

# 설명, 도덕적 책임, 그리고 양립 불가능론\*

이재호·이은홍\*\*

**주제분류** 형이상학

**주요어** 설명, 도덕적 책임, 여지 양립 불가능론, 원천 양립 불가능론

**요약문**

이 논문에서 필자들은 (반사실적) 의존, 설명, 도덕적 책임이 하나의 개념적 클러스터를 이루고, 이것에 산출(production)은 속하지 않는다는 논제를 받아들일 경우 자유의지의 문제에 대한 양립 불가능론을 옹호하기가 훨씬 쉬워진다는 것을 논증한다. 필자들은 양립 불가능론을 옹호하려는 사람들은 여지 양립 불가능론보다는 원천 양립 불가능론을 옹호하는 것이 유리하다는 것을 논증한 후, 원천 양립 불가능론에 대한 대표적인 비판인 J.M. Fishcer의 비판은 위의 논제를 받아들일 경우 쉽게 무력화된다는 것을 논증한다.

---

\* 이 논문은 2014년도 중앙대학교 신입생생성적우수장학금의 지원에 의하여 작성되었음.

\*\* 중앙대학교

## 1. 서론—의존, 설명, 도덕적 책임

이 논문에서 필자들은 **인과** 개념을 중심으로 진행되어 왔던 도덕적 책임과 자유의지에 관한 논의들을 **설명** 개념을 중심으로 전환했을 때 어떤 차이가 발생하는지를 추적하고자 한다. 보다 구체적으로, 필자들은 이 논문에서 이런 논의의 전환을 받아들일 경우, 자유의지와 관련한 양립 불가능론을 옹호하는 것이 보다 쉬워진다는 것을 보이고자 한다. 따라서 필자들이 이 논문에서 옹호하고자 하는 주장은 기본적으로 조건적인 주장이다.

필자들은 이 조건적인 주장이 자체로도 흥미롭다고 생각한다. 하지만 필자들이 이 조건문을 옹호하는 것이 매우 큰 의미를 갖는다고 믿는 보다 직접적인 이유는 이 조건문의 전건이 참이라는, 또는 최소한 대단히 그럴듯하다는, 것이 이미 다른 곳에서 논증된 바 있기 때문이다.<sup>1)</sup> 따라서 필자들이 이 논문에서 옹호하는 것은 조건적 주장이지만 필자들이 이 논문에서의 논증을 통해서 내리고자 하는 결론은 이 조건적 주장의 후건, 즉 자유의지에 관한 양립 불가능론이 대단히 그럴듯하다는 것이다.

이런 이유로, 이 “전건”에 대해 약간의 설명을 제공하는 것은 적절해 보인다. 이재호(2016)의 핵심 주장을 요약하면 다음과 같다. J. 쉐퍼나 N. 홀이 강조하듯이 기존의 (환원적) 인과 이론에는 기본적으로 두 개의 큰 흐름이 존재한다. 쉐퍼는 이를 각각 “확률-증가”와 “과정-연결”이라고 부르고 홀은 “의존”과 “산출”이라고 부른다.<sup>2)</sup> 비록 이들이 사용하는 명칭

1) 이재호, “인과의 두 개념, 설명, 그리고 도덕적 책임,” *과학철학* 19, no. 1 (2016). 앞으로 이 논문은 “이재호(2016)”으로 표기될 것임.

2) J. Schaffer, “Causes as Probability Raisers of Processes,” *Journal of Philosophy* 98, no. 2 (2001); Edward J. Hall, “Two Concepts of Causation,” in *Causation and Counterfactuals*, ed. John David Collins, Edward J. Hall, and L. A. Paul (Cambridge, Mass.: MIT Press, 2004).

과 성격 규정이 상이하지만 이들의 아이디어는 기본적으로 대동소이하며, 이들 사이의 기술적인 차이는 현재의 맥락에서는 무시될 수 있다. 의존(또는 확률-증가)에 있어서 핵심은 반사실적 의존성이다. 이 아이디어에 따르면 a가 b를 야기했다는 것은 a가 발생하지 않았다면 b도 발생하지 않았거나 최소한 b의 발생 확률이 낮아졌을 것이라는 것이다. 반면 산출(또는 과정-연결)에 있어서 핵심은 연속적인 물리적 과정을 통한 연결이다. 이 두 개의 구별된 흐름이 존재한다는 것과 관련해서 쉐퍼와 홀은 의견이 일치하지만, 이런 현상으로부터 이들은 상이한 결론을 이끌어 낸다. 쉐퍼는 이 현상으로부터 확률-증가와 과정-연결을 결합한 새로운 하이브리드 이론을 제안하고자 하는 반면, 홀은 의존과 산출을 두 개의 별개의 인과 개념으로 받아들일 것을 제안한다. 이재호(2016)는 이들의 제안 자체에 관심을 두기 보다는 이들의 논변이 모두 인과와 설명이 밀접하게 연결되어 있다는 전제에 기반하고 있다는 사실에 주목한다. 이 분석에 따르면, 인과와 설명이 별개의 문제라는 것을 받아들일 경우 순수하게 과정-연결에 기반한 인과 이론에 대한 쉐퍼의 공격도 무력화되고, 의존도 정당한 인과 개념의 하나로 받아들여져야 한다는 홀의 논변도 무력화된다. 이 사실로부터 이재호(2016)는 설명은 산출(또는 과정-연결)보다는 의존(또는 확률-증가)에 보다 연결되어 있다는 것을 논증한다. 왜냐하면 인과와 설명이 별개의 문제라는 것을 받아들여도 산출(또는 과정-연결)에 기반한 인과 이론은 전혀 손상되지 않으며 오히려 더 옹호하기 쉬워지기 때문이다. 더 나아가 이재호(2016)는 산출(또는 과정-연결)이 설명과는 별로 상관 없다는 것이 상당한 정도로 그럴듯하다는 것도 주장한다. 다음으로 이재호(2016)는 도덕적 책임이 인과 보다는 설명에 보다 밀접한 관련을 갖는다는 것을 주장한다. 인과적 직관은 명확하지 않지만 설명적 직관은 명확한 곳에서 우리의 도덕적 책임에 대한 직관은 명확한 반면, 인과적 직관은 명확하지만 설명적 직관은 명확하지 않은 곳에서 우리의 도덕적 책임에 대한 직관은 명확하지 않기 때문이다. 또 도덕적

책임은 설명과 마찬가지로 정도를 허용하는 개념이며, 배경 지식이나 대조의 초점과 같은 문맥 상대적인 요소에 대단히 민감한데 이런 특징들은 인과에서는 뚜렷이 나타나는 특징이 아니다. 이는, 이재호(2016)에 따르면, 도덕적 책임이 인과 보다는 설명에 연동되는 개념이라는 생각을 더욱 그럴듯하게 만들어 준다. 이런 이유로 이재호(2016)는 의존-설명-도덕적 책임이 하나의 개념적 클러스터를 이루고 이 개념적 클러스터에 산출은 포함되지 않는다고 결론 내린다.

이제 필자들이 앞서 “전건”이라고 부른 것, 즉 “설명 중심의 도덕적 책임과 자유의지에 관한 논의로의 전환”이라고 부른 것이 보다 구체적인 내용을 갖고 기술될 수 있게 되었다. 필자들이 “설명 중심의 도덕적 책임과 자유의지에 관한 논의”라고 부른 것은 바로 이 의존-설명-도덕적 책임이 하나의 개념적 클러스터를 이루고, 산출은 이 개념적 클러스터에 포함되지 않는다는 (앞으로 “의존-설명-책임 클러스터 논제”라고 불릴) 논제를 받아들이면서 진행되는 논의이다. 반복하자면, 필자들은 이 논문에서 이 논제를 받아들일 경우 자유의지에 관한 양립 불가능론을 옹호하는 것이 훨씬 쉬워진다는 것을 논증할 것이다.

이 논문의 나머지 부분은 다음과 같이 진행될 것이다. 2절에서는 양립(불)가능론이 현재 여지 양립(불)가능론(leeway (in)compatibilism)과 원천 양립(불)가능론(source (in)compatibilism)이라는 두 가지 형태로 나타나며 이들 가운데서 여지 양립 불가능론의 형태를 주장하는 것은 양립 불가능론자의 입장에서는 현명하지 않다는 것이 논증된다. 3절에서는 원천 양립 불가능론에 대한 일반적인 논증과 그 논증의 배후에 있는 직관의 원천을 드러내는 것을 통해서 원천 양립 불가능론이 왜 전망이 밝은 전략인지를 설명한다. 4절에서는 원천 양립 불가능론에 대한 대표적인 비판자인 J. M. 피셔의 논변을 살펴 보고, 그의 원천 양립 불가능론에 대한 공격이 왜 우리의 “전건”을 받아들였을 경우 무력화되는 지 살펴볼 것이다.

## 2. 양립 불가능론의 두 형태와 여지 양립 불가능론의 곤궁

오늘날 많은 철학자들이 양립 불가능론(또는 양립 가능론)에 두 개의 구분되는 형태가 있다는 것을 받아들인다. 종종 이 구별되는 양립 불가능론은 각각 “여지 양립 불가능론”(leeway incompatibilism)과 “원천 양립 불가능론”(source incompatibilism)으로 불린다. 이 절에서 필자들은 양립 불가능론을 주장하려는 사람들은 원천 양립 불가능론을 주장할 좋은 이유가 있다는 것을 논증하려 한다.

