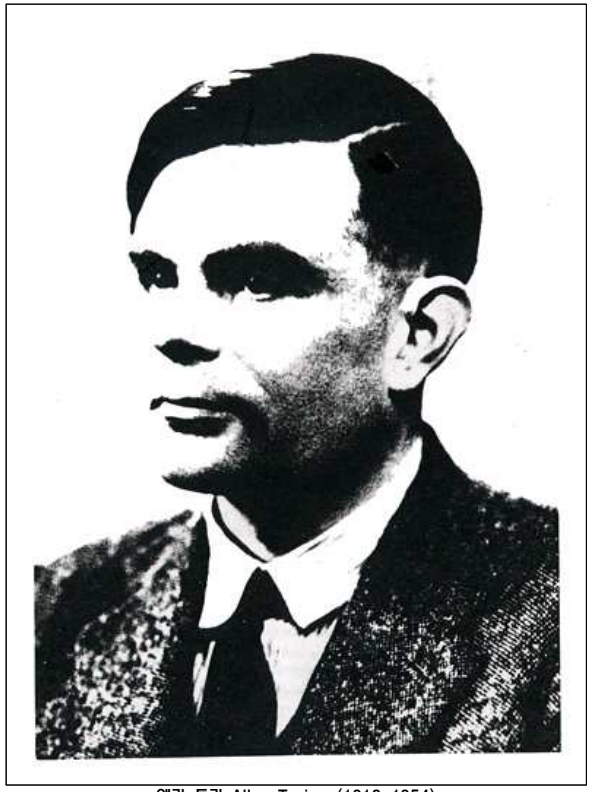
이 번역은 인간의 지능을 기계로 모의할 수 있음을 논의하여 인류 지성사에 새로운 길을 열어 놓은
튜링(1950) "계산 기계와 지능" 그리고 하춰즈(1995) "앨런 튜링과 튜링 기계"를
제가 가르치는 경상대학교 국어교육과 학생들에게 읽히기 위한 것입니다.
자유롭게 이용하시되, 만일 글로 인용할 경우에는 출처를 밝혀 주시면 고맙겠습니다.
김 지 홍 (http://nongae.gsnu.ac.kr/~jhongkim)



앨런 튜링 Allan Turing (1912~1954)

Allan Turing앨런 튜링(1950) "계산 기계와 지능"
《정신 : 심리학·철학에 대한 계간 비평》 제 59호[1)]

§ 1.모방 게임

"기계가 생각할 수 있을까?"라는 질문에 대해 생각해 보자. 이 물음에 대해 논의하기 전에 먼저 '기계'와 '생각하다'라는 말을 명확하게 정의해야 한다. 그 단어들(기계·생각하다)의 일반적인 용법을 가능한 한 많이 반영하는 정의를 내릴 수도 있지만, 이러한 태도는 위험하다. 만약 '기계'와 '생각하다'라는 단어의 의미가 그 단어의 일상적 쓰임을 살펴봄으로써 정의 내릴 수 있다면, "기계가 생각할 수 있을까?"라는 물음의 의미와 그 물음에 대한 대답 또한 갤럽 조사와 같은 통계적 조사에서 찾아질 수 있다는 결론에 이르게 된다. 이것은 불합리하다. 그렇게 정의를 내리기보다는, 그 물음(기계가 생각할 수 있을까)과 밀접하게 관련 있으면서도 비교적 덜 애매한 말로 표현되는 다른 것으로써 그 물음을 대신하고자 한다.

그 문제의 새로운 형태를 '모방 게임'(흉내내기 놀이)이라고 부를 수 있는 게임으로써 표현할 수 있다. 게임은 세 사람, 즉 한 남자(A), 한 여자(B), 성별에 관계없는 질문자 한 명(C)으로 행해진다. 질문자는 다른 두 사람(남자 A·여자 B)과는 따로 떨어진 방에 머물러 있다. 질문자에 대한 게임의 목표는, 나머지 두 명중 누가 남자이고, 누가 여자인지를 결정하는 것이다. 질문자는 그들을 X와 Y 표시로 구별하고, 끝에 가서는 "X가 A이고, Y가 B이다."라고 하든지, "X가 B이고, Y가 A이다."라고 말하게 된다. 질문자한테는 A와 B에게 질문하는 것이 허용되어 있다.

C : 당신 머리카락의 길이에 대해 이야기해 주시겠습니까?

자, 지금 X를 실제로 A라고 가정하자. 그러면 A는 대답을 해야 한다. 이 게임에서 A의 목표는 C가 틀린 판단을 하도록 만드는 것이다. 그러므로 그는 "나의 머리는 단발이며, 가장 긴 것이 약 3.5㎝(9인치)입니다."라고 대답할 수도 있다. 목소리 내용이 질문자의 판단에 도움 주지 않도록, 그 대답은 글로 씌어지거나 타이프로 쳐져야 한다. 이상적인 방법은 두 방 사이에 전신 타자기를 두는 것이다. 그 매개물(전신타자기)은 질문과 대답을 반복할 수 있다. 세 번째 역할자(B)가 이 게임에서 해야 할 임무는, 질문자를 돕는 것이다. 아마도 그녀가 취할 수 있는 가장 좋은 전략은, 진실된 대답을 하는 것이다(여자는 진실된 대답을 하고, 남자는 거짓말을 하기 때문이다). 그녀는 자신의 대답에 덧붙여 "나는 여자이며, 그가 하는 말에는 귀를 기울이지 마시오."와 같이 말을 할 수 있다. 그러나 그것은 남자도 비슷한 말을 할 수 있기 때문에 쓸모없다.

우리가 지금 "기계가 이 게임에서 A의 역할을 할 때 어떻게 될까?"라는 질문을 던진다. 게임이 이와 같이 행해질 때, 질문자는 A역할을 남자가 맡았을 때만큼 자주 잘못 판단할까? 이 물음은 처음 질문 "기계가 생각할 수 있을까?"를 대신하게 된다.

---

1) 이하에서 번역자의 각주는 따로 표시하지 않고, 원저자의 각주를 <원저자 주석>라고 따로 표시하였다.
이 글은 인류 지성사 흐름에서 처음으로 기계가 생각할 수 있다는 착상을 가장 구체적으로 완벽히 구현해 낸 논문이다. 오늘날 인공 지능에 이용되는 컴퓨터를 '범용 컴퓨터' 또는 '튜링 기계'라고 부르는 이유도(또는 합쳐서 Universal Turing Machine으로도 부름), 기계가 정신을 갖추기 위한 필요·충분조건을 처음으로 형식화해 내었기 때문이다. 기계가 입출력 장치와 기억 창고, 그리고 이를 조정하는 중앙 연산 부서만 갖고 있으면 정신의 원초적인 모습을 갖춘 것으로 보지만, 튜링은 기계가 '자기 학습 가능성'(self-learnability)을 구현하는지 여부에 의해서 인간과 기계가 구분될 수 있을 것으로 가정하였다. 최근의 컴퓨터나 로봇 전시장에서는 이런 진술 때문에 기계도 자기 학습성을 구현함을 보여 주려는 시도를 꾸준히 보여 주고 있고, 일부 상당한 성공을 거두고 있다. 기계가 완벽히 자기 학습성을 구현한다면, 더 이상 인간과 기계를 구분시켜 줄 척도가 없게 된다. 요 며칠 전 영국의 물리학자 스티븐 호킹은 기계가 인간을 지배할 날이 올 것임을 경고한 바 있다. 자기 학습성이란 구분점이 무너진다면 이를 대체할 만한 구획성을 아무도 아직 제시해 보지 못하고 있으며, 이러 노선에서는 기계나 인간이 모두 정신적 존재의 하나일 뿐이다. 이 글을 읽으면서, 우선 천재는 어떻게 사고를 하는지 그 핵심을 이해하는 일에 힘을 쏟기 바란다. 최소한 10번 이상을 정독해야 할 것이다. 세계적으로 그의 1936년 논문('계산 가능한 숫자에 대하여, 결정 가능성 문제에 응용과 함께') 출간 50주기를 기념하여 출간된 튜링의 생각에 대한 재평가는 Rolf Herken ed.(1995)《보편 튜링 기계: 반세기 연구 *The Universal Turing Machine : a half-century survey*》(Springer-Verlag)에 들어 있다.

§ 2.새로운 질문에 대한 비판

"새로운 질문에 대한 답은 무엇인가?"라는 물음뿐만 아니라, "새로운 질문이 연구할 가치가 있는 것인가?"라는 물음을 던지기도 한다. 우리는 뒤에 있는 물음을 추구하기로 하겠는데, 더 따지지 않고 그럼으로써 끝없는 질문들로의 퇴행을 차단하려고 한다.

이 새로운 문제는 인간의 물리적 능력과 지적 능력을 아주 날카롭게 구분해 주는 장점을 지니고 있다. 어떠한 기술자나 화학자도, 인간의 피부와 구별할 수 없는 물질을 만들 수 있다고 주장하지는 않는다. 그것이 미래의 어느 날 가능하게 될지 모른다. 설사 이러한 기술이 이용될 수 있다 하더라도, 기계에다 그 같은 인공적인 살을 입힘으로써 '생각하는 기계'를 더 인간답게 만들려는 시도가 쓸데없음을 깨닫게 될 것이다. 질문자가 다른 피실험자(피험자)들을 보거나 만지거나, 또는 그들의 목소리를 듣지 못하게 하는 조건에서, 모방 게임(우리가 문제를 설정한 형태)은 이러한 사실을 보여 준다. 또 다른 장점은 다음과 같은 질문과 대답을 통해 알 수 있다.

질문 : '포스 강의 다리' (Forth Bridge)라는 소재로 18행의 정형시 한 편을 지어 줄래요?
대답 : 이것으로 나를 시험하지 마십시오. 나는 시를 쓸 수가 없습니다.
질문 : 34,957와 70,764를 더하십시오.
대답 : (30초정도 있다가 대답을 함) 105,621입니다.
질문 : 서양 장기를 둘 줄 아십니까?
대답 : 예.
질문 : 나는 K1 자리에 K를 가지고 있고 다른 것은 없습니다. 당신은 K6 자리에 K를,
R1 자리에 R를 가지고 있습니다. 당신 차례입니다. 어떻게 하시겠습니까?
대답 : (15초 정도 후) R8 자리에 R을 놓겠습니다.

이들 질문과 대답은 인간의 어떤 면들을 알고자 할 때, 그것을 보여 주는 데 적합한 듯하다. 우리는 기계가 미인 대회에서 꼴찌를 도맡는다고 하여 기계를 처벌하기를 원하지도 않으며, 비행기와의 경주에서 인간이 진다고 하여서 인간을 처벌하기를 바라지도 않는다(즉, 물리적 능력은 중요하지 않다). 우리의 게임 상황은 이러한 무능력을 무의미하게 만든다(물리적 능력은 모방 게임에서는 무의미하다는 뜻이다). 만약 도움이 된다면 '피험자들'은 그들의 매력이나 힘, 또는 영웅심을 자랑할 수 있다. 그러나 질문자는 실제로 증명을 요구할 수는 없다(방 사이에 있는 타자기로 질문과 대답만 주고받기 때문임).

이런 게임은 승부가 기계한테 너무 불리하다는 점에서 비판받을 수도 있다. 만약 인간이 기계인 척한다면, 인간은 분명 매우 서툰 연기를 할 수밖에 없다. 인간은 산수에서 부정확하고 느리기 때문에 금방 발각될 것이다. 인간의 사고와는 아주 다르지만 그래도 사고라고 불러야 하는 어떤 것을 기계가 해서는 안 되는 것일까? 여기에 대한 반대는 매우 강하다. 그럼에도 불구하고, 기계가 모방게임에서 만족스럽게 행동하도록 만들어진다면, 적어도 우리는 이런 반대 때문에 속을 태울 필요는 없다.

'모방게임'을 할 때, 기계가 할 수 있는 최선의 전략은 아마 인간의 행동을 모방하는 것 이외의 다른 어떤 것이라고 주장할 수도 있다. 그럴지도 모르지만, 나는 그것이 큰 효과가 있으리라고는 생각하지 않는다. 여하튼, 여기서 게임의 이론을 고찰할 생각은 전혀 없다. 기계가 할 수 있는 최선의 전략은 인간이 자연스럽게 대답할 수 있는 내용을 모방하는 것이라고 생각한다.

§ 3. 게임에 관련되는 기계들

우리가 1장에서 제기한 물음은, '기계'라는 단어의 의미를 상세하게 설명하고서야만, 아주 명확해질 것이다. 우리가 만들고자 하는 기계 속에, 모든 종류의 기술이 쓰일 수 있기를 바라는 것은 당연하다. 또한 우리는 작동은 되지만 시험적 방법으로 만들었기 때문에, 그것을 만든 사람들조차도 작동 방식을 만족스럽게 기술할 수 없는 기계를 만들 가능성이 있음을 인정한다. 끝으로, 우리는 평범하게 태어난 사람들을 '기계'의 개념에서 배제시키고 싶다. 이 세 조건을 만족시키는 정의를 내리는 것은 어렵다. 예를 들어, 어떤 사람들은 일관성 있게 생각하는 기계를 만들기 위하여 기술자 팀이 모두 동일한 성(性)이어야 한다고 주장할지도 모른다. 그렇지만, 아마 사람의 피부 세포 하나로부터 완전한 개체를 길러내는 것이 가능하기 때문에, 이 점은 만족스럽지 못하다. 그렇게 하는 것은 아주 칭찬 받을 만한 생물학적 기술의 개가일지 모르나, 우리는 그것을 '생각하는 기계를 만들어 내는' 경우라고 고집하지 않을 것이다. 이것은 우리로 하여금 모든 종류의 기술이 허용되어야 한다는 요구를 포기하도록 부추긴다. '생각하는 기계'에 대한 현재의 관심은, 보통 '전자 컴퓨터' 또는 '디지털 컴퓨터'[2]라 불리는 특정한 종류의 기계에 의해 고조되어 왔다는 점에서 더욱 그러하다. 이 제안에 따라, 우리는 디지털 컴퓨터만 게임에 참여하도록 허용한다. 이러한 제약은 언뜻 보기에 매우 극단적인 것처럼 보인다. 나는 그것이 실제로 그렇지 않다는 것을 보여 주려고 한다. 그러기 위해서는 이들 컴퓨터의 본질과 특성을 짤막하게나마 설명할 필요가 있다.

또한, '사고'(思考)에 대한 우리의 기준과 같이, 기계와 디지털 컴퓨터를 동일시하는 것은, (나의 신념과는 정반대로) 디지털 컴퓨터가 모방 게임에서 만족스러운 결과를 보일 수 없다고 판명될 경우에만 흡족하지 못할 것이다.

이미 많은 수의 컴퓨터가 정상적으로 작동하고 있다. 그래서 "왜 그 실험이 곧장 행해지지 않는가? 게임의 조건을 만족시키기는 쉽다. 많은 질문지(質問紙)들이 이용될 수 있고, 얼마나 자주 옳은 판단을 내리는지에 대하여 통계가 나올 것이다."와 같이 질문할 수 있다. 이 질문에 대한 간단한 대답은 다음과 같다. 우리는 모든 디지털 컴퓨터가 모방 게임에서 임무를 잘 수행하는지, 또는 현재 이용될 수 있는 컴퓨터들이 기능을 잘 수행할 수 있는지에 대해서 질문하고 있는 것이 아니라, 그 임무를 잘 수행할 수 있는 컴퓨터가 있는지를 묻고 있는 것이다. 그러나 이것은 단지 간단한 대답에 불과하다. 우리는 뒤에 이 문제를 다른 관점에서 살펴볼 것이다.

§ 4. 디지털 컴퓨터

디지털 컴퓨터의 바탕을 이루는 생각은, 이러한 기계들(디지털 컴퓨터)이 휴먼컴퓨터[3]가 모든 작동들을 수행하도록 고안되어 있다는 사실을 말함으로써 설명될 수 있을 것이다. 휴먼컴퓨터는 정해진 규칙에 따라 작동한다. 휴먼컴퓨터는, 세세한 것에 이르기까지, 이들 규칙에서 벗어날 수 없다. 우리는 이들 규칙들이 휴먼 컴퓨터가 새로운 일을 수행할 때마다 변경될 수 있는 규칙 목록집 속에 들어 있다고 가정한다. 또한 휴먼 컴퓨터는 계산할 때, 필요한 종이를 무제한으로 공급받을 수 있다. 기계는 자그마한 기계 상태로 곱셈과 덧셈을 할 수도 있다. 그러나 이것은 중요하지 않다. 만약 우리가 위의 설명을 정의로서 사용한다면 우리는 순환논법에 빠질 위험이 있다. 바람직한 결과를 도출해 내는 방법을 간략히 뼈대만 이야기함으로써 순환논법을 피할 수 있다. 디지털 컴퓨터는 대충 세 부분으로 구성되어 있다고 볼 수 있다.

2) 디지털 컴퓨터는 지금 쓰고 있는 개인용 컴퓨터들이다. 정보를 0,1이라는 정보로 복사해 줄 수 있으며, 어떤 내용이든지 있고 없음의 합으로 파악함으로써, 전체 내용을 기억 창고에 담아둘 수 있다는 점이 특징이다. 개인용 컴퓨터나 정보 통신망에서 주고받는 내용들이 모두 디지털 방식으로 이루어지고 있기 때문에, 정보화 시대는 디지털 시대라고도 부른다.

3) 인간처럼 피부와 감정 따위를 갖는다고 묘사되는 컴퓨터임.

(ⅰ) 저장(장치)
(ⅱ) 실행(장치)
(ⅲ) 제어(장치)

저장장치는 정보를 저장하는데, 그 종이가 계산을 하는 종이이든, 규칙 묶음을 인쇄하는 종이이든 간에 상관없이 휴먼컴퓨터의 종이에 해당한다. 휴먼컴퓨터가 머릿속으로 계산을 하는 한, 저장장치의 일부는 그것의 메모리에 해당한다.

실행 장치는 연산에 관련된 개개의 다양한 동작을 수행하는 부분이다. 이들 개개의 동작들이 무엇이냐 하는 것은 기계마다 다를 것이다. 보통 "35406754445에다 7076345687을 곱하시오."와 같이 꽤 긴 연산도 수행할 수 있지만, 어떤 기계는 "0을 쓰시오."와 같이 아주 단순한 동작만이 가능하다.

휴먼컴퓨터에 들어 있는 '규칙의 책'(Book of rules:프로그램 순서도)은 기계(디지털 컴퓨터)에서는 저장부분에 의해 해당된다고 말했는데, 그것은 '지시표'라고 불린다. 이들 지시들이 정확하고 올바른 순서로 수행되는지를 살펴보는 것이 제어장치가 하는 일이다. 제어장치는 이런 기능을 하도록 구성되어 있다.

저장부(저장장치)에 들어 있는 정보는 대개 적당하게 작은 크기의 묶음으로 나누어진다. 예를 들어, 어떤 기계에서 한 묶음은 10진 단위로 구성될 수 있다. 숫자들이 다양한 정보묶음이 저장된 저장부분에 어떤 체계적인 방식으로 할당된다. 지시표는 전형적으로 다음과 같다.

"6809 위치에 저장된 수를 4302 위치에 저장된 수에 더하고, 그 결과를 4302 위치에 넣으시오."

물론 두말할 필요 없이 기계에서 지시는 영어로 표현되지 않는다. 그것은 아마도 6809430217과[4] 같이 기호화된 형태를 띨 것이다. 여기서 17은 두 숫자로 표현될 수 있는 다양한 동작 중의 하나를 말한다. 이 경우에 그 동작은 위에서 기술된 것 즉 "그 수를 더하시오."일 것이다. 지시가 10진 단위로 표시됨으로써 한 묶음의 정보를 형성한다는 것을 매우 편리하게 알 수 있다. 일반적으로 제어는 지시표들이 저장된 위치에 따라 순서대로 수행되도록 하나, 때로는

"5606에 저장된 지시를 따르고 거기서부터 계속하시오."

또는

"4505 위치에 00이 저장되어 있다면 6707 위치에 저장된 그 다음 지시를 따르고, 그렇지 않을 경우 곧장 계속하시오."

와 같은 지시표를 만날지도 모른다. 후자 형태의 지시표들은 어떤 조건이 충족될 때까지 일련의 동작들을 계속해서 반복할 수 있기 때문에 매우 중요하다. 그런 경우, 매번 반복할 때마다 새로운 지시를 따르는 것이 아니라, 똑같은 명령이 계속해서 반복된다. 이것을 가정생활에 비유해 보자. '토미'가 매일 아침 학교 가는 길에 구두수선공 집에 들러 어머니가 자신의 구두가 다 수선되었는지 알아보기를 바란다면, 어머니는 매일 아침마다 환기시킬 수 있다. 혹은, 토미가 학교 갈 때 보게 될 벽에다

"신발에 대해 알아보고 신발을 찾아오면 벽보를 떼어 버리시오."

라는 내용을 한번 붙여 둘 수도 있다.

독자들은 디지털 컴퓨터가 위에서 기술했던 그러한 원리에 따라 구성될 수 있고, 구성되어 왔다는 사실을 받아들여야 한다. 사실, 디지털 컴퓨터는 휴먼컴퓨터의 행동을 매우 비슷하게 흉내낼 수 있다.

4) 6809와 4302는 입력 내용이 되고, 17은 이 입력을 연산하는 부호에 해당함(더하고 저장하시오).

휴먼 컴퓨터가 이용한다고 말했던 '규칙의 책'(규칙에 대한 책)은, 물론 편리함을 위해 꾸며낸 것이다. 실제의 휴먼컴퓨터는 정말로 자기가 했던 것을 기억한다. 만약 복잡한 동작으로 이루어진 휴먼컴퓨터의 행동을 기계가 모방하기를 바란다면, 우리는 휴먼컴퓨터에게 그러한 행동을 했던 방법을 물어 본 다음, 그 대답을 지시표(명령표)의 형태로 바꾸어야 한다. 지시표를 만드는 것을 대개 '프로그래밍'이라 한다. 기계가 A라는 동작을 수행하도록 프로그램하는 것은, 기계가 A를 할 수 있도록 그것에다 적절한 지시표를 넣는 것을 의미한다.

디지털 컴퓨터에 대한 생각을 흥미롭게 변형시킨 것이, 임의의 무작위 요소를 가진 디지털 컴퓨터이다. 이들은, 주사위를 던지는 것이나 또는 그에 상응하는 전자적 과정을 포함하는 지시표를 가지고 있다. 예를 들어 그러한 지시 중 하나는

"주사위를 던져 그 결과로 나온 숫자를 저장장소 1000에 넣으시오."

일 수 있다. 때때로 그 같은 기계는 (내 스스로 이 말을 쓰고 싶지는 않지만) 자유의지를 가진 것처럼 기술된다. 기계를 관찰함으로써 그것이 임의의 무작위 요소를 가지고 있는지 어떤지를 결정하는 것은 정상적으로 가능하지 않다. 왜냐하면 비슷한 효과(현상)가 π의 소수점의 숫자 자리에 따라 선택을 하는 장치에 의해 일어난다.[5]

실제 디지털 컴퓨터는 대부분 제한된 저장 용량을 가지고 있다. 무한한 저장 용량을 가진 디지털 컴퓨터가 존재할 수 있다는 생각은 이론상으로 가능하다. 물론 한 번에 제한된 부분만이 사용된다. 제한된 부분만이 사용되지만 그러나 필요하다면 더 많은 부분을 사용할 수 있다. 그 같은 컴퓨터는 이론적으로 매우 흥미로운 것이며 무한한 능력을 가진 컴퓨터라 불린다.[6]

디지털 컴퓨터에 대한 생각은 오래되었다. 1828년부터 1839년까지 캠브리지 대학 루카치 교수직의 수학교수를 역임했던 찰스 배비쥐(Babbage)라는 수학자가 (산술 연산을 위한) 분석적인 동력기관을 설계했다. 그러나 그것은 완성되지 못했다. 배비쥐가 모든 필수적인 생각들을 가지고 있었음에도 불구하고, 그의 기계는 그 당시로서는 사람들의 관심을 끌 만한 장래성을 가지지 못했다. 그 기계의 속도가 분명 휴먼컴퓨터보다 빨랐으나 맨체스터(Manchester) 기계보다는 100배 느렸는데, 맨체스터 기계는 현대 기계 중에서 아주 느린 것에 속한다. 그 저장 장치는 톱니와 카드를 사용하여 순전히 기계적이었다.

배비쥐의 분석기관이 전적으로 기계적이었다는 사실은, 우리로 하여금 미신에서 탈피하는 데 도움을 줄 것이다.[7] 현대 디지털 컴퓨터는 전자식이며 그 신경 체계 역시 전자식이라는 사실에 종종 중요성이 부여된다. 배비쥐 기계는 전기로 작동되지 않으며, 또한 모든 디지털 컴퓨터가 어떤 의미에서 모두 똑같기 때문에, 전기를 사용한다는 것이 이론상으로 중요하다고 할 수 없다. 물론 전기는 빠르게 신호를 보내는 것과 관련된다. 그래서 이 둘 사이에서 어떤 관련성을 발견하는 것은 놀라운 일이 아니다. 신경체계에서 화학적 현상은 적어도 전기적 현상만큼 중요하다. 어떤 컴퓨터에서는 저장 체계가 주로 청각적이다. 그러므로 전기를 사용한다는 특징은 매우 피상적인 유사성에 지나지 않는다. 만약 우리가 그 같은 유사성을 발견하고자 한다면, 오히려 함수와 같은 수학적 유추를 해야 한다.

