

애쉬비의 인공 뇌, 호메오스탯 연구: 사이버네틱스의 역사에 위치한 그의 입지에 관하여

고규흔†

알파고(AlphaGo)와 이세돌 9단의 제2국이 끝났을 때, 인공 지능 개발자이자 프로 기사인 김찬우 6단은 알파고가 보여준 포석의 ‘자유로운 발상’을 예찬한 바 있다.¹ 하지만 1국 종료 시점부터 언론들은 이미 알파고의 변칙적 선택이 묘수로 판명됐음을 강조하며 이를 직관과 창의성의 소산으로 기술해 나갔다.² 상황에 맞춰 승률을 최대화하려는 착점의 연산이 독창성의 발현으로 치환됐을 때, 이 같은 언어적 선택은 언술의 주체에게로 질문의 방향을 되돌린다. ‘기계도 독창적일 수 있는가?’라는 의문은 어느덧 인간의 고유성에 관한 회의로 귀환되면서, 창의성이나 직관과 같은 부류의 영묘한 개념들이 사실상 신경 세포의 거대 집합과 이와 인접한 분자들 간 상호 작용에 불과하다는 점을 상기시키기 때문이다.³ 그렇다면 신경망으로 구현된 알파고의 바둑 역시 고수의 직관이 작용한다 해도 모순은 없다. 학습을 통해 심층에 각인된 정보의 특수한 패턴이라는 차원에서, 양측이 보인 수읽기의 예리함은 서로 구별되지 않기 때문이다.⁴

† 성균관대학교, sugarcats@hitel.net

¹ 엄인용, “인간이 두지 않는 ‘충격의 수’... 5,000년 패러다임 바꾼 ‘신의 수’”, 『경향신문』, 2016. 3. 12, 6면.

² 석남준, “알파고의 창의적 수순에 깜짝 놀랐다”, 『조선일보』, 2016. 3. 10, <http://goo.gl/ZmLZtg> (2016. 12. 15 접속); 서완석, “알파고에 대한 3가지 의문점”, 『국민일보』, 2016. 3. 11, <http://goo.gl/usFzwB> (2016. 12. 15 접속).

³ 프랜시스 크릭, 김동광 옮김, 『놀라운 가설』 (궁리, 2015), 19.

사이버네틱스의 고전적 야심을 지탱했던 관점이 이와 흡사했다. 눈으로 확인되는 반상의 상황에 비중을 둘 때 알파고에게서 이세돌에 필적할 창의성이 발견됐듯, 기계가 인간/동물의 행동 양식을 그대로 복제할 수만 있다면 양측을 구분 짓는 기성의 경계는 무력화될 것이다. 일단 이를 위해선 의사 결정에 따른 지적 행위의 합성이 필요했고, 이 계획을 성사시킬 만한 단서들이 1940년대 중반, 이미 전쟁을 통해 성숙기에 접어든 통신학, 제어공학의 선례로부터 발굴될 수 있었다. 그들은 각 분야의 성과를 종합해 수학 증명이나 보편 이론을 정립해 나갔고, 이를 생명체의 대사와 행동, 심리에 대입, 이와 은유적 접목을 시도함으로써 사이버네틱스라는 광범위한 학제를 성립시켰다.

그들이 내린 잠정 결론에 따르면 지적 행위의 주체들은 하나같이 정보를 처리한다는 것, 통신 패턴과 메시지의 순환을 통해 정교한 행동을 구현한다는 것이다. 이렇게 인간/기계를 관통하는 통합 범주가 ‘정보 처리 시스템’으로 선포됐지만, 이는 대등한 통합이라기보다 피드백 증폭, 서보메커니즘 이론과 같은 자동 장치를 구현하는 원리에 기준점을 두고 있었다.⁵ 그 방법론은 일단 내부를 블랙 박스로 묶은 상태에서, 입·출력 가시 구간을 공학적으로 유사하게 끼워 맞추는 것이다. 그런데 이러한 방법은 기계·전기적 시스템과 생리학적 체계 사이에 벌어진 현격한 이질성의 무게를 모두 유실시킨다는 데 문제가 있었다. 결국 사이버네틱스의 고전 모델은 인공적 구성 체계보다 실상 훨씬 더 조밀하고 복잡하게 엮히는 생물학적 유기체 내부의 모호한 동세를 무마했고, 이 손실분은 신호 송수신과 정보 처리와 같은 비물질성을 전면에 배치함으로써 떠안은 모종의 등가교환 같은 거래의 결과였다.

간과했던 내부와 이를 메운 질료에 생각보다 더 큰 비중이 실려 있을지 모른

4 조환규, “이세돌과 인공지능 알파고의 반상결투”, 『경향신문』, 2016. 2. 29, 29면.

5 캐서린 헤일즈는 몇 세기에 걸쳐 인간과 기계가 유사한 존재로서 탐구되어 왔지만, 양쪽 모두가 정보 처리 시스템으로서 인식되기 시작한 계기로써 메이시 회의(The Macy Conferences)를 지목한다. N. K. Hayles, “Boundary Dispute: Homeostasis, Reflexivity and the Foundations of Cybernetics,” R. Markley, ed., *Virtual Realities and Their Discontents* (Baltimore: Johns Hopkins University Press, 1996), 11-38 중 11-12 참조. 이 회의는 메이시 재단(Josiah Macy Jr. Foundation)의 후원으로 열렸으며, 재단 의학부장인 프랭크 프레몬트-스미스(F. Fremont-Smith)에 의해 처음 제안되었다. 1946년에서 53년 까지, 총 10회에 걸쳐 진행된 이들 메이시 회의는 모두 뉴욕에서 열렸으며 그 회의록 가운데 6-10차 회의들만 정식으로 공개됐다. 이 학술 회의를 계기로 사이버네틱스 특유의 간학문적 성격과, 생명체와 인공 시스템의 경계를 넘나드는 전체 논의의 골격이 갖춰졌다.

다는 각성이 표면화된 것은 이 같은 탈물질화 경향이 인지주의(cognitivism)에 의해 정점을 찍은 직후였다. ‘뇌 = (디지털)컴퓨터’의 유비로써 이 학문적 유파가 50년대 중반부터 약 십 수 년 이상 공들인 기호(표상) 처리 인공 지능이 점차 한계를 드러내고 있을 때, 한편에서 사이버네틱스는 변화의 기회를 맞고 있었다. 인지주의가 이제 막 뿌리내리던 시기, 자동 제어학, 컴퓨터 과학, 인지 과학 등이 독립 영역을 구축하며 분리돼 나감에 따라 학제의 폭은 전보다 협소해졌지만, 그 색깔은 과거와 다른 빛을 발하게 된다. 이는 잔류파들이 생명에 대한 일괄적 수학 해석이나 공학적 접근을 접고, 기계가 아직 흉내 내지 못한 자연 생성 조직의 고유성에 집중한 결과였으며, 이로써 ‘2차 사이버네틱스(2nd-order cybernetics, 이하 2차 진영으로 표기)’라는 토양이 마련될 수 있었다.

그 경향은 다음과 같다. 메시지, 정보의 흐름에 의한 명료한 통제가 1차 진영을 지배했다면, 사이버네틱스의 2차 진영은 밀도 높게 얹히는 생명 조직의 복잡성과 여기서 발생하는 불명료한 창발을 주제로 삼았고, 이들은 교정 피드백이 지배하는 시스템 상의 인과적 투명성보다, 세포와 세포, 조직과 조직이 맞물린 체계가 외부로부터 난잡한 변수를 수용해 이를 자기 조직화와 재귀 생산의 동력으로 삼는 비선형 복잡계의 예측 불가능성을 취급하려 했다. 이 같은 혼돈의 질서가 생명의 근원이자 뇌가 작동하는 원리라는 것이다. 이 역설은 마치 1943년 슈뢰딩거(E. Schrödinger)의 연속 강연 “생명이란 무엇인가”에 포함된 2가지 문제 중, ‘무질서로부터의 질서’가 당대에 주목받지 못했던 역사를 향한 후대의 응답처럼 보였다.⁶ 마침 폰 포에르스터(H. von Foerster)는 1959년에 이 논제를 “노이즈로부터의 질서”로 수정해 발표함으로써, 환경으로부터 유입되는 방향성 없는 에너지와 이로써 구현되는 질서의 모순적 상관성을 밝힌다.⁷ 그리고 이는 2차 사이버네틱스의 개략적 틀이 갖춰지는 계기이기도 했다.

하지만 과거와의 단절로서 이 흐름을 해석하기는 어렵다. 이 역사를 개괄한

6 더블린 트리니티 대학에서 있었던 이 슈뢰딩거의 강연은 생명의 신비를 열역학으로 끌어와 물질계와 하나로 묶는 통합 이론을 도출하려 했다는 점에서 사이버네틱스의 방향성과 교집합이 있었다. 우연이지만 그의 강연이 있었던 1943년은 사이버네틱스 역사의 시작점을 새긴 논문 “행위, 의도, 목적론” [A. Rosenblueth, N. Wiener, and J. Bigelow, “Behavior, Purpose and Teleology,” *Philosophy of Science* 10:1 (1943), 18-24]이 발표된 해이기도 하다.

7 H. von Foerster, “On Self-Organizing Systems and Their Environments,” M. C. Yovits and S. Cameron, eds., *Self-Organizing Systems* (London: Pergamon Press, 1960), 31-50.

존스톤(J. Johnston)은, 후기 산업 사회를 지탱하는 정보 네트워크의 파생과 컴퓨터 제반 요소의 결집을 추진한 역사적 계기로서 사이버네틱스를 파악한다.⁸ 그가 견지하는 연속성의 역사에서 생명체의 적응과 복잡성, 자기 생산과 진화는 이 역사의 보편 주제로 확대된다. 한편, 헤일리겐(F. Heylighen) · 조슬린(C. Joslyn)은 2차 진영의 참신함이 흐름상의 단절로서 과잉 해석되는 것을 경계한다.⁹ 이들은 1차 진영의 주역들이 이미 자기조직화 문제를 직간접적으로 인식했다는 사실, 마투라나(H. R. Maturana)와 같은 인물은 2차 진영의 주역이면서도, 1차 사이버네틱스의 전개에 깊숙이 관여했음을 밝혀 말끔한 세대 구분의 불가를 강조한 바 있다.

‘2차’는 허구의 개념이 아니다. 다만 초기 구성원들이 취한 논의의 다양성으로 비춰볼 때 그 경계는 모호해진다. 이 논문의 핵심 인물인 로스 애쉬비(W. R. Ashby, 1903-1972)가 이에 부합할 것이다.¹⁰ 그가 제작한 ‘호메오스탯(Homoeostat)’은 생물학적 뇌의 복제를 표방했으나, 이 뇌는 정보 처리 시스템과 거리가 있어 사이버네틱스 초기 성향과 이격을 보였다. 또한 그의 인공 뇌는 시기상 1차 진영에, 성격상 초기의 보수성을 답습하면서도, 2차 논의의 쟁점들 상당 부분을 응축해 당대 사이버네틱스 논의의 폭을 확장했다. 그리고 그 입지의 특수성은 후대에 아래처럼 함축된 적이 있다.

... 호메오스탯은 초기 사이버네틱스의 아이콘이 되었다. 하지만 아이러니하게도 그것은 메이시 집단 내의 견고한 담론을 지탱하는 안정적 설계물이 아니었다.¹¹

본 논의는 그의 호메오스탯을 조명해 이에 얽힌 ‘2중 입지’의 정체성을 탐구

⁸ J. Johnston, *The Allure of Machinic Life* (Cambridge, MA: MIT Press, 2010), 25-26.

⁹ F. Heylighen and C. Joslyn, “Cybernetics and Second-Order Cybernetics,” R. A. Meyers, ed., *Encyclopedia of Physical Science & Technology* (3rd ed.) (New York: Academic Press, 2001), 155-170.

¹⁰ 정신과 의사이자 신경 생리학자로서 사이버네틱스에 동참할 당시 그는 영국 정신 의학센터 반우드 하우스(Barnwood House)의 연구 책임자였다. 전공 분야는 뇌기능 관련 생화학, 내분비 계통이었다. 1949년에는 영국의 사이버네틱스 연구 모임인 레이션 클럽(Ratio Club)의 구성원으로 참여했으며, 미국 일리노이대 전기공학부로 옮겨 1961년부터 폰 포에르스터와 함께 생물리학 연구를 이어갔다.

¹¹ R. R. Kline, *The Cybernetic Moment* (Baltimore: Johns Hopkins University Press, 2015), 4.

할 것이다. 여기서 2중 입지란 피드백이 주관하는 항상성에 묶여 발생과 자기생산의 생물학 조직의 독창성을 거의 다루지 못했음에도, 물리적 체현과 외부 에너지에 의한 복잡성, 환경과의 상호 작용에 집중했던 후대 성향의 탈이원론적 면모가 이미 호메오스탯에 내재됐음을 가리킨다. 또한 이 인공 뇌가 사이버네틱스 초기 모델로서 등장했지만, 폰 포에르스터, 고든 패스크(G. Pask)가 2차 사이버네틱스의 특징으로 강조한 ‘관찰자’의 문제 역시 내포했음을 의미한다.¹²

2차 진영 논의 전개의 디딤돌은 투입-산출의 선형성과 여기서 생성된 ‘마치 ~처럼 ~한다’식의 모델이 실제 블랙 박스의 수수께끼를 해명하지 못한다는 전제였다. 바로 1차 모델의 맹점이었고, 모든 과학적 진술의 원천인 관찰자(실험자)를 지목해 그를 선험적일 수 없는 존재, 전체 계에 접속된 또 다른 성분으로 규정한다. 요컨대 관찰자로부터의 모든 진술은 그를 에워싼 환경과 인지 체계가 얹혀 구성되는 것이다. 그렇다면 가시 구간의 응시로써 존재를 목적론적 인과성으로 해명하거나, 정신의 본질을 정보/제어의 차원으로 환산해 기술하는 것은 은폐된 내부의 무지를 위장하는 선에서만 유효하다. 마찬가지로 ‘학습’이란 개념이나 알파고를 통해 맨 처음 살폈던 ‘창의성’ 역시 관찰자의 경험과 목적, 이해가 뒤엎킨 진술의 차원이자 내적 실체와 상관이 없다. 관찰하는 우리의 뇌가 기계의 선택을 창의적인 묘수로 해석했을 뿐, 실제 알파고의 내부는 다수의 확률 연산만을 수행하고 있었기 때문이다. 그렇다면 인간의 학습, 창의성 역시 마찬가지 아닐까? 관찰을 통한 추정과 내적 원리 사이의 무관계성, 호메오스탯의 작동 원리는 바로 이점을 시사하고 있었다.

1, 2차 성향의 차이를 비교한 여러 서술에서 애쉬비를 1차 사이버네틱스의 일원으로 간주했지만, 본 논의는 1차 진영에 터를 잡은 채 일궈낸 그의 성취가 사이버네틱스 다음 단계로의 도약에 어떤 역할을 했는지 초점을 맞추고자 한다.¹³ 특히 호메오스탯이 평형에 도달하기 위해 가동하는 ‘무작위’ 메커니즘과 이로써

¹² 사이버네틱스의 관찰자 문제는 다음을 참고하라. H. von Foerster, “Cybernetics of Cybernetics(1979),” *Understanding Understanding* (New York: Springer-Verlag, 2003), 283-286; G. Pask, “Introduction: Different Kinds of Cybernetics,” G. de Vijver, ed., *New Perspectives on Cybernetics* (Dordrecht: Kluwer Academic Publishers, 1992), 11-31.

¹³ B. Clarke and M. B. N. Hansen, “Introduction: Neocybernetic Emergence,” *Emergence and Embodiment* (Durham: Duke University Press, 2009), 1-25; Pask, op. cit.

구현되는 ‘초안정성(ultrastability)’을 논의의 중심에 놓음으로써 그가 펼친 연구의 본질에 접근할 것이다. 이들 두 개념은 기존 사이버네틱스 기계와의 차별성을 드러내면서도, 비교적 최근 인공 지능 갈래인 반표상주의 진영과 광범위하게 관점을 공유하고 있어 그가 후대에 끼친 학문적 여파를 확인할 수 있는 부분이기도 하다.

최종적으로 본 연구는, 최근 인공 지능 논의에서 다소 잊힌 감 있는 사이버네틱스의 기원을 확인해 그 원형적 발상에 접근하는 기회를 갖고, 이로써 이곳 논의의 역사적 시야를 넓히는 데 기여코자 한다.

1. 애쉬비 연구의 역사적 특수성

비글로(J. H. Bigelow): 그런데요, 평형에 도달하기 위해 이 기계가 펼치는 무작위적 탐색이 어떻게 학습에 상응한다는 거죠?

애쉬비: 난 그게 문제가 된다고 생각지는 않습니다. 당신 의견이나 내 의견이 결국 마찬가지로진데요.

비글로: 내가 묻고자 하는 건 이런 겁니다. 만일 미로에 어떤 동물을 풀어놔서 때, 그 동물은 통상 우리가 학습이라 일컫는 일을 벌입니다. 자, 과연 어떤 점에서 이 기계가 그런 유사한 일을 벌인다는 겁니까?

