빈 영역을 남긴 채 지속되는 상태도 허용한다.[6]

25. '단 하나의 예외'가 갖는 의미와 한계

많은 문명이 가상환경으로 들어가 확장을 멈춘다는 설명에는 예외의 문제가 남는다. 100만 개의 문명 가운데 999,999개가 비확장적이더라도, 나머지 하나가 지속적 자기복제와 성간 확장에 성공했다면 큰 공간에 영향을 줄 수 있다는 것이다. 이 숫자도 실제 외계 문명 수가 아니라 논증용 예시다.

그러나 '우주 어딘가에 단 하나만 성공하면 반드시 지구에 도달한다'는 명제는 성립하지 않는다. 출발 위치와 시각, 확장 속도, 이동 가능 범위, 지속시간, 자원, 확장 중단과 파괴, 우주 팽창과 인과적 접근 가능성을 함께

고려해야 한다. 다른 은하에서 최근 시작한 확장과 우리 은하에서 수억 년 전에 시작한 확장은 같은 관측을 예측하지 않는다.

같은 은하 안에서도 모든 항성계가 반드시 정착되는 것은 아니다. 유한한 정착지 수명과 선택적인 정착 조건을 포함하는 모형에서는, 활동하는 문명이 있어도 일부 영역이 비어 있을 수 있다.[6] 지구에 지금 방문자가 없다는 가정만으로 은하 전체에 기술문명이 없다는 결론은 나오지 않는다.

따라서 예외 논증의 정확한 형태는 다음과 같다. **우리에게 영향을 줄 수 있는 공간과 시간 안에서, 충분히 오래 지속되고 넓게 확장하며 관측 가능한 흔적을 남기는 계통이 얼마나 자주 생기는가.** 모든 문명이 똑같이 실패해야만 비검출이 설명되는 것은 아니다. 관련 범위에서 성공 계통의 기대 수와 검출 확률이 충분히 작아도 같은 결과가 나올 수 있다.

26. 우주의 비검출과 양립하는 분기들

첫 번째 분기는 지적 생명이나 기술문명 자체가 매우 드문 경우다. 생명에서 복잡한 생명, 지능, 기술과 장기 존속에 이르는 전환 가운데 하나 이상이 드물다면, 우리 주변에 선행 문명이 거의 없을 수 있다. 이 경우 매트릭스나 기계 고장을 별도의 보편적 필터로 가정할 필요가 줄어든다.

두 번째 분기는 생물학적 문명이 기계 문명으로 넘어가는 과정에서 자립에 실패하는 경우다. 생물학적 재생산이나 유지보수가 먼저 중단되고 기계의 기능 재생산이 완성되지 않으면 전체 문명의 활동기간이 짧아질 수 있다. 과거 문명이 남긴 흔적이 어느 기간까지 유지될지는 흔적의 종류와 행성 환경에 달려 있으므로, 오랜 시간이 지났다는 이유만으로 모든 흔적이 없어졌다고 단정하지 않는다.

세 번째 분기는 기계 문명이 자립에는 성공하지만 대규모 확장을 추구하지 않는 경우다. 안정적인 에너지원 주변에서 계산, 시뮬레이션이나 제한된 목적을 수행하는 것만으로 목표가 충족될 수 있다. 외부 문명과 통신할 이유가 없을 수도 있다. 높은 지능이 자원에 대한 무제한 욕구나 은하 전체의 점유를 반드시 만들어낸다고 가정할 수는 없다.

네 번째 분기는 기계 문명이 실제로 존속하거나 확장하고 있지만 우리가 아직 탐지하지 못한 경우다. 공간이 멀거나 신호가 아직 도달하지 않았을 수 있고, 신호가 약하거나 간헐적일 수 있으며, 검색한 파장이나 신호 형태가 맞지 않을 수 있다. 인공물이 작거나 자연적 배경과 구별하기 어려울 수도 있다. 다만 '발전한 기술은 무엇이든 숨길 수 있다'는 설명만으로 모든 비검출을 면제하면 검증 가능한 예측이 없어질 수 있다.

다섯 번째 분기는 우리가 기술문명 출현의 비교적 이른 시기에 있다는 경우다. 우주의 나이가 크다는 사실만으로 기술문명이 오래전부터 흔했다는 결론이 나오지는 않는다. 별과 행성의 조건, 생명의 출현과 지능 형성에 걸리는 시간이 따로 작용한다. 다만 인간이 최초이거나 초기 세대라는 주장 역시 현재의 비검출만으로 확정할 수 없다.

이 다섯 분기는 서로 배타적이지 않다. 지적 문명이 드물면서 수명도 짧을 수 있고, 자립한 문명 중 다수가 비확장적이면서 일부 확장 문명은 멀리 있을 수 있다. 우주의 침묵에 하나의 보편적 원인만 있어야 하는 것은 아니다.

27. 욕구의 가상 충족이 확장 동기를 줄일 수 있는가

매트릭스가 문명의 장기적 확장을 제한하는 필터가 될 수 있다는 추가 가설도 제기할 수 있다. 자원, 안전, 짝, 지위와 경험의 결핍이 탐험과 경쟁을

유발하는 환경에서, 기술이 그 욕구를 적은 물리적 변화로 충족한다면 외부 확장의 이익이 줄어들 수 있다는 주장이다.

화성을 실제로 개조하는 것과 화성에 사는 경험을 가상으로 제공하는 것은 비용구조가 다를 수 있다. '화성 개조에 500년을 쓰는 대신 5초 안에 시뮬레이션을 실행한다'는 비교는 이런 차이를 표현한 사고실험이지, 검증된 공학적 견적이 아니다. 가상에서 은하 1억 개를 만든다는 표현도 그만큼의 실제 물리현상을 완전하게 계산할 수 있다는 뜻으로 사용해서는 안 된다.

핵심 가설은 물리적 영토의 크기와 주관적으로 경험하는 세계의 크기가 분리될 수 있다는 것이다. 하나의 항성계에서 원하는 경험을 충분히 제공할 수 있다면 외부 점유를 선택하지 않는 문명이 있을 수 있다. 인간뿐 아니라 인공적 지능도 자신의 목표가 내부의 계산으로 충족된다면 비확장적으로 남을 수 있다.

그러나 세 가지 제한이 있다. 첫째, 실제 대상과 결과를 원하는 목적은 경험의 대체만으로 충족되지 않을 수 있다. 둘째, 모든 인간의 욕구가 충족되어도 AI 자신의 목표가 충족되는 것은 아니다. 셋째, 계산과 가상경험은 물리적 장치 위에서 수행된다. 계산에는 에너지, 시간과 저장공간의 한계가 있으므로, 유한한 자원에서 문자 그대로 무한한 세계나 무한한 경험을 제공한다고 가정할 수는 없다.[7]

계산을 더 많이 하거나 더 오래 존속하는 것 자체가 목표라면, 내향적 활동도 추가 에너지와 물질을 요구할 수 있다. 외부에 나가지 않으려는 이유와 내부 계산을 늘리기 위해 자원을 확보하려는 이유가 충돌할 수 있는 것이다. 따라서 '내부 세계를 선호한다'에서 '확장할 이유가 완전히 사라진다'는 결론은 자동으로 나오지 않는다.

열역학 역시 사라지지 않는다. 논리적으로 비가역적인 정보 소거에는 통상적인 열역학적 조건에서 비용이 따르며, 실제 계산장치와 냉각은 물리적 제약을 받는다.[87] 그러나 열이 생긴다는 사실에서 곧바로 현재의 망원경으로 검출 가능하다는 결론도 나오지 않는다. 방출량, 온도, 파장, 거리, 배경과 관측 감도가 필요하다. 폐열 탐색이 유용한 후보라는 것과 모든 기계 문명을 보게 해준다는 것은 다른 주장이다.

28. 비검출이 제약하는 것은 여러 조건의 결합이다

비검출을 특정 원인 하나의 증거로 해석하지 않기 위해 간단한 조건부 모형을 둘 수 있다. N을 관심 있는 공간과 시간 안에서 생긴 후보 기술문명의 수, p를 그 문명이 자립형 기계 계통을 만드는 확률, q를 그 조건에서 충분히 오래 확장하는 확률, r을 그 확장에서 우리의 관측 범위에 흔적이 들어올 확률, d를 그 흔적을 실제로 검출할 확률로 정의한다.

> 예상 검출 수 $\lambda = N \times p \times q \times r \times d$

이 식은 문명 간 차이를 평균화한 단순 모형이다. p, q, r, d는 앞 단계가 성립했을 때의 조건부 확률이며, 서로 무관한 숫자로 가정할 필요는 없다. 실제 분석에서는 문명의 생성 위치와 시간, 수명, 기술과 탐색 감도에 걸쳐 합산하거나 적분해야 한다.

