

인과적 설명 이론의 딜레마와 평형 설명*

이 재 호[†], 박 재 희[‡]

이 논문에서 필자들은 E. 소버에 의해서 인과적 설명이론에 대한 반례로 제시된 평형 설명에 대한 최근의 인과적 설명 이론가들의 대응을 검토한다. 필자들은 인과적 설명이론이 본질적으로 가질 수밖에 없는 딜레마가 존재하며 평형 설명을 인과적 설명이론과 양립 가능하게 만들려는 시도들이 이 딜레마로부터 탈출할 수 없다는 것을 논증한다. 필자들은 이 과정에서 인과적 설명이론이라는 것이 무엇인지에 대한 보다 명확한 기준을 제시하려 하며 아울러 평형 설명이 만들어지는 보다 근본적인 메커니즘을 조명할 것이다.

주요어 : 인과, 설명, 인과적 설명 이론, 평형 설명

* 이 논문은 2023년도 중앙대학교 연구장학기금 지원에 의한 것임. 이 논문의 초고에 유익한 논평을 해준 두 명의 익명의 심사자에게 감사한다.

† 중앙대학교 철학과 교수 (jaeho.jaeho@gmail.com)

‡ 중앙대학교 철학과 석사과정 (parkjaehui0312@gmail.com)

1. 서론

스트레븐스가 지적하듯이 인과적 설명 이론은 오늘날 설명에 대한 통일 접근과 더불어 가장 영향력 있는 설명 이론이다.¹⁾ 그러나 모든 철학 이론들이 그렇듯이 인과적 설명 이론도 다양한 방식으로 도전을 받아 왔다. 이 논문에서 필자들은 인과적 설명 이론에 대해 강력한 도전을 제기하는 것으로 여겨져 온 E. 소버의 평형 설명(equilibrium explanation)에 대한 최근의 인과적 설명 이론가들의 대응을 검토한다. 필자들은 이 논문을 통해서 인과적 설명이론의 정체성에 대해서 보다 분명한 기준을 제시하고자 하며, 이것을 기반으로 인과적 설명 이론이 어떤 근본적인 딜레마를 갖는다는 것을 보일 것이다. 마지막으로 필자들은 평형 설명에 대한 최근의 인과적 설명 이론가들의 대응이 이 근본적인 딜레마로부터 탈출할 수 없다는 것을 보일 것이다. 이 과정을 통해서 필자들은 평형 설명이 작동하는 구조에 대한 보다 명확한 이해가 가능해질 수 있을 것이라고 생각하며 이것이 설명 이론 일반에 적지 않은 함축을 가질 것이라고 믿는다.

2. 인과적 설명 이론과 평형 설명

2.1 인과적 설명이론의 배경과 정체성

분석 철학적 전통 하에서 과학적 설명에 대한 본격적인 논의는 C. 험펠과 P. 오펜하임의 고전적인 논문(Hempel & Oppenheim 1948)에서 출발한다. 이 논문에서 제시된 D-N모델은 다양한 반례에 직면해 결국 붕괴했지만 설명과 관련한 철학적 논의에서 험펠의 영향력은 아직도 건재하다. 왜냐하면 험펠 이후의 사실상 모든 설명 이론들은 D-N 모델이 드러낸 다양한 문제점들을 해결하는 것을 그 이론의 출발점으로 삼고 있기 때문이다. 이 논문에서 다루고자 하는 인과적 설명 이론 역시 이 점에서 다르지 않다.

1) Strevens (2004), p. 154.

인과적 설명이론은 D-N 모델에 대한 반례를 만들어 내는 정형화된 레시피에 대한 통찰에서 출발한다. D-N 모델에 따르면 어떤 것을 설명하는 것은 그것이 예측 가능했다는 것을 보여주는 것이다.²⁾ 험펠은 이 아이디어를 설명이 기본적으로 피설명항을 법칙과 단칭 사실들로부터 연역적으로 도출하는 논증의 형태를 띤다는 모델을 제시하면서 정식화했다. 이 이론에 대해 반례를 만들어 내는 하나의 방식은 피설명항이 법칙과 단칭 사실로부터 도출되기는 하지만 그 피설명항이 언급되는 단칭 사건들과 인과적으로 연결되지는 않는 경우들을 찾는 것이다. 그런 경우들이 잘 알려진 소위 비대칭성(asymmetry) 사례, 무관성(irrelevance) 사례, 그리고 공통 원인(common cause) 사례와 같은 것들이다. 비대칭성 사례로 가장 잘 알려진 것은 브롬버거의 소위 '깃대 사례'(Bromberger 1966)이다. 깃대의 높이와 태양의 고도로부터 그림자의 길이를 도출하는 것은 설명적이지만 태양의 고도와 그림자의 길이로부터 깃대의 높이를 도출하는 것은 설명적이지 않다. 이 두 도출은 험펠이 주목하는 논리적 특징에 있어서 구분되지 않는다. 아마도 이 설명적 비대칭성에 대한 가장 자연스런 설명은 깃대의 높이는 그림자의 길이의 원인이 되지만 그림자의 길이는 깃대의 높이의 원인이 되지 않는다는 것일 것이다. 대표적인 무관성 사례인 키버그의 주문 걸린 소금 사례 역시 유사한 설명이 가능하다. 모든 주문 걸린 소금은 물에 넣을 경우 녹으므로 어떤 소금이 주문이 걸린 후에 물에 넣어졌다는 것으로부터 그 소금이 녹을 것이라는 것을 연역적으로 추론하는 데는 아무런 문제가 없다. 그러나 이 도출은 설명적이지는 않은데, 아마도 그것은 소금이 물에 녹는데 주문이 아무런 인과적 역할을 하지 못하기 때문일 것이다. 대표적인 공통 원인 사례인 기압계 사례 역시 다르지 않다. 기압계의 눈금이 급격히 떨어질 때마다 폭풍우가 오므로 기압계의 눈금의 급격한 하락으로부터 폭풍우를 도출하는 것에는 아무런 문제가 없으며 이 도출은 예측과 관련해서도 아무런 문제가 없다. 그럼에도 불구하고 기압계의 눈금의 급격한 하락을 통해

2) 이 점에서 험펠은 설명적 이해의 핵심이 법칙적 예측 가능성(nomic expectability)을 보여주는 것이라고 생각했다(Salmon 1989, p. 57).

서 폭풍우를 설명하는 것은 적절하지 않은데, 아마도 그것은 이 둘 사이에 상관 관계(correlation)는 존재하지만 인과 관계는 존재하지 않기 때문일 것이다.

헵펠의 이론에 반례를 만들어 내는 또 다른 방법은 인과적 연결은 존재하지만 법칙적 예측 가능성은 제공되지 않는 경우를 찾아내는 것이다. 아마도 이런 범주의 반례로 가장 잘 알려진 것은 스크라이븐의 부전마비 사례³⁾일 것이다. 치료되지 않은 매독 환자의 25% 정도만이 부전마비를 겪는다. 따라서 치료되지 않은 매독을 통해서 부전마비를 예측하는 것은 정당화되지 않는다. 그러나 우리는 치료되지 않은 매독을 통해서 부전마비를 설명할 수 있다고 생각한다. 아마도 이 직관에 대한 하나의 자연스런 설명은 치료되지 않은 매독이 부전마비의 (최소한 부분적인) 원인이라는 것일 것이다.

헵펠의 설명 이론이 인과적 관계를 고려하지 않으며⁴⁾ 그것을 이용해 다양한 반례를 만들어 낼 수 있다는 것은 설명적 관계의 분석에 인과적 고려가 들어가야 한다는 것을 보여주는 것으로 보이며 인과적 설명 이론은 이 아이디어에서 출발한다.⁵⁾ 앞으로 드러나겠지만, 인과적 설명 이론에는 다

3) Salmon (1989), p. 49.

4) 김재권에 따르면, 헵펠은 본인이 분석하고자 하는 설명이 인과적 설명이라고 생각했다. 그러나 그는 적어도 자신의 공식적인 이론에 인과 개념을 도입하지는 않았다(Kim 1999, pp. 7-8).

5) 헵펠의 이론이 인과적이지 않은 도출 관계를 설명적 관계로 허용하며 그것이 그의 이론에 문제를 일으킨다는 것으로부터 반드시 설명적 관계를 인과적 관계로 분석해야 한다는, 즉 인과적 설명이론을 받아들여야 한다는 결론이 나오는 것은 아니다. 인과적 설명이론을 받아들이지 않더라도 인과적이지 않은 도출 관계를 배제할 수 있는 적절한 장치가 있을 경우 헵펠의 문제는 해결 가능하다. 예를 들어 P. 키처(Kitcher 1989)는 인과적 설명 이론을 받아들이지 않지만 인과적이지 않은 도출 관계를 설명 관계로 인정하지 않는다. 키처는 통일적인 도출이 설명적 도출이며 인과라는 것은 결국 설명적 도출이라는 개념으로부터 파생되는 개념이라고 주장한다. 따라서 그의 이론에 따를 경우 통일적인 도출은 결국 인과적인 도출이 되어 (수학적인 설명이나 형이상학적인 설명이 아닌 헵펠이 주목하는 인과적 설명에서) 인과적이지 않은 설명적 도출은 존재하지 않게 된다.

양한 형태가 있을 수 있다. 그러나 인과적 설명이론의 기본적인 아이디어는 피설명항을 설명하는 것은 피설명항의 원인을 드러내는 것이라는 것이다. 이 아이디어에 따르면 왜-질문(why-question)은 무엇-질문(what-question)으로 이해된다. 즉, 어떤 사건이 왜 발생했는지를 묻는 것은 그 사건의 원인이 무엇인지를 묻는 것이다.

일단 설명적 관계가 기본적으로 인과적 관계라는 것을 받아들일 경우 우리는 다음의 두 논제를 받아들여야 하는 것으로 보인다.

설명→인과: 어떤 것이 어떤 사건의 발생에 대한 설명이기 위해서는 피설명항의 원인을 제시해야 한다.

인과→설명: 어떤 것이 어떤 사건의 원인을 제시한다면 이는 그 사건의 설명으로 성립한다.⁶⁾

이 두 논제는 일견 인과적 설명 이론의 자연스런 귀결로 보이지만 사실 예외를 허용한다. 특히 인과→설명 논제에 대한 예외는 꽤나 일상적이다.⁷⁾ 예를 들어 우리는 왜 이 건물에 화재가 발생했는가라는 질문에 “이 건물 주변의 공기에 산소가 존재했기 때문이다”라고 대답했을 경우 그 대답이 적절한 설명이라고 생각하지 않는다. 마찬가지로 (고문을 당한 것으로 의심되는) 철수에 대해서 누군가가 “철수는 심정지 때문에 사망했다”라고 말했다를 경우 우리는 그가 제대로 된 설명을 제공했다고 생각하지 않는다. 그리고 우리의 이런 설명적 직관은 우리가 산소의 존재가 화재의 원인이 아니

6) 앞으로 우리가 중점적으로 살펴볼 B. 스코우의 설명 이론 역시 이 논제를 받아들이는 이론이다. 그는 그의 (거의) 공식적인 이론을 다음과 같이 제시한다.

(T1) 필연적으로, R과 Q가 사실일 경우, 만약 R이 Q의 원인이라면 왜 Q인가에 대한 하나의 이유는 R이라는 것이며 왜 <왜 Q인가에 대한 하나의 이유가 R이다>인지에 대한 유일한 이유는 R이 Q의 원인이라는 것이다(Skow 2016, p. 38).

