

가상세계와 현실세계의 구별가능성 고찰

정 기 도

(덕성여대 교양과정부 강사)

1. 가상 세계에 대한 논의

고도 컴퓨터 문명 시대에 '가상 세계'란 말은 대중화되어 사용되고 있다. 우리는 신문지상에서, 또는 컴퓨터나 과학 계통의 잡지에서, 또는 영화나 만화 등에서도 가상 세계를 주제로 글들이 꾸려져 나가는 것을 자주 본다. 가상 세계란 무엇인가? 글자대로라면 현실 세계의 대조 개념으로 보인다. 실제로 그런 용법으로 사용되고 있다. 그러나 단지 그것만인가? 아니면 어떤 다른 중요한 것이 있을 수 있는가?

나는 가끔 이런 질문을 던져본다. 만약 인간이 아무런 존재가 아니라면, 즉 만물의 영장이라는 인간이 허울뿐인 껍데기라면, 과연 나는 내가 인간임에 대해 어떻게 생각해야 하는가? 나 뿐만 아니라 대부분의 사람들이 이러한 존재론적 혹은 실존적 질문을 던질 것이다. 어지간한 사람이 아니고서는 이러한 질문에 상당한 불안을 느끼게 된다. 이것이 철학의 시작일 수도 있지만, 문제는 컴퓨터의 발전과 더불어 나타나는 가상 세계에 대한 탐구가 이러한 존재론적 질문과 그에 따른 불안을 가중시킨다는 것이다. 만약 우리 스스로가 아무런 의미가 없다면, 만약 많은 의미와 가치가 부여된 인간이 무엇인가에 의해 어떤 계획에 따라 만들어졌고, 단지 그런 존재일 뿐이라면, 과연 우리는 우리 자신에 대해, 또 우리가 만들었다고 생각한 모든 것에 대해 어떤 판단 내지는 생각을 하게 될까?

다음에서 보다 자세히 살펴 보겠지만, 가상 세계에 대한 논의는 조심스럽기도 하고, 특히 현실 세계와의 구별가능성에 있어서는 매우 중요한 문제를 야기한다. 이들 간의 구별의 문제는 단순히 현실 세계

란 어떤 것이고 가상 세계란 어떤 것이라든 규정할 수 없다. 이 문제는 어쩌면 인간이 지금까지 생존해 오면서 만들어온 모든 존재론적 기반을 송두리째 무너뜨릴 수 있을 정도의 공포스러운 문제다. 특히 종교와 관련하여 인간은 한낱 장난감 또는 유희의 도구에 지나지 않을 수도 있고, 역으로 종교 자체 또는 종교의 바탕이 되는 신성성 자체가 뿌리째 부정될 수도 있다.

우선 나는 논의의 전개를 위해 세 가지의 이야기를 구성할 것이다. 그리고 세 가지의 이야기를 통해서 가상 세계와 현실 세계의 구별가능성 이야기를 다룰 것이다. 결론부터 말하자면 다음과 같다.

1) 가상 세계와 현실 세계는 구별가능하다. 왜냐하면 우리가 구성하는 가상 세계란 현실 세계와 그 위계에 있어서 차이가 나기 때문이다.

2) 가상 세계와 현실 세계는 구별가능하지 않다. 우리가 살고 있는 세계가 가상 세계인지 현실 세계인지 우리는 결정할 수 없다. 왜냐하면 우리는 가상 세계와 현실 세계를 구별할 수 있는 아무런 논리적인 기준을 가지고 있지 않기 때문이다.

얼핏 모순되는 주장처럼 보이지만 1)과 2)는 서로 다른 논의이다. 1)은 서로 다른 위계에서의 주장이고, 2)는 동일한 위계에서의 주장이기 때문이다. 따라서 우리가 우리의 세계에서 만드는 가상 세계는 아무리 정교하게 만들었다고 하더라도 현실 세계와 여전히 구별가능하지만, 우리가 기반하고 있는 이 세계가 현실인지 가상인지는 구별 불가능하다.

만약 이러한 주장이 가능하다면, 그러한 한에 있어서 이러한 주장은 세계의 존재 방식과 인간 존재, 종교, 문화 등에 대한 전면적인 입장의 재정리까지도 함축하게 된다.

2. 세 가지 이야기

이야기 1 : 어느날 장주가 꿈을 꾸었는데, 꿈 속에서 그는 팔랑거리며 아름답게 날아다니는 한 마리의 호랑나비였다. 평소의 뜻대로, 마음 내키는 대로 즐거워할 수 있어 그 자신이 장주라는 인간임을 전혀 돌이켜 볼 수 없었다. 그런데 갑자기 잠에서 깨니, 놀란듯 호리명칭한 눈으로 두리번거리며 둘레를 둘러보고 있는 그는 장주였다. 어찌된 일일까? 장주가 호랑나비가 된 꿈을 꾸었을까, 아니면 호랑나비가 장주가 된 꿈을 꾸었을까? 어느 쪽인지 알 수가 없다. 분명히 장주와 호랑나비 사이에는 구별이 있을 것이다. 그렇지만 어떻게 하여 그것을 구별할 것인가?

이야기 2 : 어느 순간엔가 엘리스는 꿈 속에서 엘리스가 자신을 꿈꾸는 것을 꿈꾸는 잠자는 하트왕의 꿈을 꾸었고, 잠자는 하트왕은 꿈 속에서 잠자는 하트왕을 꿈꾸는 엘리스를 꿈꾸었고, 엘리스는 그런 하트왕을 꿈꾸었고……. 과연 잠자는 하트왕이 자신을 꿈꾸는 엘리스의 꿈을 꾸었을까, 아니면 엘리스가 자신을 꿈꾸는 잠자는 하트왕을 꿈꾸었을까?

이야기 3 : 컴퓨터 프로그램을 잘 짜는 은석은 시중 오락실에서 유행했던 프로그램인 '스트리트 화이터'를 많은 비용과 노력을 들여 가상 기계(virtual machine)로 꾸며놓고 친구들이 즐길 수 있게끔 하였다. 특히 은석은 프로그램 안에 그 게임을 즐기는 사람들 자신을 투영해서 게임을 할 수 있도록 조작해 놓았는데, 그 기계는 최대한 현실적이게 만들었다.

