# MATHEMATICAL FOUNDATIONS OF COGNITIVE SCIENCE

Woosik HYUN

Hoseo University, Anseo 268, Cheonan, Korea
godel@hoseo.edu

**ABSTRACT**

A would-be cognitive scientist needs to get familiar with three mathematical landmarks: Turing machines, Neural networks, and Gödel's incompleteness theorems. This paper explores the mathematical foundations of cognitive science, focusing on these historical works. We begin by considering cognitive science as a metamathematics. The following parts address two mathematical models for cognitive systems; Turing machine as the computer system and Neural networks as the brain system. The last part investigates the retrospective and prospective implications of Gödel's works for cognitive science.

**Keywords:** mathematical models of cognitive system, Turing machines, neural networks, Gödel's incompleteness theorems, quantum cognitive science

## 1 수학에서 인지과학으로

현대 과학의 가장 특징적인 면은 창의적 융합(creative convergence)을 강조하는 노력에서 찾아 볼 수 있다. 융합은 일정한 분야와 그와는 다른 분야의 경계를 넘나들면서 일어나는 사건이며, 창의성은 다른 사람이 보지 못한 새로운 연결과 생성을 함의한다. 융합이 수평적인 차원의 문제라면, 창의성은 수직적인 차원의 문제이다. 이러한 창의적 융합을 위하여 수학 분야에서는 일정한 하부시스템의 부족한 성질들 자체를 새로운 개념으로 추가하여 활용함으로써 보다 풍부한 능력을 갖춘 새로운 상위시스템의 확장(extensions)을 이룩해 왔다 ([32][33][34]).

인지과학(cognitive science)은 20세기 중반에서 후반 사이에 성립되었고, 창의적 융합을 추구하며 인지시스템을 연구하는 학문이다([25]). 인지과학의 역사를 살펴볼 때, 어떤 인지과학자도 수학자와 수학의 공헌으로부터 자유로울 수 없다는 사실을 발견할 수 있다([16][18][19][23][26][27][28][29]). 본 연구에서는 수학사의 전망에서 먼저 인지과학의 생성과 발전 과정을 분석한다. 그 다음으로는 인지과학의 기본을 이루는 두 가지 수학적 구조를 밝히고, 인지과학의 과제를 전망하고자 한다.

## 2 메타수학으로서의 인지과학

현재의 과학이 현실세계(real world)에서 다룰 수 있는 인지시스템은 (1) 계산하는 기계(Computers), (2) 뇌(Brain), 그리고 (3) 인간의 마음(Mind) 세 가지이다. 인지과학은 컴퓨터와 뇌와 마음을 융합적으로 다루는 학문이며, 그 연구의 결과는 전체 과학에서 수렴가능하다는 비전을 전제하고 있다. 이러한 비전이 실현될 수 있기 위해서는 무엇보다도 먼저 학문적 의사소통을 위한 언어가 필요했다. 다양한 분야 상호 간에 내용을 정확하게 전달할 수 있고, 특정 분야에 편중되지 않으며 보편적으로 충분히 사용할 수 있는 도구로서 추상적 언어(formal language)가 필요했는데, 그러한 조건을 만족시킬 수 있는 것은 바로 수학이었다. 인지과학 연구결과의 소통을 위해서는 내용이 수학적으로 표현가능하고 수학적으로 이해될 수 있어야 했다.

하나의 인지시스템으로 마음과 뇌와 컴퓨터를 이해하고 그러한 시스템을 구현하기 위해서는 융합이 가능한 수학적 모델을 찾거나 구성하는 일에 성공해야 한다. 곧 인지시스템에 대한 타당하고 포괄적인 수학적 모델이 존재할 때, 인지시스템의 구조와 기능에 관한 충분한 이해를 추구하는 인지과학의 목표는 달성될 수 있다. 여기에서 수학적 모델은 현실세계(real world)가 아니라, 형식세계(formal world) 또는 추상적 세계(abstract world)에 속하는 대상이다. 그러므로 현실세계의 뇌를 형식세계의 수학적 모델과 대응시켜 다루는 것, 또는 현실세계의 컴퓨터를 형식세계의 계산기계와 대응시켜 다루는 것이 인지과학에서의 수학적 해석(interpretation)이자 검증작업이다.