[중략]

영(J. Z. Young): 이 기계는 상태의 변화를 보여주지 못합니다. 학습에서의 본질은 어떤 과정을 거치면서 시스템이 이전과 다른 특성을 획득하는 것에 있습니다.

애쉬비: 동물이 뭔가를 배우고 난 후, 전과 다르게 행동한다는 점에는 동의하시죠?

비글로: 네, 동의합니다.

애쉬비: 그렇다면, 호메오스탯도 마찬가지인데요.

비글로: 대체 어떻게요? 설명해주시겠어요?¹⁴

다소 격양된 언쟁으로 읽히는 이 기록은 1952년 봄, 메이시 9차 회의에서 애쉬비가 발표를 진행하는 동안 그와 참석자들 사이에 주 받은 질의응답 일부다.¹⁵ 그는

¹⁴ W. R. Ashby, "Homeostasis," H. von Foerster, ed., *Cybernetics, Transactions of the 9th Conference, March 20-21 1952* (New York: Josiah Macy Jr. Foundation, 1953), 103-104. 이하 "Homeostasis"로 표기.

¹⁵ 각주 5 참고.

십 수 년 간 자신이 진행해 온 연구, 즉 환경과 상호 작용에 임하는 뇌 기능을 명시하는 동시에 이에 관한 가설이 집약된 인공 뇌 호메오스탯을 소개하고 있었다. 그러나 이곳 참석자들은 어이없다는 듯 표정을 지으며 계속 의혹어린 질문을 던졌다.¹⁶ 발표의 요점은 이랬다. 네 개의 유닛이 한 조를 이룬 단순한 외관의 기계가 작동하는 방식, 그리고 각 유닛 위에서 좌우로 움직이는 바늘이 중앙값을 찾아가는 무작위적 선택의 프로세스가 뇌가 환경에 적응하는 과정이며, 또 대뇌가 학습하는 원리에 상응한다는 것이다.

무슨 얘기일까? 상식을 뒤집는 발표였기에 당혹스런 분위기가 무리는 아니었고, 특히 참석자들 가운데 가장 격한 반응을 보인 비글로는 기계를 가리키며, “뭔가 그럴듯하게 복제해 놓은 것 같긴 하지만, 그게 무엇인지 누가 짐작조차 할 수나 있겠느냐”며 난감함마저 표명한다.¹⁷ 그가 일찍이 사이버네틱스의 성립에 있어 중요한 동기가 된 “행위, 의도, 목적론”의 공저자임을 고려할 때, 그의 이런 날 선 반응은 이곳 영토 내에서 애쉬비의 연구가 기존 담론의 무난한 연장선상에 있지 않음을 추측케 한다.¹⁸ 하지만 이 논문의 또 다른 저자이자 ‘사이버네틱스’를 조어하고 정초했던 위너(Norbert Wiener)가 자신의 기념비적 저서에 쓴, “학습 과정을 거쳐 스스로 목표를 탐색하는 비목적성의 무작위 메커니즘에 대한 애쉬비의 탁월한 발상이야말로, ... 위대한 철학적 공헌 가운데 하나”¹⁹라는 찬사를 고려한다면, 전체 학문 지형 내에서 그의 위치를 결정하기는 어려워진다. 이는 사이버네틱스 초기 흐름을 일군 대표적 두 사람으로부터 상반된 논평이 나왔음을 의미하기 때문이다.

그런데 후대의 평가를 살핀다면 당시 이런 견해차가 애쉬비의 연구에 내재된 복합적 성질에서 비롯됐음을 파악하게 된다. 일단 사이버 문화 비평가인 캐서린 헤일즈(Katherine Hayles)는 발표에 임한 그의 수사로부터 2차 대전의 외상을 감지한다. 그녀는 메이시 회의 개최 시기가 전후와 맞닿아 있는 점, 그리고 배에 승선한 엔지니어와 호메오스탯의 기능이 그의 언어 속에서 자연스럽게 병치된 곳을 짚어낸다.²⁰ 예컨대 엔지니어가 필사적으로 계기반을 응시하며 내부를

¹⁶ T. Froese, “From Cybernetics to Second-Order Cybernetics: A Comparative Analysis of Their Central Ideas,” *Constructivist Foundations* 5:2 (2010), 75-85 중 78.

¹⁷ “Homeostasis,” 95.

¹⁸ Rosenblueth et al., “Behavior” (cit. n. 6), 18-24.

¹⁹ N. Wiener, *Human Use of Human Being* (Boston: DaCapo Press, 1954/1988), 38.

²⁰ Hayles, “Boundary Dispute” (cit. n. 5), 23-25.

통제하는 모습, 이로써 유지된 배의 안정성이 다시 그의 생존을 위한 환경이 되고, 이 환경 내에서 연명하는 엔지니어의 체내 균형이 배의 순탄한 항해에 기여하는 상황을 시종일관 평형에 도달하려는 호메오스탯의 분주한 움직임과 일치시킨 것이다. 이런 동일시로 규정된 항상성이란, 환경의 우발적 변화에 대적하는 반엔트로피의 투쟁으로 완성되며, 이는 곧 전쟁의 대격변 속에서 치러지는 생사를 건 사투와 결부된다. 헤일즈는 호메오스탯으로부터 사이버네틱스 초기 모델의 보수적 편향을 감지한 것이다.

그러나 톰 프로스(Tom Froese)의 평가는 이와 다르다.²¹ 복잡계 연구자인 그에 따르면 당시 애쉬비의 발표와 메이시 참석자들 사이의 논쟁은 학문적 경향과 흐름에서 특정 시기의 종언을 고하는 장면이었고, 결국 이 뇌 기계는 사이버네틱스를 새로운 국면으로 전환시킨 주역으로 부상한다. 애쉬비가 그날 회의장에 불러일으킨 기성 진영의 노골적 반감은 그들이 미처 각성치 못한 부분을 표면화한 것에 대한 당혹감이었기에, 이후 생명과 기계의 사고를 연구하는 학자들로서 이에 대해 스스로 방어 기제를 발동시켜야 했다는 것이다. 이로써 그들은 모종의 선택을 강요받는 형국이 된다. 프로스는 ‘정보 처리’와 ‘통신’을 기반으로 완전한 물질 토대의 기계에 작인을 입혀 이를 마치 목표에 접근해가는 이성으로 여겼던 그들 전통이, 애쉬비의 공식 발표 앞에 무력화될 위험에 처했음을 기술한다.

그 강요된 선택이란 다음과 같다. 우선 물질 구조와 아예 결별한 채 주체의 층위를 구현하거나, 아니면 반대로 그 물질 저변의 틀 내에서 주체의 발현을 해명해야 하는 결정이 그것이다. 첫 번째는 이후 인지주의 기반의 기호 처리 인공지능 연구로, 다른 한 쪽은 마투라나, 바렐라(F. J. Varela) 등 산티아고 학파 주도의 인지 생물학 계보와 함께, 이를 포함한 2차 사이버네틱스의 흐름으로 경로가 잡힌다. 양측 모두 애쉬비 이후 그들 연구 체계의 합리적 기반을 새로이 다지기 위한 필연적 선택의 소산이란 것이다.

하지만 그의 연구가 기성의 흐름과 확연한 경계선을 긋거나 새로운 지경을 개척했다고 말하기엔 석연찮은 부분이 있다. 예컨대 로널드 클라인(R. Kline)은, 사이버네틱스의 초반부 20년 간 논의의 전형으로 애쉬비를 꼽는다.²² 생명

²¹ Froese, op. cit., 75-85.

²² R. Kline, “Where are the Cyborgs in Cybernetics,” *Social Studies of Science* 39:3 (2009), 331-362 중 351.

체와 기계의 불연속성을 부정하면서 신경계의 정신 작용과 전기·전자적 제어 계통을 하나의 유비로 묶는 모델을 제시한 대표적 인물이라는 측면에서의 평가다. 실제로 애쉬비는 발표 시작부터 “우리는 본질상 살아있는 쥐를 태엽 장치의 쥐와 유사한 존재로 간주할 수 있으며 동일한 물리 법칙, 일괄적 실증의 방법으로 양측 모두를 연구하는 것이 가능하다”라고 선언한다.²³ 곧이어 유기체가 벌이는 다채로운 행동과 기계의 움직임의 관통하는 저변의 원리를 파헤친다.

그가 살아있는 뇌의 기계적 복제를 학문적 목표로 삼을 수 있었던 것은 그 보편 원리에 대한 확신 때문이었는데, 만일 그 원리가 고스란히 구현된 모델을 제작할 수만 있다면, 동물이나 인간 행위를 지배하는 미지의 역학 또한 실험실에서 설계된 기계만큼이나 투명화 될 수 있을 것이다. 이로써 능동적 행위의 불가해한 구석은 사라진다. 그리고 이 목표 달성을 위한 첫 단계로서 애쉬비는 모종의 전제를 세운다. 모든 유기체가 향하는 최종 목표는 스스로의 ‘생존’이라는 것, 그리고 그 ‘생존이란 대체 무엇인가’에 관한 규정이었다.

간략히 살펴보면 이렇다. 시시각각 환경이 변화하고 때론 적대적으로 엄습할 때, 생명체가 감내할 궁극의 과제는 바로 생존이며, 개체의 소멸 직전까지 이 과업에 매달려야 한다. 그런데 이 생존이란 것은, 대사를 결정짓는 몇몇 핵심적 변수들이 한계치를 이탈하지 않도록 규제하는 것, 안정 범위 내에 이 변수들을 계속 묶어두는 것과 동일하다는 전제가 그것이다. 만일 이에 실패해 정해진 범위에 미달, 또는 초과가 일정 시간 이상 지속될 때 필연적으로 죽음이라는 대가가 따르며, 반대로 환경의 요동치는 변화에도 불구하고 이 변수들을 정상권 내에 가둘 수 있을 때 곧 ‘적응’이 성립한다. 그렇다면 뇌는 ‘항상성’ 유지에 총력을 기울여 개체의 적응을 추진하는 기관이며, 항시적으로 이들 변수를 붙잡아두기 위해 벌어진 오차의 만화에 매진해야한다. 호메오스탯은 이 같은 뇌 기능을 그대로 모의한 기계였다.

그런데 그가 역점을 둔 이 항상성의 문제가 사이버네틱스 초기 구성원들의 공통적 관심사와 일치한다는 점에 주의가 필요하다. 앞서 언급한 1943년 논문 “행위, 의도, 목적론”에서, 기계와 인간/동물의 행위 일부를 통합해 그것을 동일 범주로 묶어내기 위해 로젠블루스(A. Rosenblueth), 위너, 비글로(이하 R-W-B)가 택한 방법은 바로 ‘목적 추구 행위’를 규정하는 것이었고, 여기서의

²³ “Homeostasis,” 73.

핵심은 ‘교정 피드백’으로 그 행위의 본질을 압축하는 데 있었다. 여기엔 불규칙한 행동을 일삼는 타자들, 시간에 따라 변화를 꾀하는 외계의 유동성에 존재가 어떻게 탄력적으로 응대할 수 있는지에 관한 기계/생물학적 저변의 원리가 명시된다. 직전 행위가 빚어낸 결과로부터 목표와 떨어진 오차 신호를 획득, 이를 곧바로 이어질 다음 행위의 통제 정보로 되돌리는 교정 피드백이 그것이며, 이로써 주체는 외부의 총체적 변화에 굴하지 않고 행위의 결과를 목표한 바에 안착시킨다. 이 원리를 렌즈삼아 관찰한다면, 도망가는 쥐의 현재 위치와 도주로를 파악해 추격 방향과 속도를 조절함으로써 조금씩 격차를 줄여나가는 고양이 행동 패턴과, 이미 발사된 어뢰가 지정된 표적으로부터 시그널을 탐지해 향해 중인 적함을 자동으로 추적, 격침하는 행위는 모순 없이 일치하게 된다. 저자들은 “이 같이 균질화된 행동주의적 분석은 행위의 복잡성 정도에 상관없이 기계와 생명체를 가로질러 무차별 적용될 수 있다”고 말한다.²⁴

그렇다면 이들의 오차 감쇄 조절과 호메오스탯의 항상성은 과연 차별성이 있을까? 이 지점에서 애쉬비는 R-W-B가 터를 잡은 사이버네틱스의 동일 구역 내에 배속된다. 대표적으로 그레고리 베이트슨(G. Bateson)은 오류(오차), 또는 노이즈에 관한 태도를 놓고 애쉬비와 위너의 차별성에 회의를 품었다. 위너의 사이버네틱스에서 제어계의 효율은 전적으로 노이즈의 상쇄에 의해 좌우되며 그의 피드백 메커니즘은 줄기차게 목표와의 오차를 소거하는 데 매진한다. 베이트슨의 관점에서 호메오스탯 역시 이와 본질상 차이가 없다.

호메오스탯에서 오차는 회로 전체에 걸친 진동을 통해 감쇠한다. 바꿔 말하면 오차는 회로의 모든 구간을 여행하는 동안 점차 쇠잔하여 소멸해버린다. 이는 교정 피드백과 일치한다고 할 수 있다.²⁵

극복 대상으로서의 오차라면 시스템은 한계에 갇힌다. 1973년 지인에게 쓴 편지에서 베이트슨은 다음 견해를 밝혔다. 생물학적 조직에서 노이즈는 적정 조건하에 새로운 질서를 촉발하며, 만일 적응에 있어 그 오차를 포섭할 수 없거나 노이즈라는 원천을 묘용(妙用)하지 못한다면, 이에 내재된 방대한 변화의 잠재

²⁴ Rosenblueth et al., “Behavior” (cit. n. 6), 22.

²⁵ P. Harries-Jones, *A Recursive Vision* (Toronto: University of Toronto Press, 1995), 115에서 재인용.

성은 소실된다. 그렇다면 노이즈는 배척해야 할 대상이 아니라, 외계의 변동을 수용하는 적응의 실마리이자, 조직의 창조적 대응력을 개방하는 열쇠다. 애쉬비의 모델은 이런 면에서 경직되어 있다는 것이다. 그는 벌어진 오차를 생존의 위협으로 간주했고, 그의 호메오스탯은 적응 불능에 빠진 ‘나쁜’ 조직에서, 적응 가능한 ‘좋은’ 조직으로의 전환을 통해 상시 안정을 모색하기 때문이다.²⁶

애쉬비가 자기 조직화의 원리를 규정하고, 이후 이 개념의 부적절성을 다시 언급했던 것은 이와 관련이 있다. 그의 1947년 논문은 사이버네틱스 그룹 처음으로 신경계를 자기 조직화로 기술하면서, 구조적 결정성에 위배된 없는 재조직화의 가능성을 수학으로 입증하고 있었다.²⁷ 그러나 1962년 논문 “자기 조직화 시스템의 원리(Principles of the Self-Organizing System)”에서는 이 개념의 본령을 해체하는 데 주력한다.²⁸ 그에 따르면 자기 조직화는 엄밀한 의미에서 성립 불가하다. 조직의 변화는 항상 다른 시스템과의 결속 하에 성립된다는 것, 이렇게 전체 배치에 끼워진 채 진행되는 변화는 필연적으로 접촉 상대(혹은 환경)의 힘에 의존한다. 그렇다면 조직화의 영토는 그 자체가 아니라 경계가 맞물린 전체이므로, 이것을 고립된 체계의 자율 변화라 말할 수 없다. 더구나 그가 일컫는 조직화는 변이나 초월에 대한 약속이 없다는 점에서 특별하다. 이미 주어진 내적 요건 하에서 성분 간 연결과 배치만이 바뀌므로 이는 선결된 가능성의 실현일 뿐 발생적 변화가 아니기 때문이다. 이를테면 과학사학자 피커링(Andrew Pickering)이 “우리에게 호메오스탯이 제공하는 것은 오직 선택과 조합 뿐”이라면서 그 한계를 지적한 것은 이와 연관된다.²⁹ 같은 취지에서 프리초프 카프라(Fritjof Capra)는 애쉬비와 애쉬비 이후의 자기 조직화 모델 간 차이를 이렇게 정리한다.

사이버네틱스의 자기 조직화 초기 개념과 보다 정교하게 다듬어진 후기 모델 사

²⁶ W. R. Ashby, “The Physical Origin of Adaptation by Trial and Error,” *The Journal of General Psychology* 32:1 (1945), 13-25.

²⁷ W. R. Ashby, “Principles of the Self-Organizing Dynamic System,” *The Journal of General Psychology* 37:2 (1947), 125-128.

²⁸ W. R. Ashby, “Principles of the Self-Organizing System,” H. von Foerster and G. W. Zopf, eds., *Principles of Self-Organization* (London: Pergamon Press, 1962), 255-278.

²⁹ A. Pickering, *The Cybernetic Brain* (Chicago: University of Chicago Press, 2010), 154.