우선 여지 양립 (불)가능론과 원천 양립 (불)가능론에 대한 대략적인 설명이 필요하다. D. 패러봄에 따르면, 여지 양립 불가능론이란 그것에 따라 경우 대안적인 가능성의 존재가 도덕적 책임을 설명하는데 있어서 결정적이게 되는 양립 불가능론이고, 원천 양립 불가능론이란 행위자의 인과적 역사에 보다 중요한 역할을 부여하는 양립 불가능론이다. 원천 양립 불가능론에 있어서, 어떤 행위자의 행위에 대한 그 행위자의 도덕적 책임은 그에게 가능했던 대안적 가능성에 의해서 설명되는 것이 아니라 그 행위자가 특정한 방식으로 자신의 행위의 원천이 될 수 있게 만들어주는 특정한 종류의 인과적 역사를 그 행위자가 갖는다는 사실에 의해서 설명된다.<sup>3)</sup> 이들의 핵심 아이디어를 직관적으로 표현하면 다음과 같다. 우선, 여지 양립 불가능론의 주장은 다음과 같이 진행된다. 결정론이 참일 경우 우리는 다르게 행동할 수 없게 된다. 그런데 다르게 행동할 수 있느냐의 여부가 도덕적 책임과 자유 의지에 있어서 핵심적인 것이다. 따라서 결정론이 참일 경우 우리는 자유 의지를 가질 수 없게 된다. 반면에 원천 양립 불가능론의 주장은 다음과 같이 진행된다. 결정론이 참

---

3) Derk Pereboom, “Hard Incompatibilism,” in *Four Views on Free Will* (Blackwell, 2007), 85-86.

일 경우 우리의 행위는 우리가 통제할 수 없는 것이 분명한 사건들, 예컨대 우리가 태어나기도 전에 발생했던 사건들에 그 원천을 갖게 된다. 그런데 어떤 행위가 자유로운 행위이기 위해서는 그것의 궁극적인 원천이 내 안에 있어야 한다. 따라서, 결정론이 참일 경우 우리는 자유 의지를 가질 수 없다.

과거에는 양립 가능론과 양립 불가능론 사이의 논쟁이 주로 대안적인 가능성의 존재, 또는 대안적인 가능성의 의미에 관련해서 이루어졌다. 양립 불가능론자들은 결정론이 참일 경우 대안적인 가능성이 존재할 수 없게 된다고 주장하면서 양립 불가능론을 옹호했고(여지 양립 불가능론), 반면에 양립 가능론자들은 결정론이 참이라고 하더라도 어떤 매우 중요한 의미에서 대안적인 가능성이 여전히 존재한다고 주장하면서 양립 가능론을 옹호했다(여지 양립 가능론). 그러나 최근의 논의는 대안적 가능성을 중심으로 논의되기 보다는 행위의 모든 원천이 내 안에 있어야 하는지의 문제와 관련해서 주로 논의가 이루어진다. 양립 불가능론자들은 결정론이 참일 경우 우리의 행위의 궁극적 원천이 내 밖에 있게 된다는 이유로 양립 불가능론을 옹호하고(원천 양립 불가능론) 양립 가능론자들은 우리의 행위의 모든 원천이 내 안에 있어야 한다고 주장할 필요는 없다는 이유로 양립 가능론을 옹호한다(원천 양립 가능론). 이런 식으로 논의의 초점이 전이된 것은 기본적으로 여지 양립 가능론이 더 이상 진지하게 옹호되기 어렵다는 일종의 공감대가 성립했기 때문이다. 예를 들어, M. 바르가스는 다음과 같이 말한다.

보다 논쟁의 가능성이 있기는 하지만, 나는 자유 의지의 문제에 대한 접근들에 관련해 어떤 접근이 전망이 밝고 어떤 접근이 그렇지 않은지에 대해 공감대가 커지고 있다고 생각한다. 예를 들어 오늘날 자유 의지 문제가 “할 수 있다”에 대한 단순 조건문 분석에 호소하는 것을 통해서 해결될 수 있다고 생각하는 양립 가능론자들은 거의 없다. 또한 자유로운 행위자에 대한 자신들의 설명에서 자연적이지

않은 속성들에 호소하는 것에 편안함을 느끼는 자유주의자들은 더더욱 없다. 40년 전에는 이런 문제들에 대해 오늘날 우리가 갖고 있는 이런 정도의 수렴은 없었다.<sup>4)</sup>

여기서 언급되는 단순 조건문 분석이, 앞으로 보게 되듯이, 바로 여지 양립 가능론자들이 즐겨 사용해온 대안적 가능성에 대한 분석이므로 단순 조건문 분석에 대한 바르가스의 진단은 바로 (초기 형태의) 여지 양립 가능론에 대한 진단이기도 하다. 그러나 필자들은 바르가스의 진단대로 여지 양립 가능론이 사실상 죽은 이론이라고 봐야 하는 것은 아니라고 생각하며, 바로 이런 사실이 우리가 여지 양립 불가능론보다는 원천 양립 불가능론에 집중할 좋은 이유가 된다고 생각한다.<sup>5)</sup> 이를 논증하기 위해서 우선 필자들은 단순 조건문 분석에 호소하는 여지 양립 가능론을 옹호하는 대표적인 철학자인 A. J. 에이어의 논변을 살펴 보고, 그것에 대한 표준적인 비판을 살펴본 후 왜 그 표준적인 비판에 문제가 있는지 살펴볼 것이다.

에이어의 논증은 기본적으로 두 가지 논변을 통해서 진행된다. 첫 번째 논변은 결정론이 참이라고 하더라도 “다르게 행동할 수 있었다”가 참이 될 수 있는 이것에 대한 해석이 존재한다는 것을 보이는 것이다. 두 번째는 결정론이 참일 경우 “다르게 행동할 수 있었다”가 거짓이 되는 해석이 왜 자유의지와 관련해서 크게 우려할 필요가 없는 해석인지를 보이는 것이다. 에이어의 첫 번째 논변에서 핵심적인 역할을 하는 것이 바로 단순 조건문 분석이다. 그에 따르면, 여지 양립 불가능론적인 직관이 발생하는 것은 우리가 자유와 인과성(causality)이 대조되는 개념이라고

4) Manuel Vargas, “Response to Kane, Fischer, and Pereboom,” *ibid.*, 219.

5) 사실 필자들의 이런 주장은 바르가스에 대해서는 완벽하게 공정한 것이라고 볼 수는 없다. 왜냐하면 그는 **단순** 조건문 분석이 사실상 죽은 이론이라는 공감대가 있다는 것을 주장했을 뿐, 조건문 분석 자체가 죽은 이론이라는 공감대가 있다는 것을 주장하지는 않았기 때문이다.

생각하기 때문이다. 그러나 그에 따르면, 적어도 일상적인 의미에서 자유와 대조되는 것은 강제(constraint)이다.<sup>6)</sup> 그리고 어떤 사람의 행위가 강제된 행위인지 아니면 자유로운 행위인지는 다음과 같은 조건문의 진리값을 생각해 보는 것을 통해서 결정할 수 있다. “어떤 사람이 다르게 행동하기로 결심했었다면 그는 다르게 행동할 수 있었다.” 예를 들어, 직관적으로, 병적 도벽이 있는 사람이 물건을 훔치는 것은 자유로운 행동이 아니다. 그리고 이에 대한 에이어의 분석은 다음과 같다.

병적 도벽자(kleptomaniac)는 도둑질과 관련해서 자유로운 행위자가 아니다. 왜냐하면 그는 도둑질을 할 것인지에 관해 결정을 내리는 어떤 과정도 거치지 않았기 때문이다. 다르게 말하면, 만약 그가 그런 과정을 거쳤다고 하더라도 그것은 그의 행동에 무관하다. 그가 무슨 결심을 했건 그는 마찬가지로 도둑질을 했을 것이다. 그리고 그를 [자유로운 행위자인] 일상적인 도둑과 구분 지워주는 것이 바로 이것이다.<sup>7)</sup>

이런 방식으로 강제된 행위와 강제되지 않은 행위를 구분하는 것은 결정론이 참이라도 가능하다. 왜냐하면 결정론이 참이라고 하더라도, 일상적인 도둑의 경우에, “그가 도둑질을 하지 않기로 결심했었다면 그는 도둑질을 하지 않았을 것이다”는 참이었을 것이고, 에이어의 구분법에 따르면 그 도둑은 강제되지 않은 채로 도둑질을 한 것이 될 것이기 때문이다.

에이어는 어떤 의미에서 결정론이 참이라면 일상적인 도둑의 도둑질도 필연적인 것이 되며 따라서 다르게 행동할 수 없었던 것이 된다는 것을 받아들인다. 그러나 이런, ‘강제되었다’와 구분되는 것으로서의 ‘인과적으

---

6) A. J. Ayer, “Determinism, Free Will, and Responsibility,” in *Free Will*, ed. Gary Watson (Oxford, [Oxfordshire] ; New York: Oxford University Press, 1982), 19.

7) *Ibid.*, 20.

