철학적 분석 8 (2003) pp.93-113.

펜로우즈 읽기: 환원주의자인가, 비환원주의자인가*†

송 하 석‡

비환원적 유물론의 매력은 인간의 자율성을 구하면서 심적 현상을 물리적으로 설명할 수 있다는 것이다. 최근 펜로우즈는 뉴런의 하위 단계인 미세관의 현상을 통해서 인간의 의식을 설명하고자 시도한다. 그는 미세관에서 발생하는 현상은 양자역학적으로 설명될 수 있을 것이라고 기대되는데, 이에 따라 인간의 의식도 양자역학적으로 실현될 것이라고 주측한다.

김광수는 이러한 펜로우즈의 논변에서 성공적인 비환원적 유물론의 가능성을 발견하고, 지금까지 신비에 묻혀있던 의식의 뿌리를 설명할 수 있으리라고 기대한다. 나아가서 김광수는 펜로우즈의 이론을 인간의 정신의 자율성을 포기하지 않으면서 심적 현상을 물리적으로 설명할 수 있는 성공적인 비환원적 유물론으로 해석하고자 한다. 반면에 백도형은 그러한 김광수의 해석을 비판하면서, 펜로우즈는 환원적 유물론자라고 해석해야 한다고 주장한다.

이 논문은 펜로우즈를 비환원주의자로 읽어야 하는 근거를 그의 1994년 저서 *Shadows of the Mind*에서 제시된 주장을 자세히 분석함으로써 제공할 것이다. 필자는 펜로우즈가 미세관의 양자역학적 속성에 의해서 의식이 창발된다고 주장하는 비환원적 유물론자임을 보일 것이다. 그러나 펜로우즈의 논변은 많은 문제점이 있기 때문에 김광수가 기대하는 것처럼 성공적인 비환원적 유물론라고는 할 수 없음을 보임으로써, 인간 정신의 자율성을 유지하면서 심적 현상을 과학적으로 설명할 수 있기 위해서는 또 다른 비환원적 유물론을 기다릴 수밖에 없음을 논증한다.

주 제 심리철학, 과학철학

주요어 의식, 환원주의, 비환원적 유물론, 미세관, 양자역학적 현상, 펜로우즈, 김광수, 백도형

---

* 접수완료: 2003. 9. 4. / 심사 및 수정완료: 2003. 11. 17.

† 이 논문을 읽고 심사를 해주신 익명의 심사위원들께 감사를 드린다. 그들의 지적은 이 논문을 보다 명료하게 하는 데 큰 도움이 되었다. 그리고 이 논문을 쓰기까지 많은 분들의 도움이 있었다. 펜로우스의 글과 처칠랜드의 글을 함께 읽어 준 심철호 교수(동해대), 현대 물리학에 대한 질문에 답해준 고인석 박사, 현대 생물학에 관해 설명과 자극을 준 김한집 교수(아주대)와 아내 김영에게 감사를 드린다.

‡ 아주대학교 교양학부 강의교수.

1. 무엇이 문제인가

의식에 대한 펜로우즈(R. Penrose)의 저술이 발표된 이후, 그에 대한 여러 가지 평가와 논쟁이 있었다. 최근 우리나라에서 심신문제에 대한 비환원적 유물론, 혹은 속성 이원론적 입장을 유지하고자 하는 몇몇 철학자들은 펜로우즈를 통해서 자신의 입장을 뒷받침할 수 있으리라고 기대하면서 펜로우즈를 비환원주자로 해석하고 평가하는 한편, 속성 이원론에 대하여 비판적인 입장을 취하는 철학자들은 펜로우즈의 의식에 대한 이론은 환원주의적이라고 주장한다. 전자의 입장에 서는 대표적인 철학자는 김광수이고, 후자의 입장을 대표하는 철학자는 백도형이다.

이 글의 목적은 김광수의 펜로우즈 읽기와 백도형의 펜로우즈 읽기를 비교하여 검토하고, 펜로우즈의 정확한 정체를 밝혀 보는 것이다. 필자는 펜로우즈를 비환원주자로 읽는다는 점에서는 김광수에 동의한다. 그러나 김광수가 제시한 펜로우즈의 논증은 지나치게 단순화된 것이어서 오해의 여지가 있으며, 심지어 펜로우즈의 주장으로 여길 수 있는지 의심스럽기조차 하다. 또한 필자는 펜로우즈의 입장을 비환원주의적이라고 읽는다고 할지라도 그것이 김광수의 기대와 달리 인간의 자율성을 포기하지 않으면서 인간의 의식을 물리적으로 설명하는 데 성공하지 못했다고 생각한다. 그런 의미에서 펜로우즈에 의지해서 인간의 자율성을 구하려는 물리주의적 시도는 성공적일 수 없다고 보는 점에서는 백도형의 주장이 옳다. 요컨대 필자는 펜로우즈는 의식에 대한 비환원적 물리주의자라고 보지만, 그의 논증이 성공적이지 못하기 때문에 그에 의존해서 인간의 자율성을 확보하면서 비환원적 물리주의를 옹호하려는 시도는 성공적일 수 없다고 생각한다.

그러므로 이 글은 우선 김광수와 백도형의 펜로우즈 읽기를 비판적으로 살펴보고(2절, 3절), 인간의 의식에 대한 펜로우즈의 설명이 성공적인지 검토해 볼 것이다.(4절) 그의 논증 중에서 전반부는 괴델(K. Gödel)의 불완전성 정리에 의지하여 인간의 의식이 기계적 계산과정일 수 없다고

주장하는 부분인데, 이에 대한 논의는 굉장히 많이 이루어진 반면, 나머지 부분에 대해서는 과학적 전문성이 요구되는 탓인지 철학계에서 별반 논의되지 못하고 있는 것이 사실이다. 그러므로 이 글은 주로 의식에 대한 설명이 현대 물리학의 범위를 넘어서 미세관의 양자역학적 행태에 의존한다는 그의 논증의 후반부에 대한 비판적 평가를 중심으로 그의 논증을 살펴볼 것이다.

2. 김광수의 펜로우즈 읽기

20세기 중반 환원적 유물론이라고 불리는 심신 동일론이 등장한 이래, 인간의 의식을 과학적으로 설명하는 유물론(혹은 물리주의)을 포기하지 않으면서 인간에게 부인할 수 없는 것처럼 보이는 자율성, 즉 인간의 자유의지를 유지할 수 있는 이론의 등장을 기대하는 많은 철학자들이 있었던 것같다. 데이빗슨(D. Davidson)의 무법칙적 일원론은 그러한 기대에 부응하는 것처럼 보였고, 따라서 지난 세기말 대단한 관심의 대상이 되어 수많은 찬반 논쟁을 불러일으킨 것은 그런 의미에서 당연한 것이었다.

