마음과 정보처리형식체계의 논리적 동치성: 괴델의 선언결론과 불완전성 정리를 중심으로*

현 우 식
연세대학교 대학원 인지과학

Equivalence of Mind and Information Processing Formal System: Gödel's Disjunctive Conclusion and Incompleteness Theorems

Woo-Sik HYUN
Department of Cognitive Science,
The Graduate School, Yonsei University

요 약

마음과 기계의 관계에 대한 Gödel의 선언결론(disjunctive conclusion)은 마음과 정보처리형식체계의 논리적 동치성을 함의하고 있다. 그리고 Gödel의 불완전성 정리(Incompleteness Theorems)에 따르면 마음과 정보처리형식체계의 논리적 동치성은 무모순이며, 동치성 반증의 이론은 그 모델을 가질 수 없다.

1. 서 언

형식체계 내에서 참이면서 증명불가능한 특별한 논리식이 존재한다는 괴델(K.Gödel)의 증명은 수리논리에서 비롯된 것이지만 컴퓨터 과학 이론의 중요한 위치에 있다. 일반적으로 괴델의 증명은 튜링기계의 Halting Problem과 동치인 것으로 인식되어 왔다. 그래서 루틴(routine)을 따라 수행할 수 있는 문제 중 어떤 특별한 형식의 문제는 루틴에 의해 수행될 수 없다는 증명이 '인간의 마음은 디지탈 컴퓨터 이상' 임을 보여 주는 이론적 증거로 간주되기도 하였다.

마음과 기계의 논리적 동치성 문제에 관하여 J.R.Lucas [1961]가 괴델의 불완전성 정리를 적용하여 마음과 기계가 동등할 수 있다는 기계론(mechanism)이 옳지 않다는 것을 주장한 이후 이 주제는 논리학, 인공지능(AI), 그리고 인지과학의 주요 논쟁의 대상이 되어 왔다. 사실 이러한 이론적 작업에 관한 비판적 검토는 '괴델 정리 적용'의 타당성과 '기계론 반증(소위, anti-mechanism)'의 타당성으로 엄밀하게 구분되어 진행되어야 한다.

* 이 연구는 대우재단의 연구비지원에 의해 이루어 진 것임

이 논문에서는 마음과 기계의 관계에 대한 괴델의 후기 선언결론(disjunctive conclusion)을 분석하여 그 동치성과 괴델의 불완전성 정리(Incompleteness theorems)의 관계를 다루며, 또한 괴델의 불완전성 정리에 의해서 마음과 정보처리형식체계(Information Processing Formal System)의 논리적 동치성을 반증(disproof)하는 주장이 인정될 수 없음을 증명하고자 한다. 이 글에서의 '정보처리체계(IPS)'는 튜링기계(Turing Machine, A.M.Turing,1950), 컴퓨터, 일반적 상징조작이 가능한 정보처리형식체계(A.Newell & H.A.Simon, 1972), 혹은 이해 능력과 인지 상태를 보유한 강인공지능(strong AI, J.R.Searle,1980)과 논리적으로 대치가능한 용어로 사용할 수 있다. 연구의 범위는 이론적 영역 내에서의 논리적 동치성 문제에 집중될 것이다.

2. 괴델의 제1불완전성 정리와 선언결론

2.0 괴델의 제1불완전성 정리 Gödel's 1st Incompleteness Theorem

체계 S가 무모순이면 결정불가능한 논리식 G가 존재한다.

S: *consistent* $\rightarrow \exists$ *formula* $G(S\nvdash G \wedge S\nvdash \sim G)$

2.1 괴델은 1951년 12월 26일 브라운 대학에서 열린 미국 수학회(the American Mathematical Society)가 주최한 제25회 깁스 기념 강연(Josiah Willard Gibbs Lecture) 에서 "수학의 기초와 함의에 관한 몇 가지 기본적 공리들(Some basic theorems on the foundations of mathematics and their implications)"을 발표하였다. 이 강연에서 괴델은 마음과 기계의 관계설정에 관한 다음의 선언결론(disjunctive conclusion)이 피할 수 없음을 표명하였다.

> "수학의 명확한 공리들이 유한한 규칙 내에서 결코 구성될 수 없다는 점에서 수학은 불완전하거나, 즉 인간의 마음이 (순수 수학의 영역 내에서도) 임의의 유한기계의 능력을 무한히 능가하거나 또는 절대 해결 불가능한 특별 유형의 디오판틴 문제들이 존재한다.[K.Gödel,1951,]"

수학의 명확한 공리들이 유한한 규칙 내에서 결코 구성될 수 없다는 점은 이미 괴델의 제1불완전성 정리에 의해 증명된 바 있다. 수학 내에서 수학의 완전성(completeness)과 무모순성(consistency)은 동시에 유지될 수 없음이 메타수학으로 증명되었으므로 선언(disjunction) 전반부의 내용은 불완전성 정리에 의해 명확히 증명된 것이다. 그러므로 메타수학을 포함한 인간의 마음은 수학의 무모순성을 구현한 임의의 유한기계에 대하여 무한히 능가할 수 있음이 증명된다.