이야기 1과 이야기 2는 우리가 일상에서 한 번쯤은 스스로에게 물어볼 수도 있고, 또 한 번쯤은 경험할 수도 있는 것들이다. 이에 반해 이야기 3은 과학 문명이 발달함에 따라 만들어진 오락 게임을 말하고 있다.

이 이야기들 각각은 서로 다른 상황을 보여주고 있다. 이야기 1은 현실과 가상이 과연 구별가능한지를 묻고 있고, 이야기 2는 현실과 가상이 서로 무한 퇴행이라는 얽힘 속에 빠져있음을 보이고 있다. 이

야기 3은 프로그래머이자 제작자인 은석과 그가 빚어낸 피조물들 사이에서 만들어낸 인위적인 가상의 세계를 그리고 있다. 그런데 이 세계의 이야기는 한 가지 면에서 공통점을 갖는다. 즉 이 세 개의 이야기는 모두 가상 세계를 구성하고 있다는 것이다. 이제 질문을 던져보자.

- 1) 장주는 자신이 꿈을 꾸어서 나비가 되었는지 나비가 꿈을 꾸어서 장주가 되었는지를 구별할 수 있을까?
- 2) 과연 무한하게 얽혀 있는 엘리스와 하트왕의 꿈은 누구의 꿈으로부터 비롯된 것인가?
- 3) 은석이 스스로 그 게임을 한다고 했을 때, 과연 실제의 은석과 게임 속의 은석을 은석은 구별할 수 있을까?

이 질문들에 대한 답을 찾음으로써 나는 내 주장에 이르고자 한다.

3. 크립키 논변¹⁾의 원용

‘더하기(plus)’라는 산술의 개념과 ‘+’라는 수학적 기호를 가정하자. ‘더하기’라는 말과 ‘+’라는 기호는 수학적 함수를 지시하는 것으로 사용된다. 이 함수는 양의 정수의 모든 쌍들에 대해 정의될 수 있다. 사람들은 외적인 기호의 표현이나 내적인 정신적 표현에 의해 더하기라는 규칙을 파악한다. 중요한 것은 비록 사람들이 과거의 한정된 수의 더하기만을 해보았을 지라도 규칙은 사람들이 이전에 전혀 고려해보지 않았던 무한하게 많은 새로운 더하기에 대한 답을 결정하여 준다는 것이다. 사람들은 이러한 규칙을 파악한다. 그래서 사람들의 더하기에 대한 과거 경력은 미래의 무한하게 많은 새로운 경우들에도 동일한 답을 결정하게 한다.

‘ $68+57=125$ ’를 예로 들어보자. 크립키에 따르면, 125는 68과 57의 더하기라는 산술적 의미에서, 또 ‘더하기’는 68과 57이라는 수에 적

1) Kripke, 1982

용될 때 125라는 값을 산출하는 함수를 지시한다는 메타 언어적 의미에서 옳은 답이 된다. 여기서, 즉 메타 언어적 의미에서 역설은 발생한다.

어떤 사람이 과거의 유한한 경우들의 더하기에서 57보다 작은 수만을 가지고 더하기를 해본 경우를 가정하자. 그런데 이 가정은 과거의 더하기 경험을 만족시키면서 동시에 “과거에 나는 ‘더하기(plus)’와 ‘+’를 ‘터하기(quus)’나 ‘#’로 기호화되는 함수를 사용했다”라는 이야기도 성립가능하게 한다. 이 때 과연 ‘68+57’에 대해 ‘125’라는 답을 정당하게 제시할 수 있겠는가?

문제의 결론은 다음과 같다. 57보다 큰 수면 ‘더하기’를 했는지, ‘터하기’를 했는지를 결정할 수 있지만, 57보다 작은 수면 어느 규칙을 따랐는지를 결정할 수가 없게 된다. 편의상 ‘x#y’를 다음처럼 정의하자.

만약 x 나 y 가 57보다 작으면 $x\#y$ 는 $x+y$ 이고, 그렇지 않으면 5이다.

이러한 정의 하에서 과연 우리는 ‘더하기’를 하였는가 ‘터하기’를 하였는가? 더하기와 터하기는 다른 함수임에도 불구하고 어떤 사람의 과거의 모든 더하기 경험들과 터하기 경험들이 서로 양립하게 되고, 57까지의 경우에 있어서는 이 두 함수에 의해 모두 만족된다. 따라서 그 사람의 과거의 모든 경험에 의해서는 이 계산이 더하기인지 터하기인지를 분명히 구별해줄 사실이란 없게 된다.²⁾

이것은 역설이다. 확실히 역설은 발생한다.³⁾ 이러한 크립키의 논변을 원용해보자.

- 1) 장주의 꿈--->나비
- 2) 나비의 꿈--->장주

2) Wiigenstein, p.15. Kripke, 1982, p.55

3) 크립키는 이 역설을 현대적 회의주의로 보고 해결하고자 노력한다. 그러나 그것에 대한 논의는 이 글의 영역밖이다.

꿈이 있었고, 이야기 1이 성립한다면, ‘더하기’나 ‘터하기’의 경우에서 처럼 우리는 1)의 상황인지 2)의 상황인지를 구별할 수 없게 된다. 1)과 2)를 구별할 수 있는 과거, 즉 ‘장주가 꿈을 꾸었다’가 문제를 해결하는 자료로서 제시될 수도 있지만 이것은 이야기 2로 대별되는 무한의 문제에 걸려들게 된다. 앨리스가 먼저였는지 하트왕이 먼저였는지가 문제이듯 역시 장주가 먼저인지 나비가 먼저인지를 우리는 결정할 수 없고, 그럴만한 기준도 우리는 갖고 있지 못하다. 따라서 우리는 현실의 세계와 가상의 세계를 구별할 수 없다. 더군다나 우리가 살고 있는 이 세계가 어떤 세계인지(예를 들어 57보다 큰 수인지 작은 수인지)도 모른다면, 우리는 우리의 세계가 현실 세계인지 가상 세계인지를 묻는 원초적 질문조차도 가능하지 않게 된다. 그러므로 우리는 우리 세계가 현실 세계인지 가상 세계인지도 모른다.