김광수도 인간의 자율성을 유지할 수 있는 심신 문제에 대한 설득력 있는 이론을 모색한 결과, 펜로우즈로부터 자신의 요구를 충족시켜줄 수 있는 이론을 얻을 수 있으리라고 기대하는 것 같다. 김광수에 따르면, 펜로우즈는 인간의 정신은 계산적, 알고리즘적이지 않으며, 따라서 인간의 정신은 기계적 계산절차일 뿐인 컴퓨터 프로그램일 수 없다는 주장을 제기하고 나아가서 비계산적인 인간의 의식을 과학적으로 설명하고자 시도했다. 펜로우즈의 매력은 그의 이론이 옳다면 환원적 유물론이 지니는 의식에 대한 과학적 설명력을 유지한 채, 인간의 자율성을 인정할 수 있다는 데 있다. 즉 그의 이론은 일반적으로 환원적 유물론과 자유의지 사이의 다음과 같은 갈등을 해소할 수 있는 길을 제시해주는 것처럼 보인다는 것이다.

(논증1) 환원적 유물론은 결정론을 함축한다.
자유의지는 결정론과 양립할 수 없다.
그러므로 자유의지의 인정과 환원적 유물론은 양립할 수 없다.

이 논증에 따르면, 우리는 환원적 유물론을 포기하거나 자유의지를 포기해야 한다.[1] 즉 환원적 유물론은 인간을 포함해서 "존재하는 모든 것은 **자연의 인과법칙에 따라** 물질로부터 발생하는 것"이라고 주장하는 반면, "인간은, **자연의 법칙에 따라 기계적으로 작동하는 타율적 존재가 아니라**, 자신의 방식으로 문제 상황을 인식하고 자신의 신념과 소망으로부터 가장 합리적인 판단을 도출하여 행위하는 자율적 존재"[2]라는 믿음을 갖는 정신의 자율성을 주장하는 것 사이에는 명백한 갈등이 있다는 것이다. 그런데 김광수는 유물론은 하나의 이론, 혹은 가설인 반면, 인간 정신의 자율성은 사실이기 때문에 두 개가 양립할 수 없다면 자유의지를 인정하고 유물론을 포기하거나 자유의지를 인정할 수 있도록 유물론을 수정하는 것 이외의 선택의 여지는 없다고 단언한다. 심지어 그는 사실로서의 인간의 자유의지를 포기하고 유물론을 고수하는 것은 '철학적 스캔들'이라고 주장하기조차 한다. 그리하여 김광수는 펜로우즈에

---

1) 사실 논증1의 두 전제는 모두 논쟁적이다. 자유의지와 결정론이 양립가능하다고 주장하는 양립론자(compatibilist)들이 있기 때문에 두 번째 전제도 자명한 것으로 생각할 수 없다. 또 논증의 첫 번째 전제도 의심할 수 있다. 환원주의도 약한 결정론(soft determinism)과 양립가능하다고 주장하는 이들이 있기 때문이다. 그러므로 이 논증이 자유의지를 인정하면서 환원적 유물론을 견지하려는 철학자들을 심한 곤경에 빠뜨리지는 않을 것이라고 생각된다. 익명의 심사위원 중의 한 분은 필자가 자율성과 환원주의를 양립불가능한 대립쌍으로 보는 김광수의 문제제기를 쉽게 받아들이고 있다는 지적을 해주셨다. 그러나 필자는 김광수의 문제의식이 옳다고 보지 않았기 때문에 위와 같은 주를 붙인 것이고, 심사위원이 지적한 것처럼 필자도 심신동일론의 매력 중의 하나는 자율성과 인과결정론의 해묵은 대립의 문제를 일거에 해소하는 것이라는 데 동의한다. 결국 필자도 자율성과 환원주의의 문제는 분리되어야 한다고 생각한다. 다만 이 주제는 이 논문의 요지가 아니기 때문에 주에서만 간단히 지적했던 것이다.

2) 김광수, (2003), p.4. (두 입장 사이의 갈등을 분명하게 드러내기 위해서 강조는 필자가 한 것임).

의지하여 인간 정신의 자율성을 인정하면서 과학적으로 인간의 의식을 설명할 수 있으리라는 낙관적인 기대를 피력하고 있다.[3)]

필자도 인간 정신의 자율성은 포기할 수 없는 것이고, 만약 그것이 유물론과 갈등을 일으킨다면 유물론이 포기되거나 수정되어야 한다고 믿는다. 만약 펜로우즈의 논증이 옳다면, 그것은 이런 의미에서 대단히 매력적일 것이다. 그러면 김광수가 요약하여 제시하는 펜로우즈의 논증을 살펴보자.

(1) 인간의 정신은 물질의 산물이다.
(2) 인간의 정신은 물질을 지배하는 법칙으로부터 자유롭다.
(3) 의식을 실현시키는 물리적 상태는 계산불가능하다.
(4) 양자역학적으로 공존하는 여러 상태들 중에서 하나가 결정되는 과정은 계산 불가능하다.
∴ (5) [3,4로부터] 의식은 양자역학적으로 실현된다.
(6) 뇌 안은 매우 '시끄럽다.'
(7) 미세관은 뇌세포 내에서 뇌 안의 소음이나 미세관을 이루고 있는 단백질 분자들에 의해서 방해 받지 않고 양자역학적으로 공존하는 다수의 정보를 전달한다.
(8) 만일 미세관이 뇌세포 내에서 뇌 안의 소음이나 미세관을 이루고 있는 단백질 분자들에 의해서 방해받지 않고 양자역학적으로 공존하는 다수의 정보를 전달할 수 있다면, 뇌 안이 매우 시끄러 울지라도 이는 의식이 양자역학적으로 실현된다는 가설에 대한 부정적 증거가 되지 못한다.
∴ (9) 의식은 미세관 속에서 양자역학적으로 실현된다.[4)]

펜로우즈의 이론을 이렇게 정리하면서, 김광수는 (1)과 (2)는 전제로

3) 김광수, (2000), p.135 이하.
4) *Ibid.*, p.138.

나와 있지 않지만, (1)은 펜로우즈가 그렇게 생각하고 있고, (2)는 대체로 받아들여지기 때문에 이와 같이 요약된 것을 펜로우즈의 논증으로 여겨도 좋을 것이라고 말한다. 여기서 (1)과 (2)를 덧붙인 이유는 그의 논문의 관심사가 인간의 자율성에 대해 논증하는 것이기 때문일 것이다. 그리고 김광수는 펜로우즈가 자신의 가설 (5)에 대한 부정적 증거인 (6)을 미세관에 대한 해머로프(S. R. Hameroff)의 연구에 의존하여 제거하고, 새롭게 (9)라는 가설을 제시한다고 말한다.

