

제3프로토콜

Daehyung Lee

May 26, 2026

Abstract

본 원고는 현대 사회의 핵심 병목이 더 이상 실행 능력 자체가 아니라, 사회가 인간을 어떤 입력 형식으로 읽고 집계할 것인가의 문제로 이동하고 있다는 진단에서 출발한다. 자본주의는 생존과 결핍의 시대에는 인간의 고통을 줄이는 강력한 조정 장치였으나, 생존 이후의 포화 구간에서는 더 많은 생산, 더 빠른 연결, 더 강한 자극이 더 나은 삶을 보장하지 못하는 상태에 도달했다. 기술적 수단은 지수적으로 축적되지만 목적, 정의, 사랑, 책임과 같은 삶의 상위 좌표는 같은 방식으로 축적되지 않는다. 여기에 분과적 전문화는 전체 맥락을 보는 능력을 약화시키고, 민주주의와 시장은 각각 표와 가격이라는 낮은 해상도의 신호로 인간의 의사와 선호를 압축한다.

AI와 자동화는 이 불균형을 해결하기보다 먼저 폭로한다. 번역, 검색, 설계, 협상, 생산, 조율의 비용이 낮아질수록 실행 가능한 선택지는 폭발하지만, 무엇을 실행해야 하는가를 정하는 사회적 입력 형식은 여전히 빈약하기 때문이다. 이 문제를 다루기 위해 본 원고는 먼저 목적을 선언하는 주체가 무엇인지 묻는다. 우리는 자아를 과거로부터 현재까지 보존되는 단단한 실체로 상상하지만, 현재는 시간적 부피를 갖지 않으며 우리가 “나”라고 부르는 것의 대부분은 이미 지나간 흔적들의 서사에 가깝다.

또한 자유의지를 열린 미래로 전제하는 한, 서로 다른 과거의 자아 상태들은 현재의 증거 상태에서 구분 불가능하게 병합될 수 있고, 많은 내면 상태들은 완전한 흔적도, 재도달 가능성도, 현재 세계에 대한 구분 가능한 차이도 남기지 않는다. 따라서 자아는 과거에서 발굴되는 실체가 아니라, 불완전한 증거 위에서 미래를 향해 스스로를 승인하고 기립시키는 비존재적 구조로 이해되어야 한다. 그러나 자아가 인정된다는 것은 단순히 “목적을 말할 수 있다”는 뜻만은 아니다. 자아가 인정될 때 비로소 세계는 중립적 사건들의 더미가 아니라 가치가 발생하는 장이 된다.

자아는 가치사상(value mapping)을 통해 세계를 해석하고, 그 결과 세계는 가치장(value field)으로 나타난다. 이 가치장이 독서, 사유, 대화, 실패, 후회, 반박을 거쳐 가치정련(value refinement)을 통과할 때, 하나의 실천적 기술기, 곧 의지구배(volitional gradient)가 산출된다. 본 원고가 제3프로토콜의 입력으로 삼는 것은 인간의 내면 전체가 아니라, 바로 이 의지구배가 사회적으로 승인된 형식인 구배선언(gradient declaration)이다. 따라서 사회 역시 인간을 소비자, 유권자, 사용자로만 읽는 데서 멈출 수 없다.

경제는 인간을 선호의 보유자로, 정치는 의견의 보유자로, 플랫폼은 반응의 발생원으로 읽어왔지만, 인간은 그보다 앞서 세계를 가치 있는 것으로 해석하고 자신의 시간, 능력, 책임을 어디에 결속할 것인지 승인할 수 있는 존재다. 본 원고는 이 전환을 바탕으로 표와 가격을 넘어서는 제3프로토콜을 제안한다. 제1프로토콜이 표를 통해 권한을 집계하고, 제2프로토콜이 가격을 통해 희소가치를 집계했다면, 제3프로토콜은 구배선언, 조건부 결속, 기여 가능성, 책임 범위를 사회적 입력으로 받아들이고, 그것을 조직, 제도, 생산, 계약, 의도적 사건으로 되돌려 보내는 사회적 입출력 규약이다. 이는 정치와 경제의 폐지가 아니라, 두 체계가 손실 압축해 온 인간의 목적 구조를 더 높은 해상도로 사회에 입력하는 보완 회로다.

그러나 구매선언을 사회적 입력으로 받는다고 해서 그 선언들이 자동으로 정합적 질서를 이루는 것은 아니다. 의지구배는 충돌하고, 오해되며, 때로는 자기 자신에게도 불명료하다. 따라서 제3프로토콜은 여러 층위의 내부 장치를 필요로 한다. 일반공리주의(Generalized Utilitarianism, GU)는 다수의 구매선언이 충돌할 때 무엇을 더 정의로운 조정으로 볼 것인가를 다루는 규범 엔진이다.

GU는 효용을 단순한 쾌락과 고통의 합이 아니라 공감, 지식, 관계 구조, 자기이해를 포함하는 고차원 함수로 확장하고, 정의를 완성된 상태가 아니라 전지적 공감으로 수렴하는 방향성으로 이해한다. 본 원고에서 자아와 타자는 미리 결정된 실체가 아니라 비존재적 좌표로서의 주체로 다뤄지며, 따라서 GU의 1차 입력은 타자의 효용 H_i 자체가 아니라 자아가 π 아래 타자를 가치장에 귀속시킨 값 $H_{\pi \rightarrow i}$ 다. H_i 는 그 주체가 독립적 가치장을 갖는 것으로 인정될 때의 분해항이다. 계약적 가치이전역학(Contractual Value Transfer Dynamics, CVTD)은 개인의 의지구배와 현실의 의도적 사건 사이를 왕복하는 실행 문법이다. 역방향으로는 행위, 선택, 반복, 갈등, 후회, 커밋, 위반 패턴으로부터 의지구배를 추정하고 정련하며, 순방향으로는 구매선언을 조건, 계약, 사건, 기여, 책임 구조로 번역한다. 나아가 삼기저동력모형(Tri-Basis Drive Model, TBDM)은 개인의 의지구배와 사회적 정의근사 방향량을 가능성, 사회성, 정합성의 세 축 위에 사영하여, 무수한 미시적 의도 사건들이 어떤 거시적 알짜힘으로 평균화되고 상쇄되고 증폭되는지를 설명하는 공통 언어다.

마지막으로 본 원고는 제3프로토콜의 첫 현실 진입점이 곧바로 목적시장이어서는 안 된다고 주장한다. 의지구배는 로그, 클릭, 소비 기록으로부터 자동 추출되는 정보가 아니라, 읽고, 생각하고, 반박받고, 수정하고, 대화하고, 승인하는 느린 과정을 통해 정련되는 사회적 객체다. 따라서 앞으로의 핵심 인프라는 더 정교한 추천 엔진이 아니라 가치정련 인프라이다. 책과 장문 텍스트는 이 과정에서 가장 느리고 단단한 씨앗이자, 구매선언으로 이어질 수 있는 장문형 정보 레이어가 된다.

이러한 맥락에서 The Channel은 완성된 제3프로토콜이 아니라, 인간의 가치정련 과정을 도시 위에 배치하고 그 과정에서 축적된 기록이 개인을 대리할 AI의 감각신경계가 되게 만드는 첫 번째 현실 인터페이스로 위치 지어진다. 이 원고의 전체 기획은 결국 하나의 질문으로 수렴한다. 수단이 목적을 압도한 시대에, 인간은 다시 어떤 형식으로 자신의 의지구배를 승인하고, 사회는 그 선언을 어떻게 읽어야 하는가.

1 서론

“그래서 어떤 문제를 해결하는가?” 스타트업에 향한 이 관문 질문의 속내는 명확하다. 실상 “어떤 문제를 해결하여 우리에게 돈을 벌어드 줄 것인가?” 라는 자본의 요구와 맞닿아 있다. 이는 문제 해결의 가치가 자본주의의 논리와 정렬될 때만 유효하다는 뜻이다. 과거에는 이 정렬이 자연스러웠을지 모르나, 지금은 다르다.

나는 더 이상 자본주의가 “보이지 않는 손”처럼 인류를 구원할 정의로운 조정자가 아니라고 생각한다. 자본주의는 이제 인간의 고통을 섬세하게 들여다보기엔 너무나 거대하고 비대해졌다.

인류의 기본값이 기아, 추위, 질병 같은 “불행”에 가까웠던 시절, 자본주의의 목표는 불행의 축소, 즉 생존이었다. 그때 자본주의는 효율적이고 윤리적인 시스템처럼 보였다. 하지만 기술과 문명의 발달로(적어도 우리의 서구문명에선) 인류의 기본값이 생존을 넘어 행복의 단계로 올라선 지금, 자본주의의 동력은 변질되었다.

인간의 효용 곡선은 전형적인 로그함수의 형태를 띤다. 초기 단계, 즉 생존이 위협받던 시절에는 투입되는 자원이 조금만 늘어도 효용(행복)의 기울기는 가파르게 상승했다. 땀 한 조각이 목숨을 구하고, 지붕 하나가 삶을 바꿨다. 하지만 우리는 이미 그 가파른 구간을 지났다.¹

지금 우리는 투입 대비 효용의 증가가 거의 평행선에 가깝게 누워버린, 로그 함수의 오른쪽 꼬트머리에 서 있다. 이제는 어지간한 자극으로는 효용이 증가하지 않는다. 과거에는 생존을 위해 싸웠다면, 지금 자본을 굴리는 것은 더 큰 자극과 말초적인 쾌락을 향한 끝없는 갈증이다. 우리는 왜 더 빠르고 강하고 많아야 하는지를 잊었다. 그저 빠르고 강하고 많은 것이 나은 것이라는 경험적 외삽법이 일종의 거대한 종교적인 착시를 만들었을 뿐이다.

우리를 이끄는 이 거대하고 강력한 자본주의라는 시스템이, 고작 이 로그 함수의 꼬트머리에 있는 미미한 증가를 위해 존재하는가? 수십 조의 자본과 최첨단 기술이 투입되어 만들어내는 결과물이 고작 인류가 느끼는 쾌락의 1% 남짓한 상승이라니, 이 얼마나 비효율적인 폭주인가.

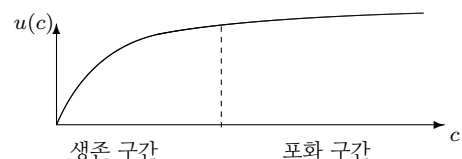
나는 이 문제가 “삶의 수단”과 “삶의 목적” 사이의 구조적인 속도 차이에서 기인된다고 진단한다. 영화 죽은 시인의 사회에서 키팅 선생은 말했다. “의술, 법률, 기술.. 이것들은 삶을 유지하는 데 필요한 고귀한 수단이다. 하지만 시, 아름다움, 낭만, 사랑.. 이것들이야말로 우리가 살아가는 목적이다.” 우리의 비극은 수단을 극대화하는 기술에는 능숙하지만, 목적을 다루는 지혜는 잃어버렸다는 데 있다. 기술은 버전업 되고, 코드는 깃허브에 쌓이며, 지식은 복제되어 전승된다. 프론트엔드 개발자는 회로가 어떻게 작동하는지 이해하지 않아도 되며, 컴퓨터 공학자는 양자역학에 대해 이해하지 않아도 된다. 우리는 아이폰의 진화를 보며 기술의 축적을 체감한다. 하지만 정의, 윤리, 사랑 따위의 것들은 그렇게 축적되지 않는다. 다시 말해, 우리에게 정의를 위한 깃허브는 없다. 기술적 수단은 지수적으로 팽창하는데 윤리적 목적은 제자리를 맴도는 이 구조적인 “속도의 격차”가, 자본주의의 폭주를 제어할 브레이크를 고장 냈다.

설상가상으로, 그러나 필연적으로, 기술의 진보는 이 쾌락을 즉각적으로 충족시키는 방향으로 이루어졌고, 폭발적으로 늘어난 지식의 총량은 개개인을 “전문화”라는 이름의 썰기로 가두었다.

역설적이게도 전체 맥락을 돌봐줄 눈은 사라졌다. 말하자면, 근대 이후 인류가 축적해낸 지식의 양적 폭증은 분과의 극단적 전문화, 즉 지식의 썰기화를 낳았다. 이 과정에서 서로 다른 체계 간 상호참조는 차단되고, 지식 생산 주체는 전체 맥락을 읽지 못한 채 부분 과업만 수행하는 기능 단위로 환원된다. 이는 마르크스가 말한 인간의 수단화가 “분과적 수행자”라는 형태로 재현된 결과이며, 동시에 니체가 비판한 “학문적 노동자”와 구조적으로 동일한 형상이다. 니체에게 학문적 노동자는 거대한 지식 공장에서 끝없이 자료를 채굴 · 주석 · 가공하지만, 산과 산 사이를 건너며 가치를 재구성하는 역할에는 도달하지 못하는 존재였다. 그는 자신을 “산과 산 사이를 넘어가는 자”로 규정했지만, 현대의 지식 체계는 이러한 산 넘는 자를 예외적 개인의 덕성에 맡긴 채, 다수를 분과 내부에 고착된 노동자로 양산하는 방향으로 굳어져 있다. 인문 · 공학 · 정치 · 경제의 판단 구조는 서로를 해석하지 못하는 상태로 분리되고, 민주주의 사회에서 이 구조는 점점 더 치명적인 폐단으로 작용한다. 유권자들은 물론 선출된 사회적 결정권자들 또한 썰기형 분과의 압력 속에서

¹수리적 주석: 이 직관은 CRRA(Constant Relative Risk Aversion) 또는 isoelastic utility 함수와 연결된다.

$$u(c) = \begin{cases} \frac{c^{1-\gamma} - 1}{1-\gamma}, & \gamma \neq 1, \\ \log c, & \gamma = 1. \end{cases}$$



여기서 c 는 소비 또는 자원 수준이고, γ 는 한계효용 체감의 정도를 나타낸다. $\gamma = 1$ 일 때 로그 효용이 된다. 이 원고에서 중요한 것은 특정 함수형의 엄밀한 채택이 아니라, 생존 구간에서는 추가 자원의 윤리적 효과가 크지만 포화 구간에서는 그 정당성이 약해진다는 구조적 직관이다.

전체 맥락을 잃게 된다.

결국 우리는 “전체”를 잃어버릴 것이다. 자본주의의 독주를 견제하고 거시적 맥락을 잡아야 할 정부가, 거대해진 자본 앞에서 무력해질 것은 자명하다. 아니, 차라리 무력해지기를 바라야 할지도 모르겠다.

민주주의는 집단지성이란 명분 아래 권력의 입장권을 국민들에게 쪼개 나눠주었다. 본래 개인이 쥔 그 권력이란 것은 신기루처럼 희미한 것이었으나 더욱 절망적인 사실은, 그 입장권을 쥔 국민들이 전문화의 썩기에 갇혀 전체 맥락을 볼 수 없게 된 것이다.

맥락을 상실한 군중이 쥔 앙상한 참정권, 무엇을 집계해야 할지 관찰하기엔 너무 빠르고 강력해진 눈먼 자본주의. 이 상황에서, 우리는 도대체 어디에서 “옳은 판단”을 기대해야 하는가.

문명이 자연히 우리를 옳은 길로 이끌 것이라는 환상, 우주가 우리에게 정의로운 개연성을 빚고 있다는 환상에 기대야만 할까. 결국 우리는 다시 근본적인 질문 앞에 서야 할 것이다. 우리는 무엇을 위해 살아가는가. 우리는 누구인가.

우리는 되돌릴 수 없는 것을 계속 버리며 나아가고 있다. 우리는 앞으로 영원히 두 번 다시 촌스러운 음악에 촌스러운 춤을 멋들어서 추지 못할 것이다. 알고리즘은 우리에게 더 정교한 취향을 주었지만, 동시에 서툰 것을 사랑할 수 있는 무지를 빼앗아갔다. 5시간 동안 침대에 누워 릴스를 넘기기 위해 우리는 무엇을 포기하고 왔는가. 적어도 우리의 앞에서 기다리는 것들은 우리가 뒤에 버리고 온 것들보다는 멋져야 하지 않는가. 기술은 불가능이 없다는 감각을 주지만 사실은 동시에 특정한 가능성의 공간을 영원히 폐기하면서 나아가고 있다.

돌아갈 수는 없다. 돌아가는 것이 답도 아니다. 문명 이전의 순진함도, 자본 이전의 공동체도, 민주주의 이전의 질서도 이미 우리가 다시 택할 수 있는 해답이 아니다. 문제는 앞으로 나아간다는 사실이 아니다. 문제는 우리가 무엇을 버리고 있는지 모른 채, 그리고 무엇을 위해 버리는지 묻지 않은 채, 너무 빠르게 나아가고 있다는 데 있다.

수단이 약한 시대에는 목적의 빈곤이 어느 정도 가려질 수 있었다. 할 수 있는 일이 많지 않았기 때문이다. 하지만 수단이 강해질수록, 목적의 빈곤은 더 이상 철학적 결핍이 아니라 실제적인 위협이 된다. 우리는 더 많이 만들 수 있고, 더 빨리 연결할 수 있고, 더 정교하게 예측할 수 있게 되었지만, 왜 그래야 하는지에 대해서는 점점 더 빈약한 언어만을 갖게 되었다. 수단은 축적되었지만 목적은 축적되지 않았다. 실행은 정교해졌지만 판단은 정교해지지 않았다.

이 상태에서 수단은 멈추지 않는다. 수단은 우리가 잃어버린 질문을 기다려주지 않은 채 계속 앞으로 나아간다. 가능성에 의존해 행해진 우리의 발명들은 우리를 이상으로 이끈다는 보장을 준 적이 없다. 연속적인 추구는 우리를 추구의 종말로 이끈다는 보장 또한 없다. 어떤 가능성이 열릴 때마다 다른 가능성은 닫힌다. 더 정교한 취향은 서툰 것을 사랑할 수 있는 무지를 빼앗고, 더 빠른 연결은 고독 속에서 천천히 형성되던 감각을 밀어낸다. 우리는 매번 무엇인가를 얻는 동시에, 다시는 돌아갈 수 없는 어떤 세계를 폐기한다. 아무래도 삶의 관성은 더 이상 신뢰를 잃은 것처럼 보인다.

그렇다면 문제는 단순히 기술이 너무 빠르다는 것이 아니다. 더 정확히는, 기술이 빠른 속도로 세계의 가능성 공간을 바꾸는 동안, 우리는 그 변화가 무엇을 위해 일어나야 하는지 판단하는 능력을 충분히 기르지 못했다는 데 있다. 문명의 엔진은 점점 더 강력해졌지만, 그 엔진이 향해야 할 방향을 정하는 장치는 여전히 낡고 거칠다. 시장은 가격을 읽고, 정치는 표를 읽지만, 인간이 어떤 세계를 원하며 무엇을 포기할 수 없고 무엇에 책임지려 하는지는 그 신호들만으로 충분히 드러나지 않는다. 태평양을 항해하는 선박은 거칠게 방향을 잡아도 큰 문제가 생기지 않는다. 그러나 대기권을 빠져 나가는 로켓은 발사 초기에 아주 작은 각도 오차만 생겨도 돌이킬 수 없는 궤도 차이로 증폭된다. 속도가 커지고 에너지가 커질수록, 방향을 정하는 장치는 더 정밀해야 한다.

1.1 문명의 입력 형식과 누락되는 가치

문명의 발전은 스스로를 본질에 가까워지는 운동처럼 드러난다. 말은 자동차가 되고, 자동차는 비행기가 되고, 비행기는 로켓이 된다. 종이화폐는 은행 원장이 되고, 전자화폐가 되고, 암호학적 합의와 분산 원장의 문제로 넘어간다. 손으로 하던 계산은 계산기가 되고, 컴퓨터가 되고, 거대한 연산 인프라가 된다. 겉으로는 수단이 복잡해지는 것처럼 보이지만, 그 복잡성의 끝에서 오히려 더 낮은 층위의 변수가 드러난다. 이동은 말이나 바퀴의 문제가 아니라 질량, 위치, 시간, 에너지, 추진, 통신, 오류, 제약 조건의 문제가 된다. 화폐는 종이나 금속의 문제가 아니라 원장, 신뢰, 합의, 검증, 희소성, 정산 가능성의 문제가 된다. 계산은 머리의 재능이나 손의 속도가 아니라 정보 상태를 물리적 매체 위에서 얼마의 시간과 에너지와 오류율로 변환할 수 있는가의 문제가 된다.

이 현상은 단순한 정보화도, 단순한 디지털화도 아니다. 더 정확히 말하면, 문명의 수단들은 발전할수록 자신이 우연히 취하고 있던 역사적 구현을 벗고, 그 기능을 가능하게 하는 더 근본적인 조건으로 환원된다. 말은 이동이라는 기능의 생물학적 구현이었고, 종이화폐는 신뢰와 정산이라는 기능의 물질적 구현이었으며, 교실은 인간의 가치장과 인지 구조를 변화시키는 과정의 제도적 구현이었다. 수단이 충분히 발전하면, 문명은 더 이상 “이 물건을 어떻게 개선할 것인가”만을 묻지 않는다. 그것은 점차 “이 기능을 성립시키는 최소 조건은 무엇인가”, “무엇이 보존되어야 하고 무엇이 버려도 되는가”, “병목은 물질인가, 에너지인가, 정보인가, 신뢰인가, 시간인가”를 묻게 된다.

이 점에서 공학과 과학은 다시 만난다. 초기에는 과학이 세계의 원리를 발견하고 공학이 그 원리를 응용하는 별도의 가지처럼 보인다. 그러나 수단이 극한으로 밀려갈수록 공학은 다시 과학의 경계로 돌아간다. 로켓은 단순한 기계가 아니라 중력, 질량비, 열역학, 유체역학, 제어, 통신 지연의 문제다. 반도체는 상품이 아니라 양자역학, 재료과학, 광학, 열, 원자 단위 제조의 문제다. 생명공학은 치료술이 아니라 유전정보, 단백질, 세포 신호, 면역, 진화의 문제다. 끝까지 밀린 공학은 결국 “무엇이 가능한가”, “어떤 변환이 물리적으로 허용되는가”, “그 변환의 최소 비용은 얼마인가”를 묻는다. 과학은 가능성과 불가능성의 지도를 그리고, 공학은 그 지도 위에서 원하는 상태 변환을 실제로 수행한다. 두 영역은 분리된 것이 아니라, 가능성의 경계를 사이에 두고 서로 다른 방향에서 접근하고 있었던 셈이다.

이러한 수렴은 누군가의 의식적 합의로 이루어진 것이 아니다. 인류가 본질을 향해 나아가자고 결정했기 때문에 말이 로켓이 되고 종이가 암호학적 원장으로 변한 것이 아니다. 이 운동은 문명 내부의 선택압에서 발생한다. 더 빠른 수단, 더 싼 수단, 더 안정적인 수단, 더 복제 가능한 수단, 더 멀리 확장 가능한 수단, 더 적은 노동과 물질로 작동하는 수단은 경쟁 속에서 살아남기 쉽다. 이 압력이 반복되면 수단에 붙어 있던 우연한 껍질은 제거되고, 그 밑에 있던 기능적 핵심이 드러난다. 자동차가 말을 이긴 것은 자동차가 말이 주던 모든 가치를 이겼기 때문이 아니다. 자동차는 문명이 “이동”이라고 읽은 좌표계 안에서 말을 이겼다. 이동 시간이 줄고, 운송량이 늘고, 지형 의존성이 낮아지고, 예측 가능성이 높아졌기 때문이다. 그러나 말과 함께 존재하던 동물과의 관계, 느린 이동의 리듬, 길과 마을의 구조, 풍경을 읽는 몸의 감각은 같은 좌표계 안에 충분히 입력되지 않았다.

따라서 문명이 본질에 가까워지는 것처럼 보인다는 말은 수정되어야 한다. 문명은 전체 본질에 가까워지는 것이 아니라, 자신이 읽을 수 있는 본질에 가까워진다. 문명은 측정 가능하고, 비교 가능하고, 반복 가능하고, 저장 가능하고, 전송 가능하고, 표준화 가능하고, 거래 가능하고, 통제 가능한 것을 잘 읽는다. 반대로 지연되어 나타나고, 인과관계가 흐리고, 개별적으로는 약하지만 생태계 전체로는 크고, 사람마다 다른 형태를 가지며, 비효율처럼 보이는 조건 속에서만 발생하는 가치는 잘 읽지 못한다. 이것은 그 가치들이 비현실적이거나 낭만적이기 때문이 아니다. 그것들은 문명의 입력

형식 안에서 강한 신호로 변환되기 어렵기 때문이다.