이에서 발견된 가장 중요한 차이는, 후자의 과정이 새로운 구조, 전과 다른 행동 양식의 발생을 포함한다는 점이다. 애쉬비에게 있어 모든 구성의 변화는 주어진 다양성의 여지 내에서 일어나며 시스템의 생존 가능성은 그 여지가 확보한 풍부함이나 필수 다양성에 의해 좌우된다. 여기에선 발생과 성장, 진화가 없다. 반면 후기 모델들의 경우 성장과 학습, 그리고 진화의 과정에서 새로운 구조와 행동 양식의 출현을 포함한다.³⁰

호메오스탯은 애쉬비 초기 이론의 결정체였던 만큼, 뇌의 조직화를 명쾌하게 해명하지만, 이렇게 기계의 동역학과 수학적 방법론의 공조가 치밀했기에 치려야 할 대가가 있었다. 그것은 이 해명이 생명체를 비롯한 자연 발생 조직의 성질을 섬세히 포착하지 못한다는 점이다. 이를테면 생명체 네트워크의 재생산과 자연계의 창발과 같은 현상을 다루기엔 추상적이며 너무 형식화되어 있었다는 것이다. 그러나 동시에, 그 한계가 사이버네틱스를 새로운 국면으로 이끌었다는 견해 또한 성립한다. 정치사회학자 장-피에르 더피(Jean-Pierre Dupuy)는 애쉬비의 1962년 논문의 파장을 다음처럼 해석한다.

이 “증명”, 즉 자기 조직화의 불가에 대한 주장은 62년 애쉬비에 의해 공식화되었다. 이는 특히 2차 사이버네틱스에 상당한 영향을 미쳤고, 이 학문적 경향은 그의 논거가 세운 장애물을 극복하는 방식으로 자기 조직화를 재고해야만 했다.³¹

더피는 그의 이 부정적 증명을 두 가지 패러다임의 상징적 분기점으로 지목한다.³² 그 중 하나는 “자기 생산(*autopoiesis*, 마투라나·바렐라)”, 다른 하나는 “노이즈로부터의 질서(폰 포에르스터)”였다. 양측 모두 수학 기반의 논리 형식 체계나 인공으로 제작된 기계, 프로그래머의 알고리즘에서 벗어나 어느 초월자의 설계가 없는 자연 생성 조직을 1차 모델로 삼았고, 이로써 생명체의 세포 네트워크나 자연계의 물리 현상, 또는 사회 조직의 발생과 진화에 대한 논의가 본격화될 수 있었다. 이렇게 애쉬비가 모순으로 귀결시킨 자기 조직화의 장벽을

³⁰ F. Capra, *The Web of Life* (New York: Anchor Books, 1997), 85.

³¹ J-P. Dupuy, *On the Origins of Cognitive Science* (Cambridge, MA: MIT Press, 2009), 152.

³² J-P. Dupuy, “On the Supposed Closure of Normative Systems,” G. Teubner, ed., *Autopoietic Law: A New Approach to Law and Society* (Berlin: European University Institute, 1988), 51-69 중 56.

우회할 수 있게 되자, 교정 피드백과 목적론적 제어로 무장한 기성의 사이버네틱스가 생명 조직의 자율성을 해명하는 데 역부족일 수 있다는 각성 또한 비등한다. 바로 2차 사이버네틱스의 흐름이 활성화되는 지점이었다.

그러나 유의할 사항은, 이 분야 역사에 섬세히 접근할수록 ‘애쉬비의 극복’으로서의 2차 사이버네틱스가 아니라, 오히려 그 여파와 유산을 강조한 서술의 비중이 높다는 점이다. 더피가 60년대 생성된 양대 흐름 중 하나로 지목한 마투라나의 학문적 성취는, 프로스·스튜어트의 논의에서 호메오스타트의 전제하에 일궈진 성과로서 피력된다.³³ 애쉬비가 지능적 행위의 근간을 파헤치기 위해 신경계의 작용에 집중했다면, 마투라나는 그의 논의를 생물학적 대사 전체로 확대시켰다는 견해다. 한편 다른 논의들은, 자극에 개방됐었지만 그 작동은 닫혀있는 자기 조직화 시스템의 폐쇄적 성질이 그에 의해 선택됐다는 점,³⁴ 관찰자가 시스템 일부로 포함돼야할 당위 역시 애쉬비의 논의로 제기됐음을 지적한 바 있다.³⁵

2. 적응의 계량화

9차 회의가 열린 그 해, 『뇌의 설계』가 출판되면서 그 원리와 면모가 상세히 소개되지만, 호메오스타트의 완성은 4년 전의 일이었다.³⁶ 종전 후 3년이 경과한 시점이

³³ T. Froese and J. Stewart, “Life After Ashby: Ultrastability and the Autopoietic Foundation of Biological Autonomy,” *Cybernetics and Human Knowing* 17:4 (2010), 7-50.

³⁴ S. A. Umpleby, “Ross Ashby’s General Theory of Adaptive Systems,” *International Journal of General Systems* 38:2 (2009), 231-238; B. Scott, “Second-order Cybernetics: An Historical Introduction,” *Kybernetes* 33: 9/10 (2004), 1365-1378; R. Glanville, “Second Order Cybernetics,” F. Parra-Luna, ed., *System Science and Cybernetics, Vol. III* (Oxford: Eolss Publishers, 2009), 59-86; “Black Boxes,” *Cybernetics and Human knowing* 16(1/2)(2009), 153-67; “Radical Constructivism and Second order Cybernetics,” *Journal of Space Philosophy* 1 (2012), <http://goo.gl/YM7SjP> (2016. 12. 15 접속).

³⁵ 이들의 판단은 1960년대까지 이어지는 애쉬비의 저술들을 고려한 결과이지만, 본 연구에서는 호메오스타트를 중심으로 두고, 이 모델이 성립되기까지의 1940년대, 메이시 발표 직후의 저술들을 중심으로 살펴도록 할 것이다. 호메오스타트 자체가 모델로서 이런 논의들을 포함하고 있었기 때문이다.

³⁶ W. R. Ashby, *Design for a Brain: The Origin of Adaptive Behavior* (New York: John Wiley & Sons, 1960). 1952년 출판 후 1954년 재판이 발행되었고, 1960년에 개정판이 나왔

었고, 정신과 의사인 애쉬비가 글로체스터의 반우드하우스 병원 연구 책임자로 있을 당시, 영국 공군의 폭탄 제어 장치를 변통해 틈틈이 제작한 결과물이었다. 2차 대전 직후, 함포 사격과 대공 화기 조준을 위한 건 디렉터, 탄도 계산을 위한 기계식 컴퓨터가 다수 생산돼 인간 계산원들의 집단 작업을 대체해나가던 당시의 상황을 의식한 듯, 애쉬비는 이 책에서 자신의 기계가 다른 계산 전담 기계들과 전적으로 차별됨을 밝혔다.

계산기 설계자는 특정 유형의 임무를 수행하는 기계, 살아있는 뇌보다 맡은 바를 더 뛰어나게 처리할 수 있는 기계를 원한다. 이 기계가 실제 뇌의 방식을 따르고 있는지는 그에게 부차적 문제다. 반면 나의 목표는 실제의 뇌를 그대로 복제하는 것이다. 만일 살아있는 뇌가 어떤 특정 방식으로 실패한다면, 나의 인공 뇌 역시 실패하길 원한다. 왜냐하면 이 같은 실패가, 구현된 모델이 진본에 위배되지 않는 복제임을 입증하는 증거이기 때문이다.³⁷

여기서 성패가 갈리는, 호메오스탯이 풀어야 할 숙제는 스스로의 ‘적응’이었다. 뇌의 인지 능력 일부를 자동화하거나 그것을 확장하려는 시도가 아니라, 환경과 대면한 개체의 투쟁과 순응의 프로세스를 모의하는 것이 목표였고, 추가적으로 뇌가 적응에 실패하는 경우까지 기계로 해명할 수만 있다면 동물과 인간 뇌 고유의 생물학적 메커니즘을 가감 없이 재현했음이 증명된다는 계산이었다. 그는 주어진 환경에서의 적응을 인간/동물 행위의 저변을 관통하는 동기로 전제하면서, 생리학자 제닝스(H. S. Jennings)의 말을 빌려 “유기체의 행위는 자신의 복리를 증진하기 위해 빚어진다”는 점, 그리고 “환경이 변할 때 유기체 역시 그 새로운 조건에 응대해야 함”을 확인시킨다.³⁸ 따라서 변화와 자극에 따라 개체 스스로를 보존하기 위한 반응과 선택 모두가 적응의 행위다. 이를테면 허기를 채우기 위해 먹이를 사냥하려는 시도, 갈증 해소를 위해 물을 찾는 행동, 산소 결핍을 만회하기 위해 숨을 가쁘게 몰아쉬는 행위, 열기나 추위를 피해 은신하려는 선택 모두는 변화에 따른 응대며 만약 그것이 성공적일 때 적응이 성사된다고 할 수 있다.

다. 본 연구에서는 1960년 개정판을 바탕으로 하되 필요에 따라 1954년 판을 활용한다. 이하 DB로 표기, 1954년도 판의 경우에만 DB(1954)로 한다.

³⁷ DB(1954), 11/1.

³⁸ Ashby, “Physical Origin” (cit. n. 26), 14.

그런데 이 같은 활동들을 조금 다른 방식으로 기술해 보자. 이들 행위는 소모된 칼로리를 신체에 재공급하고 혈당치나 혈액 pH 농도를 조절, 또는 체내 수분 비율이나 체온을 적정 수준으로 맞추려는 시도이며, 세포 내 산소 농도를 일정하게 유지하기 위함이다. 그렇다면 적응의 움직임 저변은 이처럼 체내 요소들이 외적 요인으로 변화를 겪을 때, 그것을 원래의 상태로 복구하려는 시도로 일관된다. 애쉬비는 그 복구 상태를 파블로프의 말을 빌려 ‘평형’으로 규정한다.

명백히 규정된 물리 체계로서 동물은 그 외부로부터 가해지는 힘과 지속적 평형을 유지하는 선에서 존재를 지속시킬 수 있다.³⁹

연구를 진행하면서 그가 최우선의 과제로 삼은 부분이 이 ‘적응’이란 개념을 다른 용어로 치환하는 문제였다. 개념 정의의 명료성에 집착하는 애쉬비 특유의 강박이 발휘된 1940년도 논문 “적응과 평형”에서, 그 의미가 막연하다는 점, 그리고 전혀 계량적이지 못하다는 이유로 이 용어가 과학적 논의에 부적합함을 피력했고, 따라서 이것을 ‘안정 평형(stable equilibrium)’으로 대체할 것을 제안한다.⁴⁰ 논문에 따르면, 평형은 개념으로서 “1)순수히 객관적이며, 2)목적에 관한 모든 형이상학적 모호함을 일소할 뿐만 아니라, 3)정의에 있어 명확하고, 4)그 자체로서 계량적 연구에 곧바로 적용될 수 있기 때문이다.”⁴¹ 외부 변인으로 대상에 가해진 효과가 전체 합력 내에서 다시 0으로 상쇄됨을 일컫는 이 역학 개념을 생물학적 적응과 일치시킴으로써, 이제까지 심리 활동으로 간주된 뇌의 문제를 공학적 분석과 수학적 해명의 영토로 편입시킨 것이다. 일단 동역학 프로세스를 분석하는 기본 도구로써 뇌를 향한 접근로가 개척된 상태라면, 동일 원리를 적용해 그것을 기계 장치로 모델링하는 것 또한 가능하므로 이 같은 개념 치환은 차후 연구 행보를 지탱하는 불가결한 포석이었다. 그리고 여기서부터 뇌는, 자체 구성 요소들이 서로 얽혀 영향을 주고받도록 조직화된 기계 시스템

³⁹ I. P. Pavlov, *Conditioned Reflexes* (1927). W. R. Ashby, “The Nervous System as Physical Machine: With Special Reference to the Origin of Adaptive Behavior,” *Mind* 56:221 (1947), 44-59 중 46에서 재인용.

⁴⁰ P. A. Cariani, “The Homeostat as Embodiment of Adaptive Control,” *International Journal of General Systems* 38:2 (2009), 139-154 중 140.

⁴¹ W. R. Ashby, “Adaptiveness and Equilibrium,” *Journal of Mental Science* 86:362 (1940), 478-483 중 482.

과 완전히 일치하는 것으로 간주될 수 있었다. 요컨대 안정 평형에 도달한 기계는 환경에 적응한 것이다.

궁극적으로, 모든 물리적 본체에 상시 작용하는 작은 요동에도 불구하고, 계속 존속하는 시스템이 있다면 그 시스템은 안정 평형 상태에 있음이 틀림없다고 일말의 의심 없는 결론을 이끌어낼 수 있다. 독단적으로 들릴 수 있을지 모르지만, 나는 이 같은 추론에서 벗어나는 사례를 찾을 수 없었다.⁴²

애쉬비에 의해 종종 인용됐던 “내부 환경(milieu interieur)의 고정성이야말로 자유롭게 독립한 생의 조건이다”라는 명제는 19세기 생리학자 클로드 베르나르(C. Bernard)가 제시한 이론을 함축한다.⁴³ 그리고 그의 연구를 토대로 월터 캐논(Walter B. Cannon)은 1932년에 항상성(homeostasis)을 제시했다.⁴⁴ 예컨대 혈당치가 높아지면 갈증을 부추겨 물을 섭취케 하고, 이로써 증가된 체내 수분이 당의 응집을 희석하는 원리처럼, 신체는 그 어떤 부조화를 유발하는 변화의 경향이 감지되면 자동적으로 이를 저지하는 요인과 충돌시켜 상대적 부동의 정상 상태(steady-state)를 유지한다는 것이 개념의 요체였다. 내부 환경에서 빚어지는 변화와 불균형에 대한 저항, ‘호메오스탯’이란 명칭은 캐논의 이 개념에서 비롯됐으며, 이로써 ‘안정 평형’, ‘적응’, ‘항상성’은 그의 논의 상당 부분에서 동일 의미로 묶인다. 캐논이 환경의 치명적 공격에도 견디도록 설정된 안전 계수의 범위를 언급했던 것처럼, 애쉬비에게 이 안정성의 영토는 모든 핵심 변수(essential variables, 이하 EV)가 정상 범위(생리학적 한계치)를 이탈하지 않는 위상 공간으로 규정된다.⁴⁵ 여기서 뇌는 EV의 변화에 맞서 그들이 적정 범위를 이탈할 경우 보정 메커니즘을 맹렬히 가동한다. 결국 이 같은 방비 체계는 환경의 요동과 급격한 체내 변화에 직면하더라도 비교적 미미한 장애만 감수하게 하면서 시스템 전체 기능을 계속 유지시킨다.

여기서 EV는 신체상 모든 변수가 아니라 계의 안정성을 좌우하는 요인만이 포함된다. 그리고 이 위중함은 부단히 평형을 추구하는 시스템으로 하여금 복잡

⁴² Ibid.

⁴³ P. De Latil, *Thinking by Machine* (London: Sidgwick & Jackson, 1956), 295.

⁴⁴ 월터 B. 캐논, 정해영 옮김, 『인체의 지혜』 (동명사, 2003).

⁴⁵ 여기서 그 위상 공간은 핵심 변수로 이뤄진 다차원의 영토로서 각각의 변수가 계의 상태를 표시하는 실수의 축들로 축조된다.

한 행동을 일으킨다. 예컨대 혹한 속에서 체온이 떨어지면 인간은 자신의 팔과 팔을 엇갈려 가슴을 압박하듯 자세를 취하고 무릎을 가슴으로 당겨 허벅지와 복부를 밀착시킬 것이다. 이런 태도는 노출되는 피부 면적을 최소화함으로써, EV로서의 체온을 가까스로 정상 범위 안에 가둘 수 있다. 그러나 추위가 지속되는 가운데 주변에 체온을 복구할 만한 장소가 포착되면 즉시 그곳으로 위치를 옮기는 것은 필연적이다. 이렇게 변화에 대한 모든 대응은 상황의 구체성으로 좌우되며, 환경이 제공하는 것과 제한하는 것 사이의 특수성을 규정함으로써 조직의 대응이 어떠한지를 명시할 수 있다. 그가 지적하는 바는, 유기체를 향한 공허한 입출력의 기술을 지양하는 것, 환경의 요건과 이에 대한 반응을 묶어 실질적으로 논의하자는 것이다.⁴⁶

행동에 있어 그 어떤 목적론적 해석도 없을 것이다. 기계나 동물이 어떤 순간에 특정 방식으로 행동을 취하는 이유는 바로 그 시점의 물리, 화학적 성질이 그 밖의 다른 행동을 허용하지 않기 때문임을 전제하려한다. ... 왜냐하면 우리의 취지는 마치 목적론 차원에서 지시된 것처럼 여겨지는 행위의 기원을 해명하는 것에 있기 때문이다.⁴⁷

그런데 여기서 그가 가리키는 행동의 단위는 반드시 개체의 것일 필요가 없으며, 신체 각 기관들의 생리학적 반응이나 신경 세포 집단 사이의 상호 작용일 수도 있다. 기본적으로 호메오스탯은 대뇌 신경 조직의 시뮬레이션을 표방하지만, 그 설계는 환경과 조직이 맺는 관계의 총체를 대변하는 것으로 간주하는 편이 정확하다. 이 둘 사이를 결속하는 명분이 적응이며 이 적응의 메커니즘을 객관화함으로써 기계와 동물/인간의 생물학적 조직을 균질화할 수 있었던 것이다. 한편 환경은 일반 의미로서의 외적 요인과 차이가 없지만, 이 영토 역시 다른 개체들을 포함하면서 이들 전체의 수많은 변수가 맞물려 끊임없이 균형점을 찾기 위한 부산한 활동이 펼쳐진다. 그렇다면 이 현장을 단일 개체의 내부로 지정할 수도 있을 것이다. 조직 구성 성분들이 서로가 서로에게 환경으로 작용하면서 각각의 변화를 자극으로 수용하는 것은 마찬가지이기 때문이다. 따라서 애쉬비가 말하는 조직의 경계는 유연하며, 환경의 변화와 이에 대응하는 주체의 규

46 "Homeostasis," 73. 애쉬비는 시종일관 개체가 처한 상황의 구체성에 역점을 두어 뇌의 기능을 설명한다.