인지과학은 마음과 뇌와 컴퓨터라는 인지시스템(cognitive system)을 연구하는 과학이다. 그러나 현재까지 인지(cognition)에 대한 단일한 정의는 확립되지 못했다([19][23]). 역설적으로 인지에 대한 진정한 과학적 정의가 내려지는 순간이 인지과학이 그 임무를 다하는 시점이 될 것이다. 심리학자들은 인지를 정신적 세계 내의 정신적 기능(mental function)으로 정의한다. 신경과학자들은 생물학적 세계 내의 물리적 기능(physical function)으로 정의한다. 인공지능학자들은 기계적 세계 내의 계산적 기능(computational function)으로 인지를 정의한다. 그러나 인지과학자들은 더 나아가서 인지를 마음, 뇌, 컴퓨터와 관련하여 하나의 틀 내에서 추상적으로 정의하고자 시도한다([23]). 이것은 인지과학자들이 마음에서의 인지라는 기능을 '뇌'와 '컴퓨터'의 어떤 수준(level)에서의 정보처리(information processing)로 이해하고 있음을 의미한다. 인지는 의향이나 목표지향적 행위(goal-oriented behaviors)를 내포해야 한다. 예를 들면, 조건반사에 따른 감각이나 행동들은 인지에 포함되지 않는다. 일반적으로 지능, 추론, 학습, 직관, 언어, 기술, 이해, 문제해결, 의사결정, 기억, 지각, 동기화, 감성, 정서가 인지에 포함된다([18][19]). 논의의 집중을 위해서 이 글에서는 인지의 범위를 수학적 사고(mathematical thinking)에 한정하여 다룰 것이다([21]).

인지과학의 역사에서 인지에 대한 가장 중요한 정의는 계산(computation)의 개념으로부터 시작되었다([18][19][23]). 물론 이러한 이해는 명백하게 수학의 개념으로부터 도출된 것이다. 인지를 일종의 계산으로 이해하면서 인지과학의 성립이 가능했고, 계산은 인지과학의 키워드이자 연구방법을 결정하는 도구였다. 예를 들어, '표상'과 '처리'라는 인지과학의 핵심 개념에 가장 영향력이 있고 표준적인 정의를 소개하면 다음과 같다([15]).

표상(representation). 표상은 어떤 정보의 타입과 실체를 명시화해주는 형식체계(a formal system)로 정의된다.

처리(process). 처리는 한 형식체계로부터 다른 형식체계로의 함수(a mapping)로 정의된다.

여기에서 인지란 하나의 정보처리과정(information processing)으로서 이해되고 있고, 이것은 형식체계에서 형식체계로의 함수와도 같이 이해되고 있다. 여기에서 함수는 일정한 구조들 사이의 구조함수에 해당된다고 할 수 있다.

신경생리학과 컴퓨터과학을 전공했던 인지과학자 마(D. Marr)는 인지과학을 위하여 (1)계산이론, (2)표상과 알고리즘, (3)하드웨어 구현이라는 세 가지 수준의 연구가 꼭 필요하다고 강조하였다([15]). (1)과 (2)의 관계는 타입 동일(type identity)의 관계로 볼 수 있고, (2)와 (3)의 관계는 실현(realization)의 관계로 볼 수 있다. 수학과 뇌과학을 전공했던 인지과학자 아비브(M. Arbib)는 인지적 행동(높은 수준을 의미)과 물리적 뉴론(낮은 수준을 의미) 사이의 연결 관계를 설명해 줄 수 있는 계산의 스키마(schema)를 통해서 기능적 수준과 구조적 수준 사이의 연결 관계까지 설명할 수 있어야 한다고 강조했다([1]).

과학에서 행동(action)은 시간($Time$)에서 공간으로의 함수(mapping)로 이해된다. 그렇다면 인지($cognition$)는 시간에서 마음이라는 특정 공간($cognition$)으로의 함수로 이해될 수 있다.

$$Time \xrightarrow{cognition} Mind$$

그리고 현재 인지과학에서 전제하는 마음이란 시스템은 뇌와 컴퓨터의 곱(product)으로 설정되어 있다.

$$Mind = Brain \times Computer$$

여기에서 뇌에 대한 연구를 통하여 마음을 이해하려는 접근법이 가능하며, 또한 컴퓨터를 통하여 마음을 이해하려는 접근법이 가능하다. 인지과학에서는 전자를 연결주의적 접근 $c$(connectionist approach)라 하고, 후자를 기호주의적 접근 $s$(symbolic approach)라고 한다([20][23]). 두 접근은 적절한 사영함수(projection)

$$Brain \xleftarrow{c} Mind \xrightarrow{s} Computer$$

를 찾는 과정으로 볼 수 있다. 1980년대와 1990년대 중반까지는 기호주의적 접근과 연결주의적 접근이 서로 대결의 구도를 이루고 있었다. 그러나 그 후로부터 현재까지는 서로의 장점을 인정하면서 부분적인 연합 또는 새로운 인지시스템을 구축하고자 노력하는 양상이었다. 향후 인지과학의 궁극적인 목표는 두 접근 사이의 동일구조관계(isomorphism)를 밝히는 것이라고 할 수 있다.

## 3 컴퓨터의 수학적 모델

역사적으로는 먼저 기호주의적 접근이 인지과학을 주도하였다. 기호주의적 접근은 순차적 처리(sequential process)를 전제한다. 기호주의적 접근에 의하면, 인지시스템은 순차적으로 정보를 처리하는 시스템으로 설명된다. 이러한 시스템은 기호를 조작하여 인지기능을 수행한다([18]).

