

결정론의 환상과 기능류어로서의 자유의지

모 기 룡[†]

연세대학교 인지과학협동과정 박사과정

결정론과 양립가능론은 그동안 자유의지론이 과학적인 견해와 맞지 않다고 생각해왔다. 이 논문은 친-과학적인 입장을 견지하면서 자유의지가 존재한다는 것이 아무런 문제나 모순을 발생시키지 않으며, 오히려 결정론이 과학적, 논증적 근거가 부족하다는 것을 논증한다. 먼저, 자아와 물리적인 신체(두뇌)의 동일론을 가정함으로써 이원론적인 자아를 거부한다. 자유의지는 다수실현이 가능한 기능적 개념이자 속성이다. 그래서 외계인이나 로봇의 행동 관찰을 통해서도 그들의 자유의지 유무를 판별할 수 있다. '자유의지'는 '소화' 처럼 기능류어이다. 우리가 소화기능을 가진다는 것을 아는 것처럼, 인간-자유의지의 미시적 구조를 파악해서 존재의 필연성을 알기 전이라도 우리에게 자유의지가 있다는 것은 알 수 있다. 어떤 개체가 자유의지 기능을 가지는가의 판단은 그것이 특정한 행동들을 가지는가에 달려있다. 그러한 행동은 상황에 맞게 행동을 조절하고 절제하며, 미래를 예측하고, 정교한 조건문을 만드는 등의 행동이다. 자유의지의 실체는 그러한 행동을 만들어내는 기능의 물리적인 담당자이다. 결정론은 갖가지 오해와 착각에 기초하고 있다. 결정론자들은 종종 물리법칙의 필연성과 결정론을 혼동한다. 또 다른 큰 착각은 자유의지가 한 시점에서 동일한 조건과 서로 다른 결과를 가진다고 생각한 것이다. 그러나 자유의지는 그 두 시점이 다르다는 것을 가정한다. 결정론은 종교의 영향을 받은 불필요한 추가적 가설이므로 오컴의 면도날에 따라 제거되어야 한다.

주제어 : 자유의지, 결정론, 양립가능론, 기능주의, 기능류어, 책임

[†] 교신저자: 모기룡, 연세대학교 인지과학협동과정, 연구세부분야: 심리철학

Tel: 02-2123-3240, E-mail: keoky@naver.com

도 입

진호는 다가오는 대학 축제 때 핫도그를 만들어 팔기로 했다. 그런데 핫도그 가격을 1500원으로 할지 1700원으로 할지 2000원으로 할지 고민이 되었다. 어떤 가격으로 정해도 법적, 도덕적인 문제는 발생하지 않는다. 진호는 핫도그의 가격을 정하는 것이 순전히 자신의 자유의지에 달려 있다고 생각한다. 왜냐하면 그는 본인의 마음대로 1500원으로 할 수도 있고 2000원으로 할 수도 있기 때문이다. 선택을 잘못하면 나중에 '다른 가격으로 팔걸 그랬다.' 라는 후회를 할 수도 있을 것이다.

그런데 결정론(determinism)에 따르면 그러한 생각은 일종의 착각이다. 진호는 가격을 1500원으로 정할 수도 있지만 다른 가격으로 할 수도 있다고 생각한다. 그러나 결정론에 따르면 그는 **다른 선택을 할 수 없다**. 여러 가지 선택이 가능하다고 믿는 것은 실은 착각일 뿐이고 그가 할 수 있는 것은 오직 한 가지밖에 없다. 이것은 결정론이 실제로 주장하는 것이며, 그래서 결정론은 상식과 어긋나게 된다.

그런데 결정론은 '하나의 미래만이 실현된다.'라는 주장을 하는 이론이 아니다. 그러한 생각은 모두가 알고 있는 상식이다. 즉, 핫도그를 1500원에 팔면서 **동시에** 2000원에 팔수는 없다. 핫도그의 가격은 그 때가 되면 결국 하나로 정해질 것이다. 주사위 던지기에서 우리는 여섯 개의 면이 모두 나올 수 있다고 생각한다. 하지만 주사위를 던지면 그 중 하나의 면만 나오게 될 것이다. 미래가 결국 하나의 모습으로 결정된다는 것은 어린아이도 상식적으로 아는 일이다. 다만 우리는 미래가 '무엇으로' 결정될 것인지에 대해 우리의 선택이나 행동에 의해 다르게 될 수 있다고 생각한다. 그러나 결정론은 '무엇으로' 결정되는 것이 우리의 선택에 의해 결정되는 것이 아니라 그전부터 이미 결정되어 있거나, 그게 아니라면 단지 우연으로 결정되는 것이라고 주장한다. 그래서 결정론에 따르면, 진호는 핫도그의 가격을 스스로 여러 대안들 중에 선택할 수 없다. 또한 결정론은 '우리는 안드로메다운하로 갈 수 있는 자유가 없다' 와 같은 모두가 아는 또 다른 상식적 주장을 하는 것도 아니다. 결정론은 우리가 일상생활에서 자유의지로 선택할 수 있다고 생각하는 사소한 것들마저 모두 부정하는 것이다.

자유의지가 없다는 주장에 따르면 이렇게 비상식적인 상황이 발생하게 된다. 이

것이 너무나 비상식적이었는지, 결정론을 인정하는 사람들 중에 상당수는 이러한 비상식이 일어나지 않을 수 있는 교묘한 방법을 채택한다. 그것은 ‘약한 결정론(soft determinism)’ 또는 ‘양립가능론(compatibilism)’이다. 이 이론은 자유의지의 논의에서 최근에 많은 지지를 얻고 있는 주장이다. 하지만 나는 이 논문에서 결정론을 비판함과 동시에 양립가능론도 비판할 것이다. 왜냐하면 우리는 결정론을 받아들여야 할 필요가 전혀 없기 때문이다. 자유의지가 없다는 주장은 잘못된 전제와 오해로 인해 생겨난 억측이다. 본문에서 자세히 다루겠지만, 그들은 자연 법칙에 대한 잘못된 해석을 하고 있으며, (모두가 아는)상식적 결정론으로 자유의지를 제거하려는 잘못된 시도를 하고 있다.

많은 사람들은 결정론이나 양립가능론이 친-과학적인 입장에 가깝고, 자유의지론은 그렇지 않다고 생각한다. 나는 이것을 반박하기 위해, 가장 친-과학적인 입장을 취하더라도 자유의지론(libertarianism)을 믿는데 아무 문제가 없다는 것을 밝힐 것이다. 그리고 결정론에 어떠한 문제점이 있는지를 밝혀낼 것이다. 우리에게 자유의지가 있다는 것은 상식이다. 어떠한 상식을 깨뜨리기 위해서는 합당한 근거가 있어야 한다. 만약 합당한 근거가 없다면 상식을 받아들여야 한다.

자아에 대한 친-과학적 관점

나는 결정론을 반박하기 위해 철저하게 친-과학적인 입장을 취할 것이다. 그리고 친-과학적인 입장은 실제로 나의 입장에 가깝다.

자유의지는 자아(self)와 깊은 관련을 맺고 있다. 우리는 상식적으로 ‘자아’ 또는 ‘자신’이 주도적으로 판단해서 선택하는 것을 자유의지라고 생각한다. 그것이 아니라면 자유의지에 의한 것이라고 하기 어렵다. 예를 들어 타인에 의해 강제로 어떤 행동을 하거나, 최면에 걸린 상태에서 나도 모르게 행한 일은 자유의지에 의한 것이 아니다. 다른 사람이 시켜서 어쩔 수 없이 한 일이라면 그것은 ‘자아에 의해서’ 한 것이 아니고, 자유의지에 의한 것이 아니다. 그래서 자유의지를 설명하기 위해서는 자아가 무엇인지를 설명해야 한다. 또한 자유의지는 책임의 문제와도 직결이 되는데, 자아에 의한 것이 아니라면 책임의 귀속은 어렵게 된다. 자유의지와 책임

이 발생하기 위해서는 ‘진정한 나’, 즉, ‘자아’에 의해 발생한 일이어야 한다.

‘자아’는 어디에 있을까? 자유의지론의 고전적 입장인 데카르트는 이원론을 주장하면서, 물리세계와 다른 정신의 세계가 있으며, 자아는 정신의 영역에 속한다고 보았다. 자유의지론은 ‘자아’의 역할을 강조하는 이론이고, 데카르트식 이원론은 자아의 존재를 설명하는 가장 손쉽고 고전적인 방법이다. 최근까지도 자유의지에 대한 완벽한 설명을 하지 못한 자유의지론자들이 데카르트식 이원론자로 의심되어 왔다. 자유의지론을 주장한 로더릭 치즘(Roderick Chisholm)은 모든 사건은 다른 사건을 원인으로 가진다는 사건인과론의 규칙을 깨고, 어떤 사건은 ‘행위자(agent)’ 자체가 원인이 될 수 있다고 주장했다. 그에 따르면 어떤 뇌 상태(사건)의 원인은 어떤 다른 물리적 상태가 아닌 ‘행위자’이다[1]. 하지만 이러한 주장은 이원론적인 자아를 주장하기에 충분하므로, 그에 비판적인 많은 학자들은 치즘이 주장한 ‘행위자’가 데카르트식의 정신적 실체를 가정하는 것이 아닌가 하는 의심을 가지게 되었다.

현대에 데카르트식의 실체 이원론을 받아들이는 학자는 거의 없다. 현대의 주류가 정신적인 실체를 따로 두지 않는 일원론이 된 까닭은, 아마도 ‘물리계의 인과적 폐쇄성 원리’ 때문일 것이다. 그 원리에 따르면, 물리계에 일어나는 모든 사건의 원인은 물리계 안에서만 갖는다. 이것을 거부한다면 물리학의 근본 자체가 거부되는 것이나 마찬가지이다. 그것이 과학의 일반적인 입장이자 전제이므로, 과학적 관점에서 이원론은 부정되며, 물리적이지 않은 것은 어떤 사건의 원인이 될 수 없다.

친-과학적인 학자들은 물리적이지 않은 ‘자아’의 존재를 부정하게 될 것이고, 모든 물리적 행동은 정신이 아닌 물리적인 것에 의해 일어나며, 따라서 정신이 원인으로 작용한다는 생각을 부정할 것이다. 그러면 자유의지가 부정되는 것일까? 그렇지 않다.

나는 물리계의 인과적 폐쇄성을 받아들이기 때문에, 이원론을 부정하며, 정신적 실체의 존재를 가정하지 않는다. 모든 물리적 사건들은 다른 물리적 사건의 작용으로 발생하게 된다. 나는 자유의지를 실현시키는 ‘자아’가 정신적 실체라고 보지 않고, 완전히 물리적 실체라고 본다. 자유의지를 가지는 행위자는 정신적인 실체, 즉, 데카르트식의 본체적 자아(noumenal self)가 아니어도 된다. ‘행위자’는 물리적이

며 물질적인 몸과 뇌를 지칭할 수 있다. 이것은 학문적으로 일원론적 관점일 뿐 아니라, 상식적으로도 이상한 생각이 아니다. ‘자아와 ‘신체(뇌)’는 위치적으로 동일하며, 나의 팔이 다치게 되면, ‘내가(나의 일부가) 다쳤다’라고 말하게 된다. 그 사람의 신체를 그 사람으로 여기는 것은 상식에 맞는 이야기이다.

나는 최대한 친-과학적인 입장을 가정하기 위해, 자아와 신체 간에 동일론의 입장을 취하겠다. 그리고 의식(consciousness)이 물리적인 것으로 완전히 환원되는 것을 가정하겠다.¹⁾ 이 입장은 실체이원론 뿐만 아니라 의식의 속성이원론까지 배제하는 것이므로 가장 친-(자연)과학적 입장이라 할 수 있다. 그래도 자유의지가 없다는 것으로 나아가지 않을 수 있다.

