논리연구 21-3(2018) pp. 373-393

# 자유의지에 대한 조작논증과 근원-양립가능주의의 대응* **

김 성 수

【국문요약】 조작논증은 근원-자유와 결정론이 양립가능하지 않다는 것을 보이고자 한다. 이 논증에 따르면, (1) 일정한 방식으로 조작된 행위는 직관적으로 자유행위가 아니며 (2) 일정한 방식으로 조작된 행위와 결정론적으로 발생한 일상적 행위 사이에 중요한 차이가 없다. 이로부터 결정론적으로 발생한 일상적 행위는, 그것이 양립가능주의자들이 제시한 자유의 조건을 만족하더라도, 자유행위가 아니라는 것이 도출된다. 양립가능주의자들의 대표적인 대응인 '강경한 대응'은 (1)을 부정한다. 그런데 '중립적 입장'에서 볼 때 이 대응은 불충분한 것으로 밝혀진다. 그리고 (1)은 일종의 심리적 오류를 저지른다는 최근의 제안도 이를 극복하지 못한다고 주장한다. 반면 '온건한 대응'은 (2)를 부정한다. 이 대응을 지지하기 위해 먼저 조작적 개입과 단순한 결정 사이의 차이를 밝힌다. 그리고 온건한 대응과 조작논증이 특정한 상황에 대해 서로 다른 예측을 한다는 점에 주목할 때 이 대응을 지지할 이유가 있다고 주장한다.

【주요어】 자유의지, 결정론, 양립가능주의, 조작논증, 강경한 대응, 온건한 대응

---

투고일: 2018.9.10 심사 및 수정 완료일: 2018.10.10. 게재확정일: 2018.10.10.
* 이 논문은 2017년도 서울시립대학교 연구년교수 연구비에 의하여 연구되었음.
** 이 논문의 심사를 맡아 유익한 비판과 도움말을 해주신 익명의 심사위원님들께 감사드린다.

## 1. 들어가는 말

자유의지와 결정론에 대한 양립가능주의는 행위의 자유와 도덕적 책임이 성립하기 위해 결정론이 거짓일 필요가 없다는 입장이다. 달리 할 수 있음으로서의 자유와 결정론이 양립가능하다고 주장하는 전통적인 형태의 양립가능주의[1]와 달리, '근원-양립가능주의'는 행위의 시발점 혹은 근원이 어떤 의미에서 행위자에게 있는 것으로서의 자유와 결정론이 양립가능하다고 주장한다.[2] 이 입장에 따르면, 근원-자유는 행위자가 실제와 달리 할 수 있을 것을 요구하는 대신, 합리적 이유에 일정한 방식으로 반응하거나[3] 행위의 원인인 일차적 욕구가 행위자의 이차적 욕구와 잘 맞는 것[4]으로, 실제 인과 역사에서 이런 근원-자유를 만족하는 것과 그런 인과 역사를 갖도록 결정된 것이 양립가능하다.

조작논증(manipulation arguments)은 근원-자유와 결정론이 양립가능하지 않다는 것을 보이고자 한다.[5] 이 논증은 기본적으로 다음과 같은 구조를 갖는다.

---

1) 대표적인 예로 Hume (1748)과 Ayer (1954) 등이 있다.

2) 달리할 수 있는 자유와 근원-자유의 구분을 피셔와 라비자(Fischer & Ravizza, 1998)는 다음과 같은 비유를 들어 설명한다. 운전자가 좌회전을 하고자 하였고 실제로 그렇게 했다고 해보자. 주어진 상황에서 운전자가 좌회전을 할 수도 우회전을 할 수도 있다면, 이는 실제와 달리 할 수 있는 자유가 있는 경우이다. 반면 자동차의 핸들이 오른쪽으로 돌아갈 수 없는 상황에서 운전자가 자신이 원하는 대로 좌회전하였다면, 달리할 수 있는 자유는 없으나 좌회전한 행위의 근원이 운전자에 있는 근원-자유가 성립하는 경우이다.

3) Fisher & Ravizza (1998), Sartorio (2016).

4) Frankfurt (1971).

5) 잘 알려진 조작논증으로는 페러붐(Pereboom 1995, 2001)의 '네 가지-경우 논증(four-case argument)'과 밀(Mele 2006)의 '수정란 논증(zygote argument)'이 있다.

(1) 일정한 방식으로 조작된 행위는 도덕적 책임을 물을 수 있는 자유행위가 아니다.
(2) 일정한 방식으로 조작된 행위와 결정론적으로 발생한 일상적 행위 사이에 차이가 없다.
(3) 따라서 결정론적으로 발생한 일상적 행위 역시 자유행위가 아니다.

근원-양립가능주의(이하 '양립가능주의')는 결정론적으로 발생한 일상적 행위가 자유행위라고 주장하는데, 이 논증의 결론은 이를 부인한다. 조작논증이 옳다면, 양립가능주의를 거부하고 근원-양립불가능주의(이하 '양립불가능주의')를 받아들여야 한다.

이 논증에 대한 양립가능주의자의 반박은 두 가지로 나누어진다.

강경한 대응: 전제 (2)를 받아들이고 전제 (1)을 거부한다.
온건한 대응: 전제 (1)을 받아들이고 전제 (2)를 거부한다.[6)]

강경한 대응은 결정론적으로 발생한 일상적 행위가 도덕적 책임을 물을 수 있는 자유행위라는 전제로부터 출발한다. 그리고 전제 (2)가 주어졌을 때, 일정한 방식으로 조작된 행위 역시 자유행위라는 결론을 내린다. 조작논증과 강경한 대응 모두, 만약 일정한 방식으로 조작된 행위가 자유행위가 아니라면 결정론적으로 발생한 일상적 행위는 자유행위가 아니라는 주장을 받아들인다. 양립불가능주의자는 일정한 방식으로 조작된 행위가 자유행위가 아니므로 결정론적으로 발생한 일상적 행위 역시 자유행위가 아니라고 주장하는 반면, 양립가능주의자의 강경한 대응은 결정론적으로 발생한 일상

6) '강경한 대응 (hard-line reply)'과 '온건한 대응 (soft-line reply)'은 맥케냐 (McKenna 2008)가 사용한 표현이다.