기호주의적 접근의 계산모델은 바로 '튜링기계'(Turing Machines)로 불리는 형식시스템(formal system)으로부터 직접적으로 유래되었다([3]). 그런데 튜링기계는 수리논리학자 튜링(Alan Turing)에 의해서 고안된 수학적 시스템이다(On Computable Numbers with an Application to the Entscheidungsproblem, 1936). 튜링은 계산하는 인간에 대한 수학적 모델링을 시도하였고 그때 '계산하는 인간'을 '컴퓨터(computer)'라고 명명했다. 그러므로 튜링이 사용했던 용어 '컴퓨터'는 오늘날 '계산하는 기계'를 지칭하는 것과는 전혀 다른 의미로 사용된 것이었다.

튜링에 의하면 생각하는 인간을 수학적으로 분석하면 세 가지의 요소에 의해서 설명될 수 있다((1) a list of states, (2) a finite alphabet of symbols, (3) a finite list of instructions). 이 요소에 의해서 인간의 생각을 수학적으로 설명할 수 있으며, 그 설명을 모은 것이 튜링기계에 해당된다. 튜링기계의 수학적 정의는 다음과 같다.

정의(Turing Machine). $\sum$를 기호들의 유한집합이라 하고, $B$(Blank를 표시)와 |(Stroke을 표시)는 $\sum$의 원소라 하자. $q_1, q_2, \ldots$를 상태(states)의 기호라고 하자. (단 $q_1, q_2, \ldots$는 $\sum$의 원소가 아니다). 그때에 튜링기계 $TM$은 오원 $(q_i, s, t, \Phi, q_i)$의 유한집합이다. 여기에서 $s$와 $t$는 $\sum$의 원소이고, $\Phi$는 기호 $L$(왼쪽으로 하나 이동) 또는 기호 $R$(오른쪽으로 하나 이동)의 하나이다. 기호 $q_i$는 $i$상태를 나타낸다.

그러므로 튜링기계 $TM$은 자연수 $0, 1, 2, \ldots, n$에 대하여 다음과 같은 함수이다.

$$TM : \{0, 1, 2, \ldots, n\} \times \{B, |\} \to \{B, |\} \times \{L, R\} \times \{0, 1, 2, \ldots, n\}$$

예를 들어, $x + u$를 계산하는 튜링기계 $TM_{x+u}$는 다음과 같이 정의될 수 있다.

$$TM_{x+u} = \left\{ \begin{array}{ccc} (q_1, |, |, R, q_1), & (q_1, B, |, R, q_2), & (q_2, |, |, R, q_2), \\ (q_2, B, |, L, q_3), & (q_3, |, B, L, q_4), & (q_4, |, B, L, q_5), \\ (q_5, |, B, L, q_6), & (q_6, |, |, L, q_6), & (q_6, B, B, R, q_0) \end{array} \right\}.$$

정의(Universal Turing Machine). 리커시브 함수 $F$에 대하여, 다음을 만족하는 유니버설 튜링기계 $UTM$이 존재한다. 임의의 $TM_n$과 임의의 자연수 $n$과 $x$에 대하여,

$$F_{TM_n}(x) = F_{UTM}(n, x)$$

유니버설 튜링기계 $UTM$은 임의의 튜링기계 $TM$을 구현할 수 있는 튜링기계를 의미한다. (이하, 튜링기계는 유니버설 튜링기계를 의미한다.)

튜링은 튜링기계가 계산하는 인간의 수학적 모델이므로, 튜링기계라는 형식시스템을 생각하는 기계라고 보았다. 1948년 수리논리학자 폰노이만(von Neumann)은 튜링기계에 내장된 프로그램을 구상함으로써 자기 증식(self-reproduction)이 가능한 최초의 컴퓨터를 제작할 수 있었다([2]). 기호주의적 접근을 하였던 다수의 인지과학자들은 튜링의 수학적 모델을 물리적으로 구현한 폰 노이만의 컴퓨터를 '생각하는 기계' 로 여겼다. 그들에게는 인류 역사상 처음으로 생각하는 기계, 즉 생각하는 도구를 가지게 된 것을 의미했다. 이 도구를 통하여 인간이 생각하는 방법과 구조를 이해할 수 있을 것으로 기대했다. 이러한 기대는 '물리적 기호 시스템 가설(physical symbol system hypothesis)' 에 잘 반영되어 있다. 이 가설은 기호주의적 접근의 기본적인 전제에 해당하며, 그 내용은 물리적 기호 시스템에는 일반적인 지능 행위를 위한 필요충분조건을 만족하는 수단이 있다는 것이다([18]). 즉 이 가설에 따르면 인간의 인지행위는 튜링기계의 처리과정을 물리적으로 구현한 것과 동등하다. 결국 수리논리학의 연구과정에서 나온 튜링기계는 오늘날 컴퓨터의 모델이 되었을 뿐만 아니라 인지시스템의 주요 모델이 되어, 생각하는 기계의 모델과 연구의 방향을 제공하였다.