김광수가 펜로우즈의 논증으로 제시한 위의 논증은 우선 몇 가지 점에서 문제가 있다. 김광수는 (3)이 증명된 것이 아니라 일종의 가설이라고 말한다. 그러나 펜로우즈는 괴델의 불완전성 정리를 원용하여 인간의 마음은 계산적일 수 없다고 논증한 루카스(J. Lucas)의 논변을 받아들이면서 (3)에 대한 증명을 제시한다. 그 증명 과정이 옳은지는 별개의 문제이고, 적어도 펜로우즈는 (3)을 단순히 가설로 제시한 것이 아니라는 것은 분명하다. 또 김광수가 제시한 (3)-(5)의 논증은 타당한 논증이 아니다. 펜로우즈가 그런 부당한 논증을 실제로 자신의 논변으로 제시했다고 해석하는 것은 근거도 없고, 옳지도 않다. 펜로우즈의 논증을 정확하게 요약하자면 다음과 같이 될 것이다.

(3) 의식을 실현시키는 물리적 상태는 계산불가능하다.
(4') 뉴런 단계에서 발생하는 물리적 현상은 계산가능하다.
∴ (5') 의식은 뉴런 단계에서 발생하는 물리적 현상에 의해서 실현되지 않는다.

그리고 펜로우즈는 이 논증의 결론인 (5')으로부터 (5)를 추리, 혹은 추측하고 있다고 보아야 할 것이다.

김광수도 (3)-(5)의 논증이 타당하지 않다는 것을 인정하지만, 그럼에도 불구하고 (5)를 포기할 수 없는 이유는 그것이 갖는 몇 가지 긍정적인 면 때문이라고 말한다. 그리고 그 긍정적인 면이란 (5)가 인간의 자율성

을 보장하는 옳은 방향에 서있다는 사실과 해머로프의 미세관 현상이라는 증거가 있다는 것이라고 말한다. 그러나 엄격히 말하면, 인간의 자율성을 보장하는 데 기여하는 전제는 (3), 즉 인간의 의식을 실현시키는 과정은 계산불가능하다는 것이지, 의식이 양자역학적으로 실현된다는 (5)가 아니다. 다시 말해서 인간의 의식을 실현시키는 과정이 계산 불가능할지라도, 인간의 의식이 양자역학적으로 실현되지 않을 수 있는 논리적 가능성은 얼마든지 있다. 그러므로 인간의 자율성을 유지하기 위한 옳은 방향이기 때문에 (5)를 받아들여야 한다는 김광수의 주장은 설득력이 없다. 또한 해머로프 자신이 인정하고 있는 것처럼, 초복사와 같은 양자 정합성이 미세관에서 발생한다는 주장에 대한 명백한 어떠한 경험적 증거도 없고, 그것은 단지 그럴 가능성이 있다는 추측일 뿐이기 때문에, 해머로프의 증거가 의식이 양자역학적으로 실현된다는 펜로우즈의 주장을 받아들이게 할 만한 근거가 아니며, 따라서 그로부터 의식의 뿌리가 밝혀질 수 있으리라고 기대하는 것은 성급한 기대가 아닐 수 없다.[5]

요컨대 김광수가 제시한 펜로우즈의 요약 논증은 지나치게 단순화되어 있을 뿐만 아니라, 펜로우즈의 논증이라고 보기도 어렵다. 또한 그에 대한 김광수의 기대는 펜로우즈의 논증이 성공적이지 못하는 한, 무망한 것이다. 펜로우즈 자신의 논증은 나중에 자세히 소개하겠지만, 그는 적어도 몇 단계의 논증 과정을 통해서 (i) 인간의 의식은 계산 불가능하고 따라서 현재의 물리학의 이해를 넘어서며, (ii) 미래의 물리학 이론은 비알고리즘적인 과정을 설명할 것이고, (iii) 미세관은 양자역학과도 관계하고 또한 의식과도 관계하기 때문에 의식의 비알고리즘적인 성질을 설명

5) 김광수는 다시 김광수 (2003) 이라는 글을 통해서 백도형의 비판에 대해서 응답하면서, 자신은 펜로우즈를 통해서 자율성을 구하려고 하는 것이 아니라고 말한다. 그는 "아무도 자율성을 구하지 않는다. 모두 가지고 있기 때문이다. 다만 펜로우즈의 연구가 정신의 자율성을 설명하는 데 도움이 되지 않는다면, 이는 다름 아닌 유물론자들에게 실망스러운 일이 될 것이다"고 말한다. 결국 김광수 (2000)에서 펜로우즈를 통해서 의식의 뿌리를 밝힐 수 있으리라고 기대하면서 제시한 자신의 논증은 그다지 심각한 평가를 필요로 하는 것이 아니라고 물러서고 있는 셈이다. 김광수 (2003), p.23.

할 수 있을 것임을 보이고 있다.

3. 백도형의 펜로우즈 읽기

백도형은 김광수를 비판하면서, 펜로우즈가 의식의 계산불가능성을 주장했다고 할지라도 그것이 정신의 자율성의 충분조건이 아니라고 비판한다. 그러나 김광수는 그의 논문에서 의식의 계산 불가능성이 정신의 자율성을 위한 충분조건이라고 주장하지 않는다. 오히려 김광수의 주장은 "의식이 계산 가능한 것이라면, ... 인간의 자율성은 보장되지 못할 것"[6] 이라고 말함으로써, 의식의 계산 불가능성은 인간의 자율성의 필요조건임을 시사하고 있다. 김광수가 주장하는 것은 인간의 자율성을 인정하기 위해서는 의식이 계산 불가능하다는 것을 받아들이지 않을 수 없다는 것이지, 의식의 계산불가능성이 인간의 자율성을 충분하게 확보해준다는 것은 아니다.[7]

계속해서 백도형은 펜로우즈가 거부하는 것은 심적 속성의 기능적 속성으로의 환원이지, 심적 속성의 물리적 속성으로의 환원은 아니라고 말한다.

> ... 펜로우즈가 비판하려는 것은 마음에 관한 계산주의(computationalism) 입장, 그 중에서도 강한 인공지능(Strong AI)을 옹호하는 입장이다. ... 펜로우즈는 <의식의 계산불가능성>을 옹호함으로써 모종의 비환원주의를 옹호하는 듯하다. 하지만 여기서 주의해야 할 점은 이 때의 '비환원'은 심신 비환원이 아니라 심리-기능(혹은 계산) 간의 비환원을 말하는 것이다.[8]

---

6) 김광수 (2000), p.140.

7) 물론 의식이 계산 불가능하다는 펜로우즈의 주장이 설득력있는가는 또한 별개의 문제이다. 필자는 4절에서 살펴보겠지만, 펜로우즈의 이 주장이 성공적이지 않다고 생각한다. 따라서 필자도 백도형처럼 펜로우즈의 논변에 근거해서 의식의 자율성을 확보하려는 시도도 성공적이지 않다고 생각한다.

