인 지 과 학
Korean Journal of Cognitive Science
Vol.6, No. 3(1995)

튜링 기계로서의 마음과 괴델의 정리

Gödel's Theorem and Mind as Turing Machine

선우환†
Hwan Sunwoo

요 약

루카스의 이른바 괴델 논변에 의하면, 괴델의 정리는 기계론 논제 즉 인간 인지 체계가 튜링 기계라는 논제를 반박한다. 이 논문에서 필자는 이 논변이 성공적이지 못하다는 것을 보이려고 한다. 그러나 필자는 또한 괴델 논변에 대한 기존의 많은 반론들 역시 받아들일 만하지 않다는 것을 주장한다. 그리고 나서 필자는 괴델 논변에 대한 "일관성" 반론을 강화한다. 그렇게 해서 얻어진 필자의 딜레마 반론에 의하면, 괴델 논변은 (1) 우리가 "전반적" 진리 개념을 가질 경우 거짓 전제를 가지고 (2) 우리가 그러한 진리 개념을 갖지 않을 경우 진술될 수 없으므로, 어떤 경우이든 성공적이지 못하다.

주제어 괴델의 정리, 튜링 기계, 기계론, 기능주의, 루카스

ABSTRACT

According to a well-known argument (so-called the Gödelian argument) proposed by Lucas, Gödel's theorem refutes the thesis of mechanism, that is, the thesis that human cognitive system is no more than a Turing machine. The main aim of this paper is to show

† 프린스턴대학교 철학과
Department of Philosophy, Princeton University, Princeton, NJ 08544, USA
e-mail:hsunwoo@phoenix.Princeton.EDU

that this argument is not successful. However, I also argue that many pre-existing objections (by Benacerraf, Slezak, Boyer, Hofstadter etc.) to Gödelian argument are not satisfactory, either. Using Tarski's theorem, I then strengthen what I call the consistency objection to Gödelian argument. In my dilemma objection obtained, Gödelian argument doesn't work because the argument has a false premise if we have the concept of global truth and the argument cannot be stated if not.

Keywords Gödel's Theorem, Turing Machine, Mechanism, Functionalism, Lucas.

1. 들어가는 말

인간은 일종의 기계인가? 인간 마음의 활동은 일정한 프로그램에 따른 정보 처리 과정인가? 계산 기계는 인간 인지 체계의 적절한 모델이 될 수 있는가?

이런 물음(들)은 인간 마음의 본성에 대한 근본적인 물음(들)이다. 이런 물음에 대해서, 인간은 할 수 있지만 기계는 할 수 없는 어떤 일이 있다는 근거를 댐으로써, 부정적인 대답을 하려는 여러 논변들이 존재한다.[1] 그런데, 그러한 논변들 중에서도 인간 마음의 문제와 아무 상관없어 보이는 하나의 메타수학적 정리, 즉 괴델(Gödel)의 불완전성 정리로부터 제기된 한 논변은 상당히 흥미롭다.

루카스(Lucas)에 의해 제시되고 그에 의해 끈질기게 옹호되고 있는 이 논변에 따르면, 어떠한 (일관적인) 기계도 자기가 구현하는 형식체계의 괴델 문장(괴델 정리의 증명에서 핵심적으로 사용되고 있는)이 참임을 알 수 없지만 인간은 그것이 참임을 알 수 있다고 한다. 따라서 인간은 기계일 수 없다는 것이다.

이 글의 목적은, 루카스의 이 이른바 괴델 논변(Gödelian argument)이 성립하지 않는다는 것을 보이려는 데에 있다. 이미 루카스의 논변에 대해 수없이 많은 반론들이 쏟아져 나왔다. 그러나 그러한 반론들 모두가 루카스에 대해 공

1) 예를 들어, 인간은 말을 할 수 있는데 기계는 말을 할 수 없다(Descartes); 인간은 언어에 대한 의미론적 이해를 할 수 있는데 기계는 그렇지 못하다(Searle); 인간은 창조성을 가지는데 기계는 그렇지 못하다; 등등의 논변들이 있다.

정한 반론이었던 것은 아니다. 사실 그 중 적지는 않은 것들이 논점을 벗어난, 혹은 부당한 반론들이었다. 필자는 그런 반론들 중 그래도 가장 설득력 있어 보이고 대표적인 몇 가지 유형의 반론을 살펴보고 그것들이 어떤 점에서 부당한지, 그리하여 그것들이 어떻게 극복될 수 있는지 보겠다(3절). 그리고 나서 괴델 논변에 대한 반론들 중 필자가 생각하기에 올바른 노선에 있다고 여겨지는 유형의 반론-필자는 그 유형에 속하는 반론들을 '일관성 반론(consistency reply)'이라고 부를 것이다-을 살펴 보고 그 반론을 옹호할 것이다(4절). 필자가 이 논문에서 행하려는 가장 중요한 일은, 종래의 일관성 반론의 연장선 상에서 괴델 논변에 대해 보다 결정적인 새로운 반론을 구성하는 일이다. 그러한 반론을 구성하기 위해 필자는 또 다른 메타수학적, 메타논리학적 정리인 타르스키(Tarski)의 정리를 우리의 논의에 도입할 것이다(5절).

이 글의 목적이 괴델 논변을 비판하는 데에 있으므로, 즉 괴델의 정리가 (인간) 기계론을 부정하는 함축을 가진다는 주장을 비판하는 데에 있으므로, 기계론 자체를 주장하는 것은 이 글의 범위를 벗어나는 일이다. 필자가 주장하려는 것은 단지, 괴델의 정리 혹은 괴델의 증명은 기계론이 거짓임을 보여 주지 않는다는 것이다.

2. 기계론 논제와 루카스의 괴델 논변

기계론(mechanism)이 주장하는 바는 한 마디로, 인간(마음)이 일종의 기계라는 것이다. 여기서 '기계'라는 말을 할 때 그것의 가장 중요한 성격은 다음과 같은 것이다: 즉, 기계에는 일정한 규칙들이 있고, 기계의 모든 상태는 그 기계의 이전 상태들(과 입력들)로부터 그 규칙들에 의해 결정된다는 것이다. 기계의 이러한 성격은, 규칙 표상이 명시적으로 나타나는 경우와 그렇지 않은 경우 양쪽에 있어 모두 가능하다.

이런 성격을 갖지 않은 기계가 있을 수 있는가? 그런 기계가 있다면, 그런 기계는 우리 논의의 대상이 아니다. 즉, 여기서 대상으로 하는 기계론은, 인간(마음)이 위에서 기술된 성격을 갖는 기계라는 주장이다. 이런 성격의 기계는 모두 튜링 기계(Turing machine)의 사례이거나 튜링 기계에 의해 모의될 수 있으므로, (인지적인 측면과 관련해) 기계론의 중요한 함축은 다음과 같이 표현될 수 있다. (튜링 기계에 대해서는 Boolos and Jeffrey(1980)를 참조)

인간의 인지과정은 튜링 기계의 계산과정과 동등하다.