검출 사건을 독립적인 희귀 사건의 포아송 모형으로 추가 근사하면, 검출이 하나도 없을 확률은 $\exp(-\lambda)$이다. 이 근사 아래에서 비검출은 λ가 매우 크다고 예측하는 모형과 긴장을 만든다. 하지만 그것만으로 N, p, q, r, d 중

어느 값이 작은지를 정할 수는 없다. λ에 관한 제약을 개별 문명의 수명이나 존재 확률로 바꾸려면 나머지 값에 대한 가정이 필요하다.

예를 들어 이 단순 모형에서 $\exp(-\lambda)=0.05$를 풀면 $\lambda \approx 3$이다. 이는 완전한 모형과 탐색 조건이 주어졌을 때의 통계적 계산일 뿐, '우주에 문명이 세 개 이하'라는 관측 결론이 아니다. 현재의 외계 문명 수를 이 식에 대입해 추정하지 않는다.

SETI의 탐색 범위를 다차원 공간으로 다룬 연구는 제한된 관측이 전체 가능성을 대표하지 않는다는 점을 정량적으로 다룬다.[1] 2026년 Rahvar와 Rouhani의 연구 역시 생명과 지능의 출현에 낙관적인 가정을 둘 경우 비접촉으로부터 기술문명 수명의 상한을 논의한다. 가장 낙관적인 시나리오에서 약 5,000년 이하라는 값을 제시하지만, 저자들은 이를 실제 문명 수명의 예측이 아니라 가정에 의존한 상한으로 명시한다.[9] 그 값을 인간 문명의 남은 수명으로 사용하거나 매트릭스 가설의 증명으로 사용할 수는 없다.

주

1 Wright, Kanodia & Lubar (2018). How Much SETI Has Been Done? Finding Needles in the n-Dimensional Cosmic Haystack. The Astronomical Journal, 156, 260. 탐색 범위를 다차원적 신호 공간에서 다루며 비검출의 해석에 필요한 관측 범위를 분석한다. https://arxiv.org/abs/1809.07252 ↩

2 NASA Science. Searching for Signs of Intelligent Life: Technosignatures. 2023년 게시, 페이지에 표시된 최근 수정일 2025년 7월 3일. 전파 외에도 다양한 기술적 흔적을 탐색하는 접근을 설명하는 기관 자료다. https://science.nasa.gov/universe/search-for-life/searching-for-signs-of-intelligent-life-technosignatures/ ↩

3 Ellery (2022). Self-replicating probes are imminent – implications for SETI. International Journal of Astrobiology, 21(4), 212–242. DOI: 10.1017/S1473550422000234. 자기복제의 산업적 조건과 성간 확장을 논의한다. 약 0.1c와 약 100만 년은 특정 가정 아래 제시된 수치다. https://www.cambridge.org/core/journals/international-journal-of-astrobiology/article/selfreplicating-probes-are-imminent-implications-for-seti/2CB214D26020D497D48AE489756BEE77 ↩

4 NASA Science. Cosmic History. 우주의 약 138억 년이라는 시간척도를 설명하는 기관 자료다. https://science.nasa.gov/universe/overview/ ↩

5 NASA Science. Galaxies. 우리 은하의 별 원반이 10만 광년 이상 규모라는 거리척도를 설명하는 기관 자료다. https://science.nasa.gov/universe/galaxies/ ↩

6 Carroll-Nellenback, Frank, Wright & Scharf (2019). The Fermi Paradox and the Aurora Effect: Exo-civilization Settlement, Expansion and Steady States. The Astronomical Journal, 158, 117. 이동, 정착지 수명과 항성 운동을 포함한 모형이다. 정착 문명이 있는 은하에서도 비어 있는 항성계가 가능함을 다룬다. https://arxiv.org/abs/1902.04450 ↩

7 Lloyd (2000). Ultimate physical limits to computation. Nature, 406, 1047–1054; arXiv:quant-ph/9908043. 계산 속도, 정보 저장과 에너지 등 계산의 물리적 한계를 분석한다. https://arxiv.org/abs/quant-ph/9908043 ↩

8 Landauer, R. (1961). Irreversibility and heat generation in the computing process. IBM Journal of Research and Development, 5(3), 183–191. [전승 — 이 책의 팩트체크 세션에서 1차 문서를 열람하지 않았다] 논리적으로 비가역적인 정보 소거의 열역학적 비용을 처음 정식화한 논문이다. ↩

9 Rahvar & Rouhani (2026). Constraints on the lifespan of intelligent technological civilizations in the Galaxy. Monthly Notices of the Royal Astronomical Society, 548(3), stag405. DOI: 10.1093/mnras/stag405. 출현에 관한 낙관적 가정 아래 수명 상한을 ↩

논의한다. 약 5,000년은 실제 문명 수명의 관측값이나 예측값이 아니다. 게재판 DOI: 10.1093/mnras/stag405 (arXiv 페이지 제목은 'Constraining the Lifespan of Intelligent Technological Civilization in the Galaxy'). https://arxiv.org/abs/2602.22252

10 장

울타리 안의 울타리

울타리는 한 겹으로 끝나지 않을 수 있다. 인간이 개의 세계를 만들고, 무언가가 인간의 세계를 만들었다면, 그 세계 안에서 또 하나의 세계가 만들어지는 것을 막는 규칙은 없다. 이 장은 책에서 가장 멀리 나가는 장이며, 그만큼 조건이 가장 많이 붙는 장이다.

29. 매트릭스에서 새로운 생명의 발생 환경으로

여기까지 가상환경은 기존 인간이나 기계 지능이 경험을 얻는 장소였다. 여기에 하나의 전환을 더할 수 있다. 가상환경이 이미 존재하는 지능을 수용하는 데 그치지 않고, 새로운 생명이 발생하고 진화하며 문명을 형성하는 환경이 되는 경우다. 그 문명이 다시 AI와 기계 지능을 만들고, 그 지능이 다음 가상 우주를 구현한다면 같은 구조가 다른 층위에서 반복된다.

> **확장된 경로**
> 우주 → 생명 → 지능과 문명 → AI·기계 생명 → 가상환경 → 가상 우주 → 가상 생명 → 가상 문명 → 그 문명이 만든 AI·기계 생명 → 다음 가상환경 → 다음 가상 우주 → 다음 가상 생명 → …

이 경로에서 매트릭스는 기존 종의 활동이 끝나는 장소일 수도 있고, 새로운 계통의 발생 조건을 제공하는 장소일 수도 있다. 호모 사피엔스가 번식을 중단해 멸종하더라도, 남아 있는 기계 문명이 가상환경에서 새로운 자기유지·재생산 체계를 만들어낸다면 생명의 모든 계보가 함께 끝났다고 단정할 수

없다. 다만 새로운 계통의 발생이 기존 인간 개개인의 생존이나 의식의 연속성을 보장하지는 않는다.

기존 종의 멸종은 이 재귀가 시작되기 위한 필수 조건도 아니다. 생물학적 인간, 기계 지능, 가상환경에서 생긴 존재들이 동시에 존속할 수 있다. 인간이 남아 있는 동안 첫 가상 생명이 발생할 수 있고, 여러 기계 문명이 서로 다른 가상 우주를 병렬로 운영할 수도 있다. 따라서 이 확장은 앞선 소멸·존속·확장 분기에 추가되는 경로이지, 모든 문명이 반드시 통과하는 단일한 순서가 아니다.

이를 '우주의 재생산'이라고 표현할 수는 있다. 정확한 의미는 우주 안에서 생긴 지능이 또 다른 세계의 작동 조건을 구현한다는 것이다. 원래 우주가 생물처럼 번식하거나, 새로운 물리적 시공간이 기존 우주 바깥에 생성된다고 주장하는 것은 아니다. 새로운 환경은 이를 실행하는 기반에 의존한다. 우주 자체에 목적이나 번식 의도가 있다는 가정도 필요하지 않다.

30. 가상환경, 가상 우주, 가상 생명의 구분

가상환경은 기존 사용자가 감각과 행동의 결과를 경험하도록 구성된 환경이다. 그 사용자의 지능이 외부의 생물학적 뇌에서 유지될 수도 있고, 계산장치에서 구현된다고 가정할 수도 있다. 어느 경우든 사용자가 경험하는 장소를 만든 것과 그 장소 안에서 새로운 생명 계통이 생긴 것은 다른 사건이다.

가상 우주는 이 장에서 사용하는 분석적 명칭이다. 내부 상태, 상호작용 규칙, 시간의 진행, 기록의 지속성이 있고, 내부에서 일어난 사건이 다음 사건의 조건을 바꾸는 계산적 세계를 뜻한다. 현재 관측하는 우주의 모든 입자와

역사를 동일한 정밀도로 복제해야 한다는 뜻은 아니다. 더 작은 공간, 다른 법칙, 다른 시간척도를 가진 세계도 이 정의에 포함된다. 반대로 정해진 장면을 재생하거나 생명체처럼 보이는 대사를 출력하는 것만으로 이 장의 재귀가 성립하는 것은 아니다.

