

## 세계의 인과적 구조와 설명적 구조

이재호<sup>†</sup>

이 논문에서 필자는 세계의 인과적 구조와 세계의 설명적 구조의 관계에 대해서 고찰한다. 필자는 우선 우리가 선언적 설명의 존재를 받아들이 수밖에 없으며 선언적 설명에 대한 우리의 직관을 가장 잘 설명하는 것이 필자의 설명이론이라고 주장할 것이다. 다음으로 필자는 필자의 설명 이론을 통해서 세계에 대한 다층적인 설명 구조와 평평한 인과 구조를 조화시키는 모델을 만들어 낼 수 있다는 것을 논증할 것이다. 마지막으로 필자는 이 모델이 주어질 경우 우리는 과학이 제공해 주는 그림이 세계의 인과적 구조가 아니라 설명적 구조를 드러내 준다고 볼 좋은 이유가 있다는 것을 논증할 것이다.

주요어 : 인과, 설명, 인과적 구조, 설명적 구조, 선언적 설명

---

<sup>†</sup> 중앙대학교 철학과 교수 (jaeho.jaeho@gmail.com)

## 1. 서론 – 세계의 수평적 의존 구조에 대한 일원론적 접근과 이원론적 접근

과학은 우리에게 세계에 대한 어떤 그림을 제공해 준다. 과학이 제공해 주는 세계의 그림에는 단순한 사실 또는 사건들, 예컨대 ‘언제 어디서 어떤 사건이 발생했다’와 같은 것도 포함되어 있지만 이 그림이 단순히 개별적인 사건들의 나열인 것은 아니다. 이 그림은 사실 또는 사건들의 패턴화된 의존 구조를 포함한다. 사실 과학의 보다 핵심적인 가치는 세계 속에서 일어나는 일들이 갖고 있는 이런 다양한 패턴화된 의존 구조를 드러내는 것에 있는 것으로 보인다. 예를 들어, 우리는 서기 946년 백두산이 대규모로 분화했다는 사실을 밝혀낸 것이 중요한 과학적 성과라고 생각하지만 그것이 만유인력의 법칙의 발견만큼 중요한 과학적 성과라고 생각하지 않는다. 과학은 종종 새로운 법칙이 발견되었을 때에 중요한 발전을 이루었다고 여겨지는데 이 생각의 배후에는 법칙의 발견이 패턴화된 의존 관계의 발견과 밀접하게 연결되어 있다는 생각이 있는 것으로 보인다.

패턴화된 의존 관계는 “설명” 개념, 또는 “때문에” 개념을 통해서 가장 잘 파악된다. 이 의존관계는 공시적(수직적)이거나 통시적(수평적)인데 우리는 종종 전자를 “형이상학적 설명 관계”라고 부르고 후자를 “인과적 설명 관계”라고 부른다. 형이상학적 의존 관계를 만들어 내는 것이 무엇인지에 대해서는 실현(realization), 정초(grounding), 진리 확정(truth making), 창발(emergency), 환원적 설명(reductive explanation), 수반(supervenience) 등등 다양한 이론들이 있으며 종종 이 이론들은 상호 배타적이지 않은 것으로 여겨진다. 즉, 우리는 수직적 의존 관계에 대한 다양한 모델을 수용할 의사가 있는 것으로 보인다. 반면 우리는 일반적으로 수평적 의존 관계의 경우 그 배후에 언제나 인과 관계가 존재한다고 생각한다.<sup>1)</sup> 우리가 수평적 의존 관계에 호소하는 설명을

---

1) 모든 수평적 설명의 배후에 인과 관계가 있다는 것은 이들의 배후에 인과관계만 있다 는 것을 의미하는 것은 아니다. 예를 들어 형이상학적 의존관계와 인과 관계가 모두 배후에 있는 수평적 설명은 얼마든지 있을 수 있으며, 예컨대, 정신 인과적 설명은 그런

통칭해 흔히 “인과적 설명”이라고 부르는 것은 바로 이런 생각에 기반하고 있는 것으로 보인다.<sup>2)</sup> 그러나, 이후의 논의를 통해 보다 분명해지듯이, 모든 수평적 의존 관계의 배후에 인과 관계가 있다고 말하는 것과 모든 수평적 의존관계가 인과 관계라고 말하는 것은 동일한 것이 아니다. 따라서 우리는 인과적 설명관계와 인과 관계의 관계에 대해서 물을 필요가 있다.

이 질문은 과학이 우리에게 제공하는 인과적 설명 관계의 구조가 다층적이라는 사실 때문에 더욱 흥미롭게 된다. 우리가 알고 있는 한 현재의 과학은 세계를 다층적으로 기술한다. 가장 하부에는 근원 물리학적 세계 기술이 존재하며, 그 위에는 생물학이나 지질학과 같은 특수 과학 수준의 세계 기술이 존재한다. 그리고 그 위에는 심리학과 같은 더 높은 수준의 특수 과학 수준의 세계 기술이 있을 수 있다. 그리고 과학자들은 이 다양한 수준에서 각각 인과적 설명을 제공하고 있는 것으로 보인다. 따라서 인과적 설명 관계와 인과 관계의 관계에 대한 질문은 필연적으로 과학이 제공하는 인과적 설명의 다층 구조가 세계의 인과적 구조와 얼마나 일치하느냐는 질문으로 이어지며, 필자가 이 논문에서 다루고자 하는 문제가 바로 이 문제이다.

필자의 질문을 조금 더 구체화하기 위해서, 우리는 다음과 같이 이 문제에 대한 일원론적 접근과 이원론적 접근, 그리고 준이원론적 접근을 구분할 수 있다.

**일원론적 접근:** 과학이 제시하는 다층적 인과적 설명 구조는 세계의 인과적 구조와 일치하며 세계에는 다층적 인과 구조가 존재한다.

**이원론적 접근:** 과학이 제시하는 다층적 인과적 설명 구조는 세계의 인과적 구조와 밀접하게 연결되어 있기는 하지만 일치하지 않으며 인과적 설명의 다층 구조가 인과관계의 다층 구조를 요구하지 않는다.

---

설명일 가능성이 매우 높다.

- 2) 앞으로 필자는 “인과적 설명”으로 그 설명적 관계의 성립이 어떤 종류의 인과 관계의 성립에 최소한 부분적으로 기반하고 있는 수평적 설명을 총칭해서 부를 것이다.

준이원론적 접근: 과학이 제시하는 다층적 인과적 설명 구조가 세계의 인과적 구조와 일치하는지 알 수 없으며, 따라서 우리는 일원론적 접근을 정당화할 수 없다.

필자가 이 논문에서 주장하고자 하는 것은 다음과 같다: 필자가 지금까지 응호해온 설명에 관한 이론, 특히 선언적 설명에 관한 이론을 받아들일 경우 우리는 일원론자가 될 수 없다. 일원론은 잘못되었다고 믿을 좋은 이유가 있거나(이원론) 아니면 최소한 옳다고 믿는 것이 정당화되지 않는다(준이원론).

이 논문은 이후에 다음과 같이 진행될 것이다. 우선 필자는 이 논문에서 제시되는 논증에서 필자가 전제하고 있는 몇 가지를 명확히 할 것이다. 그 후 다음 절에서 필자는 필자의 설명 이론 하에서 세계가 갖는 다층적인 인과 설명 구조가 어떻게 평평한, 즉 다층적이지 않은 인과적 구조와 양립할 수 있는지 설명할 것이다. 3절에서 필자는 이렇게 필자의 설명 이론을 통해 이해된 이원론적 접근이 우리가 반 흡주의적 형이상학적 그림을 받아들일 경우 필수적이게 된다는 것을 논증할 것이다. 4절에서 필자는 설명 우리가 흡적인 형이상학적 그림을 받아들인다고 하더라도 일원론적인 접근은 정당화되기 힘들다는 것을 논증할 것이다. 결론에서 필자는 필자의 이원론적 접근이 갖는 함축과 이 이론의 가능한 적용 사례를 암시하는 것을 통해 필자의 작업이 왜 유의미한 작업인지를 보여주고자 할 것이다.