인지과학의 역사상 중요한 사건으로 평가되는 'Logic Theorist'(1955)는 뉴웰(A. Newell), 쇼(J. Shaw)와 사이몬(H. Simon)이 만든 최초의 인공지능 프로그램이었는데(Empirical Explorations with the Logic Theory Machine: A Case Study in Heuristics), 러셀(B. Russell)과 화이트헤드(A. Whitehead)의 Principia Mathematica의 제2장을 증명할 수 있는 프로그램이었다. 그 후에 뉴웰과 사이몬이 일반적인 문제까지 확장시켜 일반문제를 해결할 수 있는 프로그램을 제시하고자 한 것이 프로그램 'General Problem Solver'(1961)였다. 생성시스템(Production system), 전문가시스템(Expert system), ACT, SOAR 등은 기호주의적 접근에서 이론적으로 주요한 모델들이다([18]). Oracle 튜링기계, 계산복잡성(complexity) 등 튜링기계와 관련된 다양한 수학적 계산능력에 관하여는 [13]에 잘 소개되어 있다.

## 4 뇌의 수학적 모델

인지과학 내에서 연결주의 접근은 뇌를 통하여 마음을 이해하려는 과학적 연구들을 묶어서 표현한 것이다. 이 접근에서는 뇌의 구조와 계산방식에 의해 인지시스템을 규명하고자 한다. 연결주의에 의하면 인지시스템은 뉴론과 같은 단위들의 연결망과 병렬분산적 동시처리 방식에 의해 설명될 수 있다([17]). 그러므로 연결주의적 계산 시스템은 다음의 표와 같이 튜링기계의 구조와 순차적 계산방식과는 여러 가지 면에서 대조된다.

| | 기호주의적 접근 (Symbolic Approach) | 연결주의적 접근 (Connectionist Approach) |
|---|---|---|
| 수학적 모델 | 튜링기계 | 신경망 |
| 모델의 기초 | 논리 시스템 | 뇌 시스템 |
| 정보처리방식 | 순차적 처리 | 병렬적 처리 |
| 추론 방식 | 연역적 추론 | 귀납적 추론 |
| 기반형 | 규칙 기반형 | 사례 기반형 |

| | | |
|---|---|---|
| 프로그래밍 분량 | 많은 프로그래밍 | 적은 프로그래밍 |
| 학습 | 역동적 | 고정적 |
| 지식 표상 | 명시적 | 암묵적 |
| 알고리즘 성격 | 정확한 일치 | 근사적 일치 |
| 수학적 근거 | 수리논리 | 통계, 확률 |
| 필요 사항 | 전문가 필요 | 데이터 필요 |
| 고장방지 능력 | 약함 | 강함 |
| 사용자 인터페이스 | 화이트 박스형 | 블랙 박스형 |

연결주의 접근이 인정을 받으며 인지과학의 주요한 축으로 자리잡은 것은 1980년대이지만, 역사적으로는 1943년 신경생리학자 머컬로크(Warren McCulloch)와 수학자 피츠(Walter Pitts)에 의해서 뇌의 수학적 모델이 구성되었다(A Logical Calculus of the Ideas Immanent in Nervous Activity, 1943). 이것은 뇌에 대한 연구

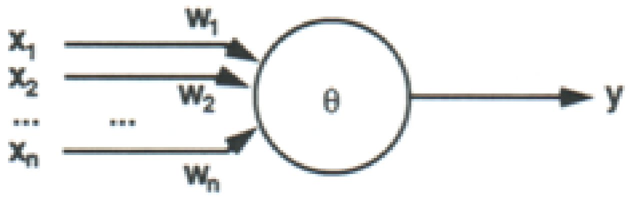
의 역사에서도 획기적인 사건이었다. 이 논문에서 뉴론(neurons)의 수학적 모델로 임계논리단위(threshold logic unit)가 제공되었다. 그들의 연구는 튜링의 연구결과에 자극을 받아 이루어진 것이었으며, 화이트헤드(A. N. Whitehead)와 러셀(B. Russell)의 수리논리학에서 사용된 용어를 채택한 것이었다. 머컬로크와 피츠는 튜링기계의 계산을 구현할 수 있는 신경망(neural networks)의 수학적 모델을 제공해 주었다([9]).

머컬로크-피츠 신경망(McCulloch-Pitts Neural network)은 머컬로크와 피츠가 정의한 뉴론의 연결망이다.

정의(McCulloch-Pitts Neuron). 뉴론은 $m$개의 입력 $x_1, x_2, \ldots, x_m$과 하나의 출력 $u$를 가진 임계논리단위이다. 뉴론은 시간 스케일 $t = 1, 2, 3, \ldots$에 따라 동작하며, 뉴론의 입력값이 임계값보다 클 경우에 발화(fire)한다. 이러한 규칙은 다음과 같다.

$$y(t+1) = 1 \leftrightarrow \sum_i x_i(t) \geq \theta$$

이러한 구상은 보다 발전되어 다음과 같이 뉴론의 수학적 모델로 일반화되었다.