이제 이야기 3이 가능한지와 그것이 함축하는 바를 살펴보자.

4. 통 속의 뇌 논변

퍼트남은 다음과 같은 과학적 공상의 예를 든다.

어떤 인간이 사악한 과학자에 의해 수술을 받았다. 그 사람의 뇌는 육체에서 분리되어 뇌를 계속 살아 움직이게끔 해줄 영양분이 가득 담긴 통 속에 옮겨졌다. 신경 조직은 그대로 초과학적 컴퓨터에 연결되어 이 컴퓨터가 그 사람으로 하여금 모든 것이 완벽히 정상적인 듯이 보이는 환각을 일으키도록 만든다. 이런 상황에서 사람들, 사물들, 하늘 등은 모두 존재하는 것처럼 보이지만 그 사람이 경험하는 모든 것은 컴퓨터로부터 신경 세포에로 이어지는 전기 자극의 결과가 된다. 컴퓨터는 그 사람이 손을 올리려고 한다면 손이 올려짐이 자동적으로 ‘보이고’, ‘느끼게’ 되도록 교묘히 장치되어 있다. 그뿐 아니라 그 사악한 과학자는 여러 가지로 프로그램을 변형시킴으로써 그 사람으로 하여금 과학자가 원하는 어떠한 상황이나 상태일지라도 ‘경험’하도록 할 수 있다. 과학자는 또한 뇌수술을 하였다는 기억을 삭제해 버림으로써 그

사람이 항상 그러한 상황에 있어온 것처럼 보이게 할 수도 있다. 심지어는 그 사람으로 하여금 그가 의자에 앉아 어떤 사악한 과학자가 사람들의 뇌를 떼내어 뇌를 계속 살아 움직이게 할 영양분이 담긴 통 속에 집어 넣는다는 재미나면서도 불합리한 가정을 기술한 글을 읽고 있는 것 같은 착각을 일으킬 수도 있다. 신경 세포는 그 뇌의 주인공으로 하여금 이러저러한 환상을 일으키도록 하는 초과학적 컴퓨터에 연결되어 있다. 뇌 하나가 아니라 모든 인간이 통 속에 들어있는 두뇌라고 상상할 수도 있다. 물론 그 사악한 과학자는 통밖에 있어야 할 것이다. 아마도 사악한 과학자란 존재하지도 않고, 우주는 온통 뇌와 신경 조직으로 가득찬 통만을 만들어 내는 자동 기계로 구성되어 있는지도 모른다. 이번에는 그 자동 기계가 우리로 하여금 서로 연관성이 없이 각각 분리되어 있는 환각들이 아니라 하나의 집단적인 환각을 일으키도록 만들어져 있다고 상상해보자. 그런 상상 속에서는 내가 당신에게 말하고 있는 것처럼 내 스스로에게 보일 때, 당신에게는 당신이 나의 말을 듣고 있는 것처럼 보인다. 물론 이 때 나의 말이 당신의 귀까지 실제로 다다르는 것은 아닐 것이다. 왜냐하면 당신에게는 실제의 귀가 없고 나에게서는 입과 혀가 없기 때문이다. 내가 말을 내 뱃을 때 정말로 일어나는 일은 나의 뇌에서 발생한 전기 자극이 컴퓨터로 나가서 나는 나 자신의 음성을 '듣고', 내 혀의 움직임을 '느끼게' 하고, 당신은 내 말을 '듣고', 내가 말하는 것을 '보게' 하는 것들이다. 이 경우 어떤 의미에서는 당신과 나는 실제로 의사소통을 한다고 할 수도 있다. 당신이 정말 존재한다는 나의 생각은 틀릴 수 없다. 어떤 관점에서 보면 '전세계'가 하나의 집단적인 환각이라 할지라도 문제가 되지 않을 수 있다. 왜냐하면 말하고 듣는 기계적 과정이 우리가 평소 생각하는 바와는 다를지라도 내가 당신에게 무엇을 이야기 할 때 결국 당신은 정말로 나의 말을 듣고 있다고 할 수 있기 때문이다.⁴⁾

다소 길지만 자세한 이 예를 통해 퍼트남은 다음과 같은 질문을 던진다.

정말 우리가 통 속에 들어있는 뇌라고 한다면 그와같은 사실,

4) Putnam, 1981, pp. 5-7.

즉 우리가 통 속의 뇌라는 사실을 우리가 말하거나 생각할 수 있을 것인가?

이러한 질문에 대해 퍼트남은 부정적인 대답을 한다. 그는 우리가 정말로 통 속에 들어 있는 뇌에 불과하다는 가정이 물리적 법칙과 위배되지 않고 우리가 경험해 온 모든 것과 전혀 모순되지는 않더라도, 이 가정은 절대 참이 될 수 없는 가정이라고 주장한다. 왜냐하면 그 가정은 스스로를 논박하는 자기 논박적인 가정이기 때문이다.

자기 논박적인 가정이란 그것이 참이면서 또한 거짓도 함축하고 있는 가정을 말한다. 그 과학자가 구성한 가상 세계에 있는 사람들은 우리가 생각하는 방식대로 생각하고, 우리가 사용하는 어떤 말이라도 '발언'할 수 있다. 그러나 적어도 현실의 우리는 지시할 수 있는 것들을 가상의 그들은 지시할 수 없는 것들이 있다.⁵⁾ 특히 그들은 그들이 통속의 뇌라는 사실을 생각할 수도 말할 수도 없다. 결론적으로 퍼트남은 문제의 가상 세계가 정말로 현실 세계가 되어 우리가 실제로 통 속의 뇌라고 한다면, '우리는 통 속의 뇌'라는 말이 의미하는 것은 우리가 이미지 속의 통에 들어 있는 뇌 또는 그러한 종류의 어떤 것이라는 것이다. 그러나 우리가 이미지 속의 통에 든 뇌가 아니라면, 우리는 통 속의 뇌라는 가정의 일부를 차지한다. 즉 우리가 지금 '환각을 일으키고' 있는 것이 우리가 통 속의 뇌라는 사실은 아니다. 따라서 우리가 만일 통 속의 뇌라면 '우리는 통 속의 뇌'라는 문장은 거짓인 어떤 것을 말한다. 즉 우리가 통 속의 뇌라면, '우리는 통 속의 뇌'라는 것은 거짓이다.