일원론에 대한 비판을 본격적으로 전개하기 전에 이 논문에서 필자가 별다른 논증 없이 전제하고자 하는 몇 가지를 분명히 하고자 한다. 이들은 모두 논란이 있을 수 있는 전제들이다. 그러나 필자가 보기에도 오늘날 많은 형이상학자들, 그리고 과학철학자들이 이들을 받아들이며 따라서, 별다른 논증 없이 이들을 전제하는 것이 큰 문제가 되지는 않을 것으로 생각된다. 우선, 필자는 형이상학적 실재론을 전제할 것이다. 즉, 필자는 우리의 마음과 독립되어 객관적으로 존재하는 세계가 있다는 것을 전제할 것이다. 다음으로 필자는 그런 객관적으로 존재하는 세계에 객관적인 인과적 구조가 존재한다는 것을 전제할 것이다. 형이상학적 실재론을 받아들인 이후에도 인과 관계 자체를 부정하는 입장을

받아들이는 것이 가능하며<sup>3)</sup>, 인과 관계 자체는 부정하지 않더라도 인과 관계라는 것은 세계의 객관적인 구조가 아니라고 주장하는 것이 가능하다.<sup>4)</sup> 필자는 이런 입장들이 철학적으로 진지하게 고려될 수 있는 입장이라는 것은 인정하지만 이 논문에서 그런 고려는 배제될 것이다. 불필요한 혼란을 막기 위해, 필자는 세계에 객관적인 인과적 구조가 존재한다는 것은 전제할 것이지만 그 (객관적) 인과 관계의 본성이 무엇인지에 대해서는 최대한 중립적이고자 할 것이라는 것을 분명히 하고자 한다. 예를 들어 인과 관계를 일종의 단순한 규칙성으로 보는 흡수의적 입장이나 그에 반대하는 반흡수의적인 입장 가운데 어떤 것도 이 논문에서는 전제되지 않을 것이다. 다음으로 필자는 인과가 사건들 사이의 관계라는 것을 전제할 것이다. 인과의 관계항에 사건이 아니라 실체가 올 수 있다는 입장도 있지만 이 논문에서 그런 입장은 고려되지 않을 것이다. 아울러, 필자는 선언적 사건은 존재하지 않으며 따라서 선언적 인과도 존재하지 않는다는 것을 전제할 것이다. 마지막으로 필자는 수평적 설명과 인과 관계는 매우 밀접하게 연결되어 있다는 것, 보다 구체적으로, 수평적 설명 관계의 배후에는 항상 인과 관계가 존재한다는 것을 전제할 것이다.

## 2. 평평한 인과적 구조와 다층적 설명적 구조는 어떻게 양립 가능한가?

이 절에서 필자는 평평한 인과적 구조와 다층적 인과 설명적 구조가 어떻게 양립 가능한지 보여주고자 한다. 필자가 이 절과 이후의 절들에서 주장하는 많은 것들은 필자가 이미 출판한 다양한 논문에서 주장한 것에 기초하고 있다. 이 논문들에서 제시된 논증들을 여기서 다시 반복하는 것은 논의를 지나치게 방대하게 만들 것이므로 이 논문에서는 다

3) 대표적으로 B. 러셀이 이런 입장을 받아들인다(Russell 1912-13).

4) (실재론자로 해석된) 칸트는 이런 입장을 받아들이는 것으로 해석될 수 있으며, P. 키쳐의 이론(Kitche 1989)도 넓은 범위에서 이런 이론으로 해석될 수 있다.

른 곳에서 논증된 것들을 반복하지 않고 최대한 간략하게 요약하는 방식으로 진행할 것이다. 필자의 요약이 불충분하다고 느끼거나 필자의 주장이 논증을 통해서 충분히 뒷받침되지 않는다고 생각하는 독자들에게는 필자의 이전 논문들을 읽어볼 것을 추천한다.

## 2.1 새로운 조작주의적 설명 이론<sup>5)</sup>

필자의 새로운 조작주의적 설명 이론에 따르면 어떤 현상을 설명하려는 우리의 욕구는 그 현상을 제어하려는 우리의 욕구로부터 나온다. 이 기본적인 아이디어를 우리는 다음과 같이 설명할 수 있다.

비유적으로 표현해서, X를 설명하고자 하는 욕구는 신이 X를 발생시키기 위해서 실제로 사용했을 것으로 생각될 수 있는 레시피를 훔치려는 욕구이다. 이 욕구의 배후에는 X와 유사한 것들을 유사한 상황에서 제어하려는 욕구가 존재한다(이재호 2023b, p. 15).

필자는 이 기본적인 아이디어를 “확장된 루이스적 다이어그램”을 통해 꽤나 복잡하게 정식화한 바 있다. 그러나 이 논문에서는 그런 복잡한 정식화를 생략하고 다음과 같이 엄밀하지 않은 방식의 정식화를 사용할 것이다.

- A, 또는 A의 진리 확정적 부분이 실제로 B를 생성하고, A를 발생시키는 것이 자체로 B를 발생시키는데 충분할 경우 A는 B에 대한 강한 긍정적 레시피가 된다.
- A가 자체로 B에 대한 강한 긍정적 레시피는 아니지만 B에 대한 어떤 강한 긍정적 레시피의 군더더기적이지 않은(non-redundant) 부분이라면 A는 B에 대한 약한 긍정적 레시피가 된다.
- A가 발생하는 것을 막는 것을 통해 B가 발생하는 것을 막을 수 있다면 A는 B에 대한 강한 부정적 레시피가 된다.

---

5) 이 부분의 내용은 필자의 이재호 (2023b)에 기반하고 있다.

- A가 자체로 B에 대한 강한 부정적 레시피는 아니지만 이들을 둘러싸고 있는 현재의 인과적 네트워크와 동형적인 어떤 인과적 네트워크 하에서 B에 대한 강한 부정적 레시피가 될 수 있다면 A는 B에 대한 약한 부정적 레시피가 된다.
- A는 B에 대해서 어떤 강도의 긍정적 레시피이면서 동시에 어떤 강도의 부정적 레시피일 경우 [오직 그 경우에만] 설명적이게 되며, 이때 설명의 정도는 긍정적 레시피와 부정적 레시피의 강도에 의존한다(이재호 2024a).

이 이론의 핵심 아이디어는 마지막 문장, 즉 A가 B에 대해서 긍정적 레시피가 되는 동시에 부정적 레시피가 될 경우 오직 그 경우에만 설명적 관계가 성립하게 된다는 것이다. 일차적으로 긍정적 레시피와 부정적 레시피는 현상의 제어가 갖는 두 측면을 각각 드러내 준다. 어떤 현상을 제어할 수 있게 된다는 것은 우리가 그 현상을 어떻게 유도할 수 있는지(긍정적 레시피) 알고 동시에 그 현상을 어떻게 방지할 수 있는지(부정적 레시피) 아는 것이다. 그러나 이 두개의 측면은 완벽하게 독립적인 것은 아니다. A가 B에 대한 (강한) 부정적 레시피가 된다는 것은 대략적으로 A가 발생하지 않게 만들 경우 B가 발생하지 않을 것이라는 것을 의미하며, 이는 A가 B의 확률을 높여준다는 것을 의미한다. 따라서 A가 B에 대해 강한 부정적 레시피가 된다는 것은 자체로 A가 B에 대한 어떤 강도의 긍정적 레시피가 된다는 것을 함축한다.<sup>6)</sup> 반면에 A가 B에 대해 유의미한 긍정적 레시피가 된다는 것은 A가 B에 대해서 어떤 의미의 부정적 레시피가 된다는 것을 의미한다. 예를

---

6) 필자가 부정적 레시피가 긍정적 레시피를 함축한다고 생각함에도 불구하고 설명적 관계를 부정적 레시피로만 정의하지 않고 긍정적 레시피인 동시에 부정적 레시피여야 한다고 주장하는 이유는 좋은 부정적 레시피가 좋은 긍정적 레시피를 함축하지는 않기 때문이다. B에 대한 사소한 필요조건 A는 그것을 막을 경우 B가 발생하지 않게 되지만 A를 발생시킨다고 해서 B의 발생 확률이 크게 높아지는 것은 아니다. 이런 경우 우리는 종종 매우 약한 설명적 직관만을 갖게 된다. 가장 이상적인 설명은 강한 긍정적 레시피와 강한 부정적 레시피로 모두 작용할 수 있는 어떤 것이 주어질 때 성립하며 이를 강조하기 위해서 우리는 설명적 관계를 성격 규정할 때 긍정적 레시피 개념과 부정적 레시피 개념을 모두 사용할 필요가 있다.