신체를 움직이는 원인이 되는 영역이 뇌라고 생각되기 때문에, 이제부터는 자유의지를 발생시키는 물리적 영역을 뇌로 한정해 보겠다. 과연 뇌는 자유의지를 만들어낼 능력이 없을까? 그러한 증거는 아직 발견되지 않았다. 우리의 뇌는 우리가 아는 단순한 기구들처럼 단순하지 않다.

자유의지에 반대하는 몇몇 학자들은 정신적인 의식 같은 것이 물질적인 것을 움직이는 것이 불가능하므로 자유의지가 없다는 결론에 이르게 되는 것 같다. 그들은 자유의지론자들이 원인이 없는 원인(uncaused cause)을 가정한다고 비판할 것이다. 그러나 자아가 물리적인 것(뇌)과 동일하다면, 모든 행동의 물리적인 원인을 찾을 수 있을 것이다. 우리 행동의 원인은 뇌와 뇌 상태에 있다. 그리고 뇌 상태는 의식과 동일하다. 그렇게 되면 의식이 행동을 일으키는 것도 설명할 수 있다. 심지어, 의식보다 앞서서 뇌 상태가 일어나도 상관없다. 그러한 뇌가 바로 자아이기 때문이다. 이에 관한 논란의 유명한 예로 벤자민 리벳(Benjamin Libet)의 실험이 있는데, 그에 관해서는 5절에서 자세하게 다루겠다.

내가 심신동일론의 입장을 취함으로써 논의하고자 하는 것은, 친-과학적인 심신동일론의 전제하에 자유의지가 존재할 수 있음을 논증하기 위함이다. 그리고 고전적인 자유의지개념인 이원론적 자아로서의 자유의지를 논의에서 배제하기 위함이다. 앞에서 보았듯, 결정론을 주장하는 사람들은 종종 자유의지론을 이원론으로 의심한다. 어떤 이들은 이원론에 의존해서 자유의지론을 주장할 수도 있을 것이다.

1) 실체일원론을 받아들이지만 의식에 물리로 환원되지 않는 속성이 있다고 주장하는 속성이원론도 존재한다.

그러나 이 논문은 이원론에 의존하지 않고 가장 친-과학적인 입장에 서면서도 자유의지의 존재에 대해 충분한 설명이 가능하다는 것을 보이고자 한다.

기능류어로서의 자유의지

‘자유의지’에 관해 우리가 가지는 상식을 이야기를 해보자. 대중에게 ‘자유의지’라는 용어가 생소하다고 해도, 진호의 경우처럼 자신에게 다양한 선택지가 있고, 그것들을 자신이 취하는 것이 모두 가능하다고 생각한다면, 그는 ‘자유의지’가 스스로에게 있다고 믿는 것과 같다. 그러한 ‘실현될 수 있는 다양한 가능성을 인지하고 그것을 선택하는 능력’을 자유의지라고 했을 때, 우리는 자신 뿐 아니라 (정상적인) 타인들도 자유의지를 가지고 있다고 상식적으로 믿는다. 반면에 결정론은 그것이 실제로 불가능하기 때문에 ‘환상’이라고 여긴다.

주목해야 할 또 다른 상식은 다른 동물들, 식물들, 무생물은 자유의지가 없는 것으로 보인다는 것이다. 돌멩이나 나무 같은 것은 말할 것도 없고, 다른 동물들은 고차원적 사고에 의한 것이 아니라 단순히 본능에 의해 행동하기 때문에 자유의지가 없다고 여겨진다.²⁾

그러나 사실 ‘자유의지’의 개념에 ‘인간만이 가지는 것’이 함의되어 있는 것이 아니다. 그것은 우리가 ‘자유의지’의 존재 유무를 판단하는 상식적 경향을 생각해 보면 알 수 있다. 이제부터 ‘자유의지’의 개념(또는 속성)에 대한 분석을 해보자.

우리가 관찰한 바에 따르면 인간만이 자유의지를 가지는 것 같지만, 만화나 영화를 보면 자유의지를 가진 인간이 아닌 것들이 종종 등장한다. 인간처럼 말을 하고 생각도 하는 동물들은 만화의 단골 소재이며, 인공지능을 가진 로봇이나 외계인들은 인간처럼 말을 하고 생각을 하며 감정표현을 한다. 그러한 것들은 인간은 아니지만 자유의지를 갖고 있는 것처럼 보인다. 따라서 자유의지의 개념에는 인간

2) 이와 관련해서 양립가능론자인 해리 프랭크퍼트(Harry Frankfurt)는 단순히 어떤 것을 원하는 욕구를 1차적 욕구(first-order desires), 1차적 욕구를 선택하거나 조절하기를 원하는 욕구를 2차적 욕구(second-order desires)로 정의하고, 2차적 욕구를 갖는 것은 (다른 동물과 비교해서)인간만이 가지는 특징이며, 그것이 자유의지를 가지는 것이라고 주장하였다[2].

만이 가진다는 것이 함의되어 있지 않다. ‘자유의지’는 그 ‘대상’이 무엇인지로 정의되는 것이 아니라, 그 대상이 가지는 ‘기능’으로 인해 정의되는 것이다.

이로써 나는 자유의지를 부정하는 사람들이 제기하는 주장 중에 하나를 논박하려고 한다. 어떤 이들은 “인간의 미시적인 뇌 구조를 관찰했을 때 아직까지 자유의지가 실현된다는 증거가 발견되지 않았다. 그러므로 자유의지는 없을 확률이 높다.”라고 주장할 것이다. 이에 대해, 나는 우리에게 자유의지가 있는지는 우리의 미시적인 뇌구조를 조사해야만 알 수 있는 것이 아니라는 것을 보이고자 한다.

영화 <가위손>의 주인공 에드워드는 기계로 만든 인조인간이다. 그의 뇌구조와 재질은 인간과 다를 것이다. 영화이긴 하지만, 에드워드의 행동이나 말을 관찰했을 때 자유의지를 가지고 있다는 생각이 들 수 있다. 그는 약간 어리숙하긴 하지만 상당히 똑똑하고 인간처럼 말하고 심지어 사랑과 같은 감정도 가지고 있다. 많은 사람들은 이 영화를 보고 에드워드가 자유의지를 가지고 있다고 생각할 것이지만, 몇몇 사람들은 그것에 반대할지도 모르겠다.

그렇다면 에드워드가 자유의지를 가지고 있는지 아닌지를 어떻게 알 수 있을까? 올바른 방법은 그의 행동과 언어를 좀 더 면밀히 관찰해 보고 행동적인 테스트를 해보는 것이다. 거기에는 어떠한 테스트의 기준이 필요할 것이다. 만약 에드워드가 그 테스트 기준을 통과한다면, 그는 자유의지가 있다는 결론에 이르게 된다. 그가 자유의지가 있는지를 확인하기 위해서 그의 몸을 열어서 내부의 미시적 구조를 살펴봐야 하는가? 그럴 필요가 없다. 인간의 경우도 마찬가지이다. 우리는 어떤 개체가 자유의지를 가지고 있는지를 알기 위해 그의 행동과 말을 보고 판단한다. 그리고 그 기준은 옳은 기준이다. 자유의지의 유무를 판별하는데 그 개체 내부의 미시적 구조를 살펴볼 필요가 없다. 왜냐하면 ‘자유의지’는 그것을 실현시키는 내부의 미시적 구조를 **직접 지시하지 않기** 때문이다.

‘자유의지’는 어떤 자연적 대상을 가리키는 용어가 아니다. 대상의 행동을 관찰했을 때 알 수 있는 특정한 공통적 기능적 속성을 가리키는 용어이다. 그래서 ‘자유의지’는 ‘기능류어’(functional kinds term)이다. 기능주의는 다수실현(multiple realization)을 가지는 특징이 있다. 그것은 대상의 재질이나 종류가 무엇인지에 관계 없이, 미시적으로 서로 다른 구조를 가진다고 해도 행동을 관찰해서 입력과 출력의 사이에 있는 기능의 담당자(혹은 인과적 역할의 담당자)가 공통된 기능을 하는

경우에 그 존재를 인정하게 된다. '자유의지'가 기능류어인 것은, 인간중이라도 금치산자나 정신박약, 혹은 좀비로 변환 사람은 자유의지가 없다고 여기고, 지능이 발달한 외계인을 상상한다면 그도 인간처럼 자유의지가 있다고 여겨진다는 사실로써 알 수 있다.

기능류어의 다른 예로는 '소화' 와 '시각'을 들 수 있다. '소화', '시각', '자유의지'에는 그 뒤에 '기능'이라는 단어를 붙여도 본질적인 개념에는 차이가 없게 된다. 인간은 소화기능과 시각기능을 가지고 있다. 하지만 인간만 가지고 있는 것은 아니다. 다른 많은 동물들과 곤충들도 소화기능과 시각기능을 가지고 있다. 그런데 그것을 실현하는 미시적 구조들은 서로 다르다. 인간 소화기관의 구조와 소, 지렁이의 소화기관의 구조는 다르고, 인간의 시각구조와 부엉이, 파리의 시각구조는 다르다. 하지만 인간뿐 만 아니라 다른 동물들도 '소화' 와 '시각'을 가지고 있다. 왜냐하면 공통적인 기능적 속성이 있기 때문이다.

기능류어는 특정한 미시적 구조를 지칭하지 않으므로 그것을 가지는지 판단하는 기준은 미시적 조사를 통할 필요가 없다. 우리는 오래전부터 우리가 소화기능이 있고 시각기능이 있다는 것을 알고 있었다. 반면에 인간의 소화기관과 시각기관의 미시적 구조가 알려진 것은 그리 오래되지 않았다. 인간의 소화기관과 시각기관의 미시적 구조를 깨닫기 전까지는 인간이 소화기능과 시각기능을 가진다고 말하면 안되는 것인가? 그것은 명백히 잘못된 생각이다. 인간의 미시적 소화구조를 아는 것은 우리에게 소화기능이 있다는 것을 알기 위해 필요한 것이 아니다. 이와 마찬가지로, 만약 인간의 미시적 뇌구조를 조사해서 인간에게 자유의지가 있다는 것을 깨달았다면, 그것은 인간-미시구조적 자유의지를 이해한 것이다. 그것은 우리에게 자유의지가 있다는 것의 충분조건이 되지만, 필요조건은 아니다. 그러므로 우리에게 자유의지가 있는지를 확인하기 위해서 반드시 인간-미시구조적 자유의지를 조사해야할 필요는 없다.

다수실현을 가지는 기능주의는 미시적 구조를 지시하는 것이 아니라 행동관찰이 중요한 판단요소가 된다. 그래서 현대 기능주의의 창시자라 할 수 있는 앨런 튜링(Alan Turing)은 어떤 개체가 특정한 기능을 가지는가를 알아보기 위해 행동적 테스트를 해보면 된다고 주장하였다. 튜링은 '기계가 인간처럼 생각할 수 있는가?'를 알아보기 위해 튜링테스트를 제안하는데[3], 그 테스트는 두 방에 컴퓨터와 사

람이 각각 들어가고, 그 방 안을 볼 수 없는 질문자(interrogator)는 질문을 통해 누가 컴퓨터이고 누가 사람인지를 판단하도록 한다. 튜링은 질문자가 키보드(타이핑)를 통해 그들과 수많은 대화를 나눈 뒤에, 컴퓨터와 인간을 구분해내지 못한다면, 그 컴퓨터는 인간처럼 생각하는 것으로 봐야한다고 주장했다.

튜링 머신식 기능주의는 힐러리 퍼트넘(Hilary Putnam)이 계승하는데, 퍼트넘의 기능주의는 '기능적 속성'이라는, 물리적 속성이 아닌 '속성이원론'의 가능성을 제시했다[4]. 반면에 데이비드 루이스(David Lewis)는 기능주의가 물리주의의 강한 환원 또는 동일론이 가능하다는 것을 주장했다[5]. 또한 루이스는 우리의 상식적 심리이론 또는 통속 심리이론이 실제적 가치가 있다고 주장했는데, 그 이유는 우리의 상식적 심리이론이 물리적으로 환원되지 않는다면, 우리의 심적 용어들은 전부 다 지시하는 것이 없게 되어버릴 것(예를 들어 '고통'의 실체가 존재하지 않을 것)이기 때문이다[5].