이를 함축하는 기계론은 '기계'라는 말을 느슨한 의미로 사용하는 기계론보다는 강한 형태

의 기계론이다. (따라서 필자는 이런 강한 의미에서의 기계론조차도 괴델 논변에 의해 반박되지 않는다는 것을 보여야 할 것이다).

기계론은 심리철학에서 사실상 기능주의와 인지주의의 공통된 가정이기도 하다. 전통적 AI 역시 이 가정에 기초하고 있다고 할 수 있다. 그러므로 기계론이 반박된다면 심리철학에서의 흥미로운 많은 생각들이 손상을 입을 것이다.

괴델의 정리가 기계론을 반박한다는 논변이 루카스(Lucas.1961)에 의해 제시되었다. 괴델의 (제1불완전성) 정리는 다음과 같은 내용이다.(G ödel,1931)

산술을 포함하면서 오메가-일관적인 어떠한 형식체계도 불완전하다.

'형식체계가 불완전하다' 라는 말은 일반적으로 두 가지 다른 의미로 사용된다. 하나(불완전성1)는 '그 형식체계에서 증명할 수도 반증할 수도 없는 문장이 존재한다' 는 것이고, 또 다른 하나(불완전성2)는 '참이면서 그 형식체계에서 증명불가능한 문장이 존재한다' 는 것이다. 그런데 실제로 괴델은 산술을 포함하는 형식 체계에 대해 이 두 가지 의미의 불완전성을 모두 증명했다. 산술 체계의 불완전성1은 오메가-일관성 가정을 요구하지만[2] 산술 체계의 불완전성2는 그보다 약한 가정인 일관성 가정만을 요구한다. 따라서 괴델은 다음과 같은 내용을 (제1불완전성 정리의 일부분으로서) 증명한 셈이다.

산술을 포함하면서 일관적인 어떠한 형식체계에도, 참이면서 그 체계에서 증명불가능한 (산술)문장이 존재한다.

불완전성 정리를 증명하기 위해 괴델은 형식체계의 모든 식과 식들의 나열에 고유한 자연수를 부여하는 방식을 개발했다. 이렇게 부여된 자연수를 그 식 또는 식들의 나열의 '괴델수(G ödel number)' 라고 한다. g(F)를 F의 괴델수라고 하자. 그러면, 식들의 나열 F1,...,Fn이 식 Fn의 증명일 경우 오직 그 경우에 g(F1,...,Fn)이 g(F)에 대해 성립하는 수적 관계(numerical relation)가 존재한다는 것이 보여질 수 있다. 이 수적 관계를 증명 관계라고 부르자. 더 나아가 증명 관계가 산술을 포함하는 체계 내에서 표현가능(expressible)하다는 것까지도 보여진다.[3] 증명 관계를 표현하는 술

2) 실제로는 롯서가 불완전성1도 일관성 가정만을 써서 증명할 수 있는 방법을 보여 주었다.(Rosser,1936)

3) x와 y 간에 관계 R이 성립하면 B(x,y)가 체계 속에서 증명가능하고, x와 y 간에 관계 R이 성립하지 않으면 -B(x,y)가 체계 속에서 증명가능할 경우, 술어 B가 관계 R을 그 체계

어를 B라고 하자. 그러면 g(G)를 괴델수로 가지는 다음 문장 G를 구성할 수가 있다.

G : (x)-B(x,g(G))

이 문장은 이 문장 자신이 증명불가능할 경우 오직 그 경우에 성립하는 산술적 사실을 진술한다. 즉 이 문장은 직관적으로 표현해 다음의 것을 의미한다고 할 수 있다.

이 문장은 증명불가능하다.

이 문장을 괴델 문장(Gödel sentence)이라고 한다. 괴델 문장 G가 증명불가능하다는 것은 비교적 쉽게 보여질 수 있다. G가 증명가능하다고 하자. 그러면 증명 관계가 표현가능하므로, G의 증명가능성을 표현하는 문장이 증명가능할 것이다. 그런데 G의 증명가능성을 표현하는 문장은 바로 G의 부정이다. 즉 -G가 증명가능하게 된다. 그리하여 그 체계는 비일관적임이 따라나온다. 그러므로 체계가 일관적이라면 그 체계에서 G는 증명불가능하다. G가 G 자신이 증명불가능하다는 것을 표현하고 있고 실제로 G가 증명불가능하므로 G는 또한 참이기도 하다. 즉 G는 참이면서 증명불가능하다.

루카스는, 괴델의 증명이 기계론을 다음과 같은 방식으로 반박한다고 논변한다. 즉, 내(인간 마음)가 기계(특히, 정리를 증명할 줄 아는 기계)라고 하자. 그러면 그 기계의 정리 증명 과정은 일정한 규칙들에 따라 이루어질 것이다. 그러면 그것에 대응해 그 기계의 규칙들(과 초기 조건들)로 이루어진 형식 체계(T)가 존재할 것이다. (그것을 그 기계가 구현하는 형식 체계라 부르기로 하자). 괴델의 방법에 의해 나는 T의 괴델 문장 G를 구성할 수 있고 G가 T에서 증명불가능하다는 것을 알 수 있다. 그리고 G가 표현하는 바에 의해 그것이 참이라는 것도 알 수 있다. 결국 T를 구현하는 기계는 G를 증명할 수 없는데 나는 그것이 참임을 알 수 있는 것이 된다. 그러므로 나는 나와 동일하다고 가정된 그 기계와 동일할 수 없다. 즉 나는 기계가 아니다.

(루카스의 논변은 보다시피 귀류법적 방식으로 전개되었다. 즉, 기계론이 참이라고 가정하고 기계론을 부정하는 결론을 이끌어내었다. 이 논변의 귀류법적 성격이 뒤에서 다시 논의될 것이다).

이같은 논변은 이미 네이글과 뉴만(Nagel and Newman,1958)에 의해 간단한 형태로 개진된 바 있을 뿐만 아니라 괴델 자신도 이런 종류의 함축을 받아들이고 있었던 것으로 보인다(Wang,1980). 또한 비교적 최근에는 펜로

속에서 표현한다고 한다. 그리고 관계 R을 그 체계 속에서 표현하는 술어가 존재할 경우, 관계 R은 그 체계 속에서 표현가능하다고 말한다.

즈(Penrose,1989)도 이 논변을 받아들여 강한 AI를 공격하는 한 무기로 사용하고 있다. 실제로 이 논변이 가지는 설득력은 단순히 무시될 수 있는 성질의 것이 아니다.