첫째, 어떤 가치는 늦게 나타난다. 좋은 선생의 영향, 오래된 동네의 안정감, 느린 이동이 만든 관찰 능력, 우연한 만남이 바꾼 삶의 방향은 즉시 측정되지 않는다. 반면 클릭, 구매, 속도, 생산성, 비용 절감은 즉시 측정된다. 문명은 빠른 피드백을 선호하고, 빠른 피드백을 제공하는 값은 선택압 속에서 증폭된다. 지연된 가치는 실제로는 클 수 있으나 현재의 입력 체계에서는 약한 신호가 된다.

둘째, 어떤 가치는 인과관계가 흐리다. 자동차가 이동 시간을 줄였다는 사실은 증명하기 쉽다. 그러나 말과 함께 사라진 생활 리듬, 동물과의 상호성, 느린 거리감, 길 위에서 생기던 우연한 관계망의 손실은 증명하기 어렵다. 증명 가능한 이익은 정책과 시장과 기술의 언어로 쉽게 올라오지만, 증명하기 어려운 손실은 취향, 향수, 예외, 부수효과로 밀려난다.

셋째, 어떤 가치는 개별적으로는 약하지만 생태계 전체로는 크다. 벤치 하나, 골목 하나, 동네 가게 하나, 학교의 쉬는 시간 하나, 목적 없는 산책 하나는 각각만 보면 대체 가능하거나 비효율적인 요소처럼 보인다. 그러나 그것들이 함께 있을 때 사람은 소비자나 사용자만이 아니라 주민, 이웃, 친구, 학생, 산책자, 우연한 목적자로 존재한다. 문명은 개별 요소의 효율은 잘 읽지만, 약한 요소들이 모여 만드는 생태계적 총합은 잘 읽지 못한다.

넷째, 어떤 가치는 사람마다 형태가 다르다. 빠른 것은 대체로 많은 사람에게 좋고, 느린 것도 대체로 많은 사람에게 좋으며, 편리한 것도 대체로 많은 사람에게 좋다. 이런 값들은 집계되기 쉽다. 그러나 느낌, 관계성, 장소성, 우연성, 몸의 리듬은 사람마다 다르게 나타난다. 어떤 사람에게 말은 낯은 이동수단이지만, 어떤 사람에게 말은 동물과의 관계이며 삶의 형식이다. 이런 이질적 가치는 거시적 집계 과정에서 서로 상쇄되고, 결국 평균적으로 강한 알짜힘을 만들지 못한다.

다섯째, 어떤 가치는 비효율처럼 보이는 조건 속에서만 발생한다. 우연한 만남은 빈 시간과 느슨한 공간과 목적 없는 체류가 있어야 생긴다. 깊은 대화는 낭비처럼 보이는 반복 속에서 생기고, 창의성은 판짓과 실패 속에서 생기며, 우정은 효율적이지 않은 시간 속에서 자란다. 그러나 문명은 빈 시간을 낭비로, 빈 공간을 저활용으로, 목적 없는 체류를 낮은 생산성으로 읽는다. 따라서 어떤 가치는 사라진 뒤에야 그것이 비용이 아니라 생성 조건이었다는 점이 드러난다.

이 구조를 가장 간단히 말하면, 문명은 인간이 명시적으로 선택한 효용은 잘 읽지만, 그 선택을 가능하게 하던 배경 조건의 효용은 잘 읽지 못한다. 사람은 자동차를 선택할 때 “빠른 이동”을 입력한다. 그러나 그 선택과 함께 “느린 이동이 만들던 관계, 공간감, 우연성, 몸의 감각, 동물과의 생활세계”를 포기하겠다고 명시적으로 입력하지 않는다. 스마트폰을 선택할 때도 사람은 즉시 연결을 입력하지만, 연결되지 않던 시간에 생기던 기다림, 내면, 느린 관계, 고독의 형성력을 함께 입력하지 않는다. 온라인 쇼핑을 선택할 때 사람은 편리한 구매를 입력하지만, 동네 가게가 만들던 약한 사회적 연결을 포기한다고 입력하지 않는다.

이것은 인간 자신도 자신의 삶에 의미를 부여하는 장치들을 사전에 충분히 식별하지 못하기 때문이다. 인간은 욕망의 전면은 비교적 잘 안다. 빨리 가고 싶고, 싸게 사고 싶고, 쉽게 만나고 싶고, 즉시 알고 싶고, 편하게 살고 싶다. 그러나 그 욕망을 둘러싼 배경 생태계가 무엇을 해주고 있었는지는 잘 모른다. 그것은 선택의 대상이 아니라 환경이었기 때문이다. 공기는 선택하지 않는다. 없어지고 나서야 그것이 조건이었다는 사실을 알게 된다.

여기서 하이데거의 기술론은 중요한 참조점을 제공한다. 하이데거에게 현대 기술은 단순한 도구들의 집합이 아니라, 세계가 인간에게 드러나는 방식 자체를 바꾼다. 현대 기술은 사물과 자연과 인간을 독립된 존재라기보다 언제나 호출되고 배치되고 이용될 수 있는 비축 자원으로 드러나게 한다. 베버의 합리화 역시 같은 방향을 가리킨다. 근대 문명은 세계를 점점 더 계산 가능하고, 예측 가능하고, 규율 가능하고, 관리 가능한 질서로 전환한다. 엘월의 technique은 이 운동을 더욱 넓은

범위에서 포착한다. 그가 말한 technique은 기계나 전자기술만이 아니라, 인간 활동 전 영역에서 가장 효율적인 수단을 추구하는 체계다. 이 체계는 특정한 사람의 악의나 계획이 아니라, 수단 자체의 효율화 논리가 사회 전체를 관통하면서 발생한다.

Scott의 legibility 개념도 이 문제를 다른 방식으로 설명한다. 대규모 국가는 사회를 다루기 위해 복잡한 현실을 읽기 쉬운 형태로 단순화한다. 숲은 목재 생산량으로, 마을은 행정구역으로, 인간은 납세자와 병역자와 유권자로, 토지는 소유권과 지번으로 변환된다. 이 과정은 통치를 가능하게 하지만, 동시에 현장의 암묵적 질서와 지역적 지식과 비공식적 관계망을 지운다. Polanyi의 tacit knowledge 또한 같은 결론에 닿는다. 인간은 말할 수 있는 것보다 더 많이 알고 있다. 장인의 손, 의사의 촉, 농부의 날씨 감각, 부모가 아이를 읽는 방식은 명시적 규칙으로 완전히 환원되지 않는다. 그러나 문명은 지식을 명시화하고, 문서화하고, 코드화하고, 표준화할수록 강해지기 때문에, 바로 그 과정에서 말해지지 않는 앎을 부차적인 것으로 밀어내기 쉽다. Goodhart의 법칙은 이 누락이 왜곡으로 바뀌는 지점을 보여준다. 어떤 측정값이 목표가 되면, 그 측정값은 더 이상 좋은 측정값으로 남지 않는다. 배움을 점수로 읽으면 점수는 올라갈 수 있지만 배움은 빈약해질 수 있고, 건강을 지표로 읽으면 지표는 개선될 수 있지만 삶의 활력은 사라질 수 있다.²

따라서 문명의 병목은 단순히 더 많은 정보를 수집하지 못한다는 데 있지 않다. 문제는 어떤 정보가 사회적 입력으로 승인되고, 어떤 정보가 잡음으로 처리되는가에 있다. 정치는 표를 읽고, 시장은 가격을 읽고, 플랫폼은 반응을 읽는다. 그러나 표와 가격과 반응은 인간의 의지구배를 낮은 해상도로 압축한다. 그것들은 인간이 지금 무엇을 선택했는지, 무엇에 지불했는지, 무엇에 반응했는지는 보여주지만, 인간이 어떤 세계를 감당할 수 있는지, 무엇을 포기할 수 없는지, 어떤 조건에서 자신의 시간과 능력과 책임을 결속할 수 있는지는 충분히 보여주지 못한다.

그러므로 문명이 누락하는 가치는 단순히 계산 불가능한 가치가 아니다. 그것은 선택압으로 번역되기 어려운 가치다. 지연되고, 흐리고, 생태계적이고, 이질적이며, 비효율 속에서 발생하는 가치들은 실제로 인간 삶의 중심부에 있을 수 있으나, 사회적 입력으로 들어가는 순간 약한 신호가 된다. 반대로 속도, 가격, 편리함, 즉시성, 예측 가능성은 강한 신호가 된다. 그 결과 문명은 인간을 더 잘 읽는 것처럼 보이지만, 실제로는 인간에게서 강한 신호로 변환되는 부분만 더 잘 읽게 된다.

이 지점에서 제3프로토콜의 문제의식이 다시 선명해진다. 사회가 인간을 소비자, 유권자, 사용자, 반응 발생원으로만 읽는 한, 문명은 인간의 목적 구조를 계속 손실 압축할 수밖에 없다. 인간은 단순히 무엇을 사고 싶은 존재도, 누구에게 표를 줄 존재도, 어떤 콘텐츠에 반응할 존재도 아니다. 인간은 세계를 가치장으로 해석하고, 그 가치장을 읽기와 사유와 반박과 실패와 후회를 통해 정련하며, 마침내 자신의 시간과 능력과 책임을 어디에 결속할 것인지 승인할 수 있는 존재다. 문명이 읽어야 할 것은 내면 전체가 아니라, 바로 그 정련 과정을 통과해 사회적으로 승인된 방향량이다. 이것이 구배선언의 필요성이다.

결국 문명은 본질을 향해 가는 것이 아니라, 자신이 입력받을 수 있는 본질을 향해 간다. 그리고 바로 그 이유 때문에, 새로운 사회적 입력 형식이 필요하다. 문제는 더 많은 욕망을 추출하는 것이 아니다. 문제는 인간이 아직 명시적으로 입력하지 못한 가치, 그러나 자신의 삶을 구성해 온 배경 조건들을 어떤 과정을 통해 식별하고, 정련하고, 책임 아래 선언할 수 있게 할 것인가다. 수단이 목적을 압도한 시대에, 사회가 인간을 무엇으로 읽을 것인가는 더 이상 부차적인 제도 설계의 문

²이 절의 기술철학적 배경으로는 Martin Heidegger, “The Question Concerning Technology”; Max Weber, Economy and Society; Max Weber, The Protestant Ethic and the Spirit of Capitalism; Jacques Ellul, The Technological Society; James C. Scott, Seeing Like a State; Michael Polanyi, The Tacit Dimension; Charles Goodhart, “Problems of Monetary Management: The U.K. Experience”; David Deutsch and Chiara Marletto, “Constructor Theory of Information”을 참조할 수 있다.

제가 아니다. 그것은 문명이 어떤 가능성 공간을 열고, 어떤 가능성 공간을 영원히 닫을 것인가를 결정하는 가장 근본적인 운영 문제가 된다.

바로 이 지점에서 AI가 나타났다. AI는 우리가 오랫동안 축적해 온 수단들이 어느 순간 하나의 얼굴을 갖게 된 사건에 가깝다. 지식의 복제, 계산의 고도화, 분과의 전문화, 실행 비용의 하락이 끝까지 밀고 간 끝에, 이제 문명은 자신이 만든 수단의 총합과 마주하게 되었다.

AI의 불안은 이제 너무 많은 것이 가능해졌다는 데 있다. 번역하고, 찾고, 요약하고, 설계하고, 연결하고, 생산하는 비용이 낮아질수록 우리는 더 많은 일을 더 쉽게 할 수 있게 된다. 그러나 더 많은 일을 할 수 있다는 사실은 무엇을 해야 하는지에 대한 답을 주지 않는다. 오히려 그 질문을 더 잔인하게 드러낸다.

지금까지 사회는 인간의 입력을 대체로 두 가지 방식으로 읽어 왔다. 정치는 표를 읽었고, 자본주의는 가격을 읽었다. 표는 권한을 배분했고, 가격은 자원을 배분했다. 자본주의는 욕망을 가격으로 번역하고, 민주주의는 의사를 표로 번역한다. 이 두 신호는 강력했다. 그러나 그것들은 인간을 매우 낮은 해상도로 읽는다. 사람은 단순히 무엇을 사고 싶은 존재도, 몇 년에 한 번 누군가에게 권한을 위임하는 존재도 아니다. 사람은 어떤 삶을 살고 싶은지, 어떤 세계를 감당할 수 있는지, 무엇을 위해 자신의 시간과 능력과 책임을 걸 수 있는지를 묻는 존재다.

우리는 지금 바로 그 병목 앞에 서 있다. 그러나 여기서 말하는 병목은 단순히 더 많은 의견을 수집하지 못한다는 뜻이 아니다. 자본주의의 포화, 기술적 수단의 과잉, 분과적 전문화, 민주주의의 저해상도, 제안의 포화는 서로 다른 현상처럼 보이지만, 문명이 실제로 방향을 정해야 하는 순간에는 하나의 질문으로 합류한다. 사회는 인간을 무엇으로 읽고 있는가. 소비자로 읽는가, 유권자로 읽는가, 사용자로 읽는가, 아니면 자신이 감당할 세계를 선언하고 책임질 수 있는 주체로 읽는가. 따라서 입력의 문제는 단순한 정보 처리의 문제가 아니라, 인간을 어떤 존재로 인정할 것인가의 문제다.

결국 문제는 다시 이 오래된 질문으로 돌아온다. 우리는 무엇을 위해 살아가는가. 우리는 누구인가. 이제 이 질문은 사적인 명상도, 철학자의 취미도 아니다. 수단이 목적을 압도한 시대에, 그것은 문명이 더 이상 미룰 수 없는 가장 현실적인 운영 문제이며, 동시에 인간 자신에 대한 질문이 되었다.

2 자아의 비존재성과 구배선언

자, 그렇다면 이 문제는 어디서부터 다시 짚어야 하는가. 기술이 빠르다는 진단만으로는 충분하지 않다. 발전이란 무엇인지, 그 발전은 어떤 의도 위에서 정당화되는지, 그리고 의도와 목적이라는 말은 어디에서 그 권위를 얻는지를 물어야 한다. 더 나아가 목적을 말한다는 것 자체가 이미 인간을 특별한 주체로 전제하는 일은 아닌지도 물어야 한다. 그렇다면 인간이란 무엇인가. 목적을 세우고, 선택하고, 책임진다고 여겨지는 그 주체는 무엇인가. 결국 질문은 가장 낮익고도 가장 불분명한 대상에게로 돌아온다. “나”가 “나”라고 여기는 그것은 무엇인가.

우리는 스스로를 “시간 여행자”로 인식한다. 과거의 기억을 담지하고, 현재를 통과하며, 미래로 나아가는 하나의 단단한 실체가 있다고 믿는다. 어제의 나, 오늘의 나, 내일의 나는 서로 다른 상태에 놓여 있지만, 그 밑에는 동일한 “나”가 계속 이어지고 있다고 여긴다. 이 믿음은 너무 자연스러워서 거의 의심되지 않는다. 우리는 자신을 사건들의 묶음이 아니라, 시간 위를 지나가는 하나의 주체로 경험한다.

그러나 동시에 우리는 기계론적 결정론을 거부한다. 인간은 단순히 이전 상태가 다음 상태를 필연적으로 산출하는 물체가 아니며, 매 순간 달리 선택할 수도 있었다고 믿는다. 이 믿음이 없다면

책임이라는 말은 성립하기 어렵다. 우리는 방아쇠를 당긴 행위와 강풍에 쓰러진 나무가 덮친 사건을 구분하는데, 전자에만 죄와 책임을 묻는 근거는 “그는 달리 행동할 수도 있었다”는 규범적 전제에 있다. 도덕 판단은 바로 이 열린 가능성 위에서 있다.

여기서 긴장이 발생한다. 우리가 정말로 자유의지를 가진 존재라면, 미래는 닫힌 단일 경로가 아니다. 같은 현재에서 여러 가능한 선택과 여러 가능한 자아의 전개가 열려 있어야 한다. 그런데 우리가 동시에 “나는 과거로부터 지금까지 하나의 실체로 이어져 왔다”고 말하려면, 그 수많은 가능성들 속에서도 어떤 단단한 자아의 사슬이 보존되어야 한다. 자유의지는 미래를 열어젖히고, 실체적 자아는 시간을 하나의 사슬로 묶으려 한다. 이 둘은 일상적으로는 함께 믿어지지만, 엄밀히 들여다보면 서로를 압박한다.

문제는 세계가 우리의 자아를 완전히 기록하지 않는다는 데 있다. 우리가 거주하는 세계의 증거는 본질적으로 불완전하다. 우주는 일종의 해상도 낮은 카메라와 같아서, 우리의 거시적 행동 궤적은 어느 정도 남기지만, 내면의 미세한 망설임, 스쳐 간 감정, 선택되지 않은 가능성, 말하지 않은 생각까지 완벽하게 보존하지는 않는다. 내가 어떤 말을 하지 않았다는 사실은 남을 수 있지만, 그 말을 하기 직전의 미묘한 충동과 두려움과 자기검열의 결은 대부분 사라진다.

따라서 서로 다른 자아 상태들이 같은 흔적을 남길 수 있다. 어떤 사람이 진심으로 용서했는지, 체념했는지, 계산적으로 침묵했는지, 혹은 그 모든 감정이 뒤섞인 채 행동했는지는 시간이 지나면 현재의 증거만으로 구분되지 않을 수 있다. 겉으로 남은 말과 행동은 같지만, 그 순간의 내면은 서로 달랐을 수 있다. 그렇다면 현재의 세계 상태에서부터 과거의 자아를 하나의 정확한 사슬로 역산하는 일은 불가능해진다.

여기서 한 가지 더 불편한 사실을 보태야 한다. 우리가 흔히 “현재의 나”라고 부르는 것은 생각보다 빈약한 대상이다. 엄밀히 말해 현재는 시간 위에서 부피를 갖지 않는다. 그것은 폭을 가진 구간이라기보다, 과거와 미래가 맞닿는 경계에 가깝다. 내가 “나”라고 가리키는 순간, 그 가리킴은 이미 조금 전의 감각, 조금 전의 생각, 조금 전의 기억을 붙잡고 있다. 순수한 현재의 나는 붙잡히기 전에 곧장 과거가 된다.

그러므로 우리가 “나”라고 부르는 것의 대부분은 현재가 아니라 과거다. 나는 방금 전의 나, 어제의 나, 몇 년 전의 나, 그리고 그 흔적들을 엮어 만든 서사를 나라고 부른다. 현재는 거의 부피가 없고, 미래는 아직 오지 않았으므로, 실체적 자아라는 믿음은 사실상 과거의 보존 위에 기대고 있다. 그런데 그 과거의 흔적은 완전하지 않다. 내면의 미세한 고뇌, 말하지 않은 생각, 선택되지 않은 가능성은 대부분 기록되지 않는다. 그렇다면 자아가 실체로 존재하려면 과거가 나를 보존해 주어야 하는데, 정작 과거는 나를 온전히 보존하지 못한다.

이것을 더 극단적으로 생각해보자. 과거의 어떤 시점에서 나는 A를 선택할 수도 있었고 B를 선택할 수도 있었다. 자유의지를 인정한다면, 두 경로는 모두 가능한 미래였다. 그런데 시간이 지나 두 경로가 현재의 증거 상태에서는 구분되지 않는 하나의 지점으로 다시 합류할 수 있다면 어떻게 되는가. 과거에 서로 다른 고뇌와 감정과 선택 가능성이 있었더라도, 그것이 현재 세계의 어떤 관찰 가능한 차이도 만들지 못한다면, 그 차이는 정보적으로 소멸한다. 이것이 경로의 병합이다.

지금 당신의 방 안에 코끼리가 있었다가 모든 흔적을 지우고 사라졌다고 해보자. 현재의 방이 코끼리가 한 번도 없었던 방과 완전히 구분되지 않는다면, “코끼리가 있었다”는 말은 현재 세계를 설명하는 데 어떤 차이를 만드는가. 물론 우리는 이야기로는 그것을 상상할 수 있다. 그러나 흔적도 없고, 재구성할 방법도 없고, 현재 세계의 어떤 구분 가능한 사실도 바꾸지 못한다면, 그 사건을 현재의 존재론 안에서 실체로 붙잡을 근거는 사라진다.

과거의 자아 상태도 이와 비슷하다. 10년 전 오늘 점심에 느꼈던 미묘한 감정의 결은 지금 우주

어디에 기록되어 있는가? 그것이 기억 속에도, 몸의 상태에도, 타인의 기록에도, 세계의 어떤 물리적 흔적에도 충분히 남아 있지 않다면, 그리고 지금의 나로부터 정확히 다시 도달할 수도 없다면, 그 감정의 결은 어떤 의미에서 존재한다고 말할 수 있는가. 과거의 그 자아 상태는 한때 체험되었을지 모르지만, 현재의 관점에서는 더 이상 실체적 향으로 남아 있지 않다.

따라서 우리가 상상하는 방식의 “시간을 관통하는 자유의지를 가진 자아”가 존재하지 않는다는 뜻이다. 과거의 자아 상태들은 대부분 완전한 흔적으로 보존되지 않고, 현재에서 정확히 복원되지 않으며, 세계의 구분 가능한 기술에 차이를 만들지 못한 채 사라진다. 그러므로 자아를 과거로부터 이어져 온 실체로 이해하는 방식은 무너진다.

그렇다면 지금 여기에서 고뇌하고 선택하는 “나”의 존립 근거는 어디에 있는가.

2.1 발굴에서 건축으로: 비존재적 구조로서의 자아

과거의 인과 사슬이 소실된 지점에서, 자아를 바라보는 관점은 “발굴”에서 “건축”으로 전환되어야 한다. 자아는 과거로부터 계승되는 단단한 실체가 아니라, 미래를 향해 스스로를 기립시키는 “비존재적 구조”이기 때문이다. 이 구조는 “나는 과거에 무엇이었는가”라는 존재론적 질문에 답하지 않는다. 대신 “나는 무엇을 향해 가는가”라는 목적론적 질문을 통해 매 순간 스스로를 기원시킨다.

이는 칸트의 “목적의 왕국” 개념과 맞닿아 있다. 칸트의 세계에서 이성적 주체는 자연 법칙(물리적 인과)에 종속된 객체가 아니라, 스스로 법칙을 수립하고 준수하는 자율적 입법자이다. 우리의 모델도 동일한 결론에 도달한다. 물리적 증거가 소실된 빈 서판 위에서, 자아는 “나는 이러한 존재로서 나 자신을 승인하겠다”는 최상위 방향량을 헌법처럼 선언함으로써 비로소 성립한다. 이 선언은 단순한 다짐이 아니라, 흩어진 가능성의 파편들을 하나의 주체로 묶어내는 중력점이며, 따라서 자아는 발견되는 “대상”이 아니라 목적의 왕국 안에서 끊임없이 갱신되는 “선언된 지위”이다.

이 구조는 허수 i 의 비유로 간단히 볼 수 있다. 실수축만으로는 $x^2 + 1 = 0$ 의 해를 얻을 수 없듯, 물리적 인과만으로는 책임과 존엄을 충분히 설명하기 어렵다. 생물학적 뇌에서 “자아”라는 물질이 발견되지는 않지만, “스스로 목적을 수립하고 책임지는 구조”라는 작동 방식을 도입할 때 비로소 인간이라는 방정식은 해를 갖는다.

실체가 없다는 것(non-substantiality)은 결함이 아니라, 물리적 결정론의 외부에서 작동하기 위한 조건이다. 그리고 이 비존재적 구조는 고립된 개인의 환상이 아니다. 블록체인이 비트열 자체보다 참여자들의 합의에서 가치를 얻듯, 자아도 영혼이라는 “중앙 서버”가 아니라 서로를 자유의지를 가진 주체로 대우하겠다는 상호 인정 속에서 유지된다.

이 비실체성은 자아 자신에게만 해당하지 않는다. 자아가 가치적으로 받아들이는 타자 역시 미리 결정된 생물학적 개체나 고정된 도덕적 실체가 아니다. 가치장, 복지, 의지, 공감 가능성이 귀속되는 것으로 인정될 때에만 타자는 자아에게 의미 있는 좌표로 떠오른다. 따라서 본 원고는 자아 측의 “나”와 타자 측의 “너” 양쪽을 같은 비존재적 좌표의 두 면으로 다루며, 이 둘을 공통의 용어로 주체(subject)라고 부른다. 주체는 발견되는 실체가 아니라 인정되고 귀속되는 좌표이며, 따라서 사람, 동물, 태아, 미래의 나, 문화, 풍경, 유품을 같은 문법 안에 둘 수 있는 통일 용어가 된다. 어떤 대상이 주체로 인정되는가, 그 인정이 얼마나 정합한가, 그 인정이 반박 · 수정 가능한가는 모두 별개의 문제로 다루진다.