퍼트넘과 루이스의 기능주의는 물리주의와 환원의 관점에서 보았을 때, 다른 입장이라고 보여지지만, 사실상 심적 용어의 속성을 무엇으로 보느냐의 차이일 뿐, 물리적인 다수실현을 가지고, 물리적 상태 또는 행동으로 기능적 속성을 파악한다는 것은 공통적이다. 그리고 본질적인 기능주의 개념은(존재론적) 물리주의나 심신 환원과 모순되거나 충돌하지 않는다. 기능적 용어의 속성은 그 자체로 물리적인 것이 아니라고 볼 수도 있지만, 그것이 모든 경우에 물리적으로 실현된다고 가정하면 물리주의와 조화될 수 있다.³⁾ 기능주의의 두 입장이 배척관계가 아니라, 포함관계일 수 있으므로, 김영정([7] p.176)은 튜링- 퍼트넘식 기능주의와 루이스식의 기능주의는 본질상의 구분에 큰 의미가 없다고 했다.

'자유의지'는 기능류어이고, 행동 또는 물리적 상태(미시물리적 상태는 아님)를 파악함으로써 그 유무를 확인하게 된다. 그러므로 어떤 개체에 자유의지가 있는지를 확인하는 공식은 다음과 같다.

X에게 자유의지가 있다 = X가 행동 1 & 행동 2 & 행동 3... 을 가진다.

3) 김재권은 물리주의를 주장하면서 기능주의 관점을 채택한다([6] pp.248-251).

여기서 행동들은 물리적 상태로 변환될 수 있다. 이러한 행동들을 확인하기 위하여, 관찰을 하기도 하고 행동적인 테스트를 할 수도 있다. 여기서 확인하고자 하는 행동은 우리가 자유의지가 있는 것으로 상식적으로 판단하게 되는 상식적 심리이론에 따라 결정된다. 상식적, 통속적 심리학에 따라, 자유의지는 '자기 통제력'이 있는 개체에게 부여될 것이다. 어떤 개체가 '자기 통제력'이 있고, 더 나아가 '자유 의지'가 있다면, 상식적 심리이론에 따라 다음과 같은 행동을 보일 것으로 예상할 수 있다.

1. 상황과 맥락을 파악하고 그에 따라 행동과 욕구(지향성)를 절제, 조절, 선택하는 행동
2. 미래의 상황을 의도적(의식적)으로 예측하는 행동
3. 정교한 조건문(if... then...)을 만드는 행동

다른 동물들이 자유의지를 갖지 않는 것으로 보이는 까닭은 이러한 행동을 하지 못하기 때문이다. 다른 동물들은 때와 장소에 맞게 자신의 욕구를 선택하거나 조절하는 능력이 부족하기 때문에 장기적인 계획이 없이 본능적인 욕구에 우선적으로 반응하게 된다. 반면에 인간은 어떤 시점에 본능적인 식욕과 성욕이 일어나더라도, 더 큰 목적을 위해 때와 장소에 맞게 행동을 조절할 수 있다. 그리고 다른 동물들은 복잡한 언어를 사용하지 못하므로 미래의 상황을 예상하는 조건문도 만들 수 없다.

행동들의 연언만으로 '자유의지'의 유무를 판단한다고 해서 물론 이 주장이 행동주의(Behaviorism)는 아니다. 왜냐하면 그러한 행동들의 묶음이 '자유의지'와 동일한 것이 아니기 때문이다. '자유의지'는 그러한 행동을 만들어내는 기능의 담당자(occupant of the function) 혹은 인과적 역할의 담당자(occupant of the causal role)이다. 그러므로 '자유의지'는 특정한 물리적 상태 또는 행동을 만드는 **기능의 담당자와 동일하다**. 그러한 기능의 담당자는 실제로 물리적인 실현자로 환원될 수 있다. 예를 들어 자유의지를 가진 인간 뇌의 작동이 그러할 것이다. 그리고 그 담당자는 로봇이나 외계인의 두뇌구조와 같은 실현을 가질 수도 있다. 즉, 자유의지 판단의 기준이 되는 행동들이 동일하게 나타난다면, 그것을 실현하는 담당자의 미시적 구

조는 다양해도 상관없다.

그런데 한 가지 제기 될 수 있는 의문점은, 왜 실제 존재자들 중에 아직까지 ‘자유의지’를 가지는 대상이 인간 중에 한정된 것처럼 보이는가 하는 점이다. 아직까지 다수 실현된 대상이 실재하지 않는다는 사실로 인한 반론이다. 그런데, 어떻게 우리는 그러한 사실(인간에게만 있어 보임)을 알 수 있는 것일까? 그것은 자유의지기능이 인간에게 ‘실재한다’라는 것을 오히려 도와주는 사실일 수 있다. 왜냐하면 실재하는 것이 아닌 단지 비유적, 추상적 기능만으로 사용하는 용어라면, 더욱 폭넓은 사용과 적용이 가능할 것이기 때문이다. 우리가 관찰한 결과, **인간은 실제로 ‘특별한 행동’을 할 수 있기 때문에 자유의지를 가진다. 다만 인간이라고 하더라도 그러한 행동을 하지 못한다면 자유의지가 없다.** 금치산자나 정신박약, 유아의 경우가 그러하다. 그러면, 왜 실제로 인간이 그러한 자유의지기능을 가지게 되었는지를 설명할 필요가 있어 보인다. 그러한 설명은 자유의지가 실재한다는 것의 개연성을 높여줄 것이다. 그것은 다음 절에서 다루도록 하겠다.

나는 자유의지를 철저하게 기능주의적으로 설명하였다. 그런데 현대 심리철학에서 기능주의는 몇몇 비판에 직면해 있기도 하다. 나의 기능주의적 설명이 그러한 비판에 동일하게 적용되는가? 그렇지 않을 것이다. 여기서 주장하는 기능주의는 기존에 심리철학계에서 주장되었던 기능주의와 다른 점이 있기 때문이다.

현대 심리철학은 심적 현상(정신과 의식)이 물리적 현상(뇌와 신체 상태)과 어떤 관계를 맺고 있는지를 주로 다룬다. 그러한 논의에서 가장 많이 등장하는 심적 현상의 예는 ‘고통(pain)’일 것이다. 학자들은 ‘고통’과 같은 심적 상태가 뇌 상태와 동일한지에 대해 논쟁을 벌이면서 ‘개념적 분석’을 하기도 한다. 그것은 ‘고통’ 같은 심적인 용어의 본질과 속성, 개념이 무엇인지를 밝혀내는 것이다.

개념적 분석에서 커다란 역할을 한 것은 술 크립키(Saul Kripke)의 ‘고정지시어(rigid designator)’ 개념이다. 고정지시어란 모든 가능 세계에서 동일한 대상을 지시하는 용어를 뜻한다([8] p.48). 가장 전형적인 예는 고유명사다. ‘닉슨’은 어떤 가능 세계에서는 대통령이 되지 않고 농부가 되었을 수 있다. 그러나 일단 그가 ‘닉슨’으로 명명되면, 어떤 가능한 세계에서도 여전히 같은 사람을 지시한다. 고유명사는 그래서 어떤 대상을 직접 지시하는 용어이고, 그것은 필연적으로 그 대상을 지시하게 된다. 크립키는 이러한 고정지시어 개념을 자연류어(natural kinds term)에 확장

시킨다. 그가 든 예는 ‘물’과 ‘열’이었다. 과학적 발견에 따라 우리는 물이 H₂O와 동일하다는 것을 안다. 그것은 후험적으로 알게 된 사실이다. 그런데 크립키에 따르면, 물이 H₂O 와 동일하다는 것을 알게 되면, 그것은 ‘고정된다’ 그래서 물과 H₂O는 **필연적 동일성**을 가지게 된다([8] pp.128-131). 그래서 만약 어떤 외계 행성에 물과 똑같이 보이고 기능하는 어떠한 물질이 있는데 그것이 H₂O가 아니라 XYZ라면, 그것은 ‘물’이 아니다. ‘열’도 마찬가지이다. ‘열’이 후험적으로 ‘분자 운동과 동일하다는 것이 밝혀지면, <열 = 분자운동>은 고정되고, 필연적이 된다. 즉, 분자 운동이 아닌 열을 상상할 수 없다.

크립키는 그러면서 고통과 뇌신경상태가 동일하다고 말할 수 있으려면, 그것은 물이나 열의 동일성과 마찬가지로 ‘필연적’이어야 한다고 말한다. 그런데 고통이 뇌신경상태와 같다는 것은 필연적이지 않아 보인다([8] p.154). 고통을 가지면서 특정한 뇌신경상태가 아닌 것을 상상할 수 있기 때문이다. 이후에 프랭크 잭슨(Frank Jackson)도 그와 유사한 주장을 했다[9].

그런데 크립키와 잭슨이 ‘물이 H₂O와 필연적으로 동일하다’라고 주장하는 것은 ‘물’에서 기능적 속성을 배제하겠다는 주장이다. 그렇기 때문에 아무리 물처럼 보이고 기능하는 것이라도, 그것이 H₂O가 아니라면 물이 아니라고 주장하는 것이다. 거기에는 논란이 있을 수 있다. 실제로 블록과 스톨네이커(Block, N. & Stalnaker, R.)[10], 맥래플린(McLaughlin, B.)[11]은 물이 H₂O와 같다는 것이 필연적(선형적)이지 않다고 주장했다. 그리고 기능주의를 폭넓게 적용하는 학자들도 그것에 반대할 것이다. 하지만, 고유명사가 고정지시어라는 것에는 거의 이견이 없다.

그런데 ‘고통’이 고정지시어라는 것에는 ‘물’이나 ‘열’ 보다도 더욱 이견이 많다. 크립키는 ‘고통’도 고정지시어가 되어야 한다고 생각하지만([8] p.148), 논란이 많은 것으로 보아, ‘물’과 ‘열’ 보다는 기능적 속성을 더 많이 가지는 것으로 보인다. ‘물’과 ‘열’에 대한 직접지시론에 찬성하는 힐러리 퍼트넘은 다만 ‘고통’은 고정지시어가 아니며 다수실현이 가능한 기능적 용어로 보았다[4]. 그래서 퍼트넘에 따르면, ‘고통’은 미시적 구조에 관계없이 동물이나 로봇에까지도 적용될 수 있는 용어이다. 하지만 이러한 심적 상태의 기능주의도 여러 가지 강력한 반론에 직면한다. 심적인 ‘감각질(qualia)’을 무시한다는 문제와 존 씨얼(John Searle)의 중국어 방 사고 실험[12] 반례가 그러하다. 그리고 인간의 뇌구조를 기능적 속성으로 환원하는 ‘계

산주의(computationalism)'에 대한 반론도 만만치 않다.

그러나 기능주의로서의 '자유의지'는 그러한 반론의 공격을 받지 않을 것이다. 왜냐하면, '자유의지'는 어떤 상태(현상)나 대상을 지시하는 용어가 아니기 때문이다. '자유의지'는 '고통'처럼 특정한 상태를 지시하는 것이 아니라 순전히 기능만을 지시하는 용어이다. 그리고 나는 뇌구조의 작동 원리를 기능적으로 설명하려는 게 아니므로 계산주의를 가정하는 것이 아니다. 감각질의 문제도 발생하지 않는다. '자유의지'는 감각질과 무관하기 때문이다. 외계인이나 로봇의 감각질이 인간과 다르거나 심지어 없더라도 자유의지가 있을 수 있다. 그러므로 만약 인간과 기능적으로 **완전히** 동일한 로봇이 있다면, 그것이 고통과 의식을 가지는지에 대해서는 논란이 발생할 수 있겠지만, 자유의지를 가지는지에 대해서는 논란이 발생하지 않을 것이다.