로 결정되었다'로서의 '다르게 행동할 수 없었다'는 자유의지와는 무관한 의미라고 주장한다. 왜냐하면, 흄적인 인과에 대한 규칙성 개념을 전제하면서, 어떤 것이 인과적으로 결정되었다는 것은 그것과 그것의 원인 사이에 어떤 항상적 결합이 있다는 것 이상은 아니며, 그것은 강제와는 전혀 무관한 개념이기 때문이다.<sup>8)</sup>

에이어식의 '다르게 행동할 수 있었다'에 대한 단순 조건문 분석은 한 때 큰 영향력을 미쳤지만, 위의 바르가스의 인용문에서 주장되듯이, 오늘날 많은 비판을 받는다. 단순 조건문 분석에 대한 표준적인 비판은 이 분석이 **결정 자체가** 강제된 사람을, 우리의 직관과는 반대로, 자유로운 사람으로 만들어 버릴 수 있다는 것이다.<sup>9)</sup> 예를 들어 어떤 뇌신경학자가 다른 사람의 뇌에 칩을 몰래 심어서 그 사람의 결정을 조작할 수 있게 되었다고 하자. 그리고 그 뇌신경학자는 그 다른 사람의 뇌를 조작해 자신의 라이벌을 살해하도록 결심하게 만들었다고 하자. 그리고 이 결심은 실제로 그 라이벌의 살해로 이어졌다고 하자. 이 상황에서 직관적으로 그 뇌가 조작된 사람은 자유롭게 행동한 것이 아니다. 그러나 에이어식의 단순 조건문 분석에 따르면 그의 행동은 강제에 의한 것이 아니고, 따라서 자유로운 행동이 된다. 왜냐하면 "그 사람이 살인할 결심을 하지 않았다면 그 사람은 살인을 하지 않았을 것이다"는 명백히 참이기 때문이다.

이런 표준적인 비판이 원래 형태의 단순 조건문 분석에 치명타를 입힌다는 것은 분명하지만 단순 조건문 분석이 기반하고 있는 핵심 아이디어

---

8) Ibid., 21. 비슷한 생각은 흄적인 인과 또는 법칙 개념을 옹호하는 다른 많은 철학자들에게서 발견된다. Helen Beebe, "The Non-Governing Conception of Laws of Nature (Ramsey-Lewis Theory, Thought Experiments)," *Philosophy and Phenomenological Research* 61, no. 3 (2000): 579.

9) 이런 비판은 예컨대 다음에서 발견된다. Roderick M. Chisholm, "Human Freedom and the Self," in *Free Will - Second Edition*, ed. Gary Watson (Oxford, 2003), 28-29.

에 치명타를 입히는지는 사실 분명하지 않다. 왜냐하면 단순 조건문 분석이 기반하고 있는 핵심 아이디어는, 자유의 반대가 인과적 결정이 아니라 강제되지 않음이고, 따라서 ‘다르게 행동할 수 있음’은 ‘강제되지 않음’으로 이해되어야 한다는 것인데, 이 아이디어가 표준적인 반례에 의해서 손상되는 것으로 보이지는 않기 때문이다. 위의 반례에서 우리는 조작된 뇌를 가진 사람의 행동이 강제된 행동이라는 강한 직관을 갖는다. 이 사례가 단순 조건문 분석을 통해서 다뤄질 수 없는 이유는 단순 조건문 분석이 기본적으로 결정과정에 개입하는 강제에 대한 분석이 아니라 행동에 직접적으로 가해지는 강제에 대한 분석으로만 작동하기 때문이다. 따라서 병적 도박과 같은 **행동 자체**에 가해지는 강제와 관련해서는 잘 작동하지만 조작된 뇌의 경우처럼 결정과정에 가해지는 강제에 대해서는 작동하지 않는다. 이런 점에서 보면 표준적인 반례는 **기술적인** 반례에 불과한 것처럼 보이며, 따라서 기술적인 방식으로 대처가 가능할 수도 있다는 생각이 든다. 그러나 J. M. 피셔에 따르면, 표준적인 반례에 대한 기술적인 대응이, 즉 단순 조건문 분석을 정제된(refined) 조건문 분석으로 전환시키는 것이 생각만큼 쉬운 것은 아니다. 예를 들어 우리는 다음과 같이 단순 조건문 분석을 정제할 수 있다.

정제된 조건문 분석: 어떤 행위자 S는 다음과 같은 경우에 한해 X를 행할 수 있다. (i) 만약 S가 X를 하기로 선택했다면, S는 X를 했을 것이고, (ii) 그 행위자는 비밀리에 최면에 걸리지도 않았고, 잠재의식 광고에 노출되지도 않았고, 과거 경험으로부터 결과하는 심리적 강제에 노출되지도 않았고, 뇌에 대한 직접적 조작에 노출되지도 않았고, 낙상이나 사고에 의한 신경 손상을 받지도 않았고 등등.<sup>10)</sup>

그러나 피셔는 이 경우 “등등”이라는 바람직하지 않는 표현을 사용해

---

10) John Martin Fischer, “Compatibilism,” in *Four Views on Free Will* (Malden, MA ; Oxford: Blackwell Pub., 2007), 51.

야 하고, 더 나아가 개념적 분석 안에 원리적으로 통일되지 않은 상이한 아이템들의 나열을 집어 넣는 “불편함”을 가져야 한다고 비판한다. 여기서 피셔의 반론의 핵심은 단순 조건문 분석, 즉 (i)에서 보여지는 바의 **원리 지워지고 단순화된** 방식의 (행위 자체에 대한) 강제 배제가 결정 과정에 대한 강제 배제에서는 주어질 수 없다는 것이다.

필자들은 피셔의 원리 지워지고 단순화된 방식의 강제 배제에 대한 요구가 정당한 요구라고 생각한다. 그러나 필자들은 피셔의 생각과는 달리 결정 과정에 대한 강제 배제가 원리 지워지고 단순화된 방식으로 주어질 수 없다고 생각하지 않는다. 왜냐하면 우리는 이미 (i)에서 강제를 배제하는 원리 지워진 방법을 배웠으므로 그것을 확장해서 결정 과정에 적용하면 되기 때문이다. 이것은 다음과 같이 이루어질 수 있다.

정제된 조건문 분석\*: 어떤 행위자 S는 X를 행위 강제, 또는 결심 강제를 받지 않고 행했을 경우에 한해 X를 자유롭게 행한다.

행위 강제의 정의: 만약 S가 X를 하지 않기로 선택했어도 어떤 요소 Y 때문에 S는 X를 할 수밖에 없었을 것이라면 Y는 S가 X를 하게 만드는 행위 강제 요소가 된다. S는 X를 어떤 행위 강제 요소 없이 했을 경우 행위 강제 없이 X를 한 것이다.

결심 강제의 정의: 만약 S가 어떤 요소 Z 때문에 X를 하기로 결정했는데, Z가 S가 X하는 것에 대한 어떤 **가능한** 행위 강제 요소 Y와 **충분히 유사한** 요소라면 Z는 S가 X를 하는 것에 대한 결심 강제 요소이며, S가 어떤 결심 강제 요소 없이 X를 할 경우 S는 결심 강제를 받지 않고 X를 행한 것이다.

정제된 조건문 분석\*에 따를 경우, 위의 조작된 뇌를 가진 사람의 행동은 자유로운 행동이 아니다. 그 이유를 살펴보기 위해서 우선 뇌에 칩을 심어(y) **행위를 직접 조작하는** 것을 통해서 살인을 하는 경우를 생각해 보자. 이 경우 우리는 그 사람의 살인 행위가 자유로운 행위가 아니라고 볼 수 있는데, 그것은 그의 행위가 어떤 행위 강제 요소, y에 의해

서 발생했기 때문이다. (여기서  $y$ 가 행위 강제 요소인 것은 ‘그 사람이 다르게 결정했다고 하더라도 동일한 행동을 했을 것이다’를  $y$ 가 참이게 만들어 주기 때문이다.) 이제, 우리의 원래의 조작된 뇌의 사례에서 **결심을 조작하는** 칩을 심는 것( $z$ )은 결심 강제 요소라고 말할 수 있다. 왜냐하면  $z$ 와 충분히 유사한 요소  $y$ 가 행위 강제 요소이기 때문이다. 반면에 결정론이 보장하는 인과적 결정은 그것이 행위자의 결심을 결정한다고 하더라도 결심 강제로 볼 수 없다. 위의 에이어의 논변에서 잘 드러났듯이 그것과 유사한 결정론적인 인과적 결정이 행위 강제로 판명되지 않기 때문이다. 정제된 조건문 분석\*은 단순 조건문 자체에 주목하기 보다는 단순 조건문 분석의 배후에 있는 핵심 아이디어, 즉 강제의 배제에 집중하기 때문에 모든 강제의 배제를 조건문을 통해서 수행하려 하지 않으며, 이를 통해서 표준적인 비판을 피해갈 수 있게 된다.

혹자들은 필자들의 이런 대응 방식이 “충분히 유사한”이라는 모호한, 또는 문맥 의존적인 표현에 지나치게 의존한다고 불평할지 모른다. 그러나 이런 비판은 “강제”라는 표현 자체가 모호한, 또는 문맥 의존적인 표현이라는 것을 고려하면 근거가 없다. 우리가 자유로운 행위를 강제되지 않은 행위로 이해한다는 것은 이미 “자유로운”을 모호한 표현으로 이해하겠다는 것을 의미한다. 모호한 표현을, 그 모호성의 원천이 무엇인지 드러내는 방식으로 모호하게 분석하는 것은 문제라기보다는 성취이다.