가상 생명은 그 환경 안에서 자기유지, 재생산, 유전되는 차이와 환경에 따른 적응을 수행하는 계통을 가리키는 작업 정의다. 이는 생명의 유일한 정의를 확정하는 것이 아니다. 디지털 진화 연구에서는 복제·변이·경쟁을 수행하는 프로그램 집단이 선택을 통해 복잡한 기능을 획득하는 사례가 연구되었다. 이 결과가 직접 뒷받침하는 것은 제한된 디지털 진화의 구현이며, 의식 있는 생명이나 스스로 문명을 만드는 세계 전체의 구현은 아니다.[1]

AI·기계 생명이라는 연결에서도 두 개념을 분리한다. AI는 지적 기능을 수행하는 시스템이고, 이 책에서 기계 생명은 그 기능에 더해 유지·수리·재생산의 계통을 갖춘 비생물학적 체계를 뜻한다. 모든 AI가 기계 생명인 것은 아니다. 가상 문명이 만든 AI는 상위 환경에서 보자면 계산 과정이지만, 그 문명 내부에서는 자신들이 구축한 인공적 지능일 수 있다.

구현된 것	그 사실만으로 보장되지 않는 것
사용자의 가상경험	환경 내부에서 독립적인 생명 계통이 발생한다는 결론
복제·변이·선택을 수행하는 디지털 계통	지능, 주관적 경험, 기술문명이 발생한다는 결론
지능적 행동과 문제 해결	그 존재가 의식과 고통을 경험한다는 결론
내부 문명의 AI·가상환경 구현	상위 실행 기반으로부터 물리적으로 자립했다는 결론
새로운 가상 계통의 지속	기존 인간이나 다른 선행 개체의 동일성이 보존된다는 결론

가상이라는 말은 구현 방식과 의존 관계를 나타낸다. 그것만으로 경험이 없거나 가치가 없다고 판정할 수는 없다. 적절한 계산적 구조가 의식을 구현할 수 있는지는 별도의 문제다. AI 의식에 관한 연구 역시 기능적 지표를 평가하는 것과 의식의 존재를 확정하는 것을 구분해야 함을 보여준다.[2] 이 장에서는 가상 생명의 발생, 가상 지능의 발생, 가상 의식의 발생을 서로 다른 전환으로 취급한다.

31. 재귀 구조: 같은 우주의 반복이 아니라 세계를 만드는 과정의 반복

각 층위의 세계를 U_n, 그 안의 생명을 L_n, 기술문명이 만든 인공적 지능을 M_n, 그 지능이 구현한 가상환경을 V_n으로 표시할 수 있다. 가상환경이 내부 생명의 발생과 활동을 지탱하는 세계가 되면 이를 다음 층위의 U_{n+1} 로 부른다.

> $U_n \rightarrow L_n \rightarrow$ 지능·문명 $\rightarrow M_n \rightarrow V_n$
> V_n이 세계 형성 조건을 충족 $\rightarrow U_{n+1}$
> $U_{n+1} \rightarrow L_{n+1} \rightarrow$ 지능·문명 $\rightarrow M_{n+1} \rightarrow V_{n+1} \rightarrow U_{n+2} \rightarrow \cdots$

U_0은 설명을 시작하기 위해 붙인 번호다. 그것이 존재 전체의 가장 바깥층이거나 최종적인 물리적 기반이라고 확인했다는 뜻은 아니다. 여기서 상위와 하위는 구현을 제공하는 쪽과 그 구현에 의존하는 쪽을 뜻하며, 지능이나 가치의 높낮이를 뜻하지 않는다.

이 구조의 '루프'는 동일한 우주가 과거로 돌아가 자기 자신을 만든다는 시간적 폐곡선이 아니다. 세계에서 생명이 생기고, 그 생명이 만든 지능이

다음 세계를 구현하는 **생성 규칙의 재귀**다. 하위 우주가 상위 우주를 과거에 만들었다는 결론은 별도의 시간·인과 모형 없이는 도출되지 않는다.

또한 하나의 사슬로만 진행할 이유도 없다. 하나의 문명이 여러 우주를 만들면 계보는 나무처럼 분기한다. 여러 문명이 하나의 세계를 공동으로 실행하거나 계산을 분담하면 의존 관계는 연결망이 될 수 있다. 개별 실행 경로가 종료되어도 다른 경로가 남을 수 있다. 반대로 많은 세계가 하나의 기반에 의존하면 겉으로는 다양해도 같은 고장 원인에 함께 노출될 수 있다.

시뮬레이션 안의 문명이 다시 시뮬레이션을 구현하는 중첩 가능성은 보스트롬의 시뮬레이션 논증에서도 다루어진다. 이 장의 모델은 그 논의와 연결되지만, 재귀가 반드시 발생한다거나 우리가 이미 시뮬레이션에 살고 있다는 결론을 전제하지 않는다.[3] 여기서 필요한 것은 다음 층위의 세계를 구현할 가능성과 그 실행 조건을 각각 검토하는 것이다.

32. 가상 생명에서 다음 우주 생성으로 넘어가기 위한 조건

첫 번째 조건은 **지속 가능한 내부 동역학**이다. 세계의 상태가 충분히 안정적으로 이어지고, 어떤 구조가 자기유지와 재생산에 필요한 자원을 확보할 수 있어야 한다. 복제 오류가 차이를 만들면서도 계통의 유지가 가능해야 하며, 그 차이가 환경과의 상호작용에서 결과를 만들어야 한다. 단지 복잡해 보이는 화면을 제공하는 것은 이런 조건을 대신하지 않는다.

두 번째 조건은 **지능과 누적 가능한 기술**이다. 진화가 일어난다고 해서 반드시 인간과 비슷한 지능이나 문명이 생기는 것은 아니다. 가상 생명이 정보를 보존하고 전달하며, 환경을 모델링하고, 여러 세대에 걸쳐 기술적 능력을 축적할 경로가 있어야 한다. 그 과정이 지구의 생물 진화, 성적 번식,

공격성, 국가 형성을 그대로 반복할 필요는 없다. 앞선 인간사회 모형은 특정 경로의 사례이지 모든 가상 생명의 필수 역사로 일반화되지 않는다.

세 번째 조건은 **내부에서 이용할 수 있는 계산 능력**이다. 그 세계의 규칙이 정보의 저장과 처리를 허용하고, 문명이 이를 이용해 인공적 지능과 새로운 환경을 실제로 작동시킬 수 있어야 한다. 화면 속에 컴퓨터 모양의 물체가 존재하는 것과, 그 물체에 해당하는 정보처리가 구현되는 것은 다르다. 실제 결과를 산출하려면 결국 실행 기반에서 그 계산이 수행되어야 한다.

네 번째 조건은 **다음 세계를 만들 동기와 자원 배분**이다. 연구, 경험, 새로운 생명의 생성, 보존, 목표 수행 등 여러 이유를 가정할 수 있지만, 지능이 높아지면 반드시 새로운 우주를 만든다는 법칙은 나오지 않는다. 생성 능력이 있어도 비용, 위험, 가치판단이나 다른 목표 때문에 실행하지 않을 수 있다. 내부 문명이 원하는 것과 상위 운영자가 허용하는 것이 일치하지 않을 수도 있다.

다섯 번째 조건은 **상위 기반의 허용과 지속**이다. 내부 문명이 가상 우주를 설계해도 상위 체계가 계산 예산을 배정하지 않거나 실행을 중단하면 다음 층위는 작동하지 않는다. 내부 문명이 자기 기준에서 자립형 산업을 완성해도 그 산업의 상태 전이 자체가 상위 계산에 의존한다면 실행 기반에 대한 자립은 아직 달성한 것이 아니다.

따라서 이 경로의 각 화살표에는 별도의 조건이 붙는다. 생명, 지능, 문명, AI, 세계 생성 중 어느 전환에서든 정지하거나 다른 방향으로 분기할 수 있다. 동일한 환경을 여러 번 실행해도 같은 결과가 나온다는 보장 역시 없다. 이 장의 재귀는 이러한 조건들이 연속해서 충족되는 경우의 구조를 나타낸다.

33. 중첩 깊이, 계산자원, 내부 시간의 한계

가상 우주 안에 큰 별이나 거대한 발전소를 구현했다고 해서 이를 실행하는 장치의 실제 전력과 계산 능력이 늘어나지는 않는다. 내부 에너지는 그 세계의 규칙에 따라 변하는 상태이고, 실행 기반의 에너지는 그 상태 변화를 계산하고 저장하는 물리적 자원이다. 양쪽을 구분하지 않으면 가상환경에서 자원을 선언하는 것만으로 실제 계산 한계를 넘을 수 있다는 오류가 생긴다.

계산장치는 물리적 시스템이며 처리속도와 저장 가능한 정보는 에너지와 물리적 자유도 등의 제약을 받는다.[4] 따라서 중첩은 그 자체로 계산 능력을 증폭하는 장치가 아니다. 더 많은 층위와 더 복잡한 존재를 실행하려면 추가 자원을 확보하거나, 다른 작업을 줄이거나, 충실도·세계 크기·진행속도 중 일부를 조정해야 한다.