5) 이것은 결정 불가능한 명제이다. 예를 들어서, 정사각형이 있고 원이 있다고 하자. 하나의 정사각형은 원에 내접해 있고, 다른 하나는 원에 외접해 있다고 하자. 만일 내접한 사각형을 같은 간격으로 계속 늘여 가고, 외접해 있는 사각형을 계속해서 같은 간격으로 줄여 나간다고 하자. 이럴 때에 두 개의 사각형이 어느 지점에서 만나게 될 것인가? π가 똑 떨어지지 않는 한, 그런 사각형의 만남은 존재하지 않는다. 존재한다고 하더라고, 그 사각형의 존재는 2차원의 평면이 아니라, 휘어져 있는 3차원의 입체 모습을 지닐 수밖에 없을 것이다. 현재 우리의 지식으로는 π는 끝없이 발산하는 존재이며, 그 발산되는 숫자들에서 어떤 반복 흐름도 찾아지지 않고 있으므로 어떤 숫자가 어느 위치에 나타날 것인지에 대해서도 침묵할 수밖에 없다.

6) 지금 현재 이와 같은 기계를 '튜링 기계'(Turing machine)이라고 부르고 있다.

7) 디지털 컴퓨터가 반드시 전자식이어야만 한다는 생각을 미신이라 표현했다.

§ 5. 디지털 컴퓨터의 보편성

앞장에서 언급된 디지털 컴퓨터는 아마도 '이산 상태 기계'[8] 중의 하나로 분류될 것이다. 이들 기계들은 갑작스러운 도약이나 딸깍거림으로써 어떤 한 상태에서 명백히 다른 상태로 움직인다. 이들 상태들은, 그것들 사이의 혼동 가능성이 무시되기 때문에, 서로 분명히 구별된다. 엄밀히 말하자면 그런 기계는 없다. 모든 것들은 실제로 연속적으로 움직인다. 그러나 많은 종류의 기계들을 이산(불연속) 상태 기계로 생각하는 것이 더 유익하다. 예를 들어 점멸 스위치를 생각해 볼 때, 각각의 스위치는 명백히 켜지거나 명백히 꺼진다고 생각하는 것이 편리하다. 분명, 중간위치가 있다. 그러나 우리는 의식적으로 그것을 무시한다. 이산 상태 기계의 예로서, 초당 120도를 딸깍거리며 돌고 있으며, 외부에서 조작되는 막대(lever:지렛대)로 멈춰질 수 있는 톱니바퀴를 생각해 볼 수 있다. 게다가 톱니바퀴 중 한 위치에는 불이 켜지는 램프가 있다. 추상적이지만 이런 기계를 다음과 같이 기술할 수 있다. 기계의 내부상태(그것은 톱니바퀴의 위치에 의해 표현된다)는 q_1이거나 q_2 또는 q_3이다. 입력신호는 i_0 또는 i_1이다. (막대(지렛대)의 위치에 따라 입력 신호가 정해진다.) 어느 한 순간의 내부 상태는 다음 도표에서 보듯이 마지막 상태와 입력신호로 결정된다.

		마지막 상태		
		q_1	q_2	q_3
입력 신호	i_0	q_2	q_3	q_1
	i_1	q_1	q_2	q_3

출력신호는 내부 상태를 눈으로 볼 수 있게끔 외적으로 나타낸 것(빛)으로, 다음 도표로 기술된다.

상태	q_1	q_2	q_3
출력신호	0_0	0_0	0_1

이것은 이산 상태 기계의 전형적인 예이다. 그 기계가 제한된 수의 가능한 상태를 가질 경우에만 위와 같은 도표로 기술될 수 있다(무한한 상태를 가진다면 도표로 기술할 수가 없기 때문이다).

만약 그 기계의 초기 상태와 입력 신호들이 주어진다면, 모든 미래 상태들을 예측하는 것이 가능한 것처럼 보인다. 이것은

> "입자의 위치와 속력으로 표현되는 어느 한 순간의 우주의 완전한 상태로부터, 모든 미래 상태를 예측하는 것이 가능하다."

라는 라플라스(Laplace)의 생각을 연상시킨다. 그러나 우리의 예측은 라플라스의 예측보다 실행가능성이 더 많다. '전체로서의 우주' 체계 속에서는, 초기 상태에서의 아주 작은 실수가 훗날 엄청난 결과를 야기할 수 있다. 어느 순간 1조분의 1㎝만큼 전자 하나를 이동시킴으로써, 1년 후 한 사람을 눈사태로 죽게 할 수도, 또는 그 화를 면하게 할 수도 있다. 이러한 현상이 일어나지 않는다는 것은, '이산 상태의 기계'라 부르는 기계 체계의 본질적 특성이다. 뿐만 아니라 이상적인 기계 대신 실제 물리적 기계를 생각할 때조차도, 어느 한 순간의 상태에 대해 적절하고 정확하게 안다면, 몇 단계 후의 적절하고 정확한 지식을 얻을 수 있다.

8) 전문 용어로서, 한 단계에서 다른 단계로 넘어 가는 과정이 연속되어 있는 것이 아니라 완전히 단절되어 있으므로, 두 단계 사이에 중간 항이 없는 것을 말한다. 손목시계는 연속적으로 시침과 분침이 있으므로, 비-이산적이고 연속적이다. 그러나 전자 손목시계는 1분에서 2분 사이에 1.5분을 나타낼 수 없다. 바로 1분에서 중간 항이 없이 2분으로 넘어가 버린다. 이런 상태를 이산 상태, 또는 불연속 상태라고 한다. 전자를 다루는 분야에서는 전자 손목시계와 같은 것을 '디지털' 방식이라고 말하고, 이와 대립되는 옛날 시계(1분과 2분 사이의 중간 상태도 모두 표시할 수 있는 시계)를 '아날로그' 방식이라고 말한다.

앞서 말했듯이, 디지털 컴퓨터는 이산 상태의 기계류에 속한다. 그러나 그 같은 기계가 가질 수 있는 상태의 수는 어마어마하게 많다. 예를 들어, 현재 맨체스터에서 작동하고 있는 기계(맨체스터 컴퓨터)가 만들어낼 수 있는 가능 상태의 수는 2^{165000}, 즉 10^{50000}이다. 위에서 기술했던 딸깍거리는 톱니바퀴 — 세 가지 상태를 가짐 — 예와 이것을 비교해 보라. 왜 상태의 수가 그렇게 많은지를 아는 것은 어렵지 않다. 디지털 컴퓨터는 휴먼컴퓨터의 종이(천공 카드)에 해당하는 저장 창고를 가지고 있다. 종이(천공 카드)에 적힐 수 있는 상징들의 결합 중 어느 하나를 저장 창고에 기록하는 것은 틀림없이 가능하다. 간단히, 0에서 9까지의 십진수를 상징부호로 사용한다고 생각해 보자. 글씨체의 차이는 무시하기로 한다. 그 컴퓨터가 한 줄당 30개의 숫자를 쓸 수 있는, 50개 줄을 가진 종이 100장을 처리할 수 있다고 가정해 보자. 그럴 경우, 가능 상태의 수는 10100 × 50 × 30 즉, $10^{150,000}$이다. 이 숫자는 대략 세 대의 맨체스터 기계들의 상태 수를 합친 것과 맞먹는다. 일반적으로 'log2'(상태의 수:2의 지수임)가 그 기계의 '저장용량'이다. 그러므로 맨체스터 기계는 저장용량이 약 165,000이며, 우리가 앞에서 예로 제시한 톱니바퀴 기계는 약 $2^{1.6}$이다. 만약 두 기계를 결합시킨다면, 그 두 능력을 합한 새로운 기계가 만들어질 것이다. 이것은, 맨체스터 기계가 각각 2560 용량을 가지는 64개의 자기 트랙과 1280 용량을 가지는 8개의 전자 진공관을 포함하고 있다는 진술을 가능하게 한다. 잡다한 저장용량은 약 300이며, 전체 합쳐 총합이 174,380이다.[9]

이산 상태의 기계에 대응되는 도표가 주어진다면, 그 기계가 무엇을 할 수 있는지 예언할 수 있다. 이러한 계산은 디지털 컴퓨터를 도구로 씀으로써 수행될 수 있다. 디지털 컴퓨터가 그러한 계산을 충분히 빠르게 수행할 수 있다면, 어떤 이산 상태의 기계 행동도 모두 모방할 수가 있다. 그러면, 문제의 그 기계(B역할) —이산 상태의 기계—와 그것을 모방하는 디지털 컴퓨터(A역할), 그리고 그 둘을 구별할 수 없는 질문자로 구성되는 모방게임이 행해질 수 있다. 물론 이때 그 디지털 컴퓨터는 적절한 저장 능력을 가지고 있어야 하며, 충분히 빠르게 작동해야 한다. 더욱이, 모방되는 기계가 새롭게 바뀔 때마다 그것을 모방하는 기계(디지털 컴퓨터)도 다시 새롭게 프로그램 되어야 한다.

디지털 컴퓨터의 이런 독특한 특징(어떤 이산 상태의 기계도 모방할 수 있다는 점)은 그것이 보편기계(universal machines)라고 말함으로써 설명될 수 있다. 이러한 특징을 가진 기계들이 존재한다는 점은, 여러 가지로 다양한 계산을 수행하기 위해 그 각각마다 새로운 기계를 고안해 낼 필요는 없다(이때, 속도는 별개의 문제로 생각한다)는 결론을 이끌어 낸다. 그 다양한 계산들은 모두, 각각의 경우마다 적절하게 프로그램된 디지털 컴퓨터 한 대가 처리할 수 있다. 이러한 결과를 볼 때, 모든 디지털 컴퓨터는 어떤 의미에서 동등하다는 것을 알 수 있다.

이제 우리는 다시 3장의 마지막에서 제기된 문제를 생각해 보기로 하자. "기계가 생각할 수 있을까?"라는 물음은 "인간이 상상할 수 있는 컴퓨터 중에서 모방게임을 잘 할 수 있는 디지털 컴퓨터가 존재하는가?"라는 물음으로 대체되어야 한다고 잠정 제안했다. 만약 원한다면, 우리는 피상적이나마 좀더 일반적인 물음 —"모방게임을 잘 할 수 있는 이산 상태의 기계가 있는가?"—을 던질 수 있다. 그러나 보편적인 특질에 중점을 두어 본다면, 그 물음들 중의 하나는 거의 다음과 같은 뜻임을 알 수 있다.

> "자, C라는 어떤 특정한 디지털 컴퓨터에 초점을 맞추어 보자. 이 컴퓨터가 충분한 저장능력을 가지도록 개조하고, 그 수행 속도를 적당하게 증가시키며, 적절한 프로그램을 입력시킴으로써, B역할을 인간이 맡더라도, C가 모방게임에서 A의 역할을 만족스럽게 해 낼 수 있을까?"

§ 6. 근원적인 물음에 대한 반대 견해들과 이것들에 대한 방어

우리는 지금 "기계가 생각할 수 있을까?"라는 우리의 물음과 앞장의 끝에서 인용된 그 물음의 변형에 대한 논쟁을 시작할 준비가 되어 있다. 그렇다고 하더라도 그 문제의 원래 형태를 전적으로 내버릴 수는 없다. 왜냐하면 그 물음을 적절하게 적용했는지에 대해 의견이 서로 다를 것이고, 우리는 적어도 이런 관련 속에서 무엇이 언급되어야 할 것인지에 대해 귀를 기울여야 하기 때문이다.

9) 174,380 = 64 × 2560 + 8 × 1280 + 300

이 문제와 관련하여 먼저 내가 믿는 바를 설명한다면, 독자를 위해 문제를 단순하게 만들 수 있을 것이다. 우선 그 물음의 더 정확한 형태를 생각해 보자. 약 50년쯤 후에는 모방 게임을 아주 잘 하도록 함으로써 일반적인 질문자가 질문을 하고서 5분 뒤에 정확한 판단을 내리는 기회가 70% 이상 넘지 않도록 할 수 있는 컴퓨터를 프로그램할 수 있을 것인데, 대략 10^9의 저장 능력을 갖고 있을 것이라고 믿는다. "기계가 생각할 수 있을까?"라는 원래 질문은 토론하기에는 무의미하다고 생각된다. 그럼에도 불구하고, 지금은 말하기 힘들지만 20세기 말에는 이에 대해 더 잘 설명할 수 있을 것이다. 나아가, 이러한 신념들을 감추고서는 어떤 유용한 효과도 거둘 수 없으리라 생각한다. 과학자들은 어떤 증명되지 않은 추측이나 선입견에도 영향을 받지 않고, 잘 정립된 사실로부터 잘 정립된 사실로 확고부동하게 나아간다는 생각이 널리 퍼져 있지만 이것은 명백하게 틀린 생각이다. 만약 어느 것이 증명된 사실이며 어떤 것이 어림짐작인지를 명백하게 가려낼 수 있다면 어떤 해로운 일도 일어나지 않는다. 어림짐작(추측, 선입견)은 연구를 할 때 윤곽을 제시해 주기 때문에 매우 중요하다. 나는 지금 내 의견에 반대하는 견해들을 살펴보고자 한다.

(1) 신학상의 반대

생각한다는 것은 인간이 가진 불멸의 영혼이 하는 작용(기능)이다. 신은 모든 남녀에게 불멸의 영혼을 주었지만 다른 어떤 동물이나 기계에게는 그렇게 하지 않았다.[10] 그러므로 다른 어떤 동물이나 기계도 생각할 수 없다.

나는 이러한 견해를 전혀 받아들일 수 없지만 그런 견해를 가진 사람들이 이해하기 쉽도록 신학상의 용어로 그 견해를 반박해 보겠다. 동물을 인간과 한 부류로 분류한다면 그 주장이 더욱 확신성 있으리라 생각한다. 왜냐하면 내 생각으로는 인간과 그 외의 동물 사이보다는 전형적인 생물과 무생물 사이에 더 큰 차이점이 있기 때문이다. 정설로 신봉되는 그 관점(신학상의 관점) 이 다른 종교사회의 구성원에게는 어떻게 나타나는지를 생각해 본다면, 자의적이고 독단적이라는 사실을 더 명확히 알 수 있을 것이다. 여자는 영혼이 없다는 회교도의 생각을 과연 기독교도들이 받아들일 수 있겠는가? 이 문제는 제쳐 두기로 하고 주된 논의로 돌아가기로 하자. 위에서 인용된 주장(신학상의 주장)이 내게는 전지자의 전능에 대해 중대한 제약을 내포하고 있는 것처럼 생각된다. 물론 아무리 신일지라도 1과 2를 같게 만들 수는 없다. 그렇다고 해서, 적절하다고 생각한다면 신은 코끼리에게도 영혼을 불어넣을 수 있는 자유로운 힘을 가지고 있다는 것을 믿지 말아야 하는가? 우리는 그 전지전능한 존재가, 코끼리가 돌연변이를 일으켜 영혼을 필요로 하는 진보된 두뇌를 갖게 되었을 경우 그의 힘을 시험하리라 생각한다. 기계에 대해서도 이와 아주 비슷한 형태의 주장이 제기될 수 있다. 이 주장은 색다르다고 느껴지는데, 왜냐하면 그것을 '받아들이기'가 힘들기 때문이다. 그러나 이것은 단지, 기계가 영혼을 부여받을 적절한 환경을 가지고 있지 않다고 우리가 생각하고 있음을 의미할 뿐이다. 여기서 문제가 되는 환경은 이 논문의 나머지 부분에서 논의될 것이다. 아이를 출산하는 것이 신의 힘을 불경스럽게 빼앗는 것이 아닌 것처럼 그 같은 기계를 만드는 것도 그러해야 한다. 오히려 이 두 경우에서 우리는 그가 창조하는 영혼을 위해 집을 마련하는, 그의 의지의 도구에 불과하다.

그러나 이것은 단순한 생각에 불과하다. 나는 아무리 그 신학상의 주장이 지지되곤 했을지라도 그것으로부터 깊은 감명을 받지 못한다. 그 같은 주장들은 과거에서조차 자주 불만족스럽게 생각되어져 왔다. 갈릴레오가 살던 시대에는 "태양은 변함없이 고정되어 있어... 하루 내내 움직이지 않는다"(여호수와 10-13)와 "신이 지구를 중심으로 만들었으므로 어떤 경우에도 지구는 움직이지 않는다"(시편 105장)라는 의견들은 지동설의 충분한 반박거리였다. 오늘날 우리가 가진 지식으로는 그러한 논쟁은 쓸모없는 것 같다. 그러한 지식이 쓸모없을 때 매우 엉뚱한 영향을 미친다.

10) 원저자의 각주임: 아마 이 의견은 이단이다. 토마스 아퀴나스(버틀란트 러셀이 인용한 Summa Theologica 480 페이지)는 신은 영혼을 가지지 못한 인간을 만들 수 없다고 말한다. 그러나 이는 그의 힘에 실제적 제약이 있음을 의미하는 것이 아니라 인간의 영혼은 불멸이며, 그래서 파괴될 수 없다는 사실에 대한 결론일 뿐이다.

(2) '사실을 인정하려 하지 않는 태도'에서 기인된 반대

"기계가 생각한다는 결론은 너무 끔찍하다. 기계는 생각할 수 없다고 바라고, 믿자."

이 주장은 좀처럼 위와 같은 형태로 아주 공공연하게 표현되지는 않는다. 그러나 그러한 생각을 조금이라도 하고 있는 우리들 대부분에게 그 주장은 영향을 미친다. 우리는 인간이 다른 창조물보다도 약간 더 우월하다고 믿고 싶어 한다. 인간이 필연적으로 우월하다는 것이 증명될 수 있다면 더할 나위 없이 좋다. 왜냐하면, 명령을 내릴 수 있는 지위를 잃어버릴 위험이 없기 때문이다. 신학상의 주장이 널리 퍼져 있다는 점은 분명 이러한 감정과 관련 있다. 그 주장은 지식인들 사이에서 아주 강하다. 왜냐하면 그들은 다른 것들보다는 생각하는 힘에다 더 큰 가치를 두며, 이 힘을 바탕으로 인간이 우월하다는 신념을 가지려하기 때문이다.

나는 이 주장이 반박할 만한 가치가 있는 중요한 것이라고는 생각하지 않는다. 위로(위안)가 더 적당할 것이다(즉 그 주장은 반박할 가치도, 고려해 볼 가치도 없는 쓸데없는 생각이라는 말임). 아마 위로(위안)는 윤회에서 찾아져야 할 것이다.

(3) 수학 상의 반대

불연속 상태 기계의 힘에는 한계가 있음을 보여주는데 이용될 수 있는 수리논리학의 결과는 많다. 그 중에 가장 잘 알려진 것으로 괴델의 정리가[11] 있다. 그 정리는, 아무리 충분히 강력한 논리체계가 있다하더라도, 체계가 모순적이지 않다면, 그 체계 속에서 증명할 수도 없고 반증을 들 수도 없는 명제들이 있을 수 있다는 것이다.

춰어취(Church)나 클리니(Kleene), 롸저(Rosser)와 튜링(Turing)이[12] 도출해 낸, 어떤 점에서는 서로 비슷한, 또다른 결과들도 있다. 맨 마지막 사람의 결과가 더 생각하는데 편리하다. 왜냐하면 그것은 기계를 직접적으로 설명하는 반면 나머지 결과들은 비교적 간접적인 논증에서만 이용될 수 있기 때문이다. : 예를 들어 괴델의 법칙을 이용한다면 우리는 기계로써 논리체계를, 그리고 논리체계로써 기계를 기술하는 방법을 더 가져야만 된다. 문제의 그 결과는 무한한 용량을 가진, 본질적으로 디지털 컴퓨터인 기계의 유형에 관해서 설명한다. 그것은 그 같은 기계가 할 수 없는 것들도 있음을 말하는 것이다. 모방게임에서 질문을 받았을 경우, 아무리 충분한 시간이 주어지더라도 그 기계가 틀린 대답을 하거나 대답하지 못하는 물음이 있을 수도 있다. 물론 그 같은 물음은 많이 있을 수 있으며, 또한, 한 기계는 대답할 수 없으나 다른 기계는 만족스럽게 대답할 수 있는 물음들도 많이 있을 것이다. 우리가 가정하고 있는 물음은 '피카소를 어떻게 생각하는가'와 같은 유형의 물음이라기보다는 '예', '아니오' 대답을 요구하는 물음이다. "… 같은 속성을 갖는 기계를 생각하라. 이 기계가 어떤 질문에 한번이라도 예라고 대답할까?" 점으로 표시된 부분에는 5장에서 살펴보았던 몇몇 표준형태의 기계들에 대한 서술이 들어

11) 흔히 불완전성(incompleteness)의 정리로 알려져 있으며, 어떤 이성적인 체계를 구성하든지 반드시 결정불가능한 명제가 하나 이상 있다는 말로 요약된다. 달리 비유적으로 표현한다면, 청코너의 권투 선수와 홍코너의 권투 선수가 서로 권투를 하고 있는데, 이들을 판결할 심판이 없는 것이라고 말할 수 있다. 완벽한 논리 체계는 '공리 체계'(axiomatic system)를 근거로 하여 모든 정리와 따름 정리들이 도출되어 나온다. 그런데 정작 그 공리 체계는 완벽하다고 입증되지 않았다. 그 공리 체계의 완벽성이 입증되려면, 반드시 초 공리 체계가 상정되어야만 한다. 그런데 그 초 공리 체계 또한 완벽성이 입증되려면, 다시 초 초 공리 체계가 필요한 것이다. 이런 악순환의 고리를 끊을 도리는 논리 그 자체에는 없다. 다만, 우리가 잠정적으로 약속하여 어떤 공리 체계가 완벽한 것이라고 인정할 도리밖에 없다. 괴델의 정리는 24살 때 비엔나 대학의 박사논문으로 러셀과 화일헤드의 공저《수학 원리 *Principia Matimatica*》의 공리 체계가 완벽히 정립되지 않았음을 다룬 것이다. 이 글은 영어로 번역되어 Heijenoort(1967)《프레게로부터 괴델까지: 수리 논리의 원전 *From Frege to Gödel—A Source Book in Mathematical Logic*》(Harvard UP)에 실려 있다. Gödel(1930a) "The Completeness of the Axioms of the Functional Calculus of Logic" 외 3편.

12) 원저자 각주임 : 사람 이름이 이탤릭 글체로 된 것은 참고 문헌에 인용된 글들을 가리킨다.
번역자 각주: 튜링은 괴델과 아이슈타인이 있던 프린스턴 대학에서 박사학위를 춰어취의 지도 아래 받았다. 춰어취는 특히 유형에 대한 이론과 추상화 연산(abraction 또는 λ-operation)이 논리학과 의미론의 기초로 중요하게 거론된다. Church(1941)《람다 변환의 계산법 *The Calculi of Lamda-Conversion*》(Princeton Univ. Press)와 Church(1944)《수리 논리 입문 *Introduction to Mathematical Logic*》(Princeton Univ. Press)

갈 수 있다. 그 기계가 질문 받는 기계와 단순한 관련이라도 있다면, 대답이 틀리든지, 답이 준비될 수 없을 것이다. 이것이 수학 상의 결론이다. 그 점이 인간 지능을 가지고 있지 못한 기계의 무능력을 증명한다고 수학자들은 주장한다.

이 주장에 대해 짧게나마 답을 내린다면, 어떤 특정한 기계의 힘에는 한계가 있지만 인간의 지능에는 한계가 없다는 것이다. 그러나 여기에는 어떤 증거도 없다. 나는 이 견해가 쉽사리 꺾이리라고는 생각하지 않는다. 이들 기계에게 적절하고도 비판적인 질문을 할 때마다, 그리고 명확히 대답을 할 때마다, 우리는 이 대답이 분명히 틀릴 것임을 아는데, 그것은 우리에게 우월감을 심어준다는 것을 안다. 이러한 감정이 착각일까? 의심할 필요도 없이 그러한 감정은 진짜다. 그러나 나는 거기에 너무 많은 중요성이 부여되어야 한다고는 생각하지 않는다. 게다가 우리가 작은 승리를 거두었던 기계와 관련된 경우에만 우월감을 느낄 수 있다. 동시에 모든 기계에 대해 우월감은 느낄 수 있는 질문은 없다. 간단히 말하자면, 인간이 어떤 기계보다 더 영리할 수 있을지 모르나, 그보다 더 영리한 다른 기계가 존재할 수도 있으며 이러한 순환은 계속 될 것이다.

내 생각에는, 수학 상의 주장을 지지하는 사람들 대부분이 토론의 토대가 되는 모방게임을 기꺼이 받아들일 것이다. 앞의 두 반대 견해를 지지하는 사람들은 아마 어떠한 기준에도 관심이 없을 것이다.

(4) '의식'을 근거로 하는 반대주장

이 주장은 제퍼슨(Jefferson) 교수가 1949년에 행한 목록 기계 연설(Lister Oration)에 아주 잘 표현되어 있는데 그것을 인용하자면 다음과 같다. "기계가 사고와 감정을 느껴 협주곡을 작곡하거나 소네트를 쓴다면, 우리는 기계가 두뇌를 가지고 있다— 즉, 쓰고 쓴 것을 이해할 수 있다 —는 것에 동의할 수 있다. 그러나 어떤 기계도 성공했을 때의 즐거움이나, 자신의 멜브가 녹았을 때의 슬픔을 느낄 수 없으며, 아첨을 듣고 우쭐거리지도, 실수 때문에 비참해질 수도, 이성에 유혹될 수도, 원하는 것을 얻을 수 없을 때 화를 내거나 실망할 수도 없다."