3. 괴델 논변에 대한 반론들

3-1. 첫번째 반론의 유형: "기계도 다른 기계를 괴델화할 수 있다."

루카스의 괴델 논변에 대해 제기되는 중요한 반론들의 유형 중의 하나는, 기계도 다른 기계의 괴델 문장이 참임을 알 수 있다는 점을 지적하는 것이다(예를들어, Benaceraf,1967). 한 형식 체계가 일관적이라면 그 형식 체계 내에서 그 형식 체계의 괴델 문장을 증명할 수는 없다. 그러나 그 문장을 증명할 수 있는 규칙을 가진 다른 형식 체계가 존재하는 것이 가능하다.(물론 그 형식 체계의 괴델 문장은 따로 있을 것이고 그 새로운 괴델 문장은 다시 그 체계에서 증명불가능할 것이다). 따라서 한 기계는 자기 자신의 괴델 문장을 증명할 수 없지만, 다른 기계는 그 기계의 괴델 문장을 증명할 수 있다. 그렇다면, 루카스가 자기에게 주어진 기계의 괴델 문장이 참임을 안다고 하더라도, 그것은 루카스가 또 하나의 기계라는 것을 배제하는 것은 아니지 않은가?

슬레작(Slezak,1982) 등이 루카스가 타입/토큰 구분을 하지 못했다면서 비판한 것도 같은 맥락에서였다. 루카스의 논변은, 루카스가 그 기계 토큰이 아니라는 것은 보여 줄지 몰라도 루카스가 기계 타입에 속하지 않는다는 것까지 보여주는 것은 아니라는 것이다.

그러나 루카스가 이 정도의 구분도 하지 못하고 있는 것은 아니다. 루카스는 단순히 어떤 하나의 기계와 자기 자신(루카스)의 동일하지 않음을 논증하는 것이 아니라, 주어지는 임의의 기계에 대해 그 기계와 자기(루카스)의 동일하지 않음을 논증할 수 있다는 것을 논증(혹은 주장)하는 것이다. 루카스가 괴델 증명의 일반적 절차를 알고 있다면, 기계론자가 루카스에게 루카스가 어떠한(any) 기계와 동일하다고 말하더라도, 그 때마다 항상 루카스는 그 기계의 괴델 문장을 찾아 그 문장에 대한 인지에 있어서 자기가 그 기계와 동일하지 않다고 기계론자를 반박할 수 있을 것이다. 루카스가 하는 일은, 그러한 개별적 반박들을 위한 반박의 틀(scheme)을 제시하는 것이다.

데이빗 루이스(D.Lewis, 1969)가 지적하듯, 루카스는 유한한 존재이므로 모든 기계들의 괴델 문장을 모두 산출하여 그것들이 모두 참임을 알 수는 없다. 그런데 루이스가 보기에는, 루카스가 그것들 모두(루이스는 그것을 루카스 산술이라고 부른다)가 참임을 알 경우에만 루카스가 기계가 아님이 입증된다. 왜냐하면 그렇지 않을 경우, 루카스에 의해 참임이 알려지는 괴델 문장들 모두를 증명하는 기계가 있을 수 있

고 루카스는 바로 그 기계와 동일할 수 있기 때문이다.

이에 대해, 루카스(Lucas.1970)는 자신의 논증이 '변증법적' 성격을 가진다고 말한다. 즉 그는 모든(all) 기계들에 대해 그 기계들의 괴델 문장이 참임을 알고 있을 필요는 없는 것이다. 다만 기계론자가 '루카스가 어떠어떠한 기계와 동일하다' 고 말할 때마다 기계론자의 그 특정한 주장들을 반박할 수 있으면 된다. 그리고 사실 어떠한 기계도 기계론자의 임의의(any) 제안에 대해 그러한 일을 할 수는 없다.

그러나, 루카스가 임의의(any) 기계에 대해 그 기계의 괴델 문장을 알 수 있다는 것은 참인가? 한 기계의 괴델 문장이 무엇인지 알기 위해 우리는 그 기계의 프로그램을 알아야 한다. 베나세라프(Benaceraf,1967)는, 우리가 자기 자신의 프로그램이 무엇인지 원리적으로 알지 못하는 기계일 가능성을 제시한다. 그 경우 나는 나 자신의 괴델 문장은 산출하지 못하는 기계일 것이다.

이에 대한 루카스의 대답(Lucas.1968)은, 나와 같다고 주장된 기계의 프로그램이 어떤 것인지 말할 책임을 기계론자에게 지우는 것이다. 기계론자가 내가 하나의 기계라고 주장할 때, 그는 내가 어떤 기계인지까지 말해야 한다는 것이다. 그러나, 기계론은 인간들이 각각 기계라는 일반적 논제로서 충분하다. 그것들이 각각 어떤 기계인지는 그 일반적 논제의 일부분이 아니다.

그럼에도 불구하고 베나세라프 식의 구제책은 환영할 만한 것이 아니다. 그런 식으로 구제된 기계론은 원래 기계론의 매력 중 많은 부분을 잃은 것일 것이다. 기계론 및 그것과 공통적 부분을 지닌 기능주의나 인지주의는, 인간 마음의 인지심리학적 탐구를 위한 작업가설들로서 매력을 지닌다. 그러한 작업가설 하에서 인지심리학의 목표는 인간 마음의 프로그램을 밝혀 내는 데에 있다. '인간은 자기 자신의 프로그램이 무엇인지를 원리적으로 알 수 없는 기계이다' 라고 말하는 기계론은 그러한 탐구를 불가능하게 한다. 따라서 제한되지 않은 기계론이 괴델 논변으로부터 구제될 수 있는지 알아 보는 것이 유익할 것이다.

3-2. 두번째 반론의 유형: "기계도 어떤 방식으로는(비형식적으로는) 괴델 문장을 알 수 있다."

이 유형의 반론은, '안다' 와 '증명한다' 의 개념에 관련된 것이다. 루카스의 논변은, 나는 문장 G를 '알' (혹은 비형식적으로 증명할) 수 있는데 기계는 문장 G를 '증명할' (혹은 형식적으로 증명할) 수 없다는 근거에서 내가 기계와 서로 다르다고 말한다. 그러나 나와 기계의 비동일성을 입증하기 위해서는 둘이 서로 동일한 측면에서 다르다는 것을 말해야 한다. 즉 나는 G를 알 수 있는데 기계는 G를 알 수 없다거나,

또는 나는 G를 증명할 수 있는데 기계는 G를 증명할 수 없다거나여야 할 것이다. (다음 논변을 살펴 보라:키케로는 카텔리나를 탄핵했다. 툴리는 카텔리나를 고소하지 않았다. 그러므로, 키케로는 툴리가 아니다.) 보이어의 반론(Boyer,1983), 베나세라프의 또 다른 반론(Benaceraf,1967) 등이 모두 이런 맥락에서 제기되었다고 볼 수 있다.