8) 백도형 (2002), pp.23-24.

물론 심적 상태의 물리적 상태로의 환원과 심적 상태의 기능적 상태로의 환원은 구별되어야 한다. 그러나 펜로우즈는 심적 상태의 기능적 상태로의 환원만 거부하는가? 필자는 펜로우즈가 심적 상태의 물리적 상태로의 환원도 거부한다고 생각한다. 펜로우즈는 존재론적으로 플라톤적인 이데아계와 물리계, 그리고 정신계라는 세 개의 세계를 가정한다. 그리고 물리계는 영원한 진리의 세계인 이데아계의 어떤 부분으로부터 투사된 것으로 생각할 수 있고, 정신계는 물리계를 구성하는 뇌로부터 '창발하는(emerge)' 것이며 이데아계는 정신적 활동을 통해서 이해될 수 있다고 설명한다.[9] 물론 펜로우즈는 자신의 견해를 전통적인 창발론과 동일시하는 것을 거부하는 것같다. 그는 전통적인 창발론은 의식을 "충분한 복잡성이라는 특징과 행위의 복잡성(sophistication) 때문에만 발생하는 창발적 현상"으로 보지만, 자신은 여기에 "비생명체인 물체의 행태와 관련하여 우리가 익숙하게 알고 있는 것과는 근본적으로 구별되는 구체적이고 새로운 토대가 되는 물리과정을 요청한다"고 말한다.[10] 요컨대 펜로우즈에 따르면, 물리적 대상의 행태를 지배하는 자연법칙과는 구별되는 새로운 물리적 과정을 통해서 의식 현상은 설명되어야 한다. 즉 의식은 뉴런의 하위 단계의 구조에서 발생하는 양자역학적 현상으로부터 창발하는 것이다. 펜로우즈가 물리적인 상태로부터 의식이 창발한다고 주장하는 것은 분명히 심신 환원 가능성을 거부한다는 것을 보여준다.[11]

또한 백도형은 펜로우즈가 심신 환원을 거부하는 것은 아닐 것이라고 주장하면서 다음과 같이 말한다

---

9) Penrose (1994), p.414.

10) *Ibid.*, pp.216-217.

11) 심사위원 중의 한 분은 백도형이 말하는 "환원"이란 결국에는 "어떤 형식"으로든 의식 현상이 "설명"될 것이라는 펜로우즈의 입장을 그렇게 해석할 수 있을 것이라고 말하면서 환원주의에 대한 보다 자세한 설명이 필요하다고 지적해주셨다. 그러나 필자는 의식이 물리적으로 "설명"된다는 주장을 환원주의라고 해석할 수는 없다고 생각한다. 그렇게 되면 데이빗슨을 포함한 많은 비환원주의자들도 환원주의자로 간주되어야 하기 때문이다. 필자는 환원주의란, 심신 사이의 법칙적, 유형적 환원을 주장하는 보다 엄밀한 물리주의라고 생각한다.

> 오히려 펜로우즈는 **물리주의**를 옹호한다. 즉 두뇌의 적절한 물리적 활동에 의해 의식이 **실현**된다고 본다. 즉 의식의 존재 여부는 그것이 수행하는 물리적 활동이 무엇인지에 달려있다고 한다.[12]

아마 백도형이 여기서 '물리주의'란 단어로 의미하는 것은 환원주의적 물리주의일 것이다. 그러나 의식이 두뇌의 물리적 활동에 의해서 실현(realization)된다고 주장하는 것이 곧 환원주의를 주장하는 것이라고 해석되어서는 안 된다. 왜냐하면 데이빗슨과 같은 비환원주의자와 창발론자들, 그리고 써얼과 같은 철학자는 의식이 물리적으로 환원된다는 것은 거부하지만, 그들은 모두 의식이 물리적 현상으로부터 실현된다는 사실은 받아들이기 때문이다. 또한 백도형은 펜로우즈가 환원주의자라고 보는 다른 이유를 다음과 같이 제시한다:

> ... 펜로우즈는 ... 물리주의의 입장을 표방한다. 그것도 존재론적으로 물리적인 실현을 주장할 뿐만 아니라, 방법론적으로도 (비록 현재의 물리학이 아닌 새로운 물리학이긴 하지만) 물리학의 방법론에 의해 수행되어야 함을 인정하기까지 한다. 이렇게 본다면 오히려 이러한 펜로우즈의 입장은 심물 환원주의가 될 수도 있[다].[13]

백도형에 따르면, 펜로우즈는 존재론적으로 뿐만 아니라, 의식에 대한 설명은 (새로운) 물리학의 방법에 의존할 수밖에 없다고 주장한다는 점에서 그는 방법론적으로도 물리주의자라는 것이다. 그러나 백도형의 이러한 주장은 펜로우즈를 비환원주의라고 해석하는 사람들도 이견없이 받아들일 만한 것이다. 비환원주의적 물리주의는 방법론적 물리주의와 양립가능하기 때문이다. 다시 말해서 펜로우즈가 방법론적 물리주의자라는 주장은 그가 심물 환원주의자라는 해석을 뒷받침할 근거가 되지 못한다.

---

12) 백도형 (2002), pp.24-25. (강조는 필자).
13) *Ibid.*, p.25.

결국 백도형은 펜로우즈가 심적 상태의 기능적 환원은 거부하지만 물리적 상태로의 환원을 옹호하고 있다고 주장하고 있는 셈이다. 그런데 백도형도 인정하듯이, 펜로우즈는 심적 상태가 계산적, 기계적으로 실현될 수 없다고 주장한다. 만약 펜로우즈가 심적 상태의 물리적 환원을 주장한다면, 어떻게 그가 심적 상태의 기계적 실현을 거부할 있는지 의심스럽다. 그런 의미에서도 펜로우즈를 환원적 유물론자로 읽는 백도형의 해석은 옳지 않다.[14]

이제 어떤 근거로 펜로우즈를 심신 비환원주의자로 읽어야 하가에 대해서 대답하기 위해서 그의 논증을 자세히 살펴보자.

4. 펜로우즈 자세히 읽기

펜로우즈는 기계 기능주의와 인공지능 전반에 대해 반대하면서, 인간의 의식은 알고리즘적이지 않으며 따라서 튜링 기계와 같은 계산기계와의 유비를 통해서 설명할 수 없다고 주장한다. 그는 인간의 의식에 대한 네 가지 가능한 입장을 다음과 같이 제시한다.