여기에서 $x_i$는 입력(the inputs), $w_i$는 가중치(the weights, 생물학적 뉴론의 시냅스 연결강도(strengths)를 의미함), $\theta$는 임계(the threshold), $u$는 출력(the output)이다.

$$y(t+1) = \left\{ \begin{array}{ll} 1 & \text{if } \sum_i w_i x_i(t) \geq \theta \\ 0 & \text{otherwise} \end{array} \right\}.$$

이러한 뉴론의 네트워크에 대한 수학적 계산모델이 인지과학의 신경망(neural networks)모델이 되었다. 특히 1980년대에는 활성화 함수(activation function)로서의 다양한 임계함수의 종류와 연결패턴을 나타내는 아키텍처(architecture)에 따라 여러 가지의 신경망이 개발되었다. 여기에서 신경망의 아키텍처는 방향을 가진 그래프(a directed graph)와 같다. 기본적인 신경망은 다음과 같이 정의될 수 있다.

정의(Neural Networks). 신경망은 오원구조 $(N, X, Y, C, q)$이다. 여기에서 $N$은 뉴런들의 유한집합, $N$는 입력의 집합, $Y$는 출력의 집합, $C = N \times N$는 시냅스의 연결(connections), $q : N \to F$는 뉴런할당함수이다 ($F$는 뉴런에서의 임계함수들의 집합).

그러므로 신경망에서 $(N \cup X, C)$는 방향을 가진 (혹은 방향을 가지지 않은) 가중치 그래프(weighted graph)이다. 그래서 여러 가지 신경망의 모델은 가중치 그래프의 종류와 활성화 함수의 종류에 따라 구별될 수 있다. 예를 들면, 유명한 최적화 문제였던 여행판매원 문제(traveling salesman problem)를 풀어서 기호주의적 접근을 능가하는 모델로 많은 관심을 모았던 홉필드 신경망(Hopfield Network)은 방향을 가지지 않은 그래프(undirected graph), 다층(multi-layered) 아키텍처, $f : R^n \to \{1, -1\}$ 임계함수 ($R$은 실수의 집합, $n$은 자연수) 등의 특징을 가지고 있었다.

튜링기계와 비교해 볼 때, 신경망의 출력은 뉴런의 활성화(activation)를 설명해주고 주어진 시간단위에서 활성화의 패턴은 신경망의 구성(configuration)에 해당된다. 신경망 내에서 구성이 단기기억(short-term memory)에 해당되면, 가중치의 패턴은 장기기억(long-term memory)에 해당된다. 이러한 구조에서, 가중치의 변화 즉 시냅스 강도의 조정을 통해서 발생되는 것이 바로 신경망에서 '학습'(learning)의 정의이다. 여기에서 가중치를 조정하는 방법이 신경망의 알고리즘에 해당된다. 그러므로 기호주의 접근의 튜링기계와는 달리 연결주의의 신경망 모델은 역동적인 학습을 설명할 수 있다는 새로운 면을 보여주었다.

뇌를 수학적으로 모델링한 신경망은 튜링기계와는 달리 중앙처리장치(CPU)가 전혀 필요하지 않으며, 수많은 뉴론들의 연결로 이루어져 있으므로 고장방지능력(fault tolerance)을 지니고 있다고 평가된다. 신경망의 종류는 매우 다양하다. 예를 들면, 이론적으로 중요한 모델로는 홉필드(Hopfield) 신경망, 해밍(Hamming) 신경망, ADALINE 신경망, 자기조직화(Self-Organizing) 신경망, 다층퍼셉트론(Multi-Layer Perceptron)이 대표적이며, 신경망의 다양한 수학적 계산능력에 관하여는 [24]에 잘 소개되어 있다.

## 5 괴델의 불완전성 정리와 인지과학

인지시스템의 본질과 기능을 해명하고 검증하려는 역사적 흔적은 유클리드(Euclid)의 '공리화'(axiomatization), 아리스토텔레스(Aristotle)의 '논리학'에서 그 기원을 찾아 볼 수 있다. 라이프니츠(G. Leibniz)는 보편적 개념을 표현할 수 있는 언어, 즉 과학의 보편언어(characteristica universalis)를 기호(symbol)라고 보고, 기호의 시스템 내에서 모든 계산을 처리할 수 있을 것으로 기대했다.

부울(G. Boole, *An Investigation of the Laws of Thought*, 1854)은 논리학의 대수화를 체계화했다. 그는 사고의 연산법칙이 존재한다면, 사고에 대한 대수적 계산이 가능하다고 생각했다. 부울의 대수는 현재의 컴퓨터과학을 포함하여 모든 이항시스템과 이항연산의 중요한 기틀이 되었다.

부울과 달리, 프레게(G. Frege, *Begriffsshrift, a formal language, modeled upon that of arithemtic, for pure thought*, 1879)는 논리의 수학화보다는 수학의 논리화를 지향했다. 개념을 명료하게 표기하고자 기호를 사용하여 새로운 논리시스템을 제안했다. 그의 공헌으로 논리학은 (1) 명제함수(propositional function)와 (2) 양화(quantification)를 갖춘 술어논리학(predicate logic)으로 발돋움하여 명제논리 시대를 마감하고 술어논리 시대를 열게 되었다. 프레게는 명제함수 즉 술어와 칸토르의 집합개념을 통하여 모든 수학적 내용을 논리학으로 환원할 수 있다고 생각하였다. 이러한 구상은 러셀과 화이트헤드(Principia Mathematica, 1910-1913)에 의해 계승되었다.