이제까지 살펴 본 바에 따르면, 용어들은 대상을 직접지시하는 속성을 많이 가지는 경우도 있고, 다수실현이 가능한 기능적 속성을 많이 가지는 경우도 있다. 고유명사는 완전하게 직접지시 속성만을 가지고, '자유의지', '소화', '시각'과 같은 기능류어는 완전하게 기능적 속성만을 가진다. '물', '온도', '개', '고양이'와 같은 자연류어 속성을 가지는 용어들은 '고통'과 같은 심리 상태 용어보다는 직접지시 속성이 더 큰 것으로 보인다. 다만, 존재하는 수많은 단어들을 모두 일괄적으로 분류해서 정리하는 것은 여기서 너무 나아간 것일 수 있으므로, 여기서 살펴 본 단어들을 검토해보면 그림 1과 같은 특징을 가질 것이다.⁴⁾

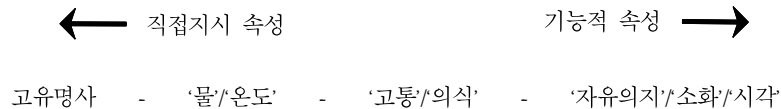


그림 1.

4) 예를 들어, '독(venom)'은 자연류어처럼 보이지만, 미시적 구조는 다양할 수 있으므로 기능적 속성이 큰 것으로 보인다. 이처럼 자연류어와 기능류어를 결정하기 어려운 것도 있다.

인간의 자유의지 능력

나는 이전 절에서 자유의지가 기능류어이고, 자유의지에 ‘기능’을 붙여서 ‘자유의지기능’이라고 불러도 그 의미에 본질적인 차이는 없다고 주장하였다. 그렇다면, ‘자유의지’와 ‘자유의지기능’은 동일한 것일까? 그것이 동일하다면, 동일성의 원리에 의해 대치시켜도 의미의 변화가 일어나지 않아야 할 것이다. 그렇다면 다음 문장이 참이어야 할 것이다.

“만약 우리에게 자유의지가 없는 것으로 밝혀진다면, 우리에게 자유의지기능도 없는 것이다.”

이것을 누구나 받아들일 수 있을까? 대부분의 사람들은 이 명제에 긍정하겠지만, 나는 약간의 불안감을 가지고 있다. 왜냐하면 실제로는 자유의지가 존재하지 않아도 ‘자유의지 역할’은 남아있다고 주장할 수도 있기 때문이다. 즉, ‘겉보기 자유의지(apparent free will)’로써의 기능을 하는 경우를 생각해 볼 수 있다. 실제로 결정론이나 양립가능론을 주장하는 많은 학자들은 자유의지를 이렇게 본다. 그래서 나는 ‘자유의지’와 ‘자유의지기능’이 완전히 동일하다고 주장하기 어려울 것 같다. ‘기능’에는 ‘역할’이라는 의미가 담겨있는데, ‘역할’은 ‘겉보기에 그럴듯하게 꾸미는’ 의미를 가질 수도 있기 때문이다. 그렇다면, ‘내면에 실제로 존재하는 기능’을 지칭하는 다른 단어를 생각해볼 필요가 있다. 그것은 ‘능력’일 것이다. 다음 문장을 생각해보자.

“만약 우리에게 자유의지가 없는 것으로 밝혀진다면, 우리에게 자유의지능력도 없는 것이다.”

이 명제를 부정할 수 있을까? 나는 이 명제는 누구도 부정하기 어렵다고 생각한다. 왜냐하면 ‘능력’은 본질적으로 ‘겉보기적인 것’이 아니라 ‘실질적으로 작동하는 기능’을 가리키는 것이기 때문이다. 그래서 위의 명제는 참이 되고, ‘자유의지’는 ‘자유의지능력’으로 바꿔 쓸 수 있다. 물론 ‘자유의지능력’도 기능류어이다. 외계인이나 로봇도 자유의지능력을 가질 수 있기 때문이다.

그래서 궁극적으로 우리가 밝히고 싶어하는 것은 ‘우리에게 자유의지능력이 있는가?’ 하는 것이다. 그것은 ‘우리에게 자유의지기능이 있는가?’ 라는 물음보다 더욱 핵심사항을 이야기 하고 있다. 우리에게 겉보기로써 자유의지기능이 존재하는

것은 분명해 보인다. 쟁점이 되는 것은 그러한 것이 아닌 진정한 자유의지, 즉, 자유의지능력이 존재하는가 하는 것이다.

미래가 하나로 고정되어 있기 때문에 대안 자체를 가질 수 없다는 일부 결정론의 주장에서는 자유의지능력 자체를 부정하겠지만, 그에 대한 비판은 다음절에서 하기로 하고, 먼저, ‘능력’이란 무엇인지부터 살펴보자. 의외로, 우리의 내면에 ‘능력’이란 것이 실제로 존재하는 것으로 인정받게 된 것은 오래된 일이 아니다. ‘능력’이 물리적으로 존재하는 것으로 인정받게 된 것은 인지과학이 태동하기 시작한 20세기 후반의 일이다. 그것은 20세기 중반 심리학의 지배적 사상이었던 행동주의를 타파한 흐름이었다. 행동주의는 내면에 있는 능력을 마치 없는 것처럼 여기면서 무시해왔다. 행동주의에 반발해서 나타난 인지주의와 노엄 촘스키(Noam Chomsky)의 ‘인간-보편적 언어능력’ 이론은 신경 과학과 결합되면서, 능력이 내면에 ‘실제로(물리적으로) 존재하는 것’으로 학계의 인식을 바꿔놓았다. 많은 인간의 능력들은 유전자로 인해 선천적인 능력으로 나타나기도 하지만, 뇌 과학적 관점에 따르면, 뇌의 가소성으로 인해 뉴런들의 연결상태가 변화하면서 새로운 능력이 생기게 된다. 그래서 ‘능력’이라는 것은 비과학적이거나 신비로운 존재가 아니다. 그것은 상식적인 것일 뿐만 아니라 우리의 신체 내에 물리적으로 존재하는 것이다. 언어학자 촘스키는 ‘언어능력’을 주장하였는데, 인지과학자 스티븐 핑커(Steven Pinker)는 더 나아가 그것이 진화에 의해 만들어진 선천적 능력에서 기인한다는 ‘언어본능’을 주장하였다[13]. 그 두 학자 사이에 진화를 바라보는 입장에서 약간의 차이가 있기는 하지만, (다윈식)진화를 통해 인간이 언어능력을 가지게 되었다는 견해는 현재 학계에서 일반적으로 받아들여지고 있다.

인간에게 어떤 능력이 있다면, 그것이 왜 생겨나게 되었는지 물어봐야 할 것이다. 인간에게 언어능력이 생겨난 것은 다른 기능들과 마찬가지로 생존과 번식에 이점이 있기 때문일 것이다. 그리고 인간에게 자유의지 능력이 있는 것도 그와 마찬가지로 마찬가지일 것이다. 자유의지에 대한 양립가능론자인 대니얼 데닛(Daniel C. Dennett)은 *Freedom evolves* 에서 생물학적 진화의 과정을 통해 인간이 자유의지를 가지게 되었다고 주장한다[14]. 나는 그 커다란 주장에 동의한다. 인간은 복잡한 사회를 이루게 되었고, 사회생활이 생존에 매우 중요해진 상황에서 생존과 번식을 위해 두뇌와 언어가 발달하게 되었다. 그와 함께 발달한 것은 좋은 선택(의사결정)능력이

었을 것이고, 또한 사회생활을 하는데 있어서 맥락과 상황, 분위기를 파악해서 자신의 행동을 현명하게 결정할 필요가 있었을 것이다. 자신이 어떤 행동을 하는가에 따라서 미래의 결과가 달라질 수 있다는 것을 알게 된 우리의 선조들은 여러 가지 미래의 시뮬레이션을 더욱 정확하게 하려고 했을 것이다. 잘못된 선택을 하면 나쁜 결과가 발생하고, 우리는 **더 좋은 선택을 할 수 있기** 때문이다. 그것이 바로 의사결정능력의 진화이자, 자유의지능력의 진화이다. 만약, 자유의지가 단지 ‘겉보기’일 뿐 허상이라면, 그런 쓸데없는 기능은 왜 생겨난 것일까? 자유의지가 우리 상식처럼 실제로 작동하는 것이라면, 진화의 최대 수혜자인 인간이 얻은 실용적인 능력으로 설명할 수 있다. 그러나 그것이 환상에 불과하다면, 회의론자들은 우리가 그러한 환상을 가지게 된 이유를 설명해야 한다.

‘능력’의 특성상, 그 능력의 ‘ 좋음(뛰어남) 과 나쁨’이 나뉘어질 수 있다. 그 기능을 잘 수행하는 것은 ‘좋은 능력’이 된다. 종 간에 능력의 차이가 있을 수도 있지만 개별자들 간에 차이가 있을 수도 있다. 개는 인간보다 후각이 더 발달되어 있다. 같은 인간종에서도 시력이 더 좋은 사람이 있고 나쁜 사람이 있다. 자유의지도 능력이기 때문에 개별적인 인간마다 좋고 나쁨의 차이가 있을 수 있다. 자유의지는 좋은 선택을 하기 위해서 진화되었을 것이므로, 좋은 자유의지능력은 아마도 선택을 잘하는 능력일 것이다. 자유의지능력이 뛰어난 사람은 좋은 상상력과 시뮬레이션 능력, 상황파악능력, 올바른 세계관, 자신의 욕구와 목표에 대한 적절한 통제를 통해 합리적인(reasonable) 선택을 한다. 그것은 그 자신에게 이로운 것이다. 대체로 그들은 ‘후회하지 않을’ 선택을 할 것이다. 이러한 능력은 사람마다 다르게 가질 수 있지만, 교육이나 훈련을 통해 계발할 수도 있을 것이다.

이제까지 나는 자유의지가 기능류어이고, 우리에게 자유의지가 있는지를 확인하기 위해서 미시적인 조사가 반드시 필요한 것이 아니라는 이야기를 했다. 그리고 자유의지기능은 진화를 통해 갖게 된 능력일 것이라고 했다. 다만 이제까지의 설명은 자유의지가 존재한다는 개연적인 설명이었지, 그 존재의 ‘필연적인’ 근거는 아니었다. 그러므로 자유의지가 실제로 존재하지 않을 가능성도 남아있을 것이다. 필연적인 근거란 미시적인 조사를 통해 과학의 법칙으로 판명되는 것을 말한다. 어떤 사람들은 자유의지론자에게 그러한 ‘필연적인’ 근거, 즉 자유의지의 미시물리적 기초(microphysical basis)를 보이기를 요구할지 모른다. 그러나 인간에게 ‘소화기

능'이 있다는 것을 알기 위해 인간의 미시물리적인 소화 구조를 알아내야만 하는가? 기능류어는 행동 관찰(테스트)만으로 그 존재 유무를 가정할 수 있다. 물론 미시적인 조사를 해서 필연적인 근거를 내놓는다면 더욱 좋을 것이다.⁵⁾ 그러나 아직 그것이 존재하는지 아닌지를 판단할 수 있을 정도의 미시적인 조사가 제대로 이루어지지 않은 상황에서, 자유의지가 존재한다는 것이 상식적 심리학에 속한다면, 필연적인 근거가 반드시 필요한 쪽은 자유의지론이 아니라 반대로 회의주의자들 쪽이다. 과연 그들은 자유의지가 없다는 것에 필연적인 근거를 내놓았는가? 아니면 그럴 확률이 크다는 개연적인 증거라도 내놓았는가? 다음 절에서는 결정론이 올바른 증거와 논거를 보여주지 못했다는 것을 논증할 것이다.

결정론은 타당한 근거를 가지고 있는가?

나름대로 똑똑한 사람들이 자유의지를 부정하고 결정론을 선호하게 되는 데에는 여러 가지 요인이 있다. 그 중 핵심이 되는 것은 자유의지가 '자연법칙(물리법칙)과 모순된다는 생각일 것이다. 먼저 단순한 반론부터 살펴보자. 일반적으로 자유의지는 '자연이나 신의 구속에서 벗어나 행동할 수 있는 능력'을 말한다. 그렇기 때문에 어떤 사람들은 인간도 자연의 일부분이고, 자연 법칙의 구속에서 벗어날 수 없기 때문에 자유의지가 불가능하다고 생각한다.