A. 모든 사유는 계산이다. 특히 의식적 자각(conscious awareness)의 느낌은 적절한 계산을 수행함으로써만 만들어진다.
B. 자각은 뇌의 물리적 활동의 특징이다. 그 어떤 물리적 활동도 계산적으로 모의될 수 있지만, 계산적 모의는 그 자체로 의식을 만들어내지는 못한다.
C. 뇌의 적절한 물리적 활동은 자각을 낳지만, 이러한 물리적 활동은 계산적으로 적절하게 모의될 수 없다.

14) 익명의 심사위원 중의 한 분은 심적 상태의 물리적 환원을 받아들이면 심적 상태의 기계적 실현도 받아들여야 한다는 필자의 논지가 거짓이거나 근거가 없다고 비판하였다. 그러나 필자는 여전히 심적 상태가 법칙적으로 물리적 상태로 환원된다면, 원칙적으로(in principle) 심적 상태가 물리적으로 실현될 수 있을 것라고 생각한다.

D. 자각은 물리적이거나 계산적이거나 혹은 그밖의 과학적인 방식으로 설명될 수 없다.[15]

펜로우즈는 전형적인 계산 기능주의, 즉 강한 인공지능의 주장인 A와 신비주의적인 주장 D는 인간의 의식에 대한 올바른 설명일 수 없다고 말한다. 그리고 B는 약한 인공지능의 주장으로서 과학적인 상식과 가장 잘 부합되는 것같지만, 모든 물리적 활동이 다 컴퓨터로 모의될 수 있는 것이 아니기 때문에, C가 가장 진리에 가깝다고 주장한다.[16] 나아가서 그는 의식을 야기한다고 믿어지는 비계산적인 물리적 상관자의 위치를 물리학과 생물학에 의지하여 찾으려고 시도한다. 그렇다면 비알고리즘적인 의식의 현상은 어떻게, 어디에서 찾아지고 설명될 수 있는가? 그는 신경자극이란 양자효과에 의해서 제한되지 않는 거시단계의 대상이기 때문에 뉴런단계에서 그러한 의식 현상이 발견될 것이라고 기대할 수 없다고 말하고, 미시신경세포의 단계인 미세관(microtubule)이 마음의 비계산적 물리 현상의 위치라고 주장한다. 그의 복잡한 논증을 간단히 요약해 보자.[17]

A1) 인간의 마음은 건전하지만(sound) 비계산적이다. 즉 알고리즘적이지 않다.

A2) 인간은 이러한 사유내용을 의식한다.

A3) 현대 물리학을 통해서 우주에서 비알고리즘적인 과정이라고 인식되는 유일한 예는 일종의 임의성(randomness)이다.

A4) 인간의 마음이 비계산적일지라도 임의적인 것은 아니다.

A5) 그러므로 인간의 의식은 현재의 물리학의 이해를 넘어선다.

---

15) Penrose (1989), p.12.

16) 사소한 문제이지만, 김광수와 백도형은 펜로우즈가 강한 인공지능(strong AI) 논제만 비판하는 것처럼 말하지만, 사실 그는 약한 인공지능의 가능성에 대해서도 비판적임을 알 수 있다.

17) 아래의 요약은 R. Grush & P. Churchland "Gaps in Penrose's Toilings"에서 나온 것을 정리한 것이다. 이 논문은 Churchland (1998), pp.205-229에 재수록되어 있다.

B1) 현대 물리학 이론은 양자역학적 파동함수의 붕괴나 무질서를 설명할 수 없으므로, 양자 인력에 대한 부가적 이론이 필요하다.
B2) 파동함수 붕괴에 대한 보다 적절한 이론은 임의적이지 않은 비계산적인(비알고리즘적인) 과정을 설명할 수 있을 것이다.
B3) 준결정체(quasicrystals)의 존재는 아직 인식되지 않은 비알고리즘적인 물리과정의 증거이다.
B4) 그러므로 양자인력과 같은 미래의 물리학 이론은 비알고리즘적인 과정을 설명하리라고 기대될 수 있다.

C1) 뉴런의 미시단계인 미세관은 일종의 양자역학적 현상과 속성을 지닌다.
C2) 미세관의 비알고리즘적이지만 임의적이지 않은 과정은 인간의 의식을 설명하기에 적절할 것이다.
C3) 미세관은 뉴런의 기능에서 핵심적인 역할을 한다.
C4) 뉴런은 인간의 의식 기능에서 핵심적인 역할을 한다.
C5) (C3과 C4로부터) 미세관은 인간의 의식 기능에서 핵심적 역할을 한다.
C6) 그러므로 미세관은 양자역학과도 관계하고 또한 의식과도 관계하기 때문에 의식의 비알고리즘적인 성질을 설명할 수 있을 것이다.

요컨대 펜로우즈에 따르면, 계산불가능한 의식이 뇌 속의 물리적 과정에 의해서 야기된다면, 의식을 야기하는 뇌의 물리적 과정도 계산불가능해야 할 것이다. 그런데 뉴런 단계에서의 뇌의 물리적 과정은 물리적 법칙에 의해 지배되는 계산가능한 것이므로, 양자역학적으로 설명될 수밖에 없는 뉴런의 하위단계에서 의식의 뿌리는 찾아져야 한다는 것이 그의 주장의 요지이다. 그런 의미에서 그는 다음과 같이 말한다.

뉴런의 역할은 아마도 **확대경**과 유사할 것이다. 거기에서 보다 작은 작

> 동 규모의 세포골격체(cytoskeletal)의 행동은 근육과 같은 몸의 다른 기관에 영향을 주는 어떤 것으로 전이된다. 따라서 뇌와 마음에 대한 지금 유행하는 그림을 제공하는 뉴런단계의 기술은 세포골격체의 행동이라는 보다 깊은 단계의 **그림자**일 뿐이고, 우리가 마음의 물리적 토대를 찾아야 하는 곳은 바로 이보다 깊은 단계인 것이다![18]

이제 이 논증을 하나하나 평가해보자. 우선 첫 번째 논증, (논증A)에 대해서 살펴보자. 이 논증에서 논란이 되는 것은 첫 번째 전제인 A1)이다. 먼저 펜로우즈가 이 논증에 대해서 이야기하는 바를 인용해 보자.