이 주장은 우리 실험(test)의 타당성을 부인하는 것처럼 보인다. <기계가 생각한다는 것을 확신할 수 있는 유일한 길은 기계가 되어 그 자신이 사고한다는 것을 느끼는 것이다>라는 견해는 위의 주장의 가장 극단적인 형태이다. 만일 그럴 수 있다면, 사람들은 그 느낌을 세상에다 알릴 수 있으나, 물론, 어떤 사람도 세상의 이목을 끌지 못할 것이다. 마찬가지로 이 견해에 따르면, 어떤 사람이 생각한다는 것을 아는 유일한 방법은 바로 그 사람이 되는 것이다. 사실 이런 생각은 자기만이 유일하다고 생각하는 유아론(唯我論)적 관점이다. 그것은 지지될 수 있는 가장 논리적인 견해일지 모르나, 생각들 간의 상호 전달을 어렵게 한다. A는 '나는 생각하지만 B는 그렇지 않다'라고 믿는 반면, B는 '나는 생각하나 A는 그렇지 않다'고 믿는다. 이러한 견해에 대해 지속적으로 논쟁하는 것보다는 모든 사람이 생각한다고 겸손하게 믿는 것이 일반적이다.

나는 제퍼슨(Jefferson)교수가 그러한 극단적이고 유아론적 관점을 받아들이기를 바라지 않는다고 확신한다. 아마도 그는 기꺼이 모방게임을 실험(test)으로서 받아들일 것이다. 그 게임(역할자 B는 생략된 채로)은 실제로 '구두시험'이라는 이름 아래, 어떤 사람이 어떤 것을 진짜로 이해했는지 또는 단순히 암기했는지를 알아보기 위해 자주 사용된다. 그 구두시험의 한 부분을 살펴보도록 하자.

질문자 : '내가 당신을 여름날에 비유해도 될까요?'라고 되어 있는 당신 소네트 첫째 줄에서 '봄날' 이라는 말이 더 낫지 않을까요?
대상자 : 그것은 운율에 맞지 않을 것입니다.
질문자 : '겨울날'은 어떠합니까? 그것은 운율에 잘 맞는군요.
대상자 : 예, 그러나 어느 누구도 겨울날에 비유되기를 원치 않습니다.
질문자 : 피퀵씨(Mr.Pickwick)를 보면 크리스마스가 생각납니까?
대상자 : 어느 정도.
질문자 : 크리스마스도 겨울인데, 나는 피퀵(Mr.Pickwick)가 그러한 비유를 싫어하리라고는

생각하지 않는데요.
대상자 : 나는 당신이 진지하게 이야기하고 있다고 생각하지 않습니다. 사람들이 생각하는 겨울이란 전형적인 겨울이지 크리스마스와 같은 특정한 날이 아닙니다.

기타 등등. 만약 소네트를 쓰는 기계가 구두시험에서 이처럼 대답할 수 있다면 제퍼슨(Jefferson) 교수는 뭐라고 했을까? 나는 그가 기계를 단순히 인위적으로(생각을 한 결과로서가 아니라 단지 외웠던 대답을 밖으로 앵무새처럼 내뱉는 것) 이러한 대답을 신호로 보내는 것으로 간주했는지 어떤지 알지 못한다. 그러나 만약 그 대답들이 위의 대화에서처럼 만족스럽게 진술되었다면, 그가 기계를 '단순한 고안품'이라고 묘사했으리라고는 생각되지 않는다. 내 생각으로는 이 말(단순한 고안품)이 가끔 스위치를 켜면 어떤 사람이 읽은 소네트를 재생해 내는 기계까지 포함하는 장치들을 포괄한다.

간단히 말해, 의식(지각)에 의한 주장을 지지하는 사람들 대부분은 유아론적 생각을 버리도록 설득될 수 있다고 생각된다. 그렇게 한다면 그들은 아마 우리의 실험(test)을 기꺼이 받아들일 것이다.

나는, 내가 의식에는 어떤 신비함도 없다고 생각한다는 인상을 주고 싶지는 않다. 예를 들면 그것은 신비로움이 없음을 구체화해 나가는 노력과 관계되는 어떤 모순[13](자가당착)이 있다. 그러나 이 논문에서 우리가 관심을 가지는 문제에 대답을 하기 전에 이들 신비가 반드시 풀려야 한다고는 생각하지 않는다.

(5) 여러 가지 무능을 근거로 하는 반박들

이 주장들은 "나는 당신이 앞서 언급했던 것들 모두를 할 수 있는 기계를 만들 수 있다는 것은 인정하지만, 당신은 결코 X를 할 수 있는 기계를 만들 수는 없다"의 형태를 취한다. 수많은 특정 X는 이러한 관련 속에 암시되어 있다. X에 들어갈 것은 다음에서 고르면 된다.

친절함, 수완이 비상함, 아름다움, 다정함(p.448), 창의력을 가짐, 유머를 가짐, 참과 거짓을 구별함, 실수를 함, 사랑함, 딸기와 크림을 즐김(p.448), 어떤 사람이 자신(=기계)을 좋아하도록 만듦, 경험을 통해 배움(p.456), 단어를 적절하게 사용함, 자기 자신의 생각에 주체가 됨(p.449), 인간처럼 다양한 행동을 함, 진실로 새로운 것을 함.(이 무능들 중에서 어떤 것에는 특별히 페이지 번호가 붙어 있다.)

사람들은 대개 이런 주장들을 지지하지 않는다. 나는 이들 주장 대부분이 과학적 귀납의 원리에 기초를 두고 있다고 믿는다. 한 사람은 일생동안 수천 종류의 기계를 본다. 그가 보았던 것으로부터 그는 많은 일반적인 결론을 이끌어낸다. "기계들은 아름답지 못해, 기계 각각은 매우 제한된 목적을 위해 고안되어 있어, 조금이라도 다른 목적을 위해 사용하려할 때는 그것들은 무용지물이 되어 버려, 그 기계들이 할 수 있는 행동은 매우 제한되어 있어." 등등. 자연히 그는 이 결론들이 일반적인 기계의 불가피한 특성이라고 생각한다. 그러나 이들 한계 중 많은 것이 대부분 기계의 매우 작은 저장용량 때문에 생긴다.(나는 저장용량에 대한 생각이 어떤 면에서 불연속 상태 기계 외의 다른 기계들을 포함하기 위해 확장되었다고 가정한다. 현재의 토론에서 수학적 정확성이 요구되지 않는 것처럼 정확한 정의는 중요하지 않다.) 몇 해 전 디지털 컴퓨터가 거의 알려져 있지 않았을 때, 만약 그 기계의 구조에 대해서는 이야기하지 않고 그것의 특징들만 말했다면, 디지털 컴퓨터에 대해 굉장히 많은 회의감(디지털 컴퓨터의 존재여부에 대한 회의감)이 있었을 것이다. 그것은 아마 과학적 귀납원리를 비슷하게 적용(응용)했기 때문이라 생각된다. 그 원리를 이처럼 적용하는 것은 물론 아주 무의식적이다. 불에 덴 적이 있는 아이는 불을 두려워하며 그래서 불을 피한다. 그것을 보고 우리는 그 아이가 과학적 귀납원리를 적용하고 있다고 말할 수 있다(두말할 나위 없이 그의 행동을 다른 많은 방법으로 기술할 수 있다). 인류의

13) 서로 다른 차원을 하나의 차원으로 간주할 때 또는 자기 자신을 자기 자신의 원소로 가질 때 이를 모순이라 한다. 예를 들면, '의식이란 무엇일까?'라는 질문에 대하여, 다시 무한 퇴행의 질문을 던지면, "의식이란 무엇일까를 의식함을 무엇일까?"의 형식이 된다. 이런 퇴행이 무한히 이어지게 된다. 이 모순 형식은 'A∈A'의 형식이 되기 때문에 모순이 된다. 'A⊆A'일 뿐이다. 'A∉A'.

행위와 관습(습관)은 과학적 귀납원리를 적용하기에 적절할 만큼 구체적인 것 같지는 않다. 믿을 만한 결론이 나오려면 아주 많은 부분의 시간과 공간이 조사되어야 한다. 반면에 우리는(영어를 사용하는 대부분의 아이들이 그러는 것처럼) 모든 사람이 영어를 사용하며, 프랑스 말을 배우는 것은 어리석다고 결론짓기도 한다(즉 단편적인 현상 또는 지식을 통해 성급한 결론을 내리기도 한다는 말임)

어쨌든 앞서 이야기된, 기계가 못하는 많은 것들에 대해 특별히 할 말이 있다. 딸기와 크림을 즐기지 못하다는 것은 독자에게 하찮게 생각되었을 것이다. 이러한 맛있는 음식을 즐기는 기계를 만들 수도 있겠지만, 기계에게 그러한 음식을 먹도록 하는 것은 바보 같은 짓이다. 이러한 기계의 무능에 대해 중요한 점은 그러한 무능이, 백인과 백인 사이 또는 흑인과 흑인 사이에서 일어나는 친밀성이 인간과 기계 사이에서도 일어나는 것을 어렵게 만든다는 것이다.

"기계는 실수하지 않는다."라는 주장은 호기심을 끈다. 어떤 사람들은 "실수한다고 해서 어디가 덧나나?"라고 곧잘 반박하기도 한다. 그러나 좀더 동정적인 태도를 취하여, 그것이 진실로 무엇을 의미하는지 알아보기로 하자. 나는 이러한 비판이 모방게임에 의해 설명될 수 있다고 생각한다. 질문자는 기계와 인간에게 많은 수학문제를 풀게 함으로써 간단히 그 둘을 구별할 수 있다. 기계는 무척 정확하므로 쉽게 정체가 드러난다. 이에 대한 대답은 간단하다. 그 기계(게임을 수행하도록 프로그램 되어 있다)는 그 수학문제들에 대한 정답을 제시하려 하지 않을 것이다. 질문자를 혼란시키기 위해 고의로 실수를 할 것이다. 기계적 결함은 계산할 때 어떤 실수를 해야 하는가에 대해 부적절한 결론을 내림으로써 드러난다. 그 비판에 대한 이러한 해석조차도 충분할 만큼 동정적이지는 않다. 그러나 우리는 그 비판에 대해 더 이상의 많은 지면을 할애할 여유가 없다. 그 비판은 두 종류의 실수를 혼동하고 있기 때문에 생긴다고 생각된다. 우리는 그 실수들을 '기능상의 실수'와 '결론상의 실수'라 부를 수 있다. 기능상의 실수는, 기계로 하여금 원래 부여된 행동을 하지 않고 다른 엉뚱한 일을 하도록 하는 기계적인 잘못이나 전기적인 잘못 때문에 생긴다. 철학적 논의에서는 그 같은 종류의 실수는 무시된다. 그러므로 '추상적 기계'를 논의하고 있는 것이다. 이러한 추상적 기계는 물리적인 실체가 아니라 수학적 허구(가공물)이다. 그러한 정의에 따른다면 당연히 그 기계는 기능상의 실수는 저지를 수 없다. 이러한 의미에서 '기계는 실수하지 않는다'고 말할 수 있다. 결론상의 실수는 단지 기계에서 나온 출력 신호에 어떤 의미가 덧붙여져 있을 때만 일어날 수 있다. 예를 들어, 기계는 결론을 수학 방정식이나 영어문장으로 출력할 수도 있다. 잘못된 진술이 적혀 나왔을 때 우리는 그 기계가 결론상의 실수를 범했다고 말한다. 기계가 이런 종류의 실수를 하지 않을 것이라고 단정할 수는 없다. 기계가 '0=1'을 단순히 반복적으로 출력해 낼 수도 있다. 좀더 공감하기 쉬운 예로서, 과학적 귀납원리에 의해 결론을 도출해 내는 경우를 들 수 있다. 우리는 그 같은 방법이 때때로 잘못된 결론을 도출할 수도 있다는 것을 감안해야 한다.

기계는 자기 자신의 생각에 주체가 될 수 없다는 주장은, 물론 그 기계가 어떤 주제에 대한 생각을 가지고 있다는 것이 전제될 때만 대답될 수 있다. 적어도 기계를 다루는 사람에게는 '기계의 작동에 대한 주된 문제'가 의미를 가지는 것처럼 보인다. 예를 들어, 기계가

$$x^2 - 40x - 11 = 0$$

이라는 방정식을 풀려고 할 때, 사람들은 이 방정식을 그때의 기계가 지닌 주된 문제의 일부라고 표현하고 싶어 한다. 이러한 의미에서 기계는 의심할 여지없이 그 자신의 주된 문제를 가질 수 있다. 그것은 자신의 프로그램을 만들 때나 또는 그 자신의 구조를 변경한 후의 결과를 예측하려 할 때 도움이 된다. 자신의 행동 결과를 관찰함으로써 기계는 어떤 목적을 효과적으로 달성하기 위해 자기 자신의 프로그램을 변경할 수 있다. 이것들은 꿈같은 이야기가 아니라 가까운 미래에 가능하게 될 것이다.

'기계는 다양한 행동을 할 수 없다'는 비판은 기계가 저장용량을 많이 가질 수 없다는 의미다. 아주 최근까지도 저장용량이 천 디지트(digit)가 되는 것조차 매우 드물었다.

여기서 다루고 있는 비판들은 종종 의식에 근거한 주장이라는 위장된 형태를 띤다. 대개, 기계가 할 수 없다고 생각되던 것들 중 하나를 그 기계가 할 수 있다고 주장하고 그것을 행하는 방법을 기술한다

할지라도, 사람들은 별 큰 감흥을 받지 못할 것이다. 그 방법(그것이 무엇이든 간에, 왜냐하면 그것은 반드시 기계적이기 때문이다)은 다소 근본적이라 생각된다. 앞의 (4)에 인용된 제퍼슨(Jefferson)의 주장에서의 괄호 안의 것과 비교하라.

(6) 러블러스(Lady Lovelace) 여사의 반대

우리는 러블러스 여사가 쓴 논문에서 배비지의 분석기관에 대한 정보를 가장 자세히 알 수 있다. 그 논문에서 그녀는 "분석기관은 어떠한 것을 창조할 수는 없다. 우리가 기계로 하여금 어떤 것을 수행하도록 시키는 방법을 아는 한, 그 기계는 무엇이든 할 수 있다."(즉 기계는 의지를 지니지 못하며 우리가 명령을 내리지 않으면 아무것도 할 수 없는 껍데기이다.) 이 주장은 하아티(Hartree p.70)가 인용했으며 거기에 덧붙여 말하길 "이 주장은, 혼자서 생각할 수 있거나 또는 그 속에 '학습'의 기초가 되는 생물학적 용어인 조건 반사가 일어나도록 설계가 된 전기적인 장치를 만들 수 없을 것이라는 것을 의미하지는 않는다. 이것이 원리상 가능한지 아닌지는 이와 같은 최근의 어떤 발달에 의해 제시되는 흥미로운 질문이다. 그러나 그 당시 이러한 특징을 가지도록 제작되거나 설계된 기계는 있었던 것 같지 않다."

나는 전적으로 하아티의 의견에 동의한다. 그는 문제의 그 기계가 그러한 특성을 가지지 않았다고 주장하지 않았다는 점에 주목해야 한다. 오히려 러블러스 여사에게 유용한 증거조차 그 기계들이 그것을 가졌다는 것을 믿게끔 그녀에게 의욕을 북돋워 주지 못했다. 문제의 그 기계는 어떤 의미에서는 이러한 특성을 가질 수도 있다. 어떤 불연속 상태 기계가 그런 특성을 가지고 있다고 가정해 보자. 분석기관은 보편적인 디지털 컴퓨터이므로, 만약 그 저장용량과 속도가 적당하다면 적절하게 프로그래밍됨으로써 문제의 그 기계를 흉내낼 수 있을 것이다. 아마 이러한 주장은 Countess(=Lovelace 여사)나 배비쥐(Babbage)에게는 떠오르지 않았을 것이다. 어쨌든, 주장될 수 있는 모든 것을 그들이 주장할 의무는 없다.

이 모든 문제는 학습기계의 첫머리에서 다시 고려될 것이다. 러블러스 여사(Lady Lovelace)의 반론이 모습을 바꾼 것이 "기계는 진실로 새로운 어떤 것을 결코 할 수 없다."라는 주장이다. 이것은 '하늘 아래 새로운 것은 없다'라는 속담을 방패막이로 삼고 있다. 이 반론이 더 잘 변형된 것은 기계는 '결코 우리를 놀라게 하지 않는다'는 것이다. 이러한 주장은 좀더 직접적인 도전이며 곧바로 부닥칠 수 있다. 나는 기계 때문에 굉장히 자주 놀란다. 이것은 내가 충분히 계산을 하지 못한 채 그들이 무엇을 할 것인지를 예상했기 때문이기도 하지만, 오히려 계산은 했더라도 그것을 위험을 무릅쓰면서까지 너무 서둘러 했기 때문이다. 아마 나는 "여기서의 전압은 저기와 분명 똑같다고 생각한다." 어쨌든 그렇다고 가정하자. 당연히 나는 종종 틀리며 그 실험이 행해질 때쯤에는 이러한 가정들이 잊혀져 버리기 때문에 그 결과는 나에게 놀라운 것이다. 이를 인정하는 일이 나의 잘못된 방식들을 주제로 삼는 강의에 귀를 기울이게 하지만, 내가 경험하는 놀라움을 검증할 때 내 자신의 신뢰성에 아무런 의심점도 부가하지 않는다.

나는 이 대답이 나의 비판을 침묵시키리라고는 생각지 않는다. 그는 아마 그 같은 놀라움은 내 자신 속에서 일어난 어떤 창조적인 정신행동에 기인하며, 그 기계에 대한 신뢰를 반영하지는 않는다고 말할 것이다. 이것은 우리를 의식으로부터의 논쟁으로 되돌아가게 하며, 결코 놀라운 생각이 아니다. 그것은 이미 판가름이 난 것으로 생각해야 하는 논쟁의 연장선에 불과하다. 그러나 아마도, 놀라운 사건이 인간, 책, 기계 또는 기타 다른 것들 중 어디에서 기인하든지 간에, 놀라운 어떤 것의 진가를 알려면 많은 '창조적인 정신 활동'이 요구된다는 것은 말할 가치가 있다.

기계는 놀라움을 야기할 수 없다는 견해는, 내가 믿건대, 철학자나 수학자가 특히 빠지기 쉬운 오류에서 기인한다. 어떤 한 사실이 정신에 주어지자마자 그 사실의 모든 결과가 동시에 마음속에서 생긴다는 가정이 바로 이것이다. 많은 경우 그것은 매우 유용한 가정이지만, 사람들은 그것이 틀렸다는 것을 너무 쉽게 잊는다. 그렇게 함으로써 일어난 자연스런 결과는 사람들이 자료와 일반적인 원리로부터

결과를 도출해내는 단순한 작업에는 아무런 장점이 없다고 생각한다는 것이다.

(7) 신경계의 연속성에 바탕을 둔 반박

확실히 신경계는 불연속 상태 기계가 아니다. 뉴런에 부딪치는 신경자극의 세기에 관한 정보에서의 아주 작은 실수가 밖으로 나가는 충격의 세기를 매우 크게 차이 나게 할 수 있다. 이러하므로 사람들은, 불연속 상태 기계를 이용하여 신경계의 행동을 모방할 수 있다고는 생각할 수 없다고 주장한다.

분명 불연속 상태 기계와 연속 상태 기계는 다르다. 그러나 우리가 모방게임의 조건들을 고수한다면 질문자는 이러한 차이점을 전혀 이용할 수 없을 것이다. 만일 우리가 몇몇 다른 간단한 연속기계를 생각해 본다면 이런 상황은 더욱더 분명해질 것이다. 미분 분석기라면 아주 좋다(미분 분석기는 어떤 종류의 계산을 위해 사용되는 불연속 상태 형의 기계가 아니다). 이것들 중 어떤 것은 답을 타이프로 친 형태로 제시하는데, 그래서 게임에 참여하는 데 적당하다. 디지털 컴퓨터가 미분 분석기가 내놓은 답을 정확하게 예측하는 것은 아마 불가능할 것이다. 허나, 정답류의 답을 제시하는 것은 분명 가능할 것이다. 예를 들어, 만약 π의 값을 내라고 한다면(π는 실제로 약 3.1416이다) 디지털 컴퓨터는 0.05, 0.15, 0.55, 0.19, 0.06의 확률을 가지고 3.12, 3.13, 3.14, 3.15, 3.16 중에서 임의로 한 값을 택하리라는 것은 당연하다. 이러한 상황에서 질문자가 디지털 컴퓨터와 미분 분석기를 구별한다는 것은 여간 어려운 일이 아니다.

(8) 행위의 비정형성에 바탕을 둔 반론(주장)

인간이 처할 수 있는 모든 상황에서 일어나는 인간 행동을 기술하는 일련의 규칙들을 만드는 것은 가능하지 않다. 예를 들어, 사람들은 빨간 불이면 서고, 파란 불이면 건너라는 규칙을 가지고 있다고 하자. 그런데 어떤 잘못으로 인해 두 불이 동시에 켜지면 어떻게 될까? 아마도 멈추는 것이 가장 안전하다고 생각할 것이다. 그러나 이러한 결정을 하고 나서는 더 큰 어려움이 생긴다. 모든 우발적인 사고를 포괄하는 행동 규칙을 만들려는 시도는 — 심지어 신호등에서 일어나는 것도 포함하여 — 불가능한 것처럼 보인다. 나는 이에 전적으로 동의한다.

이것으로부터 우리는 기계가 될 수 없다는 주장이 나온다. 나는 그 주장을 다시 하려 하지만 완벽하게 타당성을 가질 수 있게 만들지 못할까 두렵다. 그 주장은 다음과 같다. 만약 개개의 사람들이 자기 삶을 조절하는 명확한 일련의 행동규칙을 가진다면 그는 기계와 다를 바 없다. 그러나 그 같은 규칙들이 없으므로 인간은 기계가 될 수 없다. 그러나 그럼에도 불구하고 이것은 이미 사용된 적이 있는 주장이라고 나는 믿는다. 이 문제를 흐리게 하는 혼동이 있는 듯한데, '행동 규칙'(rules of conduct)과 '행위법칙'(laws of behaviour)을 혼동하는 것 같다. 내가 받아들이는 바로는 '행동 규칙'은 '만약 빨간 불을 보면 멈춰라'와 같이, 사람이 그것(행동규칙)에 따라 행동할 수 있고, 또 그것(행동규칙)을 의식할 수 있다. '행위 법칙'이란 '네가 그를 꼬집으면 그는 우는 소리를 낼 것이다'와 같이 인간의 신체에 적용되는 자연 법칙을 말한다. 만약 우리가 인용된 주장 속에서 '그의 삶을 조절하는 행동 법칙'을 행위법칙으로 대체한다면, 어벙벙한 중간 상태는 배제될 것이다. 왜냐하면 우리가 행위법칙에 의해 조절된다는 것은, 기계의 한 종류가 된다는 것을 의미할 뿐 아니라(비록 반드시 불연속 상태 기계가 되는 것은 아니지만), 반대로 그 같은 기계가 된다는 것은 그 같은 법칙에 의해 조절된다는 것을 의미한다고 믿기 때문이다. 그러나 우리는 완벽한 행동 규칙이 없는 것처럼, 완전한 행위 법칙도 없다는 것을 쉽게 확신할 수 없다.

과학적 관찰만이 그 같은 법칙을 찾는 유일한 길이며, 우리는 "충분히 찾아보았다. 그 같은 법칙은 없다."라고 말할 수 있는 상황이 없음을, 확실히 알고 있다. 우리는 그 같은 어떤 주장도 정당화될 수 없다는 것을 더욱 강력히 주장할 수 있다. 그 같은 법칙들이 존재할 경우 반드시 그것을 찾아 낼 수 있다고 가정해 보자. 그런 다음 불연속 상태 기계가 주어진다면 분명히 장차 일어날 행위를 충분한 관

찰을 통해서 예측할 수 있을 것이다. 이것은 천년 안에 실현될 것이다. 그러나 이것은 그 경우가 아닌 듯하다. 나는 맨체스터 컴퓨터에 단지 천단위의 기억 용량을 사용하는 작은 프로그램을 집어넣었는데, 하나의 16단위 숫자로 채워진 그 기계는 2초 이내에 다른 것으로 대답을 하였다. 나는 어떤 사람이 이런 대답으로부터, 시도해 보지 않은 값을 예측할 수 있는 프로그램에 관해 배우는 것을 반대하고 싶다.

(9) 초능력에 바탕을 둔 논쟁

나는 독자들이 ESP(감각외적 지각 또는 초능력) 즉, 텔레파시·투시력·예언력·염력에 대한 생각에 친숙하리라 생각한다. 이들 당혹스러운 현상은 우리의 보편적이고 과학적인 생각들을 부인하는 것처럼 보인다. 어찌나 우리는 그들을 불신하고 싶어 하는지! 불행하게도 적어도 텔레파시에 관한 한 그 통계적 증거는 굉장히 많다. 우리의 생각을 이러한 새로운 사실들에 맞추기 위해 재정립하는 것은 매우 어렵다. 사람들이 그것들을 인정한다 할지라도 그것이 유령이나 도깨비를 믿는 획기적인 비약으로는 보이지 않는다(즉 이러한 믿음을 가질지라도 아무렇지도 않다). 우리의 신체가 잘 알려진 물리 법칙과 아직 발견되지는 않았지만 다소 비슷한 법칙에 따라 움직인다는 생각은 우리가 넘어야 할 첫 단계 중의 하나이다. 내 생각으로 이 반론(주장)은 매우 강력하다. 사람들은 많은 과학적 이론들이 초능력 지각(ESP)과 충돌함에도 불구하고 여전히 실제로 유효하다고 말할 수 있다. 사실 그러한 사실을 몰라도 잘 살아 나갈 수 있다. 이것은 조금도 달갑지 않은 위안이며, 사람들은 사고가 단지, 초능력 지각(ESP)과 특히 관련 있는 현상의 일종이라는 것을 두려워한다.