들어 A와 not-A가 모두 B를 야기한다고 가정해 보자. 이 경우 A를 발생시키는 것을 통해서 B를 발생시키는 것이 가능하기 때문에 A는 B에 대한 긍정적 레시피가 된다. 그러나 이 경우 A는 B에 대한 유의미한 긍정적 레시피가 되지는 못한다. 우리가 A를 발생시키면 발생시키지 않건 B는 어차피 발생할 것이기 때문에 이 긍정적 레시피는 B의 발생과 관련한 우리의 제어력을 향상시켜주지 못한다.

필자의 새로운 조작주의적 설명이론은 기존의 설명이론이 호소해 왔던 세 가지의 요소를 모두 종합한 이론이며 특히, 설명의 정도를 설명할 수 있는 이론이라는 장점을 갖는다. 기존의 설명 이론은 크게 봐서 충분성에 호소하는 이론, 필요성 또는 의존성에 호소하는 이론, 그리고 (충분성이나 의존성으로 분석되지 않는) 인과성에 호소하는 이론으로 나누어질 수 있다. 험펠의 D-N 모델이나 키처의 통일 모델과 같이 설명을 논증으로 생각하는 이론들은 설명적 관계를 기본적으로 충분성 관계로 생각한다. 이런 이론들은 설명항이 피설명항에 충분하지 않지만 설명적인 경우를 설명하는데 난점을 갖는다.<sup>7)</sup> 반면, 새먼의 S-R 모델이나 우드워드 또는 선우환의 반사실적 의존성 이론들은 확률적 또는 반사실적 의존성 개념에 호소하며 설명적 관계를 일종의 의존성 관계로 생각한다. 이런 이론들은 설명항이 피설명항에 필요하지 않지만 설명적인 경우를 설명하는데 난점을 갖는다.<sup>8)</sup> 마지막으로 새먼의 인과적 설명이론처럼 충분성이나 필요성과 독립적인 방식으로 정의된 인과 개념에 호소하는 이론이 있다. 이런 이론은 설명적이지 않은 인과관계가 발생하는 사례들을 설명하지 못하는 문제를 갖는다.<sup>9)</sup> 이에 반해서 필자의 이론은 긍정적 레시피 개념을 통해 충분성

7) 이런 대표적인 사례가 잘 알려진 매독과 부전마비 사례이다. 치료되지 않은 매독을 앓은 환자들의 약 30% 정도만이 부전마비를 겪게 되는 것으로 알려져 있다. 따라서 치료되지 않은 매독은 부전마비에 대해 충분 조건이 될 수 없지만 우리는 부전마비의 발생을 치료되지 않은 매독을 통해서 설명할 수 있다.

8) 이런 대표적인 사례가 인과에 관한 반사실적 의존성 분석과 관련해 많이 논의된 선취 사례이다. 우리는 선취 사례를 2.2에서 논의할 것이다.

9) 예를 들어 새먼은 어떤 연속적인 물리적 과정을 통한 연결이 존재할 경우 인과 관계가 성립하며 인과 관계가 성립할 경우 설명적 관계도 성립한다고 주장한다(Salmon

이론이 갖는 직관적 힘을 수용하며, 부정적 레시피 개념을 통해서 의존성 이론이 갖는 직관적 힘을 수용한다. 아울러 레시피 개념, 특히 긍정적 레시피 개념에는 자체로 인과적 연결 개념이 포함되어 있다. 마지막으로 필자의 이론에서 긍정적 레시피와 부정적 레시피에는 모두 강한 형태와 약한 형태가 있으므로 다양한 설명적 힘의 정도를 설명할 수 있게 된다.

## 2.2 선언적 설명의 문제<sup>10)</sup>

필자의 새로운 조작주의적 설명이론은 추가적으로 선언적 설명에 대한 우리의 직관을 설명하는데 있어서 장점을 갖는다. 앞서 설명한 것처럼 이 논문의 논의는 선언적 인과는 존재하지 않는다는 것을 전제하고 있다. 그러나 우리가 이 전제를 받아들인다고 하더라도 우리는 선언적 설명을 받아들일 좋은 이유를 갖고 있다. 우선, 우리는 선취 상황에서의 우리의 설명적 직관이 인과적 직관과 매우 다르다는 것을 주목할 필요가 있다. 예를 들어 철수가 영수가 꽃병을 향해 동시에 돌을 던졌는데, 철수가 던진 돌이 먼저 꽃병에 도착해서 꽃병을 깨트렸고 영수가 던진 돌은 허공을 갈랐다고 가정해 보자.흔히 “늦은 선취” 상황이라고 불리는 이 상황에서 우리의 인과적 직관은 명확하다. 이 상황에서 철수가 돌을 던지는 사건이 꽃병이 깨지는 사건의 원인이며 영수가 돌을 던지는 사건은 꽃병이 깨지는 사건과는 인과적으로 무관하다. 그러나

---

1984). 그러나 종종 어떤 물리적 과정은 매우 사소해서 설명적이라고 보기 힘들다. 예를 들어 철수가 어떤 사형수가 총살되는 과정을 우연히 목격하게 되었다고 가정해 보자. 사형 집행인이 총을 겨누는 것을 보고 너무 놀라 철수는 소리를 질렀는데 곧이어 실제로 총이 발사되었고 사형수는 사망하게 되었다고 가정해 보자. 이 상황에서 철수가 소리를 지르는 사건과 사형수가 쓰러져 죽는 사건 사이에는 연속적인 물리적 과정을 통한 연결(소리의 전파)이 존재한다. 그러나 그렇다고 해서 철수가 소리를 질렀기 때문에 그 사형수가 죽었다고 설명할 수는 없다. 이런 문제 때문에 새먼은 이후에 다음과 같이 말하며 그의 설명이론에 확률적 의존성 조건을 다시 도입한다. “히치콕의 분석의 결과로 나는 이제 인과적 프로세스의 연결이 통계적 유관성 관계를 결여할 경우 설명적 힘도 결여하게 된다고 말한다(Salmon 1997, p. 476).”

10) 이 부분의 내용은 필자의 이재호 (2024a)에 기반하고 있다.

이 상황에서 우리의 설명적 직관은 **인과적 직관과는 달리** 그렇게 명확하지 않다. 우리는 꽃병이 철수가 돌을 던졌기 때문에 깨졌다라는 직관을 어느 정도는 갖고 있지만 그 직관은 별로 강하지 않다. 철수가 돌을 던지지 않았어도 그 꽃병은 어차피 (영수가 던진 돌에 의해서) 깨졌다를 것 이기 때문이다. 오히려 우리가 보다 명확하게 갖고 있는 설명적 직관은 철수 또는 영수가 돌을 던졌기 때문에, 또는 누군가가 그 꽃병에 돌을 던졌기 때문에 그 꽃병이 깨졌다라는 것이다. 우리가 이런 직관을 갖는 것은 아마도 이 경우 반사실적 의존성 관계가 성립하기 때문일 것이다. 즉, 철수 또는 영수가 돌을 던지지 않았더라면 그 꽃병은 깨지지 않았을 것이다. 따라서 선취 사례는 우리에게 두 개의 교훈을 주는 것으로 보인다. 첫번째 교훈은 이미 잘 알려진 교훈으로 인과 관계를 반사실적 의존 관계로 분석할 수 없다는 것이다. 두번째 교훈은 널리 주목받지는 않았던 것으로 반사실적 의존성은 인과 관계보다는 설명적 관계에 보다 직접적으로 연결된다는 것이다.<sup>11)</sup> 철수 또는 영수가 돌을 던졌다라는 것은 선언적인 성격을 갖고 있기 때문에 진정한 사건으로 볼 수 없으며 따라서 꽃병이 깨지는 것의 원인이 될 수는 없다. 그러나 이것은 선언적 성격을 갖고 있음에도 불구하고 설명적 힘은 갖는 것으로 보인다. 따라서 우리는 선언적 인과를 부정하더라도 선언적 설명은 받아들일 필요가 있다.

선언적 설명을 받아들여야 할 또 다른 이유는 소위 “평형 설명”이라고 불리는 설명의 존재이다. 평형 설명에 대한 대표적인 사례는 1:1 성비에 대한 R. A. 피셔의 설명이다.<sup>12)</sup> 만약 어떤 집단의 성비가 1:1로부터 크게 벗어날 경우 더 적은 수의 성을 더 많이 생산하는 부모가 재생 산적으로 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 부

11) 이 사실에 아마도 가장 명시적인 주목을 한 사람은 선우환일 것이다(선우환 2020, p. 141).

12) 이 사례를 대표적인 사례로 만든 것은 평형 설명에 대한 논의를 본격적으로 촉발한 다음의 논문이다. Sober (1983).