우리가 우리 세계에서 구성하는 가상 세계가 현실화되었을 때 우리가 통 속의 뇌라는 사실을 인지한다는 것은 확실히 자기 논박적이

5) 통속의 뇌가 가진 생각과 실제 세계의 인간이 가진 생각 간에 질적인 유사성이 있다고 해서 양자가 동일한 지시 기능을 가지고 있다고 말할 수 없다. 왜냐하면 통속의 뇌의 경우 내재적으로도 외부적으로도 나무를 표상하지 않는 감각 자료에 언어가 결부되도록 프로그램이 짜여져 있다면 외부적인 어떤 것을 지시하거나 표상하는 언어 체계가 뇌에 형성되는 것이 가능하지 않기 때문이다. 따라서 통 속의 뇌가 외적 사물을 지시한다고 간주할 수 있는 어떠한 근거도 있을 수 없다.

다. 즉 우리는 통 속의 뇌와 같은 가상 세계를 가정하는 것이 불가능하다. 이것을 더 확장하면 우리가 구성하는 가상 세계는 우리의 현실 세계와 구별가능하다는 점에서만 성립한다는 주장을 할 수 있다. 왜냐하면 우리 세계에서 가상 세계는 현실 세계에 대해 열려있지만(현실 세계는 가상 세계에 접근가능하지만), 현실 세계는 가상 세계에 대해 닫혀있기 때문이다(가상 세계는 현실 세계에 접근가능하지 않기 때문이다). 통 속의 뇌 가정은 현실과 가상을 동일화하는 데서 나타나는 역설이다. 따라서 우리의 세계를 기반으로 하면 현실과 가상은 구별불가능하지도 않고, 동일화 자체는 성립하지도 않는다.

5. 세계 간의 접근가능성

이야기 3을 다시 생각해보자. 은석은 게임 속에 가끔씩 혼란을 평정하기 위해서 '솔지'라는 투사를 프로그램해 놓았다. 프로그램상의 솔지는 신의 아들이라고 불리우고, 또 솔지만이 은석의 세계에 접근이 허용된다. 이러한 능력의 솔지는 혼란의 시대에 나타나서 거리를 평정하고는 홀연히 사라진다. 이것을 도식화 해보자.

은석의 세계(세계1) ---> 솔지의 세계(세계2)

세계 2의 구성원들은 세계 1을 천국이라고 부르고, 은석을 신이라고 부른다고 가정하자. 이것도 프로그램화되어 있다고 하자. 또 긴 시간 뒤에 그들에게 종교가 생기기 시작하고 거리의 투사들은 현실의 인간에서와 마찬가지로 나름의 역사도 갖게 되고 많은 상황이 진행된다고 가정하자. 그리고 이것도 프로그램화되어 있다고 하자.

세계 1은 세계 2에 대해서는 닫혀있다. 반면 세계 2는 세계 1에 대해 열려있다. 은석은 세계 2를 좌지우지할 수 있지만, 세계 1은 세계 2에 대해 어떠한 형태로든(솔지의 경우만을 제외하고는) 접근이 허용되지 않는다. 이런 상황에서 세계 2는 세계 1에 의존적이 되고, 세계 2를 구성하고 있는 구성원들의 삶과 죽음은 세계 1의 구성원, 즉 은

석의 의지에 달려있다. 따라서 세계 2의 존재란 한낱 세계 1 속에 있는 은석의 유희에 지나지 않고, 세계 2의 구성원들이 자신의 세계에 서 생각하고 느끼는 존재나 의미들은 아무런 가치가 없게 된다.

6. 괴델 정리의 응용

좀더 실제적인 논의를 위해 프로그래머 은석이 스스로를 기계 속에 구현하는 상황을 생각해 보자. 이 때 은석이 기계라면, 프로그래머 은석과 기계 속에 구현된 은석은 동일한 것이 된다. 그렇다면 기계 속에 구현된 은석은 기계 속의 환경에서 자신이 기계 속에 들어와 있다는 것을 자각할 수 있을까?.

앞에서 우리는 기계 속에 구현된 은석이 스스로가 기계 속에 들어와 있음을 자각할 수 있다는 것을 살펴보았다. 이런 점에서는 가상과 현실 사이의 구별이 가능함을 알았다. 그러나 인간이 기계라면 어떨까? 상황이 달라지겠는가?

이번에는 가상 세계와 현실 세계의 구별가능성 문제가 기계론(machanism)⁶⁾ 논쟁에서도 여전히 동일한 결과를 가져올 것이라는 점을 논증하고자 한다. 우선 '기계'라는 말이 무엇을 의미하는지 알 필요가 있는데, '기계'라는 말은 다음의 특성을 지닌다.

일정한 규칙들이 있다. 기계의 모든 상태가 그 기계의 이전 상태들(과 입력들)로부터 규칙들에 의해 결정된다. 기계의 이러한 특성은 규칙 표상이 명시적으로 나타나는 경우와 그렇지 않은 경우 모두 가능하다.

그리고 기계론의 주장은 다음과 같다.

6) 심리 철학에서 기능주의와 인지주의는 모두 기계론을 가정한다. 전통적 AI도 이 가정을 전제로 성립한다. 따라서 기계론이 반박된다면 심리 철학의 주요 주장들은 치명적인 손상을 입게 된다

인간은 일종의 기계이다. 인간의 활동은 일정한 프로그램에 따른 정보 처리 과정이다. (계산) 기계는 인간 인지 활동에 적절한 모형이 될 수 있다.