초능력 지각(ESP)에 바탕을 둔 더 구체적인 반박 주장은 다음과 같다. "텔레파시를 잘 받는 사람과 디지털 컴퓨터를 대상으로 해서 모방게임을 해 보자. 질문자는 '내 오른손에 있는 카드는 무엇인가?'와 같은 질문을 할 수 있다. 텔레파시나 투시력으로 사람은 400번 중 130번 정답을 맞춘다. 기계는 임의로 추측할 수밖에 없어서 104번 정도 맞힌다. 그래서 질문자는 옳은 판단을 내릴 수 있다." 여기서 흥미로운 가능성이 펼쳐진다. 디지털 컴퓨터가 무작위 수를 추출하는 기계(난수 생성기)를 포함하고 있다고 하자. 그러면 대답을 하기 위해 자연히 이것을 사용할 것이다. 그러나 난수 생성기는 질문자의 염력에 영향을 받기 쉽다. 어쩌면 이 염력이 기계가 확률 상으로 기대했던 것보다 더 자주 옳은 답을 하도록 만들지도 모른다. 그래서 질문자는 옳은 판단을 할 수 없을지도 모른다. 한편으로 그는 투시력으로 질문하지 않고 추측으로 옳은 판단을 내릴 수 있다. 초능력으로는 무엇이든 가능할 수 있다. 텔레파시가 인정된다면 우리의 시험을 더 강화시킬 필요가 있다. 엄격하게 하기 전의 상황은 질문자가 중얼거리는 것을 경쟁자들 중의 하나가 벽에 귀를 대고 듣는 경우 일어날 수 있는 상황과 유사하다. 그래서 경쟁자들을 텔레파시가 통할 수 없는 방에다 집어넣으면 모든 요건이 충족될 것이다.

7. 학습 가능한 기계(Learning Machine, 스스로 배워 나가는 기계)

독자는 혹 내가 주장을 뒷받침할 명확하고 확실한 근거를 가지고 있지 않다고 성급하게 단정 내릴지도 모른다. 만약 내가 그러한 근거를 가지고 있다면, 반대 주장의 오류를 지적하는 데 어려움을 겪지는 않았을 것이다. 내가 가진 증거들을 이제 제시해 보겠다.

기계는 우리가 시킨 것만을 할 수 있다는, 러블러스 여사(Lady Lovelace)의 반론을 잠시 살펴보자. 어떤 사람은, 기계에다 생각을 주입할 수 있고, 그렇게 한다면 해머로 두들긴 피아노 줄처럼 그 기계는 어느 정도까지 반응하고 나서 침묵으로 빠져든다고 말할 수 있다. 다른 비유로 임계 크기보다 작은 원자로를 들 수 있다. 주입된 생각(idea, 착상)은 외부로부터 원자로 안에 들어가는 중성자에 해당된다. 이들 각각의 중성자는 소동을 일으키다가 사라져 버릴 것이다. 그러나 만약 원자로 크기가 충분할 만큼 커진다면, 유입된 중성자가 일으킨 소동은 전체 원자로가 파괴될 때까지 계속 증가될 것이다. 인간 정신(human mind)에도 이에 해당되는 현상이 있을까? 기계에는 어떠할까? 인간정신에는 이에 해당하는 것이 있는 듯하다. 그것들의 대부분은 '임계 크기보다 작은 것' 즉, 위의 비유에서는 임계 크기보다 작은 원자로에 해당할 성싶다. 그 같은 정신 속에 주어진 사고는 반응할 때 평균적으로 하나 이상의

생각은 일으키지 못한다. 어떤 조그마한 부분들이 임계점을 넘었다고 가정하자. 그 같은 정신(임계점을 넘은 정신)에 표현된 사고는 2차, 3차, 그리고 더 먼 차원의 사고들을 연상시키면서 전체 이론을 발생시킨다. 동물의 정신은 아주 명백히 임계 크기보다 작은 듯하다. 이 비유를 염두에 두고 우리는 '기계는 임계 크기보다 크게 만들어질 수 있는가'라고 묻는다.

'양파 껍질' 비유도 또한 도움이 된다. 정신과 두뇌의 기능을 고려할 때 우리는 순수하게 기계적인 용어로 설명할 수 있는 어떤 동작을 발견한다. 우리가 언급하는 이것은 진정한 정신에 해당되지는 않는다. 그것은 진정한 정신을 발견하고자 할 때 반드시 벗겨 내어야 할 껍질의 종류이다. 그러나 그 껍질을 벗기고 나서도 벗겨내야 할 껍질들이 계속해서 더 남아 있음을 알게 된다. 이런 방식으로 '진정한' 정신에 도달할 수 있을까? 결국 아무 쓸모없는 껍질에 이르는 것이 아닐까? 후자의 경우에서 전체 정신은 기계적이다(어쨌든 그것은 불연속 상태 기계는 아니다. 우리는 이에 대해 논의했었다).

이들 마지막 두 문단은 설득력 있는 주장이 아니다. 그것들은 오히려, '신념을 형성하는 데 도움이 되는 상세한 설명'에 불과하다.

6장의 첫머리에서 제기된 견해를 20세기 말까지 실험을 계속하여야만 아주 만족스럽게 지지할 수 있을 것이다. 그러나 그 동안 우리는 뭐라 말할 수 있을까? 만약 이 실험이 성공한다면 어떤 단계가 취해져야 하는가?

내가 설명했듯이, 그 문제는 주로 프로그래밍(=프로그램을 짜는 일) 문제다. 물론 기술의 발전도 이루어져야 할 것이다. 그러나 그것만으로 요구조건이 충족될 것 같지는 않다. 두뇌의 저장용량을 측정한 수치는 이진수로 10^{10}에서 10^{15}까지 다양하다. 나는 수치가 그 같이 많으리라고는 생각하지 않으며, 사고라는 더 고차원적인 형태를 위해 아주 작은 부분만이 사용된다고 믿는다. 두뇌의 저장용량 대부분은 아마도 시각 인상을 저장하는 데 쓰인다.[14] 만약 모방게임을 만족스럽게 수행하는데 10^9 이상의 용량이 요구된다면 나는 틀림없이 놀랄 것이다(참조-대영백과사전 제 11판의 용량은 2×10^8이다). 10^7의 저장용량조차 현재의 기술로서도 매우 실용 가능성이 있다. 아마 기계의 동작 속도를 증가시킬 필요는 전혀 없을 것이다. 신경세포에 비유될 수 있는 현대 기계들 중 일부는 후자보다 약 천배 이상 빠르게 작동한다. 이것은 다양하게 발생하는 속도의 감소를 상쇄시킬 수 있는 '최소한도의 안정성'을 제공해야 한다. 그러면 어떻게 이들 기계가 게임을 수행하도록 프로그램 하느냐가 문제가 된다. 현재 나는 하루에 약 천 단위의 프로그램을 짠다. 그러므로 약 60명의 작업자들이 50년 동안 꾸준히 작업한다면 그 작업을 이루어낼 수 있다. 단, 큰 문제가 없는 한에 있어서다. 좀더 신속한 방법이 바람직한 것 같다. 어른의 정신을 모방하려는 과정에서 우리는 어른을 현재 상태로 만든 과정에 대해 많이 생각해야 한다. 우리는 세 가지 요소에 주목해야 한다.

14) 우리 인간이 지닌 여러 가지 감각과 인지 기관 중에 제일 먼저 탁월한 연구가 이뤄진 분야는 시지각이다. MIT의 데이비드 마아(Marr) 교수는 자신이 얼마 없어 타계할 것임을 알게 되자 혼신의 힘을 다 쏟아 가장 중요한 인지과학의 토대를 마련해 놓은 책을 썼다(안타깝게도 출간을 보지 못한 채 40에 타계함). Marr(1982)《시지각: 시지각 정보에 대한 인간 표상 및 처리 과정에 대한 연산적 탐구 *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*》(Freeman and Company) 우리말로 된 소개 논문은 정찬섭(1989) "시각정보 처리 계산 모형", 이정모 외 11인《인지 과학: 마음·언어·계산》(민음사)에서 읽을 수 있다. 여기서 마아는 세 가지 층위의 분석 수준을 가정한다. 연산(=계산) 층위·표상 층위·구현 층위이다(computational level·representational level·implementational level). 앞의 두 층위가 특히 인문·사회 과학을 전공하는 연구자에게 중요하다. 마지막 층위는 앞의 두 층위를 유기체에다 심어 놓을지, 아니면 화성에 사는 외계인한테 심어 놓을지, 기계에다 심어 놓을지를 다룬다. 언어를 연구하는 선도자들은 언어의 연산·표상 층위를 밝히는 것을 목표로 한다. 연산 층위는 우리가 의식하지 못하는 층위이지만, 표상 층위는 우리가 스스로 자각할 수 있는 층위이다. 연산 층위가 어떻게 구성되어야 하는지는 수학 기초론에서 터를 닦아 둔 연역 공리계를 이용하게 되는데, 괴델의 증명을 더 확장시켜 둘 이상의 공리계가 모순 없이 양립할 수 있다는 코우헌(Cohen)의 수학적 증명에 따라서, 제안된 공리계 사이에 상동성 증명 작업도 같이 진행된다. 현재까지 뛰어난 세계적 인문학자들은 마아의 노선을 따라 자신의 생각을 맞춰 가며 작업을 하고 있다고 해도 지나친 말이 아니다. 같은 대학에 있는 촘스끼(Chomsky) 교수가 특히 그러한데, 그는 언어학에서 찾아지는 연산·표상 체계는 궁극적으로 우리 인지 기관들의 정체와 기능을 드러내는 추춧돌이라고 명시적으로 언급한다. 촘스끼는 언어를 넘어서서 인지 기관을 다루는 방식을 고민하고 있다. 언어를 다루는 이들 가운데에는 세 부류가 있는 듯하다. 언어도 제대로 모르면서 언어만 제일이라고 주장질하는 이가 있다. 언어만을 알고 언어에 매달리는 이들이 있다. 언어도 알 뿐만 아니라 언어를 넘어서 있는 인간의 인지 기관에 대해 깊은 사색을 하는 이들도 있다. 응당 우리 학문은 출발은 작고 좁더라도, 도착점은 넓고 커야 할 것이다.

(a) 정신의 초기상태(태어났을 때)
(b) 받아온 교육
(c) 살아가면서 겪은 경험(교육은 아님)

어른의 정신을 모방하는 프로그램을 만드는 대신, 아이의 정신을 흉내내는 프로그램을 만드는 것이 어떨까? 만약 적절한 교육과정을 쉽사리 익히도록 만든다면 어른의 두뇌를 얻을 수 있을 것이다. 아마 아이들의 두뇌는 문방구에서 산 공책 같은 것이다. 다소 작은 기계지만, 빈 종이가 많은 공책 같은 것이다.(기계와 쓰기는 우리의 관점에서 본다면 거의 뜻이 같다.) 우리는 아이의 두뇌가 작은 기계이므로 쉽게 프로그래밍 되리라 생각한다. 기계에 대한 교육 활동의 양은, 추측하건대 아이의 것과 같다. 여기서 우리의 문제를 아이 프로그램과 교육과정으로 나누어야 한다. 이들은 매우 밀접한 관련이 있다. 첫 시도에서 좋은 아이기계를 만들어 낸 다는 것을 기대할 수 없다. 우리는 그 같은 기계를 가르치는 경험을 해야 하고, 그 기계가 얼마나 잘 배우는지를 알아야 한다. 그런 다음에야 다른 기계를 실험해 보고 기계가 더 좋아졌는지 더 나빠졌는지를 알 수 있다. 이러한 과정과 진화 사이에는 분명한 관련이 있다. 아래 것을 머릿속에 인식하라.

아이 기계의 구조 = 유전적 요소
아이 기계의 변화 = 돌연변이
실험자의 판단 = 자연 선택

그러나 사람들은 이러한 과정이 진화보다는 더 신속하게 진행되기를 바란다. 적자생존은 그 이점을 측정하기에는 느린 방법이다. 지능을 훈련함으로써 질문자는 그것의 속도를 빠르게 할 수 있다. 그가 임의의 돌연변이에 구애받지 않는다는 사실 또한 중요하다. 만약 어떤 약함(劣性)에 대한 원인을 추적할 수 있다면, 그는 아마도 그것을 개선할 수 있는 돌연변이를 생각해낼 수 있을 것이다.

정상적인 아이에게 하는 것과 똑같은 교수과정을 기계에게 적용하는 것은 불가능할 것이다. 예를 들어 기계는 다리가 없어서 밖에 나가 석탄 통에 석탄을 채우라는 명령을 수행할 수 없다. 아마 기계는 눈도 없을지 모른다. 그러나 아무리 이러한 역할들이 더 발달된 기술로 극복된다 할지라도, 기계를 학교에 수업 받으러 보낸다면 반드시 아이들의 극심한 놀림감이 될 것이다. 기계는 가르침을 받아야 한다. 그렇다고 해서 우리는 두 다리와, 눈 기타 등등에 대해 지나치게 신경 쓸 필요는 없다. 헬렌 켈러(Miss Hellen Keller)의 경우를 통해 볼 때, 어떤 다른 방법에 의해 교사와 학생 사이의 서로 간에 의사소통이 일어날 수 있다면 교육도 이루어질 수 있다는 사실을 알 수 있다. 우리는 흔히 '교육 과정' 하면 상과 벌을 떠올리게 된다. 몇몇 간단한 아이-기계는, 이런 원리에 따라 만들어지거나 프로그램될 수 있다. 기계는 벌을 유발하는 사건은 반복하지 않고, 반면에 상(賞)을 유발하는 사건을 계속해서 하도록 만들어져야 한다. 이러한 정의는 기계가 아무런 감정을 가지지 않는다는 것을 전제하지는 않는다.

나는 그 같은 아이 기계를 가지고 몇 가지 실험을 했다. 몇 가지 것들을 가르치는 데는 성공했지만 가르치는 방법이 그 실험에 너무 적합하지 않아서 성공적이라고 할 수는 없다. 상과 벌을 이용하는 것은 가르치는 과정의 한 부분에 불과하다. 심하게 말하자면, 교사와 학생 상호간에 상과 벌 이외의 다른 의사소통 방법이 없다면, 학생에게 전달된 정보의 양은 그에게 적용된, 상과 벌 회수의 총합을 넘어서지 못한다. 아이가 'Casabianca'(카사비앙카)를 암기할 때까지 만약 그 원본(text)이 '아니오'라는 대답이 나오면 한 방 얻어맞는 '스무고개'라는 방법을 통해서만 드러날 수 있다면, 아마 아이는 정말 굉장히 화가 날 것이다. 그러므로 감정이 배제된 의사소통 통로가 반드시 있어야 한다. 만약 이러한 의사소통 통로가 이용될 수 있다면, 상과 벌로써 기계로 하여금 어떤 언어, 예를 들자면 상징 언어로 표현된 명령을 따르도록 가르칠 수 있다. 이러한 명령들은 감정이 배제된 통로를 통해 전달된다. 이 언어를 사용하면, 요구되는 상과 벌의 횟수가 상당히 줄어들 것이다.

아이 기계가 어느 정도로 복잡해야 적당한가 하는 데 대해서는 의견이 분분하다. 어떤 사람들은 시종일관 일반원리를 가지고 가능한 한 단순하게 그 기계를 만들려고 할 것이고, 반면 다른 사람들은 완벽한 논리적 추론 체계를 '만들려'[15] 할 것이다. 후자의 경우 저장 용량은 대부분 정의와 명제가 차지할 것이다. 명제는 잘 정립된 사실들, 추측, 수학적으로 증명된 정리, 권위에 의한 진술, 논리적 명제의

형태를 띠고 있으나, 믿을 수 없는 표현 등등 여러 가지로 다양하다. 어떤 명제는 '명령'이라 부를 수 있을 것이다. 그 기계는 그 명령이 '잘 정립된' 것이라고 분류되자마자 적절한 행동을 자동적으로 하도록 만들어져야 한다. 이것을 설명하기 위해, 교사가 기계에게 '지금 숙제해라'라고 말한다고 가정하자. 이것은 "선생님이 <지금 숙제해라> 하고 말한다."를 잘 정립된 사실에 포함되게 한다. 그 같은 사실을 하나 더 들면 '선생님이 말한 것은 모두 옳다'이다. 이들을 결합하면 잘 정립된 사실(증명이 잘 된 사실)중의 하나로서 '지금 숙제를 해라'는 명령이 도출되는데, 그 기계를 구성함으로써 이는 그 숙제가 실제로 시작되고 그 효과도 만족스럽게 됨을 의미할 것이다. 그 기계에 의해 이용된 추론 과정은 가장 엄격한 논리학자를 만족시킬 필요는 없다. 예를 들어, 유형(type)에[16] 대한 계층이 전혀 없을 가능성도 있다. 이것이 계층 혼합이 반드시 일어날 것임을 의미하지는 않는다. 절벽에 울타리가 없다고 해서 꼭 떨어져야 하는 것은 아니다. '만약 그것이 선생님이 말했던 것의 하위 분류가 아니라면 분류하지 말아'와 같은 적절한 명령(그 체계 안에 포함되어 있으나, 그 체계에 대한 규칙의 일부를 형성하지는 않는)은 '가장자리에 너무 가까이 가지 말아'와 비슷한 효과를 낸다.

팔다리가 없는 기계가 따를 수 있는 명령은 위에서 제시된 예—즉, 숙제하는 것—처럼 분명 다소 지적인 성격을 띨 것이다. 그 같은 명령 중에서 중요한 것은 우리가 관심을 두고 있는 논리 체계에 대한 규칙이 적용될 수 있는 순서를 조절하는 명령들이다. 왜냐하면 우리가 논리체계를 사용하는 각 단계마다, 논리 체계에 대한 규칙의 준수에 관심을 두는 한, 적용될 수 있는 세부단계들이 매우 많기 때문이다. 이러한 선택들은 건전한 것과 오류인 것을 구별하는 것이 아니라, 현명한 것과 어리석은 추론자(reasoner)를 구분한다. 이러한 종류의 명령이 됨직한 명제(규칙 적용순서를 조절하는 명제)는 "소크라테스가 언급될 때는 Barbara(바바라)의[17] 삼단 논법을 사용해라." 또는 "만약 어떤 방법이 다른 것보다 더 빠르다는 것이 입증되거든 느린 방법은 사용하지 말아" 등일 것이다. 이들 명제 중 어떤 것들은 권위에 의한 명제일지도 모르나, 다른 것들은 기계 스스로 —예를 들어, 과학적 연역법으로 — 만든 것이다.

학습기계(learning machine)에 대한 생각은 어떤 독자들에게는 역설적인 것처럼 보일지 모른다. 어떻게 기계의 동작 규칙이 바뀔 수 있을까? 그 규칙들은, 기계가 어떤 계보의 것이든, 어떤 변화를 겪든지 간에, 그 기계가 어떻게 반응하는지를 완벽하게 기술해야 한다. 그래서 그 규칙들은 시간이 흘러도 변하지 않는다. 이것은 정말이다. 그 역설(즉, 규칙은 변하지 않는데 기계의 동작규칙은 바뀐다)에 대해 설명하자면, 학습과정에서 변하는 규칙들은 단지 순간적인 타당성만을 요구하는 다소 덜 잘난 체하는 규칙들이다. 독자들이여! 미국의 정치제도와 이를 비교해 보라.

학습 기계(learning machine)의 가장 중요한 특징은, 설사 그 기계를 가르치는 사람이 어느 정도 제

15) 원저자 각주임 : 아이 기계에 프로그램된 것은 디지털 컴퓨터에도 프로그램될 것이다. 그러나 논리체계를 가르쳐서는 안 될 것이다.

16) type이란 말은 수학에서 쓰는 set와 같은 말이다. type은 tokens들로 이루어지는데, tokens는 elements와 동일한 말이다. type-token이란 용어는 영국에서 러셀에 의해 쓰였고, set-element는 독일에서 칸토르에 의해 쓰였으며(원래는 set이 아니라 congregate/class였음), 폴란드 사람들은 meta-object라는 용어를 썼다. 우연히 폴란드 사람들이 세계 대전 동안에 미국 학계에 자리를 잡았기 때문에, 미국에 거의 전적으로 영향을 받는 우리도 폴란드 논리학 표기법과 용어가 많이 보급되었다. 집합에서의 모순은 한 집합이 자기 자신을 원소로 취할 때 일어난다. 집합은 다만 자신을 부분 집합의 형식으로 포함할 수 있을 뿐이다. 수학 기초론에서는 '유형 이론'을 다루는데, 관련 글들은 연대순으로 다음과 같다.

① 버틀런트 러셀(Russell 1908) "유형 이론에 바탕을 둔 수리 논리학"
② 앨론조 춰어취(Church 1940) "유형에 대한 단순한 이론의 형식화" 《기호 논리 논문집》 제 5권
③ 앨론조 춰어취(Church 1941) 《람다 변환에 대한 미적분학(고등수학)》 프린스턴 대학 출판부
④ 앨론조 춰어취(Church 1951) "의미 분석에서 추상화 개체에 대한 필요성" Daedalus 제 80권
⑤ 조어킴 람벡(Lambek 1958) "문장 구조의 수학" 《미국 수학 월간 학술지》 제 65권 3호
⑥ 폴 R. 해모스(Halmos 1977) "A에서부터 G까지의 논리학" 《수학 학술지》 제 50권 1호
⑦ 폴 R. 해모스(Halmos 1982) "추상화의 전율" 《2년제 대학 수학 학술지》 제 13권 4호

17) 삼단 논법에서

모든 X는 Y이다
모든 Z는 X이다
따라서 모든 Z는 Y이다

와 같이 '∀x-∀x-∀x'의 전칭 양화사를 갖는 형식을 가리킨다.

자(=기계)의 행동을 예측할 수 있을지라도, 내부적으로 진행되는 것을 거의 전적으로 모를 경우가 종종 있다는 것이다. 이것은 잘 디자인한(또는 프로그램한)아이기계에서부터 이후의 수준 높은 교육에 아주 강하게 적용할 수 있다. 이는 분명 계산하기 위해 기계를 사용할 때의 정상적인 절차와는 반대다. 사람들의 목표는 계산하는 매 순간마다의 기계의 상태를 머릿속 그림(Mental picture)으로 나타내는 것이다. 이 목표는 어려운 고생을 하고서야만 달성할 수 있다. '기계는 오직(only) 우리가 시킨 것만을 할 수 있다'[18]는 생각은, 앞의 생각과 마주 놓고 보면 이상한 것 같다. 우리가 기계에다 주입할 수 있는 대부분의 프로그램들은 이치에 맞지 않는 (이해할 수 없는) 어떤 것을 출력해 내거나, 생각지도 못한 제멋대로 행동하는 결과를 낳을 것이다. 지적행동은 아마도 계산에 관련되는 훈련된 행동과 전적으로 분리되며, 마구잡이행동이나 초점 없이 반복되는 루프를 발생시키지 않는다. 배우고 가르치는 과정을 통해 기계가 모방게임에서 역할을 제대로 수행하게끔 함으로써 얻어지는 또 다른 중요한 결론은, '인간적인 오류'는 특별한 '지도'없이도 다소 자연히 없어진다는 것이다. (독자들은 [원본] 24, 25 페이지의 견해와 이 생각을 조화시켜야 한다.) 배운 과정일지라도 결과를 100% 확실하게 도출하지는 못한다. 만약 기계가 100% 확실한 결론을 이끌어 냈다면 그 과정들은 잊혀지지 않을 것이다.

아마도 학습 기계에다 임의적인 요소를 포함시키는 것이 현명한 일일 것이다. (제 4장을 보라.) 임의적인 요소는 어떤 문제에 대한 해답을 찾을 때 다소 유용하다. 예를 들어 50에서 200사이의 숫자 중에서, 각 자리 수의 합의 제곱이 50에서 200사이인 수를 찾고자 한다고 가정하면(만약 '53'을 택한다면 5+3=8이므로 82는 64이다. 그래서 이 수는 우리가 찾는 수에 포함된다) 우리는 원하는 숫자를 찾을 때까지 51에서 시작해서 52 등등 계속해서 다른 수들을 살펴볼 것이다. 또 이와는 달리, 원하는 수를 찾을 때까지 임의의 숫자들을 추측해 볼 수 있다. 이 방법은 이미 살펴보았던 숫자를 계속 알고 있을 필요가 없다는 장점이 있지만, 똑같은 수를 두 번씩 살펴볼지도 모른다는 단점이 있다. 그러나 이것은 몇 가지 해결책이 있다면 별 중요하지 않다. 51부터 시작해서 체계적으로 살펴보는 방법(하나하나 살펴나가는 방법)은, 처음으로 조사될 부분에 해답이 없다면 엄청난 고생을 하게 될 수도 있다는 단점을 가지고 있다(예를 들어 답이 199라면 이 방법은 너무 소모적이며 비능률적이다). 가르치는 사람을 만족시킬 수 있는 행동 양식(또는 어떤 다른 기준)을 찾는 것을 학습 과정(leaning process)이라 할 수 있다. 해답은 아주 많이 있을 것이므로 임의로 추출하는 방법이 체계적인 방법보다 더 나은 것 같다. 그것이 진화과정에서도 유사하게 쓰인다는 것은 주목해야 한다. 반면 체계적인 방법은 불가능하다. 이미 행했던 서로 다른 유전적 결합을 다시 시도하는 것을 피하기 위해 어떻게 그것들(이미 시도된 유전적 결합들)을 다 기억하고 있을 수 있겠는가.