베나세라프는 다음과 같이 말한다.(Benaceraf,1967:19-20)

> "… 괴델의 제 1정리에 의해, 기계(그녀를 '마우드' 라고 부르자)가 하지 못하는 것으로 규정되는 것은 무엇인가? 분명, 그것은 그녀의 규칙들을 사용해 그녀의 공리들로부터 H(그녀의 괴델 문장)를 증명하는 일일 것이다. 그러나 그것을 루카스는 할 수 있는가? 분명 그렇지 않다. 그러면 마우드는 할 수 없으면서 루카스는 할 수 있는 일은 무엇인가? 그가 할 수 있을 일은 아마도 H에 대해 비형식적(마우드의 체계에서 형식화될 수 없다는 의미에서 비형식적인) 증명을 하는 일일 것이다. 그러나 마우드 역시 이것을 할 수 있다는 것이 명백하지 않은가? 마우드는 마우드-형식적 증명들을 제시함에 있어 제한을 가진다. 그러나 이러한 제한이 그녀가 제시할 수 있는 비형식적(마우드에서 형식화될 수 없는) 증명들에 대해서도 존재하는가? 그런 제한이 분명하게 존재하는 것 같지는 않다. … 그녀는 '[H]가 비록 [마우드]에서 증명불가능하지만 그럼에도 불구하고-바로 그 이유로 해서-참이라는 것을 납득할' 수 없는가?"

이 반론은, 기계는 G(그 기계의 괴델 문장)를 형식적으로 증명할 수 없는데 인간은 G를 비형식적으로 증명할 수 있다는 것이, 루카스 논변의 유일한 근거라고 비판하는 것이라 할 수 있다. 그리고 베나세라프의 지적은, 인간도 기계와 마찬가지로 G를 형식적으로(기계의 형식체계를 써서) 증명할 수 없고, 기계도 인간과 마찬가지로 G를 비형식적으로는 증명할 수 있다는 것이다.

실제로, 인간 역시, 형식체계의 규칙들을 가지고 (그 체계의 공리들로부터) 그 체계의 괴델 문장을 증명할 수 없다는 것은 명백하다. 그것은 바로 괴델의 제1 정리가 말하는 바이다. 그러나 기계가 자신의 괴델 문장을 비형식적으로 증명할 수 있다는 주장은 성립하는가?

기계(튜링 기계)의 정의상 모든 기계 상태는 기계의 이전 상태로부터 일정한 규칙들에 의해 결정된다. 따라서, 어떤 문장에 대해 그것이 참이라고 납득하거나 그것을 '비형식적으로' 증명하는 기계 상태도 규칙들에 의해 결정되어야 한다. 그렇다면, 그 기계에 의해 이른바 '비형식적으로' 증명된 것들까지 포함해 그 기계에

의해 증명된 문장들의 집합을 정리들의 집합으로 가지는 형식 체계가 존재할 것이다. 앞서 우리가 '그 기계가 구현하는 형식 체계' 라고 부른 것은 다름아닌 바로 이 형식 체계이다. 그리고 그 기계의 괴델 문장은 바로 이 형식 체계의 괴델 문장이다. 그렇다면, 괴델의 제1 정리에 의해, 그 괴델 문장은 그 형식 체계의 정리 집합에 속해 있지 않을 것이고, 따라서, '비형식적으로' 증명된 문장들 중에 괴델 문장은 없을 것이다. 즉 한 기계의 괴델 문장은 그 기계에 의해 '비형식적으로' 조차도 증명될 수 없다. (그리고 사실, '비형식적 증명' 이라는 것이 기계가 구현하는 형식 체계 바깥에서의 증명을 의미한다면, 기계가 비형식적 증명을 한다는 것 자체가 가능하지 않다).

여기서, 우리는 다음의 두 개념 사이에 구분을 해야 한다: 기계(혹은 인간)가 '사용' 하는 형식 체계와 기계(혹은 인간)가 '구현' 하는 형식 체계는 다르다. 어떤 기계가 명제(혹은 문장)들의 표상들을 조작함으로써 최종적으로 정리인 명제들을 산출한다고 해 보자. 그리고 그렇게 하면서 조작 규칙들 자체를 표상한다고 해 보자. 즉 그 기계는 표상 차원의 명시적 규칙들을 사용해 정리들을 증명한다. 그 경우 그 기계는 그 규칙들로 이루어진 형식 체계를 '사용' 한다고 말할 수 있을 것이다. 그 기계가 만약 또 다른 명제들을 그 형식 체계(어떤 명제들을 증명하는 데에는 사용되었던)의 규칙들을 사용하지 않고서 산출할 경우, (우리가 원한다면) 그 명제들은 기계에 의해 '비형식적으로' 증명되었다는 말을 할 수 있을 것이다. 물론 그 때의 '비형식적 증명' 이라는 말은 그 기계가 (어떤 명제들을 증명하는 데에) '사용' 하는 형식 체계 바깥에서의 증명을 의미한다. 그러나 그런 의미에서 기계가 '비형식적' 증명을 한다고 하더라도 그런 증명 과정은 (기계에 고유한) 일정한 규칙들에 따라야 한다. 그렇게 기계가 겪는 어떠한 과정을 통해서든 어떤 명제 집합이 산출된다면, 그러한 과정 모두에 적용되는 규칙들이 존재하므로 그 명제 집합을 정리 집합으로 하는 형식 체계가 존재할 것이다.(즉 그 명제 집합은 '공리화가능' 하다). 그러한 형식 체계가 바로 기계가 '구현' 하는 형식 체계이다.

괴델 논변이 전개될 때에 문제가 되는 것은 기계가 '구현' 하는 형식 체계의 괴델 문장이지 기계가 '사용' 하는 형식 체계의 괴델 문장이 아니다. 이 둘 간의 구분을 하지 못한 것이 바로 베나세라프 반론의 오류를 낳았다. 그는 '기계 마우드의 형식 체계'라는 말로, 마치 우리가 산술의 정리를 증명할 때 페아노 산술 체계를 사용하듯, 마우드가 사용하는—그래서, 그것을 벗어날 수도 있는— 형식 체계를 의미하는 듯이 이야기한다.