7) 설명→인과 논제에 대한 제한은 나중에 B. 스코우(Skow 2016)의 이론을 살펴보면 서 자연스럽게 논의될 것이다.

라고 생각하거나 철수의 사망의 가장 직접적인 원인이 그의 심정지라는 것을 부정하기 때문에 생기는 것이 아니다. 우리는 이런 사례들에서 산소의 존재나 철수의 심정지가 각각 건물의 화재와 철수의 사망의 원인이라고 생각함에도 불구하고 이들은 설명적이지 않다고 생각하는 것이다.

인과→설명 논제에 대해 이런 반례가 있을 수 있다는 것은 그러나 자체로는 이 논제가 인과적 설명 이론의 정체성을 구성할 수 없다거나 인과적 설명이론이 잘못된 것이라는 것을 보여주지 않는다. 왜냐하면 이런 경우들은 설명에 화용론적인 요소가 작동한다는 것을 보여주는 것이지 설명적 관계가 인과적 관계가 아니라는 것을 보여주는 것은 아니기 때문이다. 이는 우리가 다음과 같이 말할 수 있다는 사실을 생각해 보면 분명하다.

- (1) 이 건물의 화재가 산소의 존재 **때문에** 발생한 것은 사실이지만 그런 설명은 **현재 맥락에서 우리가 원하는 설명이 아니다.**
- (2) 철수가 심정지 **때문에** 죽었다고 말할 수 있지만 그런 설명은 **현재 맥락에서 우리가 원하는 설명이 아니다.**

요컨대 이런 경우들은 엄밀히 말해서 인과→설명 논제에 대한 반례라기 보다는 원인의 제시가 화용론적으로 적절한 설명을 보장한다는 논제, 즉 “인과→문맥상 적절한 설명” 논제에 대한 반례로 보는 것이 적절하다. 물론 어떤 사람들은 문맥상 적절하지 않은 ‘설명’은 진정한 의미에서 설명이라고 볼 수 없으며 설명적 관계는 본질적으로 화용론적 요소를 포함한다고 주장할 수 있다. 필자들은 이 주장이 불합리한 주장이라고 생각하지 않는다. 그럼에도 불구하고 인과적 설명 이론가들은 인과적 관계가 성립할 경우, 설명적 관계와는 별개로, 최소한 ‘때문에’ 관계는 성립한다고 주장해야 하며 이것마저 받아들이지 않을 경우 그들의 이론은 진정한 의미에서 인과적 설명이론이라고 불릴 수는 없다. 따라서 우리는 다음과 같은 방식으로 인과적 설명 이론의 하나의 필요 조건을 제시할 수 있다.

어떤 이론이 인과→설명 논제에 대해서 화용론적 제한 이상의 제한을 받

아들일 경우 그 이론은 진정한 의미에서 인과적 설명 이론이라고 불릴 수 없다.

여기서 '화용론적 제한 이상의 제한을 받아들인다'는 것은 (1), (2)와 같은 방식의 말을 할 수 없는 방식으로 인과→설명 논제에 대한 예외를 인정한다는 의미다.

2.2 소비어의 평형 설명과 인과적 설명 이론

E. 소비어는 그의 영향력 있는 논문(Sober 1983)에서 인과적 설명 이론에 대한 강력한 반례를 제시했다. 그가 사용하는 주된 예는 인구집단의 성비에 대한 R. A. 피셔의 설명이다. 많은 종(species)에 있어서 재생산 가능한 연령대의 성비는 1:1이다. 이 사실에 대해서 피셔는 다음과 같이 설명한다. 만약 어떤 집단의 성비가 1:1로부터 크게 벗어날 경우 더 적은 수의 성을 더 많이 생산하는 부모가 재생산적으로 더 적합하게 된다. 왜냐하면 그들이 생산하는 자식들이 재생산에 성공할 확률이 높아질 것이기 때문이다. 반면에 이 과정을 통해서 성비가 다시 1:1이 될 경우 특정한 성의 자식들을 더 많이 생산하는 부모가 가졌던 재생산적 이점은 사라진다. 이런 메커니즘을 통해서 인구집단의 성비는 1:1이라는 평형상태에 도달하고 이 상태는 지속된다.

소비어는 다음과 같이 피셔의 설명에서 피설명항의 인과적 역사가 언급되지 않는다는 것을 강조한다.

주어진 시간에 어떤 인구집단에서 관찰된 1:1 비율에 대한 인과적 설명은 아마도 그 집단의 어떤 초기 상태를 기술할 것이며 그 인구집단을 현재의 상태로 움직이게 만든 진화적 힘을 기술할 것이다. 그러나 피셔의 설명은 그런 것을 기술하지 않는다. 피셔의 설명은 사실 **실제** 초기 조건들과 **실제** 선택적 힘이 왜 문제가 되지 않는지를 보여준다. 초기의 성비가 어떻게 결과되었을 선택 압은 그 인구집단을 그것의 평형 상태로 움직였을 것이다. 인과적 설명은 설명

될 사건이 실제로 어떻게 산출되었는지를 보여주는 반면 평형 설명은 어떤 종류의 인과적 시나리오가 실제로 발생했는지와 무관하게 그 사건이 어떻게 발생했는지를 보여준다(Sober 1983, p. 202, 강조는 원저자).

이런 특징을 갖는 설명은 얼마든지 생각해낼 수 있다. 예를 들어 오목한 접시에 쇠구슬이 놓였다고 가정해 보자.⁸⁾ 이 구슬은 결국 오목한 접시의 가운데, 즉 가장 낮은 부분에 위치하게 될 것이다. 왜 그 구슬이 그 부분에 위치하게 되었는지에 대한 설명에서 그 구슬이 그 위치까지 도달하게 된 실제의 인과적 과정은 언급될 필요가 없다. 왜냐하면 그 구슬이 어느 위치에 어느 속도로 놓였든 (비록 다른 인과적 과정을 거치기는 했겠지만) 결과는 동일했을 것이기 때문이다. 적절한 설명은 오히려 구슬이 거처온 실제의 인과적 과정이 아니라 접시의 형태적인 특징, 그리고 중력의 작용 등을 언급할 것이다. 일반적으로 실제의 인과적 과정 인근이 동일한 결과를 갖는 현실화되지 않은 가능성 인과적 과정들로 채워져 있을 경우 평형적 설명이 갖는 특징을 갖는 설명이 발생하며 이런 설명에서 실제의 인과적 과정은 설명적으로 무력하게 된다.

소버는 평형설명이 인과적 설명이론에 반례가 될 수 있다고 생각하는데 그가 보다 구체적으로 염두에 두고 있는 것은 설명→인과 논제이다. 그는 “내가 ‘평형 설명’이라고 부르는 것은 인과적 요구(the causal requirement)에 대한 구분되는 종류의 반례를 제공해 준다”⁹⁾라고 주장하는데, 여기서 그가 ‘인과적 요구’라고 부르는 것은 “특정한 사건 발생에 대한 설명은 그것의 원인을 언급해야 한다”는 요구이며 이는 명백히 설명→인과 논제에 해당된다. 소버가 이런 식으로 문제를 설정한 영향 때문인지는 몰라도 실제 평형설명에 관한 이후의 논의는 주로 평형 설명이 설명→인과 논제에 대한 진정한 반론이 되느냐의 문제, 즉 이것이 비인과적 설명이냐는 문제에 맞추어 졌다.

8) 이와 유사한 예는 평형설명과 관련한 논의에서 종종 사용된다(Strevens 2019, p. 728).

9) Sober (1983), p. 201.

필자들은 평형설명론의 문제를 이런 식으로 논의하는 것이 이 문제의 진정한 철학적 함축을 간과하게 만들었다고 생각한다. 평형설명론은 분명 설명→인과 논제에 대한 반례로 제시되었지만 이것은 동시에 인과→설명 논제에 대한(잠정적) 반례이기도 하다. 위의 인용문에서 소비자 자신이 지적하듯이 피셔의 설명은 성비가 1:1이 되는 상태로 이어지는 실제의 인과적 과정이 왜 “문제가 되지 않는지”를 보여준다. 여기서 ‘문제가 되지 않는다’는 설명적 힘을 갖지 않는다는 것으로 이해하는 것이 자연스럽다. 따라서 평형설명론은 (a) 실제 인과적 과정은 설명적이지 않으며 (b) 진정한 설명은 인과적 과정을 언급하지 않는다는 것, 즉 인과→설명 논제와 설명→인과 논제가 모두 잘못된 것이라는 것을 (일견) 보여주는 사례이다.¹⁰⁾ 따라서 인과적 설명 이론가들이 평형 설명론을 자신들의 틀 안에서 수용하기 위해서는 평형설명론이 설명→인과 논제에 대한 반론이 되지 않는다는 것을 보이는 것에 그쳐서는 안되고 그것이 인과→설명 논제에 대한 반론이 되지 않는다는 것 역시 보여야 한다.

지금까지의 논의를 정리해 보면 다음과 같다. 정당하게 ‘인과적 설명이론’이라고 불릴 수 있는 이론은 설명→인과 논제와 (최소한 약화된) 인과→설명 논제에 대한 커미트먼트를 가져야 한다. 그리고 소비자의 평형설명론이 이 두 논제 모두에 대해서 (일견) 반례를 제공하고 있다. 따라서 인과적 설명 이론을 옹호하려는 사람들은 평형설명론이 이 두 논제 모두에 대해서 반례가 될 수 없다는 것을 보여야 한다. 이 논문의 이후의 부분은 다음과 같이 진행될 것이다. 우선 필자들은 3절에서 인과적 설명이론은 인과 개념에 호소하는 설명 이론이라는 점에서 태생적으로 한계를 갖는 이론이며 우리가 인과 개념을 깊이 생각해보는 경우 빠져나가기 쉽지 않은 딜레마를 갖게

10) 이 점에서 스페리-테일러의 평형 설명론에 관한 최근 연구는 기존의 연구보다 진일보한 점이 있다. 그는 다음과 같이 평형설명론에 대해서 말한다. “평형 설명론은 일반적으로 비인과적 설명으로 취급된다. 이 포괄적인 논제에는 두 개의 주장이 귀속된다. 첫 번째 주장은 평형설명론이 어떤 원인도 적시하지 않는다는 것이다. [...] 두 번째 주장은 평형 설명론이 [...] 원인을 적시한다고 하더라도 [...] 그런 원인들은 시스템의 행동을 설명하는데 무관하다는 것이다”(Sperry-Taylor 2021, p. 5577, 강조는 인용자들).

된다는 것을 보일 것이다. 그런 후에 4절에서 필자들은 평형설명의 문제에 대한 최근의 인과적 설명 이론가들의 대응들을 살펴 보면서 그들의 대응이 3절에서 주장된 인과적 설명이론의 딜레마로부터 벗어나기 어렵다는 것을 주장할 것이다. 필자들은 이를 통해 인과적 설명 이론이 애초에 별로 매력적이지 않은 이론이며 실제로 잘 작동하지도 않는 이론이라는 것을 주장할 것이다.