궁극적으로 우리는 기계가 순수하게 지적인 분야에서 사람과 경쟁하기를 바란다. 그러나 어디서부터 시작하면 가장 좋을까? 이것조차도 어려운 결정이다. 많은 사람들은 체스 두는 것과 같은 매우 추상적인 활동이 가장 좋을 것이라고 생각한다. 또한 기계에게 돈으로 살 수 있는, 가장 좋은 감각기관을 준 다음, 기계가 영어를 이해하고 말하도록 가르치는 것이 가장 좋다고 주장할 것이다. 이 방법은 사물들을 지적하고 이름 붙이는 등등의 정상적인 아이를 가르치는 과정 다음에 올 수 있다. 다시 말하건대, 나는 무엇이 옳은 답인지 알지 못한다. 그래서 이 두 방법 모두 시도되어야 한다.

우리는 단지 눈앞에 있는 짧은 거리만 볼 수 있다. 그러나 앞으로 할 일이 많이 있음을 알 수 있다.

참 고 문 헌

18) 원저가 각주임 : 제 6 장 (6)에 있는 러블러스 여사의 진술과 비교해 보시오. 거기에는 'only'라는 단어가 포함되지 않았다.

COMPUTING MACHINERY AND INTELLIGENCE by A. M. Turing

Mind : A Quarterly Review of Psychology and Philosophy, vol. # 59, 1950 October

1. The Imitation Game

I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous, If the meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game." It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart front the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either "X is A and Y is B" or "X is B and Y is A." The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair?

Now suppose X is actually A, then A must answer. It is A's object in the game to try and cause C to make the wrong identification. His answer might therefore be:

"My hair is shingled, and the longest strands are about nine inches long."

In order that tones of voice may not help the interrogator the answers should be written, or better still, typewritten. The ideal arrangement is to have a teleprinter communicating between the two rooms. Alternatively the question and answers can be repeated by an intermediary. The object of the game for the third player (B) is to help the interrogator. The best strategy for her is probably to give truthful answers. She can add such things as "I am the woman, don't listen to him!" to her answers, but it will avail nothing as the man can make similar remarks.

We now ask the question, "What will happen when a machine takes the part of A in this game?" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, "Can machines think?"

2. Critique of the New Problem

As well as asking, "What is the answer to this new form of the question," one may ask, "Is this new question a worthy one to investigate?" This latter question we investigate without further ado, thereby cutting short an infinite regress.

The new problem has the advantage of drawing a fairly sharp line between the physical and the intellectual capacities of a man. No engineer or chemist claims to be able to produce a material which is indistinguishable from the human skin. It is possible that at some time this might be done, but even supposing this invention available we should feel there was little point in trying to

make a "thinking machine" more human by dressing it up in such artificial flesh. The form in which we have set the problem reflects this fact in the condition which prevents the interrogator from seeing or touching the other competitors, or hearing -their voices. Some other advantages of the proposed criterion may be shown up by specimen questions and answers. Thus:

Q: Please write me a sonnet on the subject of the Forth Bridge.
A : Count me out on this one. I never could write poetry.
Q: Add 34957 to 70764.
A: (Pause about 30 seconds and then give as answer) 105621.
Q: Do you play chess?
A: Yes.
Q: I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?
A: (After a pause of 15 seconds) R-R8 mate.

The question and answer method seems to be suitable for introducing almost any one of the fields of human endeavour that we wish to include. We do not wish to penalise the machine for its inability to shine in beauty competitions, nor to penalise a man for losing in a race against an aeroplane. The conditions of our game make these disabilities irrelevant. The "witnesses" can brag, if they consider it advisable, as much as they please about their charms, strength or heroism, but the interrogator cannot demand practical demonstrations.

The game may perhaps be criticised on the ground that the odds are weighted too heavily against the machine. If the man were to try and pretend to be the machine he would clearly make a very poor showing. He would be given away at once by slowness and inaccuracy in arithmetic. May not machines carry out something which ought to be described as thinking but which is very different from what a man does? This objection is a very strong one, but at least we can say that if, nevertheless, a machine can be constructed to play the imitation game satisfactorily, we need not be troubled by this objection.

It might be urged that when playing the "imitation game" the best strategy for the machine may possibly be something other than imitation of the behaviour of a man. This may be, but I think it is unlikely that there is any great effect of this kind. In any case there is no intention to investigate here the theory of the game, and it will be assumed that the best strategy is to try to provide answers that would naturally be given by a man.

3. The Machines Concerned in the Game

The question which we put in 1 will not be quite definite until we have specified what we mean by the word "machine." It is natural that we should wish to permit every kind of engineering technique to be used in our machines. We also wish to allow the possibility than an engineer or team of engineers may construct a machine which works, but whose manner of operation cannot be satisfactorily described by its constructors because they have applied a method which is largely experimental. Finally, we wish to exclude from the machines men born in the usual manner. It is difficult to frame the definitions so as to satisfy these three conditions. One might for instance insist that the team of engineers should be all of one sex, but this would not really be satisfactory, for it is probably possible to rear a complete individual from a single cell of the skin (say) of a man. To do so would be a feat of biological technique deserving of the very highest praise, but we would not be inclined to regard it as a case of "constructing a thinking machine." This prompts us to abandon the requirement that every kind of technique should be permitted. We are the more ready to do so in view of the fact that the present interest in "thinking machines" has been aroused by a particular kind of machine, usually called an "electronic computer" or "digital

computer." Following this suggestion we only permit digital computers to take part in our game.

This restriction appears at first sight to be a very drastic one. I shall attempt to show that it is not so in reality. To do this necessitates a short account of the nature and properties of these computers.

It may also be said that this identification of machines with digital computers, like our criterion for "thinking," will only be unsatisfactory if (contrary to my belief), it turns out that digital computers are unable to give a good showing in the game.

There are already a number of digital computers in working order, and it may be asked, "Why not try the experiment straight away? It would be easy to satisfy the conditions of the game. A number of interrogators could be used, and statistics compiled to show how often the right identification was given." The short answer is that we are not asking whether all digital computers would do well in the game nor whether the computers at present available would do well, but whether there are imaginable computers which would do well. But this is only the short answer. We shall see this question in a different light later.

4. Digital Computers

The idea behind digital computers may be explained by saying that these machines are intended to carry out any operations which could be done by a human computer. The human computer is supposed to be following fixed rules; he has no authority to deviate from them in any detail. We may suppose that these rules are supplied in a book, which is altered whenever he is put on to a new job. He has also an unlimited supply of paper on which he does his calculations. He may also do his multiplications and additions on a "desk machine," but this is not important.

If we use the above explanation as a definition we shall be in danger of circularity of argument. We avoid this by giving an outline. of the means by which the desired effect is achieved. A digital computer can usually be regarded as consisting of three parts:

(i) Store.
(ii) Executive unit.
(iii) Control.

The store is a store of information, and corresponds to the human computer's paper, whether this is the paper on which he does his calculations or that on which his book of rules is printed. In so far as the human computer does calculations in his bead a part of the store will correspond to his memory.

The executive unit is the part which carries out the various individual operations involved in a calculation. What these individual operations are will vary from machine to machine. Usually fairly lengthy operations can be done such as "Multiply 3540675445 by 7076345687" but in some machines only very simple ones such as "Write down 0" are possible.

We have mentioned that the "book of rules" supplied to the computer is replaced in the machine by a part of the store. It is then called the "table of instructions." It is the duty of the control to see that these instructions are obeyed correctly and in the right order. The control is so constructed that this necessarily happens.

The information in the store is usually broken up into packets of moderately small size. In one machine, for instance, a packet might consist of ten decimal digits. Numbers are assigned to the

parts of the store in which the various packets of information are stored, in some systematic manner. A typical instruction might say-

"Add the number stored in position 6809 to that in 4302 and put the result back into the latter storage position."

Needless to say it would not occur in the machine expressed in English. It would more likely be coded in a form such as 6809430217. Here 17 says which of various possible operations is to be performed on the two numbers. In this case the)e operation is that described above, viz., "Add the number. . . ." It will be noticed that the instruction takes up 10 digits and so forms one packet of information, very conveniently. The control will normally take the instructions to be obeyed in the order of the positions in which they are stored, but occasionally an instruction such as

"Now obey the instruction stored in position 5606, and continue from there"

may be encountered, or again

"If position 4505 contains 0 obey next the instruction stored in 6707, otherwise continue straight on."

Instructions of these latter types are very important because they make it possible for a sequence of operations to be replaced over and over again until some condition is fulfilled, but in doing so to obey, not fresh instructions on each repetition, but the same ones over and over again. To take a domestic analogy. Suppose Mother wants Tommy to call at the cobbler's every morning on his way to school to see if her shoes are done, she can ask him afresh every morning. Alternatively she can stick up a notice once and for all in the hall which he will see when he leaves for school and which tells him to call for the shoes, and also to destroy the notice when he comes back if he has the shoes with him.

The reader must accept it as a fact that digital computers can be constructed, and indeed have been constructed, according to the principles we have described, and that they can in fact mimic the actions of a human computer very closely.

The book of rules which we have described our human computer as using is of course a convenient fiction. Actual human computers really remember what they have got to do. If one wants to make a machine mimic the behaviour of the human computer in some complex operation one has to ask him how it is done, and then translate the answer into the form of an instruction table. Constructing instruction tables is usually described as "programming." To "programme a machine to carry out the operation A" means to put the appropriate instruction table into the machine so that it will do A.

An interesting variant on the idea of a digital computer is a "digital computer with a random element." These have instructions involving the throwing of a die or some equivalent electronic process; one such instruction might for instance be, "Throw the die and put the-resulting number into store 1000." Sometimes such a machine is described as having free will (though I would not use this phrase myself), It is not normally possible to determine from observing a machine whether it has a random element, for a similar effect can be produced by such devices as making the choices depend on the digits of the decimal for .

Most actual digital computers have only a finite store. There is no theoretical difficulty in the idea of a computer with an unlimited store. Of course only a finite part can have been used at any one time. Likewise only a finite amount can have been constructed, but we can imagine more and more being added as required. Such computers have special theoretical interest and will be called

infinitive capacity computers.

The idea of a digital computer is an old one. Charles Babbage, Lucasian Professor of Mathematics at Cambridge from 1828 to 1839, planned such a machine, called the Analytical Engine, but it was never completed. Although Babbage had all the essential ideas, his machine was not at that time such a very attractive prospect. The speed which would have been available would be definitely faster than a human computer but something like I 00 times slower than the Manchester machine, itself one of the slower of the modern machines, The storage was to be purely mechanical, using wheels and cards.

The fact that Babbage's Analytical Engine was to be entirely mechanical will help us to rid ourselves of a superstition. Importance is often attached to the fact that modern digital computers are electrical, and that the nervous system also is electrical. Since Babbage's machine was not electrical, and since all digital computers are in a sense equivalent, we see that this use of electricity cannot be of theoretical importance. Of course electricity usually comes in where fast signalling is concerned, so that it is not surprising that we find it in both these connections. In the nervous system chemical phenomena are at least as important as electrical. In certain computers the storage system is mainly acoustic. The feature of using electricity is thus seen to be only a very superficial similarity. If we wish to find such similarities we should took rather for mathematical analogies of function.

5. Universality of Digital Computers

The digital computers considered in the last section may be classified amongst the "discrete-state machines." These are the machines which move by sudden jumps or clicks from one quite definite state to another. These states are sufficiently different for the possibility of confusion between them to be ignored. Strictly speaking there, are no such machines. Everything really moves continuously. But there are many kinds of machine which can profitably be thought of as being discrete-state machines. For instance in considering the switches for a lighting system it is a convenient fiction that each switch must be definitely on or definitely off. There must be intermediate positions, but for most purposes we can forget about them. As an example of a discrete-state machine we might consider a wheel which clicks round through 120 once a second, but may be stopped by a lever which can be operated from outside; in addition a lamp is to light in one of the positions of the wheel. This machine could be described abstractly as follows. The internal state of the machine (which is described by the position of the wheel) may be q1, q2 or q3. There is an input signal i0. or i1 (position of lever). The internal state at any moment is determined by the last state and input signal according to the table

		Last State		
		q1	q2	q3
Input	i0	q2	q3	q1
	i1	q1	q2	q3

The output signals, the only externally visible indication of the internal state (the light) are described by the table

State	q1	q2	q3
Output	o0	o0	o1

This example is typical of discrete-state machines. They can be described by such tables provided they have only a finite number of possible states.

It will seem that given the initial state of the machine and the input signals it is always possible to predict all future states, This is reminiscent of Laplace's view that from the complete state of the universe at one moment of time, as described by the positions and velocities of all particles, it should be possible to predict all future states. The prediction which we are considering is, however, rather nearer to practicability than that considered by Laplace. The system of the "universe as a whole" is such that quite small errors in the initial conditions can have an overwhelming effect at a later time. The displacement of a single electron by a billionth of a centimetre at one moment might make the difference between a man being killed by an avalanche a year later, or escaping. It is an essential property of the mechanical systems which we have called "discrete-state machines" that this phenomenon does not occur. Even when we consider the actual physical machines instead of the idealised machines, reasonably accurate knowledge of the state at one moment yields reasonably accurate knowledge any number of steps later.

As we have mentioned, digital computers fall within the class of discrete-state machines. But the number of states of which such a machine is capable is usually enormously large. For instance, the number for the machine now working at Manchester is about 2 165,000, i.e., about 10 50,000. Compare this with our example of the clicking wheel described above, which had three states. It is not difficult to see why the number of states should be so immense. The computer includes a store corresponding to the paper used by a human computer. It must be possible to write into the store any one of the combinations of symbols which might have been written on the paper. For simplicity suppose that only digits from 0 to 9 are used as symbols. Variations in handwriting are ignored. Suppose the computer is allowed 100 sheets of paper each containing 50 lines each with room for 30 digits. Then the number of states is 10 100x50x30 i.e., 10 150,000 . This is about the number of states of three Manchester machines put together. The logarithm to the base two of the number of states is usually called the "storage capacity" of the machine. Thus the Manchester machine has a storage capacity of about 165,000 and the wheel machine of our example about 1.6. If two machines are put together their capacities must be added to obtain the capacity of the resultant machine. This leads to the possibility of statements such as "The Manchester machine contains 64 magnetic tracks each with a capacity of 2560, eight electronic tubes with a capacity of 1280. Miscellaneous storage amounts to about 300 making a total of 174,380."

Given the table corresponding to a discrete-state machine it is possible to predict what it will do. There is no reason why this calculation should not be carried out by means of a digital computer. Provided it could be carried out sufficiently quickly the digital computer could mimic the behavior of any discrete-state machine. The imitation game could then be played with the machine in question (as B) and the mimicking digital computer (as A) and the interrogator would be unable to distinguish them. Of course the digital computer must have an adequate storage capacity as well as working sufficiently fast. Moreover, it must be programmed afresh for each new machine which it is desired to mimic.

This special property of digital computers, that they can mimic any discrete-state machine, is described by saying that they are universal machines. The existence of machines with this property has the important consequence that, considerations of speed apart, it is unnecessary to design various new machines to do various computing processes. They can all be done with one digital computer, suitably programmed for each case. It 'ill be seen that as a consequence of this all digital computers are in a sense equivalent.

We may now consider again the point raised at the end of §3. It was suggested tentatively that the question, "Can machines think?" should be replaced by "Are there imaginable digital computers which would do well in the imitation game?" If we wish we can make this superficially more general and ask "Are there discrete-state machines which would do well?" But in view of the

universality property we see that either of these questions is equivalent to this, "Let us fix our attention on one particular digital computer C. Is it true that by modifying this computer to have an adequate storage, suitably increasing its speed of action, and providing it with an appropriate programme, C can be made to play satisfactorily the part of A in the imitation game, the part of B being taken by a man?"

6. Contrary Views on the Main Question

We may now consider the ground to have been cleared and we are ready to proceed to the debate on our question, "Can machines think?" and the variant of it quoted at the end of the last section. We cannot altogether abandon the original form of the problem, for opinions will differ as to the appropriateness of the substitution and we must at least listen to what has to be said in this connexion.

It will simplify matters for the reader if I explain first my own beliefs in the matter. Consider first the more accurate form of the question. I believe that in about fifty years' time it will be possible, to programme computers, with a storage capacity of about 109, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning. The original question, "Can machines think?" I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted. I believe further that no useful purpose is served by concealing these beliefs. The popular view that scientists proceed inexorably from well-established fact to well-established fact, never being influenced by any improved conjecture, is quite mistaken. Provided it is made clear which are proved facts and which are conjectures, no harm can result. Conjectures are of great importance since they suggest useful lines of research.

I now proceed to consider opinions opposed to my own.

(1) The Theological Objection
Thinking is a function of man's immortal soul. God has given an immortal soul to every man and woman, but not to any other animal or to machines. Hence no animal or machine can think.

I am unable to accept any part of this, but will attempt to reply in theological terms. I should find the argument more convincing if animals were classed with men, for there is a greater difference, to my mind, between the typical animate and the inanimate than there is between man and the other animals. The arbitrary character of the orthodox view becomes clearer if we consider how it might appear to a member of some other religious community. How do Christians regard the Moslem view that women have no souls? But let us leave this point aside and return to the main argument. It appears to me that the argument quoted above implies a serious restriction of the omnipotence of the Almighty. It is admitted that there are certain things that He cannot do such as making one equal to two, but should we not believe that He has freedom to confer a soul on an elephant if He sees fit? We might expect that He would only exercise this power in conjunction with a mutation which provided the elephant with an appropriately improved brain to minister to the needs of this sort[. An argument of exactly similar form may be made for the case of machines. It may seem different because it is more difficult to "swallow." But this really only means that we think it would be less likely that He would consider the circumstances suitable for conferring a soul. The circumstances in question are discussed in the rest of this paper. In attempting to construct such machines we should not be irreverently usurping His power of creating souls, any more than we are in the procreation of children: rather we are, in either case, instruments of His will providing .mansions for the souls that He creates.

However, this is mere speculation. I am not very impressed with theological arguments whatever they may be used to support. Such arguments have often been found unsatisfactory in the past. In the time of Galileo it was argued that the texts, "And the sun stood still . . . and hasted not to go down about a whole day" (Joshua x. 13) and "He laid the foundations of the earth, that it should not move at any time" (Psalm cv. 5) were an adequate refutation of the Copernican theory. With our present knowledge such an argument appears futile. When that knowledge was not available it made a quite different impression.

(2) The "Heads in the Sand" Objection
The consequences of machines thinking would be too dreadful. Let us hope and believe that they cannot do so."

This argument is seldom expressed quite so openly as in the form above. But it affects most of us who think about it at all. We like to believe that Man is in some subtle way superior to the rest of creation. It is best if he can be shown to be necessarily superior, for then there is no danger of him losing his commanding position. The popularity of the theological argument is clearly connected with this feeling. It is likely to be quite strong in intellectual people, since they value the power of thinking more highly than others, and are more inclined to base their belief in the superiority of Man on this power.

I do not think that this argument is sufficiently substantial to require refutation. Consolation would be more appropriate: perhaps this should be sought in the transmigration of souls.

(3) The Mathematical Objection
There are a number of results of mathematical logic which can be used to show that there are limitations to the powers of discrete-state machines. The best known of these results is known as Godel's theorem (1931) and shows that in any sufficiently powerful logical system statements can be formulated which can neither be proved nor disproved within the system, unless possibly the system itself is inconsistent. There are other, in some respects similar, results due to Church (1936), Kleene (1935), Rosser, and Turing (1937). The latter result is the most convenient to consider, since it refers directly to machines, whereas the others can only be used in a comparatively indirect argument: for instance if Godel's theorem is to be used we need in addition to have some means of describing logical systems in terms of machines, and machines in terms of logical systems. The result in question refers to a type of machine which is essentially a digital computer with an infinite capacity. It states that there are certain things that such a machine cannot do. If it is rigged up to give answers to questions as in the imitation game, there will be some questions to which it will either give a wrong answer, or fail to give an answer at all however much time is allowed for a reply. There may, of course, be many such questions, and questions which cannot be answered by one machine may be satisfactorily answered by another. We are of course supposing for the present that the questions are of the kind to which an answer "Yes" or "No" is appropriate, rather than questions such as "What do you think of Picasso?" The questions that we know the machines must fail on are of this type, "Consider the machine specified as follows. . . . Will this machine ever answer 'Yes' to any question?" The dots are to be replaced by a description of some machine in a standard form, which could be something like that used in §5. When the machine described bears a certain comparatively simple relation to the machine which is under interrogation, it can be shown that the answer is either wrong or not forthcoming. This is the mathematical result: it is argued that it proves a disability of machines to which the human intellect is not subject.

The short answer to this argument is that although it is established that there are limitations to the Powers If any particular machine, it has only been stated, without any sort of proof, that no such limitations apply to the human intellect. But I do not think this view can be dismissed quite

so lightly. Whenever one of these machines is asked the appropriate critical question, and gives a definite answer, we know that this answer must be wrong, and this gives us a certain feeling of superiority. Is this feeling illusory? It is no doubt quite genuine, but I do not think too much importance should be attached to it. We too often give wrong answers to questions ourselves to be justified in being very pleased at such evidence of fallibility on the part of the machines. Further, our superiority can only be felt on such an occasion in relation to the one machine over which we have scored our petty triumph. There would be no question of triumphing simultaneously over all machines. In short, then, there might be men cleverer than any given machine, but then again there might be other machines cleverer again, and so on.

Those who hold to the mathematical argument would, I think, mostly he willing to accept the imitation game as a basis for discussion, Those who believe in the two previous objections would probably not be interested in any criteria.

(4) The Argument from Consciousness
This argument is very, well expressed in Professor Jefferson's Lister Oration for 1949, from which I quote. "Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain-that is, not only write it but know that it had written it. No mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants."

This argument appears to be a denial of the validity of our test. According to the most extreme form of this view the only way by which one could be sure that machine thinks is to be the machine and to feel oneself thinking. One could then describe these feelings to the world, but of course no one would be justified in taking any notice. Likewise according to this view the only way to know that a man thinks is to be that particular man. It is in fact the solipsist point of view. It may be the most logical view to hold but it makes communication of ideas difficult. A is liable to believe "A thinks but B does not" whilst B believes "B thinks but A does not." instead of arguing continually over this point it is usual to have the polite convention that everyone thinks.

I am sure that Professor Jefferson does not wish to adopt the extreme and solipsist point of view. Probably he would be quite willing to accept the imitation game as a test. The game (with the player B omitted) is frequently used in practice under the name of viva voce to discover whether some one really understands something or has "learnt it parrot fashion." Let us listen in to a part of such a viva voce:

Interrogator: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?
Witness: It wouldn't scan.
Interrogator: How about "a winter's day," That would scan all right.
Witness: Yes, but nobody wants to be compared to a winter's day.
Interrogator: Would you say Mr. Pickwick reminded you of Christmas?
Witness: In a way.
Interrogator: Yet Christmas is a winter's day, and I do not think Mr. Pickwick would mind the comparison.
Witness: I don't think you're serious. By a winter's day one means a typical winter's day, rather than a special one like Christmas.

And so on, What would Professor Jefferson say if the sonnet-writing machine was able to answer like this in the viva voce? I do not know whether he would regard the machine as "merely artificially signalling" these answers, but if the answers were as satisfactory and sustained as in

the above passage I do not think he would describe it as "an easy contrivance." This phrase is, I think, intended to cover such devices as the inclusion in the machine of a record of someone reading a sonnet, with appropriate switching to turn it on from time to time.

In short then, I think that most of those who support the argument from consciousness could be persuaded to abandon it rather than be forced into the solipsist position. They will then probably be willing to accept our test.

I do not wish to give the impression that I think there is no mystery about consciousness. There is, for instance, something of a paradox connected with any attempt to localise it. But I do not think these mysteries necessarily need to be solved before we can answer the question with which we are concerned in this paper.

(5) Arguments from Various Disabilities
These arguments take the form, "I grant you that you can make machines do all the things you have mentioned but you will never be able to make one to do X." Numerous features X are suggested in this connexion I offer a selection:

Be kind, resourceful, beautiful, friendly, have initiative, have a sense of humour, tell right from wrong, make mistakes, fall in love, enjoy strawberries and cream, make some one fall in love with it, learn from experience, use words properly, be the subject of its own thought, have as much diversity of behaviour as a man, do something really new.

No support is usually offered for these statements. I believe they are mostly founded on the principle of scientific induction. A man has seen thousands of machines in his lifetime. From what he sees of them he draws a number of general conclusions. They are ugly, each is designed for a very limited purpose, when required for a minutely different purpose they are useless, the variety of behaviour of any one of them is very small, etc., etc. Naturally he concludes that these are necessary properties of machines in general. Many of these limitations are associated with the very small storage capacity of most machines. (I am assuming that the idea of storage capacity is extended in some way to cover machines other than discrete-state machines. The exact definition does not matter as no mathematical accuracy is claimed in the present discussion,) A few years ago, when very little had been heard of digital computers, it was possible to elicit much incredulity concerning them, if one mentioned their properties without describing their construction. That was presumably due to a similar application of the principle of scientific induction. These applications of the principle are of course largely unconscious. When a burnt child fears the fire and shows that he fears it by avoiding it, f should say that he was applying scientific induction. (I could of course also describe his behaviour in many other ways.) The works and customs of mankind do not seem to be very suitable material to which to apply scientific induction. A very large part of space-time must be investigated, if reliable results are to be obtained. Otherwise we may (as most English 'Children do) decide that everybody speaks English, and that it is silly to learn French.

There are, however, special remarks to be made about many of the disabilities that have been mentioned. The inability to enjoy strawberries and cream may have struck the reader as frivolous. Possibly a machine might be made to enjoy this delicious dish, but any attempt to make one do so would be idiotic. What is important about this disability is that it contributes to some of the other disabilities, e.g., to the difficulty of the same kind of friendliness occurring between man and machine as between white man and white man, or between black man and black man.