이런 주장이 성립하면, 기계를 이용해 인간에게 가상 세계를 만들어 줄 수 있고, 이 경우 인간은 그것을 구별해내지 못할 것이다. 왜냐하면 인간과 기계가 동일한 논리적 구조로 처리될 수 있고, 결국 인간은 기계라는 자기 동일적 상황에 처하기 때문이다. 또 우리가 현재 기반하고 있는 세계가 현실 세계인지 가상 세계인지를 구별할 수 없을 것이다. 왜냐하면 이것은 괴델 정리가 함축하고 있는 내용으로서, 기계가 형식적으로 무모순하게 구성된다면 기계는 자신에 대한 판단을 만들 수 없게 되고, 인간은 기계이기 때문에 결국 인간도 괴델 정리에 따라 자신에 대한 판단을 하는 것에 대해서는 결정할 수 없을 것이기 때문이다. 이 주장은 적어도 우리 세계가 형식적으로 무모순한 세계인 경우에 우리가 살고 있는 세계에 대한 판단은 동일한 위계에서 할 수 없고, 메타적인 위계에서만 가능하다는 것을 함축한다. 따라서 우리는 우리 세계 자체에 대한 판단, 즉 우리 세계가 현실 세계인지 가상 세계인지에 대한 결정을 할 수 없다.

앞서 나는 두 가지를 물었다. 첫째, 우리가 만든 가상 세계와 현실 세계는 구별가능한가? 둘째, 우리가 기반하고 있는 세계가 가상 세계인지 현실 세계인지 구별가능한가? 나는 후자는 수용하지만, 전자는 받아들이지 않는다. 그래서 나의 관심사는 전자가 될 것이고, 따라서 전자에 논의를 집중 할 것이다. 그러나 나는 '인간이 기계'라는 기계론의 타당성 여부를 먼저 검토하는 방법을 택할 것이다. 인간이 기계가 아니라면 현실-가상의 구별가능성 문제는 앞서의 논의만으로도 충분하겠지만 인간이 기계라면 상황은 달라질 것이기 때문에 기계론을 논박하는 것이 순서일 것이다. 따라서 기계론을 논박하는데 성공하면 두 질문에 대한 결론을 자연스럽게 유도할 수 있을 것이다.

1. 괴델 불완전성 정리와 루카스 논변

괴델의 제 1 불완전성 정리는 '산술을 포함하면서 오메가-일관적인 어떠한 형식 체계도 불완전하다'는 명제로 표현할 수 있다. 여기서 '형식 체계는 불완전하다'라는 말은 두 가지의 의미를 갖는다.

- 1) 그 형식 체계에서 증명할 수도 반증할 수도 없는 문장이 존재한다.
- 2) 참이면서 그 형식 체계에서 증명불가능한 문장이 존재한다.

여기서 첫 번째 의미의 산술 체계의 불완전성은 오메가 일관적인 가정을 요구하지만, 두 번째 불완전성은 그보다 약한 가정인 일관성 가정만을 요구한다. 괴델은 산술 체계에서 이 두 가지 의미의 불완전성을 모두 증명한다. 이 두 불완전성의 의미를 갖는 괴델 정리는 다음을 함축한다.

산술을 포함하면서 일관적인 어떠한 형식 체계에도 참이면서 그 체계에서 증명불가능한 산술 문장이 존재한다.

이 불완전성 정리를 증명하기 위해 괴델은 형식 체계의 모든 식과 식들의 나열(sequence)에 자연수를 부여하는 방정식을 만든다. 이렇게 부여된 자연수를 그 식 또는 식들을 나열한 '괴델수'라고 하는데, 이것을 이용 다음과 같은 관계를 구성한다.(여기서 $g(F)$ 를 F 의 괴델수라고 하자).

나열 F_1, \dots, F_n 이 식 F_n 의 증명이다 iff $g(F_1, \dots, F_n)$ 이 $g(F)$ 에 대해 성립하는 수적 관계가 존재한다

이 수적 관계를 증명 관계로 보면, 증명 관계가 산술을 포함하는 체계 내에서 표현가능하다는 것을 보일 수 있다.(R 은 관계를 나타내고, B 는 증명 관계를 표현한다).

$R(x,y)$ 가 성립하면, $B(x,y)$ 가 체계 속에서 증명가능하다. $R(x,y)$ 가 성립하지 않으면, $\neg B(x,y)$ 가 체계 속에서 증명가능할 경우, B 는 R 을 그 체계 속에서 표현한다. 그리고 R 을 그 체계 속에서 표현하는 술어가 존재할 경우, R 은 그 체계 속에서 표현가능하다.

그리고 이것을 이용, $g(G)$ 를 괴델수로 가지는 문장 G 를 구성할 수 있다.

$$G: (x)\neg B(x,g(G))$$

이 문장은 이 문장 자신이 증명불가능할 경우 오직 그 경우에만 성립하는 산술적 사실을 진술한다. 이 문장을 직관적으로 나타내면 다음과 같다.

이 문장은 증명불가능하다.

이 문장을 괴델 문장이라고 한다. 이 괴델 문장 G 는 증명불가능하다. 왜 그럴까? G 가 증명가능하다고 하면, 증명 관계는 표현가능하므로 G 의 증명가능성을 표현하는 문장은 증명가능하게 된다. 그런데 G 의 증명가능성을 표현하는 문장은 바로 G 의 부정이 되므로, $\neg G$ 가 증명가능하게 된다. 따라서 이 체계는 비일관적이게 된다. 이로부터 체계가 일관적이라면 그 체계에 있는 G 는 증명불가능함을 알 수 있다. 여기서 G 는 G 자신이 증명불가능하다는 것을 표현하고 있고 실제로 G 가 증명불가능하기 때문에 G 는 참이 된다. 따라서 G 는 참이면서 증명불가능하게 된다.

루카스는 이 괴델 증명을 기계론을 반박하는데 이용한다.⁷⁾ 어떻게? 인간은 정리를 증명할 줄 아는 기계라고 하자. 그렇다면, 그 기계의

7) Lucas, 1961. 이러한 논변은 네이글과 뉴만에 의해서도 제시되었고, 괴델 자신에 의해서도 받아들여졌던 것으로 보인다. 최근에는 펜로즈도 이 논변을 받아들이고 있다.

정리 증명 과정은 일정한 규칙들에 따라 이루어 지고, 그것에 대응해 그 기계의 규칙들로 이루어진 형식 체계 S가 존재하게 된다. 괴델 정리에 따라 인간은 S의 괴델 문장 G를 구성할 수 있고, G가 S에서 증명불가능하다는 것을 알 수 있고, G가 표현하는 바에 따라 G가 참이라는 것도 알 수 있다. 따라서 S를 구현하는 기계는 G를 증명할 수 없지만, 인간은 G가 참임을 알 수 있게 된다. 이렇다면 인간은 인간과 동일하다고 가정된 그 기계와 동일할 수 없게 되고, 이로부터 인간은 기계가 아니라는 결론을 끌어낼 수 있다.