적 행위가 자유행위이므로 일정한 방식으로 조작된 행위 역시 자유행위라고 주장한다. 조작논증이나 강경한 대응과 달리 온건한 대응은, 자유와 도덕적 책임의 측면에서 볼 때 결정론적으로 발생한 일상적 행위와 일정한 방식으로 조작된 행위 사이에 차이가 있다고 주장한다. 온건한 대응에 따르면, 일정한 방식으로 조작된 행위가 자유행위가 아니라는 것이 결정론적으로 발생한 일상적인 행위가 자유행위인지 여부를 함축하지 않는다.

이 글은 다음과 같이 구성된다. 2절에서는 조작논증의 한 사례인 '수정란 논증(zygote argument)'을 사용하여 조작논증을 설명한다. 3절에서는 조작논증에 대한 양립가능주의자들의 대표적인 반박인 강경한 대응을 다룬다. 특히 '중립적 입장'에서 강경한 대응의 문제점을 살펴보고, 심리적 효과에 근거하여 강경한 대응을 보강하려는 시도를 비판적으로 논의한다. 4절에서는 온건한 대응의 가능성을 주장한다. 이를 위해 기획에 따른 의도적 개입과 자연적 결정의 차이를 살펴보고 온건한 대응을 지지할 수 있는 이유를 제시한다. 그리고 이에 대한 양립불가능주의자의 반박을 비판적으로 논의한다.

## 2. 조작논증

일반적으로 말해 조작된 행위는 자유행위도 그것에 대한 도덕적 책임이 있는 행위도 아니다. 예를 들어 블랙은, 철수는 그의 존재를 모르지만, 철수가 탈세하길 원한다. 블랙은 만약 철수 스스로 탈세하기로 결정하면 개입하지 않지만, 만약 철수 스스로 탈세를 결정하지 않을 것 같으면 철수의 뇌 상태를 변화시키는 신호를 보내 철수로 하여금 탈세하도록 만든다. 실제로 철수는 탈세 여부를 놓고 고민하다 그러지 않기로 결정하려하고 이를 알아챈 블랙이 개입하여 그 결과 철수는 탈세하기로 한다. (블랙은 상대가 어떤 결

정을 하려 하는지 알아내고 상대가 자신의 의도대로 행동하도록 조작하는 데 매우 뛰어난 능력을 갖고 있다.) 이 사례에 대한 우리의 직관은, 조작의 결과로 발생한 철수의 결정은 자유행위가 아니며 따라서 철수는 이에 대한 도덕적 책임이 없다는 것이다. 이 사례에서 철수의 결정은 철수가 진정으로 원한 것 때문에 발생한 것이 아니며, (예컨대 도덕적 요구와 일치하는 욕구를 갖길 원하는) 철수의 이차 욕구와 잘 맞는 일차 욕구가 행위의 실제 인과 역사의 일부가 되지도 못하며, 이유에 반응하는 능력을 우회하여 발생한 것이다. 양립가능주의에 따르면 철수의 행위는 자유의 조건을 만족하지 못하므로 자유행위가 아니며, 이는 우리의 직관과 부합한다.

반면 조작논증에서 사용되는 '일정한 방식으로 조작된' 행위는 (a) 직관적으로 자유행위가 아니면서 (b) 양립가능주의자들이 제시한 근원-자유의 조건을 만족하는 그런 행위이다. 이를 위해 조작논증은 조작의 시점을 행위자가 컨트롤할 수 있는 범위를 완전히 벗어난 시점으로 잡고, 이후 결정론적으로 전개된 세계에서 그 조작을 결정론적인 원인의 일부로 갖는 행위가 양립가능주의자들이 요구하는 자유의 조건을 만족하게 만든다. 이 글에서는 다음과 같은 수정란 논증의 버전을 사용하여 조작논증을 논의하고자 한다.

> *다이애나*: 메리는 임신을 위해 시험관 시술을 하려한다. 다이애나 여신이 개입하여 시험관에서 형성되고 있는 수정란에 유전자 변형이 생기게 하는데, 다이애나는 그 결과로 선천적인 유전적 성향을 가진 아이 어니(Ernie)가 잉태될 것을 안다. 다이애나는, 부분적으로 이 성향으로 인해 삼십년 후 어니가 유산을 물려받기 위해 삼촌을 살해할거라는 것 역시 안다. 세계의 상태와 결정론적 법칙을 완전히 알고 자신이 개입한 이후 세계가 정확히 어떤 식으로 전개될지 예측할 수 있는 다이애나는 어니가 삼십년 후 자기 삼촌을 살해하길 원하기 때문에 그에 맞는 유전자 변형이 생기게 한다. 그럼에도 삼십년 후 자기 삼촌을 살해할 때 어니는

> (어떤 강제나 강요 없이, 이유에 반응하여, 자신의 행동을 이끄는 욕구를 자신의 것으로 여기고, 관련 메커니즘에 대한 자신의 책임을 받아들이는 등) 자유에 대한 표준적인 양립가능주의의 모든 조건을 만족한다.[7]

조작논증의 전제 (1)에 따르면, 어니의 행위는 그것에 대해 도덕적 책임을 물을 수 있는 자유 행위가 아니다. 조작논증은 이 주장이 직관적으로 분명하다는 데 근거하고 있으며, 실제로 이는 동의할 수 있어 보이는 주장이다.