> 그것은 일종의 귀류법(reductio ad absurdum) 논증인데, 그 논증에서 나는 우리가 가지고 있는 이해력을 지닌 로봇을 만들려고 한다면 어떤 일이 발생하게 되는가를 보여주고자 한다. 괴델 논증은 기본적으로 이해와 관련된 것이기 때문이다. ... 그것은 기호의 의미에 대한 질문과 관련된 것인데, 기호의 의미는 계산 체계가 가질 수 없는 차원의 것이다. 계산 체계는 자신이 따라야 할 규칙만을 가질 뿐이다. 기호의 의미를 이해함으로써 우리가 수학에서 할 수 있는 것은 그러한 형식적 규칙을 넘어서는 것이다.[19]

펜로우즈는 A1)에 대해서 괴델의 불완전성 정리의 튜링 버전이라고 할 수 있는 멈춤의 문제(halting problem)의 논증으로 증명한다. 이에 대하여는 이미 수많은 논의가 이루어졌기 때문에[20] 간단히 그 요점만 지적하기로 하자. 펜로우즈가 그 논증을 통해서 보이고자 하는 것은 알고리즘을 수행하는 어떤 컴퓨터도 우리는 알고 있는 사실, 즉 어떤 계산절차가 멈추지 않는다는 사실을 보여주지 못하기 때문에 컴퓨터가 지능에 있어서 인간과 동일한 것일 수 없다는 것이다. 즉 인간의 정신으로 도달할 수 있는 이해와 지능이 모두 기계적 과정으로 실현될 수는 없다는 것이 그

18) Penrose (1999), p.376.
19) *Ibid.*, p.19.
20) 손병홍, 송하석, 심철호 (2002), pp.13-15 참고.

의 논변의 핵심이다. 또 펜로우즈는 보다 엄격한 논변을 통해서 무모순적인 F라는 형식체계는 인간으로서 '나'와 동일할 수 없다는 것을 증명해 보인다.[21] 그러나 이 논변들에 대한 비판은 이미 많이 나와 있는데, 주된 비판은 "우리가 우리 자신이 건전함을 안다"는 모순을 낳은 전제에 의존하고 있다는 것이다.[22]

이제 두 번째 논증(B 부분)에 대해서 살펴보자. 펜로우즈는 논증 A가 옳기 때문에 현재의 물리학은 이론은 불완전하고, 양자역학 이론에 의해서 확장되어야 한다고 주장한다. 그러나 앞에서 살펴본 것처럼 논증 A는 여전히 논쟁적이고, 오히려 많은 철학자들에 의해서 비판되고 있다. 그런 의미에서 현대 물리학 이론의 완전성을 믿으면서, 논증 A를 거부할 수도 있을 것이다. 그러므로 펜로우즈는 임의적이 아닌 비알고리즘적인 물리적 과정의 존재에 대한 증거를 제시해야만 한다. 그는 그러한 증거로서 준결정체의 존재를 제시한다. 그것이 바로 두 번째 논증의 B3)이다. 그런데 그 전제를 주장하는 펜로우즈의 논거를 정리해 보면 다음과 같다.

B3-1) 평면에 오직 비주기적으로(nonperiodically) 붙여지는 일련의 타일이 존재하기 때문에 무한한 유클리드 평면에 주어진 타일을 붙일 수 있는가 라는 문제는 결정될 수 없다 (알고리즘적이지 않다).

B3-2) 이러한 비주기적인 타일들은 5중 대칭으로 그 평면에 붙여진다.

B3-3) 그것의 격자 구조가 유사한 5중 대칭을 보이는 준결정체(quasicrystals)가 있다.

B3-4) 이러한 결정체가 계속되어지는 것은 비알고리즘적인 과정에 의존한다.

---

21) Penrose (1999), pp.133-135.

22) 펜로우즈의 첫 번째 논증에 대한 평가는 앞의 손병홍, 송하석, 심철호의 글을 참조할 것. 그리고 펜로우즈의 이 논변과 관련한 많은 비판적인 논문이 있기 때문에 이 논변에 대한 자세한 비판은 생략한다. Putnam (1975), Benacerraf (1967), Chalmers (1995), K. Coleman "Gödel Propositions for the Mind" 등이 있고, 국내에서 발표된 것으로는 이병덕 (1999)이 있다.

B3) 그러므로 준결정체의 존재는 아직 인식되지 않은 비알고리즘적인 물리적 과정의 증거이다.[23]

그러나 이 B3)을 위해서 제시된 이 논증은 설득력이 없어 보인다. 첫째 준결정체의 구조와 비주기적으로 타일 붙이기 사이의 유비가 매우 밀접해 보이지만, 그 둘의 계산가능성 문제는 각각 다른 문제이다. 즉 타일 붙이기의 경우에 결정할 수 없는 특징은 일정한 지역에 타일을 붙이기 위해서 그 타일들을 어떻게 놓아야 하는가의 문제가 아니라, 이 타일들이 빈틈없이 그리고 겹치지 않고 무한한 유클리드 공간 전체를 덮을 수 있는가의 문제이다. 또한 그럴듯해 보이는 이 유비는 결정적인 결함을 갖는다. 준결정체는 무한한 유클리드 공간과 달리 실제로 유한하기 때문에 준결정체의 계속은 의심할 바 없이 계산가능하다. 실제로 펜로우즈 자신도 그러한 준결정체의 지속의 결과에 대해서 알고리즘이 주어질 수 있음을 지적한 바가 있다.[24] 즉 B3-4)는 옳지 않다. 이에 대해서 그러쉬와 처칠랜드는 "적절한 배열을 결정하기 위해서 필요한 정보가 그 격자 구조에 있는 개별 원자에 국지적으로 주어지지 않을 수 있을지라도 알고리즘에 관한 한, 그것은 문제가 되지 않는다"[25]는 점을 지적하면서 이 전제들이 옳지 않음을 강조한다. 결국 비알고리즘적인 과정이 있다는 펜로우즈의 주장은 준결정체가 비알고리즘적인 과정에 관한 확고한 증거일 수 없기 때문에 단순한 추측 이상일 수 없다.

끝으로 마지막 논증(C 부분)을 살펴보자. 해머로프-펜로우즈의 가설로 알려진 이 논증의 첫 번째 전제는 경험적 증거를 결여한 추측 이상이 아니다. 현대 생물학에 따르면, 세포 골격체는 세포질에 널리 퍼져있는 단백질 섬유의 복잡한 네트워크으로서 세포의 여러 가지 형태의 협조적이며, 목표지향적인 운동을 가능하게 하는 것이다. 세포 골격체는 뼈로 이

23) Grush & Churchland (1998), p.221.
24) Penrose (1989), p.449.
25) Grush & Churchland (1998), p.221.

