튜링기계-1

이정모(편), [인지심리학의 제 문제(I): 인지과학적 연관]. 12장. pp. 265-283.

마음은 기계인가?: 튜링기계와 괴델 정리

이 정 모

과학은 역사적으로 그 연구대상에 대한 보다 효율적이고 엄밀하고 단순한 설명을 부여하려고 노력해왔다. 이러한 노력에서 자연현상에 대한 가장 엄밀하고 단순한 설명으로 대두된것이 기계론적(mechanistic) 설명이었다. 기계론은 처음에는 물리학에서 형성되고 이것이 화학, 생물학, 생리학을 거쳐(Dijksterhuis, 1961) 심리학에 도입되었다(Cohen, 1980). 17세기 유럽에서는 공학의 발달과 Galileo와 Newton의 물리학이 대두됨에 따라,우주를 기계론적으로 보려는 입장이 과학자들에게 널리 퍼졌다. 이 기계론적 입장은 물리적 자연현상이 기계적 결정론(Mechanistic Determinism)에 따라 일어나며 법칙적이고 예언 가능하며 관찰,실험, 측정가능하다는 생각으로 형성되었다.

뒤를 이어 이러한 입장을 인간에게도 적용하여 설명하려는 시도가 나타났고 Descartes와 La Mettrie등의 기계론적 견해가 인간의 신체와 마음의 과정에 대하여 제기되었다(이정모,1984). 17, 18세기의 대륙의 기계론과 그 이후의 기계론은 크게 둘로 나누어 볼수 있다. 하나는 Descartes와 Malebranche등과 같이 심신 이원론을 전제하고 신체는 기계이나 정신은 기계가 아니라는 입장이었으며 다른 하나는 Gassendi나 La Mettrie의 입장처럼 신체 뿐 만 아니라 심리적 과정(마음)도 기계적이라는 입장이다.

인간의 신체를 하나의 기계로 보는데에는 이의가 거의 없어 왔다. 하지만 마음 또는 심적 과정을 기계적이라고 볼 수 있느냐에 대해서는 La Mettrie의 "L'Home Machine"이래 계속적으로 논쟁이 되어져 왔다. 현재에 이르러서는 마음의 본질을 기계의 전형인 컴퓨터에 유추하여 설명하려는 정보처리적 접근과 이에 대립되는 견해가 심리철학과 인지과학에서 첨예화된 문제로 논의 되고 있다(Dennett, 1986). 본 논문에서는 '마음이 기계적인가'하는 문제가 인지과학에서 부상하게 되는 데에 결정적 출발점이 된 Allan Turing과 Kurt Gödel의 입장을 중심으로 정리해보고, 인간의 마음을 하나의 Turing기계로 간주할 수 있는가에 대한 논의를 전개하고자 한다.

일반적으로 기계론적인 심리학 이론의 입장은, 마음이 물질인 두뇌의 산물이며 두뇌란 수 많은 세포체들이 모여 이루어지며 이들 세포체들은, Arbib(1964)등이 기술했듯이, 기계적 법칙에 의해 상호작용하며 그 상호작

용의 결과로 나타나는 것이 마음의 내용이라는 것이다. 즉 마음이란 두뇌의 작용을 반영하며 두뇌의 작용이 기계적인 한에서 마음의 작용도 기계적이며 따라서 마음은 하나의 기계로 간주할수 있다는 것이다. 또한 기계의 정의를 '유한히 기술할수 있는 현상'(Jackson, 1985)과 같은 의미로 규정한다면, '기계적'이란 유한히 기술될 수 있다는 의미, 또는 알고리즘에 의해 효율적으로 규정될 수 있다는 의미가 된다(Cutland, 1980). 따라서 기계론에서 '마음이 기계적이다'라는 논지는 마음의 작용을 효율적으로 알고리즘에 의하여 유한 기술할 수 있다는 의미가 된다.

마음에 대한 이러한 입장의 기계론은 20세기 이전에는 다분히 사변적이며 엄밀한 개념규정이 결여되었고 체계화되지 못했다. 이러한 결함이 1930년대를 기점으로하여 A.Turing을 비롯한 수학자들에 의해 가다듬어져 새로운 기계론이 제기되었고 또 기계론에 대한 체계화된 형식적 반론도 나타나게 되었다.

1. Turing의 기계론

Turing의 기계론을 설명하기에 앞서서 설명되어야할 것은 현상에 대한 수학적 기술이론(theory of mathematical description)이다. 수학적 기술이론에 의하면 과학적 방법은, 본질적으로, 자연현상을 수학적으로 기술하는 기술 유형의 하나를 선택하는 방법이라고 볼 수 있다.

자연현상은 무한할 수도 있고 유한할 수도 있다. 현상이 유한하면서 수학적으로 유한히 기술될 수도 있고 무한 하면서 유한히 기술될 수도 있다. 그러나 무한히 기술될수는 없을 것이다. 자연현상을 수학적으로 기술하려는 입장의 핵심은 무한하거나 유한한 현상을 유한히 기술할 수 있다고 가정하는 것이다.

현상을 기술하기 위하여는 현상의 발생에 대한 관찰이 선행되어야 한다. 현상이나 과정이란 현실에서 발생하는 사건들이며 발생이란 '일련의 상황들'로 간주할수 있으며 현상이란 '상황들이 발생할 수 있는 모든 가능한 양식의 집합'으로 규정될 수 있을 것이다. 이는 한 현상은 다른 현상으로 구성될수 있음을 의미한다. 따라서 현상에 대한 수학적 기술의 첫 단계는 그 현상의 모든 가능한 상황의 집합 X와 시간 척도 T를 규정하는 것이다. 상황 집합을 구성하는 요소 상황을 θi, 시간 척도 T내의 한 요소 시간을 Ti, 현상의 발생을 θ라고 한다면 발생 θ는 Ti와 θi를 상응시켜 주는 함수라고 할 수 있으며, 현상이란 곧 θi의 집합, 즉 <θ1, θ2, θ3... θi>이라 할 수 있다. 따라서 어떤 현상의 기술이란 발생 θ의 집합을 기술하는 것이다. 현상에 대한 완벽한 기술이란 발생θ의 가능한 무한 집합을 기술하는 것이며,유한 기술을 그 근본으로 하는 수학적 기술이란 그 현상의 발생집합을 유한하게 기술하는 것이다. 이는 모든 발생θi들을 열거하거나 나열하는 것이라기 보다는 , 이θi들을 임의의 정확수준 정도로 계산 할수있게하는 유한 규칙 또는 함수를 위한 기술로서 받아들임을 의미

한다. 이때의 계산이란 '정확히 효율적으로 규정할수 있다'는 의미가 된다.

확률이론에 의하면 연속적 현상을 비 연속적 현상을 사용하여 임의의 정도까지 추정하고 기술할 수 있다. 이는 연속함수 현상을 이산(비연속)함수인 계단 함수로 기술하는 것이며, 후자는 다시 상징의 계열이나 상징의 집합 단위인 스트링(String)으로 기술할 수 있음을 의미한다. 환원하여, 연속적 현상은 비연속적 현상으로 기술할 수 있으며, 이 비 연속적 현상의 발생은 유한히 기술할 수 있고, 이는 특정 상징 집합단위 또는 스트링의 집합을 유한히 기술하기만 하면 된다고 하겠다. 기술하고자 하는 현상의 발생 집합이 유한 하다면 과학자는 그저 그에 해당하는 모든 스트링을 열거하면 된다.

그 집합이 무한 하다면 그 모든 경우에 대해 제 i번째 스트링을 계산해 낼수 있는 규칙(유한히 효율적으로 기술할수 있다는 규칙), 또는 함수를 찾아내면 된다.

그렇다면 현상에 대한 기술의 개념이 무한 기술의 필요성에서부터 유한 기술로, 유한 기술에서 함수들 또는 규칙 집합(rule-sets)의 기술의 개념으로 옮아간다. 그런데 모든 비연속적 현상은 특정 스트링에 자연수를 연결시켜주는 함수에 의해 표상될 수 있다. 그리고 어떤 자연수도 스트링에 의해 표상될 수 있다(예: 자연수 3을 0011로 표상하거나 C로 표상할 수 있다). 따라서 결론적으로 말하자면, 현상의 기술의 문제란, 함수 또는 규칙집합에 대한 유한 기술의 문제, 다시 말하여 한 스트링 집합과 다른 스트링 집합을 대응(mapping)시켜주는 문제이다. 그런데 한 스트링 집합과 다른 스트링집합을 대응시키는 함수을 다루는 이론이 자동기계(automata)이론이며, 자동기계이론 중에 가장 대표적인 이론으로서 계산가능성(computability)의 개념을 명확히 해주고 심리 현상의 기술에 까지 적용된 이론이 '튜링'기계이론이다. Allan Turing이 발전시킨 '튜링기계이론'과 '마음에 대한 기계론'을 정리하면 다음과 같다.