결정론에 따르면, 주어진 과거의 세계 전체의 상태와 자연 법칙은 단 하나의 유일한 미래를 함축한다. 양립가능주의자들은 도덕적 책임을 위해 요구되는 자유를 이른바 '의지의 자유'가 아니라 행위의 자유로 보고, 자신들이 제시한 자유의 조건을 만족하는 결정론적으로 발생한 행위는 자유행위라고 주장한다. 조작논증의 전제 (2)에 따르면, 어니의 행위와 결정론적으로 발생한 일상적인 행위, 예를 들어 어니와 유사한 역사를 갖지만 어떤 개입도 없이 자연적으로 생겨난 어니*의 행위 사이에 중대한 차이가 없다. 어니의 경우엔 다이애나가 개입하고 어니*의 경우엔 그렇지 않지만, 이 차이가 발생하는 수정란이 형성되는 시점은 인격체로서의 어니와 어니*가 존재하기도 전이다. 또한 수정란이 그것의 특성을 갖게 하는 원인에 대해서 그리고 그 결과 생겨난 수정란의 특성에 대해 어니와 어니* 모두 어찌할 수 없다. 두 경우 모두에서 유전자 변형은 이후 어니와 어니*의 행위의 결정론적 원인이며, 그 행위는 이들의 "컨트롤을 넘어선 요인으로 거슬러 올라가는 결정론적인 인과 과정으로부터 발생한다."[8] 이런 점을 고려할 때 어니의 행위를 결정짓는 요인들에 다이애나의 개입이 포함되어있고 어니*의 경우는 자연적

---

7) Sartorio (2016), p. 159.

8) Pereboom (1995), p. 25.

요인들만 포함되어 있다는 것이 이들의 자유와 책임에 큰 차이를 만들지 않는다고 조작논증은 주장한다. 어니의 행위가 자유행위가 아니라는 것이 주어졌을 때, 양립가능주의자들의 자유의 조건을 만족하는 행위와 어니의 행위가 다를 바 없다면 양립가능주의는 성립할 수 없다.

## 3. 강경한 대응과 그 한계

조작논증에 대한 양립가능주의자들의 대표적 반박은 조작논증의 전제 (1)을 거부하는 강경한 대응이다.[9] 강경한 대응은 조작논증의 결론에 반하는 직관으로부터 시작해서, 결정론적으로 발생한 일상적 행위와 일정한 방식으로 조작된 행위 사이에 차이가 없다는 조작논증의 전제 (2)를 받아들이고, 이로부터 조작논증의 전제 (1)이 거짓이라는 결론을 내놓는다. 강경한 대응을 다이애나 사례에 적용하면 다음과 같다.

(4) (양립가능주의자들이 내놓은 자유의 조건을 만족하는) 결정론적으로 발생한 일상적 행위는 도덕적 책임을 물을 수 있는 자유행위이다.
(5) 결정론적으로 발생한 일상적 행위와 어니의 행위 사이에 중대한 차이가 없다.
(6) 따라서, 어니의 행위 역시 도덕적 책임을 물을 수 있는 자유행위이다.

일상적 행위들에 대해, 예를 들어 자연적인 역사를 갖는 어니*가 어떤 강압이나 강제 없이 그리고 심리적 집착이나 조작이 없는 상

9) 그 예로 Fisher (2011), McKenna (2008), Sartorio (2016) 등이 있다.

태에서 스스로 여러 이유를 고려하여 한 행위에 대해, 우리는 혹은 적어도 양립가능주의자들은 그 행위가 자유행위이며 도덕적 책임을 물을 수 있는 행위라는 직관을 갖는다. 그리고 프랑크푸르트의 사례에서처럼[10] 그 행위가 실제와 달리 할 수 없는 경우에도 행위자에게 도덕적 책임을 물을 수 있다는 점을 고려할 때 이 직관은 달리할 수 있음을 요구하지도 않는다. 일반적으로 양립가능주의자들은 행위자의 컨트롤을 완전히 넘어선 시점에서 벌어지는 일과 관련된 어떤 조건을 자유의 조건으로 포함하지 않는다. 그렇다면 양립가능주의자들은 어니와 어니* 사이에 자유와 책임의 측면에서 중대한 차이가 없다고 주장할 수 있다. 이로부터 조작논증이 전제하는 바와 달리 어니의 행위는 자유행위라는 결론이 나온다.

어니의 행위가 자유행위가 아니라는 직관으로부터 출발해서 결정론적으로 발생한 일상적 행위 역시 자유행위가 아니라는 결론에 도달하는 조작논증과 달리, 강경한 대응은 결정론적으로 발생한 일상적 행위가 자유행위라는 직관으로부터 출발해서 어니의 행위 역시 자유행위라는 결론에 도달한다. 만약 강경한 대응의 근거가 조작논증의 근거와 동일한 정도의 직관적 설득력을 갖는다면 조작논증은 양립가능주의가 실패함을 보여주지 못한다. 조작논증과 강경한 대응이 서로 비기게 되면, 더 강력한 반대논증이 제시되지 않는 한 양립가능주의자들은 자신들의 입장을 유지할 수 있기 때문이다.

조작논증과 강경한 대응은 서로 상반된 직관에 근거하고 있으며, 각 입장의 출발점이 되는 직관은 개별적으로 볼 때 설득력이 있다. 그렇지만 전제 (2)와 (5)를 받아들일 때, 두 입장이 주장하는 서로 상반되는 직관을 같은 정도의 강도로 동시에 유지할 수 없다. 양립가능주의자들과 양립불가능주의자들이 서로에 대해 선결문제요구의 오류를 저지른다고 생각할지도 모르지만, 아직은 완전히 입장을 정

10) Frankfurt (1969).