(정제된) 조건문 분석을 통한 여지 양립 가능론이 아직 죽은 것으로 볼 수 없다는 것과 더불어 여지 양립 불가능론의 전망을 어둡게 하는 또 하나의 문제는 흔히 “프랑크푸르트 사례”라고 알려진 사례들을 통해서 여지 양립 가능론을 옹호하는 가능성이 존재한다는 것이다. 전형적인 프랑크푸르트 사례는 다음과 같다. 철수는 국회의원 선거에서 여당 후보에 투표할 지 아니면 야당 후보에 투표할 지 고민하고 있다. 그는 결국 투표장에 가서 야당 후보에 투표를 했다. 그의 투표에는 어떤 비정상적인 과정도 개입되지 않았다. 그러나 그의 뇌에는 비밀리에 작은 칩이 심어

져 있었는데, 그 칩은 철수가 야당 후보에 투표하는 경우 아무런 작동도 하지 않지만 철수가 여당 후보에 투표하려고 하면, 바로 작동을 시작해서 철수가 야당 후보에 투표하게 만든다.<sup>11)</sup> 이 경우 철수는 “다르게 행동할 수” 없었던 것으로 보인다. 왜냐하면 그가 다르게 행동하려고 하면 바로 칩이 작동해서 그가 실제로 했던 행동을 어떻게든 하게 만들었을 것이기 때문이다. 그러나 그런 이유로 그의 행동, 즉 야당 후보에 투표하는 것이 자유로운 행위가 아니라고 볼 수는 없어 보인다. 또는 최소한 그의 행위가 도덕적 책임을 결여 하는 행위라고 볼 수는 없어 보인다.<sup>12)</sup> 왜냐하면 그는 칩의 작동에 의해서 투표를 한 것이 아니고 지극히 **정상적인** 방식으로 결정을 내려서 투표한 것이기 때문이다.

필자들은 여기서 프랑크푸르트 사례를 통한 여지 양립가능론의 정당화 가능성에 대해서 더 이상 깊이 들어갈 생각은 없다. 왜냐하면 프랑크푸르트 사례, 또는 이것들의 변형된 사례들이 정말로 여지 양립 가능론을 옹호하는데 사용될 수 있는지에 대한 많은, 그리고 **매우 혼란된** 논란들이 있기 때문이다.<sup>13)</sup> 필자들이 보기에 이 문제에 관한 한 어떤 종류의 합의도 존재하지 않으므로, 적어도 프랑크푸르트 사례를 통한 논변이 성공한다는 주장과 그것이 실패한다는 주장 모두 현단계에서는 정당화될 수 없다. 이 사실은, 적어도 양립 불가능론을 옹호하는 다른 방식이 존재한다고 믿는 필자들과 같은 사람들에게는, 프랑크푸르트 사례를 통한 여

11) 이 사례는 J. M. 피셔가 선호하는 프랑크푸르트 사례를 각색한 것이다. Ibid., 58.

12) 여기서 도덕적 책임의 문제와 자유의 문제를 구분하는 것은 J. M. 피셔처럼 이 두 문제를 구분해서 프랑크푸르트 사례가 전자의 문제에만 관여한다고 생각하는 사람들이 있기 때문이다. 필자들은 이런 구분에 별로 우호적이지는 않지만 그 구분의 가능성을 원칙적으로 배제할 필요는 없다고 생각한다.

13) 프랑크푸르트 사례들에 관한 논란들은 다음을 참고할 것. John Martin Fischer, “Frankfurt-Type Examples and Semicompatibilism: New Work,” in *The Oxford Handbook of Free Will 2nd Edition*, ed. Robert Kane (2011); David Widerker, “Frankfurt-Friendly Libertarianism,” *ibid.*; Ishtiyaque Haji, “Obligation, Reason, and Frankfurt Examples,” *ibid.*

지 양립 가능론을 공격할, 즉 여지 양립 불가능론을 추구할 동기를 현격하게 약화시킨다.

지금까지의 논의만으로도 필자들은 여지 양립 가능론의 대표 논변들을 완벽하게 제압하는 것이 쉽지 않으며, 여지 양립 가능론에 관한 논의들은, 의미론적으로, 그리고 직관과 관련해서 매우 혼탁하다는 것을 충분히 보였다고 생각한다. 물론 이 사실로부터 여지 양립 불가능론을 추구하는 것이 불가능하다거나 무의미하다는 것이 따라 나오는 것은 아니다. 그러나 다음 절들(3절, 4절)의 내용이 주어질 경우, 필자들처럼 양립 불가능론을 옹호하고자 하는 사람들에게 있어서 원천 양립 불가능론을 옹호하는 것이 보다 현명한 전략이라는 것은 분명해 보인다. 앞으로 보여지겠지만 원천 양립 불가능론은 보다 명확한 직관에 호소하고 있고, 그것에 대한 대표적인 반론을 제압하기가 보다 용이하기 때문이다.

### 3. 원천 양립 불가능론의 직관적 원천

이 절에서 필자들은 원천 양립 불가능론에 대한 하나의 잘 알려진 논변을 소개한 후 그 논변의 배후에 있는 직관의 원천을 드러내고자 한다. 우선, 대표적인 원천 양립 불가능론자들 가운데 하나인 R. 케인의 논변을 살펴보자. 그의 논변의 핵심에는 다음의, 그가 ‘궁극적 책임 원리’(UR)라고 부르는, 원리가 있다.

궁극적 책임 원리(UR): 어떤 행위자는 어떤 사건 또는 상태 E의 발생에 대해서 오직 (R) 그 행위자가 그가 자발적으로 행하거나 행하지 않은 것이면서 그가 다르게 자발적으로 행할 수도 있었던 것이 E의 발생이거나 또는 E의 발생에 인과적으로 기여했고 E가 발생할 것인지 아닌지에 대해서 차이를 만들었을 경우에 한해, 그리고 (U) 모든 X와 Y에 대해서(여기서 X와 Y는 사건 또는 상태의 발생을 의

미), 만약 그 행위자가 개인적으로 X에 대해서 책임이 있다면, 그리고 Y가 X에 대한 아르케(또는 충분한 근거 또는 원인 또는 설명)라면 그 행위자는 Y에 대해서도 개인적으로 책임을 져야 하는 경우에 한해 궁극적으로 책임을 진다.<sup>14)</sup>

케인은 어떤 행위자는 그가 자신의 행위에 대해서 UR의 의미에서 궁극적인 책임을 져야만 자유로운 행위를 했다고 볼 수 있다고 주장한다. 이렇게 이해된 UR을 받아들였을 경우 자유의지가 결정론과 양립 불가능하다는 것은 쉽게 보여진다. 왜냐하면 결정론이 참일 경우 우리의 임의의 행위에 대해서 우리가 태어나기도 전의 어떤 사건 또는 상태들이 충분한 원인 또는 설명이 될 것이고, 우리가 우리가 태어나기도 전의 어떤 사건 또는 상태들에 대해서 책임을 진다는 것은 불합리하므로, UR, 보다 구체적으로 (U)에 따라서, 우리는 우리의 모든 행위에 대해서 책임을 지지 않게 되며 따라서 우리의 행위는 자유롭지 않게 된다.

현재 논문의 주제와 관련해서 UR에는 두 개의 주목할 만한 특징이 있다. 그 첫 번째는 UR이 인과와 설명의 구분에 대해서 상당히 모호한 태도를 취한다는 것이다. 위의 인용문에서 잘 드러나듯이, 케인은 어떤 것이 다른 것의 원천이 된다는 것과 관련해서 그 관계가 인과적 관계로 파악되어야 하는지 아니면 설명적 관계로 파악되어야 하는지에 대해서 모호한 태도를 취하고 있다.<sup>15)</sup> 따라서 UR 자체는 인과를 중심으로 놓고 진행되는 논의와 설명을 중심으로 진행되는 논의의 차이에 관련해서 중립적이다.

다음으로, UR에 포함되어 있는 보편 양화된 진술로서의 (U)는 가장 적절한 방식으로 기술된 것으로 보이지 않는다는 것이 지적될 필요가 있

---

14) Robert Kane, *The Significance of Free Will* (New York: Oxford University Press, 1996), 35.

15) 필자들이 보기에 케인의 이런 태도는 그가 인과 중심의 논의와 설명 중심의 논의 사이의 중요한 구분을 인식하지 못했다는 것에 대한 좋은 증거이다.

다. 왜 그런지 알기 위해서 다음의 경우를 생각해 보자. 이 사례는, 그것이 가능한 것일 경우, 앞으로 중점적으로 논의될 피서의 논변에 우호적으로 작동할 사례이므로 “피서 친화적 사례”라고 불릴 것이다.

(피서 친화적 사례) 철수의 행동 A의 **일차적인** 원인 또는 설명이 철수가 책임 지어야 하는 사건 B이지만, 이 사건 B가 철수의 책임이 아닌 다른 사건 C에 의해서 (충분히) 야기 또는 설명된다.

이런 경우에 우리는 철수가 A에 대해서 책임을 갖는다고 생각한다. 왜냐하면 우리는 다음의, “책임의 순방향 전이 원리”라고 부를 법한 원리를 직관적이라고 생각하기 때문이다.