루카스는 어떠한 일관적인 기계도 자기가 구현하는 형식 체계의 괴델 문장, 즉 괴델 정리를 증명할 때 핵심적으로 사용되는 괴델 문장이 참임을 알 수 없지만 인간은 그것이 참임을 알 수 있기 때문에 인간은 할 수 있지만 기계는 할 수 없는 어떤 일이 존재하게 되고, 따라서 인간은 기계가 아니라고 주장한다.

여기서 다시 은석이 스스로 자신이 만든 가상 세계 속에 들어갔다는 것을 생각해보자. 루카스가 맞다면 인간과 기계를 동일하게 보는 근거는 사라지게 되기 때문에 프로그래머 은석(현실의 은석)은 기계 속에 구현된 은석(가상의 은석)과 본질적으로 구별될 수 있는 근거를 갖게 된다. 프로그래머 은석은 프로그래머 은석 이외에 아무것도 아니다.

2. 루카스에 대한 반론과 응답

루카스 논변은 상당히 재미있고, 설득력있어 보인다. 그러나 이에 대한 반론도 만만치 않다. 여기서는 베나세라프나 루이스, 보이어 등의 반론을 중심으로 다룸으로써 여전히 기계론이 성립할 수 없음을 보이도록 하겠다. 반론으로 두 가지 경우를 생각해보자.

1) 기계도 다른 기계를 괴델화할 수 있다.⁸⁾

8) Benacerraf, 1967. 이 주장은 베나세라프 외에 슬레작(Slezak)에 의해서도 제시된다. 그의 논변의 핵심은 루카스의 괴델 논변이 마음이 개별적 기계가 아니라는 것을 보여줄지는 몰라도 마음이 기계 유형에 속하지 않는다는 것까지 보여주지는 않는다는 것이다.

- 2) 기계도 어떤 방식으로는(비형식적으로는) 괴델 문장을 알 수 있다.

먼저 1)에 대해 살펴보자. 만약 한 형식 체계가 일관적이라면 그 형식 체계 내에서 그 형식 체계 내의 괴델 문장을 증명할 수 없다. 그러나 이로부터 괴델 문장을 증명할 수 있는 규칙을 가진 다른 형식 체계가 존재하는 것이 불가능하다는 주장이 나오는 것은 아니다.(물론 그 형식 체계의 괴델 문장은 따로 있을 것이고, 그 새로운 괴델 문장은 다시 그 체계에서 증명불가능할 것이다.) 왜냐하면 한 체계가 다른 체계를 메타적으로 증명할 수 있기 때문이다. 따라서 한 기계는 자기 자신의 괴델 문장을 증명할 수는 없지만, 다른 기계는 그 기계의 괴델 문장을 증명할 수 있다. 그렇다면 프로그래머 은석이 자기에게 주어진 괴델 문장이 참임을 안다고 하더라도, 그것이 프로그래머 은석이 또 하나의 기계라는 것을 배제한다는 것은 아니다.

이 반론이 타당하려면 다음의 상황을 고려 해야 한다.

기계론이 맞다면,

- 1) 프로그래머 은석(기계)은 프로그램 속의 은석(기계)을 괴델화할 수 있다.
- 2) 프로그램 속의 은석(기계)이 프로그램 속에 있는 모든 면에서 동일한 또 다른 쌍둥이 은석(기계)을 괴델화할 수 있다.
- 3) 프로그램 속의 은석(기계)은 프로그램 속에서 또 다른 상황을 설정해서 다른 은석(기계)을 괴델화할 수 있다.

그러나 이러한 반론은 우선 루카스가 제기한 괴델 논변의 핵심이 단순히 어떤 하나의 기계와 프로그래머 은석이 동일하지 않음을 논증하는 것이 아니라 주어진 임의의 기계에 대해 그 기계와 프로그래머 은석이 동일하지 않음을 증명할 수 있다는 것을 증명하는 데 있음을 간과하고 있다. 따라서 1)과 3)은 동일 유틸리티 때문에 루카스에 대한 반론이 되지 못한다. 다음으로 2)의 경우에 예를 들어 쌍둥이 은석 1이 쌍둥이 은석 2를 괴델화한다면, 그 둘은 모든 면에서 갖기 때문에, 은석 1은 은석 2를 괴델화하는 순간 자기 논박적인 상황에 빠

지게 된다. 따라서 2)의 경우도 성립하지 않기 때문에 반론은 성립하지 않게 된다.

그렇다면 이런 경우는 어떤가? 루카스 산술이라 불리는 것들 모두가 참임을 알 경우에만 은석이 기계가 아니라는 것이 입증된다. 왜냐하면 그렇지 않을 경우, 은석에 의해 참임이 알려지는 괴델 문장들 모두를 증명하는 기계가 있을 수 있고, 은석은 바로 그 기계와 동일할 수 있기 때문이다.⁹⁾ 그런데 은석은 유한한 존재이기 때문에 모든 기계들의 괴델 문장을 모두 산출하여 그것들이 모두 참임을 알 수 없다. 따라서 루카스 논변은 틀렸다.

이에 대해 우리는 은석이 기계들에 대해 그 기계들의 괴델 문장이 참임을 알고 있을 필요는 없다는 것을 생각해야 한다. '은석이 어떠한 기계와 동일하다'고 말할 때마다 그 주장을 반박할 수 있으면 된다. 그리고 사실 어떠한 기계도 기계론자의 임의(any)의 제안에 대해 그러한 일을 할 수는 없을 것이다.¹⁰⁾

그러나 루카스가 임의(any)의 기계에 대해 그 기계의 괴델 문장을 알 수 있다는 것은 참인가라는 물음은 가능하다. 사실 한 기계의 괴델 문장이 무엇인지를 알기 위해서는 기계의 프로그램을 알아야 한다. 그런데 우리는 자기 자신의 프로그램이 무엇인지 원리적으로 알지 못하는 기계일 수 있다. 그 경우에 은석은 자신의 괴델 문장을 산출하지 못하는 기계일 수 있다.¹¹⁾

이 질문은 상당히 미묘한 것이다. 그러나 우선 은석과 같다고 주장된 기계의 프로그램이 어떤 것인지 말할 책임은 기계론자에게 있는 것이지 루카스가 져야 할 부담은 아니다. 그리고 은석이 자기 자신을 구현한 프로그램이 무엇인지 원리적으로 알지 못하는 기계일 수 있다면, 첫째 은석은 처음부터 자기 자신을 구현하는 프로그램을 구성할 수 없었을 것이고, 둘째 따라서 인간이 기계인가의 질문이 형성될 수도 없을 것이다.