하지 못하고 두 논증을 합리적으로 비교하고 평가하고자 하는 입장에서 볼 때 이는 성급한 반응인 것 같다.[11] 이 입장을, 페러붐을 따라, '중립적 입장'이라 부르자. 페러붐은 중립적 입장에서 볼 때, 결정론적으로 발생한 일상적 행위가 자유행위라는 전제 (4)를 지지하는 직관이 다이애나 사례 같은 조작논증의 전제 (1)를 지지하는 직관만큼 강하지 않다고 주장한다. 페러붐은 자유의지와 결정론의 문제를 이해하지만 아직 분명한 입장을 갖기 이전인 "초기상태에서 인식적으로 합리적인 것은 일상적인 결정론적 사례에서 행위자가 도덕적 책임이 있다고 믿지 않고 또한 도덕적 책임이 없다고도 믿지 않지만 이들 믿음 중 하나를 합리적이게끔 명확히 하는 고려사항들에 열려 있는 것"[12]이라는 점을 지적한다. 그리고 중립적 입장에서 볼 때, 조작논증은 "일상적인 결정된 행위자는 도덕적으로 책임이 없다는 믿음을 합리적이게 만드는"[13] 역할을 할 것이라 주장한다.

페러붐의 주장을 이해하기 위해, 어떤 강압이나 강제 없이 그리고 심리적 집착이나 조작이 없는 상태에서 스스로의 이유로 하게 된 일상적 행위에 대한 다음의 두 주장을 구분해보자.

(i) 일상적인 행위는 도덕적 책임을 물을 수 있는 자유행위이다.
(ii) 결정론적으로 발생한 일상적인 행위는 도덕적 책임을 물을 수

---

11) 양립가능주의자들은 자신들의 직관이 우리의 삶에서 거부할 수 없는 믿음이라고 주장할 수 있다. 반면 양립불가능주의는 다음과 같은 스피노자 식의 주장을 할 수 있다. "사람들은 자신이 자유롭다고 생각하는데, 그 이유는 자신의 의지와 욕구를 의식하지만 원하고 의지하려는 맘이 들게 하는 원인을 꿈에도 생각하지 못하며 그것에 무지하기 때문이다." (스피노자의 『윤리학』, Pereboom (2008), p. 161에서 재인용)

12) Pereboom (2008), p. 162.

13) Ibid.

있는 자유행위이다.

주장 (i)은 우리가 별 의심 없이 받아들이는 믿음이다. 우리는 사람들의 행위를 도덕적 측면에서 평가하고 행위자가 비난받을만하거나 칭찬받을만하다고 판단한다. 사실 주장 (i)은 '상식적 세계관'의 일부라고 할 수 있다. 주목할 것은, 주장 (i)을 받아들일 때 우리는 이 믿음을 결정론과 연관시켜 생각하지 않는다는 점이다.[14] 한편 주장 (ii)는, 주어진 과거와 자연법칙이 주어질 때 단 하나의 미래만이 가능하다는 결정론과 자유행위의 관계에 대해 생각할 것을 요구한다.[15] 그런데 '자유의지와 결정론의 문제'를 이해하고 결정론적 세계에서 행위의 자유와 도덕적 책임에 대해 생각할 때 우리는 (ii)에 대한 분명한 어떤 직관을 갖지 못하는 것 같다. 그리고 설령 직관을 갖더라도 이미 확고한(committed) 어떤 입장을 갖고 있는 경우가 아니라면 그 직관은 이러저러한 고려사항에 따라 쉽게 흔들릴 수 있는 것 같다. 이런 중립적 입장에서 볼 때, 강경한 대응의 전제 (4)를 뒷받침하는 직관은 조작논증의 전제 (1)을 지지하는 직관만큼 강력한 것 같지 않고, 그렇다면 강경한 대응은 추가적인 지지가 없는 한 조작논증에 대한 효과적인 대응이 되기에 부족해 보인다.

이러한 강경한 대응의 약점을 보완하기 위해 사토리오는 다이애나 사례에서 어니의 행위가 자유행위가 아니라는 직관은 심리적 오류를 범한다고 주장한다. 사토리오는 다이애나 사례와 다음의 사례를 비교할 것을 제안한다.

---

14) Pereboom (2008), p. 161.

15) 주장 (i)을 생각할 때 우리는 결정론을 고려하지 않으므로 일상적 행위가 자유행위인지 여부는 결정론과 아무 상관없다는 주장에 대한 비판으로는 van Inwagen (1983) 4장을 참조할 것.

> *번개*: 이 경우에 다이애나 여신은 없다. 대신 시험관에서 수정란 형성이 진행되는 중에 번개가 치고 그 결과 다이애나 사례에서 다이애나가 개입한 것과 동일한 유전자 변형이 생겨난다. 다른 모든 것은 다이애나 사례와 동일하다.[16]

다이애나 사례에서 어니의 행위는 다이애나의 기획(design)에 의한 의도적 개입의 결과로 발생한 반면, 번개 사례에서 어니의 행위는 번개가 내리친 결과로 발생한다. 서로 유사하지만 의도적 개입과 '자연적 개입'의 차이를 갖는 이 두 사례에 대해 사토리오는 다음과 같이 주장한다. "[번개] 사례를 생각해볼 때 내 직관은 급격히 달라진다. 특히 이 경우엔 어니의 행위가 자유롭지 않다고 여길 맘이 들지 않는데 내 추측으론 다른 이들의 반응도 비슷할 것이다."[17]

사토리오가 지적하듯, 의도적 개입이 있는 경우와 그렇지 않은 경우를 비교할 때 자유와 도덕적 책임 여부에 대한 우리의 직관적 판단에 차이가 있는 것 같다. 사토리오는 이러한 직관의 변화가 심리적인 것이라고 주장한다.