책임의 순방향 전이 원리: 어떤 행위자 X가 사건 Y에 대해서 책임을 갖고 있고 어떤 사건 Y가 사건 Z를 충분하게 야기하거나 설명한다면, 예외적인 상황이 아닐 경우, X는 Z에 대해서도 책임을 갖는다.<sup>16)</sup>

이제 위의 피서 친화적 사례가 가능하다면 원천 양립 불가능론을 옹호하기는 쉽지 않아진다. 앞에서 살펴보았듯이, 원천 양립 불가능론의 옹호에 핵심적인 아이디어는 우리의 행동의 원천이 우리가 책임을 지지 않는, 예컨대 우리가 태어나기도 전의, 사건들에 있다면 우리는 우리의 행동에 책임을 질 수 없게 되며 따라서 자유의지도 없게 된다는 것인데, 위의 피서 친화적 사례가 가능할 경우 우리의 행동의 원천이 아주 먼 과거로 소급되어 간다는 사실이 우리의 자유의지에 어떤 손상을 주는지는 분명

---

16) 여기서 “예외적인 상황이 아닌 경우”가 삽입된 것은 X에게 A가 B를 야기할 것이라는 것이 전혀 예측될 수 없었던 경우처럼 특수한 상황에서는 책임이 순방향으로 전이되는 것이 분명하지 않기 때문이다. 이런 “예외적인 상황”들을 정확하게 성격 규정하는 것은 중요하고 또 어려운 작업이지만 현 논문의 핵심 주장에 결정적인 것은 아니므로 이 논문에서는 다루지 않기로 한다.

하지 않기 때문이다. 하지만 우리가 UR, 보다 구체적으로는 (U)를 받아들일 경우 이런 문제는 사라지는 것으로 보인다. 왜냐하면 (U)가 **보편 양화된** 주장이기 때문에, (U)를 받아들일 경우, C가 A를 최소한 간접적으로 야기 또는 설명할 수 있다는 사실과 C가 철수가 책임지지 않는 사건이라는 것, 그리고 책임의 순방향 전이 원리로부터 B가 철수가 책임지지 않는 사건이라는 것이 귀결되기 때문이다.<sup>17)</sup> 이런 점에서 볼 때, 케인이 (U)를 **보편 양화된 문장의 형태로** UR에 포함시킨 핵심적인 이유는 바로 위의 피서 친화적 사례의 가능성을 차단하기 위해서인 것으로 보인다.<sup>18)</sup>

그런데, 지금까지 우리가 확인한 것은 (U)를 받아들일 경우 피서 친화적 사례가 가능하지 않게 된다는 조건문적인 주장일 뿐이다. 우리는 왜 (U)를 받아들여야 하는지에 대해서도 아직 명확한 답을 듣지 못했고, 따라서 피서 친화적 사례가 왜 불가능한지에 대해서도 명확한 답을 듣지 못했다. 이것은 사소한 문제가 아닌데, 필자들이 보기에 (U)가 직관적으로 아주 명확한 것은 아니며, 피서 친화적 사례가 불가능하다는 것도 직관적으로 명확한 것은 아니기 때문이다. 이런 이유에서 필자들은 케인의 (U)보다는 다음의 “무책임의 순방향 전이 원리”에 호소해서 피서 친화적 사례를 차단하는 것이 보다 나은 방식이라고 생각한다. 불필요한 오해를

17) 조금 더 구체적으로 말하면, (U)와 C가 A를 야기 또는 설명한다는 것, 그리고 철수가 C에 대해서 책임을 지지 않는다는 것으로부터 철수가 A에 대해서 책임을 지지 않는다는 것이 귀결되고, 철수가 A에 대해서 책임을 지지 않는다는 것과 철수가 B에 대해서 책임을 질 경우, 책임의 순방향 원리에 따라, 철수는 A에 대해서도 책임을 진다는 것으로부터, 후건 부정법에 따라서, 철수가 B에 대해서 책임을 지지 않는다는 것이 귀결된다.

18) 만약 우리가 (U)의 존재 양화된 형태, 즉 다음의 (U')을 사용한다면 피서 친화적 사례가 차단되지 않는다는 것은 분명하다.

(U') 어떤 X와 Y에 대해서, 만약 그 행위자가 개인적으로 X에 대해서 책임이 있다면, 그리고 Y가 X에 대한 아르케(또는 충분한 근거 또는 원인 또는 설명)라면 그 행위자는 Y에 대해서도 개인적으로 책임을 져야 한다.

막기 위해서 침언하자면, 필자들이 제안하는 무책임의 순방향 전이 원리는 앞으로 설명되듯이 **사실상** 케인의 (U)와 같은 것이며, 따라서 필자들의 주장이 케인의 주장과 근본적으로 상이한 것은 아니다. 필자들이 제안하는 것은 케인의 (U)보다는, 그것과 결국 사실상 동일한 것인, 무책임의 순방향 전이 원리가 직관적으로 더 분명한 느낌을 준다는 것이다.

우선, 무책임의 순방향 전이 원리는 다음의 두 가지 형태로 표현될 수 있다.

무책임의 순방향 전이 원리1: 어떤 행위자 X가 사건 Y에 대해서 책임을 갖지 않고 어떤 사건 Y가 사건 Z를 충분하게 야기하거나 설명한다면 X는 Z에 대해서도 책임을 갖지 않는다.

무책임의 순방향 전이 원리2: 어느 누구도 사건 Y에 대해서 책임을 갖지 않고, 어떤 사건 Y가 사건 Z를 충분하게 야기하거나 설명한다면, 어느 누구도 Z에 대해서 책임을 갖지 않는다.

이 두 원리 가운데 첫 번째 원리는 두 번째 원리보다 더 강한 주장이다. 왜냐하면 첫 번째 원리로부터 두 번째 원리는 논리적으로 도출되지만 그 역은 성립하지 않기 때문이다. 이는 기본적으로 첫 번째 원리가  $(\forall x)(Fx \rightarrow Gx)$ 의 형식을 갖고 있지만 두 번째 원리는  $(\forall x)Fx \rightarrow (\forall x)Gx$ 의 형식을 갖고 있기 때문이다.<sup>19)</sup> 필자들의 생각에는 두 원리 모두 직관적이지만 두 번째 원리가 조금 더 직관적이다. 예를 들어서 우리는 (인간이 할 수 있는 모든 대비를 했음에도 불구하고) 불가항력적인 천재지변에 의해서 사람들이 죽었을 경우 그들의 죽음에 대해서 누군가가 도덕적으로 책임을 져야 한다고 생각하지 않는다. 즉, 우리는 이 경우 이것이 인

19) 보다 엄밀하게 표현하면 첫 번째 원리는  $(\forall x)(\forall y)((Ex \& Ey) \& Cxy) \rightarrow (\forall z)((Az \& \sim Rzx) \rightarrow \sim Rzy)$ 로 형식화될 수 있는 반면, 두 번째 원리는  $(\forall x)(\forall y)((Ex \& Ey) \& Cxy) \rightarrow (\forall z)(Az \& \sim Rzx) \rightarrow (\forall z)(Az \& \sim Rzy)$ 로 형식화될 수 있다. 여기서 Ax: x는 행위자이다, Ex: x는 사건이다, Cxy: x가 y를 충분하게 야기 또는 설명한다, Rxy: x가 y에 대해서 책임이 있다.

재(人災)가 아니라 천재(天災)이며 따라서 누구를 비난하는 것은 정당하지 않다고 생각한다. 이것에 대한 가장 그럴듯한 설명은 천재지변의 발생에 대해서 누구도 책임을 질 수 없고, **은전히** 천재지변 때문에 사람들이 죽음을 당했기 때문이라는 것으로 보인다. 그리고 이 설명은 위의 두 번째 원리에 호소하고 있는 것으로 보인다.

다음 단계는 무책임의 순방향 전이 원리들로부터 다음의 책임의 역방향 전이 원리를 정당화하는 것이다.

책임을 역방향 전이 원리: 어떤 행위자 X가 사건 Z에 대해서 책임을 갖고 있고 어떤 사건 Y가 사건 Z를 충분하게 야기하거나 설명한다면 X는 Y에 대해서도 책임을 갖는다.

무책임의 순방향 전이원리<sup>1</sup>로부터 책임의 역방향 전이 원리를 정당화하는 것은 쉽다. 이 둘은 대우 관계에 있기 때문이다. 그러나 (보다 직관적인) 무책임의 순방향 전이원리<sup>2</sup>로부터도 책임의 역방향 전이 원리는 정당화될 수 있다. 만약 책임의 역방향 전이 원리가 거짓이라면 어떤 행위자 X가 어떤 사건 Z에 대해서 책임을 갖고, 사건 Y가 사건 Z를 충분하게 야기하거나 설명하지만 X는 Y에 대해서 책임을 갖지 않는 경우가 가능해야 한다. 그런데 이런 경우에 사건 Y에 대해서는 아무도 책임이 없거나 아니면 X 이외의 어떤 다른 사람(들)이 책임을 가져야 한다. 그러나 전자의 경우는 불가능하다. 왜냐하면 Y에 대해서 아무도 책임이 없다면, 무책임의 순방향 전이원리<sup>2</sup>에 의해서 Z에 대해서도 아무도 책임이 없어야 하는데 이는 가정에 위배되기 때문이다. 반면 후자의 경우에는 논리적 모순이 발생하는 것은 아니지만 매우 반직관적인 결과가 발생한다. 왜냐하면 X이외의 다른 사람(들)이 Y에 대해서 책임을 진다면 책임의 순방향 전이 원리에 의해서 그 다른 사람(들)은 Z에 대해서도 책임을 져야 하기 때문이다. 이 경우 Z에 관해서 X와 그 다른 사람(들)에게

이중으로 책임이 부과된다. 이 사실은 자체로 이상하며 다음의 완전한 책임의 순방향 전이 원리를 받아들일 경우 더욱 이상해진다.