사람들은 각종의 논리적 추론을 하며 창의적 사고를 한다. 또한 수학자들은 각종 수학적 문제를 제기하고 이에 대하여 연역적으로 추론하여 문제를 해결한다. 그들은 어떤 한 명제의 진위를 증명을 통하여 밝힌다. 이러한 추론적 사고 , 수학적 추론은 어떻게 이루어지는 것일까? 이러한 사고 과정을 이어가는 절차들을 명확히 설명할수는 없을까? 즉 어떤 완전한 형식적(formal) 틀을 갖춘 일관성있는 공리적 체계에 의해 이 과정들을 단계 단계 정확하고 엄밀하게 기술 할수는 없을까? 이러한 물음은 옛부터 수학자, 논리학자, 철학자, 심리학자들이 그리고 최근에는 인공지능학자들이 관심을 가져온 문제이다. 각종 수학적 문제의 제기와 증명의 추론과정을 완전히 형식화된 어떤 공리와 규칙의 체계에 의해 설명 할 수 있다면, 그것이 지니는 의의는 큰 것이다.

이러한 물음에 대하여 수학자들은 기계론적 입장을 취한다. 그들에 의하면 수학적 추론이란 기계적이다. 수학적 개념, 명제를 형식화할 수 있으며 주어진 비공식적 증명을 어떤 형식체계내에서 점검할수있는 형태로 형

식화 할수있으며, 수학적 증명 방법을 그 형식체계내의 잘 규정된 절차로 환원시킬 수 있다는 것이다. 그런데 이러한 형식체계는 기계로 간주할 수 있으며 이 체계 내에서의 정리와 산출과정은 기계적 조작(operation)으로 간주할 수 있다. 그러므로 모든(수학적) 추론은 기계화, 형식화할 수 있다. 동시에 이를 확대하여 해석하면 수학적 추론의 상위 체계인 인간의 마음도 기계로 간주할 수 있으며, 마음의 작용도 기계화, 형식화할 수 있다고 주장될 수 있다. 이러한 기계론을 체계적으로 강력히 전개한 것이 Allan Turing 등이었으며, 이에 반대되는 입장으로 해석된 것이 Kurt Gödel정리의 확대해석이다. Turing 등은 형식체계와 기계와 인간의 마음에 대해 다음과 같은 주장을 전개한다.

일반적으로 형식체계란 일련의 상징요소들 (알파벳)의 집합과 이들을 구성하고 변환시키는 명백하고 유한한 알고리즘 규칙 또는 절차로 구성된다. 알고리즘 또는 효율적 절차(effective procedures)란 어떤 수리적 조작을 수행하는 기계적 규칙 또는 자동적 방법 또는 프로그램을 의미한다. 효율적 절차란 그 절차가 단계적으로 수행되면 일정한 유한 수의 단계를 거친 후에 출력이 나온다는 의미를 가진다. 형식체계란 유한히 기술될 수 있는 체계를 의미한다. 유한히 기술 될수 있다라는 것은 한 형식체계내에서 ,알고리즘적 유한 절차들을 기계적으로 적용하여, 주어진 상징 스트링이 규칙에 잘 맞는가(well-formed) 또는 공리인가를 결정할 수 있고, 또 어떤 진술이 규칙에 잘 맞는 유한 진술 집합에서 도출될 수 있는가를 결정할 수 있다는 것이다.

이러한 형식체계 T의 알고리즘적 규칙들만 주어진다면, 체계 T의 정리들을 기계적으로 하나씩 산출하여 열거할 수 있는 기계T_m을 구성할수 있다.

마찬가지로,무한한 기억을 지닐 수 있는 컴퓨터가 있다면, 그 기계의 출력이 형식체계 T_m의 정리들과 동일한, 어떤 특정 형식체계 T_j를 발견할 수 있을 것이다. 그렇다면 어떤 형식체계도 기계로 간주할 수 있으며 역으로 어떤 기계도 형식체계로 간주 할 수 있다는 정리를 세울수 있다.

부연하여 설명하자면, 기계는 명백한 일련의 규칙에 의해 일련의 조작을 수행하는 기구이다. 기계체계에 내장된 조작의 유형과 기본가정이 일정한 수라면 우리는 이들 모두를 적절한 상징으로 표상하여 종이에 적어 볼 수 있다. 최초의 기본 가정들은 공리라든가 계(系, corollary)와 같은 기본 공식으로 표상될 수 있으며, 하나의 조작은 조작이 일어나기 전 상태와 후의 상태를 나타내는 공식과 어떠한 규칙이 적용되었는가를 명시하므로써 표상될 수 있다. 기계의 조작들이 아무리 많고 복잡하더라도 충분한 시간만 주어진다면 이러한 조작계열의 아날로그를 기록할 수 있다. 그리고 이러한 아날로그는 형식적 증명이 된다. 즉 기계의 조작 하나 하나가 규칙의 적용에 의해 표상된다. 또 일정상황에서 기계가 어떤 조작을 수행할 것인가 여부를 결정하는 조건들은, 이 표상에서 일정 공식에 어떤 규칙이 적용될 수 있는가 여부를 결정해 주는 조건, 즉 적용성의 형식적 조건(formal

conditions of applicability)이 된다. 이러한 규칙을 추론의 규칙으로 간주함으로써, 하나 하나의 공식이 이전의 공식 또는 공식들에 어떤 형식적 추론 규칙을 적용하여 도출되는 증명계열을 획득 할 수 있다. 따라서 한 기계가 산출해 낼 수 있는 조작 계열, 즉 결론들이란 그 기계와 대응하는 형식체계내에서 증명될 수 있는 정리와 상응한다. 즉 기계의 조작 결과인 출력은 한 형식체계에서 도출된 정리에 해당한다는 것이다.

일반적으로 기계란 유한 자동기계(finite automaton)를 칭한다(Arbib, 1964; McNaughton, 1982). 유한 자동기계란 어떤 유한 수의 입력(input)을 받아들일 수 있고 유한 수의 내적 상태를 지니고 있으며 어떤 유한 수의 출력을 내어놓을 수 있는 체계이다. 유한 자동기계를 A, 유한 수의 입력을 I, 유한 수의 내적상태를 q, 유한 수의 출력을 O, q와 I가 상호작용하여 결정하는 A의 다음 상태($q \times I \dashrightarrow q_i$)의 함수를 λ, 다음 출력의($q \times I \dashrightarrow O$) 함수를 δ라 한다면,

$$\text{자동기계 } A = (I, O, q, \lambda, \delta)$$

의 형식으로 표현될 수 있다. 이러한 자동기계는 불연속적 시간 척도상에서 작용하는데, 만일 시간 t에서 상태 q에 있었고 입력 a를 받는다면, 시간 t+1에서는 상태$\lambda(q_i, a)$로 바뀌고 $\delta(q_i, a)$를 출력으로 내놓게 된다. 이러한 기계를 이산상태(discrete state) 기계라고 할 수 있다.

Turing(1936)은 이러한 이산상태기계의 논리를 근거로 튜링기계의 이론을 제시 하였다. 튜링기계란 디지탈 컴퓨터가 할 수 있는 계산을 모두 할 수 있는 유한자동기계이다. 튜링기계에는 입력기구인 테이프가 있고, 무한 용량의 기억고가 있으며, 출력을 테이프에 하고, 테이프 위의 상징들을 훑어서 읽고, 상징을 테이프에 인쇄하며, 또 테이프 위로 좌우로 움직이는 기구(D)가 있다. 튜링기계는 유한 수의 어떤 상태들 중의 한 상태에 처할 수 있다(그림1 참조). 입력기구인 테이프는 선형테이프로서 좌우 양방향으로 무한하며, 나뉘어져 있고 한 정방형은 공란이나 한 상징을 지닐 수 있다. 테이프는 이 자동기계가 매 순간 t에 정방형 하나만을 판독하여 하나의 일정한 상태에 있을 수 있도록 움직여진다.

** 그림 1 삽입 **

판독되는 정방형의 내용과 각 순간의 기계의 상태가 그 순간의 기계의 전체 형태(configuration)을 결정한다. 전체 형태란 한 순간의 기계의 상태와 입력된 정보와 이를 판독하고 있는 입력기의 부분으로 구성되며,유한 수의 내적 전체형태들이 있을 수 있다. 현재의 전체형태가 다음에 어떤 연산(조작, operation)을 할가를 결정한다. 연산은 판독되는 정방향의 내용을 바꾸거나, 그 정방향을 좌우로 움직이거나, 현재상태를 다른 상태로 바꾸거나, 정지하거나 한다. 정지한 때의 테이프의 내용을 출력이라 한다(그림 2 참조).

튜링기계-6

Turing기계는 알파벳의 유한 집합에서 추출된 상징들이 테이프에 주어지면 이를 하나씩 판독하고 현재의 상태를 점검하고, 기억내의 기계표(machine table)에서 이 두 조건이 규정하는 지시, 알고리즘, 추론 규칙을 조회하여 출력δ(q, a)를 내어 놓고 상태 λ로 옮겨가는 유한자동기계이다. 이 기계는 임의의 복잡한 계산을 또는 기계적 연산을 몇개의 단순한 기계적 연산들의 조합 또는 반복에 의해 수행 할수 있음을 보여준다. 이는 알고리즘의 복잡성이 질적으로 증가하는 것을 기억의 크기와 알고리즘을 수행하는 시간의 양적 증가로 대치한다. 튜링기계가 기계적인 까닭은 이 기계의 연산들이 본질적으로 순환함수라는 의미에서이다. 즉 효율적 알고리즘 절차가 있으며 순환적으로 셀 수 있는(recursively enumerable) 함수이기 때문이다.