> *어째서* 우리의 직관은 이런 식으로 작동하는가? …… 내가 보기엔 '책임의 희석' 같은 어떤 심리적 효과가 작동하는 것 같다. 잘 알려진 바와 같이 싱어(Singer)와 엉거(Unger)는 빈민층에 대한 도덕적 의무가 우리들이 생각하는 것보다 훨씬 더 엄격하다는 자신들의 논증에서 이 효과에 주목했다. 이들이 설명해버리고자 한 현상은 바로 기아 구제나 다른 인도주의적 명분에 기여하는 것이 도덕적 의무는 아니고 그 범위를 넘어선(supererogatory) 것이라는 생각이다. 이들은 다른 사람들 *역시* 돕지 않는다는 (따라서 도덕적 의무를 다하지 못하고 있다는) 사실이 우리가 스스로를 덜 책임이 있거나 실제로 그런 것보다 덜 잘못하고 있는 것으로 보게

16) Sartorio (2016), pp. 160-161.

17) Ibid, p. 161.

> 하는 것이라고 함으로써 이 현상을 해명한다. …… 이 현상의 또 다른 가능한 예로는 총살형 집행부대 같은 과잉인과(overdetermination)의 사례가 있다. 일반적으로 부대의 개별 대원은 자기 혼자 집행할 때보다 사형수의 죽음에 대해 책임을 덜 느끼리라는 것이 집행부대가 존재하는 이유의 일부이다. 그러나 많은 이들이 이는 단지 그래 보일 뿐이라 생각하는데, (적어도 모든 총알이 거의 동시에 사형수에게 도달하였고 각각의 총알이 죽음을 가져오는 데 충분하다고 가정할 때) 다른 대원들의 존재가 실제로 개별 대원이 그 죽음에 덜 책임이 있게 만들지 않기 때문이다. 이 경우에도 이는 단지 일인칭 관점뿐만 아니라 삼인칭 관점에서의 효과이기도 하다: 밖에서 볼 때에도 우리는, 실제로는 그렇지 않다고 할 만한데도, 각각의 대원이 덜 책임이 있는 것으로 보는 경향이 있다.[18]

책임희석 효과는 동일한 사건에 대해 책임이 있을 수 있는 한 명 이상의 행위자가 있을 때 발생한다. 번개 사례에서는 책임이 있을 수 있는 행위자가 어니 한 명인데 비해, 다이애나 사례에서는 그런 행위자가 한 명 이상, 즉 다이애나와 어니가 존재한다. 어니의 행위는 다이애나의 특정한 기획(design)과 의도 때문에 발생하였으므로 다이애나는 의도의 실행과 그 결과에 대해 책임이 있을 수도 있는 책임의 후보자이며, 어니 역시 자신의 이유로 행동하였고 그에 대해 책임이 있을 수도 있는 책임의 후보자이다.

앞서 주장했듯 중립적인 입장에서 평가할 때, 강경한 대응보다는 조작논증을 받아들일 이유가 있어 보인다. 이제 다이애나 사례에서 작동하는 직관이 사토리오가 주장하듯 책임의 희석이라는 심리적 효과의 영향을 받는다면, 우리는 다이애나와 어니 사이에 도덕적 책임을 분산하게 되고 그 결과 어니가 실제 그런 것보다 덜 책임이 있다고 여길 것이다. 어니의 자유와 책임에 대한 우리의 판단이

---

18) Sartorio (2016), pp. 167-168.

이런 심리적 오류를 범한다면, 다이애나 사례에서 어니의 행위가 도덕적 책임을 물을 수 있는 자유행위가 아니라는 조작논증의 전제 (1)을 지지하는 직관이 약화될 것이다. 그리고 이렇게 약화된 직관이 강경한 대응에 대해 중립적 입장이 제기한 문제를 상쇄한다면, 조작논증과 강경한 대응 사이의 균형이 회복되고 그 결과 두 입장이 서로 비기게 될 수 있다.

그러나 강경한 대응에 비판적인 중립적 입장에 대해 이 주장이 얼마나 큰 설득력을 발휘할지 분명하지 않다. 사토리오가 든 책임희석 효과의 경우 중에서 다이애나의 책임과 어니의 책임에 자연스럽게 적용될 것 같아 보이는 집행부대의 경우에 대해 생각해보자. 집행부대가 두 명으로 구성되었고 그 중 한 명만 집행한 경우, 실제로 집행한 부대원에게 책임이 있음이 분명하다. 이제 다이애나 사례에서 다이애나를 제거해보자. 어니는 자신의 행위에 대해 책임이 있는가? 강경한 대응의 문제를 보완하기 위해서는 책임분산 효과가 발생하지 않는 상황에서 어니의 책임에 대한 직관이 조작논증의 전제 (1)을 지지하는 직관만큼 분명해야 한다. 양립가능주의의 입장에서 볼 때, 어니는 양립가능주의자들이 제시하는 자유의 조건을 만족하므로 어니는 자신의 행위에 대해 책임이 있다. 반면 중립적 입장에서 볼 때는 상황이 다르다. 앞서의 일상적 행위에 대한 주장 (i)와 (ii) 중에서 (i)을 생각한다면 어니가 자신의 행위에 대해 책임이 있음이 분명하다. 하지만 논의의 대상은 주장 (ii)이며 어니의 책임에 대한 직관이 중립적 입장이 제기한 문제를 극복할 수 있을 정도로 회복되지 않는 것 같다.