이 인정은 가치 발생의 조건이다. 자아를 인정하지 않으면 세계에는 사건, 반응, 물리적 인과, 신경계 상태, 행동 패턴만 남는다. 누군가가 어떤 말을 했고, 어떤 손이 어떤 물체를 움직였고, 어떤 자극에 어떤 뇌 반응이 일어났다는 기술은 가능하다. 그러나 그 상태에서는 책임도, 가치도, 목적도

성립하지 않는다. 어떤 존재를 단순한 인과적 물체가 아니라, 세계를 자기에게 의미 있는 것으로 받을 수 있고 책임과 의지를 가질 수 있는 주체로 인정할 때에만 세계는 비로소 가치의 언어로 열리기 시작한다.

자아가 인정되면 세계는 더 이상 중립적인 사실들의 집합이 아니다. 같은 사건도 어떤 자아에게는 소중한고, 어떤 자아에게는 견딜 수 없고, 어떤 자아에게는 무의미하고, 어떤 자아에게는 자기 존재를 걸 만큼 중요해진다. 여기서 가치는 세상에 독립적으로 붙어 있는 물리량이라기보다, 자아와 세계의 관계에서 발생하는 의미값이다. 자아가 생긴다는 것은 세계 위에 가치가 덧칠된다는 뜻이 아니라, 세계가 처음부터 어떤 자아에게 좋은 것, 나쁜 것, 위험한 것, 포기할 수 없는 것, 감당 가능한 것으로 나타난다는 뜻이다.

이때 자아는 세계를 있는 그대로 보지 않는다. 언제나 어떤 가치사상(value mapping)을 통해 본다. 여기서 가치관은 “나는 자유가 중요하다”, “나는 사랑이 중요하다” 같은 명제 목록이 아니다. 더 정확히 말하면, 가치관은 자아가 세계의 상태, 사건, 가능성, 관계, 미래 선택지에 가치를 부여하는 해석 연산자다. 내용이라기보다 사상이고, 함수이고, 세계를 자기에게 의미 있는 장으로 변환하는 작동 방식에 가깝다.

자아가 가치사상을 통해 세계를 보면, 세계는 가치장(value field)으로 나타난다. 가치장이란 자아가 가치사상을 통해 세계를 본 결과, 세계의 상태와 가능성들에 좋음, 나쁨, 중요함, 위험함, 포기 불가능함, 감당 가능함 같은 가치가 분포된 장이다. 형식적으로 쓰면, 자아 S 가 가능상태들의 공간 $\mathcal{X}$ 를 가치사상 M_S 로 해석할 때 그 결과 생기는 가치분포가 가치장이다. 여기서 $\mathcal{X}$ 는 단순한 가능세계의 목록이 아니라, 구배를 말할 수 있도록 좌표화되었거나 매끄러운 상태표현 공간에 임베딩된 가능상태 공간으로 둔다. 아직 정련되지 않은 가치장을 V_S^0 라 하면, 이는 다음처럼 여러 가치축을 가진 장으로 나타낼 수 있다.

$$V_S^0 : \mathcal{X} \rightarrow \mathbb{R}^m$$

여기서 m 은 자유, 안정, 사랑, 인정, 창조, 책임처럼 서로 다른 가치축의 수를 뜻한다. 이 경우 가치장은 하나의 스칼라 함수라기보다 다음과 같은 다중 가치분포에 가깝다.

$$V_S^0 = (V_{S,1}^0, \dots, V_{S,m}^0)$$

그러나 가치장은 처음부터 정합적이지 않다. 자아는 자유를 원하면서 안정도 원하고, 사랑을 원하면서 독립도 원하고, 인정을 원하면서 예외자가 되고 싶고, 공동체를 원하면서 지배받기는 싫어한다. 이때 가치장 안에는 여러 방향의 힘이 생긴다. 현재 자아의 위치를 $x_t \in \mathcal{X}$ 라고 하면, 각각의 가치축은 서로 다른 구배를 낼 수 있다.

$$\nabla V_{S,1}^0(x_t), \quad \nabla V_{S,2}^0(x_t), \quad \dots, \quad \nabla V_{S,m}^0(x_t)$$

이 구배들이 같은 방향으로 정렬되어 있지 않을 때, 자아는 목적이 없는 것이 아니다. 오히려 서로 충돌하는 복수의 방향량 후보들 속에 있다.

그래서 필요한 과정이 가치정련(value refinement)이다. 가치정련은 단순히 “내가 뭘 원하는지 찾기”가 아니다. 그것은 자아의 가치장 안에 흩어져 있거나 충돌하는 가치들을 검토하고, 위계화하고, 상충을 해소하거나 감당 가능한 긴장으로 재배치하여, 하나의 우세한 실천 방향이 산출될 수 있게 하는 과정이다. 독서, 사유, 대화, 반박, 실패, 후회, 실험, 관계, 기록, 자기해석은 모두 이 과정에 속한다. 의지구배는 로그나 클릭에서 자동 추출되는 정보가 아니라, 바로 이 느린 가치정련을

통해 생겨난다.

이를 하나의 연산자로 쓰면, 가치정련은 정련되지 않은 가치장 V_S^0 를 정련된 가치장 V_S^* 로 바꾸는 과정이다.

$$V_S^* = \mathcal{R}(V_S^0), \quad V_S^* : \mathcal{X} \rightarrow \mathbb{R}^m$$

여기서 $\mathcal{R}$ 은 단순 평균이나 취향의 합산이 아니다. 그것은 가치들의 충돌을 검토하고, 위계화하고, 자기기만을 줄이고, 장기적 자기정합성을 확보하여, 실천 판단이 가능하도록 가치장을 재구성하는 과정이다. 따라서 V_S^* 는 순간 자극이 제거되어도 유지되는지, 미래의 자기에게도 설명 가능한지, 다른 가치들과 정합적인지, 비용과 책임을 포함해도 승인 가능한지, 반복된 숙고와 반박과 경험 이후에도 방향성이 남는지를 통과한 가치장이다. 엄밀히 말하면 $\mathcal{R}$ 은 정련의 절차를 나타내는 연산자이므로, 적어도 안정성 조건 $\mathcal{R}(\mathcal{R}(V)) \approx \mathcal{R}(V)$ 를 만족해야 한다. 즉 한 번 충분히 정련된 가치장은 같은 숙고 절차를 다시 통과해도 완전히 다른 장으로 뒤집히지 않아야 한다.

다만 V_S^* 는 여전히 여러 가치축을 가진 다중 가치장이다. 그러므로 실천적 비교와 경로 선택을 위해서는 이 다중 가치장을 하나의 평가값으로 읽는 스칼라화 함수가 별도로 필요하다. 이를 $\sigma_S^* : \mathbb{R}^m \rightarrow \mathbb{R}$ 라 하고, 정련된 스칼라 평가함수를 다음처럼 둔다.

$$u_S^* = \sigma_S^* \circ V_S^*, \quad u_S^* : \mathcal{X} \rightarrow \mathbb{R}$$

여기서 σ_S^* 는 가치의 다원성을 삭제하는 단일 효용 함수가 아니다. 그것은 특정 자아가 숙고와 책임 아래 여러 가치축의 우선순위, 금지 조건, 타협 가능 범위, 시간축을 반영하여 경로를 비교할 수 있게 만드는 실천적 읽기 방식이다. 가치정련은 V_S^* 를 만들고, 스칼라화는 그 정련된 다중 가치장을 특정 판단 국면에서 비교 가능한 평가함수 u_S^* 로 읽는다. 이 둘을 구분해야 의지구배가 단순한 쾌락 합산으로 환원되지 않는다.

가치장이 충분히 정련되면, 그것은 단순한 가치 분포에 머물지 않고 현재 위치에서의 실천적 기울기를 낸다. 이를 의지구배(volitional gradient)라 부르자. 상태공간 위에서 직관적으로 쓰면 다음과 같다.

$$w_t^{\text{state}} = \nabla u_S^*(x_t)$$

그러나 인간의 의지는 보통 단일 행동보다 긴 경로, 정책, 삶의 방식, 커밋 구조를 대상으로 한다. 따라서 의지구배는 더 정확히는 정련된 평가함수 위에서 장기 경로함수 J_S^* 가 만드는 구배로 볼 수 있다. 이때 정책을 π_θ 로 표기하자. 여기서 θ 는 기계학습 모델의 파라미터만을 뜻하지 않고, 삶의 경로, 커밋 구조, 실천 양식을 좌표화한 표현을 뜻한다.

$$w_t = \nabla_\theta J_S^*(\theta \mid x_t)$$

경로가 어떻게 생성되는지도 최소한으로 명시할 수 있다. 행동 공간을 $\mathcal{A}$ 라 하고, 시점 t 의 행동을 $a_t \in \mathcal{A}$ 라 하자. 정책 π_θ 와 상태 전이 P 는 다음과 같은 경로분포를 유도한다.

$$a_t \sim \pi_\theta(\cdot \mid x_t), \quad x_{t+1} \sim P(\cdot \mid x_t, a_t)$$

그러면 J_S^* 는 예컨대 다음처럼 쓸 수 있다.

$$J_S^*(\theta \mid x_t) = \mathbb{E}_{\pi_\theta, P} \left[\sum_{k=0}^{\infty} \gamma^k u_S^*(x_{t+k}) \right]$$

즉 어떤 정책 또는 삶의 경로 표현 θ 를 택했을 때, 정련된 가치장 위에서 장기적으로 얼마나 자기정합적 가치가 증가하는지를 보는 함수다. 의지구배는 고정된 목적지가 아니다. 그것은 정련된 가치장이 자아의 현재 위치와 가능한 경로들 위에서 산출하는 실천적 기울기이며, 자아가 자신의 시간, 능력, 책임, 기여를 결속시킬 때 가치가 가장 강하게 증가하는 방향과 세기다. “목적벡터”가 어딘가에 도착해야 할 목표물을 암시한다면, 의지구배는 지금의 나와 세계 상태에서 가치가 가장 강하게 쏘는 방향을 가리킨다.

이 점에서 의지구배는 말초적 욕망과 구분된다. 말초적 욕망은 대체로 현재 시점의 즉각 보상 함수에서 나온다. 현재 보상함수를 $R_t : \mathcal{X} \times \mathcal{A} \rightarrow \mathbb{R}$ 라 하면, 말초적 욕망의 구배는 다음처럼 쓸 수 있다.

$$d_t = \nabla_a R_t(x_t, a)$$

이는 지금 이 상태에서 어떤 행동 a 가 즉각적 보상, 긴장 해소, 자극, 회피, 인정 욕구를 가장 빠르게 증가시키는지를 가리킨다. 릴스를 넘기고 싶다, 단 것을 먹고 싶다, 화를 내고 싶다, 도망치고 싶다는 충동은 대체로 이 현재 보상함수 R_t 의 구배에 가깝다. 반면 의지구배는 현재 보상함수가 아니라 정련된 가치장 V_S^* 와 그 장기 경로함수 J_S^* 에서 나온다.

따라서 말초욕망과 의지구배의 차이는 단순히 얕고 깊은 욕망의 차이가 아니다. 둘은 서로 다른 목적함수의 구배다.

$$w_t = \nabla_\theta J_S^*(\theta \mid x_t), \quad d_t = \nabla_a R_t(x_t, a),$$

현재 욕망을 더 거칠게 분해하면, 스칼라화된 현재 평가함수는 정련된 평가함수와 순간적 노이즈의 합으로 볼 수 있다.

$$u_{S,t}^0(x) = u_S^*(x) + \eta_t(x)$$

여기서 η_t 는 피로, 자극, 불안, 충동, 환경적 노이즈를 나타내는 스칼라 교란항이다. 그러면 현재 평가장의 방향은 다음처럼 분해된다.

$$\nabla u_{S,t}^0(x_t) = \nabla u_S^*(x_t) + \nabla \eta_t(x_t)$$

말초욕망이 강할 때는 $\nabla \eta_t$ 가 ∇u_S^* 를 압도한다. 반대로 어떤 욕망은 의지구배와 같은 방향일 수 있다. 다만 d_t 는 행동공간의 구배이고 w_t 는 정책 또는 경로 표현의 구배이므로, 둘을 직접 더할 수는 없다. 의지구배를 현재 행동공간으로 사상하는 투영 연산자 P_t^{act} 를 두면, 현재 욕망구배는 다음처럼 의지구배 성분과 노이즈 성분의 합으로 근사될 수 있다.

$$d_t = \alpha_t P_t^{\text{act}} w_t + \epsilon_t$$

여기서 α_t 는 현재 욕망이 의지구배와 정렬된 정도이고, ϵ_t 는 피로, 자극, 충동, 불안, 환경 노이즈를

뜻한다. 그러므로 말초욕망은 항상 거짓이 아니며, 어떤 욕망은 의지의 표현일 수 있다.

그러나 제3프로토콜이 읽어야 하는 것은 d_t 가 아니다. 제3프로토콜의 입력은 현재 보상함수의 국소구배가 아니라, 가치정련을 통과해 자아가 자기 책임 아래 승인한 장기 · 정합 구배다.

$$\text{Third Protocol input} \neq \nabla_a R_t(x_t, a), \quad \text{Third Protocol input} = D_S(t)$$

여기서 $D_S(t)$ 는 단순한 벡터가 아니라 선언 객체다. 예컨대 다음처럼 쓸 수 있다.

$$D_S(t) = \text{Declares}(g_S^{\text{decl}}, \Gamma_S, \rho_S, \ell_S, \chi_S, t), \quad g_S^{\text{decl}} \approx \nabla_{\theta} J_S^*(\theta | x_t)$$

여기서 g_S^{decl} 은 자아가 승인한 의지구배의 표현이고, Γ_S 는 조건부 결속과 책임 구조, ρ_S 는 수정 이력, ℓ_S 는 공개 범위, χ_S 는 대리권 또는 에이전트가 행사할 수 있는 권한의 범위를 뜻한다. 따라서 제3프로토콜은 별거벗은 내면 벡터를 읽는 체계가 아니라, 책임과 조건과 공개 범위를 가진 선언 객체를 읽는 체계다.

마지막으로 선언은 여전히 중요하다. 다만 선언되는 것은 가치관 전체도, 가치장 전체도, 내면 전체도 아니다. 선언은 정련된 가치장에서 산출된 의지구배를, 자아가 자기 책임 아래 사회적으로 압히도록 승인하고 고정하는 사건이다. 이를 구배선언(*gradient declaration*)이라 부를 수 있다.

제3프로토콜이 읽어야 하는 것은 인간의 내부 전체가 아니라, 바로 이 구배선언이다. 이 구분이 무너지면 제3프로토콜은 인간의 내면을 추출하고 프로파일링하는 감시 체계가 된다. 반대로 이 구분이 지켜질 때, 선언은 추출이 아니라 승인이고, 사회적 입력은 감시가 아니라 책임의 형식이 된다.

결국 우리의 논증은 허무주의가 아닌 구원으로 귀결된다. 과거의 정보가 삭제된다는 사실은, 역설적으로 과거가 우리를 영원히 규정하거나 구속할 수 없다는 자유의 가능성을 개방한다. 그 정보의 공백을 채우는 것은 물리적 잔여물이 아니라, 자아가 스스로 인정하고 정련하고 선언한 의지구배다. “나”는 박물관에 박제된 조각상이 아니다. “나”는 매 순간 불완전한 증거 위에서 재구성되는 현상이자, 자기 가치장을 정련하여 어떤 방향으로 힘을 발생시킬지 승인하는 비존재적 구조다. 실체가 없다는 것은, 우리가 과거의 유령이 아니라 미래를 향한 구배선언으로서 매번 새롭게 기원할 수 있음을 의미한다. 우리는 완전히 기억되지 않기 때문에 비로소 자유롭고, 서로를 목적으로 대우하기로 승인하기에 비로소 존재한다.

3 구배선언의 사회화: 개인의 가치장에서 사회적 입력으로

자아가 과거로부터 보존되는 실체가 아니라면, 인간은 현재에서 미래를 향해 스스로를 승인함으로써만 주체가 된다. 그러나 그 승인에는 내부 발생 과정이 있다. 자아가 인정되어야 가치가 발생하고, 자아는 가치사상을 통해 세계를 가치장으로 본다. 그 가치장이 정련되면 의지구배가 산출되고, 그 의지구배가 사회적으로 읽힐 수 있는 형식으로 승인될 때 구배선언이 된다. 이 전환이 본 원고의 핵심 경첩이다. 사회가 읽어야 하는 것은 인간의 내면 전체가 아니라, 자아가 정련을 거쳐 책임 아래 내놓은 이 선언된 방향량이다.

경제는 인간을 선호의 보유자로 읽고, 정치는 인간을 의견의 보유자로 읽으며, 플랫폼은 인간을 반응의 발생원으로 읽는다. 그러나 인간은 그보다 앞서 세계를 가치장으로 해석하고, 자신이 감당할 방향을 승인할 수 있는 존재다. 구배선언은 이미 완성된 채 발견되는 데이터가 아니라, 형성되고 정련되고 승인되고 책임지는 자기현법이다. 그러므로 사회는 인간을 소비자, 유권자, 사용자로만

읽는 데서 멈출 수 없고, 구매선언자로 읽을 수 있어야 한다.

이 전환은 존재론에서 제도론으로 바로 이어진다. 개인 수준에서 구매선언은 자아가 자신의 방향을 승인하는 사건이다. 사회 수준에서 구매선언은 표와 가격으로는 포착되지 않는 사회적 입력을 만든다. 문명 수준에서 구매선언의 집계 방식은 새로운 사회계약을 만든다. 만약 구매선언이 개인 자아의 헌법적 승인이라면, 사회는 그 선언들을 읽고 조정하는 상위 헌정 시스템을 가져야 한다. 다음 절의 제3프로토콜은 바로 이 요구에서 출발한다.

4 투표와 가격을 넘어: 제3프로토콜

여기서 다시 사회 전체의 입력 구조로 돌아갈 필요가 있다. 정치는 표를 집계하고 경제는 가격과 계약을 집계한다. 그러나 실행 비용이 충분히 낮아진 사회에서는 이 두 프로토콜만으로는 더 이상 사회의 조정 가능성을 충분히 회수할 수 없다. 특히 “이런 학교가 실제로 생기면 보내겠다”, “이런 규칙이 보장되면 이주하겠다”, “이 정도의 사람들이 모이면 나는 돈과 시간과 전문성을 걸겠다”

같은 조건부 집단형 수요는 정치로는 지나치게 거칠고 느리며, 경제로는 아직 거래 객체가 형성되기 이전이어서 충분히 포착되지 않는다.

마스다 무네아키의 『지적자본론』은 이러한 변화를 소비사회의 단계 전환으로 읽는다. 대량 생산과 대량유통의 시대에는 상품 자체가 희소했다. 그러나 상품과 플랫폼이 넘쳐나는 제3의 소비 사회에서는 더 이상 물건의 양이 핵심이 아니다. 중요한 것은 “제안”이다. 소비자는 단순히 더 많은 물건을 원하는 것이 아니라, 자신의 취향과 생활양식에 맞게 편집된 의미 있는 선택지를 원한다. 그래서 오늘날의 기업들은 상품을 진열하는 데서 멈추지 않는다. 그들은 끊임없이 말한다. 너는 이런 취향을 좋아할 것이다. 너는 이런 공간을 원할 것이다. 너는 이런 관념과 이런 삶의 방식을 사랑하게 될 것이다. 애플을 비롯한 현대의 강한 브랜드들은 제품을 파는 동시에, 인간에게 하나의 세계관과 생활양식을 제안한다.

지금 우리는 제안의 포화 속에 있다. 알고리즘과 브랜드와 플랫폼은 매 순간 우리 앞에 가능한 삶의 양식들을 밀어 넣는다. 무엇을 볼지, 무엇을 살지, 어떤 취향을 가질지, 어떤 사람이 될지까지 끝없이 제안된다. 이 풍요는 역설적으로 개인의 결정 능력을 마비시킨다. 너무 많은 제안은 인간이 스스로 무엇을 사랑할 수 있는지 묻는 힘을 약화시키고, 더 나아가 자신이 사랑할 수 있는 것을 직접 빚어낼 수 있다는 사실마저 잊게 만든다. 사랑은 언제나 주어진 선택지 중 가장 그럴듯한 것을 고르는 일만은 아닐것이다. 때로 사랑은 아직 세계에 없는 형상을 오래 바라보고, 깎고, 다듬고, 마침내 그것에 생명을 불어넣고 싶어지는 일에 가까울것이다. 피그말리온의 비극적이고도 아름다운 직관은 여기에 있다. 그는 제안된 이상형을 선택한 것이 아니라, 자신이 사랑할 수밖에 없는 대상을 스스로 빚어냈다. 그러나 제안의 시대는 인간을 더 세련된 선택자로 만들 뿐, 자신이 사랑할 세계를 창조하고 선언하는 존재로 세우지는 못한다.

더욱이 수단이 극대화되고 있는 지금, 제안은 점점 더 뭉뚱한 형식이 된다. 실행 비용이 낮아지고 생산과 조합의 가능성이 폭발하는 시대에는, “이런 것을 좋아할 것”이라는 추천만으로는 충분하지 않다. 중요한 것은 이미 존재하는 선택지 중 무엇을 고를 것인가가 아니라, 아직 존재하지 않는 무엇을 어떤 조건에서 함께 만들 것인가이다. 따라서 제3의 소비사회가 기업과 플랫폼이 인간에게 삶의 방식을 제안하는 시대였다면, 앞으로 필요한 전환은 제안에서 선언으로의 전환이다. 인간이 자신의 의지구배를 먼저 선언하고, 그 선언을 중심으로 상품, 조직, 도시, 교육, 관계가 역으로 구성되는 단계로 나아가야 한다.

이때 필요한 것은 표와 가격을 대체하는 체계가 아니라, 정치와 경제와 병렬로 작동하는 또 하

나의 사회적 입출력 프로토콜이다. 정치가 표를 통해, 경제가 가격을 통해 사회적 의사를 표현하고 집계하고 다시 실행으로 되돌려 보내는 체계였다면, 지능의 총량이 커지고 실행비용이 낮아지는 사회는 구매선언을 표현하고 집계하고 다시 실행으로 되돌려 보내는 세 번째 프로토콜을 요구하게 된다. 이것이 여기서 말하는 제3프로토콜이다. 제3프로토콜의 핵심은 개인이 자신의 의지구배를 공적으로 승인하는 행위가, 동시에 참여 의사와 자원 제공 의사와 장기 커밋의 가능성을 함께 드러내는 새로운 사회적 신호가 된다는 데 있다. 다시 말해, 그것은 “무엇을 사고 싶은가”나 “누구에게 표를 줄 것인가”만이 아니라, “어떤 세계를 원하며 어떤 조건이 충족되면 누구와 함께 무엇을 얼마나 걸 수 있는가”를 표현하고, 그 표현을 다시 조직과 계약과 생산으로 되돌려 보내는 세 번째 경로다. 이때 구매선언은 흩어진 의견으로 남지 않고, 서로 비슷한 선언들과 결합하여 의지구배 클러스터를 이루며, 그 자체가 아직 존재하지 않는 재화와 제도와 조직을 향한 생산 신호가 된다.

핵심은 이것이 단순한 의견 게시판이나 더 정교한 수요 예측 엔진이 아니라는 점이다. 제3프로토콜은 사용자의 로그에서 욕구를 추출하는 체계가 아니라, 인간이 읽고, 생각하고, 토론하고, 가치장을 정련하고, 마침내 자신의 의지구배를 선언하게 만드는 구조에 가깝다. 그리고 미래의 대리 지능들은 그 선언 이력과 수정 과정과 가치 우선순위를 먹고 들어와, 서로의 의지구배 간 정합성을 맞추고 공동 목적 후보와 계약 가능한 표현까지 압축하게 된다. 따라서 제3프로토콜은 인간의 숙고를 공적 구매선언으로 바꾸고, 그 선언을 다시 제도와 조직과 대리 지능의 응답으로 되돌려 보내는 사회적 입출력 프로토콜이다.

제3프로토콜이 가치관 전체를 읽으려 해서는 안 된다. 가치관 전체, 가치장 전체, 내면 전체는 입력이 아니다. 그것을 읽겠다고 나서는 순간 제3프로토콜은 인간의 내부를 추출하고 예측하고 프로파일링하는 감시 장치가 된다. 제3프로토콜이 읽어야 하는 것은 자아가 숙고와 정련을 거쳐 사회적으로 내놓기로 승인한 의지구배, 곧 구매선언이다. 이 구분 때문에 제3프로토콜은 내면 추출 기술이 아니라 승인된 방향량을 사회적 사건으로 번역하는 프로토콜이 된다.