순환적으로 셀 수 있다는 것은 , 어떤 상징 스트링들에 대해 각각이 특정 집합에 소속하는 지를 가려낼 결정 절차가 있으며, 이 스트링 집합과 다른 스트링 집합을 대응시켜주는 알고리즘이 있으며 계산함수에 의해 유한히 기술가능함을 의미한다. 튜링기계가 순환적으로 셀 수 있는 함수를 다룬다는 것은, 어떠한 함수이건 한 스트링 집합과 다른 스트링 집합을 대응시킬 수 있는 계산함수 모두를 튜링기계가 다룰 수 있다는 의미이다(McNaughton, 1982; Cutland, 1982; Jackson, 1985). 또한 어떠한 유한히 기술가능하며 계산가능한 함수도 부분순환함수로 표현할 수 있다는 정리에 따른다면, 어떠한 유한히 기술가능한 계산가능 함수도 튜링기계가 계산할 수 있다는 결론에 도달하게 된다.

「튜링」은 이에서 한걸음 더 나아가,보편 튜링기계(Universal Turing Machine) 정리를 제시하고 이를 증명하였다. 보편 튜링기계 정리란 어떠한 튜링기계 T_m에 대해서도 이를 모사(simulate)할 수 있는 보편 튜링기계 U_m이 존재한다는 것이다.

튜링기계란 실상은 하나의 절차 또는 그 집합이라고 볼 수 있다. 한 튜링기계 T_mTm은 그 기계표와 테이프 내용만 주어진다면 그림 2의 C에서 기술한 바와 같이 그 절차를 손으로도 모사할 수 있다. 이는 그 모사 절차가 알고리즘적이며 그 알고리즘적 절차가 다른 기계에 의해 수행될 수 있음을 의미한다. 이러한 알고리즘적 수행은 한 튜링기계 T_m의 전체 형태(configuration)를 새로운 상징들(예를 들어 0와 1의 조합)로 표상함에 의해서 이루어질 수 있다(이와 같이 스트링들에 다른 상징 또는 수를 부여하는 것을 Gödel화라고 한다).

어떤 튜링기계 T_m의 기계표에서 "(I_i, q_i, O_i, qj, Ii)"라는 내용, 즉 "IF 입력이 I_i이고, 현재상태가 q_i이면, THEN 출력은 O_i로 출력하고, 현재상태는 qj로 바꾸고, 다음에 테이프를 Ii로 옮기라는>" 내용 전체를 0과 1을 조합하여 표현할 수 있을 것이다(예: 01011). 이를 D_m이라고 하자. 보편튜링기계 U_m의 행동은, 한 튜링기계 T_m의 행동은, 한 튜링기계 T_m의 실제 현재상태 q_i와 입력내용 I_i를 확인한 후에 D_m에 따라 수행될수 있다. 즉 "If I_i q_i, D_m: THEN DO <O_i, Oi, Ii>"의 절차를 수행할 수 있을 것이다.

이러한 수행이란 인간이 튜링기계 T_m의 행동을 추적하는 것과 본질적으로 같은 절차에 의해 이루어진다고 하겠다. 그렇다면 보편 튜링기계란 D_m과 같은 상징내용에 의해 다른 튜링기계를 흉내낼 수 있는 기계라고 할 수 있다. 이는 보편 튜링기계가 일반목적적 디지탈 컴퓨터와 대등하다는 의미가 된다. 보편튜링기계에 주어진 D_m이란 스트링은 디지탈 컴퓨터에 넣는 프로그램으로 볼수 있고 이 프로그램이란 튜링기계 T_m의 한 스트링과 다른 스트링을 대응시키는 함수로 볼 수 있다. 그렇다면 이러한 프로그램을 지닌 컴퓨터란 유한히 기술할 수 있는 상징조작과정을 구현화 하는 보편 튜링기계라고 할 수 있게 된다.

보편 튜링기계가 존재한다는 것은 상당히 큰 의의를 지닌다. 덧셈하는 기계, 장기두는 기계, 문자 복사기계, 문제 해결 기계 등을 그 하는 일에 따라 서로 다른 기계를 따로 만들어야 할 필요가 없어지기 때문이다. 서로 다른 여러 일들을 하는 기계들을 모사할수 있는 하나의 보편튜링기계만 있으면 되기 때문이다. 각 튜링기계들의 행동이 유한히 기술될 수 있으며 보편 튜링기계가 각 튜링기계를 합한 기억 및 처리능력만 보유하고 있으면 되는 것이다. 또한 그러한 보편튜링기계로서 인간의 마음을 접근할 가능성이 시사될수 있다.

Turing은 보편튜링기계의 구성이론을 제시한 후, 더 나아가 마음을 기계로서 보는 기계론을 강력히 전개했다(Turing, 1950). 우리가 타인의 마음을 이해한다는 것은 타인의 외현적 행동을 보고 아는 것이다. 그런데 인간의 행동을 그대로 모사할수 있는 어떤 기계가 있을 수 있다. 더우기 인간의 행동이 상징으로써 표출된다고 했을 때, 그 상징 표출 조작과정을 충분히 흉내 내어 모사할 수 있는 기계가 있을 수 있다. 즉 인간의 행동을 그대로 모사 할수 있는 기계로써의 유한 자동기계가 있을 수 있다. 그런데 기계란 앞서 논의한 바에 의하여 형식체계이다. 그렇다면 인간도 상징을 받아들여 이에 대해 추론 규칙을 적용하여 출력을 내어 놓는 형식체계로 간주할 수 있으며, 형식 체계에 적용되는 논리와 법칙을 사용하여 인간의 마음을 설명할 수 있지 않는가? 그렇다면 수학적 직관을 포함한 인간의 모든 마음의 과정은 기계화, 형식화할 수 있으며 곧 인간은 기계 이상의 무엇을 할수 있는 체계가 아니라고 할 수 있다. 즉 '인간은 기계이다'라고 결론지을수 있다(Turing 자신은 '인간은 기계이다'라고 강력하게 주장하지는 않았지만 단지 '인간의 마음의 작용들을 모두 모사할수 있는 기계를 구성할 수 있다'는 약한 표현을 사용하였다).

Turing은 이러한 결론의 확인 방법으로 Turing Test라는 방법을 제시하였다. Turing Test란, 한 튜링기계와 한 인간 피험자와 한 질문자의 세 개체가 있으며, 이 세 개체 사이의 의사소통이 텔레타이프를 통해서만 이루어진다고 할 때,질문자가 자신의 질문에 대한 답이 튜링기계에서 나오는 것인지, 인간 피험자에게서 오는 지를 구분할 수 없다면 그 튜링기계는 튜링테스트를 통과한 것이며 그 기계는 인간과 같이 사고할 수 있다고 결론짓는 것이다.

이러한 튜링기계론과 튜링테스트의 개념은 인공지능학의 형성, 인지심리학에서의 정보처리관의 발전, 심리철학에서의 기계론의 대두등에 직접 간접으로 커다란 영향을 주었다.

튜링기계론에 대한 반론도 전개되었다. 「튜링」은 자신의 기계론과는 대립되는 주장을 제기할 수 있는 반론들을 9개 제시하고 이들을 차례로 공박하였다. 그중의 하나가 Kurt Gödel이 제시한 Gödel정리에 근거한 反기계론이었다. 이 Gödel정리는 마음이 튜링기계이다고 주장하는 입장에 대한 치명적 반론을 제기하였다.

2 . Gödel의 정리

수학자 Kurt Gödel은 기계적 절차(mechanical procedure)의 의미는 '튜링기계에 의해 수행되어질 수 있는 것'이라는 개념, 또는 '일반 순환함수'의 개념으로 규정한다는 것은 커다란 의의를 가진다고 보았다. 또한 형식체계는 정리산출을 위한 기계적 절차이상의 것이 아니며, 이러한 형식체계에 의하여 추론이 Turing기계 내의 기계적 조작들로 완전히 대체될 수 있다는 Turing의 주장을 Gödel은 높이 평가하였다. 그러나 그는 '완전히' 대체될 수 있다는 주장에는 동조하지 않았다. 또한 Turing기계이론을 확대해석하여 '마음은 기계이다' 또는 '모든 마음의 내용은 기계화, 형식화할 수 있다'는 주장에 대하여 반대의 입장을 제기하였다.

Gödel(1931)은 그 이전의 수학자들이 제기하였던 보편적 진리기계(Universal Truth Machine)가 존재할 수 없음을, 즉 수학적 진리에 대한 완전한 공리 집합이 있을 수 없음을 증명하였다. 어떤 기계이건 모든 해답을 결코 제시 할수 없다는 것이다. 이러한 증명은 그의 불완전 정리(Incomplete Theorem)의 제시에 의해 이루어 졌는데 그의 불완전 정리를 약술하자면 다음과 같다.