이러한 움직임에 대하여, 힐베르트(D. Hilbert)는 논리적 공리로부터 자유로운 공리를 지닌 형식시스템(axiomatic formal system)을 수학의 기초로 삼아야 한다고 생각했다. 그는 이 형식시스템이 세 가지 조건 (1) 무모순성(consistency), (2) 결정가능성(decidability), 그리고 (3) 완전성(completeness)을 만족하면, 수학의 유토피아를 건설할 수 있을 것이라고 확신했다.

괴델(K. Gödel, 'On Formally Undecidable Propositions of Principia Mathematica and Related Systems I,' 1931)은 제1불완전성정리를 통하여 러셀과 화이트헤드의 논리시스템($PM$) 내에서 결정불가능한 명제가 존재함을 증명하였다. 이는 곧 힐베르트의 프로그램이 달성될 수 없음을 의미했다. 한편, 괴델은 불완전성정리를 증명하면서, 정수 위에서의 인코딩(encoding)과 디코딩(decoding)기법을 소수를 이용하여 완벽하게 구사함으로써 정보이론과 컴퓨터 과학에 중요한 공헌을 남기기도 하였다.

괴델의 제1불완전성정리(1931). 형식시스템 $PM$이 무모순(consistent)일 때, $PM$ 내에서 증명될 수도 없고 반증될 수도 없는 결정불가능한 명제(undecidable propositions)가 존재한다.

괴델의 제2불완전성정리(1931). 형식시스템 $PM$이 무모순(consistent)일 때, $PM$ 자체의 무모순성은 $PM$ 내에서 증명될 수 없다.

인지과학을 위하여 계산하는 기계의 수학적 모델을 제공하게 되었던 튜링의 1936년 논문은 사실 괴델이 증명했던 $PM$의 불완전성이 튜링기계 형식시스템 내에서 정지불가능 문제(unhalting problem)와 수학적으로 동등함을 보이기 위한 것이었다. 괴델이 직접 정의했던 형식시스템(1934)과 재귀함수(general recursive function)는 인지과학을 구성하는 일에 결정적인 공헌을 하게 되었다. 뒤를 이어 나왔던 처치(A. Church)의 람다시스템, 튜링(A. Turing)의 튜링기계, 포스트(E. Post)의 생성시스템(production system)과 같은 형식시스템들과 그 시스템들이 정의하는 계산가능 함수들은 모두 수학적으로 동치인 것으로 증명되었다.([6][7]). 여기에서 함수들의 계산가능성과 함수를 정의하는 형식시스템은 인지과학의 기호주의적 접근에서 기본적 모델이 되었던 것이다. 가령, 포스트의 생성시스템은 인지과학의 성립에 중요한 역할을 하였던 언어학자 촘스키(N. Chomsky, *Logical Structure of Linguistic Theory*, 1955)의 생성문법에 직접적인 영향을 주었다.

괴델의 불완전성정리는 인지과학의 연구에 가장 핵심적인 문제와 직결되어 있었다([1][14][21][29]). 괴델의 정리에 따르면, 인지시스템이 모순이 없는 형식시스템이라면, 인지시스템은 결코 완전할 수 없다는 것을 의미한다. 그렇다면 인공지능의 프로그램과 그 물리적 구현이 제한적이거나 불가능함을 함의하게 된다. 그래서 인지과학의 기호주의적 접근이나 연결주의적 접근이 피해갈 수 없는 과제가 되었던 것이다. 괴델의 정리와 관련하여 이른바 ‘인공지능 논쟁’이 벌어졌다. 이 논쟁은 다음과 같이 두 가지로 범주화될 수 있다.

(1) 괴델의 정리에 의해서 수학자와 동등한 인공지능 시스템은 불가능하다는 주장: 루카스(J. Lucas), 펜로즈(R. Penrose) 등이 대표적 학자이다.

(2) 괴델의 정리에 의해서 (1)의 결론을 얻을 수 없다는 주장: 호프슈태터(D. Hofstadter), 아비브(M. Arbib), 생커(N. Shankar) 등이 대표적 학자이다.

괴델 자신은 다음의 두 명제가 모두 필요한 진리명제라고 생각했다([10]). (A) “수학자의 마음은 어떠한 유한기계도 무한히 능가한다.” (B) “절대 해결될 수 없는 수학적 문제가 존재한다.” 그러므로 괴델에게는 다음의 논리합이 성립한다. (C) “수학자의 마음은 어떠한 유한기계도 무한히 능가하거나(or), 절대로 해결될 수 없는 수학적 문제가 존재한다.” 명제(A)와 명제(B)는 서로 배타적이지 않으며, 논리합(C)에서 양립가능하다. 논리합(C)의 양립가능성은 괴델의 불완전성 정리 이후 수학자와 튜링기계의 동등성에 관한 괴델 자신의 결론이었다. 그러므로 괴델은 어느 편의 손도 들어주지 않은 셈이다([31]). 괴델의 명제에도 불구하고 괴델의 정리를 둘러싼 인공지능의 논쟁은 아직도 계속되고 있다.