그런데 '구속'이라는 말은 '통제(control)' 또는 '적용(application)'을 의미할 수 있다. 자유의지로써 벗어날 수 있는 것은 '통제'를 뜻한다. 반면에 우리가 결코 벗어날 수 없는 자연 법칙은 그것의 '적용'이다. 본능에 따라 움직이는 동물들이나 좀비 같은 것들은 자신을 스스로 통제한다기 보다는 자연에 통제당하는 것처럼 보인다. 우리가 다른 사람의 강압에 의해 억지로 행동을 해야 한다면 그것은 타인에 의해 통제되는 것이다. 반면에 자연법칙은 언제 어느 때라도 우리에게 적용되고 있다.

5) 자유의지론자인 로버트 케인(Robert Kane)은 *The Significance of Free Will* [15]에서 자유의지가 발생하는 미시물리적 기초에 관해 어느 정도 설명하고 있다. 자세한 설명은 본 논문 7절에서 나올 것인데, 그러한 미시물리적 설명이 옳은 것인지는 확실치 않다. 미시물리적 구조를 확인하는 것은 자연과학의 발전으로 가능해질 것이다.

자유지론은 자연법칙의 그러한 ‘적용’을 받아들일 수 있다. 대니얼 데닛은 ‘자연은 우리를 통제하지 않는다.’라고 했다.⁶⁾

그러면 좀 더 철학적으로 다듬어진 입장에 대해 살펴보자. 강한 결정론자인 반인와겐(van Inwagen)은 다음과 같은 전제에서 출발함으로써 자유지와 결정론이 모순됨(양립불가능함)을 주장하고, 따라서 결정론이 옳음을 주장한다[17].

“만약 결정론이 옳다면, P_0 와 L의 연언(conjunction)은 P를 함축한다.”(P_0 는 세계의 이전 상태 표현명제, L은 물리법칙들, P는 세계의 나중 상태 표현명제)

쉽게 설명하면, 결정론에 따르면 세계의 이전 상태에 물리법칙을 적용시키면 세계의 나중 상태를 연역적으로 알 수 있다는 것을 말한다. 물론 여기서 물리법칙이란 우리가 아직 알지 못할 수도 있는 ‘완성된 물리법칙들’을 말한다. 이 조건문은 그 자체로는 문제가 없어 보인다. 결정론은 실제로 그것을 의미하는 것이 맞다. 이것으로 출발해 반인와겐은 자유지와 결정론은 양립할 수 없고, 물리법칙의 성질을 받아들이는 한, 결정론을 받아들여야 함을 주장한다. 그런데 ‘물리법칙의 성질’에 문제가 있다. 물리법칙은 ‘필연성(불가피성)과 ‘결정론’을 전제로 하고 있는 것일까?

물리법칙은 필연성(불가피성)을 전제로 하는 것으로 보인다. 그러나 그것은 어떤 물리법칙이 **발견**되었을 때, 그것의 ‘적용’이 필연성을 가진다는 것을 뜻한다. 비유하자면, 물이 H_2O 라는 것이 밝혀지면 그것은 그 후로 필연적으로 동일하다고 할 수 있을 것이다(3절 참조). 그런데 어떤 물리법칙의 성질이 ‘비결정적인 것으로 밝혀진다면, 그것은 **필연적으로 비결정적(불가피하게 비결정적)**이 된다. 사람들은 종종 물리법칙의 필연성, 불가피성, 결정론을 혼동한다. 뉴턴 물리학의 영향으로 많은 사람들은 물리 법칙이 모두 결정론적인 것이어야 하는 것으로 생각하기 쉽다. 그러나 양자이론은 물리법칙이 결정론적이라는 것을 부정한다. 현재 주류의 관점에 의하면, 양자 운동의 불확정성은 단지 우리 인식의 한계(정보부족) 때문에 발생하는 것이 아니라, 미시 세계의 물리적 원리가 실제로 그러하다고 상정한다([15] p.9). 양자의 불확정성과 비결정성은 현재까지 80여 년 간 수많은 실험을 거치며 경험적으로 확인된 사실이다([18] p.54).

6) “The environment, not being an agent, does not control us.”[16] p.61

유사한 사례로, 어떤 사람들은 ‘모든 사건에는 원인이 있다.’ 라고 하는 인과법칙을 따름으로써 결정론을 옹호하기도 한다. 그러나 설령 그러한 인과관계를 받아들인다고 해도, 그것은 (비결정론을 포함한) 물리법칙적인 원인을 따르는 것일 뿐, 결정론을 따를 필요는 없다.

우리의 물리계가 결정론적으로 작동한다는 생각은 무생물의 거시적 운동과 같은 자연의 일부분을 관찰했을 때 후험적으로 가지게 된 것이다. 그런데 그것을 인간의 자유의지에까지 확장시켜 성급한 일반화를 해선 안된다. 우리는 돌맹이나 행성에 자유의지가 있다고 생각하지 않는다. 거시적인 측면을 보기를 원한다면 인간의 행동을 관찰해야 한다. 자유의지는 경험적으로 인간에게 있는 것이기 때문이다.

결론적으로, 자유의지론자는 물리법칙의 ‘필연성’은 받아들일 수 있다. 그러나 물리법칙의 ‘결정론’은 받아들이지 않는다. (일원론을 전제로 하는) 자유의지론은 물리법칙의 필연성을 받아들일 수 있으므로 물리학의 기본 전제나 이제까지의 성과와 충돌하지 않는다. 오히려 현대물리학의 성과와 충돌하는 쪽은 결정론이다.

다음 반론으로, 자유의지 논쟁에서 유명한 리벳(Libet, B)의 반례에 대해서 살펴보자[19]. 그의 실험에서는 피험자의 두피에 전극을 부착하고 뇌에서 준비전위(RP : readiness potential)가 나타나는 것을 측정했다. 준비전위는 자발적인 행동이 나타나기 약간 전에 나타나는데, 리벳은 자발적 의식과 준비전위 간에 선후관계를 살펴보고자 했다. 피험자는 돌아가는 초시계를 보면서 본인이 원할 때 손목(또는 손가락)을 움직이게 된다. 피험자는 ‘최초의 의도나 의식의 자각(first aware of wish or urge to act)’이 일어나는 시각을 보고하였다. 그러한 의식의 시점은 행동이 나타나기 약 200msec 전에 보고되었다. 그런데 놀랍게도, 준비전위는 그 의식의 시점보다도 약 350msec 전에 일어났다. 그것은 우리가 결정하는 의식 이전에 이미 뇌에서 결정을 내린 것이며, 의식은 뇌의 결정을 따라가는 것처럼 보인다. 그래서 자유의지가 존재한다는 우리의 상식에 반례를 보여주는 실험으로 여겨지게 되었다.

물론 이 실험의 해석을 비판하는 여러 가지 시도들이 있어왔다. 데닛은 그 시간 차이를 의식이 일어난 뒤에 뇌의 시각중추에서 시계의 장면을 확인하는데 따르는 뇌 안의 작동 과정에서 발생하는 시간차이라고 해석했다([14] p.234). 그 간격은 매우 짧기 때문에 그러한 해석도 일리가 있다. 그러나 나는 데닛의 그러한 해석에 의존하지 않고도, 리벳의 실험은 자유의지가 존재한다는 것에 아무런 반례가 되지

못한다고 생각한다.

고전적인 자유의지 개념(이원론적 자유의지론)은 그림 2의 a)와 같은 것이다. 의식이 먼저 있고, 그 의식의 결정에 따라 뇌가 움직이는 것이다. 그런데 리벳의 실험결과는 b)를 보여줬고, 자유의지가 존재한다는 생각이 위협받게 되었다. 여기에서 실선화살표는 ‘출발점이 목표물을 결정한다.’는 의미이다. 그래서 그 중간 과정은 결정된 과정이고, 결정론이 지배하는 과정이다.

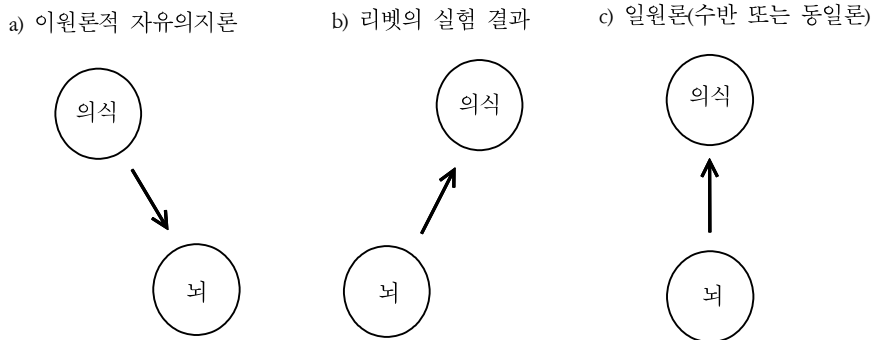


그림 2.

나는 이 중간 과정에서 있을지 모를 양자 불확정성이나 카오스현상 같은 것도 무시하겠다. 또한 리벳은 준비전위가 나타난 뒤라도 의식이 일어난 뒤 짧은 기간 안에 행동을 거부(veto)할 자유의 가능성을 주장했는데,⁷⁾ 심지어 그것도 무시하겠다. 이 과정은 결정된 과정으로 가정한다.

그런데 나는 2절에서 밝힌 것처럼, 이원론을 부정하고, 물리계로 환원되는 일원론을 받아들인다. 그래서 그림 2의 c)와 같이 뇌가 의식을 결정하는 것을 받아들이고, 본체적 의식이 뇌를 결정하지 않는다는 것을 받아들인다. 그리고 자아와 뇌의 동일론을 주장하였으므로, 뇌가 의식을 결정하는 것은 자아가 의식을 결정하는 것

7) 리벳은 “Do We Have Free Will?”[19]에서 의식이 행동을 처음으로 유발할 수는 없지만, 의식이 일어난 뒤 행동이 발생하기 전까지 200msec에서 오차 50msec를 뺀 150msec 구간에서 준비전위가 행동으로 이어지는 것을 막을 수는 있을 것이라고 주장했다.

과 같다. 나는 a)와 같은 과정이 없어도 자유의지가 존재한다는 것에 문제가 없다는 것을 주장하고 있다. 그러면 자유의지와 비결정적 과정은 어디에 있는가? 그림 2.에 보이는 결정된 과정에서는 존재하지 않는다. 그보다 더 앞에 존재한다. 그러한 비결정적 과정은 그림 3에서 시간 t_0 에서 t_1 으로 가는 **점선화살표**에서 존재한다.

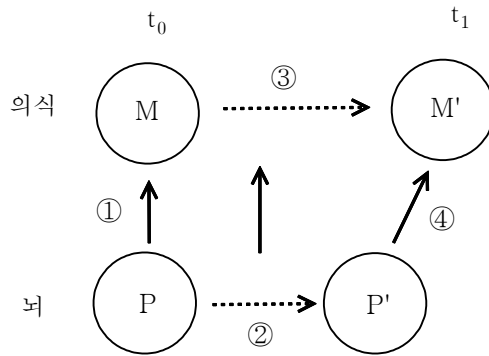


그림 3.