3. 인과적 설명 이론의 태생적 한계와 딜레마

3.1 인과적 설명 이론의 태생적 한계

인과적 설명이론이 갖는 매력은 2.1에서 설명된 바 있다. 인과적 설명이론은 험펠의 이론에 대해서 제기된 반례 상당수에 대해서 꽤나 만족스러워 보이는 해결책을 제시해 준다. 그러나 인과적 설명이론이 갖는 이런 매력에 도취되어 인과 개념을 사용하지 않는 것을 통해서 험펠의 이론이 얻게 되었던 장점을 간과하는 것은 현명하지 않다. 김재권이 지적하듯이, 험펠은 인과 개념을 여과 없이 사용하는 것이 “최소한 설명 개념 만큼이나 명료화가 필요한 개념을 통해 설명 개념을 설명하는 것”¹¹⁾이라고 생각했던 것으로 보인다. 그리고 이런 인과 개념의 불명료성에 대한 우려는 오늘날까지도 유효하다. 왜냐하면 오늘날에도 우리는 충분히 만족스러운 인과이론을 갖고 있지 않기 때문이다. 이런 점에서 인과적 설명 이론은 기본적으로 우리가 잘 알고 있지 못하는 개념을 우리가 잘 알고 있지 못하는 개념을 통해서 분석하는 이론이다.

어떤 사람들은 이것이 그렇게 심각하지 않은 문제라고 생각할 지 모른다. 그들은 우리가 만족스러운 인과이론을 갖고 있지 못하면서도 그것에 구애 받지 않고 S. 크립키의 인과적 지시 이론, A. 골드만의 인과적 지시 이론, R. 노직의 인과적 결정이론 등을 생산적으로 논의할 수 있듯이 우리

11) Kim (1999), p. 11.

는 만족스런 인과 이론을 갖지 않은 상태로 생산적인 방식으로 인과적 설명 이론을 논의할 수 있다고 주장할 지 모른다. 필자들은 인과 개념에 중립적으로 인과적 지시, 지식, 결정 이론 등이 주장될 수 있다는 생각이 그렇게 분명한 것은 아니라고 생각한다.¹²⁾ 그러나 이것이 사실이라고 하더라도 인과 개념에 중립적으로 인과적 설명이론이 만들어질 수 있다고 생각하는 것은 매우 어렵다. 예를 들어, 인과 개념이 설명 개념으로부터 파생된 개념이라는 키처식의 인과 개념¹³⁾은 애초부터 인과적 설명 이론과는 양립 가능하지 않다. 그리고 앞으로 인과적 설명이론의 딜레마에 대해서 이야기하면서 분명해지겠지만 인과 개념이 갖는 불명료성은 실제로 인과적 설명이론에 심각한 문제를 일으킨다.

험펠의 이론이 인과 개념을 설명 개념을 분석하는데 사용하지 않음으로써 갖게 된 보다 중요한 장점은 이 이론이 인과 개념을 배제하는 것을 통해서 통합적인 설명 이론으로 발전될 수 있는 잠재력을 갖게 되었다는 것이다. 험펠의 이론은, 비록 악명높은 28번 각주¹⁴⁾에서 논의된 기술적인 문제 때문에 법칙에 대한 설명 이론으로 실제 발전되지는 못했지만, 원리상 법칙의 설명에 대한 이론으로 확장될 수 있다. 예를 들어 우리는 뉴턴의 법칙을 통해서 케플러의 법칙을 설명할 수 있다고 생각하는데, 이 설명적 직관에서 케플러의 법칙이 뉴턴의 법칙으로부터 도출 가능하다는 사실이 중요한 역할을 하고 있는 것은 분명해 보인다.

더 나아가 험펠의 이론은 그것이 요구하는 법칙에 형이상학적 법칙이나 수학적 법칙을 포함시킬 경우 형이상학적 설명이나 수학적 설명에 대한 이

12) 예를 들어 김재권은 다음과 같이 말한다. “인과적 실재론과 비실재론 사이의 선택의 철학적 함축과 관련해서, 예를 들어, (지각, 기억, 지식, 행위, 사건 동일성, 지시, 시간, 지속, 속성, 그리고 그 외의 많은 것들에 대한) 소위 인과적 이론들 가운데 어떤 것이 인과에 대한 비실재론적인 개념 하에서 그것이 현재 향유하는 그럴듯함을 유지할 수 있는지를 생각해 보는 것은 흥미로운 질문이다”(Kim, 1987, pp. 237-8, 페이지는 루벤(Ruben 1993)의 것임).

13) Kitcher (1989), Sec. 8.

14) Hempel & Oppenheim (1948), p. 159.

론으로도 확장될 수 있다. 예를 들어 고통이라는 정신적 속성이 C-섭유 활성화라는 물리적 속성에 법칙적으로 수반한다고 가정해 보자.¹⁵⁾ 이 경우 우리는 철수의 고통 사건을 철수의 C-섭유 활성화 사건으로 설명할 수 있다고 생각한다. 그러나 정신적 속성과 물리적 속성 사이의 관계가 인과 관계라고 보기는 어렵다. 마찬가지로 우리는 소크라테스의 단집합의 존재를 소크라테스의 존재를 통해서 설명할 수 있다고 생각하지만 이들 사이의 관계를 인과 관계라고 생각하지는 않는다. 그러나 험펠의 D-N 모델은 최소한 원리적으로는 이런 설명적 관계에 대한 모델이 될 수 있다. 우리는 C-섭유 활성화와 소크라테스의 존재로부터 각각 고통과 소크라테스의 단집합의 존재를 법칙적으로 도출할 수 있기 때문이다.¹⁶⁾

인과적 설명 이론은 험펠의 이론이 갖는 이런 장점들을 포기하는 것을 통해서 험펠의 반례를 극복하고자 한다. 그럼으로써 인과적 설명이론은 스스로 플랜B가 되려 한다. 다른 것이 같다면 우리는 통합적인 설명 이론을 그렇지 않은 설명 이론보다 선호할 좋은 이유를 갖는다. 인과적 설명 이론은 통합적인 설명 이론이 실패할 때 비로소 주목받을 가치가 있는 이론이다. 인과적 설명 이론이 험펠의 통합적 이론의 몰락에 기대어 성행하게 되었다는 것은 이런 점에서도 우연의 일치가 아니다. 따라서 현재 논의되고 있는 통합적 설명이론¹⁷⁾이 성공적인 것으로 드러날 경우 인과적 설명 이론은 자동적으로 폐기될 수밖에 없다.

3.2 인과적 설명 이론의 딜레마

앞에서 우리는 인과적 설명이론은 애초부터 통합적 설명이론이 될 가능

15) 예를 들어, D. 차머스(Chalmers 1996)는 (의식으로서의) 고통이 물리적 속성에 법칙적으로 수반한다는 생각을 진지하게 발전시킨다.

16) 험펠의 이론이 갖는 이런 특징은 험펠의 후계자라고 할 수 있는 키처의 이론에게서도 발견되며 그는 수학적, 비인과적 설명을 자신의 이론 하에 포섭시킬 수 있다는 것이 인과적 설명이론에 대해서 갖는 자신의 이론의 강점이라고 강조한다(Kitcher 1989, p. Sec. 3).

17) 이런 이론의 예는 선우환 (2020), 이재호 (2023) 등이 있다.

성이 없는 기껏해야 플랜B 이론이라는 것을 지적했다. 물론 이것이 자체로 인과적 설명이론이 정당화될 수 없다는 것을 보여주지는 않는다. 플랜A가 성공하지 못했을 때 플랜B는 정당화될 수 있으며 아직까지 모두에게서 성공적이라고 인정받는 플랜A는 없다. 그러나 인과적 설명이론의 플랜B로서의 성공 가능성 역시 밝지 않다. 인과적 설명이론이 인과 개념에 호소하는 것으로부터 태생적으로 갖는 딜레마가 존재하기 때문이다.

이 딜레마는 인과에 두 개의 개념이 존재한다는 N. 홀의 입장에서 출발한다. 홀에 따르면 우리가 통칭해서 ‘인과’라고 부르는 개념에는 의존(dependence)과 산출(production)이라는 두 개의 개념이 있다.¹⁸⁾ 의존 개념에 따르면 인과 관계는 반사실적 의존성 관계이다. 즉, c가 e의 원인이라는 것은 c가 발생하지 않았다면 e가 발생하지 않았을 것이라는 것이다. 물론 이런 단순한 반사실적 의존성 개념을 사용하지 않는 반사실적 의존성 인과 개념도 있을 수 있다. 예를 들어 J. 우드워드(Woodward 2003)는 인과적 개입 하에서의 반사실적 의존성이라는 독특한 개념을 인과 개념을 분석하기 위해서 사용한다. 그러나 우드워드 역시 반사실적 의존성 개념을 인과 분석의 핵심으로 삼고 있다는 점에서 넓은 의미에서 의존 개념으로 인과 개념을 이해하고 있는 철학자라고 할 수 있다. 반면에 산출 개념은 이행성(transitivity), 지역성(locality), 내재성(intrinsicness)을 핵심 조건으로 갖는다. 여기서 이행성이라는 것은 c가 d의 원인이고 d가 e의 원인이라면 c는 d의 원인이라는 것이다. 지역성이라는 것은 원인과 결과는 (인과적 매개자들을 통해) 시공간적으로 연속된다는 것이다. 마지막으로 내재성이라는 것은 한 과정의 인과적 구조는 (법칙과 함께) 그것의 내재적인 (비인과적인) 성격에 의해서 결정된다는 것이다.¹⁹⁾ 요컨대 산출로서의 인과 개념에 따르면 원인과 결과의 관계는 내재적으로 결정되는 인과적 과정을 통해서 시

18) 홀의 이론은 다음을 참고할 것. Hall, (2004). J. 쉐퍼 역시 후과 유사한 구분, 즉 확률 증가자(probability raiser)와 과정 연결자(process linkage)의 구분을 도입한다(Schaffer 2001).

19) Hall (2004), p. 225.

공간적으로 연결된 사건들 사이의 관계가 된다.²⁰⁾ 물론 우리가 일상적으로 인과 관계에 있다고 생각하는 사건들은 많은 경우 의존 개념에 따라서도 인과 관계에 있고 산출 개념에 따라서도 인과 관계에 있다. 그러나 이 두 개념은 전적으로 구분되는 별개의 개념이다. 반사실적 의존성 관계는 이행적이지도 않고 지역적이지도 않으며 내재성을 갖지도 않는다.

많은 철학자들이 그렇듯이 우리도 이 두 개의 인과 개념을 구분하기로 하자. 그럴 경우 인과적 설명 이론가들은 자신의 설명 이론에 의존으로서의 인과 개념을 사용하거나 산출로서의 인과 개념을 사용해야 한다.²¹⁾ 문제는 인과적 설명 이론가들이 어떤 길을 가던 성공적인 인과적 설명이론을 만들어 낼 수 없다는 것이다.

우선 의존 개념을 이용한 인과적 설명 이론을 생각해 보자. 의존 개념으로서의 인과 개념과 인과적 설명 이론 사이의 궁합(?)이 좋다는 것은 많은 인과적 설명 이론가들이 의존 개념으로서의 인과 개념을 선호한다는 사실만 봐도 알 수 있다. 예컨대 D. 루이스는 자신의 인과적 설명 이론이 특정한 인과 이론을 전제로 하고 있지 않다고 밝히지만²²⁾ 그가 반사실적 의존 개념을 통해서 인과 개념을 분석하려고 하는 대표적인 학자라는 것은 잘 알려져 있다. J. 우드워드 역시 위에서 지적된 바 있듯이 넓은 의미에서 반사실적 인과 개념을 옹호하고 있는 사람이다. 그리고 앞으로 논의될 B. 스코우 역시 넓은 의미에서 의존성 계열의 인과 개념을 기반으로 해서 그의 인과적 설명 이론을 옹호한다. 설명 개념에 의존 개념이 산출 개념보다 밀접하게 연결되어 있다는 것은 의존 인과는 성립하지만 산출 인과가 성립하지 않는 사례들을 생각해 보면 보다 분명하게 드러난다. 이런 경우로 가장 잘 알려져 있는 사례는 소위 부재 인과 사례와 이중 방지 사례이다. 철수

20) W. 새먼 또는 P. 도우의 인과 이론이 산출 개념에 입각한 인과 이론의 대표라고 생각될 수 있다(Dowe 2000; Salmon 1984).