그렇다면 인간과 기계에 있어 어떤 동일한 측면을 비교할 수 있을 것인가? 이런 문제에 부정적으로 대답하면서, 보이어(Boyer,1983)는

루카스의 논증에 결함이 있다고 주장한다. 그는, 한 사람의 인간과 그를 모의(simulate)하는 한 기계를 설정하고서, 인간의 증명과 기계의 증명 간에 애매성이 존재한다는 점을 불평한다. 게다가 어떤 것을 인간의 증명에 상응하는 기계의 증명으로 삼을지에 대한 애매성도 존재한다고 말한다. 예를 들어 루카스가 칠판에 문장들을 증명하는 것에 대한 그림을 찍어냄으로써 기계가 루카스를 모의하고 있다고 하자. 그 경우 그 그림 속의 칠판 속에 증명되는 문장들의 집합이 기계가 증명하는 문장들의 집합으로 해석될 수도 있고 그런 그림을 찍어 내기 위해 보내는 0과 1의 전기 신호가 일차 술어논리의 문장으로 해석되었을 때의 문장들의 집합이 기계가 증명하는 문장들의 집합으로 해석될 수도 있다.

그러나 이런 것들은 아무런 문제가 아니다. 우리는 지금, 탁자 위에 놓여 프린터와 연결되어 있는 쇠로 만든 '정리 증명' 기계를 인간과 비교하고 있는 것이 아니다. 우리는 기계가 종이 위에 프린트하는 문장들과 인간이 마음 속에 믿고 있는 문장들을 비교하려 애쓸 필요가 없다. 기계론의 논제는 바로 인간이 기계라는 것이다. 그리고 기계론을 반박하기 위해 루카스가 인간과의 비교 대상으로 삼으려는 기계는 바로 기계론에 의해 인간과 동일하다고 상정된 바로 그 기계이다. 따라서 그 기계는 먹고 말하고 숨쉬는 기계이다. 그 기계에 있어서의 어떤 양태를 인간에 있어서의 앎(혹은 증명)과 비교할 것이냐고? 간단하다. 바로 그 기계인 그 인간에 있어서 우리가 앎(혹은 증명)이라고 부르는 것을 그것과 비교하라. 물론 여기에도 애매성이 있지만, 인간에 있어서 무엇을 앎이라 할 것인지의 애매성과 그 인간인 기계에 있어서 무엇을 앎이라 할 것인지의 애매성은 정확히 일치한다.

3-3. 세번째 반론의 유형: "어떤 형태의 기계는 괴델 논변을 피할 수 있다."

이 유형의 반론은 괴델 논변을 피할 수 있는 형태의 기계가 있을 수 있음을 지적한다. 그리고 인간은 바로 그런 형태의 기계라는 것이다. 죠지(George,1962)와 스마트(Smart,1963) 등은 이른바 '귀납적 기계(inductive machine)'가 괴델 논변을 피할 수 있다고 보았다. 그리고 이런 아이디어는 홉슈태터(Hofstadter,1979)에서도 다시 나타난다.

형식 체계 T1을 구현하는 기계는 T1의 괴델 문장을 증명할 수 없을 것이다. 그래서 T1의 괴델 문장을 공리로서 포함하는 형식 체계 T2를 구현하는 기계를 만들 경우 그 기계는 다시 T2의 괴델 문장을 증명할 수 없을 것이다… 등등. 이 때 T1,T2,T3,… 등을 각각 구현하는 기계들은 모두 자기의 괴델 문장을 가질 것이다. 그런데 T1에서 시작해서 자기 행동의 규칙들을 경험적으로 배울 수 있는 기계, 즉 귀납적 기계가 있다고 하자. 그 기계는 자기가 구현하는 형식

체계의 규칙들이 무엇인지 알므로 그로부터 자기의 괴델 문장을 구성할 수 있을 것이다. 그러면 그 괴델 문장을 포함하는 방식으로 자기의 규칙들을 바꿀 수 있다. 그 기계는 그런 식으로 T2,T3,… 로 스스로 나아갈 줄 아는 기계이다. 그 기계는 처음의 규칙들 아래에서는 증명할 수 없었던 괴델 문장들을 언젠가는 증명할 수 있게 될 것이다.

그런 기계는 어떻든 알고리즘적 기계이거나 알고리즘적 기계가 아니거나일 것이다. 그 기계가 결국에 있어 알고리즘적 기계라고 하자. 그러면 그 기계는 자기의 귀납 추론을 위한 규칙들도 알고리즘으로서 포함하고 있어야 한다. 그 기계는 이를테면 T1을 초기조건으로 해서 이차(second-order) 규칙들을 통해 T2, T3로 나아가는 기계이다. 그 경우 기계가 '구현'하는 형식 체계는 바로 초기조건과 그 이차 규칙들에 대응하는 형식 체계이다. T1이나 T2와 같은 것들은 기계가 '사용'하는 형식 체계는 될지 모르지만 기계가 '구현'하는 형식 체계는 아니다. 그 기계는 T1이나 T2의 괴델 문장을 언젠가는 증명할 수 있게 될지 모르지만 그 기계가 '구현'하는 형식 체계의 괴델 문장은 증명할 수 없다. 한 기계가 알고리즘적 기계인 이상에는 그 기계의 모든 종류의 상태 변화-이른바 규칙의 변화까지 포함해-를 규정하는 일정한 알고리즘이 있어야 하고, 그런 알고리즘에 상응해 형식 체계와 그것의 괴델 문장이 존재할 것이다.

그 기계의 귀납 절차가 알고리즘화될 수 없는 것이라 하자. 그런 기계는 괴델 논변을 피할 수 있을 것이다. 그러나 그런 기계는 우리가 규정한 바의 기계는 아니다. 물론 그런 기계를 기계라고 부를 수 있는 충분한 의미가 있을 수 있다. 예를 들어 병렬분산 처리 컴퓨터가 그런 기계일 수도 있다. (괴델 논변을 피하는 데 있어서의 PDP 컴퓨터의 가능성에 대해서는 Lyngzeidetson,1990 참조). 그렇지만 이것이 괴델 논변에 대해 최종적으로 남는 대답이라면, 괴델 논변은 특정한 형태의 기계론을 반박하고 기계론 중의 한 입장이 다른 입장보다 선호될 만 하다는 것을 보이는 것이 될 것이다. 따라서 원래 괴델 논변의 효력을 어느 선까지는 인정하는 것이 된다. 인간 이성이 알고리즘화될 수 있다는 흥미로운 가설을 담고 있는 것은 바로 좁은 의미의 기계론인데, 이런 좁은 의미의 기계론을 이 유형의 반론은 구제하지 못한다.

4. 일관성 반론 : 나는 내가 일관적이라는 것을 알 수 있는가?

괴델 논변에 대한 아마도 가장 대표적인 유형의 반론을 우리는 가장 마지막에 다루게 되었는데, 그것은 이 반론이 필자 자신이 받아들이는 반론이기 때문이다. 퍼트남(Putnam, 1975:366)에 의해 처음 제기되고 베나세라프(Benaceraf,1967), 굿(Good,1969), 치하라

(Chihara,1972) 등에 의해서도 되풀이된 이 유형의 반론에서 적절히 지적되었듯, 괴델 논변은 괴델의 정리를 적용함에 있어 잘못을 범하고 있다.