그러나 선언의 후반부에서 괴델이 사용한 '절대(absolutely)' 라는 표현은 임의의 공리적 체계(axiomatic system)에서 뿐만 아니라 인간의 마음이 생각할 수 있는 임의의 수학적 증명에서도 동등하게 그 한계가 적용된다는 의미였다. 그러므로 (2)가 의미하는 동치성이란 사실상 그 강조점과 기준이 마음과 기계의 각각의 한계(limitation)에 있는 것이다. 동일한 한계에 대한 동치성인 것이다. 그래서 '절대 해결불가능한 특별 유형의 디오판틴 문제'에서는 기계와 마음에 공통적으로 발생하는 결정불가능한 논리식(undecidable formula), 즉 괴델의 제1불완전성 정리 Gödel's 1st Incompleteness Theorem[1931]에 의해 증명된 다음의 괴델문장(G)의 존재가 전제된다.

$\forall HM = FM(HM = FM$: *consistent* $\rightarrow \exists G((\ HM = FM \vdash G) \wedge (HM = FM \vdash \sim G))$

괴델의 선언결론을 형식언어로 표현한다면 다음과 같다. 인간 마음을 HM, 유한기계를 FM이라 할 때,

$(1)\exists HM, \forall FM(HM \rangle FM) \vee (2)\exists G(HM \nvdash G \wedge FM \nvdash G)$

결국 괴델의 결론은 마음이 기계가 할 수 없는 일을 해낼 수 없다면, 마음과 유한기계의 (결정불가능한 문제라는 한계에 있어서) 논리적 동치성을 선언적으로 인정하고 있다(2). 따라서 (1),(2)로 부터 다음의 명제를 얻을 수 있다.

명제1 마음과 정보처리체계에서 증명불가능한 괴델문장G이 존재하지 않는다면 모든 정보처리체계를 능가하는 마음의 존재를 증명할 수 없다.

$\vdash \exists HM, \forall IPS(\exists G(HM \nvdash G \wedge IPS \nvdash G \rightarrow HM \rangle IPS)$

증명> G가 마음과 정보처리체계에서 증명불가능한 문장이라 하자.
i)마음과 정보처리체계가 무모순이면 결정불가능한 문장 G의 존재를 증명할 수 있다.
ii)마음과 정보처리체계가 모순이라면 결정불가능한 문장 G의 존재를 증명할 수 있다.
어느 경우에나 G가 존재한다.
따라서 모든 정보처리체계를 능가하는 마음의 존재를 증명할 수 없다.
그리고 모든 정보처리체계를 능가하는 마음의 존재를 증명할 수 있다.(모순)
그러므로 마음과 정보처리체계에서 증명불가능한 괴델문장 G가 존재하지 않는다면 모든 정보처리체계를 능가하는 마음의 존재를 증명할 수 없다.

2.2 마음과 물리적으로 구현된 형식체계의 관계에 대한 또 다른 형태의 괴델의 선언결론은 G.Kreisel의 미간행물「논리학과 논리학자들에 관하여」*About Logic and Logicians Vol.II.*[1993]에 나타나 있다. Kreisel은 괴델이 대화 중에 언급한 말을 다음과 같이 전하고 있다[Kreisel,p.154].

"마음이 기계적이지 않거나, 수학(사실, 산술)은 우리가 구성한 것이 아니다"

여기에서 '기계적(mechanical)' 이란 컴퓨터 프로그램화될 수 있음을 의미하며, '구성한 것(construction)'이란 우리가 알 수 있는 속성들의 집합, 즉 우리가 아는 세계를 의미한다. 그러므로 다른 용어로 표현한다면 다음과 같다.

"마음 M이 기계C 화 될 수 없거나, 또는 수학 m이 우리가 아는 세계 KW에 포함되지 않는다."

$(1)M \neq C \vee (2)m \notin KW$

그러므로 마음이 기계화될 수 있다면 수학은 우리가 아는 세계에 포함되지 않는다.

$M = C \rightarrow m \notin KW$

따라서 우리는 다음의 명제를 얻을 수 있다.

명제2 수학m이 우리가 아는 세계 KW에 포함된다면 마음HM과 정보처리체계IPS가 동치가 아님을 증명할 수 없다.