첫째로, 어떠한 형식체계일지라도 그 체계내에서 증명될 수 없는 공식 또는 명제가 존재하게 된다. 즉 어떤 명제 S가 있을 때 S도 또 S의 부정인 notS도 그 형식체계 내에서는 증명될 수 없는 명제가 있게 된다는 것이다. 예를 들어, '이 명제는 증명할 수 없다'는 명제는 참인지 거짓인지를 증명할 수 없다. 만일 이 명제 전체가 참임을 증명될 수 없다면 내용상 이 명제는 참이 된다. 즉 증명될 수 없다는 것이 이 명제가 참이기 위한 필요충분조건이 된다. 이는 결정 불가능성 또는 불완전성의 문제가 된다.

둘째로, 첫째 정리의 따름정리(corollary)에 의하면 한 형식체계의 일관성은 그 체계내에서 증명될 수 없다는 것이다. 일관성이 있다는 것은 한 체계T가 어떤 명제 S와 S의 부정을 동시에 증명할 수는 없다는 의미인데 (이를 Con(T)로 표시한다) 이를 즉 Con(T)를 형식체계 스스로가 증명하지는 못한다는 것이다.

이러한 Gödel의 정리는 수학과 논리학에 새로운 관점과 개념을 도입시켰다. 특히 보편튜링기계가 모든것을 알고리즘적으로 기술할 수 있는가

에 관하여 다음과 같은 새로운 관점을 제시하였다.

인간의 마음 또는 모든 수학적 직관들을 기계화, 형식화 할 수는 없다. 인간의 직관을 기계화 한다는 것은 한 형식체계(K)를 유한히 기술한다는 것인데, 유한한 기술에서 우리는 이 형식체계에 모순되는 것을 발견하며 그것은 이 체계가 증명할 수 없다는 것을 안다. 기계론을 용납하여 인간의 수학적 직관에 대응하는 정리들을 증명하는 기계가 존재하며(그 프로그램을 우리가 작성하지 못하더라도) 경험적으로 발견될 가능성이 있다고 하자. 그런데 우리가 이러한 기계R이란 1) 정리 증명기계이기에 유한하게 기술할 수 있으며, 2) R은 인간의 선험적이고 일관성이 있는 수학적 직관에 대등한 것이기에 R은 일관성이 있다. 3) 동일한 이유로 R은 자연수의 연산(whole-number aritlmetic)의 기초 사실들을 증명 할 수 있을 정도로 강하다. 4) 그런데 이들 조건은 Gödel정리의 기본 조건을 충족시킨다. 따라서 R은 Con(R)을 증명할 수없다. 5) 그런데 R은 인간의 수학적 직관과 대등하므로 인간도 Con(R)을 결코 증명할수 없다. 6) 그런데 인간은 Con(R)의 진위를 안다. 고로 인간은 기계와 차이가 있으며 어떠한 기계도 인간의 마음을 적절히 모델할 수 없으며 따라서 인간의 마음은 Turing기계가 아니며 기계화 또는 완전히 형식화될 수 없다.

Gödel은 더 나아가서 다음과 같이 기계론의 오류를 지적하고 있다. “Turing은 ..... 심리적 절차가 기계적 절차 이상으로 수행할 수 없다는 것을 보여주는 논지를 제시하고 있다. 그러나 이 논지는 결론적이지 못하다. 왜냐하면 그것은 유한한 마음이 변별가능한 유한 수의 상태에만 있을 수 있다는 전제에 의존하고 있기 때문이다. Turing이 완전히 무시하고 있는 사실은 마음이 정적이지 않고 끊임없이 발달하고 있다는 사실이다. 이러한 예는 집합론에서 무한에 관한 공리가 끊임없이 더 강한 공리들의 무한 계열을 이루고 있음에서 볼 수 있다.....따라서 마음의 발달의 각 단계에서 가능한 상태의 수는 유한하지만, 발달이 진행되어 가는 과정에서 이 수가 무한에 수렴되지 말라는 이유가 없다”. Gödel은 Turing 의 주장이 1) 물질과 분리된 마음은 없다는 것과, 2) 두뇌의 기능은 디지탈 컴퓨터처럼 작용한다는 두 전제하에서만 가능한데, 2)는 그럴지 몰라도 1)은 현대의 편견이며 과학에 의해 반증될 수 있으리라 본다.

Gödel의 이러한 논지를 확대해석하여 Lucas(1931)는 Gödel정리가 기계론이 거짓이라는 것과 마음은 기계로서 설명될수 없음을 증명한다고 주장한다. 그에 의하면 Gödel의 정리는 싸이버네틱스 기계에 적용해야 한다. 왜냐하면 어떤 형식체계의 구현이어야 한다는 것이 기계의 본질이기 때문이다.

기계가 형식체계에 상응하며, 기계가 출력으로 내어놓는 결론들이 형식체계의 정리와 상응되며 일관성있고 간단한 산술을 할 수 있다면, 이러한 형식체계인 기계에서 Gödel의 공식을 우리는 구성할 수 있다. 이 Gödel의 공식은 기계라는 형식체계 내에서 참 임이 증명될수 없다. 그러나 인간은 그 공식이 참임을 알 수 있다. Gödel공식이 Gödel정리에 의

거하여 그 체계내에서는 증명될수 없지만, 사실 바로 그 이유 때문에 참임을 인간은 확인할 수 있다.

Lucas의 논리를 정리하면 다음과 같다. 어떤 기계 M도 인간의 (수학적)직관 H와 같을수 없다. 이제 기계화된 형식체계 M에 의해 열거된 정리의 집합을 M이라 하고 인간의 수학적 직관 H가 참이라고 주장하는 문장들(Gödel수)의 집합을 H^*라고 하자. 그렇다면 어떤 유한한 M에 대해서도 $M^* \neq H^*$임을 주장할 수 있다. 즉 1) 만일 $M^* \subset H^*$이면 M이 참 형식적 체계인가를 H가 알수있다. 2) 만일 H가 M이 참임을 안다면 H는 M이 일관성 있음을 알며 $Con(M) \in H^*$임을 안다. 그런데 Gödel의 제 2정리는 $Con(M) \in$/ (부정)M^*이다 따라서 $M^* \subset H$이면 $M^* \neq H^*$이다. 물론 $M^* \subset$/(부정)H^*이어도 $M^* \neq H^*$이다. 고로 어떤 M도 H와 대등하지 않다.

Lucas는 계속하여 다음과 같이 진술하였다. Gödel은 일관성있는 체계가 일관성있다는 공식을 그체계내에서 증명할 수 없음을 보였다. 따라서 마음이 기계라면 마음이 일관성있는 기계라는 결론에 도달할 수 없다. 마음이 기계가 아니라면 그럴 수 있다. Gödel은 마음이 '어떤 형식체계의 일관성을 그 체계내에서 형식적으로 증명할 수 없다'는 것을 제시했으나, 그 체계를 벗어나서 나가는 것에 대해서는 반론을 제기하지 않았다. 또한 어떤 형식체계의 일관성이나 덜 형식적이고 덜 체계적인 어떤 것의 일관성에 대해 비 형식적 주장을 제기하는 것에 대해서도 반대하지 않았다. 이러한 비 형식적 주장은 완전히 형식화 할수 없을 것이다.

마음에 대한 기계론적 모델이라면 어떤 모델이건 산술적 참 여부를 밝힐수 있는 절차들을 내포해야 한다. 그것은 인간의 마음이 산술적 참을 가려낼수 있기 때문이다. 그런데 인간의 마음보다 산술적 진리를 더 많이 산출해낼 수 있는 기계론적 모델들이 있을 수 있지만, 어떤 기계이건 그 기계 자체가 참임을 밝힐 수 없는 ,그러나 인간은 밝힐수 있는 그러한 참이 있다. 따라서 어떤 기계도 인간 마음의 완전하고 적절한 모델이 수 없다. 기계가 마음이 하는 모든 것을 할 수는 없다. 왜냐하면 기계가 아무리 많은 것을 할 수 있어도 거기에는 항상 기계는 할 수 없지만 마음은 할 수 있는 것이 있게 된다. 우리는 인간 마음이 할 수 있는 모든 것을 할 수 있는 기계를 산출할 수 있다고 기대할 수 없다. 우리는 결코, 원칙적으로라도 , 마음에 대한 기계론적 모델을 가질 수 없을 것이다. '기계적'이란 본질적으로 '죽은'것이다. 그러나 마음은 실상 '살아있기에' 항상 어떠한 형식적, 화석적, 죽은 체계가 할 수 있는 것보다 한 걸음 더 나아갈수가 있는 것이다. 고맙게도 Gödel정리 덕분에 마음은 항상 마지막 결정적 말을 할수가 있는 것이다.