모가 가졌던 재생산적 이점은 사라진다. 이런 메커니즘을 통해서 인구 집단의 성비는 1:1이라는 평형상태에 도달하고 이 상태는 지속된다(이재호 2024b). B. 스코우에 따르면 이 상황에서 1:1 성비에 대한 가장 적절한 설명은 그 인구 집단의 부모들이 아들 또는 딸만 낳는 성향을 갖지 않았다는 것이다.<sup>13)</sup> 특정 인구집단의 부모들이 아들 또는 딸만 낳는 성향을 가졌다면 그 인구집단은 절멸될 수밖에 없었을 것이기 때문이다. 그러나 특정 인구 집단의 부모들이 아들 또는 딸만 낳는 성향을 갖지 않았다는 것은 1:1, 1:2, 1:3 등 다양한 비율을 선언지로 갖는 선언적 사실이며 이렇게 생각했을 때 피셔의 평형 설명은 선언적 설명이라고 볼 수 있다. 피셔 스타일의 평형 설명이 갖는 중요한 특징은 위의 선취 사례와는 달리 실제 원인은 설명적 힘을 아예 갖지 않는다는 것이다. 특정 인구 집단이 애초에 1:2의 성비를 가졌는데 이 집단이 피셔가 기술하는 메커니즘을 통해서 1:1 성비에 도달했다고 하더라도 우리는 그 인구 집단이 애초에 1:2 성비를 가졌기 때문에 1:1 성비에 도달했다고 설명할 수 없다.

평형 설명이 갖는 이런 특징은 평형 설명이 “스위치 구조”<sup>14)</sup>를 갖기 때문이다. A가 B를 야기했는데, A가 발생하지 않는 최근접 세계에서 A를 대체하는 것 역시 B를 야기한다고 가정해 보자.<sup>15)</sup> 이 경우 우리는 스위치 구조를 갖게 되며 설령 A가 B를 야기했다고 하더라도 A를 통해서 B를 설명할 수는 없다. 예를 들어 어떤 선생님이 잘못된 행동을 한 학생에게 “이게 잘못된 행동이라는 것을 알고 했니”라고 물었다고 가정해 보자. 이 상황에서 그 학생은 “예” 또는 “아니오”라고만 대답할 수 있었다고 가정해 보자. 추가로 이 선생님은 학생이 “예”라고 대답할

13) Skow (2016), p. 69.

14) 이 명칭은 N. 홀이 사용한 명칭이다(Hall 2004, p. 187).

15) 스위치 상황이 (앞서 언급된 꽃병에 돌 던지는 상황과 같은) 전형적인 선취 상황과 결정적으로 다른 점은, 전형적인 선취 상황의 경우 실제 원인이 발생하지 않았어도 동일한 결과가 발생했을 것이기는 하지만 그 동일한 결과는 실제 원인의 미발생에 의해서 발생하는 것이 아니라 (선취 되었던) 독립적인 다른 원인에 의해서 발생한다는 것이다. 반면에 스위치 상황에서는 실제 원인이 발생하지 않을 경우 그 실제 원인을 대체하는 사건 자체가 동일한 결과를 만들어 내는 상황이다.

경우 “아는 놈이 그렇게 행동했어”라고 말하면서 화를 내고 “아니오”라고 대답할 경우 “아니 나이가 몇인데 아직까지 이게 잘못된 행동이라는 것도 몰라”라고 말하면서 화를 내는 사람이라고 가정해 보자. 이 학생이 실제로 “예”라고 대답했고 선생님이 화를 냈다고 했을 때 우리는 그 학생이 “예”라고 대답했기 때문에 선생님이 화를 냈다고 말할 수 없다. 가정된 상황에서 그 선생님은 “아니오”라고 대답했어도 마찬가지로 화를 낼 사람이기 때문이다. 이 스위치 구조는 위의 평형설명에서도 마찬가지로 발견된다. 실제 원인, 즉 최초의 1:2 성비가 원인임에도 설명적 힘을 가질 수 없었던 것은 최초의 성비가 1:2가 아닌 최근접 세계에서 그것을 대체하는 것, 예컨대, 1:2.1 성비나 1:1.9 성비 역시 동일한 결과, 즉 1:1 성비를 야기하기 때문이다.

이상을 고려할 경우 인과적 설명 이론은 선취 사례와 평형 설명 사례에서의 우리의 설명적 직관을 다루는데 있어서 적절한 설명 이론으로 보이지 않는다. 인과적 설명이론의 기본 아이디어는 어떤 사건의 발생을 설명하기 위해서 우리는 그 사건의 원인이 무엇인지를 말해야 한다는 것인데, 이 사례들은 피설명항의 원인이 무엇인지를 말하는 것이 만족스런 설명이 될 수 없는 사례들이기 때문이다. 그에 반해서 필자의 설명 이론은 이런 사례들에 있어서 우리의 설명적 직관을 잘 설명할 수 있다. 어떤 인구 집단이 아들 또는 딸만 낳지 않게 만드는 것은 그 인구 집단이 궁극적으로 1:1 성비를 갖게 만드는 좋은 긍정적 레시피인 동시에 좋은 부정적 레시피이다. 따라서 필자의 이론에 따르면 우리는 1:1 성비를 그 인구집단이 애초에 아들 또는 딸만 낳지 않았다는 것을 통해서 설명할 수 있다. 반면에 그 인구집단의 초기 성비를 1:2로 만드는 것은 그 인구집단이 궁극적으로 1:1 성비를 갖게 만드는 좋은 긍정적 레시피이지만 그것은 부정적 레시피가 될 수 없다. 그 인구집단의 초기 성비를 1:2가 아니게 막는 것을 통해서 그 인구집단이 궁극적으로 1:1 성비를 갖지 못하게 만들 수는 없기 때문이다.

더 나아가 필자의 설명 이론은 왜 위의 선취 사례에서 선언적 설명과 비선언적 설명이 모두 설명적이지만 선언적 설명이 더 좋은 설명이 되는지도 설명할 수 있다. 철수가 돌을 던지게 만드는 것은 그 꽂병이

깨지게 만드는 강한 긍정적 레시피이지만 강한 부정적 레시피는 될 수 없다. 그가 돌을 던지지 않게 만들었어도 꽃병은 깨졌을 것이기 때문이다. 그러나 이 상황과 구조적으로 유사한 어떤 상황, 즉 영수가 돌을 던지기를 포기한 상황에서 철수가 돌을 던지지 못하게 만드는 것은 그 꽃병이 깨지지 않게 만드는 좋은 방법이며 따라서 철수가 돌을 던지는 것은 꽃병이 깨지는 것에 대한 **약한** 부정적 레시피가 될 수 있다. 반면에 철수 또는 영수가 돌을 던지는 것은 꽃병이 깨지는 것에 대한 강한 긍정적 레시피인 동시에 강한 부정적 레시피이다. 따라서 필자의 이론에 따르면 “철수 또는 영수가 돌을 던졌기 때문에 그 꽃병이 깨졌다”는 매우 만족스런 설명이지만 “철수가 돌을 던졌기 때문에 그 꽃병이 깨졌다”는 별로 만족스럽지 않은 설명이 된다.

필자의 설명 이론이 갖는 중요한 특징은 이 이론이 인과적 설명의 배후에 인과 관계가 있다는 것을 받아들이지만 설명 관계가 항상 인과 관계와 일치하게 만들지는 않는다는 것이다. 사실 이것은 우리가 선언적 인과는 받아들이지 않지만 선언적 설명은 받아들여야 할 경우 받아들여야 하는 필연적인 귀결인데, 필자의 이론은 이 귀결을 매우 원리 지워진 방식으로 받아들일 수 있게 해준다. 더 나아가 필자의 이론은 설명적 관계를 단순히 반사실적 의존성을 통해서 이해하는 이론에 비해서도 무시할 수 없는 장점을 갖는다. 위의 선취 사례에 대한 분석에서 드러나듯이 우리는 종종 반사실적 의존성이 없는 경우에도 (**약한**) 설명적 관계가 성립한다고 생각한다. 단순 반사실적 의존성으로 설명적 관계를 분석할 경우 이런 현상을 설명하기 어렵다. 그에 반해 필자는 **약한** 부정적 레시피 개념을 통해서 이런 문제를 극복할 수 있다. 더 나아가 단순 반사실적 의존성 관계는 무관한 선언지의 문제를 갖는다. 예를 들어 A가 B를 야기했고 A가 발생하지 않았다면 B가 발생하지 않았을 것이라고 가정해 보자. 이것이 사실이라면 B의 발생과는 무관한 C에 대해서 우리는 A 또는 C가 발생하지 않았더라면 B가 발생하지 않았을 것이라고 말할 수 있다. 그러나 그렇다고 해서 우리는 A 또는 C가 발생했기 때문에 C가 발생했다고 말할 수는 없다. 선언적 설명은 용인되어야 하지만 선언적 설명이 무제한적으로 용인

될 수는 없다. 필자의 이론은 이런 문제를 갖지 않는다. 왜냐하면 C가 B에 대해서 인과적으로 무관할 경우 A 또는 C는 B에 대해서 적절한 긍정적 레시피가 될 수 없기 때문이다. 예를 들어 공부를 열심히 하는 것은 좋은 대학에 가기 위한 좋은 긍정적 레시피가 될 수 있지만 공부를 열심히 하거나 인터넷 게임을 열심히 하는 것은 (중고등학교 학생들에게는 미안한 말이지만) 좋은 대학에 가기 위한 긍정적 레시피가 될 수 없다. 필자의 이론에 따르면 선언적 설명은 그 선언적 설명을 구성하는 선언지들 각각이 피설명항에 인과적으로 유관할 경우에 한해 용인될 수 있다.