루어진 골격 구조와 달리 세포가 형태를 변화시키고 분열하고 환경에 반응함에 따라 계속해서 재조직되는 매우 역동적인 구조이다. 세포 골격체를 구성하는 섬유 중에서 대표적인 것이 미세관인데, 미세관은 보통 한 쪽 끝은 중심체(centrosome)에 고착되어 있고 다른 끝은 세포질 속에 자유로운 상태로 된 팽팽한(stiff) 구조이다. 또 미세관은 그 미세관을 이루는 단백질인 튜불린(tubulin)의 가감에 따라 반복해서 성장과 수축을 하는 역동적인 구조이다. 이러한 사실로부터 해머로프는 세포 골격체의 미세관은 의식과 관련한 양자 정합성(quantum coherence)[26]의 훌륭한 후보자라고 추측한다. 그가 그렇게 추측하는 중요한 이유는 첫째 미세관의 개별적 하위단위인 튜불린의 구조는 소수성(hydrophobic)의 단백질 영역에서 전기적 운동이나 쌍극자(dipole) 현상[27]과 같은 양자단계의 사건과 짝을 이룰 수 있기 때문이고, 둘째 미세관의 격자구조, 대칭, 원통형 구조, 그리고 평행적인 배열은 장거리 협조와 질서를 증진시킬 수 있으며, 마지막으로 텅 빈 미세관의 내부는 양자역학적인 속성들인 도파관(wave-guide) 현상[28]이나, 초복사(super-radiance)[29] 현상, 그리고 자기유도적 투명성(self-induced transparency) 등을 가능하게 하는 것처럼 보이기 때문이다.[30]

---

26) 이는 '양자 결맞음'이라고도 번역되기도 한다. 미시적 수준의 대상에 부여되는 양자역학적 상태는 파동으로 표시될 수 있는데, 이 경우 인접한 입자들이 지니는 파동의 위상(phase)이 서로 높은 수준의 일치 관계에 있게 되면 예측하지 못한 흥미로운 현상이 발생한다. 이러한 현상을 통칭하여 양자 정합성이라고 부른다.

27) 두 극성분자가 가까이 접근할 때 그 쌍극자 사이의 정전기적 인력에 의해 일어나는 상호작용을 말한다.

28) 도파관이란 마이크로파 이상의 높은 주파수의 전기 에너지나 신호를 전송하기 위한 전송로의 일종으로 광섬유가 대표적인 것이다. 도파관의 내부는 보통 비어있고 공기로 채워져 있다. 여기서 해머로프와 펜로우즈가 도파관 현상이라고 부르는 것은 도파관 내에서 전자파 위상의 간섭으로 발생하는 양자 정합적 현상을 말한다.

29) 원자 내의 전자는 에너지가 일정하게 유지되는 준위(準位)에 있으며 이것이 다른 준위로 전이할 경우에 전자기파, 즉 광자가 방출되거나 흡수된다. 이 전이는 전하 또는 자기모멘트를 가지는 입자의 고전적인 진동에 따른 복사 등에 대응하여 생각되는 것으로서, 전기쌍극자에 의한 전기쌍극복사가 그 주된 것인데, 전기사중극복사 자기쌍극복사 등 일반적으로 다중극복사도 있다.

30) Hameroff (1994), p.105.

모든 세포체(cell body)는 미세관을 갖고, 미세관의 주요한 기능은 세포 분열을 지지하는 것이다. 뉴런의 가장 잘 알려진 기능은 세포체와 축삭돌기(axon), 세포체와 수상돌기(dendrite) 사이에 신경 전달물질(neurotransmitters)과 같은 분자를 비롯한 여러 가지 단백질을 수송하는 것이다. 일반적으로 마취제는 친수적인(hydrophile, water-binding) 것으로 뉴런막 수용기와 단백질의 통로(channels)를 변화시키는 것으로 믿어진다.[31] 해머로프는 이를 토대로 미세관을 구성하는 단백질인 튜불린은 자신의 친수적 성질에 영향을 받게 할 수 있다고 가정한다. 그의 가정이 옳다면, 그리고 튜불린 구조가 양자역학적 성질을 갖는 것이 사실이라면, 마취되었을 때 의식을 상실하는 현상을 미세관에서의 이러한 변화 때문일 수 있을 것이다.

그러나 해머로프 자신이 인정하고 있는 것처럼, 초복사와 같은 양자 정합성이 미세관에서 발생한다는 주장에 대한 아무런 경험적 증거도 없고, 그것은 다만 그럴 가능성이 있다는 추측일 뿐이다. 또한 마취의 결과 의식을 상실하는 것이 뉴런의 미세관의 변화 때문이라는 직접적인 증거도 없다. 오히려 최근 밝혀진 바에 따르면, 신경막의 여러 가지 단백질이 마취효과를 발생시키는 주된 장소인 것처럼 보인다. 그리고 만약 해머로프의 가설이 옳다면 미세관의 기능의 붕괴와 같은 현상은 수면 상태에서 깨어있는 상태로의 변화와 같은, 우리가 일상적으로 경험하는 변화의 토대가 될 것이다. 그러나 미세관이 이러한 상태 변화에 상응하는 기능을 변화시킨다고 믿을 아무런 증거도 없다. 우리에게 알려진 것은 오직 수면상태에서 깨어있는 상태로의 변화와 관련하여 어떤 뉴런, 특히 피질과 뇌간 구조에서 매우 구체적인 방식의 변화가 발생한다는 사실뿐이다. 그리고 현대 신경생리학적 연구에 의하면, 미세관을 구성하는 튜불린 튜브의 기공에 해머로프가 가정하는 것처럼 순수한 물만 들어 있다고 생각할 수 없다. 왜냐하면 그 곳에는 칼슘이나 나트륨과 같은 일상적인 세포질 이온을 들어오지 못하게 한다고 알려진 어떤 메카니즘도 없기 때문이다.

---

31) Churchland (1988), ch2 참조., 특히 pp.41-49를 참고할 것.

만약 튜불린 기공에 순수한 물만 들어 있는 것이 아니라면 해머로프가 기대한 초복사와 같은 장거리의 협동이 일어날 수 없을 것이기 때문에, 이것은 해머로프의 가정에 중요한 문제가 될 것이다.

논증 C의 첫 번째 전제인 해머로프-펜로우즈의 가정에 대해서 관대하게 받아들인다고 할지라도 여전히 이 논증은 문제점을 가지고 있다. 분명히 단 하나의 뉴런에서의 변화가 의식을 만들어 내거나 상실하게 하지는 않을 것이다. 그러므로 이제 펜로우즈와 해머로프는 다른 뉴런에 있는 미세관 덩이가 어떻게 의식을 야기하는 전체적인 신경활동의 패턴을 형성할 수 있도록 자신의 행위를 조정할 수 있는가에 대답하여 한다. 또한 공간적으로 분리된 미세관들이 사이에 양자 정합적인 상태를 예측하는 것이 어떻게 실재적일 수 있는가? 이에 대하여 펜로우즈는 과학적, 실증적인 대답을 하는 대신, "자연이 생물학적 수단에 의해서 이룩한 그러한 공적은 놀라운 것, 거의 믿을 수 없는 것이다"[32]라고 말함으로써 여전히 신비에 맡겨두고 있는 것같다.