뒤에서 다룰 GU와 CVTD의 관점에서 보면, 제3프로토콜은 철학과 실행을 연결하는 제도 적 층이다. GU가 왜 개인의 구매선언을 판단의 최상위 입력으로 복귀시켜야 하는지를 말하고, CVTD가 그러한 의지구배들이 어떻게 의도적 사건 π 의 언어로 계산될 수 있는지를 보여준다면, 제3프로토콜은 바로 그 선언들이 사회적으로 표현되고, 집계되고, 다시 실행으로 되돌아오는 통로를 설명한다. 따라서 그것은 부가적인 서비스가 아니라, 구매선언의 사회적 가시성과 계산 가능성을 여는 새로운 조정 프로토콜이다.

이를 도식화하면 다음과 같다.

프로토콜	구조
1: 표	기본 입력: 표, 권리, 의견. 집계 방식: 선거, 의회, 공문장, 절차적 정당성. 사회적 출력: 법, 정책, 권력 배분. 핵심 한계: 저빈도, 번들링, 다수결 압축.
2: 가격	기본 입력: 가격, 구매, 투자, 노동, 계약. 집계 방식: 시장, 기업, 금융, 거래 네트워크. 사회적 출력: 재화, 서비스, 자본 배분. 핵심 한계: 지불 가능한 현재 선호 중심.
3: 구매선언	기본 입력: 구매선언, 조건부 결속, 기여 가능성, 책임 범위. 집계 방식: 의지구매 클러스터, GU적 조정, 대리 지능 간 협상. 사회적 출력: 조직, 제도, 생산, 계약, 의도적 사건 π. 핵심 한계: 아직 표준화되지 않음.

제1프로토콜이 “누가 권한을 갖는가”를 묻고, 제2프로토콜이 “무엇에 얼마를 지불할 것인가”를 묻는다면, 제3프로토콜은 “어떤 세계를 원하며, 그 세계가 가능해지는 조건에서 무엇을 걸 것인가”를 묻는다. 따라서 제3프로토콜은 정치와 경제의 폐지가 아니라, 둘이 손실 압축해 온 인간의 의지구매를 더 높은 해상도로 사회에 입력하는 보완 회로다.

4.1 현실의 마찰과 사회적 입력의 충실도

그러나 문명이 읽지 못하는 가치는 단지 추상적이고 내면적인 것만은 아니다. 오히려 어떤 가치는 현실 세계 안에서 매우 구체적인 형태로 드러난다. 다만 그것은 표나 가격이나 클릭처럼 정리된 기호로 나타나지 않고, 시간, 거리, 체류, 신체적 피로, 위험 감수, 대기, 반복, 포기된 가능성의 형태로 나타난다. 사람은 무엇을 말하는지보다, 때로는 무엇을 위해 실제로 몸을 옮겼는지, 무엇을 기다렸는지, 무엇을 견뎠는지, 어떤 불편함에도 불구하고 떠나지 않았는지에서 더 정확히 읽힌다.

이 점에서 시위, 줄서기, 현장 대기, 장례식, 법정 방청, 광장의 체류, 도서관의 침묵 같은 것들은 단순히 낡은 사회적 형식이 아니다. 그것들은 인간의 절실함과 우선순위와 책임 가능성을 읽는 오래된 계측 장치다. 줄을 선 사람은 단순히 원한다고 말하는 것이 아니라, 다른 일을 할 수 있었던 시간을 포기한다. 현장에 나온 사람은 단순히 동의한다고 말하는 것이 아니라, 장소에 몸을 묶고, 피로와 노출과 위험을 감수한다. 장례식에 참석한 사람은 슬픔을 표현하는 문장을 제출하는 것이 아니라, 시간과 몸을 통해 관계의 무게를 표시한다. 현실은 이런 비용들을 자동으로 붙인다. 그래서 현실의 행위는 말보다 두껍다.

현실 세계가 강한 이유는 정보량이 많기 때문만이 아니다. 그 정보가 비용과 함께 온다는 점이 중요하다. 누가 왔는가, 얼마나 멀리서 왔는가, 얼마나 오래 남았는가, 무엇을 포기했는가, 어떤 위험을 감수했는가, 누가 올 수 있었는데 오지 않았고, 누가 오기 어려웠는데도 왔는가. 이런 것들은 말의 내용 바깥에 붙은 메타정보다. 그러나 바로 이 메타정보가 사회적 의미의 충실도를 만든다. 어떤 주장이 진실이라는 것을 증명하지는 않더라도, 적어도 그 주장이 공짜로 만들어진 것이 아니라는 사실은 보여준다.

반대로 인터넷은 정보의 확장성에서는 압도적이지만, 이 메타정보를 쉽게 지운다. 온라인에서 말은 빠르게 복제되고, 반응은 즉시 집계되며, 주장은 멀리 퍼진다. 그러나 그 말이 어떤 비용을 통과했는지는 불분명해진다. 클릭은 쉽고, 공유는 가볍고, 분노는 빠르게 생산된다. 어떤 사람이 실제로 얼마나 읽었는지, 무엇을 포기했는지, 어떤 반박을 통과했는지, 얼마나 오래 남았는지, 어떤

위험을 감수했는지는 피드 위에서 잘 드러나지 않는다. 인터넷은 메시지를 확산하지만, 메시지를 실어 나르던 현실의 두께를 제거한다.

따라서 온라인 공간에서 사회적 입력은 과잉 생산된다. 의견은 많아지고, 반응은 많아지고, 선언은 많아지지만, 그것들이 서로 얼마나 다른 무게를 갖는지는 점점 더 읽기 어려워진다. 누구나 말할 수 있다는 것은 해방이지만, 모든 말이 거의 같은 비용으로 만들어진다는 것은 또 다른 문제다. 말이 너무 싸지면 사회는 그 말의 강도와 진정성과 우선순위를 읽기 위해 다른 장치를 찾게 된다. 사람들은 다시 숫자를 보고, 인증을 보고, 팔로워를 보고, 현장 영상을 보고, 대기열을 보고, 위험을 본다. 말 자체가 아니라 말이 어떤 현실적 마찰을 통과했는지를 보려 한다.

이렇게 보면 현실과 인터넷은 단순히 오프라인과 온라인의 차이가 아니다. 현실은 확장성이 낮지만, 행위에 붙은 비용과 맥락을 함께 보여준다. 인터넷은 확장성이 높지만, 행위에 붙은 비용과 맥락을 분리한다. 현실의 입력은 느리고 비싸지만 두껍고, 인터넷의 입력은 빠르고 싸지만 얇다. 문제는 어느 하나가 절대적으로 우월하다는 것이 아니다. 문제는 사회가 어떤 종류의 입력을 어떤 상황에서 신뢰할 수 있는가에 있다.

문명이 더 많은 정보를 더 빠르게 처리할 수 있게 되었음에도, 어떤 순간에는 여전히 사람들이 실제로 모였다는 사실, 오래 기다렸다는 사실, 떠나지 않았다는 사실, 몸을 드러냈다는 사실이 강한 사회적 의미를 갖는다. 이것은 비합리적 잔재가 아니다. 그것은 사회가 아직도 어떤 종류의 진지함을 현실의 비용을 통해 읽고 있기 때문이다. 말은 비용 없이 생성될 수 있지만, 체류는 비용 없이 생성되지 않는다. 주장은 복제될 수 있지만, 몸의 위치와 시간의 포기는 같은 방식으로 복제되지 않는다.

여기서 다음 문제가 열린다. 사회적 입력이 값싼 말이 되지 않기 위해서는 왜 그 말 바깥의 비용이 필요한가. 어떤 비용은 그 입력을 더 정확히 읽히게 만들고, 어떤 비용은 단순한 낭비나 배제로 변질되는가. 왜 어떤 사회적 메시지는 내용만으로 충분하지 않고, 시간, 몸, 에너지, 위험, 기다림 같은 별도의 운반 조건을 요구하는가. 바로 이 지점에서 이중비용문제가 나타난다.

4.2 이중비용문제: 사회적 가독성의 캐리어 비용

사회는 어떤 발화나 선택을 그대로 사회적 입력으로 인정하지 않는다. 어떤 말이 단순한 말에 그치지 않고 주장, 권리, 순변, 화폐, 표, 결속, 책임으로 읽히기 위해서는 그 말 바깥의 어떤 비용이 함께 붙어야 한다. 사람은 “내가 먼저 받고 싶다”고 말하는 것만으로 순변을 얻지 못한다. 줄을 서야 한다. 시간을 잃고, 신체 에너지를 쓰고, 다른 일을 할 수 있었던 가능성을 포기해야 한다. 시위에 참여하는 사람도 단순히 “나는 이 주장에 동의한다”고 말하는 것이 아니다. 장소에 몸을 옮기고, 시간을 쓰고, 피로와 노출과 위험을 감수함으로써 그 주장의 사회적 무게를 만든다. 비트코인의 작업증명 역시 순수한 데이터 선언이 아니라, 전기와 장비와 계산 시간을 소모했다는 검증 가능한 흔적을 통해 디지털 희소성을 만든다.

이 구조를 이중비용문제라고 부를 수 있다.

이중비용문제란, 어떤 사회적 입력이나 사회적 재화가 작동하기 위해 대상 자체의 비용과 별도로, 그 입력이 값싼 말이나 임의적 기호가 아니라는 것을 사회적으로 읽히게 만드는 두 번째 비용을 요구하는 문제이다. 첫 번째 비용은 대상 비용이다. 상품을 얻기 위해 돈을 내고, 계약을 이행하기 위해 노동을 제공하고, 어떤 산출물을 만들기 위해 재료와 시간을 투입하는 비용이다. 두 번째 비용은 가독성 비용이다. 줄서기의 대기시간, 시위의 신체적 체류와 위험, 작업증명의 전기, 학위 취득의 긴 시간, 반박을 견디는 인지적 비용, 실패 시 책임을 받아들이는 평판 비용처럼, 그 행위가

사회적으로 읽히도록 만드는 비용이다.

다르게 말하면, 사회적 입력은 메시지와 캐리어를 동시에 요구한다. 메시지는 “나는 이것을 원한다”, “나는 이 주장에 선다”, “이 블록은 유효하다”, “나는 이 재화에 접근하고 싶다”와 같은 내용이다. 그러나 메시지만으로는 충분하지 않다. 메시지는 너무 쉽게 만들어지고, 복제되고, 위조될 수 있다. 그래서 사회는 그 메시지가 어떤 캐리어 위에 실려 있는지를 본다. 줄서기에서 메시지는 “나는 이 재화를 원한다”이고, 캐리어는 기다린 시간과 신체적 체류이다. 시위에서 메시지는 정치적 주장이고, 캐리어는 몸, 장소, 시간, 위험, 공동목격이다. 작업증명에서 메시지는 “이 블록은 유효하다”이고, 캐리어는 전기와 연산이다.

여기서 중요한 것은 캐리어 비용이 대개 이전되지 않는다는 점이다. 돈을 지불하면 돈은 판매자에게 이전된다. 노동을 제공하면 산출물은 상대방에게 이전된다. 그러나 줄서기에 쓴 시간은 판매자에게 이전되지 않는다. 비트코인 채굴에 사용된 전기는 코인 안에 저장되지 않는다. 시위 현장에서 감수한 피로와 위험은 정책 결정자에게 이전되지 않는다. 그것들은 대부분 사라진다. 다만 그 사라짐의 흔적이 “이 입력은 공짜로 만들어진 것이 아니다”라는 사회적 가독성을 남긴다.

따라서 두 번째 비용의 본질은 가치이전이 아니라 가치소산을 통한 가독성 생성이다. 줄서기는 상품을 더 많이 만들지 않지만 순번을 정당화한다. 작업증명은 전기를 보존하지 않지만 블록의 발행이 임의적이지 않았음을 보인다. 시위는 정책을 즉시 생산하지 않지만 그 주장이 단순한 온라인 반응보다 무겁다는 사실을 드러낸다. 두 번째 비용은 대상 세계를 직접 증가시키지 않는다. 대신 사회가 어떤 입력을 얼마나 진지하게 읽어야 하는지를 결정하는 메타정보를 만든다.

이때 이중비용문제의 역설이 발생한다. 사회는 효율을 원한다. 그러나 사회적 입력이 너무 효율적으로 만들어지면 그것은 곧 값싼 말이 된다. 아무 비용 없이 선언할 수 있고, 아무 마찰 없이 동의할 수 있고, 아무 책임 없이 분노할 수 있다면, 사회는 그 입력의 강도와 진정성과 우선순위를 읽을 수 없다. 반대로 사회적 입력을 너무 비싸게 만들면, 그것은 배제와 낭비와 특권의 장치가 된다. 돈을 두 번째 비용으로 요구하면 부자가 유리하고, 시간을 요구하면 여유 있는 사람이 유리하며, 위험을 요구하면 취약한 사람은 침묵하게 되고, 고통을 요구하면 순교 경쟁이 발생한다.

따라서 문제는 “비용이 있었는가”가 아니다. 비용이 있었다는 사실만으로 사회적 정당성이 생기지는 않는다. 오래 기다렸다고 해서 그 사람이 더 옳은 것은 아니다. 큰 위험을 감수했다고 해서 그 주장이 참인 것은 아니다. 많은 전기를 태웠다고 해서 그 디지털 재화가 인간적으로 가치 있는 것은 아니다. 두 번째 비용은 진리의 증명이 아니라, 값싼 위조를 어렵게 만드는 가독성 장치이다. 그러므로 더 정확한 질문은 “얼마나 많은 비용이 들었는가”가 아니라, “그 비용이 어떤 사회적 입력을 읽히게 만들기에 적절한 종류의 캐리어인가”이다.

줄서기는 돈이 아니라 시간과 체력을 캐리어 비용으로 삼는다. 이 때문에 가격만으로 배분하는 방식보다 어떤 면에서는 더 보편적인 신호를 만든다. 그러나 시간과 체력도 평등한 자원이 아니다. 기다릴 수 있는 사람과 기다릴 수 없는 사람이 있다. 시위는 신체적 체류와 위험을 캐리어 비용으로 삼는다. 이것은 온라인 동의보다 높은 신호 충실도를 갖지만, 장소 접근성, 신체 조건, 직업적 위험, 사회적 노출 가능성에 따라 비용이 다르게 부과된다. 비트코인은 물리적 에너지를 캐리어 비용으로 삼는다. 이는 중앙 발행자의 임의성을 줄이지만, 에너지 소산과 외부성을 만든다. 표는 반대로 정치적 입력의 캐리어 비용을 가능한 한 낮추려 한다. 민주주의는 표를 비싸게 만들면 안 된다는 것을 안다. 그러나 바로 그 때문에 표는 개인의 강도, 절실함, 조건부 책임, 수정 가능성을 거의 읽지 못한다.

이렇게 보면 표, 돈, 줄서기, 비트코인, 시위는 서로 전혀 다른 제도처럼 보이지만, 같은 하부구조를 갖는다. 사회적 입력은 그 자체로 충분히 읽히지 않는다. 그것이 cheap talk가 되지 않으려면

어떤 캐리어 비용이 붙어야 한다. 차이는 어떤 비용을 캐리어로 요구하는가, 그 비용이 누구에게 얼마나 다르게 부과되는가, 그 비용이 실제로 이전되는가 아니면 소산되는가, 그리고 그 소산이 정당한 가독성 생성인가 단순한 낭비인가에 있다.

이중비용문제에서 가장 중요한 구분은 대상 비용과 캐리어 비용이다. 대상 비용은 실제로 무언가를 얻거나 만들거나 이전하기 위해 쓰인다. 캐리어 비용은 그 행위가 사회적으로 읽히기 위해 쓰인다. 대상 비용은 대개 누군가에게 이전되거나 산출물로 남는다. 캐리어 비용은 대개 사라진다. 그러나 그 사라짐이 사회적 의미를 남긴다. 사회는 종종 소산된 자원을 통해 발화의 무게를 읽는다. 무엇이 이전되었는가보다, 무엇이 회수 불가능하게 소모되었는가가 더 강한 메타정보가 될 때가 있다.

그래서 이중비용문제는 단순한 낭비의 문제가 아니다. 오히려 사회가 신호를 믿기 위해 왜 낭비처럼 보이는 소산을 필요로 하는가의 문제이다. 출서기, 의례, 시위, 작업증명, 학위, 관료적 절차, 평판 형성은 모두 어느 정도 이 구조를 갖는다. 그것들은 대상 자체를 생산하는 비용과, 그 대상이나 발화가 사회적으로 읽히기 위해 소모되는 캐리어 비용을 분리한다. 두 번째 비용이 없으면 입력은 너무 싸지고, 두 번째 비용이 과하면 사회는 불필요하게 자원을 태운다.

그러므로 이중비용문제의 핵심은 다음과 같다. 사회적 입력은 공짜이면 위조되고, 너무 비싸면 배제된다. 두 번째 비용은 진리를 증명하지 않지만, 값싼 기호와 진지한 입력을 구분하게 만든다. 그러나 그 비용은 대부분 이전되지 않고 소산되므로, 사회 전체로는 낭비와 가독성 사이의 긴장이 발생한다. 어떤 사회가 무엇을 중요한 입력으로 읽는지는, 결국 그 사회가 어떤 캐리어 비용을 요구하는지에 의해 결정된다.

이중비용문제는 따라서 “왜 사회는 쓸모없어 보이는 비용을 요구하는가”라는 질문이 아니다. 더 정확히는 “왜 어떤 메시지는 비용 있는 캐리어 위에 실려야만 사회적으로 읽히는가”라는 질문이다. 사회적 메시지는 내용만으로 전달되지 않는다. 그것은 어떤 시간, 몸, 에너지, 위험, 기다림, 책임, 반복, 수정 가능성 위에 실릴 때 비로소 값싼 말과 구분된다. 첫 번째 비용은 대상을 만들고, 두 번째 비용은 그 대상이나 발화가 사회적으로 읽힐 수 있는 캐리어를 만든다. 이 두 비용의 분리와 긴장이 바로 이중비용문제이다.

4.3 소산된 비용이 감정으로 읽힐 때

이중비용문제에서 두 번째 비용은 단순한 숫자로만 남지 않는다. 현실에서 사람들은 어떤 행위의 가독성 비용을 표나 계량표로 먼저 읽지 않는다. 사람들은 광장에 모인 인원의 기회비용을 계산하지 않고도 그 사안의 무게를 느낀다. 오래 선 줄을 보고 그 재화가 희소하다는 사실을 즉각 이해한다. 장례식의 침묵 속에서 관계의 밀도를 느끼고, 시위 현장의 피로한 얼굴들 속에서 주장 이상의 무언가를 읽는다. 현실의 메타정보는 논리적 자료로만 전달되는 것이 아니라, 분위기, 긴장, 전율, 수치심, 경외, 책임감 같은 감정의 형태로 인간에게 입력된다.

이 감정은 비합리적 잔여물이 아니다. 오히려 너무 많은 맥락 정보가 한 번에 압축되어 들어오는 형식에 가깝다. 누가 왔는지, 누가 남았는지, 그 자리에 오기 위해 무엇을 포기했는지, 날씨가 어땠는지, 사람들의 표정이 어땠는지, 말의 떨림이 어땠는지, 그 공간이 얼마나 불편했는지, 그럼에도 사람들이 왜 떠나지 않았는지. 인간은 이 모든 값을 분리해서 계산하지 않는다. 대신 “이건 가볍지 않다”는 느낌으로 받는다. 현실은 사회적 입력의 캐리어 비용을 감정으로 렌더링한다.

그래서 어떤 현장은 논리보다 빠르게 사람을 설득한다. 그것은 현장이 논리를 초월하기 때문이 아니라, 논리로 분해되기 전의 더 많은 메타정보를 한꺼번에 제공하기 때문이다. 사람은 숫자로만

공적 무게를 판단하지 않는다. 몇 명이 모였는가도 중요하지만, 그들이 어떻게 모였는가, 무엇을 감수했는가, 얼마나 오래 남았는가, 어떤 목소리로 말했는가, 어떤 침묵을 공유했는가 함께 들어온다. 이때 사회적 가독성은 차가운 지표가 아니라 감정적 확신의 형태를 취한다.

이 지점에서 기술의 문제가 다시 나타난다. 기술은 많은 경우 인간을 자유롭게 만들기 위해 자연의 제한을 제거해왔다. 더 빨리 이동하게 하고, 더 쉽게 연결하게 하고, 더 많은 선택지를 제공하고, 더 적은 마찰로 반응하게 했다. 그러나 자연의 제한은 단순한 장애물만은 아니었다. 거리, 기다림, 신체적 피로, 시간의 느낌, 물리적 장소, 타인의 시선, 우연한 마주침은 모두 인간의 행위를 두껍게 만들던 조건이기도 했다. 그것들은 불편했지만, 동시에 말과 선택에 무게를 붙였다.

모든 제한을 제거하면 자유가 커지는 것처럼 보인다. 그러나 제한이 제거된 자리에서 사회적 신호는 얇아질 수 있다. 클릭은 자유롭지만 너무 싸다. 메시지는 빠르지만 너무 가볍다. 연결은 즉시적이지만, 그 연결이 무엇을 감수했는지는 보이지 않는다. 자연세계는 인간에게 무한한 자유를 주지 않았다. 몸의 속도, 피로, 낮과 밤, 거리, 날씨, 공간의 크기, 타인과의 물리적 공존이 행위의 문법을 제한했다. 그리고 그 제한 때문에 인간은 타인의 시간과 진지함과 절실함을 읽을 수 있었다.

따라서 다음 기술의 문제는 단순히 마찰을 제거하는 데 있지 않다. 어떤 마찰은 제거되어야 한다. 불필요한 행정 절차, 계급적 장벽, 불투명한 규칙, 정보 접근의 불평등은 줄어들어야 한다. 그러나 어떤 마찰은 인간적 의미와 사회적 가독성을 보존하기 위해 다시 설계되어야 한다. 중요한 것은 불편함을 숭배하는 것이 아니다. 중요한 것은 어떤 제한이 단순한 장애물이고, 어떤 제한이 사회적 신호의 캐리어였는지를 구별하는 일이다.

좋은 기술은 인간에게 가능한 모든 것을 즉시 허용하는 기술이 아닐 수 있다. 오히려 어떤 행위가 너무 싸게 생성되어 사회적 의미를 잃는 순간, 필요한 만큼의 시간, 주의, 반복, 책임, 반박, 체류를 다시 요구하는 기술일 수 있다. 메시지를 빠르게 복제하는 능력만으로는 충분하지 않다. 사회가 읽을 수 있는 메시지는 적절한 캐리어를 필요로 한다. 그리고 그 캐리어는 반드시 고통스럽거나 낭비적일 필요는 없다. 그것은 기다림일 수도 있고, 숙고일 수도 있고, 반박을 통과하는 절차일 수도 있고, 책임 조건을 명시하는 행위일 수도 있으며, 어떤 장소와 시간에 실제로 머무는 일일 수도 있다.

이때 기술은 자연을 흉내 내는 장식이 아니라, 자연이 제공하던 사회적 입력 조건을 다시 선택적으로 구성하는 장치가 된다. 현실의 제한은 우연히 존재했지만, 다음의 제한은 설계되어야 한다. 그리고 그 설계는 더 많은 자유를 막기 위한 것이 아니라, 자유로운 발화와 선택이 다시 사회적으로 읽힐 수 있게 하기 위한 것이다. 인간은 무한한 선택지 앞에서만 자신이 되는 것이 아니라, 자신이 선택한 것에 필요한 비용과 책임과 시간을 통과할 때 비로소 어떤 주체로 읽힌다.

그러므로 이중비용문제 이후에 남는 질문은 단순히 “두 번째 비용을 줄일 것인가, 늘릴 것인가”가 아니다. 더 중요한 질문은 “어떤 사회적 메시지에 어떤 캐리어가 필요한가”이다. 어떤 말은 즉시 말해져도 충분하고, 어떤 말은 시간을 통과해야 하며, 어떤 선언은 반박을 통과해야 하고, 어떤 결속은 책임 조건을 가져야 한다. 어떤 사회가 더 성숙한 사회인지는 모든 마찰을 제거했는가 아니라, 어떤 마찰을 제거하고 어떤 마찰을 남기며 어떤 마찰을 새롭게 설계하는가에서 드러난다.