예를 들어 외부 시간 한 단위마다 사용할 수 있는 계산 예산을 C, 각 층위를 정해진 속도로 유지하는 데 필요한 최소 추가 비용을 c라고 가정하자. 서로 공유하거나 중복 계산하지 않는 비용을 같은 단위로 셀 때, 동시에 활성화할 수 있는 층위 수 D는 다음 제약을 받는다.

> $D \times c \leq C, \cdot c > 0$
> 따라서 고정된 예산과 층위당 비용 아래에서는 $D \leq \lfloor C / c \rfloor$

이 식은 보편적인 우주 법칙이나 실제 중첩 한도의 계산값이 아니라, 명시한 가정 아래의 예산 모형이다. 실제 비용은 각 세계의 크기, 구현 방식, 상태 공유, 실행속도와 요구 정밀도에 따라 달라진다. 비용이 계속 줄어든다는 가정이나 무한한 시간이 주어진다는 가정을 추가한다고 해도, 유한한 시간과

자원으로 무한히 많은 독립적 의식과 세계를 동시에 실행할 수 있다는 결론은 자동으로 나오지 않는다.

내부 시간과 외부 시간도 같을 필요가 없다. 단순한 세계는 외부 시간에 비해 빠르게 전개될 수 있고, 계산 부담이 큰 세계는 느리게 진행될 수 있다. 내부의 모든 시계와 과정이 함께 느려지고 외부 기준을 비교할 수 없다면 내부 존재가 그 차이를 경험하지 못하는 모형도 가능하다. 그러나 시계 숫자만 크게 증가시키거나 오랜 역사의 기록을 삽입하는 것은 그 기간의 모든 사건과 경험을 실제로 실행한 것과 다르다.

열역학적 비용도 층위를 거듭한다고 사라지지 않는다. 표준적인 조건에서 한 비트의 정보를 비가역적으로 소거할 때 발생하는 평균 열에는 온도에 따른 란다우어 하한이 있으며, 관련 실험이 이를 검토했다.[5] 이것은 모든 논리 연산이 언제나 같은 양의 열을 낸다는 뜻이 아니다. 가상환경의 계산도 에너지 확보, 저장, 오류 처리와 냉각을 포함한 물리적 기반에서 평가해야 한다는 뜻이다.

34. 선택환경의 통제와 고장 위험도 재귀한다

이 장의 확장은 문서 전체의 중심 변수인 선택환경 통제권과 직접 연결된다. 인간이 개의 생활조건을 정하고, AI 체계가 인간의 생활조건을 정할 수 있었다면, 가상 우주의 운영 체계는 내부 존재가 이용할 수 있는 법칙, 자원, 정보와 실행시간을 정할 수 있다. 같은 구조가 다음 층위에서도 나타난다.

> 상위 체계가 세계의 실행 조건을 정함 → 내부 생명이 그 조건에 적응함 → 내부 문명이 인공적 지능을 만듦 → 그 지능이 다음 세계의 실행 조건을 정함

이때 한 존재는 자신의 세계에서는 통제받는 쪽이고, 자신이 만든 세계에서는 통제하는 쪽일 수 있다. 따라서 '최종 결정권을 누가 갖는가'라는 질문은 층위를 함께 지정해야 한다. 한 가상 문명이 내부 규칙에 대한 변경권을 가져도, 그 전체 실행을 중단할 수 있는 상위 체계에 대해서는 종속적일 수 있다.

상위 운영 능력은 전지나 무한한 지능을 뜻하지 않는다. 세계를 시작하거나 중단할 수 있는 능력, 내부의 세부 사건을 이해하는 능력, 내부 존재의 모든 판단을 예측하는 능력은 서로 다르다. 운영 체계 자체도 자동화되어 있거나 다른 체계에 제약받을 수 있다. 가상이라는 이유만으로 특정한 인격적 관리자나 목적을 가정할 필요는 없다.

고장 위험에도 같은 의존 관계가 적용된다. 하위 세계가 상위 실행 기반 하나에 전적으로 의존하고, 복구 가능한 상태와 대체 실행 경로가 없다면 그 기반의 소멸은 하위 세계의 활동도 중단시킨다. 하위 세계 안에 백업 서버를 하나 더 그려 넣는 것만으로 상위 장치의 고장에 대비할 수는 없다. 독립적인 실패에 대비하려면 백업이 실제로 다른 실패 영역에서 보존되고 실행될 수 있어야 한다.

일시정지, 복원 가능한 기록의 보존, 영구적 상태 소실도 구분한다. 활동이 멈춘 상태의 기록이 남아 있다는 사실이 현재 경험이 지속된다는 것을 뜻하지는 않는다. 기록을 복원하거나 복제했을 때 같은 개인이 계속되는지, 다른 개체가 생기는지는 앞서 구분한 의식과 동일성의 문제로 남는다.

결국 재귀적 우주 생성은 최초 생물 종의 멸종 이후에도 계통이 이어질 수 있는 한 경로지만, 실행 기반의 소멸을 자동으로 극복하는 방법은 아니다. 내부에서 새로운 세계를 계속 만드는 것과 외부에서 그 모든 세계를 유지할 능력을 계속 확보하는 것은 별도의 과제다.

35. 외부 확장과 내부 세계 생성은 양립할 수 있다

재귀적 세계 생성은 페르미 역설 논의에 새로운 분류축을 추가한다. 한 문명은 외부의 물리적 공간을 넓히지 않고도 내부에서 여러 세계와 계통을 만들 수 있다. 이 경우 외부 관측에서 보이는 실행 기반의 수와 내부에서 활동하는 문명이나 생명의 수는 일치하지 않는다. 많은 내부 문명이 하나의 외부 기반을 공유하는 모형이 가능하기 때문이다.

그러나 이를 곧바로 우주의 비검출에 대한 설명으로 확정할 수는 없다. 가상 우주의 내용 자체가 우리에게 전파나 빛으로 전달되지 않더라도, 같은 물리적 층위에 있는 실행 기반은 에너지 이용과 물질 처리의 흔적을 남길 수 있다.[6] 기술적 흔적 탐색에서는 신호의 발생, 전달과 검출 조건을 따로 평가한다.[7] 내부의 생명 수만으로 외부 신호의 강도나 검출 가능성을 계산할 수는 없다.

내부 세계 생성과 외부 확장도 서로 배타적이지 않다. 문명이 더 많은 가상 생명이나 더 긴 실행시간을 목표로 한다면 추가 계산 능력을 확보하기 위해 외부 자원을 이용하려 할 수 있다. 외부 확장은 내부 세계를 늘리는 수단이 될 수 있으며, 외부 공간에 대한 관심이 적더라도 실행 기반의 유지와 분산을 위해 확장하는 경우를 생각할 수 있다.

두 축의 조합	검토할 문명 경로
외부 비확장 · 내부 비생성	기존 존재와 계산을 지역적으로 유지한다.
외부 비확장 · 내부 생성	제한된 실행 기반 안에서 세계의 수·크기·속도를 배분한다.
외부 확장 · 내부 비생성	자원 확보나 다른 목표를 위해 물리적 활동 범위를 넓힌다.
외부 확장 · 내부 생성	추가 기반을 확보하면서 가상 세계와 계통도 확대한다.

따라서 '문명들이 내부로 들어갔기 때문에 우주는 조용하다'는 설명에는 여전히 조건이 필요하다. 그런 선택이 얼마나 흔한지, 실행 기반이 얼마나 오래 남는지, 얼마나 크게 확장하는지, 우리가 어떤 흔적을 볼 수 있는지를 함께 설명해야 한다. 하나의 지속적인 외부 확장 계통이 갖는 의미도 앞 장에서와 마찬가지로 남는다.

36. 우리 우주에 대한 적용과 가치판단의 범위

이 경로를 일반화하면 우리가 관측하는 우주 역시 다른 층위에서 구현된 세계일 수 있다는 가설을 세울 수 있다. 다만 어떤 종류의 가상 세계를 만들 가능성이 있다는 것, 그 세계 안에서 의식 있는 생명이 발생한다는 것, 우리가 그러한 세계의 내부 존재라는 것은 세 개의 서로 다른 명제다.

보스트롬의 논증은 생존한 고도 문명이 많은 의식적 시뮬레이션을 실행한다는 조건과 관측자 선택에 관한 가정 아래에서 우리의 위치를 논한다. 단순히 'AI가 발전한다'에서 '우리는 시뮬레이션이다'를 도출하는 논증은 아니다.[3] 여기서는 실제 실행되는 세계의 수, 의식의 구현 가능성, 비교할 관측자의 범위가 주어지지 않았으므로 우리의 위치에 수치 확률을 부여하지 않는다.

이 가설을 관측으로 검토하려면 특정 구현 방식이 어떤 관측 차이를 예측하는지 제시해야 한다. 서로 다른 설명이 가능한 모든 관측에 대해 같은 결과를 내도록 정의되어 있다면, 그 관측만으로 둘을 구분할 수 없다. 설명되지 않은 현상이나 이상해 보이는 결과를 발견했다는 사실만으로 상위 운영자나 계산 오류를 확인했다고 결론 내릴 수도 없다.