47 DB, 1/13.

정 역시 상대적이다. 그리고 호메오스탯의 독특한 열개가 그의 이 같은 관점을 표명한다.

3. 열개와 작동, 초안정성의 원리

1949년 1월의 『타임』지 기사 “과학: 생각하는 기계”는 다음 내용을 싣고 있었다.⁴⁸ 입력된 명령이 아니라 주어진 상황 변화에 맞춰 자체 행동을 취한다는 점에서 호메오스탯을 생각하는 기계로 간주할 수 있다는 애쉬비의 설명을 옮겼고, 독자의 이해를 배려한 듯 의인화해, 스스로 “불편해지면” 다수가 얹힌 전기 회로가 그 불편함에서 벗어나기 위해 재조합과 변형을 반복한다는 해설을 덧붙였다. 그러나 실상 그에게 쟁점은, 기계가 정말 사고하는지 여부가 아니라 뇌를 둘러싼 양립 불가능한 견해 양측을 모순 없이 충족시킬 수 있느냐는 점이었다.⁴⁹ 뇌는 전기적 작용으로 얹힌 정밀한 회로라는 점에서 그 자체가 기계라는 관점, 다른 하나는 생명체라면 한결같이 환경에 적응하기 때문에 뇌는 이를 위한 행동을 생산해야 한다는 견해가 그것이다. 당시까지 이들 두 관점은 화해하기 어려웠다. 환경의 변화에 적응하지 못한다면 뇌가 아니며, 따라서 뇌는 기계일 수 없다고 중론이 기울어가는 상황에서, 그가 스스로에게 부과한 숙제는 바로 환경에 적응하는 기계의 설계였다. 이것이 가능하다면 뇌는 결국 기계일 수 있다.

기사의 표현을 빌리자면, 구성 요소인 4개의 자석 모두가 중립에 위치할 때 호메오스탯은 “편안해진다”. 회전축으로 중심이 고정된 이 자석들이 좌우로 요동치거나 어느 한쪽으로 치우침을 반복하는 불안정 상태를 지나 마침내 ‘안정 평형’에 도달했을 때, 그 회전축에 연결된 와이어 끝의 작은 금속판 역시 그 회전을 따라 ‘o’를 가리킨다. 그렇다면 이 안정과 불안정의 기제는 어떠한가? 이 금속판은, 점성을 띤 용액이 담긴 부채꼴 모양의 수조에 반쯤 잠겨있고 통의 양 끝에는 -2V와 -15V의 전위차가 생성되어 있으므로, 자석의 회전으로 금속판이 접근하는 방향에 따라 이 범위 내에서 전압이 가해진다. 이렇게 금속판이 감지한 전압은 진공관에 전해진다.⁵⁰ 이어서 이 진공관은 다시 코일을 통과하게 될

⁴⁸ “Science: The Thinking Machine,” *Time* (24 January, 1949).

⁴⁹ DB, 1/1.

⁵⁰ 따라서 자석과 연결된 금속판은 그 좌우 편향 정도와 진동 빈도를 통해 호메오스탯의 상태를 모니터링 할 수 있도록 정보를 주지만, 그 자체로서 유닛 사이에서 벌어지는 상호 작용 전반의

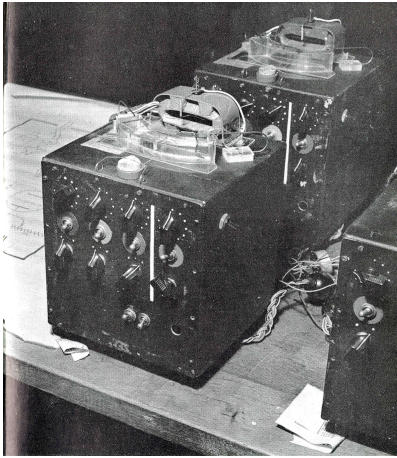


그림 1. 호메오스탯의 외관. 총 4개의 유닛이 연결되어 전체를 이룬다. 전면부에 배치된 8개의 패널 스위치로 매개 변수를 조작할 수 있다. 출처: De Latil, *Thinking by Machine*, Facing p. 274.

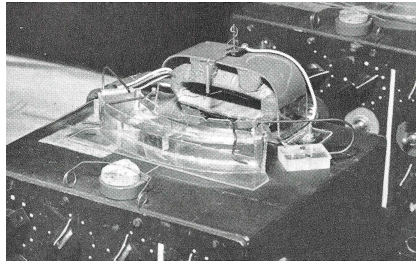


그림 2. 유닛의 상단부. 호메오스탯의 적응 프로세스가 실질적으로 진행되고, 그 과정을 확인할 수 있는 곳이다. 앞쪽의 투명 수조에 담긴 작은 금속판에 전압이 걸리면 자신과 다른 3개의 코일에 전류를 흘려보내고 이에 따라 그 코일에 둘러싸인 자석들이 회전한다.

출력 전류를 제어한다.⁵¹ 요컨대 하나의 유닛으로부터 방출돼 코일로 흐르는 전류의 세기는 금속판(이하 바늘)에 걸린 전압에 의해 좌우되며, 이는 자석이 중립에서 이탈한 정도를 반영한다.⁵² 그런데 이 자석은 코일로 둘러싸여 있고 그 코일에 걸리는 전류가 자석을 조종하기 때문에, 자석의 움직임으로 발생된 전류가 다시 자석의 회전을 결정하는 순환 구조가 된다[그림 1, 2].

그런데 문제는 총 4개의 서로 다른 유닛이 전체를 구성하는 만큼, 단일 코일로 유입되는 입력 전류는 자체 유닛 출력뿐만 아니라 각기 다른 3개의 유닛에서

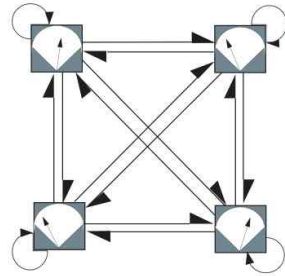


그림 3. 전체 유닛들이 주고받는 상호 작용의 양상. 각 유닛의 출력은 여타 다른 유닛들에 영향을 끼치는 동시에 자신에게 되돌리는 재귀 피드백의 성격을 띤다. 출처: <http://goo.gl/X5d2bv> (2016. 12. 15 접속)

기축을 담당하고 있다.

⁵¹ 각 유닛에는 항상 전위차가 설정돼 있지만 자석이 중립에 놓이면 저항에는 항상 일정한 전류가 걸린다. 이런 디폴트 상태에서 출력 회로는 전혀 전류를 발생시키지 않는데, 자석의 좌우 움직임에 따라 저항에 이보다 더 많거나 적은 전류가 걸리게 되면 출력 회로는 이 차이를 반영해 다른 유닛으로 전류를 흘려보낸다.

⁵² 부채꼴 물통과 여기서 좌우로 움직이는 와이어 끝의 금속 판은 마치 검류계의 표시창, 그 안에서 움직이는 바늘과 유사한 형태를 띠므로 이후 ‘바늘’로 통일한다.

방출된 전류 모두가 합류한 것이기 때문에 전체 피드백 순환이 복잡하게 얽힌다는 점이다. 이를테면 A, B, C 3개 유닛에서의 출력은 D의 코일로 유입되고 D가 방류한 전류 역시 마찬가지로 A, B, C로, 그리고 재귀 피드백으로서 D 자신에게로 되돌려지기 때문이다. 다른 유닛들 역시 D와 마찬가지로 여타 세 곳과 자신에게 출력을 배분하는 구조다. 이런 난잡한 흐름 속에서 서로가 입출력의 충돌을 반복하다 우연히 균형을 이룬 지점, 앞서 언급한 그 ‘편안함’의 상태에서 호메오스탯의 움직임은 잦아들고 마침내 적응이 성사된다[그림 3].⁵³ 이 기계를 관찰한 드 라틸(De Latil)은 유리통 속에 갇힌 바늘이 마치 우리 속 야생 동물처럼 발버둥 친다면서 다음과 같이 묘사한 바 있다.

한 유닛에서의 변동은 다른 모두에게 적응해야 할 짐이 되고, 그 각자의 적응은 다시 모두에게 반응의 의무를 지운다. 행동과 반응, 상호 작용이 얽혀 홍수로 귀결되는 상황은, 논리적으로 연쇄적 이벤트를 설계하려는 그 어떤 이성의 힘 따위를 넘어선다. ... 그들이 전체 평형에 도달할 수 있을지는 그저 추측의 문제일 따름이다. 각 유닛은 스스로와 전체를 위해 균형을 탐색해간다. 그 일은 마치 시지푸스와 다나우스의 딸에게 지워진 끝없는 노동과도 같아서, 도달에 대한 연약 없는, 목표를 향한 시련의 노동으로 보일 것이다.⁵⁴

그러나 악무한 고투가 이어지는 이 시련의 땅에 모종의 성질을 입힐 수는 있다. 전류는 코일에 들어가기 직전 정류자와 가변 저항기를 거치는데, 이들 각각은 코일 어느 극으로 전류가 진입하는지, 어떤 비율로 전류가 코일을 통과할지를 결정한다. 각 유닛 앞면에 위치한 패널 스위치(유닛마다 8개씩, 총 32개)를 오퍼레이터가 수동 조작함으로써 이들 매개 변수가 결정되고 이로써 호메오스탯 전체의 판이 바뀐다. 만일 특정 유닛의 스위치를 조작해 안정 배치를 적용하면 해당 바늘은 신속하게 중앙값으로 복귀하려는 성향을 띄지만, 그 반대의 배치에서 바늘은 정상권을 이탈해 좌우로 폭주한다. 하지만 이 반생체로서의 기계는 이런 극악의 변동까지 완충하는 체계를 구비해 자체 안정화를 추진해낸다. ‘유니셀렉터[uniselector, 혹은 스텝핑 스위치(stepping-switch)]’로 명명된 이

53 이로써 물리 역학에서의 평형, 즉 물체에 작용하는 힘이 전체 합력 안에서 제로가 되어 정지 상태가 유지된다는 정의 그대로가, 애쉬비가 의도한 대로 안정화에 돌입한 호메오스탯의 평형으로 구현된 것이다.

54 De Latil, *Thinking by Machine* (cit. n. 43), 299.

장치는 바늘이 중립에서 45도 이상 좌·우로 벗어난 채 일정 시간을 초과할 때마다 자동으로 내부 스위치의 배열을 바꿔나간다. 여기에는 난수표가 적용돼 각 유닛마다 25개의 포지션(4개의 유닛이므로 $25^4 = 390,625$ 개의 조합이 가능) 중 하나를 계속 무작위로 선택하게 된다.⁵⁵ 이것이 급격한 변화에 대한 면역 체계일 수 있는 이유는, 이 자체가 앞서 말한 오퍼레이터의 패널 스위치 조작과 동일한 성질의 매개 변수로 역할하면서 그 충격파를 상쇄하기 때문이다. 기계 외부의 패널 스위치 조작이 호메오스탯의 행동 양상을 뒤바꿨듯, 유니셀렉터가 선택한 스위치 배열 역시 이에 맞서 환경과의 상호 작용 패턴을 새로이 규정한다는 것이다.

그가 이 체계를 ‘스텝 메커니즘’으로 칭한 이유는, EV가 한 쪽으로 치우쳐 생리적 대사에 위기가 도래한 시점에만 단속적으로 가동하기 때문이다. 그런데 주목할 사항은, 적응에 적합한 스위치 조합이 지속적 시도와 실패를 거쳐 우연히 찾아진다는 점이다. 무작위 선택인 만큼 안정화를 성사시킬 조합이 발견될 때까지 연이어서 무작정 탐색된다.⁵⁶ 그리고 이렇게 계속 갱신되는 스위치 조합에서 빚어진 새로운 행동 패턴이 어찌다가 EV를 정상치로 복귀시킬 때, 비로소 그 선택이 관철되는 원리이다.⁵⁷

이들 새로운 값[유니셀렉터에 지정되는 수치]은 이전 값과 전혀 연관성을 갖지 않으며, 당장에 겪는 문제와도 특별한 상관 관계가 없다.⁵⁸

신경 조직이 이 같이 작동한다는 것이다. 보다 명백한 예시로서 애쉬비는 상자 안에 쥐를 집어넣고, 바닥면 그릴에 전기 충격을 가하는 실험을 제시한다. 내부에 설치된 페달을 밟으면 전기가 차단돼 쇼크로부터 벗어날 수 있지만, 쥐는 아직 이것을 모른다. 그릴에 걸린 전압을 서서히 높여갈 때, 처음 얼마간 특별한 변화를 보이지 않지만, 어느 수위를 넘어선 순간 쥐는 격렬하고 다급하게

55 피셔-에이즈의 표(Fisher-Yates' Table of Random Numbers)를 적용했다.

56 하나의 배열이 선택됐을 때, 일정 시간 내에 평형을 되찾지 못하면 다음 배열이 계속 시도된다. 위기가 해결되지 않는 한 유니셀렉터는 다음, 그 다음 랜덤 값을 적용해 스위치 조합을 바꾸고, 안정화에 이를 때까지 과정을 반복한다. 이런 이유로 그는 기계가 적정 유니셀렉터 조합을 ‘사냥’한다고 표현한다.

57 물론 하나의 유닛이 정상권에 안착하더라도 4개의 유닛 모두가 안정 평형에 이르지 못한다면 그 배열은 포기되고 탐색은 계속 이어지게 된다.

58 DB, 8/3.

온갖 행동을 취한다.

전기적 자극이 가해지면 쥐는 점프하거나 뛰어다니고, 비명을 지르기도 하고, 그릴을 물어뜯거나 발작하듯 버둥거리면서 계획되지 않은 다양한 행동을 내놓는다. 머지않아 쥐는 페달을 밟아 쇼크를 멈출 것이다. 열 번의 시도 끝에 전기 쇼크의 적응은 쥐로 하여금 페달로 직행시키고 그것을 누르게 만든다.⁵⁹

전기 충격으로 야기된 감각 피질에서의 변화가 일정 수준을 초과할 때, 운동 피질은 서로 다른 패턴의 신체 움직임을 연쇄적으로 일으킨다. 이는 뇌가 자극을 감내하는 동안, 스텝 메커니즘을 통해 내부를 계속 바꿔나간 결과다. 이 와중에 우연히 페달을 밟아 피부 수용체의 흥분이 악화되는 지점을 찾고 난 후에야 연속된 갱신을 멈출 수 있었다. 따라서 애쉬비가 견지하는 신경계는, 환경에 적응한 현재의 태도를 유지하며 웬만한 변화에도 이를 지속시켜 나가지만, 큰 변화가 닥쳐 균형이 깨지면 다시 복원이 가능할 때까지 기존 세팅을 계속 갈아엎는다. 따라서 뇌 상태의 쉽 없는 갱신이 변화의 충격파를 상쇄하는 해법이 되는 것이다.

그런데 이 실험 예시는 자극과 환경에 대한 반응과 적응뿐 아니라, 학습의 원리 역시 해명한다. 쥐는 전기 쇼크가 가해지는 시공간에 적응했을 뿐 아니라 그것을 회피하는 법을 익혔기 때문이다. 그가 정의한 학습이란 행동에 변화를 취하는 것이며, 변화 후 이전보다 환경에 더 잘 적응하게 된 것, 이로써 생존 확률을 끌어올리는 것을 의미한다. 그가 책에서 자주 언급한 고양이와 불의 관계가 대표적이다. 태어나서 처음으로 불을 접한 새끼 고양이는 움직이는 불꽃에 반응해 몰래 다가가 발로 낚아채려 하거나 냄새를 맡으려 접근하고, 주변을 어슬렁댈 것이다. 물론 이 행위들은 불에 데기 십상이며, 이를 경험한 고양이는 불에 쉽사리 접근하려 하지 않는다. 하지만 기온이 낮은 곳이라면 불을 회피하는 것만으론 부족하다. 체온을 일정 수준으로 맞추기 위해 너무 춥지도 덥지도 않은, 자신과 불 사이의 적정 거리를 찾아내고 유지할 수 있을 때, 적응이자 학습은 성사된다. 그리고 불에 대한 학습이 이뤄진 고양이의 뇌는 이전과 상이하다. 유사 의미로써 애쉬비는 ‘훈련’을 아래처럼 기술한 바 있다.

⁵⁹ DB, 8/7.

모든 훈련은 ‘처벌’이나 ‘보상’의 적정 활용을 포함하는데, ... ‘처벌’은 명확하다. 왜냐하면 이는 신경계에서 스텝 함수의 변화를 유발시킬 만큼의 강도로 어떤 감각 기관이나 신경 말단에 자극이 가해지는 경우를 의미하기 때문이다.⁶⁰ ‘보상’은 ... 어떤 물질(예를 들면 음식)이나 조건의 제공이 수반되며 그것의 부재가 ‘처벌’로 작동하는 경우가 이에 해당한다.⁶¹

처벌은 스텝 메커니즘을 기동시킨다. 예컨대 개가 소파 위로 뛰어오르지 못하게 훈련시키는 경우, 처벌을 통한 자극은 개에 현재 적용된 뇌의 세팅을 바꾸도록 유도하는 것이며, 주인의 의도와 일치하는 더 나은 선택을 통해 생리학적 안정을 되찾는 것과 동일한 의미가 된다.