이제 2)에 대해 살펴보자. 루카스의 논변은 "나는 문장 G를 '알'(혹

9) Lewis, 1969

10) Lucas, 1970

11) Benacerraf, 1967

은 비형식적으로 증명할) 수 있는데 기계는 문장 G 를 '증명할'(혹은 형식적으로 증명할) 수 없다는 근거에서 내가 기계와 서로 다르다"고 말한다. 이 논변은 '안다'와 '증명된다'를 구분하는 데서 성립한다. 따라서 엄밀히 하자면 나와 기계의 비동일성을 입증하기 위해서는 둘이 서로 동일한 측면에서 다르다는 것을 말해야 한다. 이 점에서 두 가지를 경우를 고려할 수 있다.

- 1) 나는 G 를 알 수 있는데 기계는 G 를 알 수 없다.
- 2) 나는 G 를 증명할 수 있는데 기계는 증명할 수 없다.

이것을 바탕으로 반론을 구성¹²⁾해 보자. 괴델의 제 1 정리에 의해 기계(기계 속에서 구현된 은석)가 하지 못하는 것으로 말할 수 있는 것은 기계의 규칙들을 사용해 기계의 공리들로부터 H (기계의 괴델 문장)를 증명하는 일일 것이다. 그러나 그것은 프로그래머 은석도 할 수 없다. 프로그래머 은석이 할 수 있는 것은 아마도 H 에 대해 비형식적(기계의 체계에서 형식화될 수 없다는 의미에서 비형식적인) 증명을 하는 일일 것이다. 그러나 기계도 역시 이것을 할 수 있다. 기계는 기계-형식적 증명들을 제시함에 있어 제한을 가진다. 하지만 이러한 제한은 프로그래머 은석이 제시할 수 있는 비형식적(기계에서 형식화될 수 없는) 증명들에 대해서는 존재하지는 않는다. 프로그래머 은석은 H 가 비록 기계에서 증명불가능하지만 그럼에도 불구하고, 바로 그 이유로해서 참이라는 것을 납득할 수 있다.

이 반론은 기계는 G (그 기계의 괴델 문장)를 형식적으로 증명할 수 없는데, 인간은 G 를 비형식적으로 증명할 수 있다는 것이 루카스 논변이 기반으로 하는 유일한 근거라고 비판하고 있다. 그리고 이 점에서 베나세라프는 두 가지를 지적한다.

- 1) 인간도 기계와 마찬가지로 G 를 형식적으로(기계의 형식 체계를 써서) 증명할 수 없다.

12) 이것은 베나세라프가 제시한 내용을 논문의 내용에 맞추어 약간 변형하였다. 베나세라프, 1967, 19-20.

2) 기계도 인간과 마찬가지로 G를 비형식적으로 증명할 수 있다.

1)에 대해 실제로 인간 역시 형식 체계 규칙들을 가지고 그 체계의 공리들로부터 그 체계의 괴델 문장을 증명할 수 없다는 것은 명백하다. 그것은 바로 괴델의 제1 정리가 말하는 바이다. 그러나 2)도 성립하는가?

기계(튜링 기계)의 정의상 모든 기계 상태는 기계의 이전 상태에서부터 일정한 규칙들에 의해 결정된다. 따라서 어떤 문장에 대해 그것이 참이라고 납득하거나 그것을 비형식적으로 증명하는 기계 상태도 규칙들에 의해 결정되어야 한다. 그렇다면 그 기계에 의해 '비형식적으로' 증명된 것들까지 포함해 그 기계에 의해 증명된 문장들의 집합을 정리들의 집합으로 가지는 형식 체계가 존재할 것이다. 이 형식 체계가 '그 기계가 구현하는 형식 체계'다. 그리고 그 기계의 괴델 문장이 바로 이 형식 체계의 괴델 문장이다. 그렇다면 괴델의 제 1 정리에 의해 그 괴델 문장은 그 형식 체계의 정리 집합에 속해 있지 않을 것이고, 따라서 비형식적으로 증명된 문장들 중에 괴델 문장은 없을 것이다. 이것은 한 기계의 괴델 문장은 그 기계에 의해 비형식적으로조차 증명될 수 없다는 것을 뜻한다. 그리고 사실 비형식적 증명이라는 것이 기계가 구현하는 형식 체계밖에서의 증명을 의미한다면 기계가 비형식적 증명을 한다는 것 자체가 가능하지 않다.

베나세라프의 이 반론은 만만치 않아 보인다. 그러나 여기서 기계(또는 인간)가 '사용하는' 형식 체계와 기계(또는 인간)가 '구현하는' 형식 체계가 다르다는 것¹³⁾을 구분해야 한다. 괴델 논변이 전개될 때

13) 어떤 기계가 명제들의 표상을 조작함으로써 최종적으로 정리인 명제들을 산출한다고 하자. 그리고 그렇게 하면서 조작 규칙들 자체를 표상한다고 하자. 그러면 그 기계는 표상 차원의 명시적 규칙들을 사용해 정리들을 증명한다. 그 경우 그 체계는 그 규칙들로 이루어진 형식 체계를 '사용'한다고 말할 수 있다. 그 기계가 만약 또 다른 명제들을 그 형식 체계(어떤 명제들을 증명하는 데에 사용되었던)의 규칙들을 사용하지 않고서 산출할 경우 우리가 원한다면 그 명제들은 기계에 의해 '비형식적으로' 증명되었다고 말할 수 있을 것이다. 물론 그 때의 비형식적 증명이라는 것은 그 기계가 어떤 명제들을 증명하는 데에 '사용'하는 형식 체계밖에서의 증명을 의미한다. 그러나 그런 의미에서 기계가 '비형식적 증명'을 한다고 하더라도 그런 증명 과정은 기계에 고유한 일정한

문제가 되는 것은 기계가 '구현'하는 형식 체계의 괴델 문장이지 기계가 '사용'하는 형식 체계의 괴델 문장이 아니다. 베나세라프는 이점을 간과하고 있다. 그는 '기계 형식 체계'라는 말을 우리가 산술의 정리를 증명할 때 페아노의 산술 체계를 사용하듯 기계가 사용하는 형식 체계를 의미하는 듯이 말한다. 그러나 그 기계 형식 체계는 기계가 구현하는 형식 체계다.