21) 두 개념을 모두 사용하는 것도 논리적으로 가능하지만 여기서는 그런 경우는 논의하지 않을 것이다. 두 개념을 모두 사용하는 것이 앞으로 필자들이 제시하는 딜레마를 극복하게 해줄 가능성은 별로 없기 때문이다.

22) Lewis (1986), pp. 213-7.

는 여름 휴가 기간 집을 비웠으며 그가 휴가에서 돌아왔을 때 그의 집에 있던 화분이 말라 죽었다고 가정해 보자. 그가 화분에 물을 줬다면 그 화분은 말라죽지 않았을 것이므로 그가 화분에 물을 주지 않은 것과 그 화분이 말라 죽은 것 사이에는 의존 인과 관계가 성립한다. 그러나 그가 화분에 물을 주지 않는 사건은 기껏해야 부정적 사건이므로 그 사건과 화분이 말라죽는 사건 사이에 적절한 과정 연결은 존재하지 않는다. 그럼에도 불구하고 우리는 그 화분이 말라죽는 사건을 철수가 물을 주지 않은 것으로 설명할 수 있다는 강한 직관을 갖는다. 이중 방지의 경우에도 우리의 직관은 유사하게 작동한다. 철수가 적군의 기지를 폭격하기 위해서 폭격기를 몰고 출격했다. 영수는 철수의 폭격기를 엄호하기 위해서 전투기를 몰고 출격했다. 철수와 영수가 목표지점에 다가가자 적국의 요격기가 출동해 철수의 폭격기를 요격하려고 시도했다. 그러자 영수가 재빨리 적국의 요격기를 격추해 철수는 무사히 폭격 임무를 완수하게 되었다. 이 상황에서 영수의 요격기 격추 사건과 철수의 폭격 성공 사이에는 반사실적 의존성 관계가 성립한다. 영수가 요격기를 격추하지 않았다면 철수의 폭격기는 요격되었을 것이며 폭격 임무는 실패로 돌아갔을 것이다. 그러나 영수의 요격기 격추 사건과 철수의 폭격 성공 사건 사이를 이어주는 적절한 물리적 과정은 존재하지 않는다. 이 상황에서 우리의 설명적 직관은 명확하다. 철수는 영수 덕에 (때문에) 폭격에 성공한 것이다.

의존 인과 개념이 인과적 설명 이론과 궁합이 잘 맞는 것은 사실이지만 의존 인과 개념은 만족스런 인과 이론이 되기 힘들다는 치명적인 문제를 갖는다. 홀이 지적하듯이 우리의 인과 개념에 보다 중심적인 것은 의존이라기 보다는 산출이다.²³⁾ 산출 인과는 존재하지만 의존 인과는 존재하지 않는 대표적인 경우인 선취 사례를 생각해 보자. 철수와 영수가 꽃병을 향해 동시에 돌을 던졌는데 철수의 돌이 먼저 도착해서 꽃병을 깬고 영수의 돌은 허공을 갈랐다고 가정해 보자. 이 경우 철수가 돌을 던지는 사건은 꽃병이 깨지는 사건과 의존인과 관계를 갖지 않는다. 철수가 돌을 던지지 않

23) Hall (2004), p. 256.

았다고 하더라도 그 꽃병은 영수가 던진 돌에 의해서 깨졌을 것이기 때문이다. 이 사례에서 우리의 인과적 직관은 확고하다. 철수가 돌을 던진 사건이 그 꽃병이 깨진 사건의 원인이다. 그러나 이 사례에서 우리의 인과적 직관과 설명적 직관을 비교해 보면 설명적 직관은 인과적 직관만큼 분명하지 않다는 것을 발견하게 된다. 철수가 돌을 던지지 않았다고 하더라도 어차피 그 꽃병은 (영수의 돌에 의해) 깨질 것이었기 때문에 철수가 돌을 던졌기 때문에 그 꽃병이 깨졌다는 것은 그렇게 분명하지 않다. 반면에 우리의 설명적 직관이 분명한 부재 사례나 이중 방지 사례에서는, 홀이 지적하듯이, 우리의 인과적 직관은 불분명하다. 요컨대 의존 없는 산출 사례에서 우리의 인과적 직관은 확고하지만 설명적 직관은 불분명하며 산출 없는 의존 사례에서 우리의 설명적 직관은 확고하지만 인과적 직관은 불분명하다.

인과에 대한 반사실적 의존성 이론의 역사를 살펴 보면 의존 개념에 호소하는 인과 이론의 난점은 더욱 분명해진다. 선취 문제를 해결하기 위해 루이스가 다양한 해결책을 제시했지만 그 어떤 해결책도 만족스럽지 않았다는 것은 잘 알려져 있다. 아마도 우드워드도 단순한 반사실적 의존성을 거부하고 인과적 개입 하에서의 반사실적 의존성 개념을 인과 개념을 분석하기 위해서 사용하는 것의 하나의 동기는 선취 상황에서 이 개념이 보다 잘 작동한다는 믿음이었을 것이다. 우드워드는 다음과 같이 실제적인 인과(AC) 개념을 정의한다.

(AC)

(AC1) X의 현실값 = x 그리고 Y의 현실값=y

(AC2) 그 위에 있지 않은 Y의 직접 원인들 Z_i 가 실제 값에 고정되었을 때, X에 대한 개입이 Y 값을 변화시키는, 그런 X에서 Y로의 경로R이 적어도 하나 존재한다.

그러면 $X=x$ 라는 것이 $Y=y$ 라는 것의 실제 원인일 필요충분조건은 (AC1)과 (AC2)가 모두 만족된다는 것이다(Woodward 2003, p. 77).

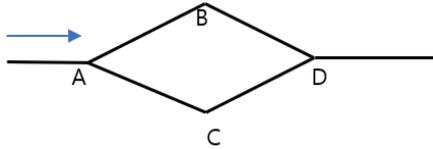
이 정의의 핵심 아이디어는 c가 e의 원인이라는 것은 c에서 e로 이어지

는 어떤 인과적 경로(R)가 존재하며 이 경로 밖에 있는 것들을 현실의 값으로 고정할 경우 c와 e사이에 반사실적 의존성이 성립한다는 것이다. 이 정의를 받아들일 경우 어떤 선취 사례들이 잘 다루어질 수 있다는 것은 사실이다. 예를 들어 위의 돌던지기 사례에서 영수의 돌이 꽃병과 충돌하지 않았다는 사실을 고정할 경우 철수의 돌던지기 사건은 꽃병이 깨지는 사건과 반사실적 의존성 관계를 갖게 되며 의존 인과 관계에 있을 수 있게 된다. 우드워드의 이론이 정말로 모든 선취 사례들에 대처할 수 있는지는 사실 분명하지 않다.²⁴⁾ 그러나 현재의 맥락에서 보다 중요한 문제는 우드워드의 이론이 정말로 모든 선취 사례에 대해서 대처할 수 있다고 하더라도 그것이 의존 개념과 결합한 인과적 설명이론의 전망을 밝게 해주지는 않는다는 것이다. 왜 그런지 알기 위해 의존 없는 산출의 또 다른 경우인 스위치 사례를 생각해 볼 필요가 있다. 스위치 사례는 실제 원인이 대안적 원인의 발생을 차단한다는 점에서 일종의 선취 사례로 볼 수도 있지만 전형적인 선취 사례, 즉 위의 돌던지기 사례와는 구분되는 뚜렷한 특징을 갖는다. 스위치 사례에서는 선취하는 원인이 발생하지 않을 경우 발생할 사건, 즉 선취하는 원인이 발생하지 않는 최근접 세계에서 선취하는 원인을 대체하는 사건이 자체로 선취하는 원인이 야기하는 결과와 동일한 결과를 야기한다. 따라서 선취하지 않는 원인이 발생하지 않을 경우 발생할 사건이 바로 선취되는 원인이 되며 전형적인 선취 사례와는 달리 선취하는 원인은 선취되는 원인을 논리적으로 차단한다. 이제 다음의 전형적인 스위치 사례를 생각해 보자.

(철길 사례) 기술자 한 명이 철도 교차로(A지점) 스위치에서 있다. 이 스위치는 왼쪽으로 가 있거나 오른쪽으로 가 있어야 하는데 현재 오른쪽에 있다. 만약 이 기술자가 그 스위치를 왼쪽으로 밀면 다가오고 있는 열차는 위쪽 철도로 진행하게 되며 B지점을 거쳐 시간

24) 우드워드의 이론에 대한 비판은 다음을 참고할 것. 선우환 (2020), 7장. 트럼핑 형태의 선취 사례에 대한 우드워드와 쉐퍼의 논쟁은 다음을 참고할 것. Woodward (2003), p. 383, 미주 45.

t에 D에 도달하게 된다. 만만 이 기술자가 스위치를 오른쪽으로 놔 둔다면 다가오고 있는 열차는 아래쪽 철도로 진행하게 되고, C지점을 거쳐 t에 D에 도달하게 된다. 실제로 그 기술자는 스위치를 왼쪽으로 밀었고 열차는 B지점을 지나 D에 도착했다(Hall 2004, p. 187) (이재호 2023, p. 11).



이 상황에서 기술자가 스위치를 왼쪽으로 미는 사건과 열차가 D에 도착하는 사건 사이에 반사실적 의존 관계는 없다. 스위치를 왼쪽으로 밀지 않아도 열차는 C를 경유해 D에 도달했을 것이기 때문이다. 그러나 이 두 사건 사이에 물리적 과정을 통한 연결은 존재한다. 이 상황에서 우드워드의 이론은 기술자가 스위치를 왼쪽으로 미는 사건이 열차가 D에 도착하는 사건의 원인이 된다고 판정한다. 왜냐하면 열차가 C를 경유하지 않았다는 사실을 고정할 경우 만약 그 기술자가 스위치를 왼쪽으로 밀지 않았다면 그 열차는 D에 도달하지 않았을 것이기 때문이다. 우드워드의 이론이 이런 판결을 내린다는 것이 자체로 우드워드의 **인과** 이론에 문제를 일으키지는 않는다. 비록 우리의 인과적 직관이 이 경우에 완벽하게 명료한 것은 아니지만 우리는 상당한 정도로 스위치를 왼쪽으로 미는 사건이 열차가 D에 도달하는 것의 원인이 된다는 직관을 갖고 있기 때문이다. 실제로 열차는 A-B-D의 경로를 거쳐 D에 도달했으며 이 경로가 인과적인 프로세스인 것은 분명하기 때문이다.²⁵⁾ 그러나 이런 결론을 갖는 우드워드의 인과 이론은 **인과적 설명이론과 결합될 수 있는** 종류의 인과 이론은 될 수 없다. 왜냐하면 이 스위치 사례에서 우리는 스위치를 왼쪽으로 밀었기 **때문에** 열차가 D에 도달한 것은 아니라는 분명한 설명적 직관을 갖고 있기 때문이다. 현재의 철길의 구조상 열차는 D에 도달할 수밖에 없게 되어 있다. 따라서 이 상

25) 이 사례에 대한 보다 자세한 분석은 다음을 참고할 것. 이재호 (2023).