$\vdash (m \in KW \rightarrow HM \neq IPS)$

증명> Peano Arithmetic이 수학에 포함된다면, 괴델 제1불완전성 정리에 의해 수학 내에는 결정불가능한 괴델논리식Gödel formula이 존재한다. 이 식은 참이면서 증명불가능하다. 그러므로 수학은 우리가 아는 세계에 포함될 수 없다. 만약 수학 m이 우리가 아는 세계에 속한다면 KW, 마음과 정보처리체계의 동치는 참이다. 따라서 마음과 정보처리체계의 동치성 반증은 증명될 수 없다.

3. 괴델의 제2불완전성 정리 그리고 마음과 정보처리체계의 동치성

3.0 괴델의 제2불완전성 정리 Gödel's 2nd Incompleteness Theorem

체계S가 무모순이면 그 체계는 자신의 무모순성을 증명할 수 없다.

S: *consistent* $\rightarrow S \nvdash Con(S)$

3.1 마음과 정보처리체계의 동치성 반증의 논리

괴델의 정리가 바로 인간이 정보처리체계와 같을 수 없음을 보여준다는 논증은 참에 대한 인지능력을 기준으로 전개된다고 할 수 있다. 임의의 정보처리형식체계 내에서는 증명할 수 없는 참의 문장(괴델문장)이 존재하지만 인간의 마음은 언제나 그것이 참임을 인지할 수 있다는 것이다. 따라서 정보처리체계는 인간의 마음과 같을 수 없다는 것이다. 이를 괴델의 제2불완전성 정리와 관련하여 루카스의 동치성 반증에 따라[Lucas,1961] 마음과 정보처리체계의 동치성 반증으로 재구성하면 다음과 같다.

괴델의 불완전성 정리를 이용한 루카스의 논증은 아래와 같다.

$HM^* = \{ f \mid HM \models f \wedge HM \vdash f \}$, $IPS^* = \{ f \mid IPS \vdash f \wedge IPS \nvdash f \}$ 라 하면

인간의 마음이 정보처리체계를 포함한다면 정보처리체계의 무모순성이 참임을 알 수 있고

HM^*은 정보처리체계의 무모순성을 포함한다. 루카스는 f를 괴델문장으로 설정하였다.

1) $IPS \subset HM \rightarrow HM \models Con(IPS) \wedge Con(IPS) \in HM^*$

그러나 괴델 제2불완전성 정리에 의해 정보처리체계는 자신의 무모순성을 포함할 수 없다.

2) $Con(IPS) \notin IPS^*$

3) $HM^* \supset IPS^* \rightarrow HM \supset IPS$ 그래서 정보처리체계는 인간의 마음과 같을 수 없다는 것이다.

그러나 괴델의 제1불완전성 정리의 의미론적 표현은 바로 $\exists f\ (HM \models f \wedge HM \nvdash f)$ 이다.

사실상 f는 괴델문장이 될 수 없고 루카스 식의 적용은 잘못된 것이다.

3.2 마음과 정보처리체계의 동치성

만일 루카스가 괴델문장을 올바로 적용하였다고 가정하자. 그래서 참인 괴델문장이 정보처리체계에는 추가될 수 없지만 인간의 마음에 추가될 수 있다면 인간의 마음이 정보처리체계를 능가한다는 것이다. 그러나 이러한 주장도 괴델의 제2불완전성 정리에 의해 반증된다.

정리1 정보처리체계에서는 증명할 수 없는 괴델문장이 있을 때, 이 괴델문장을 추가한 마음이 무모순이라는 것은 마음HM 이 괴델문장의 부정 ~G을 증명할 수 없다는 것과 같다.

$HM + \{G\}$: *consistent iff* $HM \nvdash {\sim}G$

증명> ←) 1) $HM \vdash {\sim}G$

2) $HM + \{G\} \vdash {\sim}G$

3) $HM + \{G\} \vdash G$

4) $\therefore HM + \{G\}$: *inconsistent*

→) 1) $HM+\{G\}$: *inconsistent*

2) $\exists x, HM+\{G\} \vdash x \wedge \sim x$

3) $HM \vdash G \rightarrow x \wedge \sim x$ (deduction)

4) $\therefore HM \vdash \sim G$

그러므로 정보처리체계에서 증명하지 못하는 참의 문장이 마음에 포함되어 있다고 해도 그것이 마음의 우위를 증명해 주지는 못한다. 즉 마음과 정보처리체계의 동치성을 반증하지 못한다.

정리2 마음과 정보처리체계의 동치를 부정하는 이론 내에서는 그 모델의 존재를 증명할 수 없다.

$HM \neq IPS \rightarrow HM \neq IPS \nvdash (M \models (HM \neq IPS))$

증명>

마음이 정보처리체계의 무모순성을 아는 것이 무모순이면 $Con(HM \models Con(IPS))$

그에 대한 모델을 가진다. $M \models (HM \models Con(IPS))$ 따라서 HM<M라 하자.

그리고 만일 IPS<HM 이고 HM<M 이면 IPS<M이다.