## 6 전망

역사적으로 인류에게 수학은 인간의 인지적 활동이 내부에서 개념화되고 형식화되는 과정이었다. 그리고 오늘날 인지과학은 다양한 인지적 과정들이 외부에서 개념화되고 형식화되는 메타과정이다. 지금까지 인지과학의 수학적 기초를 조명하면서, 인지과학의 형성과 발전과정에 나타난 수학적 공헌을 살펴 보았다. 수학은 형식시스템이란 이론적 기반을 제공했고, 다양한 학문의 소통을 가능하게 해주는 언어의 기능을 담당함으로써 인지과학의 형성과 발전을 가능하게 하였다. 특히, 수학자들은 기호주의적 접근에서 핵심적인 수학적 모델이 되었던 튜링기계를 제공하였고 연결주의적 접근에서 핵심적인 수학적 모델이 되었던 신경망을 제공하여, 인지과학의 이론의 양대 축을 이루는 데 결정적인 공헌을 하였다. 튜링과 피츠 외에도 인지과학의 역사에서는 다양한 수학자들이 등장하고 그들의 다양한 공헌들이 있지만, 이 글에서는 인지과학의 가장 근본적 과제와 관련하여 괴델과 그의 공헌이 인지과학 내에서 갖는 의미를 중심으로 논의하였다.

괴델 정리에 의해 제기된 튜링기계와 동등한 시스템의 한계 문제를 둘러싼 논쟁이 계속되자, 인지과학자들은 새로운 차원의 수학적 가능성을 모색하게 되었다. 괴델의 불완전성 정리의 정보이론적 의미는 정지확률의 수(halting probability number) $\Omega = \sum_{u\ halts} 2^{-|p|} (0 < \Omega < 1)$에 의해서 표현가능하다([5]). 여기에서 $p$는 프로그램을 나타내고, $|p|$는 프로그램 $p$의 비트(bits)의 크기를 나타낸다. 정보이론에서의 괴델 정리는 정지확률 $\Omega$가 정수에 대하여 우리에게 계속 복잡성의 새로운 성질, 즉 무작위성(randomness)을 보여준다는 것을 함의한다. 이것은 수학자가 단지 튜링기계와 동등하다는 튜링의 기계주의를 괴델이 비판한 바와 같다([11][30]). 비트에 근거한 기계적 방식에 의해서 수학을 정복할 수 있는 인지시스템은 존재할 수 없다.

0과 1로 구성된 이진(binary) 기반의 시스템에서는 괴델의 불완전성 정리를 피해갈 수 없으므로, 고전논리의 시스템을 넘어서는 새로운 시스템이 필요할 수밖에 없었다. 이에 가장 근접한 조건을 보여주는 대표적인 시스템은 양자논리(Quantum Logic)와 양자계산(Quantum Computing)이다. 괴델의 불완전성 정리는 바로 결정론적 계산보다는 비결정론적 양자계산의 필요성을 함의한다고 할 수 있다([4]). 버코프(G. Birkhoff)와 폰 노이만(J. von Neumann)의 기념비적인 논문(“The Logic of Quantum Mechanics,” 1936)에서 시작된 양자논리는 그동안 다양하게 정의되어 왔다. 일반적으로 양자논리는 힐베르트 공간의 닫힌 부분공간의 격자(lattice)로 해석되었고, 고정된 힐베르트 공간에서 유한한 변수들을 허용하고 $\wedge$(meet), $\vee$(join), $'$(negation)을 사용하는 다양한 방정식들을 의미했다([8]). 양자논리의 중요한 특징은 두 개의 값만을 허용하는 비트가 아니라 세 개의 값을 허용하는 큐비트(qubits)를 바탕으로 하는 상태공간(state space) $C^n$을 전제한다는 점이다.

$C^1$의 양자논리는 고전적인 부울 논리(Boolean logic)과 동등하다. 그러나 $C^2$의 양자논리부터는 분배법칙

이 성립되지 않는다. 그러므로 $C^2$의 양자논리는 표준적인 일차논리시스템(First order logic)과는 전혀 다른 계산을 하는 시스템임을 알 수 있다. 가령, 뇌와 같은 인지시스템을 양자논리시스템으로 본다면, 인지는 수학적으로 유한 차원의 힐베르트 공간을 가진 양자시스템의 양자계산으로 새로이 해석될 수 있다. 인지시스템을 양자적 관점에서 이해하는 접근(quantum cognitive science)은, 기존의 기호주의적 접근과 연결주의적 접근이 보였던 한계를 보완하고, 더 나아가 새로운 차원에서의 이론적 융합가능성을 제공할 수 있을 것으로 기대되고 있다([12][22]).