황에서 A 지점에서 스위치가 어떻게 되어 있는지는 전혀 중요하지 않다. 중요한 것은 철길이 전체적으로 어떻게 연결되어 있느냐이다. 따라서 의존 인과 개념에 대한 우드워드의 변형은 의존 인과 개념의 인과 개념으로서의 적절성은 향상시켜 줄 수 있을 지 모르나 의존 인과 개념의 인과적 설명 이론과의 궁합을 악화시킨다.

이제 인과적 설명이론의 딜레마에서 첫번째 뿔을 이루는 (하위) 딜레마가 완성되었다. 만약 인과적 설명 이론이 의존 인과 개념과 결합하게 된다면, 의존 인과 개념의 인과 개념으로서의 적합성 때문에 문제가 발생하거나, 그 문제를 해결하려고 우드워드처럼 인위적인 장치를 추가할 경우 이 조합의 가장 큰 매력인 의존 인과 개념과 인과적 설명 이론의 궁합이 훼손된다.

인과적 설명이론의 딜레마의 두번째 뿔의 내용은 첫번째 뿔에 관한 논의에서 이미 대부분 설명되었다. 우리는 위에서 의존 인과 개념이 적절한 인과 개념이 되기에는 너무나 좁은 외연을 갖는다는 것을 지적했다. 선취 상황이 보여주듯이 의존 관계가 성립하지 않으면서도 명백하게 인과 관계가 존재하는 경우들이 있기 때문이다. 따라서 의존 인과 개념은 적절한 인과 이론이 되기에는 너무나 제한적인 이론이다. 이에 반해서 산출 개념은 매우 허용적인 인과 개념이다. 산출 개념에서 중요한 것은 연속적인 물리적 프로세스에 의한 연결인데 이런 연결은 매우, 보다 적절하게는 너무, 흔하다. 예를 들어 남대문에 화재가 나는 사건은 태양이 빛을 방출하는 사건과 물리적으로 연결되어 있다. 왜냐하면 태양에서 방출된 광자가 남대문에 도달하는 연속적인 물리적 과정이 존재하기 때문이다.²⁶⁾ 이런 허용성은 인과 개념으로서의 산출 개념에 심각한 문제를 제공하는가? 그것은 우리가 얼마나 인과 개념과 설명개념을 밀접하게 보느냐에 따라 달라진다. 우리가 소비의 사례에서 살펴본 것처럼 설명적이지 않은 인과를 받아들일 충분한 준비가 되어 있다면 이런 허용성은 크게 문제되지 않을 수 있다. 세계는

26) 남대문 화재 사건이 밤에 발생했다는 것은 이 사실에 영향을 주지 않는다. 남대문 화재가 발생하던 밤에 태양에서 방출된 일부 광자는 달에서 반사되어서 남대문에 도달하였을 것이다.

정말 많은 인과적 프로세스들로 연결되어 있으며 이것들 가운데 어떤 것은 설명적이며 어떤 것은 설명적이지 않을 수 있다는 생각은 그렇게 이상한 생각은 아니다. 실제로 산출 인과를 옹호하는 대표적인 철학자 가운데 하나인 W. 새먼은 이런 생각을 받아들인다. 그는 위의 사례와 유사한 사례를 통한 C. R. 히치콕의 비판에 대해서 다음과 같이 말한다. “히치콕의 분석의 결과로 나는 이제 인과적 프로세스의 연결이 통계적 유관성 관계를 결여할 경우 설명적 힘도 결여하게 된다고 말한다.”²⁷⁾ 그러나 인과와 설명을 분리해 인과가 설명적 관계가 성립하기 위한 하나의 필요조건으로만 생각하는 것은 인과→설명 논제를 부정하는 것이며 이는 인과적 설명 이론을 포기하는 것이다.

산출 인과는 매우 허용적인 개념이지만 동시에 매우 제한적인 개념이다. 우리가 위에서 살펴보았듯이 산출 없는 의존 상황인 부재 인과나 이중 방지 상황에서 우리는 인과 관계가 성립한다는 무시할 수 없는 직관을 갖고 있기 때문이다. 이런 제한성은 얼마나 심각한 문제인가? 다시 한 번 이것은 우리가 인과와 설명을 얼마나 밀접한 개념으로 보느냐에 따라 달라진다. 산출 인과 개념을 옹호하는 또 하나의 대표적인 철학자인 P. 도우는 이런 문제를 ‘준인과(quasi-causation)’라는 개념을 도입해 해결하고자 한다. 그는 원인 또는 결과가 부정적인 사건이거나 원인과 결과가 모두 긍정적인 사건이라고 하더라도 그 중간에 부정적인 사건이 이 긍정적인 사건들을 매개하는 유사 인과 상황을 준인과 상황이라고 정의하며 부재 인과나 이중 방지 상황이 그런 의미에서 준인과 상황이라고 생각한다.²⁸⁾ 우리는 사건의 부재가 어떤 것의 원인이 될 수 있다는 생각이 의심스럽다는 무시할 수 없는 직관을 갖고 있으며 따라서 인과와 준인과를 구분하는 것은 그렇게 이상한 것은 아니다. 그러나 이런 방식으로 산출 인과 개념을 구제하는 것에 분명한 대가가 있다. 위에서 살펴보았듯이 준인과 상황에서 우리는 준

27) Salmon (1997), p. 476.

28) Dowe (2001), p. 217. 위의 이중방지 사례에서 철수의 출격사건과 철수의 폭격 성공 사건 사이에는 철수의 폭격기의 요격사건의 부재가 매개되어 있으므로, 도우의 이론에서는, 준인과 상황이 된다.

원인이 설명적 힘을 갖는다는 확고한 설명적 직관을 갖고 있기 때문이다. 도우 자신도 다음과 같이 말하면서 이것을 부정하지 않는다. “부정적인 사건들은 수단과 목적으로 기능할 수 있을 뿐만 아니라 확률을 증대시킬 수 있으므로 이들은 이들의 준결과와 준원인에 대한 증거가 될 수 있다. 그리고 이들은 설명 안에 나타날 수도 있다.”²⁹⁾ 그러나 이렇게 준인과가 설명적 힘을 가질 수 있다고 인정하는 것은 설명→인과 논제를 부정하는 것이며 따라서 인과적 설명이론을 부정하는 것이다.

이제 인과적 설명이론의 딜레마의 두번째 뿔이 완성되었다. 산출 인과 개념은 인과 개념으로서의 적절성에서 어떤 심각한 문제가 발생하지는 않는다. 그러나 산출 개념의 이런 인과이론으로서의 적절성은 인과적 설명이론의 포기를 전제로 한다. 즉, 산출 인과 개념은 애초부터 인과적 설명이론과 궁합이 맞지 않는 이론이다.

지금까지의 논의를 정리하면 다음과 같다. 인과 개념에는 의존 개념과 산출 개념이 있다. 의존 개념은 인과적 설명이론과 궁합이 잘 맞지만 자체로 인과 이론으로서의 적합하지 않은 개념이다. 반면에 산출 개념은 인과이론을 구성하기에는 큰 문제가 없는 이론이지만 기본적으로 인과적 설명이론과는 궁합이 맞지 않는다. 따라서 우리는 적절한 인과 이론과 인과적 설명이론을 동시에 가질 수 없다. 이제 필자들은 다음절에서 이 딜레마가 평형인과 문제에 대한 인과적 설명이론가들의 대응에서 그대로 재현된다는 것을 밝힐 것이다.

4. 평형 설명에 대한 인과적 설명 이론가들의 대응

4.1 스코우의 대응—의존개념에의 호소

스코우의 평형 설명에 대한 대응은 기본적으로 평형 설명이 설명→인과 논제에 대한 반례가 될 수 없다는 것을 보이는데 맞춰져 있다. 앞서 지적했

29) *Ibid.*, p. 225, 강조는 인용자.

듯이 이것은 소비 자신이 문제를 이렇게 세팅했기 때문이다. 그러나 스코우의 대응에는, 비록 그것이 큰 비중을 갖고 있는 것은 아니지만, 평형 설명이 인과→설명 논제에 대한 반례가 될 수 없다는 내용도 포함되어 있다. 필자들은 평형 설명이 설명→인과 논제에 대한 반례가 될 수 없다는 스코우의 주장은 어느정도 호소력을 갖고 있다고 판단한다. 그러나 평형 설명이 인과→설명 논제에 대한 반례가 될 수 없다는 그의 주장은 전혀 호소력을 갖고 있지 않다고 생각한다. 그의 논증을 살펴보자.

스코우의 논증에 있어서 핵심적인 것은 다음의 세 가지이다.

(3) 인과에 대한 비례성 접근

(4) 왜-질문에 대한 좋은 대답(answer)과 좋은 응답(response)의 구분

(5) 설명에서 하위 수준과 상위 수준의 구분

S. 야블로 등이 옹호하는 인과에 대한 비례성 접근의 핵심 아이디어를 스코우는 다음과 같이 설명한다.

다른 것이 같다면, 어떤 사실 C는, E가 발생하는 것을 확실하게 하기 위해서 C 이상의 어떤 것도 필요하지 않고 동시에 C 자체가 E가 발생하는 것을 확실하게 하기 위해서 필요로하는 것 이상의 것이 아닐 경우에 한해[only if] 어떤 사실 E의 원인이 된다. 이 아이디어를 이해하기 쉽게 만드는 하나의 방식은 반사실적 용어를 사용하는 것이다. 즉, C는 (i) C-가 발생했지만 C는 발생하지 않았다 하더라도 E가 여전히 발생했을 그런 보다 약한 사실 C-가 존재하지 않고, (ii) 모든 더 강한 사실 C+는 C가 발생하고 C+가 발생하지 않는다고 하더라도 E는 여전히 발생했을 것이라는 속성을 가질 경우에 한해[only if] E의 원인이 된다는 것이다(Skow 2016, pp. 68-9, 강조는 인용자).

인과에 대한 반사실적 의존성 분석과 인과에 대한 비례성 접근은 중요한 유사점과 중요한 차이점이 있다. 우선 후자는 전자와 달리 자체로는 인

과개념에 대한 분석으로 제시된 것이 아니다. 명민한 독자는 위의 인용문에서 'if and only if' 가 사용되지 않고 'only if' 만 사용되었다는 것을 발견할 것이며 또한 '다른 것이 동일하다면' 구절이 들어간 것을 발견했을 것이다. 즉, 비례성 조건은 인과에 대한 필요 조건으로 제시되었으며 그것도 보편적으로 작동하지 않는 조건으로 제시되었다. 왜 이런 제한이 필요한가? 스코우 자신이 밝히듯, 선취적 원인들은 그것의 결과들에 비례적이지 않기 때문이다.³⁰⁾ 선취적 원인들은 그것이 발생하지 않더라도 동일한 결과가 발생한다. 따라서 비례성은 여기서 성립하지 않는다. 그러나 비례성 접근은 반사실적 의존성 분석과 유사점 역시 갖는다. 위의 인용문에서 드러나듯이 비례성은 반사실적 의존성 개념을 통해 이해될 수 있으며 비례성이 성립하기 위해서는 단순 반사실적 의존성도 성립해야 한다. 만약 c가 발생하지 않아도 e가 발생할 수 있다면 c는 e에 대해서 비례적이지 않다. 이런 점에서 인과에 대한 비례성 접근은 넓은 의미에서 인과에 대한 의존 개념적 접근에 포함된다.