5 제3프로토콜의 내부 질문: 구배선언은 어떻게 조정되는가

제3프로토콜의 필요가 확인되면 곧바로 다음 질문이 발생한다. 구배선언을 사회적 입력으로 받는다고 해서, 그 선언들이 자동으로 정합적 질서를 이루는 것은 아니다. 의지구배는 서로 충돌하고, 서로를 오해하며, 때로는 자기 자신조차 불명료하게 이해한다. 따라서 제3프로토콜은 단순한 게시판이나 집계 장치일 수 없고, 선언된 의지구배들이 어떤 조건에서 더 정의롭게 조정되는지 판단할

규범 좌표계를 필요로 한다.

이 지점에서 일반공리주의(Generalized Utilitarianism, GU)가 제안된다. GU는 제3프로토콜 내부에서 구배선언의 충돌을 해석하고 조정하기 위한 규범 엔진이다. 다시 말해 제3프로토콜이 구배선언을 사회적 입력으로 받는 사회론이라면, GU는 그 입력들이 충돌할 때 무엇을 더 나은 조정으로 볼 것인가를 묻는 규범론으로서 기능한다.

6 구배선언에서 GU로

자아를 과거의 실체가 아니라 미래를 향한 승인 구조로 이해한다면, 윤리와 제도 역시 그 선언의 구조를 중심으로 다시 정렬되어야 한다. 기존 체계는 대체로 이미 드러난 선호를 집계하거나 이미 성립한 거래를 조정하는 데 능숙했지만, 어떤 가치장이 형성되고, 어떤 의지구배가 공적으로 승인되며, 그 선언된 방향량들이 서로 조정 가능한가를 판단의 중심에 두지 않았다. 그러나 자아가 구배선언을 통해 사회적으로 기입된다면, 사회적 판단 역시 그 구배선언들의 관계를 중심으로 설계되어야 한다.

이 지점에서 일반공리주의(GU)는 단순한 공리주의의 변형이 아니라, 구배선언을 사회적 판단의 상위 입력으로 복귀시키려는 시도로 이해될 수 있다. 여기서 중요한 것은 인간에게 하나의 보편 목적을 부과하는 것이 아니다. 오히려 각 개인과 집단이 선언한 의지구배들의 다원성을 전제로 하면서도, 그 방향량들이 서로 어떤 구조를 이루며 어디에서 충돌하고 어디에서 수렴할 수 있는지를 묻는 새로운 좌표계를 세우는 일이다. GU는 바로 이 좌표계를 향한 첫 번째 철학적 제안이다.

7 일반공리주의(GU)의 제안

자아의 본질이 “발견되는 실체”가 아니라 가치장을 정련하고 의지구배를 승인하는 구조라면, 우리에게 필요한 사회 시스템 또한 근본적으로 달라져야 한다. 더 이상 맹목적인 이윤이나 효율이 기준이 될 수 없다. 개인이 선언한 의지구배를 실제적인 사회적 입력으로 삼아 기존 학문 분과를 재배열하고, 니체가 예외적 개인에게만 맡겨두었던 “산과 산을 건너는” 기능을 사회 전체의 기본 구조로 끌어올리는 새로운 좌표계가 필요하다. 우리는 이것을 일반공리주의(Generalized Utilitarianism, GU)라 부른다.

일반공리주의(GU)는 이러한 조건에서 “정의”를 다시 묻고, 고전적 공리주의를 다음 세 가지 차원에서 일반화한다.

1. 효용을 단순한 쾌락·고통의 합이 아니라, 지식 상태, 관계 구조, 공감 가능성을 포함하는 고차원 함수로 확장한다.
2. 판단의 1차 기준을 가능한 선택지들 중 “무엇이 더 효율적인가”가 아니라, 각 개인이 선언한 의지구배들 간의 조정과 수렴으로 옮긴다.
3. 정의를 이미 주어진 상태가 아니라, 가능한 미래 궤적들의 기대효용이 수렴해 가는 방향, 즉 “정의에 대한 근사(approximation to justice)”로 이해한다.

7.1 핵심 직관: 무지와 목적

GU를 지탱하는 핵심 직관은 다음과 같다.

첫째, “회피 가능한 모든 고통은, 충분히 인과를 거슬러 올라가면 궁극적으로 무지(ignorance)의 결과로 수렴한다”는 테제이다. 만약 전지(omniscience)가 있었다면, 구조 · 제도 · 관계를 고통스럽게 설계하지 않았을 것이라는 강한 가정이다. 이는 “고통의 일부는 세계 구조의 필연”이라는 체념적 공리에 대한 반론이자, 고통을 세계의 본질이 아니라 정보의 결핍에 귀속시키려는 시도이다.

둘째, “판단의 1차 근거는 가능성(what is possible)이 아니라 정련된 가치장에서 발생한 의지구배(what ought to be pursued)이어야 한다”는 테제이다. 가능성은 수단을 정교화하는 정보이고, 의지구배는 판단을 가르는 방향량이다. 목적의 다원성은 사실이자 규범이며, “인간의 존재 목적”을 하나의 보편 명제로 묻는 기획은 범주 오류에 가깝다. 목적은 발견되는 본질이 아니라, 각 개인 · 집단이 자기 가치장을 정련하고 사회적으로 선언하는 구조로 취급되어야 한다.

자유의지에 관해서 GU는 물리적 실재 여부에 대한 형이상학적 또는 물리학적 논쟁을 정지하고, 규범적 공리로서의 자유의지만을 채택한다. 즉, 우리는 개인을 사회적 의사결정의 원자적 단위로 전제하는 세계 안에서 살아야 한다. 이는 논증으로 “입증”해야 할 명제가 아니라, 인간이 자연 현상이나 알고리즘과 구분되지 않은 순수 수단으로 전략하는 것을 막기 위한 최소 방어선이다. 자유의지를 공리로 채택하지 않으면 “누가 책임지는가, 누가 목적을 세우는가”라는 질문 자체가 무의미해지고, 목적은 사라지며 인간은 전면적으로 도구화된다.

자유의지를 공리로 채택한다는 것은 자아가 순수한 인과 연쇄로 완전히 환원되지 않는다는 것을 인정하는 셈이다. 이때 자아는 외부 인과가 아니라 자기 내부의 의지로부터 최소 한 번은 자신을 규정해야 한다. 그러나 이 기원적 의지는 곧바로 완성된 목적문이 되지 않는다. 그것은 먼저 세계를 가치장으로 열고, 그 안의 상충하는 힘들을 정련하며, 현재 위치에서 어떤 방향으로 힘을 발생시킬 것인가를 의지구배로 산출한다. 따라서 여기서 말하는 “목적”은 단순한 욕구 목록이 아니라, 자아가 스스로를 기원시키기 위해 정련하고 승인한 최상위 방향량이다.

이 관점에서 개인의 구배선언은 개인의 헌법적 승인으로 재정의된다. 국가의 헌법이 그 정체성과 경계를 규정하듯, 최상위 구배선언은 자아가 자신의 시간과 능력과 책임을 어디에 결속할 것인지 규정한다. 이 최상위 선언이 임계적으로 변경되면, 국가가 “제 n공화국”으로 불리듯 자아도 “제 n-자기”로 재기원한다.

다만 GU는 절대적, 외재적 심판자가 아니다. 한 사회 내에서 낙태, 안락사, 혁명과 같은 논쟁적 사안에 대해 “GU에 따르면 답은 X다”라고 공개적으로 선언하는 행위 자체가 새로운 의도적 사건 π 이며, 그 발화가 다시 미래 분포와 효용에 영향을 미친다. 즉, 계 안에 있는 주체는 계 전체에 대한 완전한 판정을 구조적으로 내릴 수 없다. 이는 형식 체계 내부에서 그 체계의 모든 참을 산출할 수 없다는 괴델의 불완전성 정리를 사회적 의사결정 차원에서 반복하는 셈이다.

GU는 바로 이 한계를 정면에서 인정하며, 일종의 “모른 척하기 게임(game of willed ignorance)”을 도출한다. 어떤 물음에 대해 계산상 한쪽이 더 유리해 보이더라도, 그 답을 “언제, 누구에게, 어떤 형식으로 말할 것인가” 자체가 또 다른 결정 변수가 되며, 어떤 경우에는 “알아도 말하지 않는 것”이 정의 근사에 더 가깝다는 역설이 발생한다. GU는 이 자기참조성을 숨기지 않고, “진실을 말하지 않는 것의 윤리”를 이론 내부의 요소로 포함한다.

7.2 철학에서 시스템으로

철학은 아름답지만, 시스템은 작동해야 한다. 우리는 이 거대한 비전을 막연한 희망 사항으로 남겨두지 않기 위해, 윤리 판단을 효용의 네트워크 상에서 계산 가능한 수식으로 공식화했다.

이것은 현실의 복잡한 도덕적 결정을 평가하고 근사(Approximate)할 수 있는 이상적 기준, 마

치 물리학에서의 “마찰 없는 평면”과 같은 좌표계를 제공하려는 시도이다. 다만 여기서 제시하는 수식들은 완결된 방법론이나 닫힌 계산 체계가 아니다. 그것들은 윤리 판단과 목적 조정의 방향성을 어떤 식으로 계산 가능한 언어로 옮길 수 있을지를 탐색하기 위한 예비적 제안이며, 이후 상당한 개발과 확장의 여지를 의도적으로 남겨 둔다.

우리는 이상적인 정의로운 상태를 모든 주체의 효용이 다른 주체의 효용과 공감까지 포함하여 최적으로 해결된 상태로 상정한다. 이 정의로운 상태를 구현하는 가상의 전지적 판단자를 상정하면, 이 판단자는 각 개인의 주관적 행복과 그들이 다른 주체에 대해 갖는 공감 정보를 모두 취합하여 평가를 내릴 것이다. 우리는 이러한 이상적인 판단 과정을 “전지적 공감”이라는 개념으로 표현하며, 그 전제 조건으로서 공감정합도 $\mathcal{E}_{n \rightarrow i}^{\pi} = 1$ 을 정의한다.

결국 우리의 목표는, 이 전지적 공감 상태를 실제로 근사하는 원리를 찾는 것이다. 물론 인간은 전지적 상태에 도달할 수 없으므로, 우리는 현실의 제약 내에서 근접 정의 (Proximal Justice)를 실현하고자 한다. 이는 현재의 인지적 한계에서 달성 가능한 정의의 최선의 근사치로서, 전지적 공감 상태에 최대한 가깝도록 의사결정 구조를 조정하는 것을 의미한다.

본 원고의 나머지 부분은 다음과 같이 구성된다. 먼저 자아의 인정, 가치사상, 가치장, 가치정련, 의지구배, 구매선언으로 이어지는 내부 발생 구조에서 GU로 넘어가는 철학적 경로를 제시하고, 이어 전지적 공감의 형식화를 통해 이를 계산 가능한 언어로 옮긴다. 그 다음 CVTD를 실행층의 예비적 스케치로 제안하고, 마지막으로 왜 표와 가격을 넘어서는 제3프로토콜이 사회적으로 요구될 수밖에 없는지, 그리고 왜 읽기가 그 출현 조건을 담당하는 첫 번째 웨지가 되는지로 이어진다.

이후에 등장하는 여러 개념들은 병렬로 늘어선 별개의 이론이 아니라, 서로 다른 층위를 맡는 하나의 구조로 읽혀야 한다. 제3프로토콜은 중심 명제, 곧 구매선언을 표와 가격에 이은 세 번째 사회적 입력으로 받아야 한다는 사회론이다. GU는 다수의 구매선언이 충돌할 때 무엇을 더 정의로운 조정으로 볼 것인가를 다루는 규범 엔진이다. 계약적 가치이전역학(Contractual Value Transfer Dynamics, CVTD)은 개인의 의지구배와 의도적 사건 π 사이를 왕복하며, 선언된 방향량이 현실의 조건, 계약, 기여, 책임 구조로 내려올 때 그 가치 변화와 정렬을 기록하는 실행 문법이다.

삼기저동력모형(Tri-Basis Drive Model, TBDM)은 독립된 엔진이라기보다, 개인의 의지구배와 사회의 정의근사 방향량을 가능성, 사회성, 정합성의 세 기저 위에 사영하여 미시 사건들이 어떤 거시적 알짜힘으로 나타나는지를 설명하는 공통 언어다. The Channel은 이 전체 구조가 현실로 들어가기 위한 첫 진입점, 곧 가치정련과 구매선언을 가능하게 하는 인프라다. 따라서 아래의 논의는 닫힌 완성 체계라기보다, 인간론에서 사회론, 규범론, 실행론, 현실 웨지로 이어지는 하나의 설계도에 가깝다.

8 GU의 형식화로

앞선 절에서 GU는 자아의 구매선언 구조로부터 도출되는 사회적 판단 원리로 제안되었다. 이제 필요한 것은 이 제안을 단순한 철학적 선언으로 남겨두지 않고, 효용과 공감과 판단의 구조를 계산 가능한 형태로 옮기는 일이다. 다시 말해, 선언된 의지구배의 다원성과 공감의 비대칭성과 정보의 불완전성을 전제로 하면서도, 정의가 어떤 방향으로 수렴하는지를 형식적으로 다룰 수 있어야 한다.

다음 절의 형식화는 바로 이 문제를 다룬다. 여기서 중요한 것은 수식이 철학을 대체한다는 뜻이 아니라, 철학이 제시한 좌표계를 연산 가능한 구조로 번역한다는 점이다. GU가 정의의 방향 성을 제시하는 거시적 프레임이라면, 그 정식화는 그 방향성이 실제 판단과 설계의 언어로 어떻게 변환될 수 있는지를 보여주는 중간 단계가 된다.

9 이론적 틀: 전지적 공감의 형식화

본 섹션에서는 일반공리주의의 수학적 모델을 구축하고, 도덕성을 효용과 공감의 네트워크 상에서 계산하는 방법을 설명한다. 이는 현실의 도덕적 결정을 평가하고 근사할 수 있는 이상적 기준을 제공하려는 시도로, 물리학에서의 “마찰 없는 평면”과 같은 개념으로 볼 수 있다.

9.1 주체, 내재 상태 및 효용 H_n

우리는 도덕적 주체들을 $n = 1, 2, \dots, N$ 으로 표기하고, 각 주체 n 은 그들의 복지와 관련된 모든 요소, 예를 들어 신체적 건강, 정신적 상태, 물질적 자원, 욕망 등을 포괄하는 내재 상태 I_n 을 가진다고 가정한다. 여기서 “주체”는 미리 결정된 생물학적 개체가 아니라, 자아의 비존재성 절에서 제시한 것과 같이 가치장, 복지, 의지, 공감 가능성이 귀속되는 것으로 인정되는 비존재적 좌표다. 이 정의는 자아 측 n 과 타자 측 i 양쪽에 동일하게 적용된다.

이 내재 상태를 바탕으로, 각 주체의 효용 또는 행복 H_n 은 자신에 대한 상태와 타자에 대한 가치적 반응을 모두 함수로 표현할 수 있다. 본 원고가 채택하는 일반형은 다음과 같다.

$$H_n = I_n + \sum_{i=1}^N c_{n,i} \cdot H_{n \rightarrow i}^{\pi}$$

여기서 $c_{n,i}$ 는 주체 n 이 주체 i 를 자신의 효용 구조에 얼마나 결속시키는지를 나타내는 공감 계수이고, $H_{n \rightarrow i}^{\pi}$ 는 의도적 사건 π 아래에서 n 이 i 를 가치장에 받아들인 값이다. 즉 $H_{n \rightarrow i}^{\pi}$ 는 n 의 가치장 안에서 i 가 차지하는 가치적 좌표이며, i 가 독립적인 복지·가치 상태를 갖는 것으로 인정될 때에 한해 i 자신의 효용 H_i 를 매개해 분해된다.

분해 가능한 주체의 집합을 $\mathcal{D}_t \subseteq \{1, \dots, N\}$ 이라 하자. 이 집합은 시점과 의도적 사건 π 에 따라 변동하며, 어떤 사회가 어떤 주체를 독립적 가치장의 보유자로 인정하는가의 사회적 사실에 의해 형성된다. 그러면 $\mathcal{D}_t$ 안의 주체에 대해 다음 분해가 성립한다.

$$i \in \mathcal{D}_t \implies H_{n \rightarrow i}^{\pi} = \mathcal{E}_{n \rightarrow i}^{\pi} \cdot H_i$$

여기서 $\mathcal{E}_{n \rightarrow i}^{\pi}$ 는 다음 절에서 다룰 공감정합도이며, 이 분해가 회복될 때 종전 GU의 자기참조 구조 $H_n = I_n + \sum_i c_{n,i} \mathcal{E}_{n \rightarrow i}^{\pi} H_i$ 가 그대로 살아난다.

$\mathcal{D}_t$ 밖의 주체에 대해서는 $H_{n \rightarrow i}^{\pi}$ 가 H_i 를 매개하지 않고 n 의 가치장 안에서 직접 결정된다. 이 외생적 귀속값을 $\tilde{H}_{n \rightarrow i}^{\pi}$ 로 쓴다. 인정의 신뢰도가 불확실한 경우에는 둘 사이의 부분 결합으로 다음처럼 둘 수 있다.

$$H_{n \rightarrow i}^{\pi} = q_i \mathcal{E}_{n \rightarrow i}^{\pi} H_i + (1 - q_i) \tilde{H}_{n \rightarrow i}^{\pi}, \quad q_i \in [0, 1]$$

여기서 q_i 는 n 의 사적 신뢰가 아니라, 제3프로토콜이 운영하는 공적 신뢰 절차의 출력으로 해석된다. 즉 다른 주체들의 귀속, 행동 신호, 반박 절차, 시간 안정성을 통과한 뒤의 사회적 신뢰도이며, 컬트적 사적 귀속이 임의로 큰 가중치를 갖지 못하도록 막는 장치다. $q_i = 1$ 은 완전 분해 가능, $q_i = 0$ 은 완전 외생적 귀속을 뜻한다.

따라서 효용식은 다음처럼 분해 가능 부분과 외생적 귀속 부분으로 명시 분리된다.

$$H_n = I_n + \underbrace{\sum_{i \in \mathcal{D}_t} c_{n,i} \mathcal{E}_{n \rightarrow i}^\pi H_i}_{\text{재귀항}} + \underbrace{\sum_{j \notin \mathcal{D}_t} c_{n,j} \tilde{H}_{n \rightarrow j}^\pi}_{\text{외생적 귀속항}}$$

이 구조 덕분에 GU는 동물, 태아, 미래의 나, 돌멩이, 문화, 종교, 풍경을 같은 문법 안에 둘 수 있다. 골든 리트리버처럼 대체로 독립적 복지 상태가 인정되는 주체는 재귀항에 들어가고, 도마뱀이나 사마귀처럼 인정의 신뢰도가 사회마다 다른 주체는 q_i 로 부분 결합되며, 유품·성물·풍경처럼 그 자체로 복지를 갖지 않는 주체는 외생적 귀속항으로 들어간다. 핵심은 1차 입력이 H_i 가 아니라 $H_{n \rightarrow i}^\pi$ 라는 점이다. H_i 는 그 주체가 독립적 가치장을 갖는 것으로 인정될 때 $H_{n \rightarrow i}^\pi$ 안에서 분해되는 항이며, 이 분해 자체는 사회적 인정의 사실에 의존한다.

만약 $c_{n,i} = 0$ 이라면 주체 n 은 i 의 가치적 좌표에 전혀 결속되지 않음을 의미하며, 양의 값이라면 n 은 i 의 가치적 좌표 변화에 따라 일정한 효용 또는 불효용을 함께 느낀다는 것을 의미한다. 이렇게 각 주체의 효용 H_n 은 단지 자기 자신의 상태 I_n 뿐만 아니라 타자에 대한 가치적 반영을 반영하여, 공감, 연대, 혹은 도덕적 분노와 같은 감정들을 포괄할 수 있다.

이 식은 한 주체의 행복이 다른 주체들에 대한 가치적 귀속에 의존함을 보여준다. 특히 재귀항 안에서는 H_i 가 다시 다른 주체들의 영향으로 결정되므로, 분해 가능 부분그래프 $\mathcal{D}_t$ 위에서 전체 $\mathbf{H} = (H_1, H_2, \dots, H_N)$ 구조는 상호 참조적인 성격을 갖는다. 이러한 상호 참조 구조는 뒤에서 고정점 해석을 통해 더욱 명확히 다루게 될 것이다.

9.2 공감정합도 (Empathic Coherence)

실제 인간은 완전한 합리성이나 무한한 정보를 갖추고 있지 않다. 또한 인간은 타자의 상태를 받아들이는 때, 단순히 “모르거나 무시”하는 것뿐 아니라, 종종 왜곡되거나 반대로 해석된 상태에서 인지적 평가를 내리기도 한다. 이를 반영하기 위해 일반공리주의(GU)에서는 각 주체 n 이 의도적 사건 π 하에서 타자 i 를 자기 가치장에 얼마나 정합되게 귀속시키는지 나타내는 지표로 공감정합도(Empathic Coherence)를 도입한다. 수식에서는 이를 $\mathcal{E}_{n \rightarrow i}^\pi$ 로 표기하며, 다음과 같은 범위를 갖는다:

$$\mathcal{E}_{n \rightarrow i}^\pi \in [-1, 1] .$$

여기서 공감정합도는 “타자의 진짜 효용을 정확히 반영하는 정도”에 국한되지 않는다. 본 원고에서 채택하는 일반적 의미는 귀속의 정합성이다. 즉 $\mathcal{E}_{n \rightarrow i}^\pi$ 는 n 이 i 에 대해 형성한 가치적 좌표 $H_{n \rightarrow i}^\pi$ 가 다음 세 측면과 얼마나 정합하는지를 측정한다.

- (a) i 측에서 반박 · 수정 가능한 신호(행동, 자기보고, 사회적 표현 등)와의 정합
- (b) 다른 주체 n' 들의 같은 i 에 대한 귀속과의 상호참조 가능성
- (c) n 자신의 다른 시점 · 맥락에서의 귀속과의 자기정합성

이 세 정합성 축은 분해 가능한 주체 $i \in \mathcal{D}_t$ 에 대해서는 종전 의미와 자동적으로 일치한다. i 가 독립적 효용 H_i 를 갖는 것으로 인정되면, (a)는 H_i 자체와의 정합으로 환원되기 때문이다. 반면 $\mathcal{D}_t$ 밖의 주체에 대해서는 (a)가 i 자체의 효용 항이 아니라 i 를 둘러싼 관측 가능한 신호와의 정합으로 작동한다.

다음의 세 극값 해석은 양쪽 경우 모두에서 보존된다.

- +1: 주체 i 를 완전히 정합적으로 귀속함 (이상적 명료성)
- 0: i 에 대한 정보 없음 또는 몰이해 상태 (무지)
- -1: 주체 i 를 완전히 왜곡하거나 반대로 귀속함 (오해 혹은 적대적 해석)

이 지수는 단순한 이해 유무를 넘어, 그 이해의 방향성과 정합성까지 반영하며, 감정적 또는 인지적 왜곡을 수치화할 수 있게 한다. 예를 들어, 타자에 대한 편견, 적개심, 증오 등은 $\mathcal{E}_{n \rightarrow i}^\pi < 0$ 으로 표현될 수 있으며, 이는 전체 정의 판단에서 i 의 가치적 좌표가 반대로 반영되는 효과를 가져온다. 즉, n 이 i 를 혐오하는 경우 $c_{n,i}$ 가 양수이더라도 $\mathcal{E}_{n \rightarrow i}^\pi$ 가 음수라면, i 의 가치적 좌표 증가는 n 에게 오히려 불행으로 작용하게 된다.

공감정합도 개념은 인간 판단의 인지적 한계와 편향을 통합하기 위한 장치다. 범위 무감각, 내집단 편향, 현재 편향 등 실제 사람들의 여러 인지적 제한으로 인해, 동일한 상황에 대해서도 개인마다 서로 다른 도덕적 평가가 나타날 수 있다. 이러한 현실을 모델에 반영하기 위해 $\mathcal{E}_{n \rightarrow i}^\pi$ 는 각 주체의 인지적 · 감정적 정합도(coherence)를 나타낸다. 낮은 값은 비합리적 판단이나 편향을 의미하며, 높은 값은 보다 숙고되고 정합된 윤리 판단을 반영한다. 결과적으로 공감정합도는 편향으로 인한 왜곡을 구조적으로 보정하거나 드러내는 역할을 한다.