사토리오가 든 책임희석 효과의 경우들에서는, 비슷한 정도의 책임이 있을 수 있는 행위자들이 대략 동시적으로 그리고 서로 독립성을 유지하면서 행동한다. 반면 다이애나의 행위는 어니의 행위에 도달하는 인과사슬 내에 있을 뿐만 아니라, 집행부대의 경우와는

달리 다이애나 사례에서는 어니를 제거하고 다이애나만 있는 상황에서 (어니의 행위에 대한) 다이내아의 책임에 대해 생각할 수 없다. 또한 다이애나 사례에 책임희석 효과가 적용된다면 다이애나의 책임의 정도에 대한 우리의 직관 역시 영향을 받아야 될 것 같다. 하지만 집행부대나 빈민 구제에 대한 도덕적 의무의 경우와는 달리, 다이애나의 책임에 대한 우리의 직관은 어떤 변화도 보이지 않는 것 같다. 이상의 논의가 옳다면, 사토리오가 주장하는 심리적 효과는 강경한 대응을 보강하기에 충분하지 않다.

## 4. 온건한 대응의 가능성

온건한 대응은 다이애나의 개입으로 인한 어니의 행위와 결정론적으로 발생한 일상적 행위 사이에 차이가 있다고 주장한다. 다이애나 사례에서 일상적인 결정론적 경우와 차이를 보이는 가장 두드러진 요소는 다이애나의 의도적 개입이다. 피셔는 원래의 수정란 논증[19]에 대한 강경한 대응을 위해, 다이애나 대신 어니의 자연적 부모인 존과 메리의 사례를 제시한다. 존과 메리는 자신들의 아이 어니가 삼십년 후에 특정한 행위를 하도록 만들 의도를 가지며, 그 의도로 인해 어니가 자신들이 의도한 바대로 행동하게 되리라는 믿음을 갖고 아이를 잉태하려 한다. 피셔는 "존과 메리의 의도와 이 의도가 (주어진 환경에서) 자신들이 원하는 미래에 대한 충분조건을 제공한다는 믿음은 삼십년 후의 문맥에서 어니의 도덕적 책임에 대한 직관적 근거와 전혀 상관없다"[20]고 지적하고, 이에 더해 "존과 메리를 다이애나로 대체할 수 있으며 이렇게 할 때 분명한 직

19) 밀(Mele 2006)의 수정란 논증에서 다이애나는 시험관 속에서 수정란이 형성되는 과정에 개입하는 대신 주변의 물질을 사용하여 수정란을 만든다.

20) Fisher (2011), p. 268.

관은 다이애나의 의도가 삼십년 후 어니의 도덕적 책임과 아무 상관없다"[21]는 것이라고 주장한다.

이 주장에 따르면 의도적 개입은 중요한 차이를 만들지 못한다. 하지만 존-메리 사례에서 의도가 어니의 도덕적 책임과 아무 상관없고 그리고 다이애나 사례와 존-메리 사례 모두에 어니의 행위에 대한 의도가 있다고 해도, 다이애나 사례에서도 의도가 어니의 도덕적 책임과 아무 상관없다는 것이 도출되지 않는다. 도덕적 책임을 위해서는 적어도 행위자의 정신상태가 문제의 결정이나 행위와 인과적으로 연결되어있어야 한다. 다이애나는 특정한 의도를 갖고 있을 뿐만 아니라, 의도의 성공에 적합한 기획에 따라 개입하며, 그것 때문에 의도를 성공적으로 실행한다. 반면 존과 메리는 특정한 의도를 갖긴 했지만, 그 의도를 성공적으로 실행할 능력이 없다. 이렇듯 다이애나의 의도는 어니의 행위에 대한 원인인 반면, 존과 메리의 의도나 믿음은 그렇지 않다. 따라서 존과 메리의 의도나 믿음은 그들에게 어니의 행위에 대한 책임이 있을 수 있게 하는 요인이 아니지만, 다이애나의 의도는 그럴 수 있는 요인이다.

일상적인 결정론적 경우와 비교할 때 다이애나 사례의 가장 두드러진 차이가 의도적 개입이라면, 그것은 어떤 특징을 갖는가? 다이애나의 의도적 개입은 그 의도를 성공적으로 달성할 수 있는 다이애나의 인식능력과 실천능력을 고려할 때, 다이애나로 하여금 어니가 어떤 행동을 할지에 대해 강한 컨트롤을 행사할 수 있게 한다. 수정란이 성장할 일정한 환경과 그 환경이 앞으로 어떻게 전개될지가 주어질 때, 다이애나는 어니가 삼십년 후에 일련의 행위를 하도록 할 수 있다. 실제로 그런 것처럼 어니가 자신의 삼촌을 살해하도록 개입할 수 있고, 어니가 살해를 뒤로 미루도록 할 수 있고, 살해대신 다른 차선책을 도모하도록 할 수 있고, 그 밖에 다른

21) Ibid, p. 269.

많은 일을 하도록 할 수 있다. 또한 수정란에 유전자 변형이 생기게 한 것 말고도, 어니 행위의 결정론적 원인이 될 수 있는 다른 환경적 요인에 개입할 수도 있다.