### 2.3 이원론적 모델

지금까지 우리는 필자의 설명이론이 긍정적 레시피와 부정적 레시피 개념에 기반하고 있다는 것, 그리고 이 이론이 우리의 선언적 설명에 대한 직관을 잘 설명할 수 있다는 것을 살펴보았다. 이제 우리는 이런 특징을 갖는 필자의 설명 이론이 이원론적 모델을 만드는데 이용될 수 있다는 것을 살펴보자 한다.

필자의 제안은 하위 수준 속성에 의해서 다수 실현 가능한 상위 수준의 속성을 하위 속성들의 선언으로 보는 친숙한 제안을 그대로 받아들이자는 것이다. 예를 들어, 인간에게 있어 고통이라는 정신적 속성은 C-섬유 활성화라는 물리적 속성에 의해서 실현된다고 가정해 보자. 더 나아가 문어에게 있어서 고통은 D-섬유 활성화라는 물리적 속성, 화성인에게 있어서 고통은 E-섬유 활성화라는 물리적 속성에 의해서 실현된다고 가정해 보자. 우리는 C-섬유 활성화, D-섬유 활성화, E-섬유 활성화 등을 선언지로 갖는 선언적 속성을 생각할 수 있으며, 이 선언적 속성을 고통으로 정의할 수 있다. 그리고 이런 선언적 속성은 기능적으로 정의 가능하다. 예를 들어 C-섬유 활성화, D-섬유 활성화, E-섬유 활성화 등이 공통적으로 다양한 고통 반응, 예컨대 P-반응, Q-반응, R-반응 등을 야기한다고 가정하자. 그렇다면 우리는 대략적으로 이 선언적 속성을 P-반응, Q-반응, R-반응 등을 야기하는 현실적, 가능성 속성들의 선언으로 정의할 수 있다. 이 소위 “선언적 전략”은 H. 퍼트남이

다수실현 논증을 통해 동일론에 반대할 당시부터 고려되었던 제안이다 (Putnam 1975). 그러나 이 전략은 퍼트남을 비롯한 많은 철학자들에 의해 거부되었는데(Fodor 1997; Kim 1992), 이들의 우려의 근본에는 선언적 속성이 과학적으로 존중받을 수 있는 속성이 아니라는 생각이 있다. 필자는 이 우려가 근거 없는 우려라고 생각하며 이는 이후의 논의에서 드러날 것이다.

이제 철수가 C-섬유 활성화를 겪었으며 이것이 그의 P-반응을 야기했다고 가정해 보자. 우리는 이 상황에서 철수가 고통을 겪었기 때문에 P-반응을 했다고 말할 수 있다. C-섬유 활성화, D-섬유 활성화, E-섬유 활성화 등이 모두 P-반응을 야기하기 때문에 철수가 고통을 겪게 하는 것은 그가 P-반응을 하게 만드는 좋은 긍정적 레시피가 될 수 있다. 반면에 철수가 C-섬유 활성화도 겪지 않고, D-섬유 활성화도 겪지 않으며, D-섬유 활성화도 겪지 않게 만들 경우 철수는 P-반응을 하지 않았을 것이기 때문에 철수가 고통을 겪지 않게 만드는 것은 철수가 P-반응을 하지 않게 만드는 좋은 부정적 레시피가 된다. 따라서 필자의 이론에 따라 철수는 고통을 겪었기 때문에 P-반응을 했다고 말할 수 있게 된다. 그러나 고통이라는 속성은 선언적 속성으로 정의되었기 때문에 인과적 힘은 갖지 않는다. 이 상황에서 인과적 힘은 (C-섬유 활성화가 선언적 속성이 아니라고 가정할 경우) C-섬유 활성화만이 갖게 된다. 즉, 철수의 P-반응의 원인은 그의 고통이 아니라 그의 C-섬유 활성화가 된다. 이 그림에서 심물 인과는 인정되지 않으며 물물 인과만이 인정된다. 그러나 이 그림에서 심물 설명은 물물 설명과 마찬가지로 인정된다. 이 그림은 심물 설명뿐만 아니라 심심 설명과 물심 설명 역시 정당화한다. B. 로워가 “김재권의 최애(favorite) 다이어그램”이라고 부르는 다음의 그림을 생각해 보자.



<그림1> “김재권의 최애 다이어그램” (Loewer 2007, p. 257)

이 그림에서 P1은 M2에 대한 강한 긍정적 레시피인 동시에 강한 부정적 레시피이다. P1은 P2를 야기하고 P2는 M2를 실현할 것이기 때문에 P1을 발생시키는 것을 통해서 M2를 만들어 낼 수 있다. 또한 P1이 발생하지 않을 경우 P2가 발생하지 않을 것이고 P2가 발생하지 않을 경우 M2도 발생하지 않을 것이기 때문에 P1은 M2에 대한 강한 부정적 레시피이다. 따라서 이 그림에서 물심 설명은 정당화된다. 마찬가지로 M1이 P2에 대한 강한 긍정적 레시피이고 P2가 M2에 대한 강한 긍정적 레시피이기 때문에 M1은 M2에 대한 강한 긍정적 레시피가 된다. 또한 M1이 발생하지 않으면 P1이 발생하지 않을 것이고 P1이 발생하지 않으면 P2도 발생하지 않을 것이며, P2가 발생하지 않으면 M2도 발생하지 않을 것이기 때문에 M1은 M2에 대한 강한 부정적 레시피이다. 따라서 이 그림에서 심심 설명도 정당화된다. 그러나 M1은 선언적 속성이고 따라서 인과적 힘은 없기 때문에 이 그림 속에서 인과적 관계는 P1과 P2 사이에만 성립한다. 이제 설명적 관계를 실선으로 인과적 관계는 점선으로 그리면 우리는 다음과 같은 그림을 갖게 된다.



<그림 2> 필자의 이론을 통해 재해석된 김재권의 최애 다이어그램

이 그림에서 우리는 실선으로 이루어진 다층적 설명적 구조와 점선으로 이루어진 단층적 인과적 구조를 확인할 수 있다. 여기서 P1과 M1, 그리고 P2와 M2를 잇는 수직 방향의 실선 화살표는 형이상학적 설명 관계이며 그 외의 수평적, 그리고 대각선 방향의 실선 화살표는 인과적 설명 관계이다. 우리가 흔히 “정신 인과” 상황이라고 부르는 상황에 적용된 이런 다이어그램을 다른 영역에까지 확장하면 우리는 다층적인 인과적 설명 구조와 평평한 인과적 구조를 갖는 이원론적 모델

을 갖게 된다.

필자가 제안하는 이원론적 모델이 갖는 몇몇 특징 또는 장점을 여기서 확인하는 것이 도움이 될 것이다. 우선, 필자의 모델은 환원주의적 모델이 아니다. 이 모델은 상위 수준 속성을 하위 수준 속성으로 환원하지 않는다. 상위 수준 속성은 하위 수준 속성을 통해 다수 실현 가능하며 따라서 동일한 속성이 아니다. 물론, 필자의 모델은 상위 수준 속성을 하위 수준 속성들의 선언으로 환원한다고 주장될 수 있다. 그러나 이런 의미의 ‘환원주의’는 설령 그것이 진정한 환원주의라고 하더라도 우리가 관심을 갖고 있는 환원주의는 아니다. 예를 들어 우리가 심리 철학에서 환원주의에 대해서 관심을 가질 때 우리는 고통과 같은 심리적 속성이 C-섬유 활성화와 같은 인간의 신경 생리학적 속성으로 환원될 수 있는지에 대해서 관심을 갖고 있는 것이다. 설령 고통이 ‘C-섬유 활성화 ∨ 특정한 형태의 실리콘 칩의 구조 ∨ 특정한 형태의 외계 생명체의 조직 구조 ∨ (비물질적) 천사들이 갖는 정신적 구조 ∨ …’로 환원되는 것이 가능하다고 하더라도 우리가 관심 갖고 있는 환원주의가 참이 되는 것은 아니다.