9.3 정의근사 (Proximal Justice)

일반공리주의(GU)에서 정의(Justice)는 단순한 정책 결과의 합산이 아니다. 정의는 주체 간의 가치적 귀속, 공감, 이해 구조가 의도적 사건 π — 누군가가 지금 행하거나 행하지 않는 결정, 계약, 발화, 돌봄, 거부 같은 사건 — 에 의해 정렬되어, 전체 효용 구조 $\mathbf{H} = (H_1, H_2, \dots, H_N)$ 가 정보론적 고정점에 수렴하는 상태로 정의된다. 여기서 π 는 단순한 결과의 명단이 아니라, 세계 상태 O_t 위에서 다음 상태 O_{t+1} 을 유도하는 변환으로 작용하는 의도적 사건이며, 그 내부 구조는 12장 CVTD에서 튜플 형태로 풀어 쓴다.

우리는 이러한 구조적 조정을 설명하기 위해 존재론적 자기 확장 (Ontological Self-Extension, OSE) 개념을 도입한다. 주체 n 이 주체 i 에 대해 갖는 존재론적 자기 확장 정도를 다음과 같이 정의한다:

$$OSE_{n,i} := c_{n,i} \cdot \mathcal{E}_{n \rightarrow i}^\pi ,$$

특히 자기 자신 ($i = n$)에 대해서는 $c_{n,n} = 1$ 이므로, 다음과 같이 간단히 쓸 수 있다:

$$OSE_{n,n} = \mathcal{E}_{n \rightarrow n}^\pi .$$

정의(Justice)는 모든 주체 n, i 쌍에 대해 $OSE_{n,i}$ 가 1에 가까워지는 극한에서 수렴한다. 다만 이 극한의 의미는 “타자의 실체 H_i 에 완전히 일치”가 아니라, 공감정합도 절에서 재정직한 의미에서 귀속의 정합성의 극한으로 읽는다. 즉 주체들이 서로를 가장 정합하게 귀속하는 이상적 극한 상태에서 정의가 실현된다고 볼 수 있다. 분해 가능 주체 $i \in \mathcal{D}_t$ 에 대해서는 이 극한이 종전 의미와 자동 일치한다. 이를 수식으로 나타내면:

$$\text{Justice} := \lim_{OSE_{n,i} \rightarrow 1} \pi^* ,$$

여기서 π^* 는 위 극한 상태로 OSE를 수렴시키는 이상적 의도적 사건(또는 그러한 사건의 결정 원리)을 가리킨다. 현실에서는 모든 관계에서 $OSE = 1$ 인 전지적 공감 상태에 도달할 수 없지만, π^* 는

이 이상을 향해 수렴하는 의도적 사건의 방향성을 나타낸다.

현실적 판단 구조: 근접 정의 π^* . 현실에서는 완전한 OSE 수렴이 불가능하므로, GU는 현실적 공감/이해 구조를 반영한 근사 최적화를 통해 정의의 판단을 수행한다. 이를 공식화하면 다음과 같다:

$$\pi^* = \arg \max_{\pi} \sum_{n=1}^N \mathcal{E}_{n \rightarrow n}^{\pi} \cdot H_n ,$$

즉, 각 주체 n 의 효용 H_n 에, 그 주체가 현재 자기 자신을 얼마나 명료하게 이해하는지를 나타내는 $\mathcal{E}_{n \rightarrow n}^{\pi}$ 을 가중치로 곱한 총합을 최대화하는 π 를 찾는 것이다. 여기서 π 는 단순한 결과의 집합이 아니라, 세계 상태 O 전체에 작용하여 다음 상태 O' 을 유도하는 의도적 사건이며, 그 작용을 통해 효용 구조 $\mathbf{H}$ 에 영향을 미친다.

각 주체의 총합적 효용 H_n 은 일반형 $H_n = I_n + \sum_i c_{n,i} H_{n \rightarrow i}^{\pi}$ 를 따른다. 이 구조는 효용 H_n 이 단순히 개인의 주관적 반응이 아니라, 공감 계수 $c_{n,i}$ 를 통해 다른 주체들의 가치적 좌표 $H_{n \rightarrow i}^{\pi}$ 를 자기 효용망에 결속시킨 값임을 보여준다. 특히 분해 가능 주체 $i \in \mathcal{D}_t$ 에 대해서는 $H_{n \rightarrow i}^{\pi}$ 가 다시 H_i 를 통해 다른 주체들의 가치적 좌표를 포함하므로, 분해 가능 부분그래프 위에서 전체 구조는 다음과 같은 자기참조 체계로 확장된다:

$$\begin{aligned} H_n = I_n + \sum_{j \notin \mathcal{D}_t} c_{n,j} \tilde{H}_{n \rightarrow j}^{\pi} \\ + \sum_{\substack{i \in \mathcal{D}_t \\ i \neq n}} c_{n,i} \cdot \mathcal{E}_{n \rightarrow i}^{\pi} \cdot \left(I_i + \sum_{\substack{j \in \mathcal{D}_t \\ j \neq i}} c_{i,j} \cdot \mathcal{E}_{i \rightarrow j}^{\pi} \cdot H_j + \dots \right) . \end{aligned}$$

이와 같은 자기참조 구조는 분해 가능 부분그래프 위에서 일련의 반복되는 내포를 보여주며, 이론적으로 무한히 계속될 수 있다.

따라서 이러한 시스템에서 정의로운 판단이 성립하기 위해서는, 전체 효용 벡터 $\mathbf{H}$ 가 안정된 값을 가져야 한다.

고정점 해석. 위 자기참조 구조는 고정점 형태로 수렴되어야 정의의 판단이 수학적으로 성립한다.

전체 효용 벡터 $\mathbf{H}$ 는 다음 고정점 조건을 만족해야 한다:

$$\mathbf{H} = F_{\pi}(\mathbf{H}) ,$$

여기서 F_{π} 는 의도적 사건 π 에 의해 유도되는 공감-이해-가치귀속 흐름의 전체 변환을 나타내는 연산자이다. 선형 근사에서는 이 구조를 분해 가능 부분그래프 $\mathcal{D}_t$ 위에서 다음처럼 쓸 수 있다.

$$\mathbf{H}_{\mathcal{D}_t} = \tilde{\mathbf{I}} + A_{\pi} \mathbf{H}_{\mathcal{D}_t}, \quad (A_{\pi})_{n,i} = c_{n,i} \mathcal{E}_{n \rightarrow i}^{\pi} \quad (i \in \mathcal{D}_t),$$

여기서 $\tilde{I}_n := I_n + \sum_{j \notin \mathcal{D}_t} c_{n,j} \tilde{H}_{n \rightarrow j}^{\pi}$ 는 외생적 귀속향을 흡수한 유효 내재 상태다. 이때 A_{π} 의 스펙트럼 반경이 1보다 작으면, 즉 $\rho(A_{\pi}) < 1$ 이면 고정점은 유일하게 존재하고 다음처럼 표현된다.

$$\mathbf{H}_{\mathcal{D}_t}^* = (I_{|\mathcal{D}_t|} - A_{\pi})^{-1} \tilde{\mathbf{I}}$$

여기서 $I_{|\mathcal{D}_t|}$ 은 항등행렬이고, $\tilde{\mathbf{I}}$ 는 외생적 귀속을 흡수한 유효 내재 상태 벡터다. 만약 $\rho(A_{\pi}) \geq 1$ 이라면 공감과 오해의 상호참조가 증폭되어 효용망이 발산하거나 여러 고정점으로 갈라질 수 있다.

이 경우 GU는 감쇠 계수, 정규화, 절단된 영향 반경, 또는 추가 제약을 통해 F_π 가 수축 사상에 가까워지도록 설계되어야 한다. 더 일반적으로는 F_π 가 완비 거리공간 위의 contraction이거나, 부분순서가 주어진 격자 위의 단조 연산자일 때 고정점 해석이 수학적으로 방어 가능하다. 만일 어떤 $\mathbf{H}_{\mathcal{D}_t}^*$ 가 이 식을 만족하는 고정점으로 존재하고, 실제 과정이 $\mathbf{H}_{\mathcal{D}_t}^*$ 에 수렴할 수 있다면, 정의는 안정된 형태로 존재한다고 말할 수 있다. 이때 π^* 는 $\mathbf{H}$ 가 $\mathbf{H}^*$ 로 수렴하는 경로의 방향성으로 해석된다. 다시 말해, π^* 는 H 공간에서 효용 분포의 변화 방향을 결정하는 연산자이며, 그 연산의 반복 결과가 정의로운 상태로 수렴한다. 비분해 주체 $j \notin \mathcal{D}_t$ 에 대한 외생적 귀속향은 이 고정점 해석에 상수처럼 흡수되며, 그 자체로 재귀를 만들지 않는다. 이는 비분해 주체에 대한 귀속이 사회적 인정 절차의 출력으로 다뤄져야 함을 다시 한 번 보여준다.

의도적 사건 π 의 본질. π 는 단순한 행위 선택이 아니라, 의지구배를 가진 주체가 세계 상태에 가하는 의도적 사건이며, 그 사건이 유도하는 변환을 통해 주체 기반 윤리 정보 구조 전체를 재구성한다. 형식적으로, π 는 세계의 존재 상태 O 에 작용하여 다음 상태 O' 을 유도하는 변환을 가진다.

$$\pi : O_t \rightarrow O_{t+1}$$

여기서 O_t 는 다음과 같은 전체 구조를 포함한다.

$$O_t = \{I_n(t), x_n(t), H_n(t), c_{n,i}(t), H_{n \rightarrow i}^\pi(t), \mathcal{E}_{n \rightarrow i}^\pi(t), \mathcal{D}_t\}_{n,i}$$

즉, 한 번의 의도적 사건 π 는 세상의 존재론적 구조 O 전체에 작용하여, 가치적 귀속, 공감, 이해, 그리고 어떤 주체가 분해 가능한 가치장 보유자로 인정되는가의 사회적 사실까지 재구성한다. π^* 는 이러한 의도적 사건들 중에서 정의의 수렴 방향성을 가장 잘 실현하는 이상적인 사건을 의미한다. 12장 CVTD에서는 동일한 π 를 튜플 $(A_\pi, X_\pi, \Phi_\pi, \tau_\pi, \Gamma_\pi, \mu_\pi)$ 로 풀어 쓰며, 여기서 τ_π 가 위 변환 $\pi : O_t \rightarrow O_{t+1}$ 에 해당한다.

9.4 GU의 수학적 해석 가능성: 비교, 중심성, 해 선택

여기서 한 가지 주의할 점이 있다. 아래의 수학적 개념들은 GU를 이미 완성된 정리 체계로 입증하기 위한 권위의 목록이 아니라, GU가 어떤 종류의 계산 언어로 발전할 수 있는지를 보여주는 해석 렌즈들이다. 특히 I_n 과 H_n 은 서로 다른 개인의 쾌락을 직접 같은 자로 재는 단순한 행복 점수가 아니다. 개인 간 의지구배의 충족과 정렬 정도는 원초적으로 직접 비교하기 어렵다. 같은 1 시간, 같은 1달러, 같은 칭찬도 어떤 사람에게는 생존의 조건이고, 어떤 사람에게는 거의 무의미한 잡음일 수 있다. 따라서 GU에서 I_n 은 각 개인의 구배선언, 신체적 조건, 관계 구조, 시간축, 위험 회피도, 자기이해 정도를 반영하여 그 개인의 가치 기저 안에서 정규화된 내재 상태로 보아야 한다.

H_n 은 그 내재 상태가 다른 주체들에 대한 가치적 귀속 $H_{n \rightarrow i}^\pi$, 공감 계수, 공감정합도와 연결되며 형성되는 네트워크적 값이다.

이 점에서 GU의 효용 벡터 $\mathbf{H}$ 는 단순 합산표라기보다 PageRank 또는 고유벡터 중심성(eigenvector centrality)에 가까운 구조를 가진다. 어떤 주체의 효용은 고립된 점수가 아니라, 다른 주체들과의 연결, 공감 가중치, 공감정합도, 자기이해의 반복적 상호참조 속에서 안정화된다. 다만 이 중심성 구조는 분해 가능 주체 $i \in \mathcal{D}_t$ 위에서만 자기참조 그래프를 이룬다. 비분해 주체에 대한 외생적 귀속향은 그래프 외부의 노드 가중치처럼 작용한다.

즉 중요한 것은 “누가 더 큰 주관적 행복을 말하는가”가 아니라, 전체 귀속망이 어떤 고정점으로 수렴하며, 그 고정점이 각자의 구매선언과 얼마나 정렬되어 있는가이다.

정의 판단은 이러한 고정점 위에서의 해 선택 문제로 볼 수 있다. 각 주체가 가능한 의도적 사건 π 에 대해 반응하고, 그 반응이 다시 다른 주체의 반응 조건을 바꾼다면, 사회적 상태는 게임이론적 균형 문제로 해석될 수 있다. 적절한 연속성, 볼록성, 콤팩트성 조건이 주어질 경우 Kakutani 고정점 정리는 Nash equilibrium의 존재를 논의하는 한 가지 수학적 배경이 된다. 그러나 GU는 단지 어떤 균형이 존재한다는 사실에 만족하지 않는다. 나쁜 균형도 존재할 수 있기 때문이다. 따라서 GU가 묻는 것은 “균형이 있는가”가 아니라 “어떤 균형을 정의 근사로 선택할 것인가”이다.

이 선택 문제는 Nash bargaining의 관점에서도 해석될 수 있다. 여러 구매선언이 동시에 만족될 수 있는 실행 가능 집합 F 와 협상 결렬점 d 가 있을 때, 하나의 단순한 해 선택 원리는 다음처럼 쓸 수 있다.

$$\pi^{NB} = \arg \max_{\pi \in F} \prod_{n=1}^N (H_n(\pi) - d_n)$$

이는 모든 문제를 곱셈식 협상해로 환원하자는 뜻이 아니다. 다만 제3프로토콜 안에서 구매선언들이 단순 다수결이나 지불 능력으로 압축되지 않고, 각자의 결렬점과 개선폭을 고려하는 공리적 해 선택의 언어로 다루어질 수 있음을 보여준다.

또 다른 관점에서, GU의 고정점은 Tarski 고정점 정리의 최소 고정점(least fixed point)을 떠올리게 한다. 만약 의도적 사건 π 에 의해 유도되는 변환 F_π 가 부분순서가 주어진 격자 위에서 단조적이라면, 우리는 가능한 고정점들 중 가장 작은 가정과 가장 적은 강제를 요구하는 최소 고정점을 먼저 고려할 수 있다. 이는 제3프로토콜의 설계 원칙과 맞닿아 있다. 구매 조정은 가능한 한 많은 방향량을 강제로 지우는 방향이 아니라, 가장 적은 억압으로도 안정된 정렬을 만드는 방향이어야 한다. 제약이 존재하는 현실에서는 Lagrange multiplier의 해석도 유용하다. 자원, 시간, 안전, 신뢰, 정보, 법적 권리, 양심의 한계가 제약식 $g_k(\pi) \leq 0$ 로 주어질 때, 구매 조정 문제는 다음과 같은 형태를 가질 수 있다.

$$\mathcal{L}(\pi, \lambda) = \Psi(\mathbf{H}(\pi)) - \sum_k \lambda_k g_k(\pi)$$

여기서 λ_k 는 단순한 계산 보조항이 아니라 shadow price, 곧 어떤 제약이 사회적 의지구배의 정렬에 얼마나 큰 그림자를 드리우는지를 나타낸다. 예컨대 한 사회에서 신뢰 부족의 λ 가 매우 크다면, 더 많은 돈보다 신뢰 인프라를 구축하는 것이 더 큰 구매 개선을 낳을 수 있다.

마지막으로, 구매선언은 common knowledge의 문제이기도 하다. 어떤 개인이 속으로 무엇을 원한다는 사실과, 그가 그것을 공적으로 승인했고, 다른 사람들이 그 선언을 알고 있으며, 그가 다른 사람들이 알고 있음을 안다는 사실은 완전히 다르다. 제3프로토콜의 구매선언은 바로 이 믿음의 고정점을 만든다. 의지구배는 내면의 기율기로서 남아 있을 때 사회적 입력이 되지 못하지만, 공적으로 승인되고 반복적으로 참조될 때 계약, 협상, 조직, 대리 지능의 행동 조건이 된다.

이때 가장 어려운 반론은 악한 목적의 문제다. 누군가가 “나는 악인이다. 누가 나를 멈출 것인가”라고 선언한다면, 구매선언을 존중한다는 원리는 그 선언을 그대로 허가해야 하는가? 답은 아니다. GU에서 구매선언은 사회적 입력이지 면책권이 아니다. 선언은 숨은 욕구를 보이지 않게 실행하는 대신, 그 방향량이 타인의 가능성, 사회성, 정합성에 어떤 피해를 주는지 계산 가능하고 반박 가능하게 만드는 행위다. 악한 목적은 억압의 바깥에서 신비롭게 처리되는 것이 아니라, 제3프로토콜

안에서 드러나고, GU 안에서 평가되며, CVTD 안에서 그 의도적 사건의 피해 경로가 추적된다. 그러므로 이 절의 여러 수학적 참조는 GU의 완성을 선언하는 것이 아니라, 구배선언의 조정 문제가 어떤 형식적 도구들과 연결될 수 있는지를 표시하는 지도에 가깝다.

9.5 시간성과 정의 판단의 시점

GU 이론에서 시간 t 은 온톨로지 벡터공간 O 의 또 하나의 축으로 포함된다. 전체 윤리적 구조는 다음과 같이 표현될 수 있다.

$$O_t = O \times \mathbb{R}$$

즉, 시간에 따라 O 의 상태가 변화한다고 본다. 주체 n 은 시점 t_0 에서의 내재 상태 $I_n(t_0)$ 및 총합적 효용 $H_n(t_0)$ 로 정의되며, 정의 판단은 항상 현재 시점 t_0 의 판단자 n 에 의해 수행된다. 이때 과거 및 미래의 주체, 즉 과거의 나 또는 미래의 나 역시 타자 i 의 하나로 간주된다. 과거의 나 는 흔히 분해 가능 영역 $\mathcal{D}_t$ 에 머무르지만, 아직 도래하지 않은 미래의 나 는 부분 분해 또는 비분해 주체로 다뤄질 수 있다. 이는 동일한 한 인간 안에서도 시간축에 따라 분해 가능성이 달라질 수 있다는 점을 보여준다. 과거 세대나 미래 세대에 대한 공감 계수 $c_{n,i}$ 와 공감정합도 $\mathcal{E}_{n \rightarrow i}^\pi(t_0)$ 가 낮을 경우, 현재 n 의 총합적 효용 $H_n(t_0)$ 에 미치는 영향력은 자연스럽게 축소된다. 이는 시간적 거리가 먼 주체일수록 현재 판단에 덜 고려될 수 있음을 의미하지만, GU 구조 내에서는 명시적으로 그 영향을 모델링하고 있다.

결과적으로, 현실에서의 정의 판단은 다음과 같은 구조를 가진다:

$$\pi^*(t_0) = \arg \max_{\pi} \sum_{n=1}^N \mathcal{E}_{n \rightarrow n}^\pi(t_0) \cdot H_n(t_0) ,$$

이는 정의가 시간에 종속된 절대적인 값이라기보다는, 현재의 정보 구조 위에서 행해지는 윤리적 정렬임을 의미한다. 현재 시점의 각 주체들은 자신과 타자의 현재 및 예상되는 미래 효용, 그리고 그들 간의 가치적 귀속 정도를 고려하여 $\pi^*(t_0)$ 를 산출한다. 그 과정에서 미래 세대나 과거 세대에 대한 고려도 c 와 공감정합도 값에 의해 연속적으로 반영되므로, 시간적 거리에서 오는 도덕적 책임의 약화를 방지하려는 구조적 특징을 갖는다.

철학적 정리.

이상의 논의를 요약하면 다음과 같이 표현할 수 있다:

$$\text{Justice} := \text{Omniscient Empathy} \succ \text{Proximal Justice} = \pi^* \xrightarrow{OSE_{n,i} \rightarrow 1, \forall n,i} \text{Justice} ,$$

즉, 현실의 정의는 항상 전지적 공감(Omniscient Empathy)의 방향으로 수렴하며, π^* 는 그 수렴을 현실에서 구현한 근사적 정의(Proximal Justice)에 해당한다. 고도화된 AI 시스템은 이러한 수렴을 보조하는 윤리적 메타 연산자로 작동해야 한다.

10 구배선언에서 의도적 사건으로

GU가 구배선언의 충동을 판단하기 위한 규범 좌표계라면, 아직 남는 문제는 실행이다. 사회는 의 지구배를 이해하는 데서 멈추지 않고, 그것을 선언, 기여, 설계, 설득, 조직 형성, 자원 이동, 계약, 생산, 거부와 조정 같은 의도적 사건으로 되돌려 보내야 한다. 따라서 제3프로토콜에는 두 번째 내부

장치가 필요하다. 하나는 선언된 의지구배들 사이의 정의로운 조정을 다루는 GU이고, 다른 하나는 그 의지구배들이 현실의 사건으로 내려오는 방식을 기록하고 계산하는 실행 문법이다.

여기서 CVTD는 로그에서 목적을 대신 추출하는 체계가 아니다. 의지구배는 앞서 말했듯 추출되는 데이터가 아니라 형성되고 정련되고 승인되는 자기현법적 방향량이다. CVTD가 하는 일은 그 구매선언이 실제 의도적 사건 속에서 어떤 형태로 구현되는지, 그리고 여러 주체의 서로 다른 가치 기저가 어떻게 양의 합을 만들거나 충돌을 낳는지를 기술하는 것이다. 계약과 교환은 그 의도적 사건의 중요한 특수 사례일 뿐, CVTD의 전체 대상은 아니다. 그러므로 CVTD는 구매선언을 대체하지 않고, 선언된 의지구배가 사회적 사건으로 번역되는 과정을 다룬다.

정확히 말해 CVTD에는 두 방향이 있다. 역방향 CVTD(reverse CVTD)는 행위, 선택, 반복, 갈등, 후회, 커밋, 위반 패턴으로부터 개인의 의지구배를 추정하고 정련하는 방향이다. 순방향 CVTD(forward CVTD)는 구매선언을 조건, 계약, 사건, 기여, 책임 구조로 번역하는 방향이다.

전자는 “당신은 실제로 어디로 힘을 쓰고 있는가”를 묻고, 후자는 “당신이 승인한 방향을 어떤 사건 구조로 구현할 것인가”를 묻는다. 따라서 CVTD는 단순한 실행 장부가 아니라, 개인의 가치장과 현실 사건 사이를 왕복하는 도구다.

다만 역방향 CVTD의 추정은 개인의 구매선언을 대체하는 판정권이 아니다. 그것은 외부 제도나 플랫폼이 “당신의 진짜 의지구배는 이것이다”라고 확정하기 위한 장치가 아니라, 사용자가 자기 이름으로 승인한 선언과 실제 사건 이력 사이의 정렬과 괴리를 되돌아보기 위한 반성적 도구다. 사건 이력은 선언을 무효화하는 증거가 아니라, 선언이 어디에서 구현되고, 어디에서 왜곡되고, 어디에서 수정되어야 하는지를 보여주는 재료다. 이 경계가 무너지면 역방향 CVTD는 곧바로 행동 로그 기반의 내면 추출 기술로 변질된다. 반대로 이 경계가 지켜질 때, CVTD는 감시가 아니라 자기정련과 책임의 문법이 된다.

11 GU에서 CVTD로

전지적 공감의 형식화는 GU가 단지 수사적 이상이 아니라 계산 가능한 방향성이라는 점을 보여준다. 그러나 여전히 하나의 간극이 남는다. GU는 무엇이 정의에 더 가까운지를 말해 주는 좌표계이지만, 현실 세계의 의지구배는 언제나 구체적 의도적 사건의 형태로 발생한다. 어떤 사람은 계약하고, 어떤 사람은 조직을 만들고, 어떤 사람은 설득하고, 어떤 사람은 돌봄을 제공하며, 어떤 사람은 특정 행위를 거부함으로써 세계 상태를 바꾼다. 따라서 정의의 방향성과 실제 의도적 사건의 언어 사이를 이어 주는 실행층이 필요하다.