리벳의 실험을 이 그림으로 설명하면, t_1 에서 M'라는 의식이 나타났고, 그에 조금 앞서서 뇌 상태 P'이 발생했다(P'를 준비전위상태로 보아도 좋고, 의식과 환원되는 뇌 상태로 보아도 좋다). 그 뇌 상태에서 의식 상태로 가는 과정은 결정된 것으로 가정한다(④). 자유의지와 비결정성은 그 과정에서 없을 것이다. 자유의지의 과정은 준비전위가 발생하기 이전의 점선화살표의 과정에 있는 것이다. 즉, 준비전위가 일어나기 전에 뇌에서 자유의지와 비결정적인 과정이 있었던 것이다. 그것은 이해하기 어렵지 않다. 피험자는 언제 손가락을 까딱할지를 결정하는 자유의지가 있었다. 결정을 마친 시점 이전의 시간이 바로 언제 까딱할지를 정할 수 있는 비결정적인 구간이다. 그 구간을 짧게 할 것인지, 아니면 한참 후에 까딱할 것인지는 결정되어 있지 않다. 그 비결정의 구간이 ②와 ③인 것이다. 반면에 ①과 ④는 결정된 구간이다. 리벳의 실험이 확인한 것은 ④의 과정일 뿐이다. 그러나 자유의지

는 그 과정에 있는 것이 아니다. 그 이전의 과정에 있는 것이다.

결국, 리벳의 실험결과는 그림 2의 a)와 같은 이원론적 모델에 대한 반례일 뿐, 일원론을 전제로 하는 자유의지에 대한 반례는 되지 못한다. 더구나 a)와 같은 이원론 모델은 현재 대체로 받아들여지지 않고 있다. 우리가 의식이 몸을 움직이는 것처럼 느끼는 까닭은 다만 의식과 동일한 뇌상태가 몸을 움직이기 때문일 것이다. 리벳의 이 실험은 자유의지의 반례가 된다고보다는 심신이원론이 틀리다는 것을 보여주는 데 의의가 있어 보인다.

②와 ③의 구간이 비결정적 구간이라고 해서 그 과정이 무작위성(randomness)을 가진다는 의미는 아니다. 우리의 행위가 오직 무작위적, 우연적으로 결정되는 것이라면 그것은 결코 자유의지에 의한 것이라고 볼 수 없다. 비결정적이라는 것은 '미래가 특정한 결과로 미리 결정되어 있지 않다'는 것을 뜻한다. 양자이론에서 불확정성이란 완전한 무작위성이 아니라 그러한 비결정성을 말한다. 양자역학이 일정한 확률분포를 따른다는 점은 물리 세계가 무작위적인 것은 아니라는 것을 보여 준다.

다음의 반론은 자유의지론에 가장 치명적인 것이다. 그 이유는 (내가 알기에) 모든 양립가능론자들이 이것을 거부하지 못하고 결정론을 인정하기 때문이다. 자유의지론은 '다르게 할 수 있었음(could have done otherwise)'을 주장하는데, 그것은 '동일한 과거와 동일한 자연 법칙을 가졌어도 다르게 할 수 있었음'을 주장하는 것과 같다. 그것을 받아들일 수 있을까? 자유의지론은 그것을 주장하는 것이 맞다. 그러나 우리들은 그것이 불가능하다는 직관을 가질 수 있다. 양립가능론자인 대니얼 데닛은 자유의지와 관련해서 1984년에 *Elbow Room*[16], 2003년에 *Freedom evolves*[14]를 썼는데, *Elbow Room*은 '다르게 할 수 있었음'을 부정하면서 그럼에도 자유의지가 가능함(양립가능론)을 주장하는 것이 중심주제였다(다만 그것은 '책임귀속'의 차원이었다). 그는 '같은 과거와 같은 자연 법칙을 가지면 다르게 할 수 없다'를 받아들이기 때문에 결정론을 받아들일 수밖에 없고, 그 후로도(2003년 저작에서도) 계속 결정론을 인정하는 양립가능론자로 남아있는 것으로 보인다.

그런데 우리는 앞에서 리벳의 반례를 해결하는 방법을 보았으므로, 이 문제를 해결하는 것이 어렵지 않다. 우리는 다른 대안이 존재하는 것(비결정성)이 어떠한 시점에 가능한 것이 아니라, 일정시간 이상의 **시간 간격**을 가져야 한다는 것을

보았기 때문이다. 예를 들어, 만약 준비전위가 일어난 뒤 행동까지 550msec가 결정된 과정이라고 한다면, '지금으로부터 550msec 후까지는 그 행동에 대안은 없다. 대안은 그 이상의 간격 후에 가능할 것이다.'

다시 '다르게 할 수 있음'이 무엇을 의미하는지 생각해 보자. 자유의지의 논의에서 '다르게 할 수 있음'은 종종 t 시점을 상정한다. 그리고 테닛은 그 때에 다른 대안이 존재하지 않기 때문에 결정론을 받아들여야 한다고 생각한다.⁸⁾ 그런데 알고 보면 그건 상식적인 이야기이다. 우리는 한 시점에 동시에 다른 것이 존재한다고 생각하지 않기 때문이다. 도입부에서 언급한 예를 떠올려보자. 진호는 핫도그 가격을 1500원으로 할 수도 있고 2000원으로 할 수도 있다고 생각하지만, 미래에는 결국 하나의 가격으로 결정될 것이다. 1500원으로 팔면서 동시에 2000원으로 팔수는 없다. 즉, 미래시점이 현재시점으로 바뀌었을 때는 서로 다른 두 세계는 존재하지 않을 것이다. ' t 시점에 다른 대안은 불가능하다'는 직관은 그러한 상식 때문에 생겨난 것이다.

' t 시점에서 같은 과거와 같은 자연법칙을 가지면서 다르게 할 수 없다.' 라는 말은 혼동을 부른다. 그것은 부사구 수식의 중의성 때문일 것이다. ' t 시점이 같은 조건(같은 과거, 같은 자연법칙)을 가지는 시점인가? 결과가 도출되는 시점인가? 아니면 전부 다인가? 결과가 도출된 시점이거나 전부 다일 때 하나의 결과만을 갖는다는 것은 모두가 아는 상식이 된다. 그것은 그림 3의 ①, ④의 과정과 같은 결정된 과정이다. 그러나 ' t 시점이 과거의 시점인 t_0 시점만을 의미한다면, 결과가 나타나는 t_1 시점에서는 다르게 되는 것이 가능하다. 다시 말해, 한 시점에서 같은 과거와 같은 자연법칙을 가지게 되면 하나로 결정되는 것이 당연하지만, 과거의 시점에서 같은 과거와 같은 자연법칙을 가지면 미래의 시점에서는 다르게 될 수 있다. 자유의지론이 주장하는 것이 바로 그것이다. 즉, 자유의지는 '동일한 조건'과 '대안이 가능한 결과가 다른 시점'이라는 것을 가정한다. 그런데 결정론자들은 이것을 같은 시점으로 오해한 것이다. 같은 시점에 동일한 원인과 조건이 있을 때 결과가

8) 케인(Kane, R.)는 대안가능성(AP: Alternative Possibilities)을 t 시점에(at t) 행위자가 A를 할 수 있고 다른 행동도 할 수 있다는 것으로 정의했고([15] p.33), 테닛은 이것을 인용하면서 t 가 한 시점인지 짧은 기간인지 불분명하고, 매우 짧은 기간으로 하면 하나로 결정되지만 그 기간을 길게 잡을수록 비결정적이 된다고 하였다([14] pp.118-120).

하나가 된다는 것은 자유의지론이 부정하지 않는 이야기이다. 그것은 '모두가 아는 상식적인 결정론'과 다를 바 없다.

철학적, 신경학적 관점에서 결정론을 긍정하는 학자들은 앞에서 살펴 본 여러 잘못된 근거들로 인해 결정론을 선호하게 되었다. 그리고 양립가능론자들은 대개 '책임귀속'과 '겉보기 자유의지'로써 자유의지의 존재도 긍정한다. 그런데 위의 반박들이 모두 오류나 착각에서 비롯된 것이므로, 결정론은 근거를 잃게 되고, 양립가능론을 주장할 필요성도 없어지게 된다. 그런데 대니얼 데닛은 *Freedom evolves* 에서 결정론을 긍정하면서도 실제로 미래가 다양할 수 있다는 주장을 한다.⁹⁾ 그것은 '한 시점에서 다른 대안은 불가능함'을 받아들이면서도 미래가 실제로 열려있다는 것을 주장하는 것으로 보인다. 그렇다면 그것은 나도 긍정하는 것이므로, 그는 결정론을 포기해야 한다. 그런 식의 결정론은 모두가 아는 상식에 불과한 것이기 때문이다.

무의식은 자유의지에 반례가 되는가?

이번에는 심리학을 기반으로 한 회의론에 대해서 살펴보자. 이들의 주장은 대개 개연성에 근거한 비판이다. 사실 우리의 행동을 살펴보면 상식적으로 인간이 자유의지를 가지고 있는 것처럼 보인다. 그런데 심리학이 무의식에 대한 탐구를 시작하면서 무의식은 자유의지에 위협으로 등장하게 되었다. 이러한 계열의 회의주의자들은 특히 프로이트(Freud, S)의 정신분석학과 행동주의 심리학에 많은 영향을 받았다. 프로이트는 자의식이 무의식이라는 거대한 빙산의 일각일 뿐이라는 점을 우리에게 각인시켰으며, 무의식의 부정적인 면을 널리 알린 학자이다. 그에 따르면, 무의식은 '원초적인 충동'이며, 사회적으로 받아들여지지 않는 것이기 때문에 억압해야 하는 것이다. 그러나 원초적인 무의식은 정신의 대부분을 차지하고 있기 때

9) 데닛은 그 책[14]에서 결정론을 긍정하면서도 겉보기 대안(apparent options)이 아닌 진정한 대안이 존재하며(p.25, p.102), 결정론은 고정된 미래를 의미하지 않으며(p.91), 결정론은 우리가 다른 것을 할 수 없었다거나 모든 사건이 원인을 지닌다거나 우리의 본성이 고정되어 있다는 것을 의미하지 않는다(p.95)고 했다.

문에 우리는 무의식의 지배를 받을 수밖에 없고, 진정한 자유의지는 설 자리를 잃게 될 것이다. 그리고 스키너(Skinner, B. F.)와 왓슨(Watson, J.)의 행동주의 심리학에 영향을 받으면, 인간이 다른 동물들처럼 입력에 자동적(무의식적)으로 반응해서 출력을 내놓는 것처럼 보일 수도 있을 것이다. 그런데 사실, 최근의 심리학계에서 그 두 가지 심리학은 많은 비판에 직면해 있다.

다만 최근의 심리학이 무의식의 역할에 대해 강조를 하고 있는 것은 사실이다. 인간 심리에 자연주의적 접근을 하는 심리학은 그 특성상 자연 법칙과 같은 것을 찾기를 원하고, 본능적, 자동적으로 일어나는 무의식에 대한 연구에 집중하게 된다. 그리고 최근에는 뇌에 대한 신경학적 접근을 많이 하고 있다. 그런데 무의식에 너무 집중하다보면, 어떤 학자들은 인간이 자유의지나 자의식이라고 믿고 있는 것도 모두 알고 보면 무의식에 의한 것이라고 주장할 것이다. 만약, 자의식이 모두 무의식에 의한 것이라면, 그것은 자의식이 스스로를 통제할 것이 아닐 것이고, 자유의지는 없는 것처럼 보인다.

‘무의식’은 의식적으로 떠오르지 않으면서 의식과 신체를 통제하는 영역으로 정의될 것이다. 그런데 이러한 무의식의 통제를 받는 것은 자유의지를 없애는 것일까? 우리의 생각과 의식은 뇌가 근원이 되는데, 물질적인 뇌는 의식과 구분되며, 그러한 뇌는 의식적으로 자각되지 못한다. 그렇다면 뇌가 통제하는 것은 자유의지가 아닌가? 나는 이미 자아가 뇌 상태와 동일하다는 것을 가정하였으므로 이러한 생각은 거부된다. 세포들로 이루어진 물질들이 무의식적 영역이므로, 그것에 의한 통제가 자유의지가 아니라는 것은 앞에서 비판한 이원론적 관점이다. 그러므로 단지 무의식적 영역이 통제하는 것이라고 해서 자유의지가 없다는 것은 옳지 않은 생각이다.