우리(인간)가 괴델 식의 추론을 통해 알 수 있는 것은 (형식 체계의) 괴델 문장이 참이라는 정언 명제가 아니다. 우리(인간)가 알 수 있는 것은, 그 형식 체계(산술을 포함한)가 일관적이라면 그 형식 체계의 괴델 문장이 참이라는 조건 명제이다. 그리고 그러한 조건 명제에 해당하는 형식 문장은 그 형식 체계 자체 내에서도 증명될 수 있다. (G가 형식 체계 T의 괴델 문장일 때)

T가 일관적이면, G가 참이다.

라는 메타언어 문장에 해당하는 형식언어 문장인

$$Con(T) \rightarrow G$$

가 T 내에서 증명가능하다.[4] (이것이 T 내에서 증명가능하다는 것은 바로 괴델의 제2정리를 증명하는 과정에서 이용되는 사실이다). 따라서 T를 구현하는 기계도 이 조건 문장은 증명할 수 있다. 그러면 괴델의 정리에 의해 존재하는 것으로 여겨졌던, 인간과 기계 간의 차이가 실은 존재하지 않는 것이 된다.

인간이 기계와 달리 G의 참을 정언적으로 알 수 있으려면 위의 조건문('T가 일관적이면, G가 참이다')뿐만 아니라 그 조건문의 전건인 'T가 일관적이다'가 참임을 알아야 한다. 그러나 인간이 T의 일관성을 알 수 있다는 근거가 있는가? (T가 일관적일 경우) T의 일관성이 T 내에서 증명될 수 없다는 것이 괴델의 제 2정리가 말하는 바이다. 그리고 이것은 T의 일관성이 유한적인 방법으로 증명될 수 없음을 보여 주는 것처럼 보인다. 인간이라고 해서 T의 일관성을 정당하게 알 수 있음을 보여 주는 힌트는 (최소한 괴델의 정리 내에) 없다.

이 문제를 피하기 위해 루카스(Lucas,1968)는 우리가 우리 자신의 일관성을 알 수 있다고 주장하고 이로부터 T의 일관성을 우리가 알 수 있다는 것을 논증하려 한다. 그 논증을 정리해 보면 다음과 같다.

(1) 나는 일관적이다. (전제)
(2) T를 구현하는 기계는 나와 동일하다. (기계론의 가정)
(3) T가 비일관적이면 T를 구현하는 기계는 나와 동일할 수 없다. ((1)로부터)
(4) T는 일관적이다. ((2)와 (3)으로부터)

4) 'Con(T)'는 T의 일관성을 진술하는 형식언어 문장의 축약이다. 이를테면 그것은 '(x)-B(x,g(f))'와 같은 문장의 축약이다.

(5) T가 일관적임을 나는 알 수 있다.
((4)로부터)

루카스는 전제 (1)을 주장하기 위한 (더 나아가 전제 (1)을 내가 알 수 있다는 것을 주장하기 위한) 나름의 논변을 펼친 바 있다. (Lucas,1961;1968)

그러나 루카스의 생각과는 달리 (1)이 나에 의해 알려질 수 있다는 것도 정당화되지 않고, 또 그것이 정당화된다고 하더라도 그것으로부터 (5)가 따라나오지도 않는다.

우선, 내가 (1)을 알 수 있다는 것으로부터 (5)가 따라나오지 않는다는 것을 살펴 보겠다. 위의 논증에서 (4)로부터 (5)로 나아가는 단계는 어떻게 정당화되는가? p로부터 '나는 p를 알 수 있다'로 나아가는 추론은 오직 정언적 맥락(특히, 그 논증의 행들이 내가 아는 명제들의 나열인 논증의 맥락)에서만 정당화된다. 어떤 가정을 세우고 그 가정으로부터의 귀결을 끌어내는 조건적 논증(conditional argument)을 하는 맥락에서는 이 추론이 정당하지 않다. 왜냐하면 조건적 논증 속에서 언명되는 어떠한 명제에도 그 명제를 내가 알고 있음이 개입(commit)되지 않기 때문이다. 위의 논증에서의 전제 (1)을 내가 알 수 있음을 성공적으로 보인다고 하더라도, 위의 논증은 귀류법적 가정인 (2)를 포함하고 있고, 특히 (4)는 이 가정인 (2)에 의존하고 있다. 따라서 (4)로부터 (5)를 끌어낼 수는 없다.

루카스는 어쩌면 기계론의 가정을

(2)′ 나는, T를 구현하는 기계가 나와 동일하다는 것을 알 수 있다.

라고 강화하려 할지도 모른다. 그러나 (2)′은 기계론 논제 이상의 것을 이야기하고 있다. 기계론이 참이라는 것에 대한 근거를 제시하는 것이 기계론자의 책임일지는 모르지만, 그것은 기계론 논제 자체의 일부는 아니다. 가정 (2)를 이렇게 강화함으로써 (전제 (1)을 내가 알 수 있다는 조건 하에서) 괴델 논변은 기계론에 대한 불가지론을 입증할 수 있을런지는 모르지만, 기계론이 거짓임을 입증할 수는 없다. (그런 조건 하에서) 괴델 논변이 증명하는 것은 인간이 알고리즘화될 수 없다는 것이 아니라 인간이 알고리즘화될 수 있음이 알려질 수 없다는 것일 것이다. (이것은 베나세라프가 제시했던 탈출구의 한 변형이다).

그러나 괴델 논변은 기계론에 대한 불가지론이나마 입증할 수 있는가? 그것은 전제 (1), 즉 나의 일관성을 내가 알 수 있음을 성공적으로 보여 주는 데에 달려 있다. 루카스(Lucas,1961:264-268;1968:157-158)는 이를 보이기 위해 많은 노력을 했지만 크게 성공적인 논변을 제시하지는 못했다. 자기의 일관성에 대한 그의 논변의 중요한 부분은 사실상 요청의 차원에 머물러 있다:

> "합리적 행위자가 자기 자신의 합리성, 그리고 따라서 자기 자신의 일관성을 믿는 것이 합리적이다. 그것은 그 이상의 사유가 가능하기 위한 가정(assumtion)이기 때문에, 그리고 그것은 사실의 문제가 아니라 결단의 문제이기 때문이다. 우리는 자신이 일관적이기를 결단할 것이며... 어떠한 것도 모두 다 언명하도록 허용하기로 결단하지 않고 참과 거짓 사이의 구분을 하기로 결단할 것이다. 나는 내가 비일관적이라는 것을 믿을 수 없다." (Lucas,1968:158)

그래서 어떠한 사유를 함에 있어서도 자기 자신이 일관적이라는 가정을 하는 것이, 혹은 자기 자신이 비일관적이라고 믿지 않는 것이 합리적일지도 모르겠다. 그러나 이것이 자기 자신이 일관적임을 우리가 '알 수 있다'는 것을 보여 주는가? 그리고 기계론에 대한 불가지론이나마 보여 주고자 한다면, 우리가 일관적임을 가정하는 것뿐만 아니라 우리가 일관적임을 우리가 알 수 있음을 보여야 한다.