그렇다면 인간과 기계에 있어 동일한 측면에서의 비교가 가능한가? 그렇지 않다는 가능성을 살펴보자.¹⁴⁾ 프로그래머 은석과 그를 모의하는 기계 속에 구현된 은석을 비교할 때, 인간의 증명과 기계의 증명 간에 애매성이 존재할 가능성이 있다. 그리고 어떤 것을 인간의 증명에 상응하는 기계의 증명으로 삼을지에 대한 애매성이 존재할 가능성이 있다. 예를 들어 프로그래머 은석이 칠판에 문장들을 증명하는 것에 대한 사진을 찍어냄으로써 기계 속에 구현된 기계가 프로그래머 은석을 모의하고 있다고 하자. 그 경우 그 사진 속의 칠판에 증명되는 문장들의 집합이 기계 속의 은석이 증명하는 문장들의 집합으로 해석될 수도 있고, 그런 사진을 찍어 내기 위해 보내는 0과 1의 전기 신호가 일차 술어 논리의 문장으로 해석되었을 때의 문장들의 집합이 기계 속의 은석이 증명하는 문장들의 집합으로 해석될 수도 있다.

그러나 이것이 바람직한 설정인가? 문제는 우리가 지금 탁자 위에 놓여 프린터와 연결되어 있는 '정리 증명' 기계를 인간과 비교하고 있는 것이 아니라는 점이다. 우리는 기계가 산출하는 문장들과 인간이 마음 속에 믿고 있는 문장들을 비교할 필요가 없다. 기계론의 논제는 인간이 기계라는 것이고, 기계론을 반박하기 위해 우리가 인간과의 비교 대상으로 삼아야 할 기계는 기계론에 의해 인간과 동일하다고 가정된 기계이다. 따라서 그 기계는 먹고 말하고 숨쉬는 기계다. 이

규칙들을 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.

14) Boyer, 1983

때 우리는 인간에 있어서의 앎(또는 증명)을 기계인 그 인간(기계 속에 구현된 인간)에 있어서 앎(증명)이라고 부르는 것과 비교한다. 물론 여기에 애매성이 있다. 그런데 인간에게서 무엇을 앎이라 할 것인지의 애매성과 그 기계인 인간에게 무엇을 앎이라 할 것인지의 애매성은 동일하다. 그렇다면 보이어가 말하는 종류의 애매성은 문제가 되지 않는다.

7. 맺는 글

이상에서 우리는 가능 세계와 현실 세계의 구별가능성에 대한 문제를 여러 각도에서 고찰해 왔다. 퍼트남의 견해에 따르면 현실 세계와 동일한 가상 세계란 가능하지도 않다. 그러나 이런 질문은 가능하다. 지금 우리의 세계가 가상 세계는 아닐까? 크립키의 논변은 이러한 회의주의적 시각이 꽤 견고한 것임을 보여준다. 이로부터 우리는 우리의 세계가 무엇인가 의해서 계획되어진 것인지, 만약 그렇다면 어떠한 연유로 인해 그렇게 된 것인지를 결정할 아무런 자료나 규준이 우리에게 없는 주장도 도출할 수 있다. 따라서 어떤 세계에 우리가 존재하게 되었을 때 그 세계가 현실 세계인지 가상 세계인지 우리는 구별할 수도, 또 그 세계가 현실 세계인지 가상 세계인지를 인식할 수도 결정할 수도 없다.

가상 세계는 현실 세계일 수 없다. 현실 세계라는 착각을 할 수는 있으나, 본질적으로 가상 세계는 현실 세계일 수 없다. 왜냐하면 현실 세계는 가상 세계에 접근할 수는 있지만 가상 세계는 현실 세계에 대한 접근이 허용되지 않기 때문이다. 즉 현실과 가상은 그 위계에 있어서 존재론적 정도(degree)가 다르기 때문이다. 이 위계에 있어서의 존재론적 정도가 다르다는 점은 괴델 정리에 의해 함축된다. 루카스는 이 괴델의 정리를 이용하여 인간과 기계가 다르다는 점을 밝힌다. 그리고 이로부터 나는 우리가 만드는 가상과 현실의 구별가능성과 우리가 기반하고 있는 세계가 가상인지 현실인지에 대한 구별불가능성

을 도출하였다.

우리가 구성하는 가상 세계란 현실 세계와는 다르고, 다른 만큼 구별가능할 뿐이다. 그러나 우리가 어느 곳에 서 있는지는 알 수 없다. 이것은 우리에게 끊이 없는 불안을 줄 것이다.

참고문헌

- Benacerraf, 1967, "God, the Devil, and Goedel", *Monist* 51, 9-32
- Boyer, 1983, "Lucas, Goedel, and Astaire", *Philosophy Quarterly* 33, 147-159
- Kripke, S.(1982). *Wittgenstein on Rules and Private Language*, Cambridge: Harvard Univ. Press.
- Lewis, 1969, "Lucas against Mechnism", *Philosophy* 44, 231-233
- Lucas, 1961, "Minds, Machines and Goedel", *Philosophy* 36, 112-127
- _____, 1968, "Satan Stultified: a Rejoinder to Paul Benacerraf", *Monist* 52, 145-158
- _____, 1970, "Mechnism, A Rejoinder", *Philosophy* 45, 149-151
- Putnam, H. (1981). *Reason, Truth, and History*, Harvard: Cambridge Univ. Press.