다음으로 필자의 모델은 (다수) 실현 관계에 대한 매우 명확한 모델을 제시해 준다. 정신적 속성이 물리적 속성에 의해서 (다수) 실현된다고 주장하는 사람들은 그 실현 관계의 본성이 무엇인지에 대해서 명확한 설명을 제시할 필요가 있다. 필자의 모델에 따르면 실현 관계는 선언지와 선언의 관계이다. 이렇게 이해했을 때 우리는 실현 관계가 어떻게 형이상학적 필연성을 갖게 되는지 추가로 설명할 필요가 없어진다. P가 참일 경우  $P ∨ Q$ 가 참이 되는 것은 논리적 참이며 여기에는 어떤 신비도 없다.<sup>16)</sup>

필자의 모델은 또한 특수 과학의 자율성을 손상시키지 않는다. 물론 필자의 모델에 따르면 특수 과학은 인과적 자율성을 누리지는 못한다. 특수 과학의 수준에서 기술된 사건들 사이에 인과적 관계는 없기 때문이다. 그러나 특수 과학은 인과 설명적 자율성을 누린다.<sup>17)</sup> 위의 그림

16) 이에 반해서 포더 등이 주장하는 **이차 속성으로서의** 기능적 속성은, 특히 인과적 속성에 대한 강한 속성 실재론적 입장을 받아들일 경우, 쉽게 이해하기 어려운 속성이다.

17) J. 포더는 특수과학의 자율성에서 핵심적인 것이 설명적 자율성이라고 주장한다

이 보여주듯이 특수 과학 수준에서 기술된 사건들은 다른 특수 과학 수준에서 기술된 사건은 물론 일반 과학 수준에서 기술된 사건들과도 인과 설명적 관계를 맺을 수 있다. 이 인과 설명적 관계는 일반 과학 수준의 설명으로 환원되지 않는다. 필자의 생각에, 인과 설명적 자율성은 특수 과학의 자율성을 보여주기에 자체로 충분하다. 과학의 일차 목표가 세계의 인과적 구조가 아니라 설명적 구조를 보여주는 것이라는 생각은 그렇게 과격한 생각이 아니며 우리는 종종 “때문에(because)”와 “원인(cause)”을 혼동해서 사용한다. 이렇게 본다면 필자의 모델은 기능주의적인 접근을 통해서 소위 “비환원적 물리주의”를 옹호하려고 하는 사람들이 원하는 거의 모든 것을 충족시켜줄 수 있는 모델이 된다. 유일한 문제는 필자의 모델이 정신적 속성의 인과적 효력을 보여줄 수 없다는 것이지만 필자의 모델은 정신적 속성의 인과 설명적 힘을 보여줄 수 있으며 필자가 보기엔 이것은 그리 큰 양보는 아니다.

앞서 필자는 많은 철학자들이 선언적 속성이 과학적으로 존중될 수 있는 속성이 아니라는 생각 때문에 선언적 전략에 반대하지만 그 우려는 근거 없다고 주장한 바 있다. 필자가 이렇게 생각하는 이유는 선언적 전략을 좌절시킬 수 있는 “과학적으로 존중될 수 없는 속성”의 의미가 존재하지 않기 때문이다. 어떤 사람들은 선언적 속성이 인과적 힘을 가질 수 없기 때문에 과학적으로 존중될 수 없다고 말할 수 있다. 그러나 필자의 이론에 따르면 선언적 속성은 인과 설명적 힘을 가질 수 있다. 인과 설명적 힘을 갖는 속성을 과학적으로 존중하지 않을 이유는 없다. 어떤 사람들은 선언적 속성이 투사 가능하지 않으며 그런 이유에서 법칙에 등장할 수 없으므로 과학적으로 존중받을 수 없다고 주장한다. 그러나 선언적 속성이 법칙에 등장할 수 없다는 생각은 그렇게 분명한 것이 아니다. 우리가 위에서 검토한 피셔의 평형 설명을 생각해 보자. 피셔의 설명이 맞다면 우리는 “아들 또는 딸만 낳는 성향을 갖지 않는 인구 집단은 궁극적으로 1:1 성비에 도달한다”를 하나의 (파생적인) 법칙으로 충분히 인정할 수 있다. 여기서 “아들 또는 딸만 낳는 성향을 갖지 않음”이 갖는 선언적 성격은 이것의 법칙성을 훼손하

는 것으로 보이지 않는다. 마찬가지로 현대 의학에서 대사 증후군은 선언적인 정의를 갖는 질병이지만<sup>18)</sup> “대사 증후군에 걸릴 경우 심혈관계 질환의 발병이 증가할 수 있다”를 하나의 법칙으로 생각하지 않을 이유는 없다.<sup>19)</sup>

선언적 속성이 포함될 경우 투사 가능성이 사라진다는 생각도 그렇게 분명한 것이 아니다. S. 월터가 강조하듯이, 인과적 프로파일이 충분히 유사한 속성들의 선언의 경우 투사 가능성을 의심할 이유는 별로 없다(Walter 2006). 예를 들어, 대사 증후군으로 진단된 사람들 사이에서 무작위적으로 샘플을 골라 심혈관계 질환의 발병이 증가한다는 것을 확인할 경우 그 유관성이 미래의 사례에서도 이어질 것이라고 생각하는 것에는 어떤 잘못도 없다. 물론, (위에서 언급된 인과적 힘과 같이) 선언적 속성이 비선언적 속성에 비해 결여하는 어떤 특징을 찾아내는 방법은 있을 수 있다. 그러나 일단 우리가 어떤 선언적 속성이 설명적 힘을 가질 수 있다는 것을 받아들였을 때, 그런 선언적 속성이 과학적으로 존중될 수 없다는 결론을 내리기는 쉽지 않다.

지금까지 필자는 필자의 설명 이론을 통해서 어떻게 이원론적인 형

18) 대사증후군의 정의는 다음과 같다: “만성적인 대사 장애로 인하여 내당능 장애, 고혈압, 고지혈증, 비만, 심혈관계 죽상 동맥 경화증 등의 여러 가지 질환이 한 개인에게서 한꺼번에 나타나는 것을 대사 증후군이라고 한다” (서울대학교병원 의학정보 <https://terms.naver.com/entry.naver?docId=926836&cid=51007&categoryId=51007>) 대사증후군의 진단 기준은 (대략적으로) 위의 질환이 세 개 이상 나타나는 경우이다. 이점에서 대사증후군의 정의는 “중증 장애인”的 정의와 비슷하며 선언적이다.

19) 한 익명의 심사자는 “어떤 선언적 속성이 그 선언적 속성을 가진 대상들 사이에 어떤 유사성을 가지게 한다면, 그 대상들은 바로 그 점에서 선언적 속성 이상의 진정한 sparse property를 공유한다고 생각할 수 있다”고 주장했다. 필자는 여기서 심사자가 “sparse property”로 정확하게 무엇을 의미한 것인지 불분명하다고 생각한다. 만약 이것이 “완벽하게 자연적인 속성”을 의미하는 것이라면 필자는 심사자의 생각에 반대한다. 대사증후군이 완벽하게 자연적인 속성이라고 생각할 이유는 없다. 반면에 이것이 “상당한 정도의 자연성을 가진 속성”을 의미한다면 심사자의 주장은 필자의 주장과 크게 다르지 않게 된다. 결론 부분에서 주장되듯이 필자는 선언적 속성이 (완벽하지 않은) 자연적 속성이 될 수 있다고 생각하기 때문이다. 다만 필자는 그 “상당한 정도의 자연성을 가진 속성”이 선언적 속성 이상의 것이라고 생각할 필요가 없다고 생각할 뿐이다.