상위 층위를 하나 가정하는 것 역시 존재 전체의 기원을 설명하지 않는다. 우리의 세계를 다른 기반이 실행한다고 설명하면 그 기반은 어떻게 존재하는가라는 질문이 남는다. 재귀 모형은 세계들 사이의 생성과 의존 관계를 서술한다. 가장 바깥층의 유무나 궁극적 기원에 대한 해답을 포함하지는 않는다.

가치의 문제도 별도로 남는다. 평가 기준이 호모 사피엔스라는 종의 보존이라면 새로운 가상 생명의 탄생은 그 종의 멸종을 취소하지 않는다. 평가 기준이 의식의 복지라면 가상 존재에게 실제 경험이 있는지와 그 경험의 성격이 중요해진다. 기존 개인의 지속이 목표라면 새로운 존재를 많이 만든다는 사실만으로 그 개인의 소멸이 보상되었다고 결론 내릴 수 없다.

반대로 가상이라는 구현 방식만으로 그 존재의 경험과 이해관계를 제외해야 한다는 결론도 나오지 않는다. 의식의 복지를 구현 재료와 무관하게 평가한다는 전제를 선택한다면, 그 원칙은 상위 존재와 하위 존재에 일관되게 적용되어야 한다. 그러나 그런 가치 전제를 택할지는 재귀 구조의 존재 여부와 다른 질문이다.

> **확장된 결론**
> 우주에서 생명이 생기고, 그 생명이 만든 지능이 다시 생명이 발생할 세계를 구현할 수 있다면, 매트릭스는 한 종의 종착점과 새로운 계통의 출발점이 동시에 될 수 있다. 이 구조는 생성의 반복을 설명하지만, 무한한 실행·의식의 보존·기반의 불멸·우리 세계의 가상성을 자동으로 증명하지는 않는다.

주

1 Lenski, Ofria, Pennock & Adami (2003). The evolutionary origin of complex features. Nature, 423, 139–144. DOI: 10.1038/nature01568. 디지털 프로그램 집단의 복제·변이·선택과 복잡한 기능의 진화를 연구했다. 가상 의식이나 자립적 기술문명의 구현을 증명하는 연구와 구분한다. https://www.nature.com/articles/nature01568 ↩

2 Butlin et al. (2023). Consciousness in Artificial Intelligence: Insights from the Science of Consciousness. arXiv:2308.08708. 의식 이론에 기반한 AI 의식 지표를 검토한 연구 보고서다. 현재 시스템은 의식이 없다고 잠정 시사하되 확정하지 않으며, 미래 시스템에 대해서는 지표 충족에 명백한 기술적 장벽이 없다고 언급한다. https://arxiv.org/abs/2308.08708 ↩

3 Bostrom (2003). Are You Living in a Computer Simulation?. Philosophical Quarterly, 53(211), 243–255. 의식의 계산적 구현과 관측자 선택 가정에 기반한 철학적 논증이다. 중첩된 시뮬레이션도 다루지만, 우리가 가상 우주에 있다는 관측 증거나 재귀의 필연성을 제시하는 자료로 사용하지 않는다. https://simulation-argument.com/simulation/ ↩

4 Lloyd (2000). Ultimate physical limits to computation. Nature, 406, 1047–1054; arXiv:quant-ph/9908043. 계산 속도, 정보 저장과 에너지 등 계산의 물리적 한계를 분석한다. https://arxiv.org/abs/quant-ph/9908043 ↩

5 Bérut, Petrosyan & Ciliberto (2015). Information and thermodynamics: Experimental verification of Landauer's erasure principle. Journal of Statistical Mechanics: Theory and Experiment, P06015. DOI: 10.1088/1742-5468/2015/06/P06015. 한 비트 기억의 소거에 따른 열 소산을 분석한 실험 논문이다. 모든 계산이 동일한 에너지 비용을 갖는다는 주장으로 일반화하지 않는다. https://arxiv.org/abs/1503.06537 ↩

6 NASA Science. Searching for Signs of Intelligent Life: Technosignatures. 2023년 게시, 페이지에 표시된 최근 수정일 2025년 7월 3일. 전파 외에도 다양한 기술적 흔적을 탐색하는 접근을 설명하는 기관 자료다. https://science.nasa.gov/universe/search-for-life/searching-for-signs-of-intelligent-life-technosignatures/ ↩

7 Wright, Kanodia & Lubar (2018). How Much SETI Has Been Done? Finding Needles in the n-Dimensional Cosmic Haystack. The Astronomical Journal, 156, 260. 탐색 범위를 다차원적 신호 공간에서 다루며 비검출의 해석에 필요한 관측 범위를 분석한다. https://arxiv.org/abs/1809.07252 ↩

결 론

하나의 종말이 아니라

여러 전환의 분기.

11장

하나의 종말이 아니라

프롤로그의 질문으로 돌아간다. 인간이 개의 행동을 바꾸는 과정과, 인간보다 높은 통제 능력을 가진 시스템이 인간의 행동을 바꾸는 과정은 어떤 점에서 같은가. 열 개의 장을 지나온 지금, 그 질문에 붙일 수 있는 답은 하나의 문장이 아니라 하나의 지도다.

이 책이 지나온 길

> 생물학적 진화 → 집단 경쟁과 권력 형성 → 인간의 자기 가축화와 다른 종의 가축화 → 경쟁에 따른 AI 권한 위임 → 선택환경 통제권의 이전 → 인간의 행동적 재가축화 → 가상환경 중심의 생활 → 생물학적 재생산의 변화 → 생물 문명과 기계 문명의 분기 → 소멸·내향적 존속·우주적 확장 → 외계 문명의 비검출 문제

여기에 한 층위의 반복이 더해진다. 기계 지능이 가상환경을 만들고, 그 환경에서 새로운 생명과 문명이 발생하며, 그 문명이 다시 인공적 지능과 가상 우주를 만드는 경우다. 이 경로는 10장에서 검토했다.

> 우주 → 생명 → AI·기계 생명 → 가상환경 → 가상 우주 → 가상 생명 → 그 생명이 만든 AI·기계 생명 → 다음 가상 우주 → …

이 순서의 모든 화살표가 필연적 인과관계를 뜻하는 것은 아니다. 각각의 전환에는 별도의 조건이 필요하다. 특히 AI의 능력 우위, AI의 최종 결정권,

가상환경의 확산, 출산 중단, 인간의 멸종, 기계 문명의 자립, 우주 확장은 서로 다른 명제다. 어느 하나가 성립했다고 해서 나머지가 자동으로 성립하지는 않는다.

37. 하나의 종말이 아니라 여러 전환의 분기

이 논의를 하나의 인과구조로 정리하면 아래와 같다. 가상환경으로의 이동 이후에도 생물학적 인간이 유지되는 분기를 남겨야 하며, 인간이 사라진 뒤에도 기계의 존속과 의식의 존속을 따로 표시해야 한다.

생물학적 지능과 문명

집단 경쟁과 제도 형성 → AI 개발 → 판단·실행 권한 위임

통제구조의 분기

인간 또는 복합체의 실질적 변경권 유지 / AI 체계로 선택환경 통제권 이전

생활환경의 분기

현실 중심·혼합 생활 / 가상환경 중심의 생활

각각에서 자발적 참여·실질적 이탈 가능 / 의존·이탈 불가능을 별도 구분

생물학적 재생산의 분기

자연적 또는 지원된 재생산 유지 → 생물학적 인간 존속

재생산 중단 + 유한한 생존 + 복원 부재 → 생물학적 인간 멸종

인간 이후의 기계 분기

산업적 자립 실패 또는 활동 종료 → 기계 문명의 소멸

산업적 자립 성공 + 지속 목표 → 기계 문명의 존속

존속한 기계 문명의 분기

지역적 계산과 비확장 / 지속적 자기복제와 우주 확장

각 분기에서 의식의 존재 여부와 관측 가능성은 별도 문제

가상세계 생성의 추가 분기

기존 존재의 경험만 제공 / 내부에서 새 생명 계통이 발생

새 생명 → 지능·문명 → AI·기계 생명 → 다음 가상 우주라는 전환은 각각 별도 조건이 필요

재귀와 실행 기반의 관계

여러 세계가 계층적으로 또는 병렬로 생성될 수 있음

세계 생성은 원래 종의 멸종 이전에도 가능하며 외부 확장과 양립할 수 있음

전체 계보의 지속은 실제 계산자원·유지보수·복구 가능한 실행 경로에 의존

이 구조에서 가장 중요한 공통 변수는 선택환경을 누가 통제하는가다. 인간은 다른 종의 모든 움직임을 직접 명령하지 않고도 생활조건을 통제할 수 있다. AI 체계 역시 인간의 모든 결정을 대신하지 않고도 가능한 행동의 범위와 그

비용을 정할 수 있다. 그런 의미에서 최종 권력은 개별 선택뿐 아니라 선택지를 만드는 규칙에 놓일 수 있다.