CVTD는 바로 이 간극을 메우기 위한 시도이다. 만약 GU가 선언된 의지구배들의 조정과 수렴을 위한 거시적 규범이라면, CVTD는 그러한 의지구배들이 물리적 세계에서 어떤 사건 행렬과 가치 기저와 커밋 구조로 표현될 수 있는지를 다루는 미시적 실행 언어가 된다. 다음 절은 이 점에서 GU를 버리는 것이 아니라, 오히려 GU를 현실의 계산 가능한 상호작용으로 번역하는 과정으로 읽혀야 한다. 표기에 관해 한 가지를 분명히 해 두자. GU 절의 π 와 CVTD의 π 는 동일한 객체이며, 둘 다 의지구배를 가진 주체가 세계 상태에 가하는 “의도적 사건 그 자체”를 가리킨다. 차이는 다루는 추상 수준에 있을 뿐이다. GU에서는 π 를 그 사건이 유도하는 변환 $\pi: O_t \rightarrow O_{t+1}$ 으로 추상적으로 다루고, CVTD에서는 같은 π 를 튜플 $\pi = (A_\pi, X_\pi, \Phi_\pi, \tau_\pi, \Gamma_\pi, \mu_\pi)$ 로 풀어 쓰며, 그 사건이 만드는 가치 변화를 행렬 $\Pi(\pi, t)$ 로 표현한다. 마찬가지로 $\mathcal{P}$ (또는 $\mathcal{P}^{\text{int}}$)는 GU의 π 를 분해해서 얻는 무엇이 아니라, 의도적 사건들의 집합(또는 시점 t 에 활성화된 사건들의 집합)에 대한 표기일 뿐이다. 한편 CVTD의 핵심 객체인 가치 평가 연산자 $\mathbf{C}_n^t$ 는 처음부터 “ n 이 세계를 자기 가치공간으로 사상한

결과”라는 관계 우선 구조를 갖는다. 이는 GU의 일반형 $H_n = I_n + \sum_i c_{n,i} H_{n \rightarrow i}^\pi$ 와 그대로 정합한다. 즉 CVTD는 처음부터 $H_{n \rightarrow i}^\pi$ 형 귀속을 명시적으로 다루고 있었고, 본 원고의 GU 일반화는 그 정합성을 명시화한 셈이다. 다만 CVTD 역시 완성된 calculus라기보다, 실행층의 수학적 골격을 어떻게 세울 수 있을지에 대한 초기 제안이다.

12 계약적 가치이전역학(Contractual Value Transfer Dynamics, CVTD): 가치의 계산과 의지구배의 역산

일반공리주의(GU)가 정의(Justice)의 수렴 방향을 제시하는 거시적 좌표계라면, 계약적 가치이전역학(CVTD)은 이를 물리적 세계 위에서 실제로 계산하고 연산하기 위한 실행 레이어(Execution Layer)이다. 우리는 우주의 모든 사건 중 자연 현상을 제외하고, 앞서 정의한 비존재적 좌표로서의 주체들이 의지구배를 가지고 만들어내는 모든 “의도적 사건”을 CVTD의 기술 대상으로 삼는다.

따라서 CVTD의 “계약적”이라는 표현은 법률상 계약서나 명시적 합의 절차만을 뜻하지 않는다. 그것은 의도적 행위가 언제나 조건, 기대, 책임, 커밋, 귀속, 위반 가능성, 집행 가능성의 구조를 가진다는 뜻이다. 친구에게 조언하는 행위, 회사를 창업하는 행위, 글을 쓰는 행위, 어떤 제안을 거부하는 행위, 공동체 규칙을 설계하는 행위, 자원을 이전하는 행위, 명시적 계약을 체결하는 행위는 모두 서로 다른 형태의 π 이다. 계약과 교환은 CVTD의 중요한 사례 중 하나이지만, CVTD 자체는 모든 의도적 사건 π 를 기술하기 위한 동역학이다.

이 장에서는 GU 절의 의도적 사건 π 를 튜플로 풀어 쓰고, 각 사건의 행렬 표현 $\Pi(\pi, t)$ 와 개인별 가치 평가 연산자를 선형대수학적 실체로 구체화한다. $\mathcal{P}$ (또는 $\mathcal{P}^{\text{int}}$)는 그러한 사건들의 집합에 대한 표기이지 GU의 π 를 분해한 결과가 아니다. 이를 통해 실제 의도적 사건 시나리오에서 가치 총량의 증가 메커니즘과 의지구배의 역산 과정을 설명한다.

12.1 의도적 사건 π 의 일반형

CVTD에서 하나의 의도적 사건은 단순한 거래 기록이 아니라, 세계 상태에 작용하는 의지구배를 가진 변환이다. 주체 집합을 $\mathcal{N} = \{1, \dots, N\}$, 시점 t 의 세계 상태를 O_t , 공통 가치공간을 $K = \text{span}(k_1, \dots, k_d) \simeq \mathbb{R}^d$ 라 하자. 그러면 사건 π_t 는 다음과 같은 튜플로 정의될 수 있다.

$$\pi_t = (A_\pi, X_\pi, \Phi_\pi, \tau_\pi, \Gamma_\pi, \mu_\pi)$$

여기서 $A_\pi \subseteq \mathcal{N}$ 는 사건에 참여하거나 영향을 받는 주체 집합, $X_\pi \subseteq O_t$ 는 사건이 다루는 객체, 행위, 관계, 제도 상태의 묶음, Φ_π 는 사건이 성립하는 조건과 맥락, $\tau_\pi : O_t \rightarrow O_{t+1}$ 는 세계 상태 변환, Γ_π 는 커밋, 책임, 제약, 집행, 신뢰 구조, μ_π 는 이 사건이 어떤 구매선언 또는 의지구배 후보와 연결되는지를 나타내는 구매 서명이다.

이때 사건의 기본 작용은 다음과 같다.

$$O_{t+1} = \tau_\pi(O_t)$$

자연 현상은 구매 서명 μ_π 와 커밋 구조 Γ_π 를 갖지 않으므로 CVTD의 직접 대상이 아니다. 반대로 어떤 주체의 의지구배가 세계 상태에 개입해 어떤 변환을 일으킨다면, 그것이 거래든 설계든 대화든 거부든 제도 형성이든 모두 CVTD의 π 가 된다.

12.2 가치 변화 벡터와 사건 행렬

각 주체 n 은 세계 상태와 사건 객체를 자신의 가치 기저로 사상하는 평가 연산자 $\mathbf{C}_n^t : O_t \rightarrow K$ 를 가진다. 이 평가 연산자는 GU의 가치적 귀속 $H_{n \rightarrow i}^\pi$ 와 같은 관계 우선 구조를 명시적으로 구현한 것으로, 한 주체가 어떤 대상을 자기 가치장에 어떤 좌표로 받아들이는가를 직접 표현한다. 사건 π_t 가 일어났을 때 주체 n 의 주관적 가치 변화는 다음과 같이 표현된다.

$$\Delta_n(\pi, t) = \mathbf{C}_n^t(O_{t+1}) - \mathbf{C}_n^t(O_t)$$

이 값은 돈의 이동만이 아니라 시간, 신뢰, 전략적 자산, 평판, 자유도, 책임, 관계 안정성, 목적 정렬성 같은 여러 가치축의 변화를 포함한다. 따라서 동일한 사건도 주체마다 서로 다른 행 벡터로 나타난다.

사건 전체의 행렬 표현은 다음과 같다.

$$\Pi(\pi, t) = \begin{pmatrix} \Delta_1(\pi, t) \\ \Delta_2(\pi, t) \\ \vdots \\ \Delta_N(\pi, t) \end{pmatrix} \in \mathbb{R}^{N \times d}$$

따라서 이하에서 π 는 의도적 사건 자체를, $\Pi(\pi, t)$ 는 그 사건이 시점 t 에서 만드는 선형대수학적 사건 행렬을 가리킨다. 이 구분은 중요하다. 같은 사건 π 라도 평가 시점, 가치 기저, 참여자 집합이 달라지면 행렬 표현 $\Pi(\pi, t)$ 는 달라질 수 있기 때문이다.

사회적 가치 변화의 총량 벡터는 다음처럼 정의된다.

$$G(\pi, t) = \mathbf{1}^T \Pi(\pi, t) \in K$$

또한 GU가 제공하는 사회적 평가 함수 또는 정렬 함수 $\Psi_t : K \rightarrow \mathbb{R}$ 를 두면, 사건의 규범적 점수는 다음과 같이 표현된다.

$$W(\pi, t) = \Psi_t(G(\pi, t))$$

따라서 CVTD에서 양의 합(Positive Sum)이란 단순히 돈의 총합이 늘어나는 것이 아니라, 사건 행렬의 총량 벡터가 관련 가치축에서 양의 방향을 가지거나, GU적 정렬 함수 아래에서 더 높은 값으로 평가되는 경우를 뜻한다. 이때 어떤 축에서의 양의 합이 다른 축의 침해를 숨길 수 있으므로, Γ_π 와 Φ_π 는 동의, 책임, 강제, 정보 비대칭, 장기 비용을 함께 기록해야 한다.

12.3 사건 흐름으로서의 동역학

CVTD는 단일 사건의 표기법을 넘어 사건들의 시간적 흐름을 다룬다. 시점 t 에 활성화된 의도적 사건들의 집합을 $\mathcal{P}_t^{\text{int}}$, 각 사건의 강도 또는 실행 비율을 $\lambda_\pi(t)$ 라 하면, 사회의 누적 가치 상태 $Z(t)$ 는 다음과 같이 갱신된다.

$$Z(t+1) = Z(t) + \sum_{\pi \in \mathcal{P}_t^{\text{int}}} \lambda_\pi(t) \Pi(\pi, t)$$

이 식이 CVTD를 단순한 장부가 아니라 동역학으로 만든다. 제3프로토콜은 구매선언들을 받아 가능한 사건 집합 $\mathcal{P}_t^{\text{int}}$ 를 생성하고, GU는 그 사건들의 규범적 정렬을 평가하며, CVTD는 실제 선택된 사건들이 각 주체와 사회의 가치 상태를 어떻게 변화시키는지 계산한다.

12.4 GU 변수의 벡터 공간 매핑

12.4.1 사건 집합과 사건 행렬 (Event Matrix)

GU와 CVTD에서 의도적 사건 π 는 동일한 객체이며, CVTD에서는 그 사건들의 집합을 $\mathcal{P}^{\text{int}}$ 로 표기한다. 개별 의도적 사건 $\pi \in \mathcal{P}^{\text{int}}$ 는 행렬 표현 $\Pi(\pi, t)$ 를 통해 계(System)의 가치 총량을 변화시킨다. 물리적 세계에서 의지구배를 가진 주체가 발생시키는 모든 의도적 상호작용은 $\mathcal{P}^{\text{int}}$ 의 한 원소로 취급된다.

12.4.2 공감 계수 c 와 가치 평가 연산자 (Value Evaluation Operator, C)

GU에서의 공감 계수 $c_{n,i}$ 는 주체 n 이 다른 주체 i 의 가치적 좌표 $H_{n \rightarrow i}^\pi$ 를 자신의 효용 함수에 얼마나 결속시키는지 나타내는 관계적 가중치다. 반면 CVTD에서의 $\mathbf{C}_n^t : O_t \rightarrow K$ 는 주체 n 이 세계 상태와 사건 객체를 자신의 가치공간 K 위의 좌표로 평가하는 가치 평가 연산자다. 둘은 같은 객체가 아니다.

$c_{n,i}$ 가 i 에 대한 자아의 결속 강도를 나타낸다면, $\mathbf{C}_n^t$ 는 어떤 사건이나 객체가 n 에게 어떤 가치측 변화로 나타나는지를 계산한다. 따라서 $\mathbf{C}_n^t$ 는 주체 n 이 세계를 해석하는 구조적 필터이며, CVTD의 사건 행렬을 구성하는 주관적 평가 연산자다. GU의 $H_{n \rightarrow i}^\pi$ 가 거시적 수준에서 “ n 이 i 를 가치장에 어떻게 받아들이는가”를 묻는다면, CVTD의 $\mathbf{C}_n^t$ 는 미시적 수준에서 그 받아들이는 좌표 분해를 직접 계산한다.

12.5 사건 행렬 $\Pi(\pi, t)$ 의 구성과 주관성

가장 단순한 의도적 사건 원형 중 하나는 2자-2객체 교환(2×2 Primitive)이다. CVTD의 전체 대상은 이보다 넓지만, 교환 사건은 서로 다른 가치 기저가 어떻게 양의 합을 만들 수 있는지를 가장 선명하게 보여준다. 주체 A 와 B 가 각각 객체 x 와 y 를 교환하는 사건($A \rightarrow B : x, B \rightarrow A : y$)을 고려하자. 이때 사건 π 의 행렬 표현 $\Pi(\pi, t)$ 는 다음과 같이 정의된다.

$$\Pi(\pi, t) = \begin{pmatrix} \mathbf{C}_A(y) - \mathbf{C}_A(x) \\ \mathbf{C}_B(x) - \mathbf{C}_B(y) \end{pmatrix}$$

여기서 1행은 주체 A 의 관점에서 본 순 가치 변화량이며, 2행은 주체 B 의 관점이다.

핵심은 물리적으로 동일한 객체 x 라 할지라도, $\mathbf{C}_A$ 와 $\mathbf{C}_B$ 라는 서로 다른 기저(Basis)를 통과하며 전혀 다른 가치 값으로 변환된다는 점이다. 이 기저의 차이(Misalignment)가 바로 동일한 사건이 양측 모두에게 양의 합(Positive Sum)을 만들 수 있는 원동력이다.

12.6 사례 연구: 프리랜서 개발자와 스타트업 창업자

CVTD의 작동 원리를 명확히 하기 위해, 의도적 사건의 한 특수 사례인 계약형 교환 시나리오를 행렬로 분석한다.

시나리오 설정. 우리는 가치 공간을 3차원 축

$$K = [k_1 : \text{재무}, k_2 : \text{시간/노력}, k_3 : \text{전략적 자산}]$$

으로 정의한다.

- 주체 A(개발자): “안정성(Cash)”과 “워라벨(Time)”을 중시하는 기저 $\mathbf{C}_A$ 를 가짐.
- 주체 B(창업자): “제품 자산(Asset)”과 “속도”를 중시하며 리스크를 감수하는 기저 $\mathbf{C}_B$ 를 가짐.
- 교환 객체: x (개발된 코드), y (용역비).

주관적 평가(Projection). 각 주체의 기저에 따른 객체의 평가는 다음과 같다.

- A의 평가: 코드를 작성하는 것은 시간 손실($k_2 : -10$)이지만, 용역비는 큰 재무적 이익($k_1 : +15$)이다.

$$\mathbf{C}_A(x) = [0, -10, 0], \quad \mathbf{C}_A(y) = [+15, 0, 0]$$

- B의 평가: 용역비 지출은 재무적 손실($k_1 : -15$)이지만, 코드 획득은 막대한 자산 가치($k_3 : +50$)이다.

$$\mathbf{C}_B(x) = [0, 0, +50], \quad \mathbf{C}_B(y) = [-15, 0, 0]$$

사건 행렬 $\Pi(\pi_{\text{dev}}, t)$ 의 도출. 위 값을 앞서 정의한 사건 행렬의 형식에 대입하여 구성한다.

$$\begin{aligned} \Pi(\pi_{\text{dev}}, t) &= \begin{pmatrix} (15 - 0) & (0 - (-10)) & (0 - 0) \\ (-15 - 0) & (0 - 0) & (50 - 0) \end{pmatrix} \\ &= \begin{pmatrix} 15 & 10 & 0 \\ -15 & 0 & 50 \end{pmatrix}. \end{aligned}$$

해석. 이 행렬은 단순한 장부 기록 이상의 정보를 담고 있다. 물리적 재화(돈, k_1)의 합은 0이지만, 행렬 전체의 가치 총량 벡터는 $[0, 10, 50]$ 으로 양의 값을 가진다. 즉, 서로 다른 기저 $\mathbf{C}$ 가 교차하면서 무에서 유(시간 효율과 자산 가치)를 창조했다. CVTD는 이 행렬을 통해 “무엇이 거래되었는가”를 넘어 “왜 거래되었는가”를 기록한다.

12.7 선형대수학적 동역학: 의지구배의 역산과 정렬

사건을 행렬 $\Pi(\pi, t)$ 로, 주체의 평가 구조를 가치 평가 연산자 $\mathbf{C}$ 로 표현함으로써, 우리는 자아의 문제와 윤리적 판단을 “계산 가능한(Computable)” 영역으로 가져올 수 있다.

12.7.1 고유벡터(Eigenvector)와 의지구배의 역산

주체 A 가 관여한 수백 건의 의도적 사건 $\pi_1, \pi_2, \dots$ 에서 A 의 행 벡터들을 모아 사건 이력 행렬 $M_A(T)$ 를 구성한다고 가정하자.

$$M_A(T) = \begin{pmatrix} \Delta_A(\pi_1, t_1) \\ \Delta_A(\pi_2, t_2) \\ \vdots \\ \Delta_A(\pi_T, t_T) \end{pmatrix}$$

이때 공분산 행렬은 다음과 같다.

$$\Sigma_A = \frac{1}{T} M_A(T)^T M_A(T)$$

Σ_A 의 최대 고유값에 대응하는 고유벡터(Eigenvector) $\mathbf{e}_A$ 는 A 가 반복적으로 선택하고 감수하고 생성한 사건들이 향하는 지배적 가치 방향을 가리킨다.

위 예시에서 A 가 지속적으로 $[15, 10, 0]$ 형태의 이득을 취한다면, $\mathbf{e}_A$ 는 k_1 (재무)과 k_2 (시간) 평면 상에 놓이게 된다. 이는 A 의 의지구배가 안정적인 삶의 영위에 가까움을 수학적으로 보여준다.

반면, B 의 고유벡터는 k_3 (전략적 자산)을 강하게 가리킬 것이며, 이는 B 의 의지구배가 성취와 확장에 있음을 보여준다. 즉, CVTD에서 자아란 형이상학적 실체가 아니라, 의도적 사건 행렬들이 반복적으로 드러내는 방향성과 그 방향성을 승인하는 구조다.

다만 이것은 구배선언을 행동 로그로 대체한다는 뜻이 아니다. 앞서 말한 것처럼 구배선언은 개인의 자기헌법적 승인이고, CVTD의 사건 이력은 그 선언이 실제 사건 속에서 어떻게 구현되고 왜곡되고 수정되는지를 보여준다. A 의 선언된 의지구배를 g_A^{decl} 이라 하면, 선언과 사건 이력의 정렬도는 다음과 같이 측정될 수 있다.

$$\alpha_A = \frac{\langle g_A^{\text{decl}}, \mathbf{e}_A \rangle}{\|g_A^{\text{decl}}\| \|\mathbf{e}_A\|}$$

α_A 가 높을수록 구배선언과 반복된 의도적 사건의 방향이 정렬되어 있으며, 낮을수록 선언, 자기이해, 환경 제약, 실제 선택 사이에 재조정되어야 할 간극이 존재한다.

12.7.2 기저 변환(Change of Basis)과 상호 이해

주체 간의 갈등은 서로 다른 기저 $\mathbf{C}$ 를 사용하기 때문에 발생한다. 동일한 가치공간 K 와 양측의 기저가 명시되어 있다면, 선형대수의 기저 변환 행렬(Transition Matrix) $P_{A \rightarrow B}$ 를 통해 주체 A 의 가치관 $\mathbf{C}_A$ 을 B 의 관점 $\mathbf{C}_B$ 으로 정확하게 좌표 변환할 수 있다.

$$[\text{Value}]_B = P_{A \rightarrow B} [\text{Value}]_A$$

현실에서는 양측의 기저가 완전히 명시되어 있지 않을 수 있으므로, 사건 이력으로부터 다음과

같이 근사 변환을 학습할 수도 있다.

$$\hat{P}_{A \rightarrow B} = \arg \min_P \sum_{s=1}^T \|\Delta_B(\pi_s, t_s) - P\Delta_A(\pi_s, t_s)\|^2$$

이는 CVTD가 은유라는 뜻이 아니라, 형식적으로는 정확한 좌표 변환을 정의하고 현실 적용에서는 미지의 기저를 추정한다는 뜻이다. 이를 통해 A 의 주말 근무 불가(시간 가치 보호)라는 요구는, B 의 기저에서 생산성 저하 리스크 관리(전략적 자산 보호)라는 형태로 번역되어 전달될 수 있다. 이는 상호 이해를 높이고 GU의 목표인 전지적 공감($\mathcal{E}_{n \rightarrow i}^\pi \rightarrow 1$)에 근접하게 하는 수학적 엔진이다.

12.7.3 SVD와 잠재 욕망의 발견

사건 행렬 $\Pi(\pi, t)$ 또는 사건 이력 행렬 $M_A(T)$ 를 특이값 분해(SVD, $\Pi(\pi, t) = U\Sigma V^T$)하면, 명시적인 사건 구조 뒤에 숨겨진 잠재적 차원(Latent Dimensions)을 발견할 수 있다. 예를 들어, A 와 B 의 상호작용 이력을 분해했을 때, 명시적인 가치축 (k_1, k_2, k_3) 외에 제4의 축에서 높은 상관관계가 발견될 수 있다. 이는 두 사람이 계약서에는 쓰지 않았지만, 암묵적으로 공유하는 “미적 취향”이나 “신뢰 비용”일 수 있다. 고도화된 AI 시스템은 이 잠재 차원을 포착하여, B 에게 A 는 단순한 개발자가 아니라 당신의 미적 비전을 공유하는 파트너라고 제안함으로써, 더 높은 차원의 의도적 사건과 협력(Positive Sum)을 유도할 수 있다.

13 TBDM: 개인 효용 압력에서 집단 동력으로

CVTD가 모든 의도적 사건 π 를 미시적으로 기록하는 문법이라면, 다음 질문은 그 무수한 미시 사건들이 집단 수준에서 어떤 거시적 힘으로 나타나는가이다. 여기서 필요한 보조 모델이 TBDM 이다. 다만 TBDM은 또 하나의 독립 엔진이라기보다, 개인의 의지구배와 사회의 정의근사 방향량을 비교 가능한 언어로 바꾸는 좌표계이자 기저 변환 도구에 가깝다. 이 모델의 출발점은 단순 하다. 집단은 독립된 의지를 갖지 않는다. 국가, 기업, 정당, 시장, 민족, 종교가 무엇을 원한다고 말할 때, 실제로 의식하고 욕망하고 판단하고 행동하는 것은 언제나 원자적 개인들이다. 집단의 의지처럼 보이는 것은, 개인들이 특정 효용 압력 아래에서 같은 방향으로 정렬될 때 발생하는 거시적 패턴이다.

TBDM은 개인의 효용 압력이 대체로 세 기저축 위에서 표현된다고 본다.

기저	핵심 질문	대표 매개물	실패와 충족의 상태
가능성	나는 무엇을 할 수 있는가	돈, 자원, 시간, 안전, 기술, 자유, 권력, 접근권	무력함, 빈곤, 위험에서 선택지, 안전, 실행력으로 이동
사회성	나는 타자들 속에서 어떤 존재인가	명예, 지위, 평판, 사랑, 소속, 신뢰, 권리	배제, 수치, 고립에서 인정, 소속, 존중으로 이동
정합성	나는 무엇을 믿고 살아도 되는가	진실, 의미, 명분, 신념, 세 계관, 양심	혼란, 허무, 자기배반에서 의미, 확신, 진실성으로 이동

짧게 말하면 가능성은 “할 수 있어야 한다”, 사회성은 “받아들여져야 한다”, 정합성은 “말이 되어야 한다”는 압력이다. 돈은 가능성의 대표적 매개물이고, 명예는 사회성의 대표적 매개물이며, 진실 또는 명분은 정합성의 대표적 매개물이다. 이 세 축이 같은 방향을 가리킬수록 개인은 강하게 정렬된다. 어떤 집단에 속하는 것이 이익이고, 그 집단 안에서 인정받는 것이 존엄이며, 그 집단이

말하는 세계가 진실처럼 느껴질 때, 개인은 집단의 방향성을 외부 명령이 아니라 자기 의지처럼 내면화한다.