정리해보자. 호메오스탯에는 결국 이중의 피드백 회로가 맞물린다. 빈도가 잦고 사소하게 발생하는 외란에 저항해 수시로 중앙값을 회복하는 회로가 하나의 충위를 이루며, 다른 충위에선 간헐적이지만 시스템의 존립을 흔드는, 보다 위력적인 변화에 맞서 기존 매개 변수 세팅을 계속 회생시키는 정책으로 평형을 수복하기 때문이다. 그리고 이 같은 이중 구조가 그가 명명한 ‘초안정성’을 지칭한다. 초안정성은 호메오스탯이 어쩌서 뇌로 간주될 수 있는지를 입증하는 개념으로서, 이를 명확히 규정하기 위해 과거에 있었던 동물 실험의 선례를 꺼내든다. 1910년대 치러진 어느 실험은 원숭이 안구의 외직근과 내직근을 절제해 그것을 엇갈려 연결한다. 이렇게 되면 외직근의 수축이 안구를 바깥이 아니라 안쪽으로 회전시켜 운동 방향이 역전돼버리지만, 시간이 지나면서 원숭이는 예전처럼 안구 운동을 재개할 수 있었다. 또 다른 1947년 실험에선 거미원숭이 팔의 굴근과 신근에 붙은 신경을 잘라 반대로 연결했는데, 신경이 재생된 후 얼마간 팔을 제대로 가눌 수 없었지만, 시간이 흐를수록 이전과 별 차이 없이 움직임을 회복했다.⁶² 뇌는 이렇게 내부가 역전된 최악의 상황까지 극복한다는 것이

⁶⁰ 애쉬비에 따르면 환경과 조직은 상호 긴밀히 연루되지만, 전체 요소들이 항시적 변화를 겪는 것이 아님을 강조한다. 만일 전체가 환경과 엮여 순간 순간 벌어지는 모든 변동을 감내해야 한다면[그는 이것을 완전 함수(full-function)로 칭한다], 각 변수에 미치는 파장은 매번 치명적일 수 있다. 마치 쟁반 위에 놓인 젤리가 국소적 흔들림에도 몸통 전체가 진동하는 것처럼, 모든 변수가 다른 모든 변수와 얽혀 즉각 연쇄적으로 영향을 주고받는 상태에선 반응의 복잡성이 폭발해 결국 적응에 실패한다는 것이다. 특히 매개 변수와 같이 행동 양상 자체를 규정하는 부분들은 반드시 어떤 시점에서만 단속적으로만[스텝 함수(step-function)로서] 작동해야 함을 설명한다.

⁶¹ DB, 8/8.

⁶² DB, 8/4.

며, 그만큼 초월적 탄력성이 잠재해 있음을 말한다. 호메오스탯은 이를 모의한 것이다.

대표적으로 정류자에 관한 부분을 상기해보자. 오퍼레이터가 패널 스위치로 정류자를 조작해 인위적으로 극성을 바꿔놓으면, 유니셀렉터는 기존 스위치 배열의 포기과 재탐색을 반복함으로써 위 원숭이들처럼 전류의 역전을 극복한다.⁶³ 그렇다면 이것이 기존 교정 피드백 시스템과 어떻게 다를까? 그는 호메오스탯에 잠재된 복원력의 특수성을 항공기의 자동 조종 장치와 비교한다. 이 장치는 기체가 왼쪽으로 쏠리면 보조익이 오른쪽으로 힘을 받도록 하고, 반대의 경우 다시 왼쪽으로 기울도록 조절해 기체의 수평을 맞춘다. 그런데 정비사의 실수로 자동 조종 장치와 보조익을 거꾸로 연결했을 경우를 가정해 보자는 것이다. 이렇게 되면 기체는 양의 피드백이 걸려 자체 교정 능력을 상실해 정위치와의 오차를 계속 벌려가다 추락할 것이 뻔하다. 하지만 호메오스탯처럼 이중 회로가 작동한다면, 그 이탈 정도가 처음엔 악화되더라도, 일정 수준을 초과하는 순간 곧바로 조직 재배열에 돌입해 스스로 편차를 줄이기 때문에 비행을 계속할 수 있다.

만일 이런 유형의 메커니즘으로 자동 조종 장치가 설계된다면, 정비사가 “보조익 어느 쪽으로 그것을 연결해야 맞는 거지?” 라고 물었을 때, 그 대답은, “상관없어, 네 맘대로 연결해도 돼. 어차피 장치가 알아서 처리하니까”가 될 것입니다.⁶⁴

기존의 교정 피드백, 서보메커니즘만으로는 ‘적응하는 뇌’를 구현할 수 없다. 그러나 애쉬비의 모델은 행동을 수정할 뿐 아니라 스스로 조직의 연결 패턴을 바꿔 대응하기 때문에 기성 피드백 제어의 안정성을 초월하며,⁶⁵ 그렇기 때문에 외부 환경이 던지는 돌발적 난제들을 탄력적으로 수습하는 복잡한 생명체의 행

⁶³ 앞서 말한 반우드 하우스의 시연회에서 애쉬비는 방문자들에게 망치로 호메오스탯의 몸체에 충격을 가하거나 유닛 사이의 일부 연결선을 끊어보도록 했고(하나의 연결이 끊겨도 다른 유닛을 경유해 연결을 유지할 수 있다), 수조의 금속판이 움직이지 못하도록 하는 등 다양한 자극의 실험을 권유했다. 이런 극단적 상황에서도 호메오스탯은 안정 평형(적응)에 도달할 수 있음을 보여주고자 한 것이다. De Latil, *Thinking by Machine* (cit. n. 43), 301.

⁶⁴ “Homeostasis,” 93.

⁶⁵ 따라서 시간의 흐름에 따라 부단히 피드백 작용 자체의 개선을 이뤄나가는 특성을 갖는다. J. O. Wisdom, “The Hypothesis of Cybernetics,” *The British Journal for the Philosophy of Science* 2:5 (1951), 1-24 중 11.

동 저변을 투명화한다. 뇌의 기계적 재현이라는 당대의 흐름을 타면서도 분명 그 이상을 성취했다는 평가가 가능한 이유는 이 점 때문이다.

4. 무작위 선택의 전략

1952년 회의장에서 비글로는 아래처럼 이의를 제기한다.

그런데요, 나는 이런 방식으로 해법을 찾는 것과 학습이란 단어를 연계시키는 것이 곤란하다고 봅니다. ... 만일 바닥에 구멍 뚫린 상자 안에 쇠구슬을 넣고 사방으로 부딪치게 굴린다면 결국 그 구멍으로 떨어지겠죠. 이게 당신이 말하는 학습이라는 거군요.⁶⁶

그의 비판은 호메오스탯의 위기 극복 방식이 전적으로 무작위 선택에 의존한다는 점을 염두에 둔 것이었다. 지적 행동과 의사 결정의 저변을 한낱 난수표 따위의 확률 변수가 지배한다는 것을 도저히 수용할 수 없었기 때문이다. 이리저리 불규칙하게 움직이는 쇠구슬이 언젠가 구멍을 빠져나가는 것처럼, 호메오스탯의 해법 역시 우연히 얻어걸린 것에 불과하다는 것이다. 그렇다면 어떻게 이걸 지능이라 할 수 있는가? 과연 이게 학습일까? 회의장을 혼란에 빠뜨린 것은 이 부분이다.

애초에 애쉬비가 신경계의 작동 기제를 명시하기 위해 개념화한 다이어그램은, ‘환경’과 ‘반응 기관’, 이 반응의 양상을 결정짓는 ‘매개 변수’, 그리고 ‘EV’, 총 4요소가 얹힌 피드백 회로로서 이는 생물과 환경의 관계를 공학적 제어계와 일치시킨 결과였다.⁶⁷ 반응 기관은 환경으로부터의 자극에 반응해 EV를 출력하고, 이 출력값 통제를 위해 환경과 반응 기관의 역학에 따라 재조작이 가해진다. 그리고 그 EV 수치가 정상 범위 내에 있을 때와 이탈했을 때, 서로 다른 피드백 루프가 대응하고, 이들 회로의 공조 하에 EV를 정상치로 되돌려 시스템의 평형을 유지한다. 자동 제어 공학 용어로 치환하면 EV는 제어의 목표치, ‘제어량’에 대응하며, 반응 기관과 환경의 상호 작용은 ‘조작량’이라는 변수에 상응한다. 그리고 매개 변수는 액추에이터와 컨트롤러의 기능을 합친 것처럼 제어 신호를 내보낸다. 애쉬비가 세운 생명-환경 상호 관계의 토대는, 그곳 메이시 회의의

⁶⁶ “Homeostasis,” 104.

⁶⁷ DB, Chapter 7.

토론자들에게 익숙한 서보메커니즘, 교정 피드백 모델과 하등의 차이가 없었다. 그러나 이 기계의 지능과 학습의 근간으로 제시된 ‘무작위 선택’은 도무지 석연찮은 것이었다.

예컨대 R-W-B 논문의 공학적 출전 격인 대공(사격) 예측기는 사이버네틱스 초반부를 장식한 합성 생명체로서 호메오스탯과 대비를 이룬다. 과학사학자 피터 갤리슨(Peter Galison)에 따르면, 위너가 군부의 발주로 개발에 착수한 뒤 연구를 심화시켜 갈수록 이 자동 시스템은 적군 파일럿과 아군 포병, 양측의 의도와 반응이 동시에 얹힌 지적 체계의 원형이자, ‘전기 공학적’ 시스템과 ‘자기 수용체’의 생리학 계통 전체가 맞물린 혼종 체계로서 개발자 스스로에게 각인되고 있었다.⁶⁸ 한편 피커링은 영공에 적기가 뜨면 즉시 각성에 돌입, 자동으로 추적하고 적기의 미래 위치를 예측함으로써 지정된 시각, 계산된 지점에서 격추 임무를 마친 뒤 다시 잠에 빠져드는 존재로 그의 완성된 대공 예측기를 가정해 의인화한 적이 있다.⁶⁹ 피커링과 갤리슨 양측 모두 1940년대 전쟁의 배경 하에 시도된 이 미완의 기계를 산 것과 죽은 것, 생명 있는 존재와 무생물 기계의 이분법을 현격히 위축시킨 계기이자, 행동과 정신의 본질을 기계에 대입하는 사이버네틱스 특유의 사고가 기폭된 지점으로 지목한다. 개발에 깔린 저변의 의도에서 호메오스탯 역시 이와 일치했다. 애쉬비는 생명 조직과 기계 사이를 가르는 관념적 경계를 부수려 했고, 그의 인공 뇌는 이 해체 작업의 도구이기도 했다.

... 복잡한 평형을 위한 선택은, 관찰자가 이 범위 내에서 적응현상을 추적할 수만 있다면, 이례적인 어떤 것이나 아니면 경이로운 사건으로 간주하는 것은 결코 적합지 않다. 그것은 법칙일 뿐이다. 우리가 여태껏 이런 사실을 확인하는데 실패한 주된 이유는, 우리가 발 딛고 있는 이 세상이 형태에 있어 지극히 이원화 돼 있기 때문이다. 멈춰버린 시계처럼, 아주 단순한 형태라 손쉽게 분해해 버리거나, 아니면 극도로 복잡해 완전히 다르게 인식한 나머지, 그것이 생명 있는 존재라 말하는 것처럼 말이다.⁷⁰

⁶⁸ P. Galison, "The Ontology of the Enemy: Norbert Wiener and the Cybernetic Vision," *Critical Inquiry* 21:1 (1994), 228-266 중 228-233.

⁶⁹ A. Pickering, "A Gallery of Monsters: Cybernetics and Self-Organization, 1940-1970," S. Franchi and G. Güzeldere, eds., *Mechanical Bodies, Computational Minds* (Cambridge, MA: MIT Press, 2005), 229-245 중 231.

⁷⁰ DB, 18/1. 강조는 원문

그러나 위너가 구상한 대공 예측기의 학습 기반은 데이터 축적을 통한 통계학적 추론이라는 점에서 호메오스탯의 지능과 확연히 갈린다.⁷¹ 우선 이 시스템은 적기를 포착해 이동 방향과 속도, 대공포와의 거리를 자동 계측하지만, 포탄이 표적까지 도달하는 시간을 고려해야 하므로 탄착점은 적기가 그럴 미래 경로상의 한 지점으로 정해진다. 그렇다면 포의 조준은 결국 허공을 향해야 하는데, 이것은 발포와 타격의 구간에서 펼쳐질 적기의 운동을 예측해 그 미지의 공백을 채워 넣음으로써 성사된다. 어떻게 가능할까? 우선은 표적의 이동 궤적을 외삽해 직선이나 매끄러운 곡선으로 미래의 위치를 추정해야 한다. 하지만 첫 번째 격추 시도가 실패하는 순간 이 방법은 효력을 잃고 사태는 복잡해진다. 자신이 공격당하고 있음을 눈치 챈 파일럿은 그것을 피하려 완만한 경로를 타기보다 지그재그식의 회피 동작을 취할 것이고, 그 불규칙성은 예측의 난도를 높이기 때문이다.

하지만 다행스럽게도 파일럿이 취하는 도피의 조종술은 제한된 폭에 갇힌다. 초고속 비행 중 기체를 급선회할 경우 엄청난 중력 가속도가 파일럿을 의식 불명에 빠뜨릴 수 있으므로 그는 항상 이 점에 유의해야 한다. 뿐만 아니라 조종 의면과 기류가 빚어내는 한계 요건, 교전 시 엄습하는 심리적 압박은 그로 하여금 복잡하고 즉흥적인 조종술을 구사하기보다 이미 훈련으로 각인된 패턴을 따르도록 유도한다는 점이다. 이런 제약 때문에 결국 회피의 조종술은 마냥 자유롭기보다 모종의 규칙성을 띄고, 이로부터 통계적 편향이 발생한다는 것이다.

모든 계산은 제어기 자체에 탑재되어야 한다. 그리고 이 계산은, 특정 유형의 항공기가 여러 가지 비행 조건 하에서 보였던 과거의 통계적 경험치의 데이터를 반드시 포함해야 한다.⁷²

개요는 다음과 같다. 관측된 표적의 운동과 부합하는 통계를 인출해 제어부에 되먹이고, 여기서 조직된 지령에 따라 재빠르게 포신을 움직임으로써, 시시각각 변하는 적기의 이동 방향과 속도에 자동 응대가 가능한 체계였다. 그는 이렇게 목표의 특정 움직임에 맞춰 조준과 발포의 대응책을 스스로 조절해 나가는 것을 학습의 본질과 연결시킨다. 그에게 학습이란 상황에 실시간 반응하면서

⁷¹ N. Wiener, *Cybernetics: Or Control and Communication in the Animal and the Machine* (Cambridge, MA: MIT Press, 1961[1948]), 5-6, 8-10.

⁷² Wiener, *Human Use* (cit. n. 19), 62.

누적된 과거의 경험치로 계획을 수정해가며 지정된 미래를 통제하는 것과 다름 없었다.

그런데 연구를 진행하면서 위너는 공습에서 조준, 그리고 사격으로 이어지는 과정 전체를 단일 정보 회로로 통합하고, 적군 파일럿과 아군 포수라는 인적 성분을 이 피드백 루프에 삽입된 서보메커니즘으로 무차별 치환함으로써, 이곳 인간과 기계가 뒤섞이는 교전의 장을 순수한 통신 공학의 영토로 변모시킨다.⁷³ 이렇게 통신 회로로서 필드가 규정된 이상, 신호의 송수신 안정화 문제가 제어의 총체를 관할했고, 노이즈의 개입은 이곳의 최대 위협 요소로 부상했으며, 이 체계에 접근하는 응용수학적 방법이 인간/기계를 관통해 일괄적 분석의 체계로 동원될 수 있었다. 예컨대 배경 소음처럼 레이더를 통해 우연히 회로에 난입, 신호의 패턴을 뭉개는 요인들은 통계역학적으로 고안된 여파기로 억제돼야 했고, 기계 작동 시 마찰과 관성으로 유발되는 진동이나 지체 등의 우발적 난조는 메시지의 민활한 순환에 장애를 일으키므로 최소치로 제압되어야 했다. 결국 우연이나 우발성은 명령과 통제의 모델을 지표삼아 조절과 안정화에 주력했던 초기 사이버네틱스의 영토에선 결코 환영받지 못했다.

애쉬비의 발표 1년 전, 메이시 8차 회의에서 벌어진 도널드 맥케이(D. M. MacKay)와 새비지(L. J. Savage) 사이의 논쟁 역시 기계와 우연성에 관한 상반된 견해차 때문이었다는 점에서 이곳 분위기를 엿볼 수 있다. 영국 킹스 칼리지의 물리학부 연구원 맥케이가 자동 기계에 무작위성을 편입시켜 추론 기반의 인간 행동을 모방할 수 있다고 제안했을 때, 통계학자 레너드 새비지는 이 전략이 기계가 현실 문제를 푸는 듯 보여주는 그럴듯한 표면 처리가 될 수 있을지언정, 절대로 문제의 근본 방편이 될 수 없다면서 맥케이를 신랄히 비판했다.