우리의 논의의 이 단계만으로도 루카스의 괴델 논변에 정당화되지 않은 부분들이 있다고 생각할 이유가 있다. 그러나 루카스는 여전히 자신의 일관성을 알 수 있다고 믿고 그것을 근거로 괴델 논변이 가능하다고 생각할지도 모른다. 그러므로 이 문제와 관련해 루카스에 대한 보다 적극적인 반론을 시도해 보자.

5. 일관성 가정과 타르스키 정리

괴델 논변에서, 우리 자신의 인지 체계가 일관적이라는 것을 우리가 알 수 있다는 가정이 중요한 역할을 한다는 것이 명백해졌다. 필자는 이제 우리가 우리 자신의 인지 체계의 일관성을 알고 있다는 루카스의 가정이 근거 없다는 것을 주장하는 것에 그치지 않고, 우리가 괴델 논변을 전개시킬 수 있다는 바로 그 사실이 우리 자신의 인지 체계의 일관성을 위협한다는 논변을 펴려고 한다.

우선, 괴델 정리 말고 또 하나의 메타논리학적 정리인 타르스키 정리가 말하는 바를 상기하는 것이 좋겠다. 타르스키 정리에 의하면, 산술을 포함할 수 있으면서(괴델의 절차에 의해, 이는 곧 자기 지시가 가능함을 의미한다), 일관적인 언어에는 그 언어의 진리 술어가 존재할 수 없다.(그 언어의 진리 술어가 그런 언어 내에서 정의될 수 다). 여기서 어떤 언어의 진리 술어란 그 언어의 모든 문장 p에 대해 ('s'가 'p'를 대치하는 문장의 이름일 때) 다음의 도식 T가 성립하는 술어 B이다.

T: B(s) iff p

이 정리가 성립하는 직관적인 이유는 자기 지

시적 장치(산술을 포함하면 이는 보장된다)와 그 언어 자체의 진리 술어를 지닐 경우, 우리가 잘 아는 다음의 거짓말장이 역설 문장을 구성할 수 있고 그로부터 비일관성이 따라나오기 때문이다.

(1) 이 문장은 참이 아니다.

우리 자연언어에서 이 문장을 구성할 수 있다는 것이 곧바로 우리 자연언어 체계의 비일관성을 의미하는 것은 아니다. 우리 자연언어에서의 거짓말장이 역설은 흔히 겉보기의 현상으로 여겨지기도 한다. 타르스키는 이 역설처럼 보이는 현상을 해소할 방안을 제시한다. 그것은 자연언어에서의 '참이다'라는 술어가 다의성을 가지는 것으로 해석하는 것이다. 즉 진리 술어는 원천적으로 언어상대적인 술어- 즉, 이항관계 술어-이고, 자연언어의 '참이다'라는 술어는 특정 언어에 대한 관계항을 감추고 있다는 것이다. 자연언어에서의 진리 개념들이 각각 특정 언어에 대한 진리 개념이고 그 특정 언어에 대한 진리 개념이 그 특정 언어 속에 있을 수 없다면 역설은 일어나지 않는다. 메타언어 M 속에 M에 대한 진리 술어가 존재할 수 없다면, 위의 (1)의 진리 술어는 이를테면 대상언어 O에 대한 진리 술어로 해석되어야 한다.

(1)′ 이 문장은 O에서 참이 아니다.

이렇게 되면 이 문장은 더 이상 역설이 아니게 된다.

그렇다면, 우리는 우리가 사용하는 궁극적인 메타언어 전체에 대한 진리 개념,즉 전반적 진리(global truth) 개념을 가지지 못하는가? 그런 진리 개념을 가진다면 진정한 거짓말장이 역설을 구성할 수 있을 것이고 우리(의 언어)가 지닌 논리 체계는 근본적으로 비일관적일 것이다. 다시 말해 우리가 일관적이라면 우리는 오직 전반적 진리 개념을 갖지 않고 오직 국지적 진리 개념만을 가지고 있을 것이다. 우리가 실제로 일관적인지 비일관적인지 또한 전반적 진리 개념을 가질 수 있는지 가질 수 없는지에는 논란의 여지가 있다. 필자는 여기서 그 어느 쪽도 가정하지 않는다.

필자는 대신 다음과 같은 딜레마 논증을 구성하겠다.

1. 우리가 전반적 진리 개념을 갖지 않는다면 우리는 괴델 논변을 구성할 수가 없다.
2. 우리가 전반적 진리 개념을 갖는다면, 우리는 비일관적이고, 따라서 괴델 논변의 전제가 거짓이다.
3. 그러므로, 괴델 논변은 구성될 수 없거나 그 전제가 거짓이다.

이 논증의 첫번째 선언지 부분을 위해, 괴델

논변이 어떻게 진행되는지 상기해 볼 필요가 있다. 우선 루카스가 누누이 강조한 대로 괴델 논변이 귀류법적(루카스가 '변증법적'이라는 단어로써 표현하고자 하는 대로) 맥락에서 제기된다는 것을 기억하도록 하자. 즉 괴델 논변은 기계론이 참이라는 것을 가정하고 그 가정 하에서 기계론의 부정을 이끌어내는 논변이다. 따라서 이 논변의 맥락 내에서는 기계론이 거짓임이 가정되어서는 안된다.

기계와 달리 우리는 괴델 문장을 증명할 수 있다는 루카스의 주장은, 다음과 같이 괴델 문장을 메타언어 체계에서 증명하는 논변을 통해 개진되었었다.(Lucas,1961:256)

> "…'이 문장이 체계 내에서 증명불가능하다'라는 문장이 거짓일 가능성을 고려하고, 그것이 불가능함을 보인다. 그로부터 그 문장이 증명불가능하다는 것이 따라나온다."

이 논변은 구체적으로 이렇게 진행된다: 괴델 문장 G가 거짓이라고 가정하고, 이로부터 G는 증명불가능하지 않고 증명가능하다는 것을 끌어낸다. 그런데 이렇게 되면 체계의 건전성[5] 위배되므로 G가 거짓이라는 가정은 옳지 않다. 따라서 G는 참이고 G는 자기 자신이 증명불가능하다고 말하는 문장이므로 G는 증명불가능하다.