완전한 책임의 순방향 전이 원리: 어떤 행위자 X가 사건 Y에 대해서 (부분적으로가 아니라) 완전한 책임을 갖고 있고 어떤 사건 Y가 사건 Z를 충분히 야기하거나 설명한다면, 예외적인 상황이 아닐 경우, X는 Z에 대해서도 완전한 책임을 갖는다.

우리가 완전한 책임의 순방향 전이 원리를 받아들일 경우, 우리는 일종의 과도결정 문제(over-determination problem)의 도덕적 책임 형태를 갖게 된다. 하나의 사건에 대해서 완전한 책임을 갖고 있는 사람(들)이 있음에도 불구하고 그 사건에 대해서 책임을 갖는 또 다른 사람을 지목하는 것이 가능하기 때문이다.

더 나아가 이 경우 X의 Z에 대한 책임이 다른 사람(들)이 Y에 대해서 책임을 갖고 있다는 것에 의존하는 또 다른 반직관적인 결과도 발생한다. 왜냐하면 만약 다른 사람(들)이 Y에 대해서 책임을 갖지 않았다면 무책임의 순방향 전이 원리2에 의해서 X도 Z에 책임을 갖지 않게 되었을 것이기 때문이다.

이제 원천 양립 불가능론의 직관의 원천을 찾는 마지막 단계는 이 책임의 역방향 전이 원리가 케인의 (U)와 동일한 것이라는 것을 깨닫는 것이다. 이렇게 볼 경우, 우리는 케인의 (U)에 대한 직관적 근원이 무책임의 순방향 전이 원리 1 또는 2에 있다고 볼 수 있다. 왜냐하면, 적어도 필자들의 눈에는, 이들이, 특히 논리적으로 1보다 더 약한 2가, (U)보다 더 직관적으로 보이기 때문이다. 더 나아가 무책임의 순방향 전이 원리는 책임의 순방향 전이 원리와 서로를 강화하는 관계에 있는 것으로 보인다. 책임이 인과 또는 설명 관계를 따라 전이된다면 무책임도 마찬가지로 전이된다고 생각하는 것은 자연스럽기 때문이다.

지금까지의 논의를 종합해 보면, 원천 양립불가능론은 궁극적으로 (책임의 순방향 전이 원리와 함께) 무책임의 순방향 전이 원리에 대한 우리의 직관으로부터 나오는 것이며, 그런 한에서 원천 양립불가능론의 전망은 상당히 밝다. 무책임의 순방향 전이 원리, 즉 책임을 지울 수 없는 사건 때문에 (부분적으로가 아니라) 온전히 발생한 사건에는 책임을 지울 수 없다는 원리는 자체로 매우 직관적이며, 마찬가지로 직관적인 책임의 순방향 전이 원리와 서로를 강화하는 관계에 있다. 따라서 (다음 절에서 보여지듯이) 우리가 원천 양립 불가능론을 거부할 아주 좋은 이유가 없다면 원천 양립 불가능론은 정당화될 수 있다.

#### 4. 원천 양립 불가능론에 대한 피셔의 공격과 그것의 문제

앞 절에서 우리는 원천 양립 불가능론이 매우 강력한 직관적인 원리, 즉 무책임의 순방향 전이 원리에 기반하고 있다는 것을 밝혔다. 이 절에서 필자들은 원천 양립 불가능론에 대한 대표적인 비판가라고 할 수 있는 피셔의 비판을 살펴보고 그 비판이 의존-설명-책임 클러스터 논제를 받아들이는 경우 왜 힘을 잃는지 논증할 것이다.

본격적으로 피셔의 비판과 그것의 문제를 살펴보기 전에, 우리는 앞 절에서 논의된 (무)책임의 순방향 전이 원리를 이 논문이 전제하고 있는 논제, 즉 의존-설명-책임 논제에 맞추어 수정할 필요가 있다. 앞서 지적된 바 있듯이, 케인의 UR은 인과와 설명을 구별하지 않고 모호하게 진술되어 있다. 이에 따라서 필자들은 앞 절에서 (무)책임의 순방향 전이 원리를 기술할 때 인과와 설명을 구별하지 않고 모호하게 진술했다. 이제, 이 부분은 수정될 필요가 있다. 필자들이 원천 양립 불가능론의 배후에 있는 직관의 원천으로 지목한 무책임의 순방향 전이 원리<sup>2</sup>는 다음과 같

이 수정되어야 하며, 다른 원리들도 마찬가지로 수정되어야 한다.

무책임의 순방향 전이 원리2\*: 어느 누구도 사건 Y에 대해서 책임을 갖지 않고, 어떤 사건 Y가 사건 Z를 (부분적으로가 아니라) 완전하게 **설명한다면**, 어느 누구도 Z에 대해서 책임을 갖지 않는다.

무책임의 순방향 전이 원리2\*은 더 이상 인과와 설명 사이에 모호한 태도를 취하지 않고 명시적으로 책임을 전이시키는 관계가 설명적 관계라는 것을 주장한다.

이제 피셔의 비판을 살펴보자. 피셔는 (G. 스트로슨, S. 스미란스키, R. 케인, D. 패러봄 등) 다양한 형태의 원천 양립 불가능론 논변을 구분한 후 그들을 각각 비판한다. 그러나 그의 이 비판들은 본질적으로 하나의 단일한 아이디어에 기반해 있다. 원천 양립 불가능론이 통제에 대한 과장된 생각에 기반하고 있다는 것이다. 이를 보이기 위해서 피셔는 자율성에 대한 다음의 파인버그의 주장에 먼저 주목한다.

자율적인 인간은 종종 “스스로 만들어진(self-made) 인간”으로 생각된다. 물론 자율적 인간은, 모순을 범하지 않고서는, 말 그대로 전적으로 스스로 만들어질 수 없다. ... 만약 우리가 진정성(authenticity)이 모든 원리들(믿음, 선호 등등)이 모든 결정의 순간에 이성적 견주어 새롭게 함께 조사될 것을 요구하는 것으로 생각한다면 합리적 반성과 같은 어떤 것은 시작될 수조차 없을 것이다.<sup>20)</sup>

우리가 일단 진정한 진정성이 말 그대로 스스로 만들어진 것이라고 볼 수 없다는 것을 받아들일 경우 자기 창조도 말 그대로의 의미를 갖는 것으로 해석되어서는 안 된다. 이런 이유로 파인버그는 “자기 창조는 가능

---

20) John Martin Fischer, “The Cards That Are Dealt You,” *The Journal of Ethics* 10 (2006): 109.

하지만 그것은 무에서부터의(*ex nihilo*) 창조가 아니다”라고 주장한다.<sup>21)</sup>

피셔는 무에서부터 나온 것으로서의 자율성 개념이 과장된 개념이며 일상적인 자율성 개념에 대한 적절한 기술로 간주될 수도 없다는 파인버그의 지적과 유사한 것을 원천 양립 불가능론자들의 주장에 대해 적용시킨다. 예를 들어, 원천 양립 불가능론자 가운데 한 명인 스밀란스키는, 앞에서 살펴 보았던 케인과 유사하게, 만약 인과적 결정론이 참이라면 우리가 우리의 행위와 결정에 대해서, 피상적이거나 얇은 통제와는 구분되는, 그리고 자유의지에 필수적인 것으로서의, **궁극적인 통제**를 할 수 없게 되며, 따라서 모든 발생하는 것은 결국 운의 문제가 된다고 생각한다. 그러나, 피셔가 보기에, 여기서 스밀란스키에 의해서 요구되고 있는 “궁극적인 통제”는 과장된 통제 개념에 의존하고 있다. 그는 다음과 같이 말한다.

우리가 결과에 영향을 미치는 모든 요소들에 대해서 통제할 수 있다고 가정하는 것은 터무니없는 환상이다. 완전한 통제는, 파인버그의 용어를 빌면, 자율성에 대한 “부풀려진(*inflated*)” 관념이다. (우리는 경제에서와 마찬가지로 자율성에 관해서도 인플레이션을 경계해야 한다!)<sup>22)</sup>

예를 들어, 내가 오늘 한 행동의 원인 가운데는 태양이 과거에 계속 빛나고 있었다는 것이라든지, 운석이 충돌하지 않았었다는 것과 같은 것도 들어 있다. 왜냐하면 태양이 계속 빛나지 않았다면, 오늘날 우리가 살고 있는 지구 환경은 없었을 것이고 따라서 내가 오늘 한 행동도 존재할 수 없었을 것이기 때문이다. 마찬가지로 거대한 운석이 며칠 전에 지구에 떨어졌다면, 아마도 나는 죽었을 것이고 따라서 나는 오늘 한 그 행동을 할 수 없었을 것이다. 그런데 우리가 이런 태양의 상태나 운석의

21) Ibid., 110.