3. Gödel입장의 문제점.

마음을 튜링기계로 본 Turing의 입장과 이 입장의 본질적 문제점을 제기한 Gödel-Lucas의 입장을 살펴보았다. Turing의 이론이 하나의 설득력있는 혁명적 이론이었으나 Gödel과 Lucas의 논지에도 문제점이 있다. 먼저 거론되어야 할 점은 Gödel정리가 심리현상에 대한 기계론을 염두에 두고 전개된 것은 아니었으며, 단지 그의 불완전정리의 의의를 분석해볼 때 Gödel정리가 Turing적 기계론에 대한 반론으로 해석될 수 있다는 것이라는 점이다. 이러한 점 이외에도 다음과 같은 문제점들이 Gödel-Lucas의 논지에 내포되어 있다.

첫째로, Gödel정리가 적용되는 체계의 본질의 문제이다. Gödel의 정리는 경계, 요소, 규칙들이 잘 규정된 일관성있는 연역적 형식체계에만 적용된다. 그런데 인간의 마음이란 그 요소와 규칙과 경계들이 잘 규정되어 있지도 않으며 연역적 사고이외의 사고도 할 수 있는 개방체계이다. 따라서 인간은 연역적 방법에 의하지 않고도 진리를 산출할 수 있는 기계를 만들 수 있을 것이며, 그렇다면 연역적 체계를 전제하여 제시된 「괴델」의 정리를 적용해서 마음의 우세성을 증명할 수도 없을 것이다.

둘째로 , Gödel식의 생각의 기본오류는 진리와 증명가능성을 동일시하고 있음에 있다고 비판할 수 있다(Wang, 1974). 인간은 모든 기계에 대해 증명 불가능한 문제의 모든 사례를 정확히 결정할 수 없으며, 인간이 알고 있다는 것은 , 형식체계 S가 일관성이 있다면, 참이지만 체계 S내에서 증명될 수 없는 진술 Hs가 있다는 것을 아는 것이지 Hs자체를 모든 경우에 다 안다는 의미는 아니다. 안다는 것 자체가 증명하는 것은 아니다. "'나는 일관성있다'(A)를 안다. 그래서 나는 '내가 일관성이 있다'를 증명할 수 있다."는 논지에는 문제가 있다. 내가 A를 증명할 수 있다면 Gödel정리의 결과로 나는 일관성 있는 Turing기계가 '아니다'. 첨가해서 A가 참이라면 나는 일관성 '없는' Turing기계일 수 없다. 즉 나는 기계일 수 없다는 강한 결론에 도달하게 된다. 그런데 이러한 추론에 있어서 기계가 나보다 더 많은 정리를 증명할 수도 있을 것이다. Hs가 S내에서 증명될 수 없음은 참이다. 그러나 그렇다면 마찬가지로 나도 Hs를 증명할 수 없을 수 있다. 그 까닭은 S의 정확한 명제를 모르거나 어떤 다른 이유로 Con(S)를 증명할 수 없기 때문이다.따라서 내가 A를 증명할 수 있다는 것과 A가 참이다라는 것은 내가 어떤 기계보다도 더 잘할 수 있다(보다 많은 정리를 증명할 수 있다는 의미에서)는 결론을 제시하는 것은 아니다. 또다른 의미에서 다음과 같이 해석할 수 있다. 기계는 순환적으로 열거할 수 있는 집합만을 모두 생성할 수 있다. 그런데 어떤 기계도 나의 정리를 정확하게 생성할 수 없기에 내 정리의 집합은 순환적으로 열거할 수 있는 것이 아니다. 고로 위의 두 가정이, 내가 기계보다 더 많은 정리를 증명할 수 있음을 보여주지는 못하지만 기계적으로 생성될수 없는 집합을 내가 생성할 수 있음을 보여주며, 그런 의미에서 모든 기계에게 거절된 능력을 나에게 부여해준다고 해석할 수 있다.

그런데 내가 A를 증명할 수 있다는 가정은 매우 애매모호하다. 첫째

로, 나는 기계와는 본질적으로 다른 증명개념을 사용하고 있다. 기계는 순환적으로 열거할 수 있는 집합에서의 정리를 증명하는 것이며 수 이론의 참 진술은 산술적 진술이 아니기 때문에 자신의 일관성 진술을 내포할 수 없다. 그런데 인간인 내가 나 자신을 정리 생성기계로 생각하며 동시에 나의 일관성을 증명할 수 있다는 것은 순환적으로 열거할 수 없는 집합에 관한 것이며 자신의 일관성 진술을 내포한 것이다. 이러한 경우의 정리는 전자의 경우의 정리와 다른 개념의 증명이 필요하다.

A를 사용하여 '내가 기계이다'는 전제에서부터 모순을 도출하여 내가 기계가 아님을 증명하는 것은 불가능하다. 나 자신의 기계표(machine table)나 프로그램의 정확한 명제를 안다거나 내가 나 스스로 모순에 빠지지 않고 항 상 정확하게 기능함을 안다는 것은 거의 있을 수 없는 일이다. 또한 A의 의미의 불명료성은 "나는 기계이다"라는 전제의 정확한 힘(의의)을 모호하게 한다. 예를 들어 어떤 의미에서 A를 증명할 수 있을 수 있다. 그러나 '증명'의 의미나 A의 의미는 나는 기계가 아니다라는 결론을 도출하기 위해 가정한 형식적 의미일 수 없다. 따라서 우리는 우리가 사물을 비형식적으로 아는 방법이 있다는 신념을 수용하게 되고 그 이유로 내가 기계가 아님을 믿게 된다. 반면에 A가 일관성있는 산술적 진술과는 다르다는 것과 A를 증명할 수 있음에 동의한다면, 그렇다면 나는 실상은 기계이지만 나 자신이 일관성이 있는지를 내가 모른다고 할 수 있다.

이 문제를 달리 접근할 수도 있다. 인간인 나의 안에 요소C가 있어서 정리 산출 기계로서 작용하며 Gödel의 정리가 적용될 수 있을 정도로 충분한 수이론을 생성 가능하다고 하자. 첨가하여 C가 일관성 있음을 내가 증명할 수 있다고 하자. C가 나의 기계적 부분의 전부가 아니라면 나는 C보다 더 낫다 (어떤 기계보다 더 많은 정리를 증명할 수 있어서가 아니다). C가 나의 기계적 부면의 전체라면 C의 일관성의 정리는 나의 C가 아닌 다른 부분에서 와야하고 따라서 나의 일부는 비기계적이라고 결론지을 수 있다. 그러나 기계(기계적 부분요소)가 스스로 모순에 도달했는가를 점검하고, 모순에 도달했으며 기본공리를 수정하도록 고안한다면 기계적요소와 나는 일관성 없는 체계라는 점에서 차이가 없어지며 기계와 마음의 차이가 모호해진다.

인간과 기계의 근본적 차이가 인간의 마음은 자신에 관한 질문에 대해 자신이외의 체계가 되지 않고도 대답할 수 있다는 것이다. 그런데 기계도 기계 자신에 대한 질문에 대답할 수 있음이 드러났다. 체계 S가 수 이론의 보통체계일때 '나는 S안에서 증명될 수 없다'가 S의 정리임이 드러났다. 또한 인간이라고 해서 자신에 대한 완전한 지식을 가질 수 있는 것은 아니다.

우리는 우리가 해답할 수 없는 우리 자신에 관한 질문들이 많이 있음을 느낌으로 알수 있다. Wang(1974)에 의하면 Gödel의 정리를 중심으로 한 '인간은 기계이상이다'의 문제에 관한 논쟁은 명백한 결론에 도달할 수는 없다고 하겠다.

셋째로, D.C.Dennett(1981)에 의하면 Gödel의 정리를 마음에 적용하려고 한 시도의 본질적 오류는 Turing기계와 이에 상응하는 구체적 대상의 규정명세(specification)를 제시할 수 있다는 생각이었다. 그러나 어떤 Turing기계를 규정할 수 있으며, 이와 상응하는 구체적 대상의 활동과 능력을 객관적이고 배타적으로 결정하여, 어떤 Turing기계의 규정명세가 이 대상에 대한 올바른 규정명세인가를 명확히 제시할 수 있다는 생각은 잘못이다. Turing기계 규정명세는 그렇게 명백히 자세히 주어져 있지 않다. 이런 오류를 인정한다면 Gödel의 정리를 마음에 적용하려는 것은 아무런 쓸데가 없는 것이다. 한 인간이 정리 증명의 특정 Turing기계의 구현이라면 Gödel문장 S를 증명하지 못 할 것이다. 그러나 이는 그의 다른 역할에서의 능력에 대해서는 아무것도 이야기하지 못한다. (인간은 단순한 정리 증명목표이상의 궁극적 목표를 지니고 있다. 기계와 동일한 정리를 산출한다는 것이 인간이 기계이어야 하는 충분조건은 아니다) 어떤 것을 Turing기계 Tm으로 규정하고 그 능력에 어떤 한계를 부여한다는 것은 그것의 다른 능력, 다른 면에 대해 아무것도 말해주는 것은 아니다. Gödel 정리에 의한 한계란 알고리즘에 의해 이뤄질 수 있는 것에 대한 한계이지 알고리즘에 의해 자기발견적으로(huristically) 할 수 있는 것에 대한 한계는 아닌 것이다.