이 메타언어 논변에는 체계에 대한 의미론적 개념, 특히 진리 개념이 핵심적으로 사용되고 있다. 그런데 이 귀류법적 맥락 안의 논변은 귀류법의 가정에 따라 '바로 나인' 기계가 구현하는 체계에 대한 것이다. 다시 말해 이 논변은 내가 구현하는 형식 체계에 대한 진리 개념을 사용해야 한다. 따라서 이 논변에서 나는 내가 사용할 수 있는 메타언어 전체의 진리 개념을 요구한다. 결국, 귀류법의 가정을 진지하게 받아들인다면, 그 가정 하에서 괴델 논변을 행하기 위해 전반적 진리 개념을 필요로 한다. 즉, 전반적 진리 개념이 없을 경우 괴델 논변은 정식화될 수조차도 없다.

물론 괴델 정리에 대한 메타언어적 추론은 보다 구문론적 방식으로 주어질 수 있다. 우리가 2절에서 보았던 증명 형태(괴델 자신의 것과 보다 유사한)는 괴델 문장이 체계 내에서 증명불가능하다는 것을 보이기 위해 의미론적 개념들을 필요로 하지 않는다. 그러나 그 때에도 괴델 문장이 증명불가능하다는 사실로부터 괴

5) 이 방식의 논변에서는 일관성보다는 강한 조건인 건전성(즉, 그 체계에서는 참인 모든 문장이 증명가능하다는)이 요구된다. 그러나 루카스는 이것을 '일관성'이라는 이름으로 부른다.(Lucas,1961:255)

델 문장이 참임을 끌어내기 위해서는 어쩔 수 없이 괴델 문장(이 포함된 언어)에 대한 진리 개념을 사용하지 않을 수 없다. 따라서 이 경우에도 우리가 전반적 진리 개념을 갖고 있지 않다면 괴델 문장이 참임을 추론할 수가 없다.

우리 딜레마 논증의 두번째 선언지가 성립함을 우리는 이미 알고 있다. 4절에서의 고찰에 의해 드러났듯이 괴델 논변은 우리 인지 체계가 일관적이라는 (더 나아가 그것을 우리 자신이 알고 있다는) 전제를 필요로 한다. 그런데 우리가 전반적 진리 개념을 가지고 있다면 우리 인지 체계는 비일관적이라는 사실이 타르스키 정리에 의해 주어졌다. 따라서, 우리가 전반적 진리 개념을 가지고 있다면 괴델 논변에서 요구되는 전제가 거짓이다.

결국, 어떤 경우든 괴델 논변은 정당화될 수 없다는 것이 우리 딜레마 논증에 의해 보여졌다.

6. 맺음말

괴델의 정리가 기계론을 반박한다고 주장하는 사람들의 표어는 "인간 이성은 형식화될 수 없다"(Nagel and Newman, 1958)는 것이다. 그리고 그들은 바로 괴델의 정리가 그것을 보여준다고 생각한다. 그러나 괴델의 정리에 대한 증명 과정으로부터 우리가 깨닫는 가장 흥미로운 사실은 산술에 대한 우리 메타언어를 산술 안에서 표현할 수 있다는 것이다. 그리하여 우리는 괴델의 정리에 대한 메타언어적 증명 자체를 산술 안에서 표현할 수 있고, 따라서 괴델의 정리에 상응하는 형식언어(산술언어) 문장이 산술 형식체계에서 증명가능하다는 것을 알 수 있다. 즉 괴델의 증명은 다름 아니라 인간 이성의 한 중요한 부분—메타수학적 추론—이 형식화될 수 있다는 것을 (오히려) 보여 준다. 물론 이것은 인간 이성의 모든 부분이 형식화될 수 있음을 보여 주는 것은 아니다. 인간 이성은 형식화될 수 없을 것 같아 보이는 많은 인지 과정을 포함하고 있다. 이런 모든 인지 과정들이 형식화될 수 있다고 주장하는 것은 기계론자가 떠맡아야 할 부담이다. 필자는 기계론을 주장하거나 논증하려고 하지 않았다. 필자가 이 글에서 보이려고 한 것은 오직, 인간 이성의 어떤 부분이 형식화될 수 없다고 주장할 때 그 근거가 괴델의 정리를 우리가 메타언어적으로 증명할 수 없다는 것에서 찾아져서는 안된다는 것이었다. 그 부분은 인간 이성에서 오히려 가장 잘 형식화될 수 있는 부분 중의 하나이다. 그리고 루카스는 괴델의 정리를 증명하는 인지 능력에 있어서 인간이 기계보다 낫다는 것을 증명하지 못했다.

〈참고문헌〉

(1) Lucas [1961] "Minds, Machines and Godel", Philosophy 36:112-127

(2) Lucas [1968] "Satan Stultified: a Rejoinder to Paul Benacerraf", Monist 52:145-58

(3) Lucas [1970] "Mechanism: A Rejoinder", Philosophy 45:149-51

(4) George [1962] "Minds, Machines and Godel: Another Reply to Mr. Lucas", Philosoohy 37:62-63

(5) Good [1967] "Human and Machine Logic", Brit J Phil Sci 18:145-6

(6) Benacerraf [1967] "God, the Devil, and Godel", Monist 51:9-32

(7) Lewis [1969] "Lucas Against Mechanism", Philosophy 44:231-3

(8) Chihara [1972] "On Alleged Refutations of Mechanism Using Godel's Incompleteness Results", J Phil 64:507-26

(9) Hofstadter [1979] *Godel, Escher, Bach* , Basic books

(10) Boyer [1983] "Lucas, Godel, and Astaire", Phil Q 33:147-59

(11) Penrose [1989] *The Emperor's New Mind* Oxford Univ.

(12) Lyngzeidetson [1990] "Massively Parallel Distributed Processing and a Computationalist Foundation for Cognitive Science", Brit J Phil Sci 41

(13) Martin & K. Engleman [1990] "The Mind's I Has Two Eyes", Philosophy 510-16

(14) G del.K.[1931]. On formally undecidable propositions of Principia mathe-matica and related systems I. in J. van Heijenoort [1967], 596-616.

(15) Hellman, G. [1981]. "How to Gödel a Frege-Russell: Gödel's incompleteness theorems and logicism." NOUS 15:451-468.

(16) Hofstadter, D.R.and Dennett, D.C.[1981]. The Mind's I. Basic Books.

(17) Nagel, E and Newman, J.R.[1958]. G ödel's proof. New York Univ. Press.

(18) Putnam, H.[1975]. "Minds and machines" in *Mind, Language and Reality*. 362-85. Cambidge Univ.Press.

(19) Slezak. P.[1982]. "Gödel's theorem and the mind." Britisch Journal for the Philosophy of Science

33:41-52.

(20) Wang, H.[1980]. *From mathematics to philosophy*. R.K.P.