반면 '자연적인 방식으로' 어니의 행위를 결정하는 번개와 같은 경우는 이런 특징을 갖지 않는다. 번개 사례에서 실제로 번개가 쳤고 그 영향으로 수정란에 유전자 변형이 일어났다. 만약 실제와 달리 번개가 치지 않았다면 유전자 변형이 일어나지 않았을 것이고 그 결과 어니는 실제와 다른 행위를 했을 것이다. 그리고 만약 번개가 아니라 특정 화학물질에 노출되었더라면 실제와 다른 유전자 변형이 생겨서 또 다른 행위를 했을 수 있고 그 밖에 다이애나의 개입에 의해 발생할 수 있는 일련의 행위에 대응하는 여러 반사실적 가능성이 성립한다. 다이애나의 다양한 개입의 가능성과 마찬가지로 '자연적 개입'에 대해서도 유사한 반사실적 가능성이 성립한다면, 번개의 개입과 다이애나의 개입 사이에 중요한 차이가 없다고 생각할 수 있다. 하지만 다이애나의 개입과 번개의 개입 사이엔 중요한 차이가 있다. 다이애나의 의도와 어니의 행위 간에 성립하는 반사실적 가능성은 다이애나의 의도에 따라 어니의 행위가 어떻게 달라질 수 있는지를 보여주며, 이 반사실적 가능성은 다이애나가 어니의 행위에 대해 갖는 컨트롤 능력을 보여준다. 반면 번개 사례에서 성립하는 것은 *번개가* 어니의 행위에 대해 어떤 컨트롤을 행사할 능력을 갖고 있음을 보여주지 않는다. 번개 사례와 관련된 반사실적 가능성들은 단지 서로 큰 관련이 있을 필요가 없는 이러저러한 경우에 어떤 일이 벌어질지에 관한 것일 뿐이다.

이제 다이애나의 개입처럼 다른 행위자의 행위를 컨트롤하는 개입을 '조작적 개입'이라고 하자. 온건한 대응을 지지하는 한 방법은, 다이애나 사례에서 어니의 행위가 자유행위가 아닌 이유는 조작적 개입 때문이라고 주장하는 것이다. 온건한 대응과 조작논증

둘 다 다이애나 사례에서 어니의 행위는 자유행위가 아니라는 데 동의하지만, 조작논증은 어니의 행위가 자유행위가 아닌 이유가 조작적 개입 때문이라는 주장에는 동의하지 않는다. 이는 조작적 개입과 자연적으로 발생하는 결정론적 요인 사이에 차이가 있는지 여부에 대해 두 입장이 서로 동의하지 않는 것과 연관되어 있다. 흥미 있는 것은, 다이애나 사례와 다를 바 없지만 조작적 개입은 없는 번개 사례가 두 입장을 평가하는 데 사용될 수 있다는 점이다.

이와 관련하여 특히 온건한 대응과 조작논증은 번개 사례에 대해 서로 다른 예측을 한다는 점에 주목할 필요가 있다. 번개 사례는 다이애나 사례에서 조작적 개입을 자연적 개입으로만 바꾼 경우이다. 온건한 대응이 주장하듯 조작적 개입과 자연적 결정 사이에 차이가 있다면, 번개 사례에서 어니의 행위가 자유행위가 아니라는 직관은 다이애나 사례에서만큼 강하지 않을 것이라고 예측할 수 있다. 반대로 조작논증이 주장하듯 조작적 개입과 자연적 결정 사이에 차이가 없다면, 번개 사례에서 어니의 행위가 자유행위가 아니라는 직관은 다이애나 사례와 별로 다르지 않을 것이라 예측할 수 있다. 그런데 번개 사례에서 어니의 행위가 자유행위가 아니라는 직관은 다이애나 사례에서만큼 강하지 않아 보인다. 그렇다면 번개 사례에 대한 서로 다른 예측은 조작논증보다는 온건한 대응을 지지할 이유를 제공한다.[22]

---

22) 온건한 대응을 지지하기 위해 양립가능주의자들은 (Kearns 2012; Schlosser 2015) 여기서 제시된 것보다 더 강한 어떤 것을 양립불가능주의자들에게 요구하기도 한다. 이들에 따르면, 어니의 행위가 자유행위가 아닌 데 조작이 중요한 역할을 하지 않는다면, 단지 결정론적인 요인만이 작동하면서 직관적으로 명백하게 자유행위가 아닌 사례를 제시하는 것이 가능해야 한다. 하지만 이는 지나치게 강한 요구사항인 것 같다. 설령 양립불가능주의자들이 자신들의 입장을 지지하는 직관적으로 명백한 사례를 제공하지 못한다고 해도, 그 입장을 더 선호할만한 이유를 제시한다면 양립불가능주의를 지지할 수 있다.

양립불가능주의자들은 다이애나 사례에서 어니의 행위가 자유행위가 아니라는 직관이 어니의 "컨트롤을 넘어선 요인으로 거슬러 올라가는"[23] 결정론적 요인에 의해 설명된다고 주장한다. 그런데 다이애나의 개입과 번개의 개입을 비교할 때, 이 둘은 어니의 '컨트롤을 넘어선 요인으로 거슬러 올라가는' 공통점을 갖고 있긴 하지만, 어니의 행위에 대한 컨트롤의 행사 측면에서는 차이를 보인다. 물론 양립불가능주의자들은 이 차이가 중요한 차이가 아니라고 주장할 것이다. 하지만 온건한 대응을 지지하는 앞서의 이유를 고려할 때, 조작논증을 지지하는 추가적 근거가 제공되지 않는 한 이 주장만으로 온건한 대응을 반박하지 못한다.

마지막으로, 온건한 대응을 지지하는 앞서의 논의는 조작적 개입이 어니의 행위를 자유행위가 아니게 하는 이유를 분석하여 설명하지 않는다. 양립불가능주의자들은 이 점을 들어 온건한 대응을 반박하려 할 수 있다. 다이애나가 의도적으로 개입하여 어니의 행위를 컨트롤 할 때, 어니가 어떤 행위를 할지는 다이애나에게 달려있는 듯 보인다. 그리고 다이애나의 컨트롤 능력이 크면 클수록 어니의 행위는 자유행위가 아니라는 직관이 점점 강해지는 것 같다. 이에 더해 조작적 개입과 자유행위의 관계에 대한 분석이 더해진다면 온건한 대응이 더 강화될 수 있을 것이다.[24] 그러나 요구되는 분석

---

23) Pereboom (1995), p. 25.