넷째로, Gödel화 하는 알고리즘 방법의 결여 문제이다. Lucas는 인간이 항상 체계를 벗어나서 체계 밖에서 Gödel화 하는 작업을 할 수 있는데 반하여, 기계는 이를 할 수 없다고 하였다. 할 수 있다고 하여 Gödel화 연산자를 기계 안에 프로그램으로 내장하여도, 끊임없는 새로운 Gödel공식(a)의 Gödel공식(b), Gödel공식(a)의 Gödel공식(b)의 Gödel공식(c)... 의 양식으로 무한 계열을 이루게 되며 그렇다 해도 결국은 그 체계내에서 증명될 수 없는 Gödel공식이 있게 된다고 하였다. 이러한 논리로 Lucas가 기계론을 반박하지만 ,이러한 논리는 ' Gödel화 할수 있다'는 것을 추상적으로 논의한 것이지 모든 개개의 사례에서 어떻게 Gödel화를 할 수 있는가를 명시해 준 것은 아니다. 이는 Dennett의 규정명세의 결여에 대한 비판과 연결되는 문제이다. Gödel-lucas의 이론은 Gödel방법을 모든 가능한 유한의 형식적 체계에 어떻게 적용할 것인가를 기술하는 알고리즘적 방법을 제시하고 있지는 않다. 수학자 Church의 이론에서 논의되는 바와 같이 모든 서수(ordinal)에 대해 이름을 줄 수 있는 표상체계란 없다. 서수가 커질수록 불규칙성이 나타나고 '불규칙성 내의 불규칙성','불규칙성 내의 불규칙성 내의 불규칙성'등으로 전혀 새로운 질서의 것이 나타나게 된다. 따라서 하나의 도식이 제 아무리 복잡하고 포괄적이라도 모든 서수를 명명할 수는 없다. 그러므로 형식적 체계 또는 프로그램이 복잡해짐에 따라 Gödel방법을 적용하기 위해 무엇을 어떻게 해야할 지를 명백히 지적할 수 없게 되고, 마침내는 Gödel방법을 적용할 방도를 전혀 강구해 낼수 없는 복잡한 경우에 마주치게 된다. 이와 같이 형식체계 또는 기계가 너무 복잡하여 Gödel방법을 어떻게 적용해야 될지를 모를 경우에

, 이 체계 또는 기계는 비록 불완전하지만 인간의 마음의 능력과 거의 같아지게 된다. 이러한 경우에 기계와 인간의 마음의 구별은 불명확해 지며 더구나 Gödel정리의 적용에 의하여 이 둘을 구별짓는다는 것은 무의미해진다.

다섯째로, Gödel-Lucas의 논지는 상징조작이 단 한 수준에서 일어남을 시사하고 있는데, 생리학,심리학의 연구에 의하면 인간의 마음이라는 체계가 다원적 수준의 체계임을 인정하지 않을 수 없다. 상징조작은 단 한 수준에서만 일어나는 것이 아니다. 여러수준에서 상징조작이 일어날 수 있으며 하위수준은 형식적 특성을 지니나 상위수준은 그렇지 못할 수도 있다. 상위수준과 하위 수준이 하나는 비규칙적이며 불완전하고 다른 하나는 규칙적이며 완전할 때 이들의 연결이 어떻게 이루어지는가에 대하여 Gödel이론은 뚜렷한 대답을 지니고 있지 못하다.

기계와 인간의 마음을 연결짓는 인공지능학에서의 각종 프로그램은 형식체계의 구현이긴 하지만, 상징조작이 여러수준에서 있을 수 있음을 보여준다. 최하위 수준이 있을 수 있고, 상위수준이 여럿 있을 수 있으며 고위수준은 비형식적일 수 있다. Gödel과 Lucas의 논지를 약하게 해석하여서 그들이 다수준을 인정했다고 해석하더라도 문제는 남는다. 그들은 수준들 상호간에 서로 독립적이며 (최)상위 수준만 따로 떼어낼수 있고, 각 수준들에 대응하는 요소들이 외계현실에 있으며, 상위수준의 의미가 하위수준의 의미에 의해 결정되지 않는다고 주장한다. 그러나 Hofstaedter(1980)에 의하면, 현실의 심리과정이란 단일 수준이 아니라 여러수준이 함께 작용하는 것이며 수준들은 상호작용하며 의존적이고 , 상위 수준을 지원하는 하위 계층으로서의 하위 수준을 포함하지 않고는 상위 수준의 의미표상이 불가능하다. 이러한 체계의 수준들은 현실의 요소에 직접적으로 대응하여 연결할 수 없는 수준들이 있으며 따라서 한 수준은 외계와의 직접적 대응-연결관계에 의해서 보다는 타 수준(특히 상위수준)과의 촉매관계에 의해서만 이해될수 있다. 인간의 두뇌-마음의 체계에서는 그 자체로서는 완전하며 규칙적이며 합리적인 하위수준(예:뉴론수준)이 ,불완전한 오류가 있고 비합리적인 상위수준(일부의 사고,감정)을 지원할 수 있다. 그러나 Gödel- Lucas의 입장에서는 이러한 연결은 불가하다.

Gödel정리는 인간이 인간 자신의 마음/두뇌를 이해할 수 없다는 것을 의미하는 것으로 해석된다. 그러나 이러한 해석은 '마음/두뇌를 이해한다'는 것이 무엇인가, 어떤 수준에서 이해한다는 것인가를 지적하지 않고는 무의미하다. 그렇긴하지만 Hofstaedter는 Gödel의 정리를 넓게 해석하여 Gödel정리가 인간의 마음과 두뇌에 대한 이해가 불가능하다고 단정한 것은 아니라고 본다. 우리가 우리의 마음을 이해 못한다고 한다면 그것은 Gödel식의 기초적 이유 때문이기 보다는 우리 두뇌가 우리 자신을 이해하기에는 너무 약하기 때문에 또는 우연적 이유로 설명불가하기 때문에 그럴수 있다. Gödel정리가 인간의 마음과 기계의 차이레 대해서 또는 인간이 인간의 마음을 이해할 수 있는가에 대해서 단정적인 진술을 제시했다

고 해석하기 보다는 오히려 인간이라는 체계 내에서의 각 수준간의 관계성에 대한 이해를 촉진시켰다고 긍정적으로 해석해야 할 것이다.

Gödel 정리를 이해하기 위해서, 임의의 상징설정, 자기 참조, 복잡한 이형동질(isomorphism), 상위 수준의 설명 등의 개념을 우리는 이해하여야 했고, 이러한 이해과정을 통해 심적 구조의 여러 수준간의 관계성에 대한 이해가 촉진되었다고 하겠다. Gödel의 정리는 어떤 체계에 대한 고위수준적 관점이 하위 수준에서는 결여되어 있는 설명능력을 부여 할 수 있다는 해석을 제시한다. 즉 어떤 사실은 하위 수준에서는 전혀 설명될 수 없지만, 고위 수준에서는 쉽게 설명될 수 있다는 점이다. 고위 수준의 개념(예:의식)은 하위 수준에서 주어질 수 없는 창발적인(emergent) 새로 솟아나는 특성인 것이다. Gödel정리의 의의는 , 한번에 단 하나의 수준에서 체계를 이해해야 한다는 것이 아니라, 한 수준이 그 초수준(mata-level)을 반영하고 이 반영한 결과에 의해, 서로 다른 수준들 사이의 관계성에서 창발적 현상으로 설명되어야 한다는 것이다.

상위수준이 하위 수준에 뻗치어 영향을 주고 동시에 하위 수준에 의해 영향을 받아 결정되는 수준간 상호작용의 '기이한 루프'로서 이해해야 하는 것이다. 이러한 수준 교차(level-crossing)가 Gödel정리라는 소용돌이(Gödel vortex)에서 이루어진다는 의미로 해석되어야 한다고 Hofstaedter는 주장하고 있다. 어떤 체계를 이해하기 위해서는 수준간의 상호작용이 강조된 다수준적 설명이 필요함과 그 설명능력을 제시한 것으로 Gödel정리를 해석한다면 이러한 해석은 인간의 마음과 기계의 연결문제를 구체적으로 모델링을 통해 연구하고 있는 다수준설명적 입장의 인지과학과 인지심리학의 연구 경향에 부합되는 것이라고 볼수 있다.

이러한 입장의 해석은 Lucas가 Gödel정리를 해석한 입장과는 다른 것이다. 기본적으로 Lucas는 Gödel정리의 불완전성과 진리 개념에 강조를 두어 해석함으로서 Gödel정리가 기계론을 반증하는 증거라고 본 것이다. 그러나 상술한 Hofstadter의 해석을 따른다면 Gödel의 정리에서 자기참조(self reference)와 미결정성(undecidablity)의 개념을 더 중요한 개념으로 간주한 것이며, Gödel정리를 오히려 기계론을 옹호한 또는 그에 가까운 의의를 지닌 것으로 해석한 것이다.