22) Ibid., 120.

운동에 대해서 통제력을 갖고 있지 않다는 것은 너무나 명백하다. 그렇지만 내가 이런 원인들에 대해서 통제력을 갖고 있지 않다는 사실이 내가 오늘 한 행동을 자유로운 행동이 아니게 만들고, 그 행동에 대한 나의 책임을 면제시켜 버린다고 주장하는 것은 불합리하다. 그렇다면 자유의지나 도덕적 책임이 나의 행동에 대한 궁극적인 통제를 요구한다는 것은, 적어도 그 **궁극적인** 통제가 **전적인(total)** 통제를 의미한다면, 불합리한 것이 된다.

피셔의 이런 비판은 자체로는, 피셔 자신도 인정하듯이, 원천 양립 불가능론자들에 대한 결정적인 반론으로 작동할 수 없다. 왜냐하면 원천 양립 불가능론자들은 자신의 궁극적인 통제가 반드시 **전적인** 통제를 의미해야 한다고 주장할 필요가 때문이다. 피셔 자신도 (태양이 계속 빛나는 것과 같은 것에 대한 통제의 기여가 자유의지를 훼손시키지 않는다는) 자신의 반례들은 인과적인 **필요조건**에 관한 것이며, 따라서 만약 원천 양립 불가능론자들이 요구하는 궁극적인 통제가 모든 인과적 **충분조건**에 대한 통제라면 자신의 반례는 자체로 결정적인 것이 될 수는 없다는 것을 인정한다. 사실 피셔의 반례가 자체로는 결정적일 수 없다는 점은 우리가 앞서 검토했던 무책임의 순방향 전이원리2\*를 살펴보아도 분명하다. 무책임의 순방향 전이원리2\*에 따르면 (인류가 탄생하기 전의) 과거의 사건(들)이 현재의 나의 행동에 대한 **완전한(또는 충분한)** 설명이 될 경우 (신이냐, 어떤 인간이 아니면서도 책임을 질 수 있는 존재가 없다는 가정 하에) 나는 나의 행동에 대해서 책임을 질 수 없게 된다. 그러나 그것은 (내가 책임질 수 없는) 과거의 어떤 사건이 나의 행동에 **부분적인** 설명이 될 경우 나는 나의 행동에 대해서 책임을 질 수 없다는 것을 주장하는 것이 아니다. 원천 양립 불가능론자들은 이런 점에서 나의 행동의 모든 (부분적) 원천이 나 또는 내가 책임져야 할 것 때문에 발생했다는 것을 주장할 필요는 없다. 그러나 바로 이 지점에서 피셔는 이런 식으로 필요조건과 충분조건을 구분하는 것이 어떤 차이를 만들어

내는지 알기 어렵다고 주장하며 증명의 부담을 이 구분이 중요한 구분이라고 생각하는 사람들에게 넘긴다. 더 나아가 그는 이 구분이 왜 중요한 차이를 만들어 내는지에 대한 원리 지워진 설명이 제공된 바 없다는 믿음 하에서 다음과 같이 말한다.

마찬가지로 나는 실제 사건들의 진행에 있어서의 인과적으로 충분한 조건의 존재가 그것 때문에 행위자가 (굳건하게) 자신의 행위에 도덕적으로 책임을 지게 되는 특징을 방해하거나 밀어낼 필요는 없다고 말하고 싶은 생각이 든다. 나는 이 두 종류의 조건들, 즉 필요 조건과 충분 조건 사이에 이 맥락에서 대칭성이 존재한다고 믿는다.<sup>23)</sup>

지금까지의 내용을 요약하면, 피셔의 아이디어는 결국 다음과 같이 정리될 수 있다.

- (1) 나의 행동이 자유로운 행동이기 위해서 나의 행동에 대한 모든 인과적 **필요조건**들에 대해서 내가 통제할 수 있기를 요구하는 것은 과도한 요구이다.
- (2) 현재 맥락에서 인과적 필요조건과 인과적 충분조건 사이에 어떤 비대칭성을 설정할 좋은 이유가 없다.
- (3) ((1)과 (2)로부터) 나의 행동이 자유로운 행동이기 위해서 나의 행동에 대한 모든 인과적 **충분조건**들에 대해서 내가 통제할 수 있기를 요구하는 것은 과도한 요구이다.
- (4) 무책임의 순방향 전이 원리, 또는 책임의 역방향 전이 원리는, 나의 행동이 자유로운 행동이기 위해서 내가 나의 행동에 대한 모든 인과적 충분 조건들에 대해서 나의 통제, 또는 책임을 가질 것을 요구한다.
- (5) ((3)과 (4)로부터) 무책임의 순방향 전이 원리<sup>2)</sup>와 같은 것을 받아들이라고 요구하는 것은 과도한 요구이다.

---

23) Ibid., 123.

(6) 따라서 원천 양립 가능론 논변의 표준적인 형태는 과도한 요구에 기반해 있다.

필자들은 피셔의 논증에서 (2)가 잘못되었다고 생각한다. 그리고 피셔가 (2)를 주장하는 것은 인과와 설명이 중요하게 다르며, 도덕적 책임, 따라서 자유 의지는 인과와 연동하는 개념이기 보다는 설명과 연동하는 개념이라는 것을 인식하지 못했기 때문이라고 생각한다. 즉, 우리가 의존-설명-책임 클러스터 논제를 받아들일 경우 (2)는 참이 아니라는 것이다. 왜 그런지 알아보기 위해서 피셔의 주된 반례를 다시 살펴보기로 하자. 내가 오늘 한 행동의 하나의 인과적 필요 조건은, 피셔가 강조하듯이, 과거에 태양이 계속해서 빛났다는 것이다. 그러나 이 사실은 내가 왜 오늘 한 행동을 했는지에 대해서 설명적으로 유관하지는 않다. 즉, 이 경우 우리는 인과적 관계는 성립한다고 할 수 있지만 설명적 관계는 성립한다고 볼 수 없다. 왜 여기서 설명적 관계가 성립하지 않는지에 대해서는 다양한 설명이 있을 수 있다. 예를 들어, 우리는 이 경우 우리의 피설명항, 즉 오늘 내가 특정한 행동 A를 했다는 것이 오늘 내가 특정한 다른 행동 B를 하지 않고 A를 했다는 대조의 초점을 갖고 있는데, 이 대조의 초점과 관련해서 과거에 태양이 계속 빛났다는 것은 어떤 설명적 차이도 만들어 내지 않는다고 말할 수 있다. 태양이 빛나지 않았다면 나는 행동 A를 할 수 없었을 뿐만 아니라 행동 B도 할 수 없었을 것이기 때문이다. 또는, 과거에 태양이 계속 빛났다는 것은 너무나 당연히 전제되고 있는 사실이라서 현재의 설명의 맥락에서 설명적 역할을 할 수 없다고 말할 수도 있다.<sup>24)</sup> 그 현상에 대한 설명이 어떤 것이건, 현재의 맥락에서 과거에 태양이 계속 빛났다는 것이 내가 오늘 행동 A를 했다는 것에 대해서 설명적으로 유관하지 않다는 사실 자체는 명백하다. 이 점은 왜 우

24) 이와 관련해서 보다 자세한 설명은 이재호, “인과의 두 개념, 설명, 그리고 도덕적 책임.”를 참조할 것.

리가 인과적 충분조건과 인과적 필요조건을 “차별”해야 하는지에 대한 좋은 근거를 제공해 준다. 어떤 것이 어떤 피설명사건에 대한 인과적 필요조건이라는 것은 그것이 그 피설명사건에 대해서 설명적으로 유관하다는 것을 보장하지 않는다. 그런데 만약 우리가 의존-설명-책임 클러스터 논제를 받아들일 경우 자유의지와 책임 귀속에 관련해서 중요한 것은 인과적 유관성이 아니라 설명적 유관성이다. 따라서 인과적 필요조건은, 그것이 설명적으로 무관한 한에 있어서, 책임의 역방향 전이의 대상이 될 수 없으며, 따라서 그것에 대한 책임, 또는 통제의 결여가 원천 양립 불가능론에 어떤 문제를 일으키지 않는다.

불필요한 오해를 막기 위해서, 여기서 필자들이 인과적 충분 조건은 무조건 설명적으로 유관하다고 주장하는 것은 아니라는 것을 강조할 필요가 있다. 적어도 인과적 충분 조건이 그것이 성립할 경우 관련된 결과가 반드시 야기되게 되는 조건이라고 정의될 경우 필자들은 인과적 충분 조건이 설명적으로 반드시 유관하다고 주장할 생각은 없다.<sup>25)</sup> 이런 의미에서의 인과적 충분 조건은 종종 설명적으로 무관한 조건을 포함하고 있을 수 있으며, 설명적 유관성 관계는 설명적으로 무관한 요소의 존재에 민감하다.<sup>26)</sup> 따라서 인과적 충분 조건이 설명적으로 무관한 경우도 가능

25) 사실 인과적 충분조건이 이렇게 정의될 수 있는지는 의심스럽다. 어떤 사람들은 J.L. 맥키가 군더더기(redundant) 요소가 배제된 “최소 충분조건(minimal sufficient condition)”이라고 부른 것이 어떤 것이 충분 원인이 되기 위한 필요조건이라고 생각할 수도 있다. John L. Mackie, “Causes and Conditions,” in *Causation - Oxford Readings in Philosophy*, ed. Ernest Sosa and Michael Tooley (Oxford, 1993), 35. 현재 맥락에서 필자들의 관심은 인과적 충분조건이 어떻게 정의되어야 하는지에 있기 보다는, 인과적 충분조건이 어떻게 정의되든, 인과적 충분조건과 인과적 필요 조건 사이의 유의미한 구분이 가능하다는 것을 보이는데 있다.