여기에서 집합 S를 $M \models n \notin S_n$ 를 만족하는 모델 M의 존재성을 나타내는 모든 수들의 집합이라 하자

g를 S의 괴델 수라 하고, A를 "$g \in S$"인 문장이라 하면

집합 U($IPS, HM, M \in U$) 안에서 다음이 증명된다.

1) $U \vdash (A \leftrightarrow \exists M(M \models \sim A))$

2) 만일 $M \models A$ 이면, $(M \models A \leftrightarrow \exists HM < M(HM \models \sim A))$이고

3) $M \not\models A$ 이면, HM<M는 참이다.

4) U가 무모순이면 U는 모델을 가진다.

5) 그러므로 임의의 M에 대하여 HM<M의 모델 M_{HM}이 존재한다.

6) 따라서 같은 방법으로 IPS<HM인 모델 M_{IPS}가 존재하므로 모순이다.

결국 마음과 정보처리체계가 동치가 아니라면 그 모델의 존재를 증명할 수 없다.

정리3 마음HM과 정보처리체계IPS에 증명불가능한 문장이 존재한다면 HM=IPS는 무모순이다.

$\forall HM, IPS(\exists G(HM \wedge IPS \nvdash G) \rightarrow HM = IPS$: *consistent*)

증명> $F=\{G \mid HM \wedge IPS \vdash G\}$라 하자

1) $HM = IPS$: *inconsistent*

2) $(HM = IPS) \vdash F \wedge \sim F$

3) $\therefore \forall G(HM \wedge IPS \vdash G)$

그러므로 동일한 상태(state)에서 마음과 정보처리체계에서 동시에 증명불가능한 문장이 존재한다면 마음과 정보처리체계의 동치성은 무모순이다.

4. 결 어

괴델의 의도와는 달리 E.Nagel & J.R.Newman[58], J.R.Rucas[61], R.Penrose[89]등의 학자들은 괴델의 불완전성 정리를 마음과 기계의 동치성을 부정하는 도구로 사용하였다. 그러나 괴델의 선언결론에 나타난 불완전성 정리의 의미는 인간의 마음과 정보처리형식체계가 동치라면 절대 해결불가능한 문제가 존재한다는 것이다. 그리고 제2불완전성 정리에 따르면 동치성 반증에 대한 모델의 존재를 증명할 수 없고, 증명불가능한 괴델문장의 존재가 동치성의 무모순성을 함의하기 때문에 동치성 반증의 주장은 타당한 것으로 유지될 수 없다. 그러나 이러한 결과가 곧 불완전성 정리가 마음과 정보처리체계의 논리적 동치성을 증명을 보장해 준다는 의미는 아니다. 오히려 괴델의 불완전성 정리는 동치성 반증으로부터 독립적인 위치에 있다고 할 수 있다. 그래서 괴델의 불완전성 정리는 마음과 정보처리체계의 논리적 동치성의 닫혀 있는 한계이자 동시에 열려 있는 가능성의 증거이다.

참고문헌

Boden,M.A.(ed)[1990] *The Philosophy of Artificial Intelligence,* Oxford: Oxford Univ. Press

Gödel,K.[1931] "On Formally Undecidable propositions of Principia mathematica and related systems I," *From Frege to Gödel.* J.van Heijenoort(ed),Cambridge:Harvard University Press

[1951] "Some basic theorems on the foundations of mathematics and their implications" ,*Kurt Gödel Collected Works Vol.III: Unpublished Essays and Lectures[1995]* ,S.Feferman(eds) N.Y:Oxford University Press

Jech,T.[1994] "On Gödel's Second Incompleteness Theorem," *Proceeding of The American Mathematical Society* Vol.121, 1

Kim,S.M.[1995] "Turing-computability and artificial intelligence: Gödel's incompleteness results," *Kybenetes* 24,6:57-63

Kreisel,G.[1993] *About Logic and Logicians Vol.II.* selected and arranged by P.Odifreddi, Manuscript

Lucas,J.R.[1961] "Minds, Machines, and Gödel," Philosophy 36:112-127

Nagel,E.& Newman,J.R. [1958] *Gödel's Proof.* N.Y.: New York University Press

Newell,A. & Simon,H.A.[1972], *Human Problem Solving.* Englewood Cliffs, N.J.:Prentice Hall.

Penrose,R[1989] *The Emperor's New Mind.* N.Y.: Oxford University Press

Searle,J.R.[1980], "Minds, Brains, and Programs" *The Behavioral and Brain Science* 3(1980):417-457

Shankar,N.[1994] *Metamathematics, Machines, and Gödel's Proof.* N.Y.:Cambridge University Press

Turing,A.M.[1950], "Computing Machinery and Intelligence" *Mind* LIX, No.2236

Yu,Q.[1992] "Consistency, Mechanicalness, and The Logic of The Mind," *Synthese* 90:145-179