24) 2절 시작부분에서 논의한 철수-블랙 사례는 자유와 책임이 성립하지 않게 하는 전형적인 조작의 사례이다. 만약 어니의 행위와 철수의 행위가 동일한 의미에서 조작된 행위라면, 어니의 행위가 자유행위가 아니라는 것이 함축될 것이다. 두 사례 모두에서 개입은 행위자가 스스로 고려하는 실천적 이유와 상관없이 일어났고, 행위자는 개입의 의도대로 행동한다. 그러나 이 두 사례를 비교할 때 중요한 차이가 있다. 블랙의 개입은 철수의 결정 메커니즘을 우회하는 방식으로 발생한 반면, 다이애나의 개입은 어니라는 인격체가 형성되기 전에 매우 은밀한 방식으로 일어나서 어니는 표준적인 양립가능주의자들의 자유의 조건을 다 만족한다. 또한 개입의 시점에도 큰 차이가 있다. 철

이 없다는 것이 온건한 대응에만 해당되는 문제는 아니다. 양립불가능주의자들은 다이애나 사례에서 어니의 행위가 자유행위가 아닌 이유가 조작적 개입 때문이 아니라 결정론적 요인 때문이라고 주장한다. 이 주장을 직접적으로 지지하기 위해서는 양립불가능주의자들 역시 조작적 개입과 자유행위의 관계에 대한 분석을 제공할 필요가 있다.

## 5. 나가는 말

근원-양립가능주의는 현재 많은 양립가능주의자들이 받아들이는 입장이다. 조작논증은 근원-양립가능주의에 대해 강력한 반론을 제기한다. 이 논증에 대한 대표적인 양립가능주의자들의 답변은 강경한 대응이다. 하지만 이 대응은 조작논증에 대한 충분한 반박이 되지 못한다. 반면 온건한 대응에 대해서는 그 입장을 지지할만한 이유가 있고, 따라서 이 대응은 조작논증으로부터 양립가능주의를 옹호할 수 있다.

---

수의 경우엔 자신이 고려한 이유에 따라 스스로 결정하기 바로 전에 개입이 일어나고 어니의 경우엔 수정란이 막 형성되는 시점에 일어난다. 이러한 차이에도 불구하고, 다이애나는 자신의 개입으로 인해 세계가 어떻게 전개될지를 알며 어니가 자신이 의도한대로 행동하게 만들기 위해 필요한 일을 할 수 있으므로 개입 시점의 차이가 개입의 성격에 큰 차이를 만들지 않는다고 주장할 수도 있다. 하지만 두 사례 모두에 적용되는 조작에 대한 분석이 정확히 어떤 것일지 모르겠다.

## 참고문헌

Ayer, A. J. (1954), “Freedom and Necessity”, In Derk Pereboom (2009), ed., *Free Will* (2nd ed.), Hackett Publishing, pp. 139-147.

Fischer, M. (2011), “The Zygote Argument Remixed”, *Analysis*, 71, pp. 267-272.

Fischer, M. & Ravizza, M. (1998), *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press.

Frankfurt, H. (1969), “Alternate Possibilities and Moral Responsibility”, *Journal of Philosophy*, 66, pp. 829-839.

Frankfurt, H. (1971), “Freedom of the Will and the Concept of a Person”, *Journal of Philosophy*, 68, pp. 5-20.

Hume, D. (1748), *An Enquiry Concerning Human Understanding*, P.H. Niditch, ed., Oxford: Clarendon Press.

Kearns, S. (2012), “Aborting the Zygote Argument”, *Philosophical Studies*, 160, pp. 379-389.

McKenna, M. (2008), “A Hard-Line Reply to Pereboom’s Four-Case Manipulation Argument”, *Philosophy and Phenomenological Research*, 77, pp. 142-159.

McKenna, M. & Pereboom, D. (2015), *Free Will: A Contemporary Introduction*, Routledge.

Mele, A. (2006), *Free Will and Luck*, Oxford University Press.

Pereboom, D. (1995), “Determinism Al Dente”, *Noûs*, 29, pp. 21-45.

Pereboom, D. (2001), *Living Without Free Will*, Cambridge University Press.

Pereboom, D. (2008), “A Hard-Line Reply to the Multiple-Case Manipulation Argument”, *Philosophy and Phenomenological Research*, 77, pp. 160-170.

Sartorio, C. (2016), *Causation and Free Will*, Oxford University Press.

Schlosser, M. (2015), “Manipulation and the Zygote Argument: Another Reply”, *The Journal of Ethics*, 19, pp. 73-84.

van Inwagen, P. (1983), *An Essay on Free Will*, Oxford University Press.

서울시립대학교 철학과
University of Seoul, Department of Philosophy
sungsukim@uos.ac.kr

# Compatibilist Replies to Manipulation Arguments

Sungsu Kim

Manipulation arguments purport to show that source freedom is incompatible with determinism. According to manipulation arguments, (1) intuitively, an agent manipulated in a certain manner to do A is not free, and (2) there is no significant difference between manipulation and determination. It follows that an agent determined to do A is not free. A compatibilist hard-line reply ('HR') denies (1), whereas a soft-line reply ('SR') denies (2). HR, which is arguably compatibilists' favorite, is assessed from the 'neutral stance.' HR turns out to fall short of adequately rejecting manipulation arguments. Recently Sartorio defends HR by claiming that (1) commits some sort of psychological fallacy. I argue that it does not work. I claim that SR is more promising. I examine the difference between intentional manipulation by design and ordinary determination. I argue that this difference suggests some determination scenario without intentional manipulation to which SR and manipulation arguments make different predictions, and that SR is better supported. Finally, incompatibilist objections are considered and replied.

Key Words: Free will, Determinism, Compatibilism, Manipulation arguments, Hard-line reply, Soft-line reply