The claim that "machines cannot make mistakes" seems a curious one. One is tempted to retort, "Are they any the worse for that?" But let us adopt a more sympathetic attitude, and try to see what is really meant. I think this criticism can be explained in terms of the imitation game. It is claimed that the interrogator could distinguish the machine from the man simply by setting them a

number of problems in arithmetic. The machine would be unmasked because of its deadly accuracy. The reply to this is simple. The machine (programmed for playing the game) would not attempt to give the right answers to the arithmetic problems. It would deliberately introduce mistakes in a manner calculated to confuse the interrogator. A mechanical fault would probably show itself through an unsuitable decision as to what sort of a mistake to make in the arithmetic. Even this interpretation of the criticism is not sufficiently sympathetic. But we cannot afford the space to go into it much further. It seems to me that this criticism depends on a confusion between two kinds of mistake, We may call them "errors of functioning" and "errors of conclusion." Errors of functioning are due to some mechanical or electrical fault which causes the machine to behave otherwise than it was designed to do. In philosophical discussions one likes to ignore the possibility of such errors; one is therefore discussing "abstract machines." These abstract machines are mathematical fictions rather than physical objects. By definition they are incapable of errors of functioning. In this sense we can truly say that "machines can never make mistakes." Errors of conclusion can only arise when some meaning is attached to the output signals from the machine. The machine might, for instance, type out mathematical equations, or sentences in English. When a false proposition is typed we say that the machine has committed an error of conclusion. There is clearly no reason at all for saying that a machine cannot make this kind of mistake. It might do nothing but type out repeatedly "O = I." To take a less perverse example, it might have some method for drawing conclusions by scientific induction. We must expect such a method to lead occasionally to erroneous results.

The claim that a machine cannot be the subject of its own thought can of course only be answered if it can be shown that the machine has some thought with some subject matter. Nevertheless, "the subject matter of a machine's operations" does seem to mean something, at least to the people who deal with it. If, for instance, the machine was trying to find a solution of the equation x2 - 40x - 11 = 0 one would be tempted to describe this equation as part of the machine's subject matter at that moment. In this sort of sense a machine undoubtedly can be its own subject matter. It may be used to help in making up its own programmes, or to predict the effect of alterations in its own structure. By observing the results of its own behaviour it can modify its own programmes so as to achieve some purpose more effectively. These are possibilities of the near future, rather than Utopian dreams.

The criticism that a machine cannot have much diversity of behaviour is just a way of saying that it cannot have much storage capacity. Until fairly recently a storage capacity of even a thousand digits was very rare.

The criticisms that we are considering here are often disguised forms of the argument from consciousness, Usually if one maintains that a machine can do one of these things, and describes the kind of method that the machine could use, one will not make much of an impression. It is thought that tile method (whatever it may be, for it must be mechanical) is really rather base. Compare the parentheses in Jefferson's statement quoted on page 22.

(6) Lady Lovelace's Objection
Our most detailed information of Babbage's Analytical Engine comes from a memoir by Lady Lovelace (1842). In it she states, "The Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform" (her italics). This statement is quoted by Hartree (1949) who adds: "This does not imply that it may not be possible to construct electronic equipment which will 'think for itself,' or in which, in biological terms, one could set up a conditioned reflex, which would serve as a basis for 'learning.' Whether this is possible in principle or not is a stimulating and exciting question, suggested by some of these recent developments But it did not seem that the machines constructed or projected at the time had this property."

I am in thorough agreement with Hartree over this. It will be noticed that he does not assert that the machines in question had not got the property, but rather that the evidence available to Lady Lovelace did not encourage her to believe that they had it. It is quite possible that the machines in question had in a sense got this property. For suppose that some discrete-state machine has the property. The Analytical Engine was a universal digital computer, so that, if its storage capacity and speed were adequate, it could by suitable programming be made to mimic the machine in question. Probably this argument did not occur to the Countess or to Babbage. In any case there was no obligation on them to claim all that could be claimed.

This whole question will be considered again under the heading of learning machines.

A variant of Lady Lovelace's objection states that a machine can "never do anything really new." This may be parried for a moment with the saw, "There is nothing new under the sun." Who can be certain that "original work" that he has done was not simply the growth of the seed planted in him by teaching, or the effect of following well-known general principles. A better variant of the objection says that a machine can never "take us by surprise." This statement is a more direct challenge and can be met directly. Machines take me by surprise with great frequency. This is largely because I do not do sufficient calculation to decide what to expect them to do, or rather because, although I do a calculation, I do it in a hurried, slipshod fashion, taking risks. Perhaps I say to myself, "I suppose the Voltage here ought to he the same as there: anyway let's assume it is." Naturally I am often wrong, and the result is a surprise for me for by the time the experiment is done these assumptions have been forgotten. These admissions lay me open to lectures on the subject of my vicious ways, but do not throw any doubt on my credibility when I testify to the surprises I experience.

I do not expect this reply to silence my critic. He will probably say that h surprises are due to some creative mental act on my part, and reflect no credit on the machine. This leads us back to the argument from consciousness, and far from the idea of surprise. It is a line of argument we must consider closed, but it is perhaps worth remarking that the appreciation of something as surprising requires as much of a "creative mental act" whether the surprising event originates from a man, a book, a machine or anything else.

The view that machines cannot give rise to surprises is due, I believe, to a fallacy to which philosophers and mathematicians are particularly subject. This is the assumption that as soon as a fact is presented to a mind all consequences of that fact spring into the mind simultaneously with it. It is a very useful assumption under many circumstances, but one too easily forgets that it is false. A natural consequence of doing so is that one then assumes that there is no virtue in the mere working out of consequences from data and general principles.

(7) Argument from Continuity in the Nervous System
The nervous system is certainly not a discrete-state machine. A small error in the information about the size of a nervous impulse impinging on a neuron, may make a large difference to the size of the outgoing impulse. It may be argued that, this being so, one cannot expect to be able to mimic the behaviour of the nervous system with a discrete-state system.

It is true that a discrete-state machine must be different from a continuous machine. But if we adhere to the conditions of the imitation game, the interrogator will not be able to take any advantage of this difference. The situation can be made clearer if we consider sonic other simpler continuous machine. A differential analyser will do very well. (A differential analyser is a certain kind of machine not of the discrete-state type used for some kinds of calculation.) Some of these provide their answers in a typed form, and so are suitable for taking part in the game. It would not be possible for a digital computer to predict exactly what answers the differential analyser would give to a problem, but it would be quite capable of giving the right sort of answer. For

instance, if asked to give the value of (actually about 3.1416) it would be reasonable to choose at random between the values 3.12, 3.13, 3.14, 3.15, 3.16 with the probabilities of 0.05, 0.15, 0.55, 0.19, 0.06 (say). Under these circumstances it would be very difficult for the interrogator to distinguish the differential analyser from the digital computer.

(8) The Argument from Informality of Behaviour

It is not possible to produce a set of rules purporting to describe what a man should do in every conceivable set of circumstances. One might for instance have a rule that one is to stop when one sees a red traffic light, and to go if one sees a green one, but what if by some fault both appear together? One may perhaps decide that it is safest to stop. But some further difficulty may well arise from this decision later. To attempt to provide rules of conduct to cover every eventuality, even those arising from traffic lights, appears to be impossible. With all this I agree.

From this it is argued that we cannot be machines. I shall try to reproduce the argument, but I fear I shall hardly do it justice. It seems to run something like this. "if each man had a definite set of rules of conduct by which he regulated his life he would be no better than a machine. But there are no such rules, so men cannot be machines." The undistributed middle is glaring. I do not think the argument is ever put quite like this, but I believe this is the argument used nevertheless. There may however be a certain confusion between "rules of conduct" and "laws of behaviour" to cloud the issue. By "rules of conduct" I mean precepts such as "Stop if you see red lights," on which one can act, and of which one can be conscious. By "laws of behaviour" I mean laws of nature as applied to a man's body such as "if you pinch him he will squeak." If we substitute "laws of behaviour which regulate his life" for "laws of conduct by which he regulates his life" in the argument quoted the undistributed middle is no longer insuperable. For we believe that it is not only true that being regulated by laws of behaviour implies being some sort of machine (though not necessarily a discrete-state machine), but that conversely being such a machine implies being regulated by such laws. However, we cannot so easily convince ourselves of the absence of complete laws of behaviour as of complete rules of conduct. The only way we know of for finding such laws is scientific observation, and we certainly know of no circumstances under which we could say, "We have searched enough. There are no such laws."

We can demonstrate more forcibly that any such statement would be unjustified. For suppose we could be sure of finding such laws if they existed. Then given a discrete-state machine it should certainly be possible to discover by observation sufficient about it to predict its future behaviour, and this within a reasonable time, say a thousand years. But this does not seem to be the case. I have set up on the Manchester computer a small programme using only 1,000 units of storage, whereby the machine supplied with one sixteen-figure number replies with another within two seconds. I would defy anyone to learn from these replies sufficient about the programme to be able to predict any replies to untried values.

(9) The Argument from Extrasensory Perception

I assume that the reader is familiar with the idea of extrasensory perception, and the meaning of the four items of it, viz., telepathy, clairvoyance, precognition and psychokinesis. These disturbing phenomena seem to deny all our usual scientific ideas. How we should like to discredit them! Unfortunately the statistical evidence, at least for telepathy, is overwhelming. It is very difficult to rearrange one's ideas so as to fit these new facts in. Once one has accepted them it does not seem a very big step to believe in ghosts and bogies. The idea that our bodies move simply according to the known laws of physics, together with some others not yet discovered but somewhat similar, would be one of the first to go.

This argument is to my mind quite a strong one. One can say in reply that many scientific theories seem to remain workable in practice, in spite of clashing with ESP; that in fact one can get along very nicely if one forgets about it. This is rather cold comfort, and one fears that

thinking is just the kind of phenomenon where ESP may be especially relevant.

A more specific argument based on ESP might run as follows: "Let us play the imitation game, using as witnesses a man who is good as a telepathic receiver, and a digital computer. The interrogator can ask such questions as 'What suit does the card in my right hand belong to?' The man by telepathy or clairvoyance gives the right answer 130 times out of 400 cards. The machine can only guess at random, and perhaps gets 104 right, so the interrogator makes the right identification." There is an interesting possibility which opens here. Suppose the digital computer contains a random number generator. Then it will be natural to use this to decide what answer to give. But then the random number generator will be subject to the psychokinetic powers of the interrogator. Perhaps this psychokinesis might cause the machine to guess right more often than would be expected on a probability calculation, so that the interrogator might still be unable to make the right identification. On the other hand, he might be able to guess right without any questioning, by clairvoyance. With ESP anything may happen.

If telepathy is admitted it will be necessary to tighten our test up. The situation could be regarded as analogous to that which would occur if the interrogator were talking to himself and one of the competitors was listening with his ear to the wall. To put the competitors into a "telepathy-proof room" would satisfy all requirements.

7. Learning Machines

The reader will have anticipated that I have no very convincing arguments of a positive nature to support my views. If I had I should not have taken such pains to point out the fallacies in contrary views. Such evidence as I have I shall now give.

Let us return for a moment to Lady Lovelace's objection, which stated that the machine can only do what we tell it to do. One could say that a man can "inject" an idea into the machine, and that it will respond to a certain extent and then drop into quiescence, like a piano string struck by a hammer. Another simile would be an atomic pile of less than critical size: an injected idea is to correspond to a neutron entering the pile from without. Each such neutron will cause a certain disturbance which eventually dies away. If, however, the size of the pile is sufficiently increased, tire disturbance caused by such an incoming neutron will very likely go on and on increasing until the whole pile is destroyed. Is there a corresponding phenomenon for minds, and is there one for machines? There does seem to be one for the human mind. The majority of them seem to be "subcritical," i.e., to correspond in this analogy to piles of subcritical size. An idea presented to such a mind will on average give rise to less than one idea in reply. A smallish proportion are supercritical. An idea presented to such a mind that may give rise to a whole "theory" consisting of secondary, tertiary and more remote ideas. Animals minds seem to be very definitely subcritical. Adhering to this analogy we ask, "Can a machine be made to be supercritical?"

The "skin-of-an-onion" analogy is also helpful. In considering the functions of the mind or the brain we find certain operations which we can explain in purely mechanical terms. This we say does not correspond to the real mind: it is a sort of skin which we must strip off if we are to find the real mind. But then in what remains we find a further skin to be stripped off, and so on. Proceeding in this way do we ever come to the "real" mind, or do we eventually come to the skin which has nothing in it? In the latter case the whole mind is mechanical. (It would not be a discrete-state machine however. We have discussed this.)

These last two paragraphs do not claim to be convincing arguments. They should rather be described as "recitations tending to produce belief."

The only really satisfactory support that can be given for the view expressed at the beginning of §6, will be that provided by waiting for the end of the century and then doing the experiment described. But what can we say in the meantime? What steps should be taken now if the experiment is to be successful?

As I have explained, the problem is mainly one of programming. Advances in engineering will have to be made too, but it seems unlikely that these will not be adequate for the requirements. Estimates of the storage capacity of the brain vary from 1010 to 1015 binary digits. I incline to the lower values and believe that only a very small fraction is used for the higher types of thinking. Most of it is probably used for the retention of visual impressions, I should be surprised if more than 109 was required for satisfactory playing of the imitation game, at any rate against a blind man. (Note: The capacity of the Encyclopaedia Britannica, 11th edition, is 2 X 109) A storage capacity of 107, would be a very practicable possibility even by present techniques. It is probably not necessary to increase the speed of operations of the machines at all. Parts of modern machines which can be regarded as analogs of nerve cells work about a thousand times faster than the latter. This should provide a "margin of safety" which could cover losses of speed arising in many ways, Our problem then is to find out how to programme these machines to play the game. At my present rate of working I produce about a thousand digits of progratiirne a day, so that about sixty workers, working steadily through the fifty years might accomplish the job, if nothing went into the wastepaper basket. Some more expeditious method seems desirable.

In the process of trying to imitate an adult human mind we are bound to think a good deal about the process which has brought it to the state that it is in. We may notice three components.

(a) The initial state of the mind, say at birth,
(b) The education to which it has been subjected,
(c) Other experience, not to be described as education, to which it has been subjected.

Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain. Presumably the child brain is something like a notebook as one buys it from the stationer's. Rather little mechanism, and lots of blank sheets. (Mechanism and writing are from our point of view almost synonymous.) Our hope is that there is so little mechanism in the child brain that something like it can be easily programmed. The amount of work in the education we can assume, as a first approximation, to be much the same as for the human child.

We have thus divided our problem into two parts. The child programme and the education process. These two remain very closely connected. We cannot expect to find a good child machine at the first attempt. One must experiment with teaching one such machine and see how well it learns. One can then try another and see if it is better or worse. There is an obvious connection between this process and evolution, by the identifications

Structure of the child machine = hereditary material
Changes of the child machine = mutation,
Natural selection = judgment of the experimenter

One may hope, however, that this process will be more expeditious than evolution. The survival of the fittest is a slow method for measuring advantages. The experimenter, by the exercise of intelligence, should he able to speed it up. Equally important is the fact that he is not restricted to random mutations. If he can trace a cause for some weakness he can probably think of the kind of mutation which will improve it.

It will not be possible to apply exactly the same teaching process to the machine as to a normal child. It will not, for instance, be provided with legs, so that it could not be asked to go out and fill the coal scuttle. Possibly it might not have eyes. But however well these deficiencies might be overcome by clever engineering, one could not send the creature to school without the other children making excessive fun of it. It must be given some tuition. We need not be too concerned about the legs, eyes, etc. The example of Miss Helen Keller shows that education can take place provided that communication in both directions between teacher and pupil can take place by some means or other.

We normally associate punishments and rewards with the teaching process. Some simple child machines can be constructed or programmed on this sort of principle. The machine has to be so constructed that events which shortly preceded the occurrence of a punishment signal are unlikely to be repeated, whereas a reward signal increased the probability of repetition of the events which led up to it. These definitions do not presuppose any feelings on the part of the machine, I have done some experiments with one such child machine, and succeeded in teaching it a few things, but the teaching method was too unorthodox for the experiment to be considered really successful.

The use of punishments and rewards can at best be a part of the teaching process. Roughly speaking, if the teacher has no other means of communicating to the pupil, the amount of information which can reach him does not exceed the total number of rewards and punishments applied. By the time a child has learnt to repeat "Casabianca" he would probably feel very sore indeed, if the text could only be discovered by a "Twenty Questions" technique, every "NO" taking the form of a blow. It is necessary therefore to have some other "unemotional" channels of communication. If these are available it is possible to teach a machine by punishments and rewards to obey orders given in some language, e.g., a symbolic language. These orders are to be transmitted through the "unemotional" channels. The use of this language will diminish greatly the number of punishments and rewards required.

Opinions may vary as to the complexity which is suitable in the child machine. One might try to make it as simple as possible consistently with the general principles. Alternatively one might have a complete system of logical inference "built in."' In the latter case the store would be largely occupied with definitions and propositions. The propositions would have various kinds of status, e.g., well-established facts, conjectures, mathematically proved theorems, statements given by an authority, expressions having the logical form of proposition but not belief-value. Certain propositions may be described as "imperatives." The machine should be so constructed that as soon as an imperative is classed as "well established" the appropriate action automatically takes place. To illustrate this, suppose the teacher says to the machine, "Do your homework now." This may cause "Teacher says 'Do your homework now' " to be included amongst the well-established facts. Another such fact might be, "Everything that teacher says is true." Combining these may eventually lead to the imperative, "Do your homework now," being included amongst the well-established facts, and this, by the construction of the machine, will mean that the homework actually gets started, but the effect is very satisfactory. The processes of inference used by the machine need not be such as would satisfy the most exacting logicians. There might for instance be no hierarchy of types. But this need not mean that type fallacies will occur, any more than we are bound to fall over unfenced cliffs. Suitable imperatives (expressed within the systems, not forming part of the rules of the system) such as "Do not use a class unless it is a subclass of one which has been mentioned by teacher" can have a similar effect to "Do not go too near the edge."

The imperatives that can be obeyed by a machine that has no limbs are bound to be of a rather intellectual character, as in the example (doing homework) given above. important amongst such imperatives will be ones which regulate the order in which the rules of the logical system concerned are to be applied, For at each stage when one is using a logical system, there is a very

large number of alternative steps, any of which one is permitted to apply, so far as obedience to the rules of the logical system is concerned. These choices make the difference between a brilliant and a footling reasoner, not the difference between a sound and a fallacious one. Propositions leading to imperatives of this kind might be "When Socrates is mentioned, use the syllogism in Barbara" or "If one method has been proved to be quicker than another, do not use the slower method." Some of these may be "given by authority," but others may be produced by the machine itself, e.g. by scientific induction.

The idea of a learning machine may appear paradoxical to some readers. How can the rules of operation of the machine change? They should describe completely how the machine will react whatever its history might be, whatever changes it might undergo. The rules are thus quite time-invariant. This is quite true. The explanation of the paradox is that the rules which get changed in the learning process are of a rather less pretentious kind, claiming only an ephemeral validity. The reader may draw a parallel with the Constitution of the United States.

An important feature of a learning machine is that its teacher will often be very largely ignorant of quite what is going on inside, although he may still be able to some extent to predict his pupil's behavior. This should apply most strongly to the later education of a machine arising from a child machine of well-tried design (or programme). This is in clear contrast with normal procedure when using a machine to do computations one's object is then to have a clear mental picture of the state of the machine at each moment in the computation. This object can only be achieved with a struggle. The view that "the machine can only do what we know how to order it to do,"' appears strange in face of this. Most of the programmes which we can put into the machine will result in its doing something that we cannot make sense (if at all, or which we regard as completely random behaviour. Intelligent behaviour presumably consists in a departure from the completely disciplined behaviour involved in computation, but a rather slight one, which does not give rise to random behaviour, or to pointless repetitive loops. Another important result of preparing our machine for its part in the imitation game by a process of teaching and learning is that "human fallibility" is likely to be omitted in a rather natural way, i.e., without special "coaching." (The reader should reconcile this with the point of view on pages 23 and 24.) Processes that are learnt do not produce a hundred per cent certainty of result; if they did they could not be unlearnt.

It is probably wise to include a random element in a learning machine. A random element is rather useful when we are searching for a solution of some problem. Suppose for instance we wanted to find a number between 50 and 200 which was equal to the square of the sum of its digits, we might start at 51 then try 52 and go on until we got a number that worked. Alternatively we might choose numbers at random until we got a good one. This method has the advantage that it is unnecessary to keep track of the values that have been tried, but the disadvantage that one may try the same one twice, but this is not very important if there are several solutions. The systematic method has the disadvantage that there may be an enormous block without any solutions in the region which has to be investigated first, Now the learning process may be regarded as a search for a form of behaviour which will satisfy the teacher (or some other criterion). Since there is probably a very large number of satisfactory solutions the random method seems to be better than the systematic. It should be noticed that it is used in the analogous process of evolution. But there the systematic method is not possible. How could one keep track of the different genetical combinations that had been tried, so as to avoid trying them again?

We may hope that machines will eventually compete with men in all purely intellectual fields. But which are the best ones to start with? Even this is a difficult decision. Many people think that a very abstract activity, like the playing of chess, would be best. It can also be maintained that it is best to provide the machine with the best sense organs that money can buy, and then teach it to understand and speak English. This process could follow the normal teaching of a child. Things

would be pointed out and named, etc. Again I do not know what the right answer is, but I think both approaches should be tried.

We can only see a short distance ahead, but we can see plenty there that needs to be done.

Bibliography

Samuel Butler, Erewhon, London 1865, Chapters 23, 24, 25, *The Book of the Machines*.

Alonzo Church, "An Unsolvable Problem of Elementary Number Theory", *American Journal of Math.*, 58(1936), 345~363.

Kurt Gödel, "Über formal unentscheidbare Stätze der Principia Mathematica und verwandter Systeme, I", *Monatshefte für Math. und Phys.*, (1931), 173~189.

D.R. Hartree, *Calculating Instruments and Machines*, New York, 1949.

S.C. Kleene "General Recursive Functions of Natural Numbers", *American Journal of Math.*, 57 (1935), 153~173 and 219~244.

G. Jefferson, "The Mind of Mechanical Man", Lister Oration for 1949. *British Medical Journal*, vol. i(1949), 1105~1121.

Countess of Lovelace, 'Translator's notes to an article on Babbage's Analytical Engine', *Scientific Memoirs* (ed. by R. Taylor), vol.3(1984), 691~731.

Bertrand Russell, *History of Western Philosophy*, London, 1940.

Allen M. Turing, "On Computable Numbers, with an Application to the Entscheidungsproblem", *Proc. of London Math. Soc.*(2), 42 (1937), 230~265.

집필자 소속 Victoria University of Manchester

“앨런 튜링과 튜링 기계”

앤드루 하쥐즈 Andrew Hodges
(영국 옥스퍼드 대학 수학 연구소)

1936년 논문 간행 50주년의 금을 그어 놓으면서, 이 책자에서는 튜링 기계 개념에[19] 대한 오래도록 지속되는 영향력을 깨닫는다. 아주 많은 분야로부터 나온 새로운 논문들을 한데 모아 놓은 것은, 그 영향력이 예외일 만큼 광범위함을 드러낸다. 앨런 M. 튜링 그 자신의 인간적 측면에 대해 회고하고 경의를 표하고자 하는 간략한 이 글에서는, 튜링의 저작에 대한 역사적 전기를 언급할 뿐만 아니라, 튜링의 그 기계 개념에 대한 도입 및 발전 밑바닥에 깔려 있는 생각의 폭을 전달해 주려고 할 것이다.

앨런 튜링은 1912년 6월 23일 태어났고, 전통적인 영국 상위 중산 계층의 교육 방식을 끝낸 뒤에 케임브리지 대학 킹스 칼리지(King's College)에서 장학금을 받았다. 1934년 수학에서 뛰어난 성적으로 졸업하였다. 1년에 채 못 되어 킹스 칼리지의 선임 연구원으로 선출되었고, 그의 학위논문이 ‘중심 극한(Central Limit) 정리’의 증명이었는데, 그 자신의 관심거리들이 확률 이론에 초점이 모아졌었던 듯이 보인다.

그렇지만 실제로 튜링의 연구는 결코 한 가지 분야에만 국한되지 않았으며, 특히 적어도 1933년 이후로는 수학의 기초에 관심을 쏟았다. 1935년 초에 탁월한 케임브리지 위상 수학자 뉴먼(M.H.A. Newman)에 의해 주어진 ‘수리 논리’를 수강하였는데, 이 강좌는 힐버트(Hilbert)가 정의해 놓은 의미에서 산수의 공리계가 어떤 것도 일관되면서[20] 동시에 완벽할 수 없다는 괴델(Gödel)의 증명에 대하여 온전히 다뤄 주는 것으로 끝을 맺었다. 그러나 수학의 기초를 위한 힐버트의 미해결 문젯거리들도[21] 또한 임의의 한 공리계가 결정 가능한지 여부에 대하여 논의를 불러 일으켰는데, 원리상으로 수학적인 주장(assertion 언명)의 참값을 결정해 주기 위하여 적용될 수 있는 명확한(definite) 방법이[22] 존재하는지 여부이다. 그리고 이 ‘결정 가능성 문제’(Entscheidingsproblem)가 여전히 해결되지 않은 채 남아 있었다. 튜링은 힐버트의 도전에 의해서 매료되었고, 홀로 그 문제를 놓고서 작업을 하였으며, 24살 되던 1936년 4월에 “계산 가능한 숫자(=기수)들에 대하여 — 결정 가능성 문제에 대한 응용과 더불어”(*On Computable Numbers, with an Application to the Entscheidungsproblem*)라는 이름을 붙인 논문을 뉴

19) 이 글은 롤프 허킨(Rolf Herken) 편(1994)《보편 튜링 기계: 반세기 간의 연구》Springer-Verlag에 들어 있는 첫 논문이다(경상대학교 중앙도서관에 있음). 튜링 기계를 가리키는 universal machine은 ‘범용 기계’라고도 번역할 수 있다(汎用: 여러 용도에 두루 쓰이는 기계). 그렇지만 아마 universal이란 꾸밈말은 유형이론(type theory)과 관련되어 있는 듯하다. 즉 다른 상위 유형을 도입하지 않고서도 내부 모순이 발생하지 않도록 조치해 나가는 방식인 것이다. 여기서는 ‘보편 기계’라는 말로 번역해 두겠는데, 정신을 갖춘 어떤 기계의 작동 방식을 그대로 모방하여 완벽히 구현해 낸다는 뜻을 지닌다(보편 적용/보편 응용). Turing(1912~1954)을 미국인들은 ‘투링’이라고 부르고, 영국에서는 ‘튜링’ 또는 ‘튜어링’이라고 한다. 여기서는 ‘튜링’으로 적어 두겠다. 이 논문에서의 각주는 모두 번역자가 달아 놓은 것이다.