다음으로, 왜-질문에 대한 좋은 대답과 좋은 응답이 구분되어야 하는 것은, 스코우에 따르면, 일반적으로 질문에 대해서 좋은 대답과 좋은 응답이 구분되어야 하기 때문이다. 예를 들어 '철수가 파티에 오니?'라는 질문에 대해서 '그는 몸이 아파'라고 말하는 것은 엄밀히 말해서 대답(answer)이 될 수 없다. 이 질문에 대한 진정한 대답은 '철수가 온다(예)' 이거나 '철수는 오지 않는다(아니오)'여야 한다. 그럼에도 불구하고 이것은 좋은 응답(response)일 수는 있다. 왜냐하면 이것을 통해서 질문을 던진 사람은, 그라이스의 화용론적 격률에 호소해³¹⁾, 좋은 대답, 즉 '오지 않는다'를 스스로 알아낼 수 있게 되기 때문이다.³²⁾ 마찬가지로의 구분이 왜-질문에 대해서도

30) Skow (2016), p. 68, 각주 5.

31) 스코우는 여기서 작동하는 그라이스의 격률이 유관성의 격률(the maxim of relation)이라고 지적한다. 우리는 대화 상황에서 유관한 것만을 말해야 하는데, 철수가 아프다는 사실은 그것이 철수가 파티에 오지 않게 만들지 않았을 경우 현재 대화에서 유관성을 잃는다(*Ibid.*, p. 73).

32) *Ibid.*, pp. 72-3.

성립한다. '왜 E가 발생했는가'에 대한 좋은 대답은 'C 때문에 E가 발생했다'의 형식을 가져야 한다. 그러나 이런 형식의 대답을 제공하지 않아도 이런 형식의 대답을 스스로 알 수 있게 해주는 정보를 제공했을 경우에 왜-질문에 대한 좋은 응답이 제공된 것이라고 볼 수 있다.

마지막으로, 상위 설명과 하위 설명의 구분은 험펠과 스크라이븐의 논쟁을 통해서 설명할 수 있다. 스크라이븐은 험펠을 비판하면서 설명은 단순히 'p 때문에 q이다'라는 형식으로 주어질 수 있으면 여기에 법칙은 필요하지 않다고 주장했다. 스킴은 기본적으로 스크라이븐의 이 통찰을 계승한다. 그러면 설명에 있어서 법칙의 역할은 무엇인가? 스킴은 전형적인 D-N 설명에서 법칙이 하는 역할은 'p 때문에 q이다'라는 형식의 하위 설명을 (상위) 설명하는 것이다. 이 구분에 따르면 하위 설명은 일차적인 피설명항에 대한 설명이며 상위 설명은 일차적인 피설명항에 대해 제시된 설명이 왜 설명인지에 대한 설명이다.³³⁾ 예를 들어 다음과 같은 D-N 설명이 있다고 가정해 보자.

$$\begin{array}{l} (\forall x)(Fx \rightarrow Gx) \\ Fa \\ \hline Ga \end{array}$$

스킴에게 있어서 왜 Ga가 발생했는가에 대한 적절한 (하위) 대답은 위의 D-N 논증이 아니라 'Fa 때문에 Ga가 발생했다'이다. 그런데 이런 대답이 주어질 경우 왜-질문을 던졌던 사람은 '왜 Fa와 Ga 사이에 때문에-관계가 성립하느냐'는 추가 질문을 던질 수 있다. 이 질문에 대한 적절한 (상위) 대답은 '그것은 모든 F가 G이기 때문이다'이다. 즉, 스킴에 따르면 험펠의 D-N 모델은 상위 설명과 하위 설명을 구분하지 못하고 하위 설명에 대한 이론에 상위 설명을 포함시키는 오류를 범한 것이다.

33) *Ibid.*, pp. 84-5.

이제 이 세 개의 개념적 자원을 갖고 스코우의 논증을 살펴보도록 하자. 우선 스코우는 피셔의 평형 설명이 왜 인과→설명 논제에 대한 반례가 될 수 없는지를 설명한다. 소비가 원인이라고 생각했던 인구 집단의 초기 성비는 결과되는 평형상태, 즉 1:1 성비 상태에 대해서 비례적이지 않다. 왜냐하면 그 초기 성비가 달랐다고 하더라도 궁극적으로 1:1 상태에 도달했을 것이기 때문이다. 따라서 인구 집단의 초기 성비는 비례성 원리에 따라서 원인이 아니게 된다.

다음으로 스코우는 비례성 원리에 따를 경우 1:1 성비의 원인은 그 인구 집단의 부모들이 아들만 또는 딸만 낳는 성향을 갖지 않았다는 것이 된다고 주장한다. 왜냐하면 그 인구집단의 부모들이 초기 상태에서 아들 또는 딸만 낳는 성향을 가졌더라면 그 집단은 1:1 평형상태에 도달하는 것이 아니라 멸종되었을 것이기 때문이다. 물론 이 사실은 사소한 사실이며 피셔의 평형 설명에 언급되지 않는다. 피셔의 평형 설명에서 언급되고 있는 것은 인구 집단이 1:1 평형상태에 도달하게 되는 메커니즘 뿐이다. 그러나 스코우는 이 사실이 설명→인과 논제가 틀렸다는 것을 보여주지 않는다고 생각한다. 스코우는 피셔의 '설명'이 사실 원래의 왜-질문에 대한 좋은 대답이 아니라 좋은 응답으로 제공된 것이라고 생각하기 때문이다. 일단 피셔의 평형 설명을 이해한 사람은 매우 쉽게 왜 초기 상태에서 그 인구집단의 부모가 아들이나 딸만 낳는 성향을 갖지 않았다는 것이 중국에 1:1 평형 상태에 도달하는 것의 원인이 되는지를 알아채게 된다. 즉 피셔의 평형 설명은 원래의 왜-질문에 대해 상위 설명을 제공하는 것을 통해서 그 질문에 대한 좋은 응답이 되게 된다.³⁴⁾

필자들은 스코우가 평형 설명으로부터 설명→인과 논제를 옹호하는 방식은 매우 흥미롭고 그의 입장에서 본다면 꽤나 정합적일 수 있다고 생각

34) 만약 우리가 왜-질문에 대한 좋은 응답 역시 넓은 의미에서의 설명이라고 본다면 평형 설명은 설명→인과 논제에 대한 반례가 될 수 있다. 그러나 이 경우 스코우는 인과적 설명 이론가들이, 앞서 살펴 본 것처럼, 인과→설명 논제를 양심의 가책 없이 일정부분 제한할 수 있었던 것처럼 설명→인과 논제도 양심의 가책 없이 제한될 수 있다고 주장할 수 있다.

한다. 그러나 그가 인과→설명 논제를 옹호하는 방식은 전혀 설득력이 없으며 정합적이지도 않다고 생각한다. 우선, 소비가 그렇게 생각하듯이, 1:1 평형 상태로 이르는 실제의 과정이 인과적 과정이 아니라고 보는 것은 극도로 반직관적이다. 따라서 이것을 부정하려는 사람은 매우 좋은 이유를 제시해야 한다. 물론 스코우는 자신이 그런 이유를 제시했다고 생각하며, 위에서 설명된 인과에 대한 비례성 접근이 그런 이유가 된다고 생각한다. 그러나 그것은 착각이다. 스코우 자신이 강조하듯이 비례성 접근은 선취 상황에서는 잘 작동하지 않는다. 문제는 평형 설명의 상황이 우리가 앞 절에서 살펴본 바 있는 스위치 상황의 일종이며 스위치 상황은 일종의 선취 상황이라는 것이다. 2.2에서 설명된 바 있듯이 일반적으로 실제의 인과적 과정 인근이 동일한 결과를 갖는 현실화되지 않은 가능성 인과적 과정들로 채워져 있을 경우 평형 설명이 발생한다. 즉 실제의 원인이 발생하지 않는 최근접 가능세계에서 실제의 원인을 대체하는 사건이 자체로 실제의 원인과 동일한 결과를 만들어 낼 때 평형 설명이 발생한다. 피셔의 설명에서 가정되는 특정 인구집단의 초기 성비가 2:1 이었다고 하자. 이 초기 성비가 발생하지 않는 최근접 세계에서 이 초기 성비를 사건을 대체하는 사건은, 피셔의 설명이 옳다면, 그것이 1.9:1이던 2.1:1이든, 동일한 결과를 만들어 내며 그런 이유로 이 상황은 일종의 스위치 상황이 된다. 이렇게 보면 평형 설명에서 실제의 원인이 설명적 힘을 상실하는 이유는 스위치 상황에서 실제의 (산출) 원인이 설명적 힘을 상실하는 이유와 동일하다. 인과는 산출 개념에 보다 밀접하게 연결되어 있고 설명은 의존 개념에 보다 밀접하게 연결되어 있기 때문에 이런 현상이 발생하는 것이다. 따라서 스코우의 논증은 앞서 우리가 확인한 바 있던 인과적 설명이론의 첫번째 뉘, 즉 의존 개념과 인과적 설명이론이 결합되었을 때 발생하는 문제와 유사한 문제를 갖게 된다. 만약 그의 비례성 원리를 단순하게 인과에 대한 ‘분석’으로 간주한다면 그 이론은 그가 사용하는 인과 이론이 선취 문제를 해결할 수 없어서 인과 이론으로서 적절하지 않게 된다는 문제가 발생한다. 반면에 그의 비례성 원리를 선취 문제를 해결할 수 있게 적절히 제한한다면 그는 더 이상 평형 설명이 인과→설명 논제에 대한 반례가 아니라고 주장할

수 없게 된다. 스코우는 평형 설명 상황이 스위치 상황이라는 것을 깨닫지 못했기 때문에 이 딜레마를 깨닫지 못한 것이다.³⁵⁾

4.2 스트레븐스의 대응—산출 개념에의 호소

스트레븐스의 설명에 관한 ‘카이레틱(Kairetic)’ 이론은 이 이론이 제시된 논문의 제목(The Causal and Unification Approaches to Explanation Unified-Causally)이 드러내듯이, 그리고 그 스스로 강조하듯이 완전한(fully) 인과적 설명 이론의 한 형태로 제시된 것이다.³⁶⁾³⁷⁾ 일단 그의 이론과 그의 이론이 평형 설명을 다루는 방식을 살펴 보자. 그의 이론은 매우 복잡한 과정을 통해서 정식화되기 때문에 여기서 충분히 자세히 설명될 수는 없다. 이후에 제시되는 요약은 이 논문의 목적에 따라서 최대한 단순화된 요약이다.

스트레븐스가 설명의 문제에 대해 접근하는 방식은 인과적 설명이론의 선구자들 가운데 하나인 새먼이 후기에 이 문제에 접근하는 방식과 유사하다. 새먼은 앞서 3.2에서 살펴 보았듯이 그의 초기 인과적 설명이론에 대한 히치콕의 비판에 직면해서 수정된 인과적 설명이론을 제시한다. 그 이론에 따르면 c 가 e 에 설명적으로 유관하기 위해서는 c 와 e 를 연결하는 인과적 과정이 존재해야 하며 그것에 추가해 c 와 e 사이에 통계적 유관성 관계가 존재해야 한다. 스트레븐스는 새먼이 ‘통계적 유관성’에 대해서 말하는 자리에 ‘차이 만들기(difference-making)’ 개념을 사용한다. 따라서 그에 따르면 c 가 e 에 대해서 설명적으로 유관하다는 것은 c 가 e 에 대해 차이를 만

35) 사실 소버 자신도 평형 설명 상황이 스위치 상황이며 그래서 실제의 원인들이 설명적 힘을 상실하게 되는 것이라는 것을 인식하지 못한 것으로 보인다. 소버는 평형 설명 상황은 가능한 인과적 시나리오들의 선언을 사용할 뿐 어떤 선언지가 실제적인 원인인지는 말하지 않는 선언적 설명이라고 분석한다(Sober 1983, p. 204). 그러나 소버는 왜 평형 설명이 실제 원인을 언급할 수 없게 되는지에 대해서는 자세히 분석하지 않는다.