끝으로 이 논증의 추론에서 C5)는 앞의 두 전제 C3)와 C4)로부터 얻어진다. 즉 이 추론은 인지와 의식에서 뉴런이 핵심적 역할을 하고 미세관이 뉴런의 기능에서 핵심적인 역할을 하기 때문에 미세관이 인지와 의식에서 핵심적 역할을 한다는 것이다. 그러나 이 추론은 타당하지 않다. 분명히 '핵심적 역할을 한다'라는 술어는 이행적(transitive)이지 않은 경우가 많다. 특히 미시세계에 속하는 대상들이 갖는 기능은 거시세계의 대상들이 갖는 기능으로 이행되지 않는다는 것은 상식이다. 펜로우즈가 설명하듯이, 미세관의 기능이 양자역학적 설명으로 파악될 수 있는 미시세계의 것이라면 이와 같은 이행논증은 타당하지 않음에 분명하다. 즉 인간의 의식 기능에서 뉴런이 핵심적인 기능을 하고, 미세관이 뉴런의 기능에서 핵심적인 역할을 한다고 할지라도 미세관이 인간의 의식 기능에서 핵심적인 역할을 하지 않을 수도 있다는 것이다.

---

32) Penrose (1989), p.373.

펜로우즈의 논증에 대한 비판을 요약해 보자. (논증 A)에서 전제 A1)은 잘못된 주장이어서 논증 A는 받아들일 수 있는 논증이 아니고, (논증 B)는 그 전제 B3)이 옳지 않기 때문에 그 논증은 의심스러운 추측에 불과하며, 그리고 (논증 C)는 전제 C1)은 증거가 없을 뿐만 아니라 오히려 현대 생물학에서 받아들여지지 않을 것같은 가정일 뿐이며, C5)는 부당한 논증의 결론이기 때문에 이 논증도 설득력이 없다.

5. 맺음말

펜로우즈는 의식이 뉴런의 하위 단계인 미세관의 양자역학적 속성에 의해서 창발된다는 주장을 분명하게 하고 있기 때문에 그를 환원주의적 물리주의자로 해석하는 것은 근거가 없어 보인다. 그러므로 백도형이 주장하는 것처럼, 펜로우즈를 심적 상태의 기능적, 계산적 상태로의 환원만을 거부하고, 심신 환원에 대해서는 긍정적인 입장을 취한다고 해석하는 것은 옳지 않다.

필자도 개인적으로 인간의 의식이 물리적으로 환원될 수 없다고 생각한다. 그러나 펜로우즈가 제시하는 논증은 성공적이지 않기 때문에 그의 논증을 비환원적 물리주의의 옹호 근거로 삼는다든지, 그를 통해서 의식의 뿌리가 밝혀질 것이라고 기대해서는 안 된다고 생각한다. 따라서 펜로우즈의 논증이 인간의 자율성을 확보해 줄 수 있는 비환원적 물리주의의 가능성을 보여주는 것이라는 김광수의 기대 또한 옳지 않다. 결국 인간의 자율성을 유지하면서 심적 현상을 물리적으로 설명하려고 시도하는 비환원주의자들은 자신의 논거를 정당화하기 위해서 펜로우즈가 아닌 다른 논거를 찾아 나설 수밖에 없다.

참 고 문 헌

김광수 (2000), "유물론과 자유" 한국 분석철학회 편 「21세기와 분석철학」, 철학과 현실사, pp.118-144.

_____(2003), "심리철학과 정신의 자율성", 「철학적 분석」 7호, pp.1-27.

백도형 (2002), "참을 수 없는 존재론의 가벼움", 「철학적 분석」 6호, pp.1-27.

손병홍, 송하석, 심철호 (2002), "인공지능과 의식: 강한 인공지능의 존재론적 및 의미론적 문제", 「철학적 분석」 5호, pp.1-33.

이병덕 (1999), "Penrose and his Gödelian Argument" *Philosophy and Culture*: 191-204. 1.

Benacerraf, P. (1967), "God, the Devil, and Gödel" *The Monis, 51*: 9-32.

Chalmers, D. (1995), "Minds, Machines, and Mathematics", *Psyche 2, No. 9*, http://psyche.cs.monash.edu.au/v2/psyche-2-09-chalmers.html.

Churchland, P. & P. (1998), *On the Contrary,* Cambridge: MIT press.

Churchland, Patricia (1988), *Neurophilosophy: Toward a Unified Science of the Mind/Brain,* Cambridge: MIT press.

Grush, R. & P. Churchland "Gaps in Penrose's Toilings" In Churchland, P. & P. (1998), pp.205-229.

Hameroff, S. R. (1994), "Quantum Coherence in Microtubules: A Neural Basis for Emergent Consciousness?" *Journal of Consciousness Studies 1*: 99-112.

Penrose, R. (1989), *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics,* New York: Oxford University Press.

_____ (1994), *Shadows of the Mind* ,Cambridge: Oxford University Press.

Putnam, H. (1975), "Minds and Machines", *Mind, Language, and Reality*, New York: Cambridge University Press, pp.362-385.

Reading Penrose: Is he a reductionist or non-reductionist?

Ha-Suk Song

The successful non-reductive materialism will be very attractive because it can account for mental states only by physical terms, saving mental autonomy. Recently, R. Penrose seems to provide such a version of materialism. He tries to explain consciousness by appealing to contemporary neuro-science. According to the science, microtubules play a crucial role in neuronal fuctioning, and neurons play a key part in consciousness. From those facts, Penrose infers that microtubules play a key role in consciousness. In addition, Hameroff suggests that microtubules have the properties that make certain quantum-mechanical phenomena possible, which is called Hameroff's hypothesis. Accepting this hypothesis, Penrose argues that consciousness will be realized in the microtubules in the quantum-mechanical way.

Dr. Kwangsu Kim suggests that Penrose's theory should be a successful version of non-reductionism, so that it can reveal the mysterious veil of consciousness. That is, Dr. Kim expects Penrose's theory to explain mental state physically and to save mental autonomy. On the contrary, Dr. Tohyung Paik criticizes Dr. Kim's interpretation of Penrose, and claims that Penrose should be read as a reductive materialist.

I will claim that Penrose is a non-reductive materialist, and provide some ground for it. In fact, he argues that consciousness is emergent form quantum-mechanical properties of the microtubules. But I conclude that his arguments for non-reductive materialism have several problems, so that they are not successful. Therefore, I think, Dr. Paik is wrong in that he interprets Penrose

as a reductionist; Dr. Kim's hope that consciousness will be successfully explained by Penrose's arguments is not plausible. So we should search for another version of non-reductive materialism in order to explain the consciousness physically, saving metal autonomy.

[Subject] Philosophy of Mind, Philosophy of Science

[Key Words] Consciousness, Reductionism, Non-reductive Materialism, Microtubule, Quantum Coherence, R. Penrose, Kwangsu Kim, Tohyung Paik