어쨌든 우리는 행위에 있어 무작위 전략이란 것이 과연 어떤 역할인지, 아니면 어떤 일을 수행하는지 심도 있게 생각해 보아야 합니다. 인간 행동을 탁월하게 모방하는 자동 기계가 내부에서 어쩔 수 없이 덜컥대며 굴러다니는 거라고 곁핥기식으로 추정해 버리기 전에 말이죠.⁷⁴

73 N. Wiener, *I am a Mathematician* (Cambridge: The MIT Press, 1964), 251-252.

74 D. M. MacKay, "In Search of Basic Symbols," H. von Foerster, ed., *Cybernetics, Transactions of the 8th Conference, March 15-16 1951* (New York: Josiah Macy Jr. Foundation, 1952), 205.

새비지의 험난 섞인 마지막 표현은 다음 해, 비글로가 비유했던 ‘바닥에 구멍 난 상자 안에서 이리저리 배회하는 쇠구슬’과 닮아 있었다. 상황 변화에 맞서 안정화를 꾀하는 것이 피드백 메커니즘 본연의 임무라는 점, 따라서 시스템이 겪는 우발성은 오류 그 자체로 인식된다는 점에서 호메오스탯은 기괴하기까지 했다. 그들에게 랜덤 요동은 시스템 붕괴의 단초이자 무질서를 의미했지만, 이 인공 뇌는 그 혼돈이 오히려 지능의 원천임을 선언하고 있었기 때문이다. 더구나 이 기계는 심각한 결핍마저 보였다. “목적 추구 행위는 곧바로 교정 피드백과 의미가 일치한다”는 명제, “목표를 지향하는 모든 행위는 교정 피드백을 요구한다”는 R-W-B의 진술처럼, 이들 모델은 목적과 피드백 서로를 불가분 관계로 묶는다.⁷⁵ 피드백으로 조절되는 행위가 목적인 바를 성사시키지만, 그 행위를 규제하는 신호 자체가 목표와의 격차에서 빚어지므로 지정된 목표는 역으로 정보 순환의 회로를 지탱한다. 그러나 호메오스탯의 피드백에는 설정된 목표가 없다. 그것은 위너의 기계처럼 적기를 격추해야 할 임무가 부여된 바 없고, 새년의 전자 쥐처럼 미로 탈출의 과제조차 프로그래밍 되지 않은 채, 뭔가 부산하게 일을 벌이면서도 오직 고요한 침잠을 수복하기 위해 뒤척임을 이어간다.⁷⁶ 영국의 또 다른 사이버네티션 그레이 월터가 냉소를 가미해 그것을 ‘잠에 빠져들기 위한 기계’로 표현한 까닭은 이 때문이었다.⁷⁷

하지만 목적 없이 구현된 계의 내부에서 의미 있는 목적이 출현할 수 있음을

⁷⁵ Rosenblueth et al., “Behavior” (cit. n. 6), 18, 24.

⁷⁶ 클로드 새넌은 1951년 메이시 8차 회의에서, 미로를 해마다 지정된 목표를 찾아내는 전자 쥐를 소개한다. 알루미늄 벽으로 미로를 만들고 출발 지점에 전자 쥐를 놓아두면, 처음 시도에서 여러 번 벽에 부딪히고 헤매면서 복잡한 경로를 그려내지만, 두 번째, 세 번째 시도에서 이 쥐는 점차 단거리의 최적 경로를 이동해 목표물(전기 치즈)을 발견한다. 또한 이후에 미로 어디에나 쥐를 놓더라도 목표물을 빠른 시간 안에 발견하는 모습을 보인다. 과거 시도의 기억을 활용하기 때문이며 초반의 시도에서 시행착오를 겪지만 무작위 시도의 해법이 적용된 것은 아니었다. [C. Shannon, “Presentation of a Maze-Solving Machine,” H. von Foerster, ed., *Cybernetics, Transactions of the 8th Conference, March 15-16 1951* (New York: Josiah Macy Jr. Foundation, 1952), 173-180] 때문에 비글로는 애쉬비와의 논쟁에서 “새년의 전자 쥐는 분명히 학습하지만, 당신의 것은 아니다”라며 비판한다.

⁷⁷ 스티브 하임스는 메이시 회의를 기점으로 인간 인지와 심리, 이것의 인공적 시스템 구현에 관한 학제적 토론을 이어간 학자 군을 사이버네티션으로 통칭한다[S. J. Heims, *The Cybernetics Group* (Cambridge, MA: MIT Press, 1991)]. ‘잠에 빠져들기 위한 기계’라는 표현은 W. G. Walter, *The Living Brain* (New York: W.W. Norton & Company, 1953), 123-124에 나온다. 이 책에서 월터는, 외부의 방해로 잠을 깬 고양이가 이리저리 몸을 뒤척이다가 다시 잠들어버리는 모양새로 호메오스탯의 행동을 묘사했다.

처음으로 보여줬다는 위너의 논평처럼, 설계자로부터 이식된 목적은 결여했으
 되 끊임없이 적응(평형)을 추구한다는 점, 각종 병기 제어 시스템이나 수치 제
 어 절삭기처럼, 1950년대 초 이제 막 자동 제어부를 탑재하기 시작한 여타 기계
 들과 비교해 그 어떤 수행적 프로그래밍이 없는 그것은 존재의 기원 자체가 판
 이했다.⁷⁸ 앞서 인용한 바처럼, 애초에 애쉬비가 구현하려 했던 것은 고난도 방
 정식 풀이와 같은 특수 용도에서 뇌의 역량을 초월하는 인공 지능이 아니라, 스
 스로 적응하기 위해 애쓰고 그러면서 오류와 지체를 배어내는 생물학적 뇌의
 본모습이었기 때문이다. 그렇다면 이 모델에게 주어진 궁극의 목적이 자기 생존
 으로 지정되는 것 자체에 모순은 없다. 살아있는 유기체는 주어진 목적을 이행
 하기보다 스스로 목적을 생성해내기 때문이다.

이런 의미에서 그날 있었던 비글로의 지적이 고스란히 애쉬비의 논거가 될
 수 있다는 점은 아이러니하다. 탈출구의 위치를 모를 때 이리저리 헤매는 것은
 지능이 아닌가? 상황이 급변하거나 낯선 환경에 처했을 때, 닥치는 대로 모든
 수단을 동원해 보다 허둥대며 실패하고, 그러다 우연히 들어맞는 방법을 발견하
 면 일단 제쳐두고 이를 밀어붙여 보는 것은 동물과 인간의 보편 습성이자 위기
 를 극복하는 현실적 책략이기도 하다. 하지만 그의 인공 뇌가 정확히 환기하는
 부분은 보다 낮은 단계의 이벤트였고, 생물학적 조직이라는 저층위에서 인지와
 학습에 관한 기성 관념의 허구성을 폭로한다. 어떻게 무작위 선택이 학습일 수
 있느냐는 반박에 대한 그의 답변은, 애초에 '학습'이란 용어가 관찰을 통해 얻
 된 것이 아니냐는 되물음이었다.⁷⁹ 이는 우리가 지능이라 일컫는 대상의 배후에
 신호 송수신을 비롯한 정보의 조작과 처리, 혹은 이를 초월하는 미지의 영역이
 작동하는 것이 아니라 단지 물리적 대상들이 맞물리는 우연적이며 복합적 이벤
 트의 연속에 불과할 수 있다는 점, 그리고 그것이 하나의 허명으로서 관찰자의
 의식에서 구성된 것임을 시사한다.⁸⁰

1945년 논문에서 그는 생리학적 적응의 근간이 조직 재배열이며 그것은 항상

⁷⁸ Wiener, *Human Use* (cit. n. 19), 38. 피커링은 외부로부터 주어진 목적이나 임무가 아니
 라 자체 내에서 발생한 목적을 갖게 된 사상 초유의 기계라고 말한다. A. Pickering,
 "Cybernetics and the Mangle: Ashby, Beer and Pask," *Social Studies and Science* 32:3
 (2002), 413-437 중 417.

⁷⁹ "Homeostasis," 103.

⁸⁰ 애초에 그가 '적응'의 관념적 모호함과 부적절성에 관해 얘기하면서 물리학 개념의 '평형'을
 취했던 것 역시 이와 동일한 동기가 반영돼 있었다.

시행착오로써 취해짐을 명시했다. 재배열이 무작정 시도되므로 이후의 시도가 이전의 것보다 나은 것임을 보장할 수 없고, 또 어떤 선택이 이루어지더라도 이는 최선의 것이 아닐 공산이 크다. 단지 지금 바뀐 배치가 우연히 평형을 탈환하는 데 기여한다면 더 이상 변환을 지속할 필요가 없으므로 그 자리에서 작업은 중단된다. 그런데 관찰자의 시점에서는 이런 맹목적 프로세스가 의도된 선택이나 계산에 따른 조직화로 읽힐 수 있음을 고려해야 한다. 신경 과학 기반 로보틱스 연구자인 디 파올로(E. A. Di Paolo)는 다음과 같이 말한다.

스텝 함수의 무작위 선택이 개념상 흥미로운 이유는, ... 아무 지적 능력을 갖추지 못한 메커니즘이 적응의 반응을 내어놓을 수 있음을 증명한다는 것, 그리고 외부 관찰자에게 그것이 꽤 영리해 보일 수 있다는 점 때문이다. 애쉬비는 지능적 수행의 행위가 이와 상응하는 지적 메커니즘의 결과로서 나타나야만 한다는 관점...에 첫 도전을 감행한 인물이었다.⁸¹

완전한 무지의 프로세스가 지능적 행위를 낳는다는 애쉬비의 도전적 모델은 뇌에 관한 일반 상식에 위배됨이 없다. 뉴런의 작용이나 그 해부학적 성질은 단순하지만 10^{11} 개의 뉴런이 백 조 단위의 시냅스로 얹히는 상호 작용으로부터 ‘사고’가 생성된다는 뇌과학계의 분석은, 오히려 그가 제시한 모델의 신빙성을 높이기 때문이다. 해부학자 요로 다케시는 신경 세포 개개의 기능이 절대 복잡하지 않다는 점, 그들은 서로 연결되어 흥분하거나 억제될 뿐이라며 그 단순성을 강조한다. 그러면서도 “도대체 거기서 어떻게 기억이나 심리, 계산이나 꿈이 생기는 걸까. 복잡 기괴한 것은 바로 그 점이다”라며 그 비대칭의 경이로움을 표명했다.⁸² 한편 1980년대부터 복잡계 과학을 이끌었던 페르 박(Per Bak), 뇌-기계 인터페이스 분야의 니코렐리스(M. Nicolelis)가 설명한 뇌의 공통점 역시 지극히 단순한 요소의 집합으로부터 고차원적 현상이 창발한다는 점이였다.⁸³ 그런데 애쉬비 역시 복잡성과 임계에 관한 학문 영역이 정착되기 이전부터 이를 명확히 기술한 바 있다. 다수의 성분들이 뒤엎히는 동역학계는 기본 단위의

⁸¹ E. A. Di Paolo, “Organismically-inspired Robotics,” K. Murase and T. Asakura, eds., *Dynamic Systems Approach for Embodiment and Sociality* (2003), <http://goo.gl/92OmVT> (2016. 12. 15 접속).

⁸² 요로 다케시, 김석희 옮김, 『유뇌론』 (재인, 2006), 70-71.

⁸³ 페르 박, 정형재 · 이형우 옮김, 『자연은 어떻게 움직이는가』 (한승, 2012), 265-268; M. Nicolelis, *Beyond Boundaries* (New York: St. Martin's Press, 2011), 11-30.

물리 작용으로 나눠 직접 계산할 수 없으므로, 기존 피드백 모델로선 더 이상 그 수학적 분석의 효력을 기대하기 어렵다는 것, 또한 이런 비선형계의 복잡 패턴 자체로부터 다른 층위의 행위가 발현될 수 있다는 것이 요점이었다.

오직 두 개의 성분만이 서로에게 영향을 줄 때, 피드백의 속성은 전체 특질에 관한 중요하고 유용한 정보를 가져다준다. 하지만 겨우 네 개 정도로 성분의 개수가 증가하기만 해도, 각 요소들 모두가 나머지 세 부분에 영향을 미치면서 전체적으로 총 20개의 회로가 확인되고, 이들 모든 회로의 개별 속성을 파악하는 것만으론 시스템 전체에 대한 완전한 정보를 얻을 수 없다. 이 같은 복잡계는 독립된 피드백 회로 몇몇이 얹힌 것으로 간주할 것이 아니라 오직 전체로서 파악해야 한다.⁸⁴

지적 행위의 시스템 내부가 영리한 개별 요소의 조합과 탄탄일 수 있다는 점은, 설계자의 입장을 연루시키면서 그가 확보한 시야의 한계를 시사한다. 전류를 차단해 그것을 죽인 다음, 신경 조직을 해부해 전선-계전기 상태를 들춰보지 않고서는 실험자가 특정 시점에서 회로 배치를 알아낼 방법이 없다는 월터의 논평처럼, 시스템 내부에서 뭐가 어떻게 되고 있는지는 관찰자, 설계자 모두에게 예측 불가다.⁸⁵ 이것은 외적 행위만으로 내부 동향을 지정할 수 없다는 뜻이며, 각 요소를 설계해 이미 구조를 섭렵한 이에게도 블랙 박스로 봉인되기는 마찬가지라는 얘기다.⁸⁶ 이 지점에서 호메오스탯의 정체성은 선명해진다. 증기 터빈의 조속기, 2차 대전의 함포 조준기, 위너가 착수했던 대공 예측기처럼 피드백을 통한 자가 조절 체계의 움직임은 인과적으로 자명하며, 또 합리적 예측의 권역을 벗어나지 않는다. 그러나 난잡한 상호 작용으로 배치가 뒤바뀌고 시시각각 무작위 선택을 거듭하면서 단속적 우연의 흐름을 생성해갈 때, 시스템에 대한 투명한 통찰은 무력화된다.

설계자의 예측과 제어가 먹혀들지 않는 기계는 신뢰도가 형편없는 불량으로 낙인 찍히겠지만, 이것이 생명체라면 그 제멋대로의 성질은 본질이 된다. 동물

⁸⁴ W. R. Ashby, *An Introduction to Cybernetics* (New York: John Wiley & Sons, 1956), 54.

⁸⁵ Walter, *Living Brain* (cit. n. 77), 124.

⁸⁶ 관찰자가 확보한 시야의 한계에 관한 설명은 다음을 참고하라. Cariani, "Homeostat" (cit. n. 40), 142; K. Krippendorff, "Cybernetics's Reflexive Turns," *Cybernetics and Human Knowing* 15:3-4 (2008), 173-184 중 175-176; Pickering, "Cybernetics and Mangle" (cit. n. 78), 417.

의 신경계는 외부에서 변화를 지정하거나 지시하는 것이 불가능하며, 환경에 의해 변화가 촉발되더라도 계의 내부에서 그 어떤 선택이 취해질지는 미지수다.⁸⁷ 하지만 이런 자율성이 환경의 끝없는 변화를 감내하는 내구력의 열쇠가 된다는 점을 애쉬비는 초안정성의 원리로 해명하려 했다. 결과적으로 기계임에도 설계자의 권능에서 벗어나 스스로 모호해진다는 것, 그리고 불가측 영역에 진입했다는 사실, 그러나 이 어지러움 속에서도 전체가 균형 회복을 끊임없이 모색할 수 있다는 점은 호메오스탯이 살아있는 뇌의 속성을 모의하는 데 소정의 성과가 있었음을 입증한다. 그리고 이 부분이 애쉬비의 인공 뇌가 사이버네틱스 기성담론을 향한 도전이 성립되는 지점이기도 했다.

5. 순수한 물리적 상태로서의 정신

카리아니(P. A. Cariani)는 유니셀렉터가 기존 배치를 계속 갈아엎을 때, 이 전환의 프로세스가 그 어떤 계획이나 이해가 전무한 상태에서 진행된다는 점에 착안해 이를 생물학적 진화에 대입한다.

호메오스탯의 인식론적 맥락은 생물학적 조직이나 뇌의 항상성, 학습, 그리고 진화의 차원에서 지극히 정상이다. 선택을 행하는 계의 부분들은 그들이 제어하는 세부 과정에 있어 그 어떤 이해나 모델을 (거의 법칙처럼) 필요로 하지 않는다. 생물학적 진화는 맹목적이다. 유전학적 기전은, 어떤 돌연변이가 생존과 복제를 증진시킬지, 혹은 어떤 변이가 그 반대를 초래할지 자체의 본보기나 그것의 환경에 대한 예상 모델이 전혀 없다는 점에서 그러하다.⁸⁸

생물 역사의 테이프를 고속 되감기한 뒤 재생 버튼을 누른다 해도, 진화의 과정이 그대로 반복 플레이될 리 없다는 굴드의 ‘생명 테이프 재생하기’ 비유처럼, 진화는 돌연변이라는 우연으로 점철되기 때문에 현재의 결과는 필연이 아니며, 지금과 상이한 종들의 조합이 살아남아 역사의 전개를 뒤집는 일이 얼마든 벌어질 수 있었다는 점을 상기해보자.⁸⁹ 동시에 그 돌연변이가 비록 눈먼 과정이긴 해도, 승자와

⁸⁷ Capra, *Web of Life* (cit. n. 30), 219.