태의 비환원적 물리주의 모델이 가능한지를 설명하였고 이 모델이 꽤나 많은 장점을 갖는다는 것을 보여주었다. 그러나 그것이 자체로 이 모델을 우리가 받아들여야 한다는 것을 보여주는 것은 아니다. 이후의 논의에서 필자는 왜 우리가 이 모델을 매력적으로 생각할 필요가 있는지 논증하고자 한다. 필자의 논증은 다음과 같이 진행될 것이다.

- (1) 흡적 형이상학적 그림이 옳거나 반흡적 형이상학적 그림이 옳다.
- (2) 반흡적 형이상학적 그림이 옳다면 김재권의 배제 논증을 거부하기 어려우며 따라서 우리는 환원주의를 받아들이던지 아니면 필자의 모델을 받아들여야 한다.
- (3) 환원주의를 받아들이는 것보다는 필자의 이원론적 모델을 받아들이는 것이 낫다.
- (4) 흡적 형이상학적 그림이 옳다면 우리는 세계의 인과적 구조에 대한 회의주의를 받아들여야 한다.
- (5) 이 경우 일원론적 접근은 정당화되지 않는다.
- (6) 따라서 일원론적 접근은 받아들일 수 없거나 최소한 정당화되지 않는다.
- (7) 따라서 우리는 이원론자가 되거나 아니면 최소한 준이원론자가 되어야 한다.

### 3. 반흡적 형이상학적 그림과 이원론적 모델

#### 3.1 필연적 연결과 인과 이론<sup>20)</sup>

세계에 대한 형이상학적 그림은 가장 근원적인 수준에서 흡적인 그림과 반흡적인 그림으로 나누어질 수 있다. 흡적인 그림은 다음의 루이스의 흡적 수반 논제에 잘 나타나 있다.

---

20) 이 부분의 논의는 필자의 다음 논문에 기반하고 있음. 이재호 (2013c).

훔적 수반이라는 이름은 위대한 필연적 연결의 거부자에 대해 경의를 표하기 위해서 붙여졌다. 이것은 세계 속에는 국지적인 특정한 사실들의 거대한 모자이크만이 존재한다는, 즉 하나의 작은 사실 옆에 또 다른 작은 사실들이 하나하나 나열되어 있을 뿐이라는 주장이다. (그러나 이런 국지적인 사실들이 정신적이라는 것은 이 논제에 포함되어 있지 않다.) 우리는 기하학을 갖는다. 즉, 점들 사이의 시공간적인 거리라는 외적 관계들의 시스템을 갖는다. … 그리고 이 점들에서 우리는 국소적인 성질들을 갖는다. 이 성질들은 완벽하게 자연적이고 내재적인 속성들이며 예회되기 위해서 하나의 점보다 더 큰 어떤 것도 필요로 하지 않는다. 짧게 말해서, 우리는 성질의 배열을 갖고 있다. 그리고 그게 다다. 성질의 배열에서의 차이 없이는 어떤 차이도 없다. 모든 다른 것들은 이 배열에 수반한다(Lewis 1986b, p. ix).

훔적 수반 논제는 다양한 논제들의 연인이며 이들 가운데서 가장 중요한 것은 필자가 NNC(No Necessary Connection)라고 부르는 논제와 SHF(Supervenience on Humean Facts)라고 부르는 논제이다. NNC에 따르면 전적으로 구분되는 훔적 사실들(국지적인 특정한 사실들) 사이에 필연적 연결은 없으며, SHF에 따르면 인과, 법칙, 설명, 객관적 확률에 관한 사실들과 같은 다른 모든 사실들은 훔적 사실들의 총체적 배열에 수반된다. 우리가 “훔적 인과 이론”, “훔적 법칙 이론” 등등으로 부를 수 있는 이론들은 기본적으로 인과, 법칙 등등에 대해서 SHF 논제를 받아들이는 이론으로 볼 수 있다. NNC와 SHF는 구분되는 논제이지만 상호 독립적인 논제는 아니다. 우리는 SHF가 NNC를 전제한다고 봐야 한다. 예를 들어, 햄펠의 D-N 모델이 주장하는 것처럼, 일반화된 규칙성으로부터 피설명항 사건을 연역적으로 도출하는 다음의 논증이 진정한 설명이라고 가정해 보자.

$$(\forall x)(Fx \rightarrow Gx)$$

$$Fa$$

-----

$$Ga$$

이 경우 우리는 아마도 설명에 관한 SHF(SHF-E), 즉 흡적 설명이론을 받아들일 수 있게 될 것이다. 우리 세계의 흡적 사실들의 모자이크가 주어질 경우 우리 세계에 어떤 일반화된 규칙성이 성립하는지가 결정될 것이고 그것이 결정될 경우 어떤 사실들이 그로부터 연역적으로 도출될 수 있는지가 결정될 것이기 때문에 우리 세계에서 어떤 설명적 관계가 성립할 것인지도 결정되게 된다. 문제는 만약 세계에 필연적 연결이 존재한다면 우리는 더 이상 이렇게 SHF-E에 커미트먼트를 갖는 D-N 모델을 받아들일 수 없게 된다는 것이다. F와 G 사이에 필연적 연결이 존재한다고 가정해 보자. 이 경우 우리는 그 필연적 연결이 우리의 설명에 어떤 역할을 해야 한다는 것을 부정할 수 없다. 우선, F와 G 사이의 필연적 연결은 왜 F와 G 사이의 일반화된 규칙성, 즉  $(\forall x)(Fx \rightarrow Gx)$ 가 성립하는지에 대한 설명에 언급되어야 한다. 더 나아가 이 필연적 연결은 왜 Ga가 발생했는지에 대해서도 언급되어야 한다. Fa가 발생했고, F가 G를 필연화하기 때문에 Ga가 발생했다는 것을 설명으로서 거부할 어떤 이유도 없다.<sup>21)</sup>

일단 우리가 SHF-E가 NNC를 전제한다는 것을 받아들이면 우리는 인과에 관한 SHF, 즉 SHF-C 역시 NNC를 전제한다는 것을 받아들여야 한다. 이 논문에서 자세한 논증을 제공하지는 않을 것이지만, 인과, 법칙, 설명 등은 서로 긴밀하게 연결된 개념이며 이들은 묶음 판매(Package deal)의 대상이다. 즉, 설명에 관해 흡적인 이론을 받아들이면 인과와 법칙에 관해서도 흡적인 이론을 받아들여야 하며 설명에 관해서 반흡적인 이론을 받아들이면 인과와 법칙에 관해서도 반흡적인 이론을 받아들여야 한다. 그리고 흡적 설명이론이 NNC를 전제한다는 것을 받아들이면, 흡적 인과 이론 역시 NNC를 전제한다는 것을 받아들여야 한다. 이 점에서 흡적 설명(인과)이론은 강한 형이상학적 전제 위에서 성립하는 이론이다. 이에 반해 반흡적 설명이론은 강한 형이상학적 전제를 갖지 않는 이론이다. 예를 들어 A를 통해 B를 설명하기 위해서는 A가 B를 필연화해야 한다고 주장하는 (극도로 소박한) 반흡적 설명이론을 생각해 보자. 설령 NNC가 참인 것으로 드러난다고 하더라도 우리는 이 이론을 포기할 필요가 없다. 이

---

21) 이 주장에 대한 보다 자세하고 정식화된 논증은 다음을 참고할 것. 이재호 (2013c), 3절.

이론의 옹호자들은 우리 세계의 국소적 사실들이 무작위적으로 발생하며 우리 세계에서 발생하는 일반적인 규칙성은 우연의 일치일 뿐이므로 우리 세계에서 진정한 설명은 불가능하다고 주장할 수 있다. 이들은 우리가 우리 세계에서 설명이 가능하다고 생각하는 것은 우리 세계가 우연의 일치에 의해서 일반적인 규칙성을 갖고 있기 때문이라고, 즉 우리 세계가 극도로 오도적인(misleading) 세계이기 때문이라고 말할 수 있다.

비록 반 흡적 설명이론이 필연적 연결의 존재에 대한 전제를 갖는 것은 아니나 우리가 반 흡적 설명이론을 받아들이면, 우리가 최선의 설명으로의 추론(IBE)를 우리의 기본적인 귀납 추론의 원리로 인정한다는 전제 하에, 우리는 필연적 연결의 존재를 받아들일 좋은 이유를 갖게 된다. 현실세계에서 발견되는 다양한 일반화된 규칙성은 설명을 요구하며 이에 대한 최선의 설명은 그 규칙성 배후에 필연적 연결이 있다는 것이 될 것이기 때문이다.<sup>22)</sup> 흡적 설명 이론은 NNC를 전제하지만 반 흡적 설명 이론은 ~NNC를 귀납적 귀결로 갖는다. 따라서 우리에게 주어지는 선택지는 사실상 흡적 설명이론(인과이론)을 받아들이고 이것이 전제하는 NNC를 받아들이는 입장과 반 흡적 설명이론(인과 이론)을 받아들이고 이를 통해 ~NNC를 받아들이는 두개의 입장 뿐이게 된다.