종합하여 말하자면 Gödel-Lucas의 논지가 반기계론을 강력히 제시하였으나, '인간의 마음은 기계 이상이다'를 증명해주었다고 하기는 힘들다. 즉 기계론을 반증하지는 못했다. 단지 기계론을 증명하려는 논리적 시도의 문제점을 표면화시켰을 뿐이다. 마음의 작용에 대한 형식적 기술에 있어서 순환적 자기참조성, 상위수준의 독특성 등을 중심으로 개념화할 필요성을 제시한 것 뿐이다.

4. 튜링기계론의 다른 문제점.

Turing의 기계론을 다시 살펴본다면 Gödel정리가 제기했던 본질적

문제점 이외에도 다음과 같은 문제점들이 있다. Turing기계론은 유한히 기술될 수 있는 기계상태로 이루어진 기계표(machine table)를 전제로 한다. 이를 마음의 상태에로 적용한다면 인간의 심적 상태들이 유한히 기술될 수 있다는 결론을 도출하게 되는데 이 결론은 받아들이기 힘들다. 또한 튜링기계는 하나의 이상적 체계이기에 현실적으로 구현화하는데에 여러가지 문제가 수반된다. 첫째로 현실세계의 튜링기계는 이상적 튜링기계처럼 무한한 길이의 테이프 즉 ,기억용량을 지닐수 없다. 인간의 기억용량은 제한되어 있다. 둘째로 인간의 마음을 포함한 어떤 튜링기계도 현실적으로 그 처리과정의 수행을 유한시간이라는 제약내에서 이루어내야 한다. 따라서 현실적 튜링기계의 처리속도에는 한계가 있고, 이러한 제약이 새로운 가외변수로 작용한다. 셋째로 현실적 튜링기계는 그 처리조작을 수행함에 있어서 절대적으로 오류가 없을 수는 없다. 제로 이상의 오류확률을 지녀야 한다. 즉 현실적 튜링기계는 이상적 튜링기계와는 달리 확률적이다.

이러한 현실적 문제이외에도 마음을 튜링기계로 간주하는 입장의 본질적 문제가 있게되고 이러한 문제들이 심리철학적으로 논의되어 왔다. 이 문제들 중에 두가지 문제점을 부각시켜 본다면 다음과 같다.

첫째 문제는 심적상태가 튜링기계의 기계상태에 대응되는 것이냐 아니면 비 물질적인 계산상태에 대응되느냐의 문제이다. 튜링기계론을 인지이론에 도입한 심리철학의 기계 기능주의와는 달리 심리 기능주의는 심적상태란 튜링기계의 기계상태에 대응되는 것이 아니라 추상적인 계산상태에 대응되는 것이라고 본다. 전술한 바와 같이 튜링기계론은 유한한 기계상태들을 전제로하는데, 인간의 심적상태는 무한하다고 할수 있다. 또한 서로 다른 기계상태를 거치고서도 동일한 심적 내용을 산출해 낼수도 있다. 하나의 심적 내용에 대해 항상 고정된 , 모든 사람에게 공동된 불변의 기계상태를 규정하기 힘들다. 고로 심적내용이란 특정기계상태가 아니라 기계상태들이 서로 지니는 관계성에 의해 이루어내는 계산상태에 의해 결정된다고 볼 수 있다. 따라서 튜링기계론은 기계상태와 심적 과정 또는 상태를 대응시키려 했을 것이 아니라 튜링기계의 계산상태와 심적상태의 대응을 시도했어야 할 것이다.

둘째는 튜링기계론은 인간의 마음이 본질적으로 규칙지배적이며 알고리즘적이고 형식화 할수 있으며 합리적이고 기계적 결정론에 의존함을 전제로 한다. 그러나 인간의 마음이란 규칙체계 즉 통사체계에 의해 설명될 수 없는 부분을 지닌다(Searle, 1980, 1984).

Searle(1984)에 의하면, 마음은 심적 내용을, 특히 의미 내용을 지닌다. 의미란 통사(규칙체계)에 의해 충분히 나타낼수 없다. 그런데 튜링기계란 전적으로 형식적 또는 통사적 구조에 의해 규정된다. 따라서 튜링기계는 인간의 마음의 의미내용을 충분히 다룰수 없다. 인간의 마음은 지향적 내용을 지니고 있다. 지향성(Intentionality)이란 심적상태가 그 자체 이외의 대상에 대하여 참조하며, 지향하는 것이다. 세상 대상에 대하여 의도하고, 믿고, 욕망하고, 바라고,두려워하고, 사랑하는 등의 모든 상태를 지칭한다.

이러한 지향성은 통사적 체계에 의해 나타내어질 수 없다고 본다. 이러한 심적 상태에 의해 물리적 대상에 변화가 일어나는 지향적 인과나 마음의 주관성 등은 튜링기계론에 의해 설명될 수 없는 비규칙지배적, 비형식적 특성을 지닌다고 본다.

한편 인지심리학과 인접분야의 연구들,특히 사고와 판단과정의 연구와 정서연구들은 인간의 마음이 알고리즘적,합리적 규칙에 의존적이지 않은 면을 지니고 있음을 지적하고 있다(Gardner, 1984; Tversky, Kahneman, & Slovic, 1983). 이러한 연구들과 Searle의 주장은 인간의 마음이 본질적으로 규칙지배적이거나 알고리즘적이지 않은 면들이 있음을 제기한다. J.Fordor, Z.Pylyshyn, P.Johnson-Laird 등은 이러한 주장들을 보다 세련된 기능주의(functionism)나 계산주의(computationalism) 에 의해 반박할 수 있다고 주장하지만(Dennett, 1986) Searle등이 던진 의문은 계속 남아있다. 이 의문의 핵심은 바로 Gödel에 의해서 제기된바 '자신을 넘어서는 작업을 할 수 있는'체계로서의 인간의 특성과 관련된 문제라고 하겠다. 이는 하위수준에서는 결여되어 있는 특성이 하위수준을 넘어서는 상위수준에서 솟아나는(emergent)특성을 인정할 것인가의 문제이며 심신론 논쟁의 핵심이기도 한다.

5 맺는 말

심신론에 대한 심리철학적 논쟁에 깊숙히 들어가지 않더라도 상술한 바와 같이 문제점이, 마음을 튜링기계로 간주하려는 관점에 내재함을 인정할 수 밖에 없다.

그렇다면 우리는 어떠한 입장을 취해야 하는가? 심리현상에 대한 이론으로서 튜링기계론을 완전히 기각하여야 할 것인가 아니면 수정된 튜링기계론을 모색할 것인가? 우리는 병행처리모델 또는 연결주의모델과 양자물리학에서 몇가지 시사점을 얻을수 있을 것이다. 기존의 튜링기계적 심리이론이 계열적 상징조작 처리체계를 출발점으로 삼음에서 오는 여러가지 문제점을 극복하기 위해서 Rumelhart와 McClelland(1986)등은 연결주의 모델(Connectionist model)또는 병행분산처리(Parallel Distributed Processing)모델을 제시하였다. 이들 신연결주의자들은 병행분산처리체계 모델이 의식의 문제, 자각의 문제, 지향성의 문제에 대한 잠정적 답변을 제시할 수 있다고 본다(Johnson-Laird, 1987). 그들은 두뇌가, 유한상태 기구들이 위계적으로 병행적으로 조직된, 비동시적 계산을 수행하는 체계라고 보며, 이러한 특성이 계열처리체계인 튜링기계의 제한점을 부분적으로 해결해 준다고 본다.

한편 물리학자 Margenau(1984)는 두뇌의 활동에 양자역학을 도입해야하며 마음이란 양자역학적 장(field)으로서 비물질적, 비에너지적, 확률적 장으로서 개념화할 수 있다고 하였다. 생리학자 Eccles(1987)는 완전히 잘 규정된 정확한 물리법칙도 때로는 비알고리즘적 행동을 일으킬 수 있으며

따라서 매우 단순한 고전물리학의 결정론적 체계내에서도 비계산적, 비알고리즘적 행동이 일어난다고 본다. 두뇌를 비롯한 하드웨어의 기능중에는 알고리즘으로 기술할 수 없는 기능이 있으며 고전물리학의 기계적 결정이론으로 완전히 기술할 수 없는 면들이 있다. 따라서 양자역학적 비결정론적 설명이 도입되어야 한다는 것이다. 두뇌의 활동에 양자역학적 설명을 도입하며 마음을 비물질적 양자역학적 장으로 개념화한다는 것은 , 두뇌의 작용이 비튜링기계적 면이 있음을 인정하는 것이며, 따라서 그러한 두뇌의 작용에 의해 발생되는 심리현상이 비튜링기계적일 수 있음을 인정하는 것이다.