**REFERENCES**


1. Arbib, M., *Brains, Machines, and Mathematics*, Springer-Verlag, 1987.
2. Aspray, W., “The Mathematical Reception of the Modern Computer: John von Neumann and the Institute for Advanced Study Computer,” Esther Phillips (ed.), *Studies in the History of Mathematics*, The Mathematical Association of America, 1987.
3. Beeson, M., “Computerizing Mathematics: Logic and Computation,” Rolf Herken (ed.), *The Universal Turing Machine: A Half-Century Survey*, Springer-Verlag, 1995.
4. Blaha, S., *The Equivalence of Elementary Particle Theories and Computer Languages: Quantum Computers, Turing Machines, Standard Model, Super String Theory, and a Proof that Gödel's Theorem Implies Nature Must be Quantum*, A Pingree-Hill Research Monograph, 2005.
5. Chaitin, G., *Thinking about Gödel and Turing*, World Scientific, 2007.
6. Davis, M., “Influences of Mathematical Logic on Computer Science,” Rolf Herken (ed.), *The Universal Turing Machine: A Half-Century Survey*, Springer-Verlag, 1995.
7. Davis, M., “Mathematical Logic and the Origin of Modern Computers,” Esther Phillips (ed.), *Studies in the History of Mathematics*, The Mathematical Association of America, 1987.
8. Dunn, J., Hagge, T., Moss, L., and Wang, Z., “Quantum Logic as Motivated by Quantum Computing,” *The Journal of Symbolic Logic*, Vol.70, No.2 (2005), 353–359.
9. Franklin, S., Garzon, M., “Computation by Discrete Neural Nets,” Paul Smolensky, Michael Mozer, David Rumelhart (eds.), *Mathematical Perspectives on Neural Networks*, Lawrence Erlabaum Associates, 1996.
10. Gödel, K., “Some Basic Theorems on the Foundations of Mathematics and Their Implications(1951),” *Collected Works, Vol.III.*, S. Feferman et al. eds., Oxford University Press, 1995.
11. Gödel, K., “Some Remarks on the Undecidability Results(1972),” *Collected Works, Vol.II.*, S. Feferman et al.(eds.), Oxford University Press, 1990.
12. Goertzel, B., *The Structure of Intelligence: A New Mathematical Model of Mind*, Springer-Verlag, 1993.
13. Herken, R.(ed.), *The Universal Turing Machine: A Half-Century Survey*, Springer-Verlag, 1995.
14. Hofstadter, D., *Gödel, Escher, Bach: An Eternal Golden Braid*, Basic Books, 1979.
15. Marr, D., *Vision*, W. H. Feeman and Company, 1982.
16. McCarthy, J., “Mathematical Logic in Artificial Intelligence,” Stephen Graubard (ed.), *Artificial Intelligence Debate: False Starts, Real Foundations*, The MIT Press, 1990.
17. McClelland, J., Rumelhart, D., and Hinton, G., *Parallel Distributed Processing*, Vol.1. The MIT Press, 1986.
18. Newell, A., 차경호 옮김, 통합인지이론(*Unified Theory of Cognition*), 아카넷, 2002.
19. Norman, D.(ed.), *Perspectives on Cognitive Science*, Ablex Publishing Corporation, 1981.
20. Graubard, S.(ed.), *Artificial Intelligence Debate: False Starts, Real Foundations*, The MIT Press, 1990.
21. Penrose, R., “On the Physics and Mathematics of Thought,” Rolf Herken (ed.), *The Universal Turing Machine: A Half-Century Survey*, Springer-Verlag, 1995.
22. Penrose, R., *The Large, the Small and the Human Mind*, Cambridge University Press, 1997.

23. Posner, I.(ed.), *Foundations of Cognitive Science*, The MIT Press, 1989.

24. Smolensky, P., Mozer, M., and Rumelhart, D.(eds.), *Mathematical Perspectives on Neural Networks*, Lawrence Erlabaum Associates, 1996.

25. 미국립연구회의, "인지과학," 과학과 기술, 대한교과서, 1985.

26. 이정모, 인지과학, 성균관대학교출판부, 2009.

27. 이정민 외, 인지과학, 태학사, 2001.

28. 한광희 외, 인지과학 : 마음·언어·기계, 학지사, 2000.

29. 현우식, "계산가능성이론 형성에서의 Church's Thesis와 Turing's Thesis," 한국수학사학회지 11 (1998), No.1, 19–26.

30. 현우식, "Gödel's Critique of Turing's Mechanism," 한국수학사학회지 17 (2004), No.4, 27–36.

31. 현우식, "Gödel's Disjunctive Conclusion," 한국수학사학회지 13 (2000), No.1, 137–141.

32. 홍성사, "Extensions and Mathematical Extensions," 한태동교수 고희기념논문집, 1995.

33. 홍성사·홍영희, "순서와 위상구조의 관계," 한국수학사학회지 10 (1997), No.1, 19–32.

34. 홍성사·홍영희, "Categorical Topology의 역사," 한국수학사학회지 10 (1997), No.2, 11–23.

35. 홍영희, "격자론의 기원," 한국수학사학회지 12 (1999), No.2, 15–23.