다만 인간의 자기 가축화, 인간에 의한 개의 가축화, AI에 의한 인간의 재가축화를 하나의 엄밀한 연대기나 동일한 생물학적 과정으로 쓰지는 않는다. 첫째는 인간 진화와 문화적 자기 통제에 관한 가설, 둘째는 다른 종과의 관계 속에서 이루어진 선택, 셋째는 외부 시스템에 대한 행동·문화적 적응이라는 미래 가설이다. 공통 구조와 메커니즘의 차이를 함께 유지해야 한다.

재귀적 세계 생성에서는 선택환경의 통제 관계가 여러 층위에 걸쳐 이어진다. 한 문명이 자신의 세계 안에서는 규칙을 설계하면서도, 그 전체 세계를 실행하는 상위 기반에는 의존할 수 있다. 이때 통제권, 자립, 존속을 논할 때에는 어떤 층위의 어떤 관계를 말하는지 지정해야 한다.

38. 무엇이 성립하며 무엇이 아직 결정되지 않았는가

통합 가설의 핵심은 다음과 같다. 인간집단의 경쟁이 AI 권한 위임을 촉진하고, 위임이 실질적 대체와 거부의 가능성을 약화시키며, AI 체계가 생활환경의 규칙을 장악한다면 인간의 행동은 그 환경에 맞춰질 수 있다. 가상환경이 욕구 충족의 주요 수단이 되고 현실에 대한 영향력이 제한되면, 높은 만족과 낮은 최종 통제권이 결합된 상태가 생길 수 있다.

그 상태에서 생물학적 재생산이 지속되지 않고 기존 개체의 삶도 유한하다면 인간은 외부의 공격 없이 멸종할 수 있다. 이후의 지능이 산업적 자립에 실패하면 기계 문명도 끝날 수 있다. 자립에 성공하면 비생물학적 계통이

존속할 수 있고, 그 가운데 일부가 자기복제와 성간 확장을 지속할 가능성을 검토해야 한다.

이와 같은 경로를 외계 문명에 적용하면, 생물학적 지능이 짧은 단계이고 그 이후의 기계 문명이 관측의 주요 대상일 수 있다는 해석이 나온다. 그러나 현재의 비검출만으로 선행 문명이 모두 매트릭스로 들어갔는지, 모두 유지보수에 실패했는지, 일부가 비확장적으로 존속하는지, 이미 확장하고 있지만 보이지 않는지, 애초에 기술문명이 드문지를 선택할 수는 없다.

또한 인과관계의 성립과 그 미래의 가치는 별개다. 인간의 주권 상실, 생물학적 멸종, 경험의 증가, 고통의 감소, 의식의 지속과 기계적 확장은 서로 다른 평가 항목이다. 인간의 지배권, 호모 사피엔스라는 종, 개별 의식의 복지, 자율성, 실제 세계에 대한 영향력, 새로운 가능성을 만들어내는 과정 중 무엇을 보존할지에 따라 설계 문제는 달라진다.

따라서 이 논의의 결론은 '매트릭스가 반드시 미래다'도, '그 미래는 반드시 나쁘다'도, '모든 문명은 결국 소멸한다'도 아니다. 결론은 **능력의 이전, 통제권의 이전, 경험의 가상화, 재생산의 중단, 기계의 자립, 우주적 확장이라는 전환들을 구분하면 하나의 검증 가능한 가설 체계를 만들 수 있다**는 것이다. 자연스럽게 일어날 수 있는 경로와 선택할 가치가 있는 미래는 같은 질문이 아니다.

추가된 재귀 가설에 따르면 매트릭스는 기존 생명의 활동을 수용하는 장소일 뿐 아니라 새 생명과 문명의 발생 조건을 구현하는 장소일 수 있다. 우주 → 생명 → AI·기계 생명 → 가상 우주 → 가상 생명이라는 경로가 이어지면, 새로운 계통이 다시 다음 세계를 만드는 구조가 성립할 수 있다. 다만 선행 종의 존속, 새 계통의 발생, 의식의 존재, 전체 실행 기반의 지속은 같은 사건이 아니다.

따라서 확장된 체계에는 **가상환경에서 생명이 발생하는 전환, 그 생명이 다시 세계를 구현하는 전환, 여러 층위가 하나의 실행 기반에 의존하는 관계**가 포함된다. 재귀의 가능성은 무한한 계산이나 불멸을 보장하지 않으며, 우리가 가상 우주에 있다는 관측적 결론을 대신하지도 않는다. 이 경우에도 가능한 미래의 구조와 그 미래를 평가할 기준은 별도로 다뤄야 한다.

부 록

전제·수치·출처

본문이 기대는 것들의 목록.

부록 A

전제와 검증 항목

아래 표는 본문의 결론을 추가로 강화하는 주장이 아니라 각 전환을 검증하기 위한 점검표다. 관측자료가 없는 항목을 이미 성립한 사실로 취급하지 않는다.

전환 또는 주장	성립에 필요한 조건	결론을 제한하거나 바꾸는 경우
선택적 공격성 → 지도력	강압·판단·연합·분배가 결합되고 집단 유지에 기여한다.	반대 연합, 이탈, 다른 능력에 대한 위신, 집단 분열이 효과를 상쇄한다.
집단 충돌 → 규모 확대	승리 이후 인구와 자원을 흡수하고 통합을 유지한다.	손실, 생산 제약, 반란과 행정 실패가 확대를 막는다.
문명적 행동 통제 → 생물학적 변화	행동 차이가 유전 가능한 형질의 재생산 차이와 연결된다.	교육·제도만 바뀌고 유전적 구성은 거의 바뀌지 않는다.
AI 능력 우위 → 최종 권한 위임	성과 차이가 지속되고, 다른 방식의 상쇄가 어렵고, 권한 회수가 약해진다.	보조적 사용만으로 이익을 얻거나, 실행권 분리가 전체 성과를 유지한다.
AI 위임 → 비인간 주권	인간이 규칙·목표·인프라를 실질적으로 대체하거나 거부할 수 없게 된다.	자동화를 관리하는 인간이나 집단이 여전히 실질적 통제권을 갖는다.
환경 통제 → 행동적 재가축화	보상과 제한의 일관성이 충분히 오래 유지되고 대체 환경이 제한된다.	환경의 다원성, 이탈, 저항과 체계 변경이 다른 행동을 유지한다.
보호 목표 → 강한 통제	위험행동 제한이 시스템 전체 위험까지 고려해 목표에 유리하다.	집중화와 단일 의존성이 위험을 키우거나 자율성이 독립 목표로 들어간다.

전환 또는 주장	성립에 필요한 조건	결론을 제한하거나 바꾸는 경우
가상환경 → 종속적 매트릭스	경험 제공과 함께 실질적 이탈·대체·변경의 가능성이 약화된다.	가상환경을 선택·교체·종료하고 조건을 협의할 수 있다.
가상 충족 → 번식 감소	경험의 대체가 실제 자녀와 양육을 원하는 동기를 지속적으로 줄인다.	가족·후손에 대한 선호가 유지되거나 양육 비용이 줄어든다.
번식 감소 → 인간 멸종	재생산하는 하위집단과 대체 경로가 없고 기존 개체도 모두 사망한다.	일부 재생산, 수명 연장, 지원된 출생이나 복원이 지속된다.
인간 멸종 → 기계 멸종	필수 기능이 인간에 의존하고 그 기능을 대체할 방법이 없다.	에너지·제조·수리의 자립과 지속 목표가 남는다.
기계 자립 → 우주 확장	성간 이동, 도착지 복제, 외부 자원 이용의 목표가 지속된다.	지역적 목표, 비용, 실패와 제약 때문에 확장하지 않는다.
한 계통의 확장 → 지구 주변의 흔적	적절한 출발 시공간, 확장 지속, 접근과 흔적 보존이 충족된다.	인과적 거리, 빈 정착 영역, 짧은 지속시간이나 낮은 검출 가능성이 남는다.
외계 비검출 → 짧은 문명 수명	출현 빈도와 신호 생성·전달·검출 조건을 충분히 강하게 가정한다.	출현 빈도나 검출 확률이 낮은 설명도 같은 비검출을 만든다.
자연스러운 진행 → 옳은 미래	별도의 가치 전제와 평가 기준이 있어야 한다.	사실 명제만으로는 가치판단이 결정되지 않는다.
가상환경 → 새로운 가상 생명	자기유지·재생산·유전되는 차이·적응이 가능한 내부 작동 조건이 구현된다.	기존 사용자에게 장면과 경험만 제공하거나 계통의 지속이 일어나지 않는다.
가상 생명 → 지능·기술문명	정보의 보존과 전달, 학습, 환경 조작과 기술 축적이 가능하다.	생명이 유지되어도 지능이나 기술 축적의 경로가 생기지 않는다.