회의주의자들이 그것을 주장하는 것이 아니라면, 아마도 그들은 인간의 본능적 욕구나 통제할 수 없는 무의식으로 인해 우리의 의식과 자유의지가 지배받고, 따라서 진정으로 자율적이거나 합리적인(reasonable) 사고는 할 수 없다고 주장할 것이다. 이러한 생각은 다음과 같은 선입견에 지대한 영향을 받았을 것이다. 1. 우리는 무의식을 통제할 수 없다. 2. 무의식은 삶에 나쁜 영향을 미친다. 3. 무의식은 본능적(선천적)이다. 그런데 이 세 가지 선입견은 모두 사실이 아니다.

우리는 무의식을 통제할 수 있다. 우리가 성욕이나 식욕, 그 밖에 모든 지향성

을 자유자재로 통제할 수 있다는 말이 아니다. 우리는 그러한 무의식들이 행동으로 옮겨지는 지점에서 통제할 수 있다. 그것이 자유의지의 핵심적 기능이다. 3절에서 자유의지의 기능적 테스트 조건을 설명할 때, 여러 가지 욕구들을 상황과 맥락에 따라 선택해서 행동할 수 있다는 것을 언급하였다. 우리는 자유의지에 의해서 때와 장소에 맞게 행동할 수 있고, 더 중요한 목표를 위해 욕구에 우선순위를 정해 행동하고 절제한다. 그것이 인간과 짐승 간의 **자연스러운** 차이점이다(윤리적인 이야기를 하는 것이 아니다). 또한 내면적으로 느끼는 욕구의 크기도 혼란이나 수양을 통해서 어느 정도 통제하는 것이 가능하다. 그것에 한계가 있을 수는 있어도 불가능한 것은 아니다.

‘통제할 수 없음’과 프로이트 심리학의 영향으로 무의식에 대한 부정적인 생각이 퍼지게 되었을 것이다. 그러나 대부분의 무의식은 우리에게 필요하며, 많은 도움을 주고 있다. 먼저 적응무의식(adaptive unconscious)에 대해서 살펴보자[20]. 우리가 정상적인 생활을 할 수 있도록 지탱해주는 것들의 대부분은 의식적으로 자각되지 않는다. 오히려 의식적으로 자각되지 않는 것이 더 도움이 되는 경우가 많다. 운동을 하거나 주의를 기울이거나 말을 할 때마다 의식적으로 일일이 통제를 해야 한다면 에너지 소비가 매우 클 것이고, 우리의 의식은 그것에 정신이 팔려 다른 일을 하지 못할 것이다. 심리학자 티모시 윌슨(Timothy D. Wilson)은 “인간의 마음은 고차원적이고 정교한 사고의 상당부분을 무의식에 넘길 때 가장 효율적으로 작동한다.” 라고 했다([20] pp.6-7).

적응무의식은 대개 선천적인 무의식이라 할 수 있다. 이처럼 무의식의 많은 부분이 선천적인 것은 사실이다. 그러나 인간은 후천적으로 가지는 무의식을 선택하고 계발할 수 있다. 인간은 여러 가지 ‘능력’을 키우고 계발할 수 있다. 예를 들어 운전 능력, 수영 능력, 언어 능력, 외국어 능력, 소방관의 능력, 의사의 능력, 감정사의 능력 등등이 있을 수 있다. 이러한 능력의 많은 부분은 무의식적인 것이다. 어떤 능력을 높은 수준으로 가진 사람을 전문가(expert)라고 하는데, 전문가는 좋은 무의식을 가지고 있으며, 그것을 적극적으로 활용한다. 반면에 초심자는 무의식적 능력이 부족하기 때문에 전문가들이 무의식적으로 하는 일도 일일이 의식적으로 떠올려서 따라해야 한다. 심리학자 게리 클라인(Gary Klein)은 전문가들이 많은 경험과 피드백을 통해 무의식과 직관을 계발시킨다고 보고하였다([21] p.104). 그는

의사결정(decision making)에 관한 전문가들의 특징에 대해 연구했다. 그에 따르면 전문가의 무의식은 초심자가 보지 못하는 것을 볼 수 있도록 하는데, 그것은 초심자가 알아차리지 못하는 패턴, 변칙적인 것, 큰 그림, 기회와 즉흥적 대처, 미세한 차이, 자기 자신의 한계점 등이다([21] pp.148-149).

심지어 전문가에게는 종종 의식적인 사고보다 무의식적 사고가 문제해결에 더 도움이 되는 경우가 있다. 데익스테르후이스(Dijksterhuis, A)와 동료들의 실험[22]에서는 축구에 대해 많이 아는 전문가 그룹과 초심자그룹을 대상으로 축구 경기 결과를 예측하는 실험을 했는데, 전문가의 경우에는 2분 동안 그 문제에 대해 곰곰이 생각하게 한 조건보다도 오히려 2분 동안 관계없는 일을 했을 경우에 더 예측이 정확했다. 즉, 무의식적 판단을 한 경우에 더 결과가 좋았다.

의사결정은 자유의지의 주요 무대이다. 우리는 중요한 의사결정을 할 때 의식적으로 곰곰이 생각하게 되고, 그것은 대체로 좋은 방법이다. 그러나 전문가의 예에서 보듯이, 의사결정에 무의식이 개입되는 것이 좋은 결과를 가져다주는 경우가 많이 있다. 다만 노벨경제학상을 수상한 대니얼 카너먼(Daniel Kahneman)처럼, 무의식적 선입견이나 휴리스틱이 의사결정에 나쁜 영향을 미칠 수 있고 합리적 판단을 방해한다는 점을 강조하는 심리학자들도 있다[23]. 그러나 무의식이 좋은 것으로 계발될 수 있다는 것은 사실이고, 많은 무의식이 우리에게 도움을 준다는 것도 사실이다. 결론적으로, 무의식을 보편화시켜서 그것이 우리의 의사결정에 방해가 된다거나 반대로 도움이 된다고 말할 수는 없다. 그리고 우리는 무의식을 (어떤 수준에서) 통제할 수 있기 때문에, 무의식으로 인해 자유의지가 제거된다는 주장의 근거는 타당하지 않다.

아직도 미련을 버리지 못하는 몇몇 사람들은 ‘우리의 자유의지는 과거의 협소한 지식과 욕구에 한정되기 때문에 진정한 자유가 없다,’고 주장할지도 모른다. 그러나 그 주장은 마치 ‘우리는 마음대로 공중부양을 할 수 없다. 그러므로 우리에게 진정한 자유가 없다.’는 주장처럼 들린다. 자유의지의 원천이 과거의 경험과 욕구에 한정된다고 하더라도, 그 시점의 조건을 활용한 자유의지를 가질 수 있다. 다만 자유의지 능력에 일정한 한계를 가질 뿐이다. 더구나 그 원천들을 확장하고 계발할 수 있기 때문에, 자유의지의 한계도 고정되어 있다고 할 수 없다. 똑똑한 사람들은 자신의 자유의지를 계발하기 위해서 그 원천을 계발해나갈 것이다.

자유의지와 책임

자유의지에 관한 논쟁은 전통적으로 책임의 문제와 매우 관련이 깊거나 혹은 그 자체이다. 일반적으로 우리는 자유의지로 행했을 때 그에게 책임이 발생한다고 생각하는 경향이 있다. 만약 어떤 사람이 선택의 여지가 없이 반드시 어떤 일을 해야만 하는 상황이었기 때문에 행위를 하게 되었다면, 그는 그 행위에 책임이 (거의) 없을 것이다. 예를 들어 엄격한 군대사회에서 상관의 총을 쏘라고 지시했다면, 총을 쏜 부하는 그 일에 책임을 거의 지지 않을 것이다. 그 행위는 자유의지에 의한 행위가 아니라 강제에 의한 행위이다. 그렇기 때문에, 만약 자유의지가 존재하지 않는다면 우리의 모든 행위는 자유의지에 의한 것이 아니라 '어쩔 수 없는 행위'가 될 것이다. 그래서 진정한 책임은 가질 수 없게 되고, 아무리 커다란 범죄를 저질렀다고 할지라도 그에게 비난을 하는 것도 부당할 것이다. 법과 우리의 일반적인 감정은 자유의지에 의한 범죄행위에 대해 비난을 하고 큰 책임을 지우지만, 그 행위가 만약 자유의지에 의한 것이 아니라면 비난을 하기 어렵고 형벌의 크기도 줄어든다.

강한 결정론자들은 자유의지가 없기 때문에 우리에게 '책임'과 같은 것도 없다고 생각한다. 샘 해리스(Sam Harris)는 어떠한 범죄자라도 '교화'나 '감금'의 대상은 될지언정, '보복'이나 '비난'의 대상이 되어서는 안된다고 주장한다. 왜냐하면 그는 단지 '불운한 사람'이기 때문이다([24] p.53). 그러나 강한 결정론자의 주장은 우리의 일반적 감정과 잘 맞지 않기 때문에 양립가능론자들은 '다르게 할 수 없었다고' 할지라도 책임은 있을 수 있다는 것을 주장한다. 해리 프랭크퍼트(Harry Frankfurt)의 반례에서는 실제로 다르게 할 수 없는 상황이라고 할지라도, 자신의 의지가 그 행위를 선호하는 것은 가능하고, 그 경우에 책임이 발생하는 상황을 묘사한다[25]. 하지만 그것을 선호하는 의지 자체가 다르게 할 수 있다는 것을 뜻한다는 반론에 직면할 것이다.

자유의지론을 주장하는 학자들은 물론 자유의지에 의해 행동이 발생할 수 있으므로 그 경우에 책임을 가진다고 주장할 것이다. 비양립가능론자들(자유의지론자, 결정론자)은 '책임'이 자유의지의 필요조건이라고 생각하는 경향이 있다. 자유의지론 중에 가장 현대적인 모델을 주장한 로버트 케인(Robert Kane)은 *The Significance of*

Free Will 에서 비양립가능론자들은 궁극적 책임(UR: Ultimate Responsibility)을 자유의지의 필수요소로 포함시킨다고 보았다([15] p.59). 그래서 그에 따르면, 자유의지는 다른 것이 가능함(AP: Alternative Possibilities)과 함께 궁극적 책임(UR)이 있음을 주장하는 것이고, 그는 그 두 가지가 실현가능한 자유의지론 모델을 설명한다. 케인은 여러 가지 대안적 욕구들이 양자 불확정성과 카오스 증폭에 의해 비결정적으로 선택된다고 주장한다. 그것은 중요한 행위 결정의 순간에서 '자기 형성 행동(SFA: Self-Forming Action)'이 되므로 이 행위는 행위자의 궁극적인 책임이 된다. 이러한 행위는 인과적으로 다른 사건의 원인이 될 것이고, 우리는 어떠한 사건이나 행위의 궁극적 책임을 알기 위해 과거에 발생했던 '자기 형성 행동(SFA)'으로 거슬러 올라갈 수 있다(backward to their source)([15] p.79). 그리고 비록 불확정성에 의해 선택되었다 할지라도, 그 선택지들은 모두 합리적일 수 있다. 그 선택지들은 이전의 경험과 사고에 영향을 받아 만들어진 것이며, 그것들은 모두 '다원적 합리성(plural rationality)'과 '통약불가능성(incommensurability)'을 가지기 때문이다([15] pp.112-113, pp.197-198).

자유의지에 의한 행위가 책임을 더욱 크게 발생시킨다는 것은 직관적인 것으로 보인다. 나는 자유의지가 책임을 발생시킨다는 점을 인정한다. 그것은 어쩌면 SFA와 같은 행동에서 비롯된 것일지도 모른다. 하지만 우리는 어떤 사람에게 책임을 물을 때 그러한 궁극적 책임의 시초를 일일이 찾아보지 않는다. 우리가 일상적으로 책임을 부여하는 원리가 어떠한 것인지를 생각해 볼 필요가 있다.