⁸⁸ Cariani, “Homeostat” (cit. n. 40), 146.

⁸⁹ 스티븐 제이 굴드, 김동광 옮김, 『생명, 그 경이로움에 대하여』 (경문사, 2004), 64.

패자를 골라내는 선택의 엄정함 때문에 개체는 결국 생존에 유용한 질서를 획득한다는 사실 또한 병행한다.⁹⁰ 이로써 ‘무작위 변이’와 ‘정교한 복잡성의 모순은 양립하며, 이는 고스란히 호메오스탯을 이해하는 단서가 된다. 이 모델의 적응과 학습 또한, 눈먼 시도들이 적합 여부 판정에 의해 계속 희생됨으로써 어느덧 성사된다는 점은 전자의 섭리와 일치하기 때문이다. 같은 관점에서, 지능이란 것이 박스 안에서 이리저리 굴러다니는 쇠구슬처럼 작동할 리 없다는 비글로의 비판은, 생물구조의 정치함이 그런 무작위 프로세스로 생겨났을 리 없다는 해묵은 오해와 닮아 있다. 이 오해들은 이면에서 빚어지는 수많은 실수, 그 폐기 처리된 시도들을 계산에 넣지 못한다.

헤일즈는 호메오스탯의 단출한 구조에는 많은 가정들이 약화되어 네트워크로 얽혀있다고 피력한다.⁹¹ 복잡한 세부를 결여했기에 오히려 물리적 행위로 규정하는 것 이상의 가정들을 대변할 수 있었다는 것이다. 실제로 그 가정의 그 물망에는 적응과 지능, 학습, 진화의 메커니즘까지 노드(node)로 합류해 있었지만, 한편 이 말을 뒤집어보면 호메오스탯을 구성하는 물리 유닛 4개의 상호작용만이 이 모두를 관통한 셈이다. 그리고 구조 자체 미학으로서의 미니멀리즘 만큼이나 이를 통해 애쉬비가 구사하는 어법 역시 간결하다. 결국 거기엔 우리가 기대했던 것이라고는 없다. 계획을 실행하는 고등의 사고 체계나 두뇌라는 중앙 통제부의 지령에 따른 목적론적 행위 등은 시스템의 덮개를 닫은 채 제한된 관점 안에 스스로를 가두어 버린 관찰자의 감탄사이나 존재할 뿐이라는 식이다.

[사이버네이션들은] 기계와 유기체 세계 사이의 불연속성을 무시하지 않았나? 애쉬비는 여기서 더 나아가, 전자석 몇 개를 엮어 만든 기술적 장치로 구현하지 못할 생명의 정수 따윈 아예 없다고까지 했다.⁹²

뇌를 정신 작용의 거점이 아니라 여타 기관들과 마찬가지로 진화를 거쳐 생존에 특화된 “활동 기계”의 일부로 간주해야 한다며 이에 얽힌 모든 관념상의 특권을 말소한 애쉬비의 관점은 근본적으로 인지와 생명 과정 자체를 구별하지 않는다.⁹³ 이 지점에서, 1970년대 이래 마투라나와 바렐라가 체계화한 구성주의

⁹⁰ 선 B. 캐럴, 김명주 옮김, 『한 치의 의심도 없는 진화 이야기』 (지호, 2008), 62.

⁹¹ Hayles, “Boundary Dispute” (cit. n. 5), 26.

⁹² Dupuy, *On the Origins* (cit. n. 31), 150.

학파의 인지 생물학 기초, 그리고 기호 처리형 인공 지능의 대안으로 1990년대 본격화된 발현형 인공 지능의 기본 발상이 착근되고 있었다. 산티아고의 신경 생리학자들이 펼친 구성주의 논의의 독창성은, 이전엔 전적으로 별개의 분야로 다뤄졌던 생물 조직 이론과 지각/인식 이론 양측을 결합했다는 점에 있었고, 이로써 마음과 생명 유지의 물질 대사가 서로 분리된 채 논의 될 수 없다는 급진적 일원론의 과학적 타당성이 확보되기 시작한다. 그들이 내건 명제 또한 애쉬비만큼이나 간명했다.

인지는 생물학적 현상이며 오직 생물학적 현상으로서만 이해될 수 있다. 그 어떤 인식론적 통찰도 이러한 이해를 전제한다.⁹⁴

특히 이들 학문 진영이 내세운 ‘구조 접속’ 개념은 어떻게 단위와 단위 사이의 접속과 작용이 계의 다음 활동 기반을 산출하는지 밝히면서, 이를 인지 과정에도 일괄 적용한다는 점에서 의미가 있다. 이 개념의 핵심은 상호 작용에 따른 자극이 각 단위 내부를 교란시키지만, 그 상태를 지정하거나 일정 방향으로 조작할 수 없다는 점이다. 접속된 서로 다른 폐쇄 체계의 자율 논리에 따라 순간 순간 내적 변화가 결정될 뿐, 이 자극은 절대로 지시 명령적 영향력을 행사할 수 없고 그 변화 또한 임의적이기에 이를 “섭동”이라 규정한다. 그런데 이 섭동의 단위는 신경 세포와 세포, 신경계와 다른 계 사이, 또는 개체와 환경 사이로도 지정될 수 있으므로 그 경계는 상대적이며, 각 단위는 스케일 상의 분류일 뿐 그것이 어떤 단위든 간에 내부 고유의 작동 법칙으로 지배되는 것은 마찬가지다. 결국 그들이 말하는 인지란, 뉴런들의 섭동으로 신경계의 재조직이 촉발되고, 개체는 이 연속된 내적 변화의 흐름을 떠안으며 다시금 환경과의 작용으로 새로운 자극을 수용해 다음 전개의 바탕이 될 자신의 내부를 거듭 확정해가는 순환적 과정으로 규정된다.⁹⁵

그런데 인지가 이렇게 단위 간 펼쳐지는 에너지 교환의 역사라는 측면은 호

⁹³ DB, 3/14; W. R. Ashby, “Design for a Brain,” *Electronic Engineering* 20 (1948), 379-383 중 379.

⁹⁴ 게르하르트 로트, “자기생산과 인지: 마투라나의 이론과 그 계속적인 발전의 필연성,” J. 슈미트 편, 박여성 옮김, 『구성주의』 (까치, 1995), 255-284 중 257.

⁹⁵ H. R. Maturana and F. J. Varela, *The Tree of Knowledge* (Boston: Shambhala, 1992); H. R. Maturana, “The Organization of the Living: A Theory of the Living Organization (1974),” *International Journal of Human-Computer Studies* 51:2 (1999), 149-168.

메오스탕의 기본 열개에 적용된 원칙이기도 하다. 살폈다시피 기계를 구성하는 유닛 각자는 나머지 세 개의 유닛과 전류를 매개로 영향력을 주고받는다. 그런데 이들은 독립된 단위로서 서로가 서로에게 있어 외부에 존재하는 다른 자극의 원천과 차별되지 않는다.⁹⁶ 기계의 몸체 안에서, 유닛 A에게 나머지 유닛 B, C, D는 서로 분리되어 있으면서도 접촉을 통해 에너지를 주고받는 외부 환경의 여타 요소들과 차이가 없다는 것이다. 실제로 애쉬비는 “신체의 해부학적 구획 만으로는 환경과 조직의 개념을 명확히 지정할 수 없다”며, “체내 변수들은 마땅히 다른 어떤 성분의 ‘환경’으로 간주돼야 한다”라고 밝힌다.⁹⁷ 본고 2절의 마지막에서 기계의 독특한 열개가 경계의 유연성을 대표한다는 대목은 바로 이와 관계한다. 엄밀히 말해 그의 인공 뇌는 두개골 내부에 안치된 고립된 유닛으로서의 뇌가 아니라 환경과 신경계가 결합된 통합 인지 단위였다.⁹⁸

이제부터 ‘시스템’은 신경 체계가 아니라 유기체와 그것이 처한 환경의 복합적 전체를 의미한다는 것에 유의해야만 한다. 따라서 ‘시스템’이 모종의 속성을 띠다는 점이 표명돼야 한다면, 이 속성은 신경계에 기인하는 것으로 간주해서는 안 된다. 그것은 전제로 귀속되어야 한다.⁹⁹

마투라나의 인지 생물학 이론과 애쉬비의 뇌 모델에는 우리가 통상적으로 기대하는 부분이 빠져 있다. 양측 모두 외부 실재의 표상을 가져온다거나, 환경으로부터 ‘입수’해 내부로 ‘전달’되는 정보 자체가 없기 때문이다. 다만 상호 접속된 폐쇄적 모듈의 다발들이 자극으로 촉발된 연쇄 반응을 주고받고, 그 요동에 따라 연결 패턴을 조절해 나갈 뿐이다. 이런 물리 조직의 재편성이 인지, 또는 학습이라는 것, 따라서 그 어떤 정보도 신경계를 도관 삼아 흐르거나 그것이 입출력될 수 없음을 피력한다. 급진적 구성주의의 핵심 명제로서 ‘생명의 체계는 외부 에너지에는 개방되어 있지만, 정보 상으로 폐쇄돼 있다’라는 진술이 이미

⁹⁶ 다른 자극의 원천이란 이를테면 패널 스위치 조작이나 금속판이 담긴 용액의 점도 변화, 자극을 인위적으로 회전시키는 등의 외적 자극 모두가 될 수 있다. 각주 68 참조.

⁹⁷ DB, 3/12. Maturana and Varela, op. cit., 74-80에서도 이와 합치되는 설명을 확인할 수 있다.

⁹⁸ A. Pickering, “After Dualism,” R. E. Lee, ed., *Questioning Nineteenth-Century Assumptions about Knowledge, III: Dualism* (New York: State University of New York Press, 2010), 70-71.

⁹⁹ DB, 3/12.

1956년 애쉬비의 저서에서 사이버네틱스의 학문적 성질을 정의하는 데 동일하게 언급된 적이 있다는 사실은 우연이 아니며, 동시에 이 명제는 외부 세계와 신경계의 상호 작용 방식에 대한 양측의 관점을 함축한다.¹⁰⁰

그러나 마투라나가 자신의 이론을 체계화한 시점은 호메오스탯의 시연이 있는 지 약 20년 뒤였고, 그는 자신의 학문적 성과에서 애쉬비로부터의 직접적 수혜를 부인한다.¹⁰¹ 그리고 1990년대 인공 지능의 신 경향 역시 그 발상의 모티브에서 애쉬비와의 연결성을 찾기 어려울 정도로 시간적 격차가 벌어져 있었다. 반표상주의, 또는 발현형 인공지능의 토대를 제공한 인지 과학 분파로서, 체화된 인지(embodied cognition)의 지지자들이 애쉬비를 자신들 연구의 직접적 기원으로 거론하는 일은 드물다. 이들은 시간의 흐름과 상황의 구성성에 결부된 생리학적 변화, 상황에 따른 응대의 실질 행위를 앞세우면서, 기존 인지주의 패러다임이 중앙 프로세서의 입출력 말단부로 좌천시킨 감각 운동 체계의 복권을 시도한다. 감각 정보를 인지 프로세서가 처리해 행위가 산출되는 게 아니라, 상황에 연루된 감각과 행동으로써 인지가 발현된다는 것이 핵심이지만, 40년 전 애쉬비의 모델과 이들 연구의 긴밀성은 자신들 진영이 강조하는 표상주의, 혹은 기호주의라는 재래식 논의와의 차별성에 묻혀 여간해선 부각되지 않는 편이다.

하지만 예외가 있다. 인공 생명 분야 초기 논의가 있었던 1992년 유럽 회의 보고서 서문, “자율 시스템의 실현을 향해”에서 바렐라·부르진(P. Bourguine)은 현재 자신들 연구의 기저에 1950년대 사이버네틱스의 학문적 유산이 관류하고 있음을 밝혔고, 그 중에서도 월터의 거북, 그리고 이 애쉬비의 뇌 기계를 자기네들 인공 생명 계보의 시작점으로 새긴다.¹⁰² 그들이 말한 1950년대 초반과

¹⁰⁰ “사이버네틱스는 에너지에는 열려있으면서도 정보와 제어에는 닫힌 시스템, 정보가 이미 뻗 뻗하게 들어찬 체계에 관한 연구로서 정의될 수 있다.” Ashby, *Introduction* (cit. n. 84), 3-4.

¹⁰¹ 프로스와 스튜어트는 Froese and Stewart, “Life After Ashby” (cit. n. 33)에서 마투라나가 애쉬비의 신경계에 대한 논의 핵심을 계승했다는 견해를 밝혔고, 이에 대해 마투라나는 논문, H. R. Maturana, “Ultrastability ... Autopoiesis?: Reflective Response to Tom Froese and John Stewart,” *Cybernetics and Human Knowing* 18:1-2 (2011), 143-152에서 이를 부정하는 의견을 개진했다.

¹⁰² 신경학자 그레이 월터는 거북 형태의 로봇 한 쌍을 제작한다. 이들은 모터 구동으로 돌아가는 3개의 바퀴로 움직이며, 터치 센서와 광 센서를 통해 스스로 장애물을 피하고 어두운 곳에서 밝은 곳을 찾아 이동하는 습성을 보인다. 월터는 이들이 다른 로봇들과 달리 고정된 패턴의 행동을 보이지 않으며, 동일한 환경에서도 매번 다른 행동을 보인다는 점을 강조하면서 그 자율성이 살아있는 생명과 닮아있음을 피력했다.

1990년대를 묶는 고리는 ‘자율성’이었다.

사이버네틱스의 이들 두 개척자들과 지금 이 시대 인공 생명 지지자들에게 있어 그 선도적 직관은 상당히 흡사했다. 살아있는 생명에 고유의 자율성을 부여한 프로세스의 층위, 이에 대한 통찰의 필요성이었다. 이 맥락에서 자율성이란 생명체의 기본이자 근본인 존재의 능력, 실존을 확인하면서 사전에 정리된 것이 아닌 지금 유의미하며 현재에 특화된 세계의 생산 능력을 가리킨다. 그러므로 자율성은 그들이 펼치는 활동과 함께, 세상을 의미 있게 구성하는 방식 양쪽 측면에서 이해될 수 있다. 이 개념적 탐구는 자율적 작인의 설계와 구성을 동반하는 동시에, 세포에서부터 사회 집단 규모에 이르는 방대한 범위로의 적용을 제안한다.¹⁰³

애쉬비가 인공지능보다 생명체의 인공적 구현에서 더 부각됐다는 것은 무엇을 의미할까. 사이버네틱스가 목적론적 피드백 시스템에서 생물학적 조직으로 초점을 이동하고, 제어 체계의 정보 투입·산출로부터 내적 복잡성과 자기 생산으로 무게추를 옮겨가는 변화의 흐름과 무관치 않다. 사이버네틱스의 2차 물결이 도래한 시기가 1970년대 중반이란 점을 고려할 때, 1950년대 초에 제시된 애쉬비의 모델은 당시의 주류와 거리감이 있었고, 시대를 한 발 앞서는 통찰이었지만 그만큼 시기 상조였다는 의미 또한 성립한다. 당대의 상황, 기술 사회의 특성과 결부해 그의 연구를 조망하는 것으로 논의를 맺고자 한다.

6. 사이버네틱스, 전쟁의 목적론과 생명의 존재론 사이에서

기술사학자 토머스 휴즈(T. P. Hughes)는 1946년에서 1953년까지 해마다 열린 메이시 회의를 가리켜, 제어와 통신, 피드백에 관한 메타포(metaphor)가 넘쳐나 흥수를 이뤘던 장으로 묘사한 적이 있다.¹⁰⁴ 이 학술회의에 동원된 풍성한 메타포들은 수학과 제어 공학, 심리학, 신경 생리학 등의 서로 다른 분야가 활기차게 교신할 수 있는 유사성의 통로를 열었을 뿐 아니라, 분야를 관통한 통합 논의가 진척

¹⁰³ F. J. Varela and P. Bourguine, “Introduction: Toward a Practice of Autonomous Systems,” *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life* (Cambridge, MA: MIT Press, 1992), xi-xvii 중 xi.

¹⁰⁴ T. P. Hughes, *Human-Built World* (Chicago: University of Chicago Press, 2005), 94-96.

되는 토대이기도 했다. 그런데 이들 메타포의 다양성만큼이나 정치적 편향에도 주목할 필요가 있다. 특히 1947년까지의 초기 사이버네틱스 성원들은 공통적으로 교전 상황에서의 관측 및 조준, 목표물 추적, 무전 송수신에 얽힌 지각과 심리에 주목하는 경향을 보였으며, 이런 관심은 대공 화기나 어뢰, 유도 미사일 같이 제2차 세계 대전 기간 개발되고 활약한 자기-제어, 서보-유도 장치 기반의 무기 체계와 병치를 이루며 발표문과 저서 곳곳에서 담론을 형성해 나간다. 폴 에드워즈(P. N. Edwards)는 그 까닭을 아래처럼 풀이한다.