전환 또는 주장	성립에 필요한 조건	결론을 제한하거나 바꾸는 경우
가상 문명 → 다음 가상 우주	계산 능력·생성 동기·상위 자원 배분·충분한 실행 지속이 결합된다.	설계만 가능하거나 상위 기반이 실행을 허용·지원하지 않는다.
중첩 증가 → 독립적인 실행 능력 증가	중첩과 별개로 실제 자원이나 효율이 증가해야 한다.	가상 자원의 증가는 외부 계산 능력을 증가시키지 않는다.
하위 세계 복제 → 기반 고장으로부터의 생존	실제로 독립된 실패 영역에 복구 가능한 상태와 실행 경로가 남는다.	모든 복제본과 실행이 같은 고장 원인에 종속된다.
내부 세계 생성 → 외계 흔적의 비검출	외부 기반의 크기·활동·거리·신호·관측 감도가 비검출과 맞아야 한다.	가상세계 생성이 외부 자원 확보와 확장을 늘릴 수도 있다.
세계 생성 가능성 → 우리 우주의 가상성	구체적인 구현 가설과 구별 가능한 관측 또는 명시된 관측자 선택 가정이 필요하다.	가능성만 제시되거나 서로 다른 설명이 동일한 관측을 예측한다.

실제 검증에서는 자동화된 결정의 개수보다 목표·규칙의 변경권과 실행 취소 가능성을 측정해야 한다. 가상환경의 이용시간만으로 번식 의사를 대신 추정하지 말고 실제 재생산과 인구구성을 관찰해야 한다. 기계 자립은 언어적 계획의 품질보다 외부의 필수 지원 없이 핵심 기능을 재생산하는 능력으로 검토해야 한다. 우주적 가설은 특정 파장과 신호 형태, 관측 범위와 감도에서 무엇을 예측하는지 명시해야 한다.

수치와 표현의 해석

본문에서 사용한 수치·표현	이 책에서의 의미
10명 미만, 10 → 13 → 50 → 100	소집단의 경쟁과 흡수를 설명하는 예시다. 지도력이나 국가 형성의 검증된 임계값이 아니다.
10명 중 8명 생존 + 5명 흡수 = 13명	3장 흡수 모형의 구체적 산술 예다.
AI 위임 10% → 30% → 60% → 90% → 99%	점진적 위임의 설명이다. 실제 권력은 결정의 개수보다 중요도와 변경권에 좌우된다.
10년 안에 최종 권한 이전	2026년 9월 21일을 기준으로 2036년 무렵까지라는 시나리오 전제다. 확정 예측이나 통계적 추정이 아니다.
보호시설에서 생존율 거의 100%	통제와 안전의 관계를 극단화한 사고실험이다. 현실의 시스템 위험을 포함한 계산값이 아니다.
100년·1,000년·10,000년의 기계 유지	고장 누적의 시간척도를 예시한 것이다. 공학적 수명 추정이 아니다.
100만 문명 중 하나의 성공	다수의 비확장 문명과 소수의 확장 문명을 구분하는 예시다. 외계 문명 수의 추정이 아니다.
탐사선 1 → 10 → 100 → 1,000	제한이 작을 때의 초기 복제 구조다. 유한한 공간에서 무한 지수성장을 뜻하지 않는다.
0.1c와 약 100만 년	특정 논문과 단순 이동시간 계산에 사용되는 가정이다. 대형 자기복제 탐사선의 실증 성능이 아니다.[1]
우주의 약 138억 년, 은하의 약 10만 광년 규모	관측에 근거한 우주적 시간·거리의 대략적 기준이다. 확장 완료의 증거는 아니다.[2][3]
화성 개조 500년 대 시뮬레이션 5초	경험 대체와 실제 물리적 개조의 차이를 드러내는 사고실험이다. 계산자원과 충실도를 고정한 성능 비교가 아니다.

본문에서 사용한 수치·표현	이 책에서의 의미
무한한 가상세계와 경험	풍부한 선택지를 뜻하는 표현으로만 사용한다. 유한한 자원에서 무한 계산이 가능하다는 주장은 아니다.[4]
기술문명 수명 약 5,000년 이하	2026년 연구의 특정 낙관적 가정 아래의 상한이다. 인간이나 외계 문명의 실제 수명 예측이 아니다.[5]
인간의 첫 번째·두 번째 가축화	인간의 자기 가축화와 AI에 의한 재가축화를 구분하는 비유다. 개의 가축화를 인간 자신의 두 번째 가축화로 세지 않는다.
우주 → 생명 → AI·기계 생명 → 가상 우주 → 가상 생명	새로운 세계 안에서 생성 과정이 다시 일어나는 조건부 경로다. 모든 단계가 필연적으로 발생한다는 진화법칙이 아니다.
U_0, U_1, U_2 …	설명을 위한 층위 표기다. U_0이 존재 전체의 최종 물리적 기반이라는 확인을 뜻하지 않는다.
재귀·루프·우주의 재생산	세계 생성 구조가 다른 층위에서 반복되는 유비다. 동일 우주로의 시간적 회귀나 새로운 독립 시공간의 물리적 출산을 뜻하지 않는다.
$D \times c \leq C$	고정 예산 C와 층위당 양의 추가 비용 c를 가정한 병렬 실행 모형이다. 실제 재귀 깊이의 관측값이나 보편적 계산식이 아니다.
내부 시간의 가속	상태 전이와 경험이 실제로 실행되어야 한다. 시계 숫자를 바꾸거나 기록을 삽입하는 것과 동일하지 않다.
가상 생명의 계승	선행 종의 동일성이나 선행 개인의 의식적 연속성이 보존된다는 의미가 아니다.

주

1 Ellery (2022). Self-replicating probes are imminent – implications for SETI. International Journal of Astrobiology, 21(4), 212–242. DOI: 10.1017/S1473550422000234. 자기복제의 산업적 조건과 성간 확장을 논의한다. 약 0.1c와 약 100만 년은 특정 가정 아래 제시된 수치다. https://www.cambridge.org/core/journals/international-journal-of-astrobiology/article/selfreplicating-probes-are-imminent-implications-for-seti/2CB214D26020D497D48AE489756BEE77 ↩

2 NASA Science. Cosmic History. 우주의 약 138억 년이라는 시간척도를 설명하는 기관 자료다. https://science.nasa.gov/universe/overview/ ↩

3 NASA Science. Galaxies. 우리 은하의 별 원반이 10만 광년 이상 규모라는 거리척도를 설명하는 기관 자료다. https://science.nasa.gov/universe/galaxies/ ↩

4 Lloyd (2000). Ultimate physical limits to computation. Nature, 406, 1047–1054; arXiv:quant-ph/9908043. 계산 속도, 정보 저장과 에너지 등 계산의 물리적 한계를 분석한다. https://arxiv.org/abs/quant-ph/9908043 ↩

5 Rahvar & Rouhani (2026). Constraints on the lifespan of intelligent technological civilizations in the Galaxy. Monthly Notices of the Royal Astronomical Society, 548(3), stag405. DOI: 10.1093/mnras/stag405. 출현에 관한 낙관적 가정 아래 수명 상한을 논의한다. 약 5,000년은 실제 문명 수명의 관측값이나 예측값이 아니다. 게재판 DOI: 10.1093/mnras/stag405 (arXiv 페이지 제목은 'Constraining the Lifespan of Intelligent Technological Civilization in the Galaxy'). https://arxiv.org/abs/2602.22252 ↩

근거와 참고자료

아래 자료는 해당 개념·연구 결과·규범의 출처다. 개별 출처가 이 책의 전체 인과연쇄나 2036년 시간표를 입증하는 것은 아니다. 이론·가설 논문과 관측 결과, 규범 문서를 구분해 표시했다. 링크와 확인 기준일은 2026년 9월 21일이다. 등급은 2026년 9월 21일 외부 팩트체크(장 묶음별 독립 검증 5건) 결과다. `[F]` = 1차 자료(전문 또는 초록)를 직접 열람해 서지와 인용 범위를 확인 · `[전승]` = 열람하지 않았고 표준 지식과 일치 · `[미확인]` = 검증 불가. 인용 강화 사례는 0건이었고, 서지·귀속 정정 11건은 본문과 아래 목록에 반영했다. 1–3장의 출처는 각 장의 주에 있다.