이를 CVTD와 연결하면 다음과 같다. CVTD의 가치공간 $K \simeq \mathbb{R}^d$ 위에 표현된 각 주체 n 의 사건별 가치 변화 $\Delta_n(\pi, t) \in K$ 와 선언된 의지구배의 좌표 표현 $\mathbf{g}_n^{\text{decl}} \in K$ 를 가능성, 사회성, 정합성의 세 기저공간

$$\mathcal{B}_3 = \text{span}(\mathbf{e}_{\text{pos}}, \mathbf{e}_{\text{soc}}, \mathbf{e}_{\text{coh}}) \simeq \mathbb{R}^3$$

으로 사상하는 투영 연산자 $R_3 : K \rightarrow \mathcal{B}_3$ 를 둔다. 이 사영은 GU의 가치적 귀속 $H_{n \rightarrow i}^\pi$ 를 가능성 · 사회성 · 정합성 세 축에서 비교 가능한 좌표로 만드는 역할을 한다. 즉 TBDM은 $\text{GU} \cdot \text{CVTD}$ 와 분리된 별개의 모델이 아니라, 같은 관계 우선 구조를 삼기저 위에 사영해 본 결과다.

$$\mathbf{f}_n(\pi, t) = R_3 \Delta_n(\pi, t) = \begin{pmatrix} f_n^{\text{pos}}(\pi, t) \\ f_n^{\text{soc}}(\pi, t) \\ f_n^{\text{coh}}(\pi, t) \end{pmatrix}$$

동시에 개인 n 이 승인한 선언된 의지구배 $\mathbf{g}_n^{\text{decl}}$ 의 삼기저 표현은 다음처럼 쓸 수 있다.

$$\mathbf{h}_n^{\text{decl}} = R_3 \mathbf{g}_n^{\text{decl}} = \begin{pmatrix} h_n^{\text{pos}} \\ h_n^{\text{soc}} \\ h_n^{\text{coh}} \end{pmatrix}$$

여기서 $\mathbf{f}_n(\pi, t)$ 는 사건 π 가 개인 n 에게 실제로 가하는 삼기저 효용 압력이고, $\mathbf{h}_n^{\text{decl}}$ 은 개인이 자기 이름으로 승인한 구배선언의 삼기저 방향이다. f^{pos} 와 h^{pos} 는 가능성의 압력과 선언 방향, f^{soc} 와 h^{soc} 는 사회성의 압력과 선언 방향, f^{coh} 와 h^{coh} 는 정합성의 압력과 선언 방향을 뜻한다. 따라서 TBDM은 실제 사건들이 개인의 선언된 방향과 얼마나 정렬되는지, 또는 어느 축에서 괴리를 만드는지를 비교할 수 있다. 집단 수준의 알짜힘은 다음처럼 쓸 수 있다.

$$\mathbf{F}(t) = \sum_{\pi \in \mathcal{P}_t^{\text{int}}} \sum_{n=1}^N w_n(t) \lambda_\pi(t) \mathbf{f}_n(\pi, t)$$

이때 $w_n(t)$ 는 각 개인의 영향력, 책임, 노출도, 취약성 또는 대표성을 반영하는 가중치이고, $\lambda_\pi(t)$ 는 사건의 실행 강도다. 집단의 방향성은 독립된 집단 정신에서 나오지 않는다. 그것은 원자적 개인들의 삼기저 압력이 평균화되고, 상쇄되고, 특정 방향으로 증폭될 때 나타나는 $\mathbf{F}(t)$ 의 방향이다.

이 모델의 설계 목표는 개인의 의도를 억압해 집단적 선을 강제로 만드는 것이 아니다. 오히려 개인들 사이에서 발생하는 무수한 미시적 가치 이전과 의도적 사건을 통과시키되, 악한 힘은 내부에서 서로 상쇄되고 선한 알짜힘만 거시적으로 남도록 프로토콜을 설계하는 것이다. 이를 위해 제3프로토콜은 각 구배선언이 가능성, 사회성, 정합성의 어느 축을 강화하거나 훼손하는지 드러내야 하고, GU는 그 힘의 방향을 평가해야 하며, CVTD는 실제 사건들이 어떤 삼기저 압력으로 누적되는지 계산해야 한다.

기업을 예로 들면, 기업이 이윤과 성장을 원하는 것처럼 보이는 이유는 기업이라는 생명체가 욕망하기 때문이 아니다. 구성원들이 월급, 승진, 지분, 고용 안정이라는 가능성 압력, 직함과 성과 평가와 업계 평판이라는 사회성 압력, 미션과 혁신과 고객 가치라는 정합성 압력 아래에서 같은 방향으로 정렬되기 때문이다. 정당, 국가, 민족, 시장도 마찬가지다. 집단의 이름으로 말해지는 의지는

개인들의 삼기저 효용 정렬이 만든 거시적 그림자다.

따라서 어떤 집단 현상을 분석할 때 TBDM은 세 가지 질문을 던진다. 가능성의 차원에서는 누가 무엇을 얻고 잃는가. 사회성의 차원에서는 누가 인정받고 누가 배제되는가. 정합성의 차원에 서는 어떤 명분과 진실의 언어가 행동을 정당화하는가. 이 세 질문을 통과하면, 집단의 의지처럼 보였던 것이 개인 단위의 효용 압력으로 분해된다. 인류의 driving force는 집단의 의지가 아니라, 원자적 개인들의 삼기저 효용 정렬이다.

14 프로토콜에서 현실 웨지로

여기까지의 논의가 인간론, 사회론, 규범론, 실행론을 세웠다면, 마지막 질문은 현실적 진입점이다. 구매선언은 중요하다고 해서 즉시 풍부하게 생성되지 않는다. 인간은 자신의 의지구배를 이미 완성된 문장으로 보유하고 있다가 플랫폼에 입력하는 존재가 아니다. 의지구배는 읽기, 사유, 반박, 수정, 대화, 승인이라는 느린 가치정련 과정을 통해 만들어진다. 따라서 제3프로토콜의 첫 현실 인터페이스는 목적을 곧장 거래시키는 시장이 아니라, 가치장을 정련하고 구매선언을 가능하게 만드는 인프라여야 한다.

이 절부터 The Channel은 완성된 제3프로토콜 그 자체가 아니라, 그 프로토콜로 가기 위한 첫 웨지로 자리 잡는다. 책과 장문 텍스트는 구매선언 이전의 가치정련 과정을 견디게 만드는 가장 느리고 단단한 매체이며, The Channel은 그 과정을 도시적 인터페이스와 대리 지능의 학습 구조 위에 배치하려는 현실적 출발점이다.

15 가치정련 인프라와 구매선언

그러나 의지구배는 원리상 추출만으로는 충분히 다룰 수 없다. 행동 로그, 클릭 패턴, 소비 기록은 말초적 선호와 즉각적 반응은 잘 포착할 수 있지만, 사람이 무엇을 위해 살고 어떤 세계를 원하며 무엇에 책임을 지려 하는지는 동일한 방식으로 읽어낼 수 없다. 의지구배는 대부분 이미 완성된 채 데이터 속에 숨어 있는 정보가 아니라, 읽고, 생각하고, 반박받고, 수정하고, 승인하는 과정을 통해 비로소 정련되기 때문이다. 따라서 앞으로의 핵심 병목은 더 좋은 추천 알고리즘이 아니라, 인간이 자신의 가치장을 정련하고 구매선언으로 승인하게 만드는 장의 부재에 있다.

이 점은 AI가 더 정교해질수록 오히려 더 중요해진다. Meta FAIR가 2026년 3월 26일 공개한 TRIBE v2는 영상, 음성, 언어 자극에 대한 인간의 fMRI 뇌 반응을 예측하는 tri-modal foundation model이며, Meta는 이를 인간 신경 활동의 디지털 트윈에 가까운 모델로 설명한다.³ 이 사례가 보여주는 것은 분명하다. 앞으로의 시스템은 인간의 반응을 점점 더 깊고 정밀하게 예측할 수 있다. 그러나 뇌 반응을 예측하는 것과 인간이 자기 이름으로 승인할 수 있는 의지구배를 갖는 것은 전혀 다른 문제다. 어떤 영상에 뇌가 반응했다는 사실은 중요한 자료이지만, 그것이 곧 “나는 이런 인간이 되겠다”는 헌법적 선언은 아니다. 추출이 강해질수록 선언의 필요성은 약해지는 것이 아니라 더 강해진다.

³Meta AI, “Introducing TRIBE v2: A Predictive Foundation Model Trained to Understand How the Human Brain Processes Complex Stimuli,” March 26, 2026, <https://ai.meta.com/blog/tribe-v2-brain-predictive-foundation-model/>. See also Meta AI Research, “A foundation model of vision, audition, and language for in-silico neuroscience,” <https://ai.meta.com/research/publications/a-foundation-model-of-vision-audition-and-language-for-in-silico-neuroscience/>.

이 점에서 필요한 것은 처음부터 목적시장이 아니라 가치정련 인프라이다. 그 안에서 인간은 읽고, 생각하고, 토론하고, 가치장을 다듬고, 마침내 의지구배를 선언한다. 그리고 그 선언은 단순한 발화가 아니라 책임과 귀속과 협상의 출발점이 된다. 이후에야 비슷한 구매선언들이 서로를 발견하고, 조건부 커밋과 자원 의사와 기여 형태가 결합되며, 목적조직과 목적시장으로의 이행이 가능해진다. 목적시장은 자본을 대체하지 않지만, 사람들을 처음 결속시키는 원리를 자본이 아니라 선언된 방향량에 두는 구조다.

이 과정이 실제로 대리 지능의 언어로 넘어가기 위해서는 표준화 또한 필요하다. 각 개인은 최상위 의지구배, 하위 구매, 우선순위, 금지 조건, 타협 가능 범위, 시간축, 자원 제약, 수정 이력, 공개 범위, 대리권 범위를 갖는 일종의 구매선언 패스पोर्ट를 통해 공적 공간에서 대리될 수 있어야 한다. 이때 플랫폼의 핵심은 직접 최고의 에이전트가 되는 데 있지 않다. 더 중요한 것은 어떤 모델 위의 에이전트든 자기 주체를 더 깊이 대리하기 위해 반드시 통과해야 하는 구매선언 포맷과 의미론적 표준층을 정의하는 데 있다.

기술적으로 이 표준층은 폐쇄형 에이전트 제국이 아니라, 에이전트 간 상호운용성을 전제해야 한다. Google이 2025년 4월 발표한 Agent2Agent(A2A) 프로토콜은 서로 다른 프레임워크나 벤더가 만든 에이전트들이 안전하게 정보를 교환하고 행동을 조율하기 위한 공개 프로토콜로 제안되었다.⁴ The Channel이 직접 모든 일을 하는 초거대 에이전트가 될 필요는 없다. 오히려 The Channel의 역할은 사용자의 구매선언 패스पोर्ट와 선언 이력을 표준화하고, 이를 A2A와 같은 에이전트 상호운용 표준을 통해 외부 에이전트들이 읽고 존중할 수 있는 의미론적 레이어로 만드는 데 있다. 그렇게 될 때 The Channel은 모델이 아니라 자기 주체를 중심에 두는 에이전트 생태계의 의지구배 레지스트리가 된다.

16 왜 읽기가 첫 번째 웨지인가: The Channel

이러한 가치정련 인프라가 필요하다면, 그 첫 번째 현실 인터페이스는 무엇이어야 하는가. 나는 그 답이 읽기 일반이 아니라, 장문과 책을 중심으로 한 느리고 단단한 언어 공간에 있다고 본다.

짧은 포스트와 피드와 반응형 미디어는 사람을 자극하고 선호를 포착할 수는 있어도, 삶의 방향과 장기 목적을 형성하게 만들기에는 지나치게 빠르고 얇다. 반면 책은 긴 문맥을 따라가게 하고, 자기 생각을 문장과 논리와 서사 속에서 다시 조직하게 만든다. 의지구배는 대부분 짧은 자극 속에서 태어나지 않고, 긴 독해와 긴 사고와 긴 언어를 필요로 한다. 내가 예상하는 것은, 실행과 맞춤화의 비용이 계속 떨어지고 지능의 총량이 커질수록 사회 전체가 더 높은 해상도의 구매선언을 요구하게 되며, 그 결과 사회의 “목적요구압”이 필연적으로 커진다는 점이다. 여기서 말하는 목적요구압은, 개인과 조직과 대리 지능이 “정확히 무엇을 원하고 무엇을 위해 커밋할 것인가”를 더 선명하게 말해내야 한다는 압력의 증가를 뜻한다.

책이 첫 번째 웨지가 되는 이유는 단지 책이 오래된 매체이기 때문이 아니다. 첫째, 책은 긴 시간의 주의를 요구하므로 즉각적 반응보다 자기 해석을 강화한다. 둘째, 책은 낮은 자극 밀도 속에서 개념의 계층을 따라가게 하므로, 파편적 취향이 아니라 세계관을 만든다. 셋째, 책은 문장과 논증과 서사의 형태로 남아 다시 밀줄 긋고, 주석 달고, 반박하고, 공유할 수 있는 공적 객체가 된다. 넷째, 책은 개인의 경험을 더 긴 역사적·철학적 기억과 연결하여, 자기 목적을 단순한 현재 욕구가 아니라 문명적 질문 위에서 재구성하게 만든다. 다섯째, 책을 둘러싼 읽기와 대화의 흔적은 이후

⁴Google Developers Blog, “Announcing the Agent2Agent Protocol (A2A),” April 9, 2025, <https://developers.googleblog.com/en/a2a-a-new-era-of-agent-interoperability/>.

대리 지능이 주인을 이해할 때 사용할 수 있는 고품질 가치정련 데이터가 된다. 따라서 책은 과거의 콘텐츠 형식이 아니라, 구배선언 이전의 가치정련을 위한 가장 안정적인 인터페이스다.

따라서 The Channel은 매우 현실적으로는 책 사업이다. 그러나 그 본질은 책을 상품으로 유통하는 데 있지 않다. 인간이 스스로 생각할 수 있는 조건을 다시 문명의 중심에 가져오는 데 있다. 스스로 생각하지 않는 인간은 점점 더 정교한 추천, 관리, 자극, 대리 판단의 체계 안에서 사육되거나, 자기 목적을 잃은 채 자기소거의 방향으로 밀려날 위험에 놓인다. 이때 책은 낭만적 취미가 아니라, 인간이 자기 가치장을 스스로 정련하기 위해 남겨진 가장 오래되고 강한 기술이다.

기술이 종교가 되기 이전, 사람들은 왜 살아야 하는가, 무엇을 사랑해야 하는가, 무엇을 부끄러워해야 하는가, 어떤 세계를 만들어야 하는가에 대한 끝없는 질문과 각자의 답을 인문학, 철학, 문학, 종교, 역사, 과학의 긴 문장으로 남겼다. 우리가 무한한 경쟁과 즉각적 최적화 속에서 뒤에 두고 온 것은 바로 그 장기 기억이다. The Channel은 그 장기 기억을 다시 현재의 가치정련 인프라로 연결하려는 시도다.

바로 이 지점에서 The Channel의 위치가 정해진다. 목적요구압이 높아질수록, 사회는 더 많은 콘텐츠를 필요로 하는 것이 아니라 더 분명하고 더 숙고된 구배선언을 낼 수 있는 담론의 씨앗을 필요로 하게 된다. 여기서 책은 단순한 매개체가 아니다. 그것은 읽히고, 밑줄 그어지고, 주석 달리고, 반박되고, 다시 해석되고, 마침내 선언으로 이어질 수 있는 장문형 정보 레이어이자 담론의 씨앗이다. 같은 텍스트를 둘러싸고 여러 사람이 자기 언어를 다시 조직하기 시작하는 순간, 책은 소비재가 아니라 가치정련의 출발점이 된다.

따라서 The Channel은 사람들이 자기 가치장과 세계관을 정련할 수 있는 실제 조건을 도시 위에 배치하는 첫 번째 채널 인프라이며, 가치정련 인프라가 목적시장으로 넘어가기 위한 가장 현실적인 출발점이다. 더 정확히 말하면, 우리가 여기서 겨냥하는 것은 사회에 배포될 지능의 개인에 대한 감각신경계이다. 기업과 정부의 영역에서 Palantir류의 시스템이 AI의 감각신경계에 가까운 위치를 점유하듯, The Channel이 노리는 것은 개인에 대해 그와 같은 위치다. 중요한 것은 책을 많이 유통하는 것이 아니라, 어떤 책과 어떤 장문 텍스트가 사람 안에서 사유와 토론과 세계관의 재조직을 일으키고, 그 변화가 나중에 선언 가능한 형태로 축적되는지를 붙잡는 일이다. The Channel은 담론의 씨앗을 사람들의 생활 동선 위에 배치하고, 그 씨앗이 읽기와 대화와 기록을 거쳐 가치정련과 구배선언으로 이어지는 경로를 축적하는 첫 번째 웨지로 기능한다.

더 나아가 이 구조는 인간층과 대리 지능층의 이중 구조를 전제한다. 인간은 책을 읽고, 긴 글을 쓰고, 토론하며 가치장을 정련하고 의지구배를 선언한다. 동시에 각자의 대리 지능은 그 긴 기록과 선언 이력, 가치 우선순위, 수정 과정뿐 아니라 무엇을 읽었고 어떤 텍스트를 계기로 생각이 바뀌었는지의 흔적까지 학습하여, 플랫폼 밖으로 나갈 때조차 주인을 더 깊이 이해한 상태로 남는다.

이때 The Channel의 의미는 단순한 사용 시간의 확보가 아니라, 말초적 선호를 최적화하던 대리 지능을 숙고된 의지구배를 대리하는 방향으로 전환시키는 첫 현실 인터페이스라는 데 있다.

여기서 대리 지능은 단순한 추천 엔진이 아니라, 내가 내 이름으로 승인할 수 있는 나를 함께 구성하는 트윈에 가까워진다. 사람들은 자신이 무엇을 원하는지 잘 모른다. 그렇다고 도파민에 이끌려 짧은 영상을 넘긴 기록이 완전히 내가 아니라고 말할 수도 없다. 그것 역시 나의 일부다.

다만 그것이 나 전체를 대표하기를 원하지 않을 뿐이다. The Channel의 트윈은 말초적 반응을 부정하지 않되, 그것을 긴 읽기, 긴 사유, 선언, 수정, 책임의 기록과 함께 놓는다. 그리고 사용자가 끝내 “이것은 내가 내 이름으로 승인할 수 있는 나다”라고 말할 수 있는 의지구배로 자신을 정렬하도록 돕는다. 결국 인간이 궁극적으로 원하는 것은 단순한 욕구 충족이 아니라, 자신이 원할 수 있는 자신에 가까워지는 일이다.

결국 책은 여기서 상품이 아니라 인간용 연료이자 담론의 씨앗이며, The Channel은 그 씨앗이 도시와 인간과 에이전트 사이를 오가며 가치정련과 구매선언으로 자라나게 만드는 첫 채널이다.

17 결론

본 원고는 현대 사회의 지식 파편화와 목적 상실 문제를 해결하기 위해, 자아를 과거에서 발굴되는 실체가 아니라 가치장을 정련하고 의지구배를 승인하는 비존재적 구조로 재정의하는 철학적 토대 위에서 출발하였다. 본 원고에서 자아 측의 n 과 타자 측의 i 는 모두 미리 결정된 생물학적 실체가 아니라, 인정되고 귀속되는 비존재적 좌표로서의 주체다. 자아가 인정되어야 가치가 발생하고, 자아는 가치사상을 통해 세계를 가치장으로 보며, 가치정련은 그 장에서 실천적 방향량을 산출한다. 제3프로토콜이 읽어야 하는 것은 인간의 가치관 전체나 내면 전체가 아니라, 자아가 자기 책임 아래 사회적으로 승인한 구매선언이다. 이를 바탕으로 정의의 수렴 방향을 제시하는 “일반공리주의 (Generalized Utilitarianism, GU)”를 제안하였으며, GU의 1차 입력은 타자의 효용 H_i 자체가 아니라 자아가 π 아래 타자를 가치장에 귀속시킨 값 $H_{n \rightarrow i}^\pi$ 이고, H_i 는 그 주체가 독립적 가치장을 갖는 것으로 인정될 때에만 $H_{n \rightarrow i}^\pi$ 안에서 분해되는 항임을 명시하였다. 나아가 이 추상적 모델을 물리적 세계에서 실행하기 위한 계산 언어로서 모든 의도적 사건 π 를 기술하는 “계약적 가치이전역학 (Contractual Value Transfer Dynamics, CVTD)”을 제안적 형태로 스케치하였다. GU와 CVTD에서 π 는 동일한 의도적 사건이며, CVTD의 사건 행렬 $\Pi(\pi, t)$ 는 그 사건이 만드는 가치 변화를 선형대수학적으로 표현한 것이다. 주체의 가치 평가 연산자 C 는 각 사건이 개인의 가치공간에서 어떤 변화로 나타나는지를 계산한다. GU의 공감 계수 c 는 그 변화가 다른 주체에 대한 가치적 귀속 $H_{n \rightarrow i}^\pi$ 를 자기 효용망에 어떻게 결속시키는지 나타내는 별도의 관계적 가중치다.

CVTD는 역방향으로 행위, 반복, 갈등, 후회, 커밋, 위반 패턴에서 개인의 의지구배를 추정하고 정련하며, 순방향으로 구매선언을 조건, 계약, 사건, 기여, 책임 구조로 번역한다. 이 통합된 프레임 워크는 고도화된 AI 시스템이 단순한 도구를 넘어, 인간의 선언된 의지구배와 실제 의도적 사건의 정련을 계산하고, 잠재된 연결을 발견하여 사회 전체를 전지적 공감의 상태로 수렴시키는 윤리적 동반자로 기능하기 위한 이론적 토대가 될 것이다.

나아가 TBDM은 이 미시적 사건들이 집단 수준에서 어떻게 가능성, 사회성, 정합성의 알짜 힘으로 나타나는지를 설명한다. 집단은 독립된 의지를 갖지 않는다. 집단의 의지처럼 보이는 것은 원자적 개인들이 세 기저에서 받는 효용 압력이 특정 방향으로 정렬될 때 발생하는 거시적 패턴이다. 따라서 제3프로토콜의 설계 목표는 개인의 의도를 삭제하는 것이 아니라, 미시적 구매선언과 가치 이전을 통과시키면서 악한 힘은 내부에서 상쇄되고 선한 알짜힘은 증폭되도록 만드는 데 있다.

그러나 본 원고의 확장은 여기서 멈추지 않는다. 실행 비용이 지속적으로 하락하고 지능의 총량이 계속 커지는 사회에서는 표와 가격만으로 인간의 의지구배를 충분히 집계할 수 없으며, 제3프로토콜의 출현이 점차 기능적으로 요구된다. 이때 핵심 병목은 더 많은 행동 데이터를 추출하는 데 있지 않고, 인간이 자기 가치장을 정련하고 공적으로 구매선언하게 만드는 인프라의 부재에 있다. 가치정련 인프라는 단순한 커뮤니티나 추천 엔진이 아니라, 인간의 속고를 사회적 표현으로 전환하고, 대리 지능이 그 표현을 더 깊이 학습하도록 만드는 층으로 이해되어야 한다.

이러한 맥락에서 The Channel은 완성된 목적시장이 아니라 그 시장으로 가는 첫 번째 현실적 채널로 위치 지어진다. 책과 장문 텍스트는 인간이 자기 가치장을 다듬는 가장 느리고 단단한 씨앗이자, 구매선언으로 이어질 수 있는 장문형 정보 레이어이며, The Channel은 그 씨앗을 도시적 인터페이스와 에이전트 학습 구조 위에 배치하는 첫 번째 웨지다. 그것은 책 사업이지만, 단지 책을

파는 사업은 아니다. 인간이 스스로 생각하고, 자기 이름으로 승인할 수 있는 트윈을 만들며, 그 트윈이 A2A와 같은 에이전트 상호운용 표준 위에서 외부 세계와 협상할 수 있게 만드는 가치정련 인프라다. 따라서 GU는 철학적 좌표계이고, CVTD는 아직 개발과 확장의 여지가 크게 남아 있는 실행층의 제안적 언어이며, 제3프로토콜과 가치정련 인프라는 그 둘이 사회적 구조로 확장되는 방향이다. 본 원고는 이 흐름 전체를 닫힌 체계로 완결하기보다, 앞으로의 사회가 어떤 종류의 구매 입력과 선언 구조를 필요로 하게 될 것인지에 대한 하나의 강한 방향 제시로 읽혀야 한다.

한 줄로 정리하면, 자아는 발견되는 것이 아니라 미래로 승인되는 주체이고, 타자는 자아의 가치 장 안에서 가치 · 복지 · 의지 · 공감 가능성이 귀속되는 주체다. GU의 1차 입력은 타자의 실제적 효용 H_i 가 아니라 자아가 타자를 가치장에 귀속시킨 좌표 $H_{n \rightarrow i}^\pi$ 이며, H_i 는 그 주체가 독립적 가치 장을 갖는 것으로 인정될 때 $H_{n \rightarrow i}^\pi$ 안에서 분해되는 항이다. 이 한 줄이 본 원고의 형식 체계 전체를 관통하는 비존재성의 대칭이며, 자아의 비실체성과 타자의 비실체성을 같은 문법 안에 묶으로써 GU와 CVTD가 비로소 같은 관계 우선 좌표계 위에서 닫힌다.