26) 설명적 관계의 성립에 설명적으로 무관한 요소가 치명적이라는 사실을 잘 보여주는 대표적인 예는 햄펠의 D-N 모델에 대한 표준적인 반례로 여겨지고 있는 주문 걸린 소금의 예이다. 어떤 사람이 소금 위에 지팡이를 휘두르며 주문을 원 다음 그 소금을 물 속에 넣었는데 그 소금은 물에 녹았다. 왜 그 소금이 물에 녹았는지를 설명하기 위해서 주문을 건 소금을 물에 넣으면 항상 물에 녹는다는

하다. 그럼에도 불구하고, 인과적 필요조건과 인과적 충분 조건 사이에는 여전히 중요한 차이가 있다. 인과적 충분 조건은 반드시 설명적으로 유관한 요소들을 포함하고 있을 것이기 때문이다. 이를 앞의 예와 관련 지어 생각해 보면, 결정론이 참일 경우, 내가 오늘 한 행동 A에 대한 충분한 원인이 인류가 태어나기도 전의 사건들 가운데 있을 것이고, 그 충분한 원인을 이루는 사건들 가운데는, 그것이 무엇인지에 대해 우리가 알 수 있느냐의 여부와 별개로, 내가 왜 행동 B가 아니라 행동 A를 했는지에 대한 설명에 유관한 요소들이 있음에 틀림없다. 그리고 그 요소는 분명 현재의 맥락에서 당연한 것으로 치부될 수 있는 것은 아닐 것이다. 따라서 결정론이 참이라면 내가 제어할 수 없는, 하지만 나의 행동 A에 대해서 설명적으로 유관한 요소가 존재함에 틀림없다는 것이 나온다. 그런데 만약 우리가 인과적 충분 조건에 대한 통제를 전적으로 결여 한다면, 그것은 설명적으로 유관한 그 충분 조건의 요소에 대한 통제와 책임도 결여 한다는 것을 의미한다. 이럴 경우, 책임의 역방향 전이 원리에 의해서 그 인과적 충분조건에 의해서 야기되는 사건에 대한 책임도 제거된다. 요약하면, 인과적 필요조건에 대한 통제의 결여는 설명적으로 유관한 요소에 대한 통제의 결여를 보장하지 않지만 인과적 충분조건에 대한 통제의 결여는 설명적으로 유관한 요소에 대한 통제의 결여를 보장한다는 것이다.

## 5. 결론

피셔가 강조하듯이, 나의 행동의 하나의 원인은 태양이 과거에 계속 빛났었다는 것인데, 그 원인에 대해서 나는 어떤 통제도 책임도 갖지 않

---

것을 지적하는 것은 잘못된 것이다. 그 소금은 **주문이 결린 후** 물에 넣어졌기 때문에 물에 녹은 것이 아니라 물에 넣어졌기 때문에 물에 녹은 것이기 때문이다.

는다는 사실에 의해서 나의 그 행동이 자유의지에 의한 행동이 아니게 되는 것은 아니다. 이런 점에서 피셔의 주장대로 **궁극적인 통제**는 **전적인 통제**를 의미할 수 없다. 그러나 만약 나의 행동이 X **때문에** 발생한 것일 경우, X에 대해서 내가 어떤 통제도 책임도 갖지 않는다면, 나의 행동은 자유의지에 의한 행동이라고 말할 수 없다. 내가 한 행동이 자유의지에 의한 행동이기 위해서는 나 또는 내가 책임져야 할 것 **때문에** 그 행동이 발생해야 하기 때문이다. 이것이 우리가 의존-설명-책임 클러스터 논제를 받아들일 경우 자연스럽게 받아들여야 하는 논제이다. 그런데 인과적 관계는 설명적 관계보다 그 외연이 훨씬 크다. 이것에는 이중의 이유가 있다. 우선, (최소한 잠재적) 의존성을 결여한 인과 관계는 설명적 관계로 여겨지기 어렵고, 다음으로 의존성을 포함하는 인과 관계라고 하더라도 설명적 관계가 갖는 화용론적 인식적 차원들 때문에 설명적 관계로 여겨질 수 없는 경우도 있다.<sup>27)</sup> 바로 이것이 왜 우리가 인과적 충분조건과 인과적 필요조건을 차별해야 하는 지에 대한 진정한 이유이다. 인과적 필요조건은 첫 번째 이유에서 설명적 유관성을 결여하지는 않겠지만 두 번째 이유에서 설명적 유관성을 결여할 수 있다. 그리고 피셔의 원천 양립 불가능론에 대한 비판은 바로 이 점을 간과하고 있다. 물론 필자들의 이 비판은 온전히 피셔에게만 향할 수는 없다. 왜냐하면 케인과 스미란스키를 포함하는 대부분의 원천 양립 불가능론자들의 문헌에서도 원천 양립 불가능론의 옹호에 (인과 중심이 아니라) 설명 중심의 논의가 결정적인 역할을 한다는 것에 대한 명백한 인식은 드러나지 않기 때문이다.

---

27) 이에 대한 자세한 논의는 이재호, “인과의 두 개념, 설명, 그리고 도덕적 책임.”의 4절을 참고할 것.

## 참고문헌

- 이재호. “인과의 두 개념, 설명, 그리고 도덕적 책임.” 『과학철학』 19, no. 1 (2016): 51-84.
- Ayer, A. J. “Determinism, Free Will, and Responsibility.” Chap. 1 In *Free Will*, edited by Gary Watson, 15-23. Oxford, [Oxfordshire] ; New York: Oxford University Press, 1982.
- Beebe, Helen. “The Non-Governing Conception of Laws of Nature (Ramsey-Lewis Theory, Thought Experiments).” [In English]. *Philosophy and Phenomenological Research* 61, no. 3 (Nov 2000): 571-94.
- Chisholm, Roderick M. “Human Freedom and the Self.” Chap. 1 In *Free Will - Second Edition*, edited by Gary Watson, 26-37: Oxford, 2003.
- Fischer, John Martin. “The Cards That Are Dealt You.” *The Journal of Ethics* 10 (2006): 107-29.
- \_\_\_\_\_. “Compatibilism.” Chap. 2 In *Four Views on Free Will*, 44-84. Malden, MA ; Oxford: Blackwell Pub., 2007.
- \_\_\_\_\_. “Frankfurt-Type Examples and Semicompatibilism: New Work.” Chap. 12 In *The Oxford Handbook of Free Will 2nd Edition*, edited by Robert Kane, 243-65, 2011.
- Haji, Ishtiyaque. “Obligation, Reason, and Frankfurt Examples.” Chap. 14 In *The Oxford Handbook of Free Will 2nd Edition*, edited by Robert Kane, 288-305, 2011.
- Hall, Edward J. “Two Concepts of Causation.” Chap. 9 In *Causation and Counterfactuals*, edited by John David Collins, Edward J. Hall and L. A. Paul, 225-76. Cambridge, Mass.: MIT Press, 2004.

- Kane, Robert. *The Significance of Free Will*. New York: Oxford University Press, 1996.
- Mackie, John L. "Causes and Conditions." Chap. 1 In *Causation - Oxford Readings in Philosophy*, edited by Ernest Sosa and Michael Tooley, 33-55: Oxford, 1993.
- Pereboom, Derk. "Hard Incompatibilism." Chap. 3 In *Four Views on Free Will*, 85-125: Blackwell, 2007.
- Schaffer, J. "Causes as Probability Raisers of Processes." *Journal of Philosophy* 98, no. 2 (Feb 2001): 75-92.
- Vargas, Manuel. "Response to Kane, Fischer, and Pereboom." Chap. 8 In *Four Views on Free Will*, 204-19: Blackwell, 2007.
- Widerker, David. "Frankfurt-Friendly Libertarianism." Chap. 13 In *The Oxford Handbook of Free Will 2nd Edition*, edited by Robert Kane, 266-87, 2011.

## Explanation, Moral Responsibility, and Incompatibilism

Lee, Jaeho, Lee, Eunhong (Chung-Ang Univ.)

In this paper, we argue that once we accept “dependence-explanation-moral responsibility thesis,” according to which the notion of counterfactual dependence, explanation and moral responsibility comprise a conceptual cluster and that of production does not belong to this cluster, it becomes much easier to defend incompatibilism concerning free will. We first argue that anybody who wants to defend incompatibilism needs to defend ‘source incompatibilism’ rather than ‘leeway incompatibilism.’ Then we argue that the most influential criticism against source incompatibilism, namely J.M. Fisher’s criticism, can be easily undermined once we accept dependence-explanation-moral responsibility thesis.

Key words: Explanation, Moral Responsibility, Leeway Incompatibilism, Source Incompatibilism

이재호 E-mail: jaeho.jaeho@gmail.com

이은홍 E-mail: lunhong@daum.net

투 고 일	2016년 03월 28일
심 사 일	2016년 04월 27일
게재확정	2016년 05월 11일