20) 흔히 consistent를 ‘무-모순’이라고 번역하지만, 아마 더 숙고해 봐야 할 듯하다. ‘non-contradictory’(무-모순)이란 용어는 다른 유형(증명이 이뤄진 더 상위의 유형)의 개념을 가리키기 때문이다. 수학에서 다루는 대상들은 어떤 것이든 상관없이 계산 가능성(computability: 전개 과정의 타당성) 및 결정 가능성(decidability: 증명의 타당성)을 충족시켜 주어야 한다고 할 때에, 앞의 개념은 임의의 공리계에서 주어진 무정의 용어(undefined terms)들로 연역될 수 있어야 함을 가리키며, 뒤의 개념은 그 도출 과정이 처음부터 끝까지 참임이 증명되어야 한다는 뜻이다. consistent(일관된)는 계산 가능성을 충족시킨다는 개념이며, non-contradictory(무-모순)는 그 계산 가능성이 완벽히 증명되어 거짓된 도출 과정이 전혀 없음을 가리킨다. 즉, consistent는 도출 과정이 오직 한 공리계의 무정의 용어들로써만 이뤄졌다는 뜻이지만, non-contradictory는 임의의 도출과정이 검증이나 증명을 거쳤으므로 그 공리계가 붕괴되지 않는 한 거짓 값이 전혀 깃들 수 없다는 뜻이다. 비록 동일한 결과가 나오지만, 서로 지향하는 차원과 동기가 다른 것이므로, 용어 사용에 엄격해질 필요가 있다.

21) 1900년 수학자 파리 국제 대회에서 개회사로서 힐버트가 특별 강연을 하게 되는데, 민코프스키(Minkowski) 및 허위츠(Hurwitz)와 상의하여 “수학적 교훈”(*L'enseignement mathématique*)이라는 제목으로 세계 수학자들이 해결하기를 희망하는 수학 문제 23개 조항을 발표하였다. 흔히 이를 ‘파리 문제’ 또는 ‘힐버트 문제’라고 한다(1902년 Newson이《미국 수학회지》제 8권에 영어로 번역해 놓음). 이일해 번역(1989)《힐버트: 수학과 삶》(민음사) 112쪽 이하에 나와 있는데, 그는 “모든 수학적 문제는 반드시 해결될 수 있다.”는 신념에 불타고 있었다.

22) 아래에서도 이 용어가 계속 나온다. 문외한으로서는 ‘엄밀한 수학적 방법’ 정도가 더 나을 듯싶다. definite (명확한, 영역과 대상이 명확히 규정된, 확정적인)이란 수식어는 수학 사전에서는 ‘정’(正) 또는 ‘정’(定)으로 번역되어 있다(정비례 definite proportions, 정적분 definite integral). 이를 응용하면, 진리에 도달하는 정법(正法 올바른 방법) 또는 정법(定法 확정된 방법)으로도 번역해 볼 수 있다. 그러나 수학도 확정된 것이 아니라 다른 사고 법칙에 의해 지배되고 종속되어야 하는 지위에 있으며, 공리계로부터 필요충분조건을 준수하면서 도출되어 나오는 일반적인 방법을 탐색하고 있다는 점에서 definite method가 필요충분조건을 준수하는 도출 방법임을 알 수 있다. 일상 언어의 느낌을 살린다면, ‘명확한 방법’이나 ‘확정된 방법’ 또는 ‘확정적 방법’ 정도가 후보로 될 듯하다.

먼에게 갖다 주어 그를 놀라게 하였는데, 이 논문에서 지금 우리가 기리고 있는 ‘추상적인 기계 구성’을 소개하였다.

결정 가능성의 질문을 풀어 나가는 데에 핵심적인 난점은, 수학 문제들을 풀기 위해서 ‘명확한 방법’에 의해 의미되는 바가 무엇인지에 대하여 만족스럽게 정의해 주는 일에 놓여 있다. 튜링이 그 자신의 가장 위대한 독창성을 보여 준 것은 바로 이 논제를 해결하는 데에 있었다. 튜링은 명확한 방법의 본질이 바로 ‘기계적으로’ 적용될 수 있어야 하며, 따라서 기계의 작동에 비추어 그런 방식을 구현하는 일을 모형으로 드러내야 한다는 착상을 얻었다. 그는 종이 띠(tape) 위에 씌어진 기호들을 읽고 쓸 수 있는 어떤 기계에 대한 그림(the picture)을 터득해 냈고, 이 착상을 표준 형태로 가다듬어 놓았다. 튜링 기계에 대한 개념은 이제 논리 및 컴퓨터 과학의 교재들에서 언급되며, 수리 논리의 발전에서 그 역사적 위치에 대한 훨씬 자세한 설명은 이 책자에 있는 다른 논문들에 제시되어 있다. 이하에서의 언급은 다만 튜링 생각의 범위를 소개하려는 의도이다. 심지어 그 독창적인 용어도 이 범위를 암시해 준다. 종이 띠에 대한 기계 ‘판독’(scanning)·유한 수의 많은 ‘구성 내용’(configurations 내부 모습, 형상)들을 지닌 기계·지정된 ‘행동 지침 표’(tables of behaviour)들을 지닌 기계는 기술·양자 물리·심리학이라는 학문 분야를 시사해 준다.

수학자로서 튜링의 경험 및 창의성은 이런 일반적인 착상들을 튜링 기계 개념에 대한 정밀한 형식화로 가다듬는 그의 능력에서 저절로 드러났는데, 오직 아주 제한된 범위의 가능한 ‘행동’들만을 요구하는 기계이다. 그러나 튜링은 아주 작은 묶음(set 집합)으로부터 복잡한 논리적 연산들을 수립해 놓는 방법을 보여 주는 데 포함된 순수히 수학적인 재간에만 만족하지 않았다. 그렇게 제한된 명령 체계(repertoire 컴퓨터 작동 지시 목록)를 처리해 내는 기계들이 ‘명확한 방법’들에 대하여 가장 일반적인 가능한 부류들을 찾아내어 확정하는 일을 보장해 주어야 한다고 논의하면서, 튜링은 순수 수학(mathematics proper) 밖으로 넘어섰다. 여러 쪽에 걸쳐서 비-수학적인 논의를 하면서, 튜링은 ‘명확한 방법’들에 대하여 두 가지 일반적인 모형을 제시하였고, 각각 튜링 기계에 의해서 모형으로 만들어질 수 있음을 논의하였다. 하나는 어떤 절차를 따르는 사람에 대한 것이었다. 모든 단계에서 수행되어야 하는 내용을 다른 사람이 그 일을 도맡아도 계속 이어나갈 수 있는 방식으로 설명해 주면서 완벽히 명시적인 ‘지시사항들에 대한 쪽지’가 씌어질 수 있다는 의미에서 확정적이다. 그러나 유한 수의 가능한 ‘정신 상태’들을 놓고서 분명히 튜링은 좀더 대담한 착상에 매료되었는데, 그 과정의 후속 단계는 어떤 것이든지 반드시 기호들·기호들의 변화 내용·정신 상태에 변화에 대한 관찰로 기술될 수 있어야 한다는 생각이다. 이 논의에 따르면 그 후속 과정에 대한 명시적인 기술은 전혀 주어질 필요가 없다. 그 과정이 어떤 방식으로 그 정신의 내부 구조 속에 내포되어 있는 것만으로 충분하기 때문이다.

계산 가능성에 대한 이런 정의를 갖고서, 계산 불가능한 숫자들의 존재를 예증하는 것이 가능해졌고, 이 점으로부터 부정적인 방식으로 ‘결정 가능성 문제’를 해결하는 쪽으로 나갈 수 있었다. 힐버트가 해결되기를 희망하였던 것과 같은 수학 문제들을 [해 집합이 존재하는지 여부를] 결정해 주기 위한 일반적인 방식은 어떤 것이든 존재할 수 없다. 그러나 이것이 그 제목에 명시되어 있듯이 다만 하나의 ‘응용’이었음에 주목해야 한다. 튜링은 자신의 논문에서 계산 가능성의 개념에 초점을 모았다. 즉, 인간의 계산 작업과 밀접한 연관을 늘 시사해 주면서 실제로 ‘계산 가능한 숫자들’에 초점을 모았다. 절대적 제약 내용들과 더불어, 계산 가능성의 이런 절대적 표준에 대한 튜링의 발견은 그 자체로 수학에서 하나의 대단한 발견이었다. 비록 더 시간이 흘러야만 그 온전한 속뜻을 보여 주겠지만, 튜링의 ‘보편’ 기계에 대한 참신한 정의도 그러하였다.[23] 판독되도록 모방될 기계에 대한 기술 내용이 그 종이 띠 위에 놓이는 경우, 어떤 것이든지 모두 다른 튜링 기계 작업을 [모방하여] 수행할 수 있는 튜링 기계이다.

그렇지만 며칠 안 되어 튜링은 처어취(Church 1936, 1936a)의 논문 한 편 또는 두 편 모두로부터 배웠다. 수학문제 해결 작업이 두 사람 사이에 나란히 진행되었고 결정 가능성 문제에 관하여 동일한 결론에 도달하였던 만큼, 가엾게도 튜링은 이 논문들에 대하여 “똑 같은 일을 다른 방식으로 행하고 있다.”고 적어 놓았다. 1936년 5월 28일 튜링은 자신의 논문을 출간하려고 제출하였지만, 자신의 계산 가능성에 대한 정의가 처어취의 ‘효과적인 계산 가능성’과[24] 동일하였음을 예시해 주는 8월 28일자가 적힌 부

23) 수학의 역설 해결 방식으로만 보면, 임의의 기계를 모방해 내려면, 더 높은 유형의 기계를 상정하였을 것이다. 그런 목적이라면 a meta-Turing machine(상위 유형의 튜링 기계)으로 불렀을 것이다. 그렇지만, universal machine(보편 적용 기계)이라고 부른 데에는 동일한 유형을 유지하면서, 내부에 다른 유형의 요소들이 전혀 스며들지 않도록 함과 동시에 자신의 내부를 명시적으로 드러낸다는 속뜻을 담고 있는 듯하다.

24) computation과 calculation은 더러 구분해서 쓰기도 한다. 충분조건만 고려하는 경우를 연산(computation)

록을 덧붙여 놓았다. 수정된 그 논문은 1937년 말에 《런던 수학 학회 논문집》에 실려 출간되었다. 《기호 논리 학술지》에 실린 그 논문에 대한 춰어취의 서평에서 처음으로 '튜링 기계'라는 낱말을 썼다.

이런 내용이 공식적인 수학 내부의 역사이다. 그러나 전적으로 수리 논리 영역 안에만 국한하여 논문들의 관계를 집중적으로 살펴본다면 잘못일 듯하다. 튜링의 성취의 알곡은 수학 밖에 놓인 착상에 뿌리를 둔 논리학에 대해 응용되는 개념의 발견이었다. 춰어취는 '춰어취의 논제'(Church's thesis)라는 가정을 제안하였다. '효과적으로 계산 가능함'은 반드시 일반적인 반복(회귀) 함수들로써 확인될 수 있어야 한다는 내용이다. 이런 견해에로 이끌어 가는 강력한 수학적 논증들이 있었다. 그러나 훨씬 더 일반적으로 실용적인 계산 문제의 본질에 대하여 생각해 냄으로써, 정신적 과정의 본질로부터 이 논제를 위하여 튜링이 정당한 증거를 찾아내었다. 튜링의 정의는 인간 존재가 현실적으로 구현할 수 있는 바에 모형을 두고 마련된 것이었다.

그 분야의 완벽히 국외자인 튜링으로서, 미해결의 그 문제를 놓고서 어떻게 하여 그렇게 근원적인 접근법을 제공하고 적용할 수 있었을까? 그 실마리에 대해서 해설해 주면서, 튜링 기계의 비공식적인 발전 과정의 역사를 짐작할 수 있을까? 불행하지만 튜링은 1930년대에 자기 생각의 발전을 설명해 주는 글을 전혀 써 놓지 않았다. 그렇지만 학부 시절에 튜링의 내적인 관심거리들에 대해 서광을 던져 주는 한 도막의 증거를 갖는 것만으로도 다행이다. 이는 "영혼의 본질"(Nature of Spirit)로 제목이 붙은 편지글이다. 아마 그가 20살 때인 1932년에 써 놓았다(하쥐즈 Hodges 1983에 모두 들어 있음). 문장 일부가 그 분위기를 전해 준다.

> 과학에서 우주에 대하여 모든 비밀이 임의의 특정 순간에 알려진다면, 그것이 모든 미래에 걸쳐서 어떻게 될 것인지 예측할 수 있다고 줄곧 저는 믿어 왔습니다. … 그렇지만 좀더 현대적인 과학에서는 우리가 원자와 전자들을 다루는 경우에, 그것들에 대한 정확한 상태를 확실히 알아낼 수 없다는 결론에 도달하였습니다.…

튜링이 써 놓은 내용은 일부 에딩튼(A.S. Eddington)에 의해 영향을 받았는데, 학교에서 그의 1928년 책자 《물리 세계의 본질》을 읽었었다. 에딩튼은 양자 역학의 출현 및 고전적인 결정론의 소멸이 물리 법칙과 갈등이 없이 정신도 다시 자율적인 힘에 조화될 수 있다는 견해를 강하게 주창한 인물이었다. 튜링은 정신적 힘에 대한 자신의 유물주의 그림을 더 확대해 놓았다.

> … 우리는 아마 두뇌의 작은 부분에서 또는 두뇌 전반에 걸쳐서 그 원자들의 행동을 결정할 수 있는 의지를 지니고 있습니다. 육신의 나머지는 이를 증폭하기 위해 행동하며 … 육신이 죽는 경우 영혼을 붙들고 있는 육신의 '기계'는 사라져 버리지만, 아마 영혼은 조만간 즉시 새로운 육신을 찾아냅니다. …

튜링 쓰는 말의 의미를 이해하기 위해서, 반드시 이 글이 누구를 위해 씌어졌는지 언급되어야 한다. 이것은 튜링의 학교 친구이며 1930년에 일찍 죽은 크뤼스토퓌 모어콤(Christopher Morcom)의 어머니에게 보낸 사적인 편지이다. 그의 죽음으로 크게 사랑하고 존경하였던 친구를 튜링에게서 빼앗아가 버렸다. 튜링은 그녀의 아들이 어떻게 여태껏 자신을 도와주면서 영혼으로 살아 있다고 믿는지를 편지로 써 보내었다. 이런 감성은 1936년 이후 부각된 신랄한 유물주의자이며 무신론자인 튜링에게는 매우 이례적인 듯하다. 하지만 놀라운 대조 밑바닥에서 공통된 맥을 인식해야 한다. 정신 현상의 순수 신비에 대한 위대한 진지성 및 정신 현상들이 반드시 과학적 세계관과 조화되어야 한다는 똑같이 진지한 확신이다.

정신의 문제는 '계산 가능한 숫자들'(1936)에 대한 열쇠가 된다. 어쨌거나 명확한 기계적 방식들에 대한 물음에서 튜링은 '결정되는 속성'(being determined)의 개념을 추상화하고 가다듬는 기회를 깨닫고 있었으며, 정신에 대한 옛 물음에다 새롭게 가다듬어진 이런 개념을 응용하였다. 하여간 수학의 기초에 대하여 어느 누구에게라도 사뭇 무관한 물음들로 보인 바와 정신에 대한 물질적(=유물주의적) 기술 사이에 어떤 연관을 지각하였던 것이다. 그 연관은 철학적이라기보다는 오히려 과학적 시각이었다. 그가 도달한 바는 새로운 유물론(=유물주의)으로서, 이산 상태에 대한 착상에 근거를 둔 새로운 기술 층위이다. 원자 및 전자의 층위라기보다는 오히려 사실상 두뇌 조직의 생리학 층위인 이런 층위가 정신 현상들에 대한 기술을 언급하는 올바른 층위라는 논의인 것이다. 그가 뒤이어 자신의 인생을 대부분 쏟아 바친 것이 바로 이런 착상을 촉진하고 탐구하는 일이었다.

이라고 부르고, 필요조건까지 고려하는 경우를 계산(calculation)이라고 하는 것이다. 여기서는 다 같이 계산 가능성으로 번역해 둔다.

그 논문(=1937)이 출간된 시기 어름에, 튜링 자신은 프린스턴 대학에 있었다. 따라서 현대 수학에서 가장 뛰어난 인물들에게 자신의 생각을 말해 볼 수 있는 위치에 있었다. 그 반응에 대해서 그는 실망을 하였다. 가령 봐일(Weyl)이 어떻게 튜링의 업적을 평가하는 데 실패하였는지 살펴보기 바란다. 그러나 자기 자신에 대해서 남에게 말하려고 하지 않는 튜링의 개인 성격 요인 이외에도, 전문적인 수학 세계에서 기대된 형식으로 자신의 생각들을 표현해 넣는 일이 아마 사실상 그의 최우선 순위는 아니었다고도 말할 수 있겠다. 1937~38년에 나온 그의 주요한 업적이 형식상 수학적이었음은 사실이다. 튜링의 '서수 논리학'에 대한 작업(Turing 1939)은, 유한 공리계들을 탐구함으로써 괴델의 불완전성 정리의 영향력(force)을 피하는 것이 가능한지 여부에 대한 탐구였다. 또한 리이만 제타 함수(Riemann zeta-function)의 이론을[25] 놓고서도 작업을 하였는데, 거의 고전 수학 중심부에서 자리를 좀더 분명히 잡지 못하는 주제이다(Turing 1943). 그럼에도 논리적 작업은 '직관'(intuition)에 대한 착상을 이해하는 일에 흥미를 지님으로써 강력하게 동기가 주어졌다. 튜링은 이를 무한 공리계를 산출하는 데에 포함된 비계산적(non-computable) 단계들을 찾아내는 일로 간주하였다. 그리고 리이만 제타 함수 문제는 튜니 바퀴들로부터 특정한 목적의 계산기를 구성하도록 동기를 마련해 주었다(Turing 1939a). 그러는 동안에 프린스턴 대학에서 그는 또한 전기 개폐 장치(electric relays)를 써서 이진법 곱셈 기계를 만들어 내었다. 1939년에 케임브리지로 되돌아와서, 그는 또한 수학의 기초(Wittgenstein 1976)에[26] 대한 비트겐슈타인의 토론에 참석하는 작은 모임에 참여하였다. 기계적 속성과 심리적 속성은 계속 그에게 대단한 매력거리로 남아 있었다. 세상일이 그러하듯이, 1936년에는 튜링이 거의 상상해 볼 수 없었던 방식으로, 이런 관심거리들을 발전시키는 전기가 마련되었다.

세계 제 2차 대전은 통상적인 학문의 관심으로부터 튜링의 경력을 바꿔놓았다. 튜링은 통상 관례를 따르는 학자가 아니었고, 그의 경력은 결코 1939년에 정주하여 끝난 것이 아니었다. 그는 미국인 지위를 추구할 가능성을 완전히 거절하였으나, 케임브리지 대학에서는 즉시 강사직에 대한 전망도 조금도 안 겨주지 않았다. 마침내 튜링은 제 2차 대전 동안 특히 독일 해군의 무전 소통 내용들을 해독해 내는 책임을 맡아 영국의 암호해독 분야에서 제 1인자의 과학적 인물이 되었는데, 그 자체로 대단한 역사적 중요성 및 지적 명성을 지닌 공적이다.

튜링은 확률 이론에서 그의 논리적 창의성 및 혁신적 내용에 몫을 충분히 다하면서, 아주 참신한 정교함을 지닌 현실적 흐름도(algorithms)를 고안해 내어 기계로 구현하고 있음을 자각하였다. 불행하게도 정부의 비밀주의가 여전히 비수치적(non-numerical) 계산에서 이러한 발전 내용들에 대하여 약간을 제외하고는 공개하지 못하도록 막아놓고 있다(Hodges 1983, Good 1979). 그렇지만 우리는 장기 두기 놀이와 같은 과제를 실행하는 기계적 방법들의 능력에 관하여, 암호 해독 작업에 포함된 튜링 및 다른 사람들 사이에서 전개된 훨씬 활발한 논의를 알고 있다. 그리고 일반적인 용어로, (실제 기계에 장착이 되어 있든지, 아니면 정형화된 과제를 수행하도록 훈련된 사람에게 들어 있든지 간에) 기계적 방법들이 지금까지 인간의 고유한 판단 영역으로 간주되어 온 상당 부분을 맡고 있음을 알고 있다. 정신적 과제를 기계로 구현해 내는 가능성에 매료되어 있는 튜링을 자극할 수 있도록 보다 더 낫게 계산된 경험이란 상상조차 할 수 없다.

튜링은 또한 당시 디지털 방식으로 운용되는 전기적 구성체의 참신한 이용을 포함하여, 1940년대 초반의 가장 진보된 기술들을 접하였다. 실제로 그는 고도로 수준 높은 그 자신의 설계를 바탕으로 손수 전기적 말소리 [도청 방지] 변환기를 만드는 일에 제 2차 대전의 마지막 해를 대부분 쏟아 부었다(Hodges 1983). 그는 전기적 구성체들이 보편 튜링 기계의 착상을 실용적 형태로 바꾸는 데 필요한 (연산) 속도를 제공해 줄 수 있을 것임을 재빨리 간파하였다. 1945년 무렵 그는 특히 '두뇌를 만들어내는 일'을 말하고 있었다. 국립 물리 실험실(NPL)에 임명되면서 기회가 신속히 찾아왔는데, 거기에서 사실상 전기 컴퓨터를 설계하는 사명이 주어졌다. 1945년 말에 써 놓은 자세한 제안서(Turing 1946)는 세계 최초라는 특권을 부여받을 수 있었다. 그 특권은 1945년 6월의 EDVAC[27] 제안서로 언급된다. 그

25) 한국사전연구원(1989:201) 《최신 수학사전》, s가 복소수이고, $s=\sigma+it$, $\sigma>1$일 때, 다음처럼 정의되는 함수.

$$\zeta(s)=\sum_{n=1}^{\infty}\frac{1}{n^s}=\prod_{p}(1+\frac{1}{p^s})^{-1}$$

26) 1956년 간행된 독일어 원판이 박정일 뒤침(1997) 《수학의 기초에 관한 고찰》(서광사)로 나와 있다.

27) 폰 노이먼(von Neumann)이 제안한 프로그램 내장 방식의 컴퓨터로서, Electronic Discrete Variable Automatic Computer(전기적 이산 변수 자동 계산기)로 불렸고, 1951년에 물리적으로 구현되었다. 오늘날 컴퓨터는 이 기계와는 달리 프로그램을 내장하지 않고 다만 소프트웨어로 운용하는데, 이는 튜링 생각을 따르고 있는 것이다.

러나 튜링은 오직 세부 내용만을 미국 보고서로부터 빌려왔다. 그의 설계는 독립적이었고, 기계에 대하여 소수의 간단한 작동 내용만을 지닌 아주 개성적인 그 자신의 개념내용으로부터 나왔는데, 하드웨어에다 추가되는 것이 아니라, 현재 우리가 이용하는 방식대로 적합한 소프트웨어를 적어 놓는 일에 의해서 다른 모든 것들이 구성될 수 있었다.

전기적 계산기의 능력을 탐구하는 방법에 대한 논의에서, 튜링의 보고서는 똑같은 미국의 시도보다도 훨씬 앞서 있었다. 그는 전쟁 시절에 얻은 정교한 컴퓨터 지시사항 경로(routines)들에[28] 대한 광범위한 경험을, 어떤 종류이든지 상관없이 상징 조작에 영향을 주는 보편 기계의 동력에 대한 이해와 결합시켜 놓을 수 있었다. 튜링은 '계산 가능한 숫자들'(1936)에서 튜링 기계 세부내용을 위하여 상징적 약호들을 썼다. 그로서는 그것이 오직 보편 기계가 그런 약호들을 늘여놓음으로써, 따라서 현대적 의미로 컴퓨터 언어 및 프로그래밍에 대한 착상을 형성하는 일을 스스로 할 수 있음을 간파해 내는 지름길이었다. (실제로 1936년 논문을 읽기가 어렵다. 튜링은 거기에서 현대적 프로그래머의 정신 상태를 채택해야 했다. 따라서 '눈먼'[29] 기계의 작동들을 통하여 스스로 생각을 해 내야만 했다. 그 당시에 컴퓨터 프로그래밍과 같은 것이 존재하지 않았음을 기억하기 바란다.)