우리가 왜 자유의지로 인한 행위에 책임을 부여하는지를 생각해보자. 아마도 자유의지에 의한 행위는 세계 내에서 그 행위의 존재를 **증가시키기** 때문일 것이다. 자유의지가 없는 '어쩔 수 없는 행위'였다면, 그 행위가 존재하는 것은 결정되어 있는 것이나 마찬가지이다. 그러나 자유의지에 의한 행위라면, 그 행위는 **존재하지 않을 수 있었다**. 그래서 자유의지에 의한 행위는 그 존재의 양을 증가시킨다. 자유의지에 의해 도둑질을 했다면, 그것을 하지 않는 다른 선택에 비해 '도둑질'의 존재를 증가시키는 것이다. 그러므로 자유의지에 의한 비도덕적 행동은 자유의지가 없는 행동에 비해 더욱 나쁜 행동이다. 그리고 그러한 직관은 자유의지에 따라 '다른 것을 할 수 있음'이라는 우리의 상식에 기초하는 것이다. 자유의지에 의한 행위는 다른 행위를 포기하고 그것을 실현시켜서 그 존재를 증가시킨 것이기 때문

에 그 행위의 존재에 대한 책임을 지게 되는 것이다.

그러면 책임은 언제, 어떻게 발생할까? 강한 결정론자들은 ‘누구라도 범죄자의 그때 그 조건과 상황에 똑같이 처한다면 다르게 할 수 없을 것이다. 그런데 어떻게 그를 비난할 수 있겠는가?’라고 주장하기도 한다([24] p.4). 그러나 그러한 생각은 5절에서 설명한 ‘한 시점에서 모든 조건이 같을 때 다른 대안이 존재할 수 없다’라는 모두가 아는 상식에 불과하다.

그 사람에게 책임이 있는지를 따져보기 위해 어느 시점에 어떤 자유의지가 작동했는지를 따져봐야 할까? 자유의지는 초기의 고정된 조건의 시점(t_0)과 결과도 출된 시점(t_1) 사이에서 작동했을 것이다. 그런데 그 초기 시점이 언제부터인지도 불분명하다. 우리의 행동은 그 전의 수많은 자유의지로 인한 선택들에 영향을 받았을 수 있다. 예를 들어 음주운전을 하다가 사고를 냈다면, 사고를 일으킨 당시에는 자유의지가 없었는지 모르지만, 술을 마셨을 때나 운전대를 잡았을 때, 어쩌면 그 전에 ‘나는 술을 마시고 운전을 해도 괜찮을 거야.’ 라는 생각을 했을 때 자유의지가 있었을 것이다.

그래서 자유의지가 작동해서 책임을 발생시키는 특정한 시점을 찾는 것은 힘든 일이 될 것이다. 하지만 우리는 그것을 깊이 따져보지 않고 대충의 상황만 보더라도 그가 책임이 있는지 없는지를 감지할 수 있다. 나는 그 원리가 다음과 같을 것이라고 생각한다.: <우리는 그가 자유의지를 가지는 개체인지를 확인하고, 그것이 확인되면 그의 행동에 특별한 제약조건이 없는 경우에 내정값(default value)처럼 자유의지에 의한 것으로 가정한다.> 그래서 만약 인간이 어떤 행위를 저질렀을 때 특별한 제약조건이 없다면, 그의 자유의지에 따른 것으로 보고 책임을 부여하게 된다. 우리는 명시적이지는 않지만 어떠한 대상이 자유의지가 있는지 없는지를 판단할 수 있는 기준을 가지고 있다. 자동차 운전자가 핸들을 돌려 가로수를 들이받았다면, 그것은 누구의 책임일까? 자동차의 책임도 아니고(자동차에 이상이 없다면) 가로수의 책임도 아니다. 운전자의 책임이다. 왜냐하면 자유의지를 가진 개체는 운전자이기 때문이다. 다만 그가 어떤 외부적 강압에 의해서 핸들을 돌리게 된 것이라면 그것은 특별한 제약조건이 되고, 그의 책임은 대폭 줄어들 것이다.

사이코패스의 예를 생각해보자. 연쇄살인을 저지른 사이코패스는 뇌의 기능이 다른 사람들과 조금 다를지 모른다. 하지만 뇌의 기능과 욕구의 지향성이 다르다

고 해서 우리가 그를 비난하지 않는 것은 아니다. 우리가 사이코패스에게 범죄의 책임을 부여할 수 있는 것은 그가 자유의지를 가지는 개체임을 확인했기 때문이다. 사이코패스가 자유의지를 가진다는 것을 확인하게 되면, 우리는 그에게 행위의 책임을 부여한다. 우리가 책임을 부여하지 않는 대상은 정신박약이나 금치산자와 같이 자유의지기능이 없는 것으로 보이는 사람들뿐이다. 우리의 그러한 습성은 자유의지의 특성에 따른 자연스럽고도 올바른 것이다.

마치며: 결정론의 환상

결정론자들은 자유의지가 환상이라고 주장한다. 우리에게 자유의지가 있다는 것은 이원론적인 영혼 같은 것을 가정하는 것처럼 보이고, 우리는 물리계에 구속될 수 밖에 없기 때문에 자유의지가 실제로는 없다는 것이다. 그러나 나는 이원론을 거부하고 모든 것이 물리적인 것으로 환원되는 것을 가정하였다. 그래도 자유의지가 존재한다는 생각에는 모순이 일어나지 않는다. 그리고 결정론을 지지하는 주장들은 모두 오류를 가지고 있었다.

자유의지는 기능류어이다. 우리는 상식적으로 자유의지가 어떠한 것인지 알고 있으며, 자유의지를 가진 대상과 그렇지 않은 대상을 구분할 수 있다. 심적 상태의 기능주의를 주장한 데이비드 루이스(David Lewis)는 우리의 상식적 심리학(common-sense psychology) 또는 통속 심리학(folk psychology)이 진정한 가치가 있으며, 그러한 개념은 실제로 물리적으로 환원되어 존재한다고 주장했다[5]. 그가 주장한 기능주의는 자유의지에 대한 설명으로 매우 적합해 보인다. 우리의 상식적 심리학은 자유의지가 실제로 존재한다는 것이다. 그것은 순전히 경험과 기능을 통해 확인 가능한 것이므로, 엄밀한 미시적 조사를 통한 **필연적인** 불가능성이 발견되기 전까지 우리는 그 존재를 믿어야 한다(나는 그 불가능성이 발견될 것이라는 데 회의적이다).

결정론자들은 자신이 철저히 과학에 기반하고 있다고 주장하지만, 나는 그러한 주장에 의심을 품고 있다. 결정론은 ‘운명론(fatalism)’과 본질적으로 동일하다. 운명론은 동서양의 각종 종교들과 관련이 깊다. 서양에서는 전지전능한 신이 모든 것

을 알고 있음(divine foreknowledge)이나 예정론의 관점, 점성술에서 나타날 수 있으며, 동양에서는 힌두교와 불교의 카르마(Karma: 업보)나 사주팔자 등이 운명론의 유혹에 빠지게 만든다. 이러한 신비스러운 주장들이 우리가 결정론을 믿도록 부채질 하고 있는 것이다. 그러나 결정론의 주장은 모두 근거가 적절하지 않으며, 우리의 상식과 어긋나고, 우리의 세계를 설명하는데 ‘불필요한 추가적 가설’에 불과하다. 그러한 가설은 환상일 가능성이 높다. 오컴의 면도날에 따라, 결정론은 제거되어야 한다.

참고문헌

- [1] Chisholm, Roderick (1964, reprint 2003), Human freedom and the self, In Gary Watson(ed), *Free Will*, Oxford University Press.
- [2] Frankfurt, Harry G. (1971), Freedom of the will and the concept of a person, *The Journal of Philosophy*, 68, 5-20.
- [3] Turing, Alan M. (1950), Computing machinery and intelligence. *Mind*, 59, 433-460.
- [4] Putnam, H. (1967, reprint 1980), The nature of mental states, In Block, N.(ed) *Readings in the Philosophy of Psychology*, Volume 1, Harvard University Press.
- [5] Lewis, David (1972), Psychophysical and theoretical identifications, *Australasian Journal of Philosophy*, 50, 249-258.
- [6] 김재권 저, 하중호 역 (2007), 물리주의, 아카넷(원서: *Physicalism, or Something Near Enough*, Princeton University Press, 2005.).
- [7] 김영정 (1996), 심리철학과 인지과학, 철학과 현실사.
- [8] Kripke, Saul A. (1972, reprint 1980), *Naming and Necessity*, Harvard University Press.
- [9] Jackson, Frank (1994), Finding the mind in the natural world, In Casati, R., Smith, B., White G.(ed), *Philosophy and the Cognitive Sciences*, Holder-Pichler-Tempsky.
- [10] Block, N. and Stalnaker, R. (1999), Conceptual analysis, dualism, and the explanatory gap, *Philosophical Review*, 108, 1-46.
- [11] McLaughlin, B. (2007), On the limits of a priori physicalism, In McLaughlin(ed),

Contemporary debates in philosophy of mind, Blackwell, 200-223.

- [12] Searle, John (1980), Minds, brains, and programs, *Behaviorism and Brain Sciences*, 3, 417-457.
- [13] Pinker, Steven (1994), *The Language Instinct*, William Morrow and Company.
- [14] Dennett, Daniel (2003), *Freedom evolves*, Viking.
- [15] Kane, Robert (1996), *The Significance of Free Will*, Oxford University Press.
- [16] Dennett, Daniel (1984), *Elbow Room: The Variety of Free Will Worth Wanting*, The MIT Press.
- [17] van Inwagen, P. (1975), Incompatibility of free will and determinism, *Philosophical Studies*, 27, 185-199.
- [18] Rosenblum, B. and Kuttner, F. (2011), *Quantum Enigma*, Oxford University Press.
- [19] Libert, Benjamin (2002), Do we have free will?, In Robert H. Kane(ed), *The Oxford Handbook of Free Will*, Oxford University Press.
- [20] Wilson, Timothy (2002), *Strangers to Ourselves*, The Belknap Press of Harvard University Press.
- [21] Klein, Gary (1998), *Sources of Power*, The MIT Press.
- [22] Dijksterhuis, A. et al., (2009). Predicting soccer matches after unconscious and conscious thought as a function of expertise, *Psychological Science*, 20, 1381-1387.
- [23] Kahneman, Daniel (2011), *Thinking, Fast and Slow*, Farrar, Straus and Giroux.
- [24] Harris, Sam (2012), *Free will*, Free Press.
- [25] Frankfurt, Harry G. (1969), Alternative possibilities and moral responsibilities, *The Journal of Philosophy*, 66, 828-839.

1차 원고접수 : 2013. 07. 17.

2차 원고접수 : 2013. 09. 12.

최종게재승인 : 2013. 09. 21.

(Abstract)

The Illusion of Determinism and Free Will as a Functional Kinds Term

Kiryong Mo

The doctoral program in Cognitive Science, Yonsei University

Determinism and compatibilism have thought that libertarianism contradicts the scientific view. This paper demonstrates that the hypothesis that free will exists does not cause any problems and contradictions, while maintaining pro-scientific view. Rather, determinism lacks scientific and demonstrative bases. At first, I reject dualistic self by assuming identity theory of the self and physical body & brain. Free will is a functional concept or property that can have multiple realization. Therefore we can recognize whether aliens or robots have free will through observation on their acts. 'Free will' is a functional kinds term like 'digestion'. As we know we have digestive function, we can know that we have free will even before knowing the necessity of its existence by understanding micro structure of human free will. The judgement that an individual has free will function depends on whether it has particular actions. Those actions are controlling and restraining actions in context, predicting future, making elaborate conditional sentences, etc. Real being of free will is the physical occupant of the function that makes those actions. Determinism is based on various kinds of misconceptions and misunderstandings. Determinists often confuse necessity of physical law and determinism. Another decisive misunderstanding is that free will postulates same conditions and different results from the same point of time. However, free will postulates that the two points of time are different. Because determinism is unnecessary redundant hypothesis influenced by religions, according to Ockham's Razor, determinism should be eliminated.

Key words : free will, determinism, compatibilism, functionalism, functional kinds term, responsibility