그러하다면 문제점이 있는 기존의 튜링기계론(고전적 튜링기계론이건 수정된 세련된 튜링기계이건)에 대한 대안으로 병행분산처리모델과 양자역학적 불확정(무선)모델의 접합을 생각해 볼 수도 있을 것이다. 병행분산처리모델은 비록 고전적 튜링기계론의 문제점을 일부 극복하였으나 여전히 계산주의의 기본 전제를 그대로 지니고 있으며 또 뉴론이라는 하드웨어의 하위 수준 중심의 모델이다. 또한 확률개념이 도입되었으나 양자역학의 불확정적 무선적 개념이라기 보다는 결정법칙에 따르는 확률의 개념이다. 따라서 튜링기계에 대한 비판들이 지적하였던 문제들을 병행분산처리모델은 모두 해결하지 못한다. 특히 Gödel의 反기계론적 비판의 논지를 극복했다고 할 수 없다. 인간의 마음의 Gödel적 특성, 비결정적, 비기계적, 수준 교차적(level-crossing) 특성을 기술 할 수 있기 위하여 양자물리학의 불확정성의 모델을 하나의 무선자(randomizer)로서 도입하는 것을 고려해 볼 수 있을 것이다.

그러나 이러한 시도가, Searle(1984) 등이 제기한 여러가지 문제점을 해결해주고 마음을 기계로 또 마음의 활동을 계산으로 보는 입장들을 의문점이 없는 입장으로 확립시킬 수 있다고 보기는 힘들다.

마음이 기계적이며 계산적인가 하는 물음은 마음의 과정이 '어떻게 진행되는가'의 문제이다. 이러한 물음에 대한 만족할 만한 답변을 얻기 위해서 우리는, 잠시 멈추어 마음의 과정들이 이루어내는 것의 본질이 과연 '무엇인가'하는 물음을 다시금 되 생각해야 할것이다.

'어떻게'심적 과정이 이루어지는가를 잠시 제쳐놓고, 그러한 과정들이 이루어내는 내용의 본질 또는 기능의 본질에 대한 물음을 던져야 할 것이다. 이는 심리현상을 단순히 통사적이고 기계적인 과정으로서만 보는 것이 아니라, 내용을 지닌 현상으로 보는 것이며, 과정과 내용을 동시에 고려함으로서 과정 중심의 기계론이 지니는 한계성을 넘어서자는 것이다. 심적과정이 이루어내는 바의 본질에 대한 물음이 함께 고려됨으로써 비로서, 우리는 하위수준에서 주어질 수 없는 새로운 특성이 상위수준에서 창발적으로 출현되는 것에 대해 보다 적절한 설명을 줄 수 있을 것이다. '마음은 기계적인가' 하는 물음이 마음의 내용의 본질에 대한 물음과 분리된 채 던져짐은 적절하지 못한 접근이라고 하겠다.

* 참고 문헌 *

이정모 (1984). Gestalt개념의 형성사(I) Descartes에서 Hamilton 까지. **한국심리학회지,** 4권 2호, 97-118.

Anderson A. R. (Ed)(1964). *Mind and machines.* Englewood. Cliffs. N. J. : Prentice-Hall.

Arbib. M. A(1964). *Brains, Machines, and mathematics.* New York: McGraw-Hill.

Cutland. N. (1980). *Computability:* An introduction to recursive function theory. Cambridge: Cambridge U. Press.

Dennet, D. C. (1981). *Brainstorms: Philosophical essays on mind and psychology. Hassocks,* Sussex: Harvest Press

Dennet, D. C. (1986). The logical geography of computational approaches: A view from the Eastpole. In Bravel, M. & Harnish, R.M. (Eds.). *Problems in the representation of knowledge and belief.* Tucson : Arizona U. Press

Eccles, J. (1987). Brains and Mind: Two or One? IN C. Blakemore & S. Greenfield.(Eds.). *Mindwaves.* Oxford Basil Blakewell.

Gardner. H. (1984). *New sciences of mind.* Cambridge. Mass.: Harvard U. Press.

Godel. K. (1931). Ueber formal unentscheidbare Satze der Principia Methematica und verwandter System I. (Translated in J. von Heijenont. (Ed.). *From Frege to Gödel: Sourcebook in mathematical logic 1879-1931.* Cambridge. Mass.: Harvard U. Press.(1966).

Hofstadter. D. (1979) *Gödel, Escher, Bach.: An eternal golden braid.* New York: Basic Books.

Jackson, Jr. P. C. (1985) *Introduction to Artificial Intelligence.* New York: Dover.

Johnson- Laird. P. (1987) How could consciousness arises from the computations of the brain? In C, Blakemore. & S. Greenfield (Eds.). *Mindwaves.* Oxford: Basil Blackwell.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.).(1982). *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Lucas. J. R. (1961). *Minds, Machines, and Gödel.* Philosophy. 36, 112-127.

McNaughton. R. (1982). *Elementary computability, formal languages, and automata.* Englewood. Cliffs. N. J.: Prentice-hall.

McGinn, C. (1987). Could a machine be conscious? In C. Blakemore. &

S. Greenfield(Eds.). (Eds.). *Mindwaves.* Oxford: Basil Blackwell.

Margenau. H. (1984). *The Miracle of existence.* Woodbridge, Conn.: Ox Bow Press.

Rucker. R. (1982). Infinity and mind: *The science and philosopy of infinite.* : Brighton, Sussex: Harvest Press

Rumelhart. D. E. (1982). & McClelland, J. L. (1986). *Parallel distributed processing: Exploration in the microstructure of cognition.* Cambridge. Mass.: MIT Press.

Searle, J. R. (1980). Minds, brains and programs. *Behavioral and Brain Sciences, 3.* 417-458.

Searle, J. R. (1984). *Minds, brains and science*. Cambridge. Mass.: Harvard U. Press.

Searle, J. R. (1987). Minds and brains without programs. In C. Blakemore. & S. Greenfield(Eds.). *Mindwaves.* Oxford: Basil Blackwell.

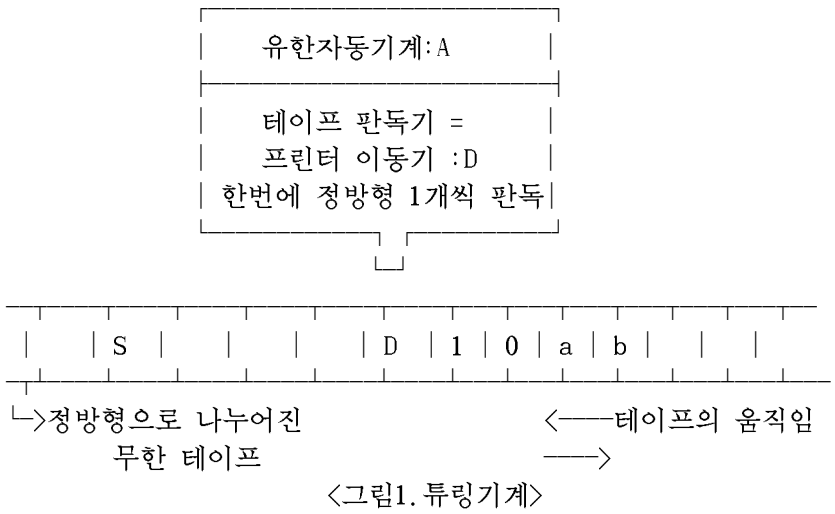
Turing. A. M. (1936). On computable numbers with an application to the Entscheidungs-problem. *Proceedings of London Mathematical Society. Vol. 42.* 230-265, & *Vol 43.* 544-546. Reappeared in M. Davis. (ed.) (1965). *The undecidable*. Hewlett, N. Y.: Raven.

Turing. A. M. (1950). Computing machinary and intelligence. *Mind.* 59. 433-460. Reappeared in A. R. Anderson. (Ed.) (1964). *Minds and Machines,* Engelwood. Cliffs. N. J.: Prentice-Hall.

Wang. Mao. (1974). *From Mathematics to philosophy.* London: Routledge & Kegan.

[그림1], [그림2] == >

튜링기계-21



<그림1. 튜링기계>

[그림 1]. 튜링기계

A: 연산과제 / 덧셈 (3 + 2) --> (5)

테이프 내용의 변화 : 0/1/1/1/0/1/1/0 --> 0/0/1/1/1/1/1/0
판독기의 위치 : (이후 테이프가 좌우로 움직임)

B:기계표 :『IF (현상태, INPUT), THEN (OUTPUT, 변화된 새상태, 테이프 움직임)』

현상태 \ INPUT :	0	1
S1	(0, S, 좌)	(0, S2, 좌)
S2	(1, S3, 좌)	(1, S2, 좌)
S3	(0, S0, 정지)	(0, S3, 좌)

C: 기계표를 사용하여 (3+2) (5)의 조작을 수행한 단계 별 내용

단계	현상태	INPUT(판독)	OUTPUT	새상태	테이프 움직인 방향	출력후 테이프의 내용

튜링기계-22

1	S1	0	0	S1	좌	0
2	S1	1	0	S2	좌	0 0
3	S2	1	1	S2	좌	0 0 1
4	S2	1	1	S2	좌	0 0 1 1
5	S2	0	1	S3	좌	0 0 1 1 1
6	S3	1	1	S3	좌	0 0 1 1 1 1
7	S3	1	1	S3	좌	0 0 1 1 1 1 1
8	S3	0	0	S0	정 지	0 0 1 1 1 1 1 0

[그림 2].「튜링」기계의 조작 수행의 한 예 : (3+2) = 5