1936년에 기술된 '지시사항들에 대한 비망록'은 현재 프로그램의 줄이[30] 되고, 그가 즉시 써 나가기 시작했던 실용적 프로그램들이 되었다. 그러나 그의 관심은 즉각적으로 '지능을 갖춘 기계'로 부른 것의 실현 가능성을 탐구하는 것이었는데, 흔히 현재는 '인공 지능'으로 불릴 수 있다. 다시 말하여, 그는 시작에서부터 '지시사항들에 대한 비망록'으로 명백하게 기술될 수 있는 과정들뿐만 아니라, 모든 정신적 과정들이 논리 기계에 의해 충실하게 똑같이 모방될 수 있다는 논제를 확장하고 탐구하려고 하였다. 경험에 비추어 특히 최초의 지시사항들을 스스로 고쳐 나가는 컴퓨터의 능력에 관심이 있었고, 이런 과정이 인간 학습과 동등하게 간주될 수 있음을 논의하였다. 이런 주제를 놓고서 튜링 자신의 분명한 글(Turing 1946, 1947, 1948)들에 유의할 필요가 있다. (이런 중요한 논문들은 튜링이 살아 있을 동안에는 출간되지 않았었지만, 지금은 모두 접할 수 있다.) '계산 가능한 숫자들'(1936)을 발표하고 나서 꼭 10년 뒤에, 컴퓨터 기술에 대한 잠재성의 강력하고 예언적인 조감 속으로 그가 그 생각들을 어떻게 번역해 놓았는지를 살펴보는 일은 아주 놀랍다. 짧은 기간 동안이나마 국립 물리 실험실(NPL)에서 튜링의 위치는 이런 새로운 기술을 창조해 내는 데에 주도적인 역할을 주는 듯하였지만, 그가 겪은 좌절로 1948년에 사임하게 된다. 대신 맨체스터 대학에 있는 자리를 받아들였는데, 비록 설계를 하지는 않았지만, 거기서 개발되고 있는 전기적 컴퓨터를 자유로이 이용하도록 해 주었다.

그 이후에 튜링은 사실상 놀랍게도 계산의 현대 과학을 수립하는 일은 거의 조금도 하지 않았다. 튜링(1946)에서 프로그래밍 언어에 대한 생각을 간단 명확히 설명해 놓았지만, 그리고 실제로 국립 물리 실험실(NPL)에서 '약호 지시사항'에 대한 그의 작업이 이들 생각을 발전시켜 놓았지만, 맨체스터 대학에서 그는 이것을 전적으로 다른 사람들에게 맡겼고, 그 자신은 매력 없는 base-32(?? 기수-32: 未詳) 표기법으로 표현된 기계 부호에 대해서만 작업을 하였다. 케임브리지 EDSAC[31] 컴퓨터의 개통을 고려하면, "거대한 지시사항 경로 점검하기"(Checking a large routine)에 대한 그의 짤막한 논문(Turing 1949)은 1960년대까지 아직 개발되지 않은 프로그램 증명에 대한 생각을 예견해 주는 것으로 간주된다. 그것은 원리를 개발해 내는 데 기여할 수 있는 수학적 힘을 쉽게 예증해 주지만, 다만 그 예시일 뿐이다.

비록 자동 계산에서의 주도적인 역할로부터 물러나 있지만, 튜링은 정신 과정들이 특정한 물리적 구현과는 독립적으로(=어떤 물리적 구현을 보이든지 상관없이) 논리 모형으로 올바르게 기술된다는 자신의 생각으로부터 퇴각한 것은 아니며, 따라서 생리적인 두뇌 말고도 다른 물리적 형태로 구현될 수가 있는 것이다. 수학의 독자보다는 철학의 독자들을 위해 씌어진 논문인 튜링(Turing 1950)에서는 특히 분명하고 구체적으로 논리 모형의 내용을 제시해 주었고, 가장 충실한 모습으로 튜링 생각의 발전을 전달해 준다. 춰어취의 논제가 때로 튜링의 동시 기여를 인정함으로써 '춰어취-튜링' 논제로 기술되지만, 다소 다른 모습을 지니고 있는 독특한 튜링 논제를 우리는 다음처럼 더 정확히 표현할 수 있다. 이산 상태의 기계 모형은 물질세계(=유물주의)의 한 측면, 곧 두뇌들의 작동에 대하여 타당한 기술이다.

28) 루틴(routine 경로)은 컴퓨터에서 특정한 작업을 실행하기 위한 일련의 지시[명령]사항들을 일부 혹은 전부를 가리킨다. 컴퓨터 용어 사전에는 50개나 되는 루틴들이 소개되어 있다.

29) 아마 'blind'는 그 기계의 구현 여부도 아직 확증되지 않았던 때이므로, 눈을 감은 채 상상으로만 그 기계의 운용을 언급해야 하는 처지를 가리키는 듯하다.

30) 특히 프로그래밍 언어로서 C나 C++에서는 프로그램 줄(행)이 전개 및 수정에 중요한 역할을 한다.

31) 전자적 지연 기억장치 자동 계산기(Electronic Delay Storage Automatic Calculator)의 약자임.

이런 견해 및 그 함의에 대하여 튜링은 견실하면서도 사실상 도발적인 방어를 하였다. 자신의 논제를 가능한 대로 멀리 확장하면서, 그는 새로운 논제와 논의를 열어나갔다. '생각하기'와 '지능'에 대한 그의 지속적인 논의는 언제나 유관한 것으로 간주된 내용의 범위를 넓혀 놓는 경향이 있었다. 1936년에 그의 논의는 흐름도(algorithms)를 실행하는 일에 초점이 모아졌다. 1946~1948년 업적에서는 장기 놀이하기(대부분 전쟁 때 작업에서 논의가 됨)가 지능에 대한 그의 모범 사례가 되었다. 성공적인 장기 놀이 기계는 결코 자신에게 명시적으로 제공되지 않은 흐름도를 개발해 나가야 할 것이라는 점이 주요한 초점이었다. 튜링(1950)에서 그 논의는 '지능을 갖춘 기계'의 궁극적인 성공이 사람들과 일반적인 대화를 해 나가는 훨씬 더 야심찬 과제로 바뀌어 제시되었다.

사실상 튜링은 '지능'으로 불릴 법한 내용을 놓고서 분명히 금을 그어 놓는 문제를 다룬 논의를 통해서 (=지능과 비-지능을 구분해 놓음으로써) 이런 그림으로 이끌어 갔다. 그의 논문에서 불경스럽기로 이름난 시작 부분은 '모방 놀이'를 기술한다. 거기에서 한 사람의 '질문자'가 안 보이는 한 명의 남자와 한 명의 여자에게 모두 질문을 던지고, 씌어진 응답을 근거로 하여 누가 여자인지를 결정한다(실제로는 원격 출력기[teleprinter]로 의사소통을 한다). 남자와 여자는 둘 모두 자신이 여자라고 주장을 하게 된다. 그러고 나서 튜링은 동일한 조건 아래에서 사람과 기계가 서로 각자 경쟁적으로 자신이 사람임을 주장하는 비슷한 놀이를 모방하는 데로 나아간다. 사실상, 겉으로 드러난 유추는 기묘하게도 부정확하며, 튜링이 실제로 이해시키려고 하는 생각으로부터 주의를 딴 곳으로 돌려놓는다. 첫 번째 놀이에서 성공적인 모방은 전혀 아무런 것도 증명해 주지 않는다. 우리는 신체상의 남녀 성별이 원격 출력기 응답에 의해서 결정되는 것이 아님을 알고 있다. 그렇지만 두 번째 놀이에서 튜링 논의의 가장 중요한 요체는, 원격 출력기 내용으로 된 지능에 대한 성공적인 모방이 뭔가를 증명해 준다는 점이다. 왜냐하면 그것(=모방)이 지능이 있기 때문이다. 성별이나 신체상의 강건함이나 다른 속성들과 대립되는 것으로서, 지능은 효과적으로 이산 기호들의 의사소통에 의해서 분명해질 수 있는 것으로 정의된다. 튜링이 언급하듯이, 그는 정신 능력 및 '미인 뽑기 모임에서 이채를 띠는 일'이나 '비행기 경주'와 같은 다른 능력 사이에 구분되는 금을 그어 놓고자 하였다. 이는 '인간의 신체 능력 및 지적 능력 사이에서 아주 분명한 금을 그어 놓는 일'로서, 그 놀이의 조건들에 의해서 달성되었다.

그러고 나서 튜링(Turing 1950)은 기계가 그런 지능을 습득하도록 이끌어갈 법한 수단들을 제안한다. '교사'로부터 '보상'이나 '벌'을 받게 될 튜링 기계 프로그램의 무작위 돌연변이를 전적으로 허용하는 것에 의해서 발전하는 학습을 기대하기보다는, 그는 오히려 명백한 프로그래밍에 의해서 규칙 체계를 제공해 주는 것이 필요할 것이라는 견해를 취한다. 그러나 정신에 대한 이산 상태 기계 모형 속에서 진행되는 방법에 대한 이들 제안의 이면에는, 바로 그 모형의 타당성에 영향을 미치는 "아주 분명한 금긋기"(fairly sharp line)이라는 구절에 의해 튜링이 제기해 놓은 문제가 놓여 있다. 과연 지적 능력 및 신체 능력 사이에서 그 금이 분명하게 그어질 수 있을 것인가? 그 논문에서 튜링에 의해 재치 있게 뒤섞이어 꾸며진 인간 및 기계 사이에 있는 가상적인 대화는, 외부 세계와 관련된 일상 언어를 채택하고 있다. 그러나 인간 두뇌는 그 세계와 관련되는 감각 및 행위의 복잡한 상호접속을 통해서 그런 언어를 배우고 응용할 듯하다. 그렇다면 그에 상응하는 상호접속이 없이도, 과연 기계가 언어를 배울 수 있을까? 외부 세계는 과연 튜링 기계에 의해서 판독된 기호들로 실패 없이 환원될 수 있는 것일까? 헬렌 켈러를 언급하면서, 결론에서 이런 질문이 중요하지 않음을 시사한 튜링은, 결국 "그 기계에 돈으로 살 수 있을 최상의 감각 기관들을 제공해 주는 것이" 최선이 될 것으로 시사한다. 경험의 문제, 다시 말하여 인간적 '지능'을 지닌 물체가 또한 정신세계 및 물질세계 사이에 있는 인간적인 연결도 지녀야 할 필요가 있는지 여부는, 집필자가 생각하기에 튜링 논제의 타당성에 대하여 핵심 물음인 듯하다. 그 물음은 튜링의 풍부하고 생생하며 인습 타파적 탐구에 의해서 제기만 되었을 뿐이지, 아직 답변되지 않은 질문이며, 결코 사고의 안전구역 속에 갇혀 있는 것이 아니다.

관련이 있지만 이상하게도 튜링이 취약한 언급으로만 남겨둔 또다른 물음이 있다. 계산 가능성에 대한 절대적 한계가 정신 및 육체(심신)의 문제와 조금이라고 관련이 있는지 여부에 대한 것이다. 기묘하리만큼 그는 이산 상태의 기계 모형 안에서 절대적으로 결정 불가능한 물음들에 대한 그 자신의 발견에 전혀 관심이 없었고, 대신 보편 기계의 힘을 강조하는 데에 훨씬 더 많은 관심을 쏟고 있었다. 예를 들면, 정신적 직관(mental intuition 심적 통찰력)을 갖는 도약이 무한 공리계를 생성하는 데에 포함된 비-계산적 단계들과 상응(일치)하는 것으로 말할 수 있는지 여부를 놓고서, 자신의 1937년 생각들에 대하여 더 뒤에서는 전혀 언급해 놓지 않았다.

그러나 튜링은 실제로 물질세계에서(=유물론에서) 논리적인 기계 모형의 구현에 관련되는 다른 물음들을 기술해 놓았다. 심지어 '계산 가능한 숫자들'(1936: 8절의 각주)에서도 유한 숫자의 허용된 기호들을 물질적으로 가능한 기호의 연속된 무한과 관련짓는 주석을 달아 놓았다. '기호 공간'의 위상론은 다시 튜링(Truing 1947)에서 기묘하게 퇴보하는 것으로 보이는데, 기억장치를 개발하고 있던 당시에 음향 지연 회선에서[32] 전기 흐름(pulses)을 재순환시키는 일의 물리적 결과만을 기술하고 있다. 관련 논의는 1947년 1월 하버드 대학 학술회의(Harvard 1948)에서 정전기(electrostatic) 기억장치의[33] 물리적 원리들에 대한 그의 논의 속에 들어 있다. 두 경우에서 모두 그는 이산 구조가 유지되는 데에 필요한 물리적 조건들만 논의하였다. 튜링(Turing 1948)에서는 보편 기계 개념을 물질적(=유물론적) 형태로 번역하는 데에 어떻게 열역학적 고려사항들이 영향을 주는지에 대한 논의를 제시해 준다.

튜링(Turing 1950)에서는 이산 상태 기계 모형을 다음처럼 소개하였다. "예를 들면, 점화 체계를 위한 스위치를 고려하는 경우, 스위치가 각각 명백히 켜져 있거나 아니면 명백히 꺼져 있어야 한다는 점은 간편한 상상이다. 분명히 중간 위치들이 있겠지만, 대부분의 목적을 위해서 우리는 그것들을 잊어버릴 수 있다." 이는 이산 상태의 기계로써 물리적 체계를 모형으로 만드는 일의 타당성에 대한 신중한 진술이었다. 따라서 이 논문 이후로부터는, 신경계의 실제적인 물리적 연속성을 튜링이 자신의 논제에 대한 심각한 결함으로 상정하였다. 그는 이산 기계가 확률 요소(random elements)들로써 연속성의 효과를 모방할 수 있을 것이라는 논의로써 그 결함을 방어하였다. 이는 다소 취약한 논의이다. 한 가지 이유로는 양자 역학의 불확실성에 대한 물음을 언급하는 데 실패하였기 때문이다. 또다른 이유로는 심지어 고전 물리학 내에서조차 작은 초기 변이의 효과들이 자의적으로 증폭되어 커지며, 이산 상태의 기계가 그런 증폭 효과를 올바르게 본뜰 수 있는지가 분명하지 않기 때문이다. 실제로, 동일한 논문의 다른 부분에서, 튜링은 사실상 이산 상태 기계의 예측 가능성 및 고전 물리학에서 예측을 얻어내는 일의 불가능성 사이에 있는 현저한 차이점에 주의를 기울였다. "어느 순간 1조 분의 1㎝ 정도 단 하나의 전자를 이동시키면, 1년 뒤에 한 사람을 눈사태에 의해 죽게 할 수도 있고, 그 화를 면하게 할 수도 있는 그 차이를 만들 가능성"이 얼마만큼 있겠기 때문이다.[34] 따라서 우리는 논리적인 기계 구조에 의해서 두뇌의 정신 기능을 모형으로 만들어 놓는 일로써 그가 의미했던 바에 대하여 튜링이 완벽한 이론에 이르렀다고 느끼지는 않는다. 그렇지만 그런 물음들을 조금이라도 제기한 것에 대해서는 튜링에게 알맞은 영예를 안겨 주어야 한다.

1949년에 튜링은 또한 소거법(cancellation)으로써 반군(semi-group)에 대한 낱말 문제의 비-가해성(unsolvability 해를 가질 수 없는 속성)을 보여 줌으로써 수학에 새롭고 중요한 기여를 이룩하였다(Turing 1950a). 그런 작업은 만일 그가 바랐었더라면 고전 수학의 심장부로 되돌아가는 길을 제공해 주었을 것이다. 그러나 튜링이 비록 흥미를 갖고서 그 발전을 지켜보았지만, 그것은 따라가지 않은 길이었다. 그의 마지막 논문(Truing 1954)은 그런 문제점들의 중요성을 설명해 주고, 새로운 발전 내용들

32) 대은출판사 1992《대은 컴퓨터 용어 대사전》에는 지연 회선(delay line)과 지연선 기억장치(delay line storage) 항목이 들어 있다.

지연 회선: 재료나 회로 매개 변수, 기계적 장치 등의 특성으로 인해 입력 및 출력 사이의 정보 흐름을 지연시키는 장치이다. 크게 두 종류가 있는데, 전기적 신호의 전반 시간을 이용한 전자 지연선, 전기적 신호를 초음파로 변환하여 전반 시간을 이용하는 초음파 지연선이다. 지연선의 재료 매체로는 음성이 점진적으로 전달되는 수은이나 동축 케이블, 전송선 등을 쓴다.

지연선 기억장치: 지연선을 이용한 기억장치이다. 일정한 시간 간격으로 지연을 만드는 회로와 증폭 정형을 하는 회로를 조합하여 폐쇄 회로를 만들고서 여기에 자료를 순환시켜 기억시키는 장치이다. 시간 지연을 만드는 수단으로서 ① 초음파의 전도를 이용하는 것, ② 자기 드럼을 이용하는 것, ③ 수동 전기 회로를 이용하는 것이 있다. 순환 기억장치라고도 부른다.

33) 대은출판사 1992《대은 컴퓨터 용어 대사전》에 보면, 자료를 표현하기 위하여 정전기 전하를 이용하는 기억 장치라고 함.

34) 이는 비선형 고전 역학의 한 사례로서 번역본 6쪽에 있으며, 앞뒤 맥락에서는 이런 일이 이산 상태의 기계에서는 일어날 수 없음을 지적하고 있다. 번역자가 보기에, 튜링은 이산 상태 기계에서는 이런 효과가 성립될 수 없다고 생각하였지만, 인용자는 정신 모형도 그런 가능성이 있는 것으로 간주하는 듯하다.

아마 오늘날 혼돈 이론의 하나로서 거론되는 '나비 효과'(Butterfly Effect), 즉 미미한 한 사건이 마침내 엄청난 결과를 빚는 경우와 동일하지 않을까 생각된다. MIT 대학에서 기상학을 연구하던 에드워드 로렌츠는 컴퓨터 모의실험에서 미미한 수치가 자신이 세운 기상 예측 방정식에서 엄청난 결과를 갖는 수치로 구현되었다. 이는 어떤 중심점에서 운동을 점차 너른 범위로 일정하게 이끌고 나가는 것으로, 그 중심점에 나비 모습의 'Strange Attractor'(기묘한 끌개)가 있다고 보았다. 혼히 브라질 아마존에 날고 있는 한 마리의 나비가 1년 뒤에 미국의 플로리다에 엄청난 허리케인으로 돌변할 수 있다고 예를 든다. 아마 시간 축에 따른 "거대 증폭 효과"로도 부를 법한 이 효과는, 특히 경제나 경영에 적용되어 아무리 사소하게 보이는 변인이라도 소홀히 다루지 말도록 하는 교훈으로 받아들여지고 있다. 번역자는 이 효과가 필연성 토대 위에 있는지, 아니면 가능성이 있는 우연으로 취급되고 있는지에 대하여 명시적인 언급이 없어 아쉽다. 반드시 나비 효과의 적용 대상과 범위가 아울러 언급되어야 할 것이다.

을 조사한 일반적인 논문이었다. 비록 전기적 계산에 대한 선구적인 이용에서 맨체스터 대학의 기계를 리이만 제타 함수(Riemann zeta-function) 계산에 응용하였지만(Turing 1953), 수 이론으로 다시 돌아가지는 않았던 것이다. 전쟁 이전과 같이, 그의 주된 관심사는 주로 수학적이라기보다는 과학적인 것들이다.[35] 대신 그는 자신의 초기 업적과 무관해 보이는 주제로 관심을 돌렸다. 형태 발생학(morphogenesis)에 대한 수학적 모형 만들기이다. 여기서 다시 그는 완벽한 국외자로서 안으로 뛰어들어 중요한 발견을 이룩하는 경험을 가졌다. 그의 화학적 모형은 공간상으로 대칭적이지만 불안정한 초기 조건으로부터 이산 공간 형태를 만들어 내는 속성을 지닌 비선형 미분 방정식으로 이끌어 갔다(Turing 1952). 튜링은 확실히 이 작업을 유물주의 설명의 한계를 넘어선 것으로 선전되기 일쑤인 현상들을 놓고서 설명해 냄으로써 "설계로부터 나온 (공상적) 논의"를 맞받아치는 공격으로 간주하였다. 그렇듯 우리는 이를 그의 초기 관심에 대한 다른 도구들로써 인생에서 가장 신비로운 대상들을 과학적으로 설명할 수 있는 구조와 관련지어 놓는 지속적인 모습으로 볼 수 있다. 그의 초기 업적들과의 좀 더 구체적인 관련은 두뇌에서 신경 연결들이 물질적으로 어떻게 만들어지는지에 대한 물음을 놓고서 그 자신이 언급 내용에 의해 드러난다(하쿄즈 1983에 인용된 영 J.Z. Young에게 보낸 편지임). 그러나 아마 우리는 또다른 연관도 볼 수 있을 것인데, 대칭성의 깨어짐이 이산 구조가 물질적 연속체로부터 생겨나는 수단을 제공해 주는 것이다. 어떤 경우이든지, 이는 수리 생물학에서뿐만 아니라, 비선형 고전 역학의 현대적 연구(앞에 있는 튜링의 눈사태에 대한 언급에서 암시된 주제임)에서도 토대를 이루는 논문이었다. 다시 한 번 튜링은 자신이 가정한 비선형계에 대한 수치적 모의를 위하여 컴퓨터를 이용함으로써 컴퓨터의 창조적인 이용을 선구적으로 보여 주었다.

1954년에 튜링은 양자 역학적 물리학의 토대를 검토하고 있었다. 이는 에딩튼이 더 일찍이 자신의 관심을 엄청나게 많이 쏟아 넣었던 주제이지만, 튜링이 논리 기계 모형에서 예측 가능성의 논의에서는 다소 기이하게도 누락시켜 놓았다. 특히 그는 연속체 상태를 관찰 가능한 값들로 된 이산 스펙트럼으로 아주 신비스럽게 "환원하는" 과정을 연구하는 데 매료되어 있었다. 튜링은 양자 역학의 표준 설명법이 그런 관찰이 얼마만큼 자주 일어날 것으로 가정하는지에 대한 지침을 전혀 전달해 주지 못함을 지적하면서, 양자 역학 상태를 지속적인 관찰과 환원 아래에 통제되도록 해 놓는 일이 역동적인 진화를 막아버리는 효과를 지닌다는 사실에 주목하였다. 그는 비선형 속성이 양자 역학 속으로 들어가야 한다는 생각을 피력하였다(Gandy 1954). (바람을 가져온 이래 죽 다른 사람들과 같이) 그가 자신의 화학적 대칭성 깨어짐과 비슷하게 보이는 비선형 과정에 의한 환원을 모형으로 만들고자 하였던 것일까? 그가 지금에라도 고전 및 양자 물리학의 근원적인 물음들을 튜링 기계 모형의 이산 결정론과 관련지으려고 시도할 것 같은가? 1954년 6월 7일 튜링의 죽음으로 그의 탐구가 막을 내렸기 때문에 우리는 다만 짐작만 할 수 있다.

튜링의 자살이 단순히 그의 시도로만 탓할 수 없으며, (당시 영국에서는 법적으로 완전히 불법이었던) 동성연애에 대한 호르몬 주사 치료를 강요받았다. 이런 사건들이 두 해 일찍 1952년에도 일어났었지만, 그들이 부끄럽게 생각하지도 않았고, 그를 을러 협박하지도 않았다. 그렇지만 튜링의 위치는 독특하였다. 그의 업적은 제 2차 세계 대전에서 극도로 비밀스런 영국과 미국 첨보 활동에서 중심에 있었으며, 사실상 1948년에 그런 작업을 재개하였다(체포된 뒤에는 그 일이 중단되었다). 1953년에 튜링은 경찰의 감시를 포함하여 똑같이 불안한 "위기"에 대해 암호 상으로 언급하였다. 필자의 생각에, 핵심 질문은 그가 도덕적으로 불가능한 위치에 있다고 느끼게 되었는지 여부, 개인의 자유에 대한 요구와 국가에 대한 충성이 화합할 수 없었는지 여부이다.

앨런 튜링은 대답을 하나의 논제로 제시했다. 정신세계를 이산 상태의 기계와 동일시하는 것이다. 그것은 현재 인공지능에서 연구가 진행되고 있듯이, 실용적인 방식으로 탐구될 수 있는 엄청나게 중요한 특성을 지닌 논제이다. 이것 하나만으로도 그 토대 논문의 50주기를 기념하는 게 마땅하다. 그렇지만 또한 열린 질문들도 많이 남겨 놓았다. 논리학과 물리학, 그리고 인간이라는 육체적 환경의 연구들 사이 여러 영역에 두루 걸쳐 있는 어려운 물음들이다. 1936년 앨런 튜링의 위대한 업적은, 심층 물음들을 완벽히 진지하게 생각하면서 비롯된 참신한 사고가, 생각에 대한 우리의 문화적 구분들을 어떻게 초월하고 바꿀 수 있는지를 보여 준다. 50년 세월은 그 본보기의 광채를 흐려 놓지 못했다. 그의 삶은 물론 그의 업적도 인간 지능의 특성을 붙잡는 데 우리가 대면하는 극단적인 어려움을 잘 보여준다. (끝)

35) 이 글 필자가 수학자이므로, 완고하리만큼 이런 관심을 부정적으로 보고 있는 듯하다. 그렇지만 수학자도 인간이며, 인간에 관한 관심을 두루 끌어안는 게 마땅하다. 또 수학이 모든 학문의 어머니라면, 오히려 이런 관심 확대가 자연스러운 것이다. 수학이 좁은 세계에만 갇혀 있다면 그야말로 '반피어'에 지나지 않을 것이다.