36) Strevens (2004), p. 154.

37) 이 논문에서 필자들은 스트레븐스의 2004년 논문을 중심으로 그의 이론을 설명할 것이다. 동일한 이론의 보다 포괄적인 형태는 다음을 볼 것. Strevens (2008).

들어 내는 원인이라는 것이다. 그가 사용하는 주된 사례인 라스푸틴의 사례를 생각해 보자. 러시아 제국 말기의 악명높은 요승이었던 라스푸틴은 그의 암살자들의 연속적인 살해 시도를 통해서 사망하게 된다. 우선 암살자들은 그가 마시는 차에 독을 탔지만 실패했고 다음으로 그에게 총 두 발을 쏘지만 그는 죽지 않았다. 마지막으로 암살자들은 그의 손발을 묶은 후 네바강의 얼음구멍 속으로 그를 밀어 넣었으며 그는 익사했다. 스트레븐스는 “인과적 접근과 관련해 제기되는 도전은 라스푸틴의 죽음에 이르는 복잡한 인과적 네트워크로부터 강물에 집어 넣는 것만을 그의 죽음에 대한 적절한 설명으로 집어내는 것”³⁸⁾이라고 진단한다. 그에 따르면 라스푸틴의 죽음에 이르는 인과적 네트워크 안에는 독을 탄 차 그리고 두 발의 총알만 있었던 것이 아니며 “라스푸틴의 수염의 길이, 그 날의 꽃가루 수, 화성의 중력의 영향 등등”³⁹⁾ 수없이 많은 것들이 포함된다. 그리고 이런 수많은 인과적 프로세스들 가운데 설명적 힘을 갖는 원인을 찾아내기 위해서 그는 두 개의 절차를 제안한다. 첫번째 절차는 그가 ‘제거적 절차’라고 부르는 것으로 인과적 네트워크 안에서 라스푸틴의 죽음을 인과적으로 산출하기에 충분한 부분을 찾아내고 그 부분에서 충분성에 기여하지 않는 것들을 드러내는 것이다. 이 과정을 통해서 J. 맥키가 ‘E에 대한 충분조건의 잉여적이지 않은 부분’이라고 부른 것을 찾아낸다. 예를 들어 화성의 중력의 영향과 같은 것은 이 과정에서 잉여적인 것으로 판명되어 제거된다.

그러나 이런 제거적 절차를 통해서 걸러진 원인들 가운데 어떤 원인들은 평형 설명에서 실제의 원인이 설명적이지 않은 것과 같은 이유로 설명적이지 않다. 예를 들어 2 킬로그램 나가는 벽돌을 유리창에 던져서 그 유리창이 깨졌다고 하자. 그 벽돌이 정확히 2 킬로그램 나간다는 사실은 설명적이지 않다. 그 벽돌이 1 킬로그램 나갔거나 3 킬로그램 나갔다고 하더라도 유리창은 마찬가지로 깨졌을 것이기 때문이다. 따라서 보다 만족스런 원인을 찾기 위해서는 단순히 제거적 절차만 수행해서는 안되고 추상화

38) Strevens (2004), p. 158.

39) *Ibid.*, p. 158.

(abstraction)라는 절차를 수행해야 한다.⁴⁰⁾ 이 추상화의 절차의 핵심 아이디어는 “[인과적] 모델을 그 이상의 추상화가 [피설명항] E의 인과적 산출을 보장할 수 없는 지점까지 계속 추상화시키는”⁴¹⁾ 것이다. 위의 벽돌 던지기 사례에서 그런 최대 추상화 지점은 벽돌이 갖는 질량(m)X속도(v), 즉 운동량(mv)이 특정 영역 안에 있다는 것이다. 따라서 우리가 원하는 가장 만족스런 설명은 “그 유리창은 특정 범위 안의 운동량을 가진 벽돌이 충격했기 때문에 깨졌다”가 된다.

스트레븐스가 평형 설명의 문제를 해결하는 방법은 바로 이 추상화 전략과 관련이 있다. 평형 설명의 상황은 위의 벽돌 던지기 상황과 비슷한 상황이다. 실제의 인과적 과정은 궁극적으로 평형 상태에 도달하는데 있어 어떤 차이도 만들어 내지 않는 원인이다. 따라서 이 실제의 원인은 설명력을 갖지 않으며 이 실제의 원인은 추상화 과정을 통해서 추상화되어야 한다. 피서의 평형설명이 벽돌 던지기 상황과 다른 점은 이 추상화 과정이 극단적인 추상화를 허용한다는 것이다. 즉, 평형 설명 상황에서 추상화 과정은 극단화 되어서 실제의 원인이 사실상 제거되어도 되는 상태가 되는 것이다.⁴²⁾

스트레븐스의 이론은, 비록 그가 어떤 인과 개념을 사용하고 있는지 명시적으로 밝히고 있지는 않지만, 화성의 중력의 영향과 같은 수많은 사소한 ‘원인’들을 모두 원인으로 인정하고 있다는 점에서 대략적으로 산출적 인과 개념과 유사한 인과 개념을 전제하고 있는 것으로 볼 수 있다. 이 점에서 그의 이론은 앞서 3.2에서 살펴 본 새먼의 이론이 인과적 설명 이론과 양립 가능하지 않은 것과 정확히 같은 이유로 인과적 설명 이론과 양립 가능하지 않다. 그에 따르면, 어떤 것이 설명적이기 위해서는 인과적이면서 차이를 만들어 내야 한다. 이는 설명적이지 않은 인과를 인정하겠다는

40) *Ibid.*, p. 167.

41) *Ibid.*, p. 168.

42) 스트레븐스는 다음과 같이 말한다. “카이레틱 이론에 따르면, [평형설명] 초기 조건들은 그들이 최종적인 상태에 어떤 차이도 만들지 않기 때문에 설명에서 생략된다. 즉 그 초기 조건이 무엇이 되든 그 시스템은 평형 상태에 이르게 된다”(Ibid., p. 174).

것이며 따라서 인과→설명 논제를 부정하겠다는 것이다. 더 나아가 스트레븐스는 평형 설명의 경우 모든 실제적 원인이 추상화되어 설명에서 제거된다고 생각하므로 설명→인과 논제 역시 부정한다.⁴³⁾

스트레븐스는 그럼에도 불구하고 자신의 평형 설명에 대한 취급이 인과적 설명이론과 양립 가능하다고 생각한다. 왜냐하면 그에 따르면 평형 설명은 실제의 인과적 과정을 추상화해서 얻어진 것이며 그것은 인과적 과정에 대해서 어떤 것을 말하는 것을 통해서 설명적 힘을 갖기 때문이다 (Strevens 2019).⁴⁴⁾ 이런 점에서 스트레븐스의 아이디어는 D. 루이스의 아이디어와 닮아 있다.⁴⁵⁾ 루이스에 따르면 어떤 것이 인과적 설명이기 위해서 그것이 실제로 원인을 제시할 필요는 없다. 설명은 인과적 역사에 대해 정보를 제공하는 것이며 실제로 인과의 부재에 대한 정보도 그런 의미에서 설명적일 수 있다.⁴⁶⁾ 그러나 이런 느슨한 인과적 설명에 대한 개념이 스트레븐스에 어떤 도움을 줄 수 있는지는 분명하지 않다. 예를 들어, 우리가 지금까지 받아들였던 인과적 설명이론이 되기 위한 두 개의 필요조건을 다음과 같이 ‘느슨하게’ 수정해 보자.

설명→인과*: 어떤 것이 어떤 사건의 발생에 대한 설명이기 위해서는 피 설명항의 **인과적 역사에 관한 어떤 정보**를 제시해야 한다.

인과→설명*: 어떤 것이 어떤 사건의 **인과적 역사에 대한 정보**를 제시한

43) 스트레븐스는 앞서 지적했던 도우의 이론이 갖는 문제 역시 가질 위험이 있다. 필자들이 의심하는 것처럼 스트레븐스가 산출 개념에 호소하는 인과 이론을 암묵적으로 채택하고 있다는 것이 사실이라면 부재 인과나 이중 부정과 같은 사례들은 인과적이지는 않지만 차이를 만들어 내는 사례라고 말할 수 있다. 이런 사례 역시 스트레븐스의 이론에서 설명→인과 논제에 대한 반례가 된다.

44) 스페리-테일러는 이런 이유에서 평형 설명이 인과적 설명이라고 부르는 것에 대해서 반대한다(Sperry-Taylor 2021).

45) 우드워드나 포토치닉 역시 설명 대상이 다른 요인에 의존하는 패턴을 보여주고 있다는 넓은 의미에서 인과적 설명 개념을 받아들이며 이런 점에서 평형 설명 역시 인과적 설명에 포함될 수 있다고 주장한다(Potochnik 2015; Woodward 2003).

46) Lewis (1986), p. 220.

다면 이는 그 사건의 설명으로 성립한다.

이런 느슨한 개념을 통해서 스트레븐스는 아마도 평형 설명이 설명→인과*에 대한 반론이 될 수 없다는 것을 보일 수 있을 것이다. 그러나 이런 느슨한 개념을 사용하는 것은 평형 설명이 인과→설명*에 대한 반론이 된다는 것을 보여주는 데는 실패한다. 피셔의 평형 설명에서 특정 인구집단의 초기 성비를 제공하는 것은 분명 그 인구집단의 궁극적인 성비의 인과적 역사에 대한 정보를 제공하는 것이지만 설명적이지 않다. 그리고 이 설명적 힘은, 스트레븐스에 따르면, 화용론적 고려 때문에 상실되는 것이 아니다. 이 인과적 역사에 대한 정보 제공은 주어진 정보가 “차이를 만들지 않기 때문에” 설명적 힘을 잃는 것이다. 즉, 스트레븐스는 인과→설명* 논제가 화용론적인 방식 이상으로 제한되어야 한다고 주장하는 것이며 이는 인과적 설명 이론과 양립 불가능하다. 이점에서 스트레븐스의 실패는 앞서 설명되었던 인과적 설명이론의 딜레마의 두번째 뿔의 하나의 사례에 불과하다. 만약 우리가 산출 개념처럼 매우 허용적인 인과 개념을 사용한다면 우리는 인과가 자체로 설명적 힘을 갖는다고 할 수 없고 인과+ α 가 설명적 힘을 갖는다고 주장해야 한다. 스트레븐스의 이론은 정확히 이런 구조를 갖고 있다. 그는 α 에 “차이 만들기”를 집어 넣기 때문이다. 그리고 이런 추가 조건을 받아들이는 순간 더 이상 인과→설명 논제는 성립하지 않으며 그 이론은 진정한 의미에서 인과적 설명 이론이라고 볼 수 없게 된다.