첫째, 이들 기계는 공유된 전쟁의 경험을 물리적으로 제시했고, 둘째로 반자동 무기 체계는 수학적 기술(記述)과 구체화된 실행 양측에서 인간과 기계를 결합, 사이보그 모델의 원형을 이룩했으며, 마지막으로 컴퓨터 등장 이전에 정보 활동 측면에서 이 기계들은 그들에게 익숙한 가장 진보적 장치들이었다.¹⁰⁵

각종 병기 체계와 맞물린 오퍼레이터, 제어 패넬을 마주한 군인들, 이들의 정신과 행동이 전쟁 수행 도구로서, 또 정보 처리 장치로서 전기/전자적 제어·통신 시스템과 연동 가능하다면 기계와 인간 사이에 벌어진 틈은 생각보다 좁을지도 모른다. 요컨대 논의에 동원된 군사 목적의 기계들은 단지 견해를 피력하기 위해 동원된 설명적 예시가 아니라, 아직 미지에 둘러싸인 지각과 인지 과정을 공식화하는 모델이자, 정신 작용의 기계적 재현이라는 야심을 합리화하는 근거였다. 카프라에 따르면 기계론적 과학과의 연결성, 군사 권력과의 유착 때문에 사이버네틱스는 여타 과학 체계들 가운데서도 그 출발부터 높은 위세를 누릴 수 있었다.¹⁰⁶ 뒤이어, 역시 전쟁의 목적에 부응하기 위해 개발된 디지털 컴퓨터가 곧바로 민간으로 확산, 산업 사회를 새로운 단계로 진입시킴에 따라 정보 공학적 사고는 시대의 패러다임으로 부상했고, 다른 한편에서 컴퓨터는 냉전 체제 핵전력 기반 군 지휘·통제 시스템의 근저를 지배함으로써 체제 존립의 주역으로 군림할 수 있었다.

이 같은 연루에 의해 컴퓨터 연구자들이 구사하는 언어는 군사 영역에서 조달된 메타포로 가득했고, 30년 이상 뇌 연구와 인지 과학의 주류가 다시 이 정보 처리의 렌즈로써 인간 정신 활동의 다층을 검증하려 시도함에 따라 각 분야

¹⁰⁵ P. N. Edwards, *The Closed World* (Cambridge, MA: MIT Press, 1996), 186.

¹⁰⁶ Capra, *Web of Life* (cit. n. 30), 69.

에서 통용되는 개념과 용어가 혼합되기에 이른다.¹⁰⁷ 이를테면 ‘외부로부터 입수한 정보를 대뇌 중앙 프로세서가 처리, 여기서 도출된 명령이 각 신경계를 통해 함으로써 지정된 목표를 수행한다’라는 식의 진술이 그것이다. 실제로 이 같은 정보 처리 도그마는 아직까지 효력을 발휘하는 중이며, 이때의 어휘들은 지금도 폭넓게 유통되는 편이다.

정보 채널로서의 뇌, 컴퓨터와 정신 작용의 유비가 아직 틀렸다고 할 수 없다. 그러나 사이버네틱스, 인지주의 양측이 채택한 것은 자신들이 이미 자신 있게 통제하고 조작할 수 있는 대상을 앞세워 이를 다시 주체에 대입하는 방식이었기 때문에, 심리와 정신의 블랙 박스를 열어젖혔다기보다 전쟁 수행과 컴퓨터 개발의 노하우로 다져진 가상의 모델로 그 미지의 공백을 채워 넣었다고 하는 편이 적합하다. 덕분에 인간 내면의 의사 결정 체계는 정연한 논리와 규칙으로 공식화될 수 있었고, 나름 유효한 방법이기엔 진척이 있었지만 한계 또한 동반했다. 대안으로 제시된 연결주의, 앞서 언급한 체화된 인지와 같은 발현형 인공지능은 바로 그 한계에 따른 반발의 기류였고, 기성 진영이 돌발 변수나 노이즈로서, 억제돼야 마땅한 속성으로 치부한 조절 불능, 예측 불가의 생물학적 본질의 모호성은 이 새로운 기류 속에서 부활할 기회를 잡는다.¹⁰⁸

그런데 애쉬비의 발상은 이런 변증법적 역사의 수순을 건너뛴다. 이제 막 사

¹⁰⁷ Ibid., 70.

¹⁰⁸ W. S. McCulloch and W. Pitts, “A Logical Calculus of the Ideas Immanent in Nervous Activity,” *The Bulletin of Mathematical Biophysics* 5:4 (1943), 115-133. 매컬록과 피츠의 이 논문으로부터 계보를 잇는 연결주의의 신경망은 기호주의에 밀려 침체를 겪지만, 80년대 비선형성 기반의 병렬 분산 처리 방식이 개척되면서 진일보할 수 있었다. 최근의 딥 러닝(deep learning)은 그 연장선상에 있다. 한편 표상 처리나 중앙 통제 장치 없이, 행동을 조절하는 단순한 병행 프로그램들의 조합으로 로봇의 지적 행동을 구현한 로드니 브룩스(R. A. Brooks)의 방식이 발현형 인공 지능의 대표적 예시로 꼽힌다. 이들의 공통적 특징은 인식의 기반을 외부로부터 주어진 것이 아닌 시스템 스스로가 구성해내기 때문에 행위 패턴을 예단하기 힘들고, 어떤 의사 결정이 있을 때, 설계자라도 그런 선택이 이뤄진 이유를 역으로 추적하기 어렵다는 점이다. 개발자 데미스 하사비스(D. Hassabis)가 알파고의 능력치에 대한 질문에 “놀라고 할 말을 잃었다”고 애기한 것이나, 이세돌 9단과의 5국 진행 와중에 “알파고가 대국 초반 안 좋은 실수를 저질렀다. … 수를 모르겠다”고 트위터에 적은 것은 이런 자생적이며 미결정적 특성과 관계한다. 이서희, “하사비스 ‘놀라서 할 말 잃어…알파고 한계 알고 싶다’”, 『한국일보』, 2016. 3. 12, <https://goo.gl/3onuEK> (2016. 12. 15 접속); 김철오, “하사비스 CEO ‘알파고, 대국 초반 안좋은 실수… 수를 모르겠다’”, 『국민일보』, 2016. 3. 15, <http://goo.gl/p2fl49> (2016. 12. 15 접속). 결국 직관과 창의성은 이 같은 체계 내부 질서의 모호함과 복잡성이 갖는 수수께끼를 해명하는 데 효과적으로 동원된 개념이었다.

이버네틱스의 체계가 갖춰지는 시기였고, 이들과 근본 기초를 공유했음에도 말이다. 호메오스탯을 비롯한 메이시 회의장에 공개된 기계 장치들은 생명체를 모의했고 실제 그들 행동을 분석해 설계됐지만, 일단 시연을 통해 외견상 유사성이 입증되면, 그 장치가 오히려 복잡한 인간 행동과 지능을 해명하는 모델로 논의되는 식이었다. 애초에 R-W-B가 논문 첫 페이지에서 “내재 조직을 생략한 채, 오직 입/출력 차원에서 접근한다”라고 선언했듯, 애쉬비 역시 “사이버네틱스는 존재가 아니라 행동하는 방식을 취급한다” 또는 “이것은 무엇인가가 아니라 이것은 무엇을 하는가의 질문을 던진다”라고 밝힌다.¹⁰⁹ 따라서 관찰 가능한 자극-반응 패턴이 겹친다면, 인간의 내적 본질이 정말 그러한지의 여부는 쟁점이 아니며, 이 같은 ‘닫힌 상자’의 연구 태도와 시대 상황이 교차하는 지점에서 다양한 전쟁의 메타포가 양산될 수 있었다.¹¹⁰ 이는 곧 목적을 추구하는 행위 주체에 유도 미사일과 엇비슷한 체계가 탑재돼있다는 의미이기도 했다.

과연 우리의 뇌가 서보메커니즘처럼 작동할까? 여기서 애쉬비는 방향타를 달리 잡는다. 폭탄 제어 장치라는 전쟁의 재료였지만, 일말의 음험함 없이 생물학적 조직에 충실한 신경계를 모의했고, 여기엔 그 어떤 목적론의 메타포도 매개되어 있지 않았다. 이 장치가 곧 적응의 뇌, 환경과 얹힌 신경계라는 것이다. 진화 생물학자 에른스트 마이어(E. Mayr)의 다음 진술은 호메오스탯을 이해하는 단서가 될 수 있다.

언제, 어떤 배를 격침할지는 어뢰가 아니라 그 어뢰를 사출하는 해군 함장이 결정한다. 교정 피드백은 목적 추구 행위의 정확도를 높이지언정 행위 자체를 지정할 수 없다. 피드백 장치는 프로그램을 번역하는 동안 가동되는 실행적 기제에 불과하다. 따라서 앞 예시들은, 이것의 존재 여부로 합목적적 프로세스를 규정한다는 점에서 강조점이 잘못 찍혀있다. 피드백 장치는 프로그램의 중개자지만, 목적 달성의 기본 원리 차원으로 따져보면 지극히 부수적이다.¹¹¹

¹⁰⁹ Rosenblueth et al., “Behavior” (cit. n. 6), 18; Ashby, *Introduction* (cit. n. 84), 1.

¹¹⁰ A. Rosenblueth and N. Wiener, “The Role of Models in Science,” *Philosophy of Science* 12:4 (1945), 316-321. 논문에 따르면 과학적 실험 대상의 완전한 투명화는 애초에 불가능하기 때문에 항상 제한된 터미널을 통해서만 접근할 수밖에 없다. 따라서 그 터미널의 입출력 데이터로 구성되는 모델이 실제 실험 대상을 대체해 온 역사가 과학 실험 전반을 관통해 왔다고 기술한다. 그들 관점에서 모든 과학적 실험과 증명은 ‘닫힌 상자’의 문제를 떠안는다.

¹¹¹ E. Mayr, “Teleological and Teleonomic: A New Analysis,” *Evolution and the Diversity of Life* (Cambridge, MA: Belknap Press, 1976/1997), 383-406 중 391.

그가 사이버네틱스 모델의 오류를 지적했다면, 오직 침잠하기 위해 부산하게 요동치는 호메오스탯의 역설은 이곳 영토에 흡수되지 않는다. 그러나 이 뇌 기계 역시 정상치로부터의 이탈을 보정해 나간다는 면에서 피드백 장치의 범주에 머무른다. 그렇지만 다시, 기성 사이버네틱스 기계들이 설계자의 통제와 프로그래밍에 기대야 했다면, 호메오스탯은 그런 의존성을 탈피한다. 메이시 발표 3년 전, 애쉬비가 컬럼에서 밝힌 내용은 이미 마이어의 진술을 상당히 포괄하고 있었다.

인공 뇌의 제작에는 현재로선 시간과 노력 이상의 것이 필요하다. 그러나 명확히 해야 할 부분이 하나 있다. 그것이 제대로 된 인공 뇌라면 영리함을 자체 발전시켜 나가야 한다는 점이다. ... 그 수행 능력이 얼마나 현란한 것이든 간에, 설계자에 의해 집행된 부분은 구체적으로 얼마큼이며, 과연 어느 정도가 기계 스스로가 기여한 부분인지 항상 의문을 가져야 한다.¹¹²

주도권 이양을 위해 그가 자신의 뇌 기계에 부여한 것은 ‘무작위적 선택’의 권한뿐이다. 오직 우연성을 해법삼아 위기를 자체 극복하도록 했으나, 그 간절한 해법이 고도로 복잡한 행동을 낳는 근원이며, 학습을 추진하는 동력임을 밝혔고, 궁극적으로 설계자의 권능에서 벗어나는 열쇠가 될 수 있음을 공표한다. 같은 회의에서 애쉬비가 이어간 다음 발표, “기계 체스플레이어”가 이 내용을 담고 있었다.¹¹³ 여기서 그는, 뇌 기계가 마치 브라운 운동처럼 제멋대로의 선택을 이어감으로써 어떻게 설계자를 능가할 수 있는지를 설명한다. 필승의 전법을 빈틈없이 구사하도록 알고리즘을 완벽히 새긴다 해도, 기계는 영원히 프로그래머의 실력을 추월할 수 없다. 반면 말들의 이동 가능한 포지션의 랜덤 결정과 무수한 경기 기회가 기계 플레이어에게 허용된다면, 빠른 패배를 몰고 온 말들의 이동 라인부터 회피를 거듭하고 이로써 추려진 선택을 계속 갱신함으로써 어느덧 프로그래머의 수준을 초월한다는 것이다. 여기서 게임의 승리와 패배는 기계가 마주한 환경의 긍정, 부정적 자극으로 활용되는 셈이다.¹¹⁴

¹¹² W. R. Ashby, “The Electronic Brain,” *Radio-Electronics* (March 1949), 79.

¹¹³ W. R. Ashby, “Mechanical Chess Player,” H. von Foerster, ed., *Cybernetics, Transactions of the 9th Conference, March 20-21 1952* (New York: Josiah Macy Jr. Foundation, 1953), 151-154.

¹¹⁴ 당시엔 이론상의 얘기였고 이런 전략을 구사하려면 엄청난 개수의 유닛이 얹힌 방대한 네트워크가 필요하다는 점, 이 때 시연된 기계식 적응 속도로는 절대 불가능하다는 사실 역시 인식

사이버네틱스 출범 당시 전면에 내세운 것은 노이즈로 일그러짐 없는 정보의 패턴이었다. 지능적 행위의 근간은 통신 체계 기반의 명령과 제어의 절차였으며, 이것은 신호의 흐름이 물질의 체계를 관통하는 모델이다. 한편, 그들이 천공 카드나 자기 테이프를 통해 제어계로 읽어 들인 추상적 정보가 원소 단위로 조합되는 행위를 연속 조합해 생명체의 복잡한 행동을 직조할 수 있다고 예견했을 때, 애쉬비는 물리적 대상들이 얹혀 불규칙적이며 파상적 이벤트를 이어나가는 분방한 관계 속에서 그것이 빚어질 수 있음을 보여줬다. 여기엔 일괄적 명령과 규칙 대신, 자연 선택과 지능의 상관성에 대한 암시가 담겨 있었다. 사이버네틱스가 물질의 상호 작용에 눈을 돌리게 된 계기가 정보의 패턴만으로 복잡성을 설명하기 어렵다는 각성 때문이었지만, 애쉬비는 너무 이른 시점에서 다른 해법을 준비하고자했던 것이다.

역사의 흐름에서 위너나 매컬록, 폰 포에르스터와 같은 인물들에 비해 애쉬비에 대한 조명이 상대적으로 약했던 이유는, 그가 1, 2차 진영의 어느 한 편에서 혁혁한 입지를 구축하기보다, 마치 단절면이 형성된 것처럼 보이는 양측을 유연한 흐름으로 메우고, 통합된 회로로서 서로를 연루시키는 역할에 더 큰 비중이 실렸기 때문일지 모른다. 물론 이 같은 역사적 판단이 맞다면 이는 그가 의도한 바가 아니며, 단지 뇌를 향한 보다 현실적이며 냉담한 접근에서 빚어진 결과였다.

(투고: 2016년 4월 8일, 심사 완료: 2016년 9월 9일, 게재 확정: 2016년 12월 8일)

하고 있었다. 하지만 오늘날 머신 러닝의 방법론으로서 행동 결과에 따라 긍정, 부정의 자극을 주어 기계 스스로 선택의 방향을 긍정적 포상을 추구하는 쪽으로 잡아가게 하는 강화 학습이 시행된다는 점, 현재 이 방법이 인공 지능에게 체스나 바둑 뿐 아니라 범용적 문제의 해결을 가르치는 주된 방향이라는 점에서, 이에 관한 어렵פות한 개요가 애쉬비에게 그려져 있었음을 추정해볼 수 있다.

A Study on W. R. Ashby's Artificial Brain, Homeostat: Focusing on His Position in the History of Cybernetics

KO Kyu-Heun (Sungkyunkwan University)

Abstract. Ashby assumes a unique position in the history of cybernetics. Alongside N. Wiener, J. Bigelow, and C. Shannon, he helped create the academic trend. With his contemporaries, he shared the goal of artificially reproducing complex human behavior by exploring the common ground of machinery and organisms. However, Ashby's approach in his demonstration at the 1952 Macy Conference was so shocking to the point that it caused the established cybernetics models to crack. Because of such novelty it was difficult for his colleagues to accept his view. At the core of the discussion was the Homeostat. With this device, Ashby integrated all his academic hypotheses on the interactions between the nervous system and the environment. Since the method in which this mechanical brain adapted itself to and learned the environment had already predicted the wave of the next generation cybernetics, his research had both ambiguous and ambivalent aspects. Focusing on Ashby's Homeostat, this article will approach the integrated essence of the biological process and cognition into which he gained insight, using the notions of 'random mechanism' and 'ultrastability' that formed the basis of this device. These two core elements not only clearly highlighted the difference between the established cybernetics models and the Homeostat, but also embraced the methodological detour of cognitive science and artificial intelligence afterward. Accordingly, this article attempts to relocate Ashby's position, typically described as a founding member of cybernetics, as the instigator of second-order cybernetics and the originator of the recent discussions in the biology of cognition and artificial life.

Keywords. Homeostat, W. Ross Ashby, ultrastability, step-mechanism, self-organization, artificial brain, second-order cybernetics, adaptation as equilibrium