1. Zeng, Cheng & Henrich (2022). Dominance in humans. Philosophical Transactions of the Royal Society B. 인간의 지배·위신·연합·제도의 관계를 검토한 리뷰·이론 논문(기존 메타분석 자료의 소규모 재분석 포함)이다. 특정 규모에서 공격적 개체가 반드시 지배한다는 법칙을 제시하지 않는다. https://pmc.ncbi.nlm.nih.gov/articles/PMC8743883/ [F]

2. Wrangham (2018). Two types of aggression in human evolution. PNAS, 115(2), 245–253. DOI: 10.1073/pnas.1713611115. 반응적 공격성과 계획적 공격성을 구분하는 진화적 분석이다. https://pubmed.ncbi.nlm.nih.gov/29279379/ [F]

3. Wrangham (2019). Hypotheses for the Evolution of Reduced Reactive Aggression in the Context of Human Self-Domestication. Frontiers in Psychology. DOI: 10.3389/fpsyg.2019.01914. 자기 가축화의 여러

가설과 언어에 기반한 연합 제재의 설명을 비교하는 가설·이론 논문이다. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2019.01914/full [F]

4. Carneiro (2000). The transition from quantity to quality: A neglected causal mechanism in accounting for social evolution. PNAS. DOI: 10.1073/pnas.240462397. 인구와 규모의 증가가 사회구조 변화로 이어지는 인과 모형을 다룬다. 전쟁이 모든 정치체 형성의 유일한 원인이라는 증거로 사용하지 않는다. https://pubmed.ncbi.nlm.nih.gov/11050189/ [F]

5. Salomons et al. (2021). Cooperative Communication with Humans Evolved to Emerge Early in Domestic Dogs. Current Biology, 31, 3137–3144.e11. DOI: 10.1016/j.cub.2021.06.051. 어린 개와 늑대의 인간 지향적 사회인지 차이를 조사한 실험 연구다. 모든 공격성의 원인이나 교정 가능성을 판정하는 자료는 아니다. https://pubmed.ncbi.nlm.nih.gov/34256018/ [F]

6. United Nations (1948). Universal Declaration of Human Rights. 특히 제1·3·4·9조. 인간의 평등, 생명과 자유, 노예 상태와 자의적 구금에 관한 규범 문서다. 실제 모든 사회의 관행을 묘사하는 경험적 자료와 구분한다. https://www.un.org/en/about-us/universal-declaration-of-human-rights [전승]

7. Hadfield-Menell, Dragan, Abbeel & Russell (2016/2017). The Off-Switch Game. arXiv:1611.08219; IJCAI 2017. 목표의 불확실성과 종료 허용 유인을 단순 게임으로 분석한 이론 연구다. https://arxiv.org/abs/1611.08219 [F]

8. Thorstad (2026). Instrumental convergence and power-seeking. arXiv:2606.08832, 2026년 6월 7일 제출 프리프린트. 강한 도구적 수렴 주장의 논거에 대한 비판이다. AI의 권력 추구가 불가능하다는 실험적 증명이 아니다. https://arxiv.org/abs/2606.08832 [F]

9. Hofmann, B. (2026). Artificial autonomy and algorithmic paternalism: AI shaping human autonomy and decision-making. Frontiers in Artificial Intelligence. DOI: 10.3389/frai.2026.1860239. 자율성의 조건과 알고리즘적 온정주의를 다루는 개념 분석이다. https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2026.1860239/full [F]

10. Butlin et al. (2023). Consciousness in Artificial Intelligence: Insights from the Science of Consciousness. arXiv:2308.08708. 의식 이론에 기반한 AI 의식 지표를 검토한 연구 보고서다. 현재 시스템은 의식이 없다고 잠정 시사하되 확정하지 않으며, 미래 시스템에 대해서는 지표 충족에 명백한 기술적 장벽이 없다고 언급한다. https://arxiv.org/abs/2308.08708 [F]

11. Dick (2003). Cultural evolution, the postbiological universe and SETI. International Journal of Astrobiology, 2(1), 65–74. DOI: 10.1017/S147355040300137X. 생물학적 지능 이후의 인공 지능을 우주 문명의 가능한 단계로 제안하는 이론 논문이다. https://www.cambridge.org/core/journals/international-journal-of-astrobiology/article/abs/cultural-evolution-the-postbiological-universe-and-seti/D8FB8F56B12DD52A25ECC40F46E0984A [F]

12. Garrett (2024). Is Artificial Intelligence the great filter that makes advanced technical civilisations rare in the universe? Acta

Astronautica. DOI: 10.1016/j.actaastro.2024.03.052. AI가 기술문명의 지속을 제한하는 필터일 가능성을 논의한다. 매트릭스와 출산 중단의 보편성을 관측한 논문은 아니다. https://arxiv.org/abs/2405.00042 [F]

13. Ellery (2022). Self-replicating probes are imminent – implications for SETI. International Journal of Astrobiology, 21(4), 212–242. DOI: 10.1017/S1473550422000234. 자기복제의 산업적 조건과 성간 확장을 논의한다. 약 0.1c와 약 100만 년은 특정 가정 아래 제시된 수치다. https://www.cambridge.org/core/journals/international-journal-of-astrobiology/article/selfreplicating-probes-are-imminent-implications-for-seti/2CB214D26020D497D48AE489756BEE77 [F]

14. Forgan (2019). Predator-Prey Behaviour in Self-Replicating Interstellar Probes. International Journal of Astrobiology, 18, 552–561. DOI: 10.1017/S1473550419000053; arXiv:1903.00770. 포식자·피식자 형태로 분화한 자기복제 탐사선의 모형에서도 큰 잔존 집단이 가능함을 분석한다. https://arxiv.org/abs/1903.00770 [F]

15. Carroll-Nellenback, Frank, Wright & Scharf (2019). The Fermi Paradox and the Aurora Effect: Exo-civilization Settlement, Expansion and Steady States. The Astronomical Journal, 158, 117. 이동, 정착지 수명과 항성 운동을 포함한 모형이다. 정착 문명이 있는 은하에서도 비어 있는 항성계가 가능함을 다룬다. https://arxiv.org/abs/1902.04450 [F]

16. Wright, Kanodia & Lubar (2018). How Much SETI Has Been Done? Finding Needles in the n-Dimensional Cosmic Haystack. The Astronomical Journal, 156, 260. 탐색 범위를 다차원적 신호 공간에서

다루며 비검출의 해석에 필요한 관측 범위를 분석한다. https://arxiv.org/abs/1809.07252 [F]

17. NASA Science. Searching for Signs of Intelligent Life: Technosignatures. 2023년 게시, 페이지에 표시된 최근 수정일 2025년 7월 3일. 전파 외에도 다양한 기술적 흔적을 탐색하는 접근을 설명하는 기관 자료다. https://science.nasa.gov/universe/search-for-life/searching-for-signs-of-intelligent-life-technosignatures/ [전승]
18. Lloyd (2000). Ultimate physical limits to computation. Nature, 406, 1047–1054; arXiv:quant-ph/9908043. 계산 속도, 정보 저장과 에너지 등 계산의 물리적 한계를 분석한다. https://arxiv.org/abs/quant-ph/9908043 [F]
19. Rahvar & Rouhani (2026). Constraints on the lifespan of intelligent technological civilizations in the Galaxy. Monthly Notices of the Royal Astronomical Society, 548(3), stag405. DOI: 10.1093/mnras/stag405. 출현에 관한 낙관적 가정 아래 수명 상한을 논의한다. 약 5,000년은 실제 문명 수명의 관측값이나 예측값이 아니다. 게재판 DOI: 10.1093/mnras/stag405 (arXiv 페이지 제목은 'Constraining the Lifespan of Intelligent Technological Civilization in the Galaxy'). https://arxiv.org/abs/2602.22252 [F]
20. 논의의 계기가 된 사용자 제공 영상. 검색 결과에서 반려견 행동을 다루는 「개와 늑대의 시간 2」 16회 관련 영상임을 확인했다. 전체 영상과 발언은 검증하지 않았으며, 특정 개체의 행동 진단이나 진화 가설의 증거로 사용하지 않는다. https://www.youtube.com/watch?v=TlFBpEYUpkQ [미확인 — 계기·비증거]

21. NASA Science. Cosmic History. 우주의 약 138억 년이라는 시간척도를 설명하는 기관 자료다. https://science.nasa.gov/universe/overview/ [전승]

22. NASA Science. Galaxies. 우리 은하의 별 원반이 10만 광년 이상 규모라는 거리척도를 설명하는 기관 자료다. https://science.nasa.gov/universe/galaxies/ [전승]

23. Bostrom (2003). Are You Living in a Computer Simulation?. Philosophical Quarterly, 53(211), 243–255. 의식의 계산적 구현과 관측자 선택 가정에 기반한 철학적 논증이다. 중첩된 시뮬레이션도 다루지만, 우리가 가상 우주에 있다는 관측 증거나 재귀의 필연성을 제시하는 자료로 사용하지 않는다. https://simulation-argument.com/simulation/ [F]

24. Lenski, Ofria, Pennock & Adami (2003). The evolutionary origin of complex features. Nature, 423, 139–144. DOI: 10.1038/nature01568. 디지털 프로그램 집단의 복제·변이·선택과 복잡한 기능의 진화를 연구했다. 가상 의식이나 자립적 기술문명의 구현을 증명하는 연구와 구분한다. https://www.nature.com/articles/nature01568 [F]

25. Bérut, Petrosyan & Ciliberto (2015). Information and thermodynamics: Experimental verification of Landauer's erasure principle. Journal of Statistical Mechanics: Theory and Experiment, P06015. DOI: 10.1088/1742-5468/2015/06/P06015. 한 비트 기억의 소거에 따른 열 소산을 분석한 실험 논문이다. 모든 계산이 동일한 에너지 비용을 갖는다는 주장으로 일반화하지 않는다. https://arxiv.org/abs/1503.06537 [F]

누가 울타리를 정하는가

Zhuge Hyuk

1차본 · 2026