이 지점에서 어떤 사람들은 인과적 설명이론의 개념을 한번 더 약화시켜서 어떤 이론이 인과적 설명이론이기 위해서 설명→인과* 논제를 받아들이는 것으로 충분하다고, 즉 인과→설명* 논제까지 받아들일 필요는 없다고 주장할 수 있다.⁴⁷⁾ 필자들은 이런 아이디어에 대해서 두 개의 우려를 갖고 있다. 우선 이런 아이디어를 받아들인다고 해도 스트레븐스의 이론은 정당화되지 않는다. 앞서 각주 43에서 지적한 바 있듯이 스트레븐스가 산출 인과 개념을 사용하는 한 그는 산출 인과 관계는 성립하지 않지만 “차이 만들

47) 필자들은 이 가능성을 지적한 익명의 논평자에게 감사한다.

기” 관계는 성립하는 사례들, 즉 부재 사례나 이중 방지 사례 등이 갖는 설명적 힘을 설명할 수 없게 된다. 이는 그의 이론이 여전히 설명→인과* 논제와 관련해 반례를 갖는다는 것을 의미한다. 이 점에서 스트레븐스의 이론에서 실질적으로 설명적 힘의 원천은 “인과”에 있기 보다는 “차이 만들기”에 있다고 보는 것이 그럴듯하다. 차이를 만들어 내지 않는 인과는 설명적 힘이 없지만 차이를 만들어 내는 (도우의 표현을 빌자면) 준인과는 설명적 힘을 갖기 때문이다. “인과”가 아니라 “차이 만들기”가 실질적인 일을 수행하는 설명 이론을 진정한 의미에서 인과적 설명이론이라고 생각하기는 쉽지 않다.

또한 인과→설명* 논제를 부정하는 것은 원인이 **원인으로서** 설명적 힘을 갖는다는 것, 그리고 설명적 관계가 인과 관계라는 것을 부정하는 것이며 이런 이론을 진정한 의미에서 ‘인과적 설명이론’이라고 부를 수 있는지는 의심스럽다. 3.1에서 지적한 바 있듯이 우리가 형이상학적 설명의 존재를 부정할 수는 없으므로 인과적 설명이론은 애초부터 포괄적인 설명이론이 될 수는 없다. 그러면 인과적 설명 이론은 어떤 설명에 관한 이론인가? 인과적 설명 이론이 과학적 설명에 대한 이론이라고 말하는 것은 그럴 듯하지 않다. 예를 들어 고통의 발생을 C-섬유 활성화를 통해서 설명하는 것, 소금의 용해성을 소금의 분자 구조를 통해서 설명하는 것 등은 일견 (형이상학적 설명인 동시에) 과학적 설명인 것으로 보인다. 그리고 이런 설명들은 분명 인과적 설명으로 보이지 않는다. 그렇다면 인과적 설명 이론은 아마도 우리가 흔히 “인과적 설명”이라고 부르는 현상에 대한 이론이라고 보는 것이 가장 그럴 듯해 보인다. 그리고 인과→설명* 논제를 포기하는 약화된 인과적 설명이론 개념이 말하는 것은 우리가 ‘인과적 설명’이라고 부르는 설명의 영역에서 인과적 역사에 대한 정보를 제공하는 것이 필수적이라는 꽤나 사소해 보이는 주장이 되어 버린다. 이런 방식으로 인과적 설명 이론을 ‘인과적 설명이론’이라는 개념을 흥미롭지 않게 만드는 것을 통해서 평형설명론의 문제로부터 구제하는 것이 인과적 설명 이론가들에게 얼마나 자랑스런 일이 될 수 있는지 알기는 쉽지 않다. 이런 의미에서 ‘인과적 설명이론’은 어떤 현상의 발생을 설명하는 것이 그 현상의 원인을 지적하는

것이라는 인과적 설명이론의 핵심 아이디어를 보존하는 이론을 위해서 쓰이는 것이 적절해 보이며, 필자들의 논증이 옳다면 스트레븐스는 이런 의미에서의 인과적 설명 이론이 평형 설명의 문제를 해결할 수 있다는 것을 보였다고 할 수 없다.

5. 결론

필자가 앞서 강조한 것처럼 소비는 평형 설명을 일차적으로 설명→인과 문제에 대한 반례로 제시했다. 그는 평형 설명이 비인과적 설명이라는 것을 강조하는 것을 통해서 그것이 인과적 설명 이론에 대한 반례가 된다고 생각했다. 소비의 이런 문제 설정은 이후의 평형 설명에 대한 논의의 틀을 형성했으며 많은 철학자들은 평형 설명이 과연 인과적 설명인지의 문제에 집중했다. 그러나 필자들의 분석이 옳다면 평형 설명은 그것이 설명→인과 문제에 대한 반론으로 여겨질 때보다 인과→설명 문제에 대한 반례로 여겨질 때 보다 강력한 반례가 된다. 평형 설명 상황에서의 초기 조건이 설명적으로 무력하다는 것을 설명하는 방식에는 두가지가 있다. 첫번째 방식은 그 초기 조건이 원인이 아니라고 주장하는 것이다. 앞서 살펴본 것처럼 스킴은 이 길을 가며, 이 길을 가는 가장 표준적인 방식은 넓은 의미의 의존 인과 개념을 도입하는 것이다. 평형 설명 상황에서 평형 상태는 초기 조건에 반사실적으로 의존하지 않는다. 그러나 이 길은 의존 인과 개념이, 비록 그것이 인과적 설명이론과 궁합은 좋지만, 인과에 대한 분석으로 적절성을 갖기 어렵다는 문제에 직면한다. 특히 의존 인과 개념은 선취 상황에서 잘 작동하지 않으며 평형 설명 상황은 일종의 선취 상황이라고 볼 수 있는 스위치 상황이라는 것이 필자들의 분석이다.

초기 조건이 설명적으로 무력하다는 것을 설명하는 두번째 방식은 초기 조건이 원인이기는 하지만 설명적이지 않은 원인이라고 주장하는 것이다. 앞서 설명된 스트레븐스의 이론이 바로 이 길을 간다. 그는 산출 인과 개념에 호소하는 매우 허용적인 인과 개념을 받아들이는 대신 원인은 원인으로

서 설명적 힘을 갖는 것이 아니라 차이를 만들어 내는 것을 통해서 설명적 힘을 갖는다고 주장한다. 그러나 원인이 **원인으로서** 설명적 힘을 갖는다는 것을 부정하는 것은 설명 관계가 기본적으로 인과 관계라는 것을 부정하는 것이며 따라서 그런 이론은 진정한 의미에서 인과적 설명 이론이라고 불릴 수 없다. 앞서 새먼과 도우의 사례에서 살펴본 것처럼 산출 인과 개념에 호소하는 인과 이론은 일반적으로 인과적 설명 이론과는 양립 가능하지 않다. 설명적이지 않은 인과, 그리고 인과적이지 않은 설명을 허용할 수밖에 없기 때문이다.

이런 점에서 보면 평형 설명을 인과적 설명 이론과 양립 가능하게 만들려는 최근의 인과적 설명 이론가들은 시도들은 모두 필자들이 ‘인과적 설명이론의 딜레마’라고 부른 것으로부터 탈출하는데 실패한다. 앞서 지적한 대로 인과적 설명이론은 기껏해야 플랜B 이론이며 따라서 애초부터 그렇게 강한 동기를 부여하는 이론은 아니다. 그리고 이 논문에서 살펴본 것처럼 인과적 설명 이론은 플랜B로서도 그렇게 잘 작동하는 이론은 아니다.

참고문헌

- 선우환 (2020), 『때문에—‘때문에’의 의미에 대한 철학적 연구』, 아카넷.
- 이재호 (2023), 「신의 레시피로서의 설명—새로운 조작주의적 설명 이론」, 『과학철학』 26권 3호, 1-37.
- Bromberger, S. (1966), “Why-Questions”, in Colodny, R. (ed.), *Mind and Cosmos*, Pittsburgh University Press.
- Chalmers, D. J. (1996), *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- Dowe, P. (2000), *Physical Causation*, Cambridge; New York: Cambridge University Press.
- _____ (2001), “A Counterfactual Theory of Prevention and ‘Causation’ by Omission”, *Australasian Journal of Philosophy* 79(2): pp. 216-26.
- Hall, E. J. (2004), “Two Concepts of Causation”, in Collins, J. D. & Hall, E. J. & Paul, L. A. (eds.), *Causation and Counterfactuals*, Cambridge, Mass.: MIT Press, pp. 225-276.
- Hempel, C. & Oppenheim, P. (1948), “Studies in the Logic of Explanation”, *Philosophy of Science* 15: pp. 135-75.
- Kim, J. (1987), “Explanatory Realism, Causal Realism, and Explanatory Exclusion”, *Midwest Studies in Philosophy* 12: pp. 225-39.
- _____ (1999), “Hempel, Explanation, Metaphysics”, *Philosophical Studies* 94(1-2): pp. 1-20.
- Kitcher, P. (1989), “Explanatory Unification and the Causal Structure of the World”, in Kitcher, P. & Salmon, W. C. (eds.), *Scientific Explanation*, Minneapolis: University of Minnesota Press, pp. 410-505.
- Lewis, D. (1986), “Causal Explanation”, in *Philosophical Papers* Vol. 2, Oxford: Oxford University Press, pp. 214-40.
- Potochnik, A. (2015), “Causal Patterns and Adequate Explanation”, *Philosophical Studies* 172: pp. 1163-82.

- Ruben, D. H. (1993), *Explanation*, Oxford; New York: Oxford University Press.
- Salmon, W. C. (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton, N. J.: Princeton University Press.
- _____ (1989), “Four Decades of Scientific Explanation”, in Kitcher, P. & Salmon, W. C. (eds.), *Scientific Explanation*, Minneapolis: University of Minnesota Press, pp. xiv-528.
- _____ (1997), “Causality and Explanation: A Reply to Two Critiques”, *Philosophy of Science* 64(3): pp. 461-77.
- Schaffer, J. (2001), “Causes as Probability Raisers of Processes”, *Journal of Philosophy* 98(2): pp. 75-92.
- Skow, B. (2016), *Reasons Why*. Oxford.
- Sober, E. (1983), “Equilibrium Explanation”, *Philosophical Studies* 43: pp. 201-10.
- Sperry-Taylor, A. T. (2021), “Reassessing Equilibrium Explanations: When Are They Causal Explanations?”, *Synthese* 198: pp. 5577-98.
- Strevens, M. (2004), “The Causal and Unification Approaches to Explanation Unified—Causally”, *Nous* 38(1): pp. 154-76.
- _____ (2008), *Depth—An Account of Scientific Explanation*, Harvard.
- _____ (2019), “Explanation, Abstraction, and Difference-Making”, *Philosophy and Phenomenological Research* 99(3): pp. 527-782.
- Woodward, J. (2003), *Making Things Happen: A Theory of Causal Explanation*, New York: Oxford University Press.

논문 투고일	2024. 02. 19.
심사 완료일	2024. 03. 14.
게재 확정일	2024. 03. 14.

A Dilemma of Causal Theories of Explanation and Equilibrium Explanation

Jaeho Lee · Jaehee Park

In this paper, we examine the responses of recent defenders of causal theories of explanation to the equilibrium explanation problem presented by E. Sober as a counterexample to causal theory of explanation. We argue that there is an inherent dilemma causal theory of explanation must have, and that attempts to make equilibrium explanations compatible with causal theory of explanation cannot escape this dilemma. In this process, we will attempt to present clearer criterion for what a causal theory of explanation is, and will also shed light on the deep mechanisms by which equilibrium explanations are generated.

Keywords: causation, explanation, causal theory of explanation,
equilibrium explanation