### 3.2 반 흡적 인과 이론과 배제 논증<sup>23)</sup>

일단 우리가 반 흡적 설명(인과) 이론과 ~NNC, 즉 필연적 연결의 존재를 받아들였다고 하자. 필자는 이 경우 우리는 이원론자가 될 수밖에 없다고 생각한다. 왜냐하면 이 경우 김재권의 배제 논증의 핵심적인 아이디어를 받아들일 수밖에 없기 때문이다. 김재권의 배제 논증은 두 개의 핵심적인 전제, 즉 어떤 물리적 사건이 원인을 갖는다면 그 사건은 물리적인 충분 원인을 갖는다는 소위 폐쇄성 원리와 체계적 과잉 결정은 존재하지 않는다는 원리(과잉 결정 금지 원리)에 기반해 있는 논증이다. 이 논증은 다음과 같이 정리될 수 있다.

---

22) 귀납에 대한 소위 “설명주의자”들의 접근은 흔히 이런 추론에 기초해 있다(Armstrong 1983, pp. 52-3; Foster 1983).

23) 이 부분의 논의는 필자의 다음 논문에서의 논의에 기초해 있다. 이재호 (2019).

- (1) 원인을 갖는 물리적 사건은 물리적 충분 원인을 갖는다. (폐쇄성 원리)
- (2) 시점  $t_1$ 에서 발생한 어떤 정신적 사건  $m$ 이  $t_2$ 에서 발생한 어떤 물리적 사건  $p_2$ 의 원인이라고 가정하자.
- (3) 그럴 경우  $t_1$ 에서 발생한 어떤 물리적 사건(이 사건을  $p_1$ 으로 부르자)이  $p_2$ 의 원인이 된다. (1과 2로부터)
- (4)  $m$ 과  $p_1$ 이 같은 사건이 아니라고 가정하자.
- (5) 그럴 경우  $p_2$ 는 두 개의 충분 원인(즉,  $m$ 과  $p_1$ )을 갖게 되어 과잉 결정된 사건이 된다.
- (6) 이것이 사실이라면,  $m$ 과 같은 유형의 사건이  $p_2$ 와 같은 유형의 사건을 야기할 때마다 (체계적으로) 과잉 결정이 발생하게 된다.
- (7) 체계적 과잉 결정은 존재하지 않는다. (과잉 결정 금지 원리)
- (8) 따라서, 2나 4 가운데 최소한 하나의 가정은 잘못된 것이다. (귀류법적 추론)
- (9) 2를 부정할 경우 부수현상론을 받아들여야 하는데 부수현상론을 받아들일 수는 없다.
- (10) 따라서, 4를 부정해야 하는데, 이는 환원주의를 받아들여야 한다는 것을 의미한다.

(이재호 2019, pp. 3-4)

흔히 배제 논증을 거부하는 사람들은 이 논증에서 (7), 즉 과잉 결정 금지 원리를 부정한다. 이들은 체계적 과잉 결정이 전혀 이상한 것이 아니며 그것이 금지되어야 할 이유가 없다고 주장한다. 보통 이들은 이 주장을 위해 흡적 인과 이론에 호소한다.<sup>24)</sup> 인과 관계라는 것이 기본적으로 넓은 의미에서의 규칙성 관계라고 가정해 보자.<sup>25)</sup> 또한, 물리적 속성 P1이 물리적 속성 P2와 적절한 규칙성 관계 있고, P1이 정신적 속성 M1을 실현한다고 가정해 보자.<sup>26)</sup> 이 경우 P1과 P2 사이의

24) B. 로워가 이런 길을 가는 대표적인 철학자이다(Loewer 2007).

25) 고전적인 인과에 대한 규칙성 이론과 D. 루이스적인 반사실적 의존성 이론들이 이런 이론에 해당된다.

규칙성 관계는 M1과 P2 사이의 규칙성 관계를 함축한다. 이 경우 인과에 대한 규칙성 이론에 따라 M1과 P2 사이에는 P1과 P2 사이에서 와 마찬가지로 인과 관계가 성립하게 되며 체계적 과잉 결정이 발생하게 된다. 그러나 여기에는 어떤 신비스러운 점도 없고 형이상학적으로 반대할 이유도 없어 보인다. 체계적 과잉 인과는 물물 인과의 원인이 정신적 속성을 실현할 경우 사실상 논리적으로 따라 나오는 귀결일 뿐이다. 그러나 우리가 반 흡적 인과 이론을 받아들이면 상황은 전혀 달라진다. 이 경우 우리는 더 이상 인과 관계가 단순한 규칙성 관계라고 말할 수 없게 되기 때문이다. 앞서 설명한 이유로 반 흡적 인과 이론을 받아들일 경우 우리는 인과 관계가 성립하기 위해서 단순한 규칙성을 넘어서는 어떤 종류의 필연적 연결이 인과 관계의 성립에 필수적이라고 말해야 한다. 이 경우 P1과 P2 사이에 인과 관계가 성립하고 P1이 M1을 실현한다고 가정하더라도 그로부터 M1과 P2 사이에 인과 관계가 성립한다고 볼 이유는 없다. P1과 P2 사이에 성립하는 인과 관계가 보장하는 것은 M1과 P2 사이의 단순한 규칙성 관계이지 인과 관계는 아니게 되기 때문이다. 따라서 체계적 과잉 인과 상황이 성립하기 위해서는 P1과 P2 사이에 필연적 연결이 있을 때마다 M1과 P2 사이에서도 필연적 연결이 있다고 말해야 하는데, 이는 감내할 수 없는 수준의 우연의 일치를 세계에 도입하는 것이며 (오컴의 면도날에 의해서 배제되어야 하는) 불필요한 형이상학적 요소를 세계에 도입하는 것이다. 즉, 우리가 반 흡적 형이상학적 그림을 받아들일 경우 우리는 반 흡적 인과 이론을 받아들여야 하며, 반 흡적 인과 이론을 받아들일 경우 우리는 과잉 결정 금지 원리를 받아들여야 한다. 따라서 일단 우리가 반 흡적 형이상학적 그림을 받아들일 경우 우리는 위의 논증에서 (8)을 받아들여야 한다. 즉 우리는 김재권식의 환원주의를 받아들이거나 아니면 부수현상론을 받아들여야 한다. 김재권은 이 상황에서 부수 현상론을 받아들일 수 없으므로 환원주의를 받아들여야 한다고 주장한다. 그러나 필자는 우리가 이 지점에서 김재권을 따라갈 필요는 없다고 생각한다. 우리는 이 지점에서 필자가 제안하는 이원론적 접근을 받아들

---

26) 위의 <그림1> 참고할 것.

일 수 있다. 즉, 정신적 속성의 인과적 힘을 부정하는 대신 정신적 속성의 인과 설명적 힘을 긍정하는 길, 즉 위의 논증에서 (9)를 부정하는 길을 갈 수 있다. 김재권은 이 길을 받아들이지 않는다. 왜냐하면 그는 선언적 설명을 받아들이지 않기 때문이다.<sup>27)</sup> 그의 선언적 설명에 대한 반대의 배후에는 설명의 구조가 인과의 구조와 일치해야 한다는 일종의 강한 형태의 인과적 설명이론이 있다. 그러나 우리는 이미 앞에서 (2.2) 선언적 설명은 받아들일 수밖에 없으며 선언적 설명에 대한 우리의 직관을 가장 잘 설명하는 이론이 필자의 설명 이론이라는 것을 확인한 바 있다.

일단 우리가 인과 관계와 인과적 설명 관계를 구분하고 심적 속성이 인과적 힘을 가질 수는 없지만 인과 설명적 힘을 가질 수 있다고 생각하면 부수현상론은 그렇게 혐오스러운 것이 될 필요가 없다. 우리는 우리의 행위를 우리의 심적 속성, 예컨대 믿음과 욕구를 통해서 얼마든지 설명할 수 있게 된다.<sup>28)</sup> 더 나아가 앞서 설명했듯이 필자의 이원론적 모델은 기존의 비환원적 물리주의 이론이 갖는 거의 모든 매력을 그대로 유지하는 이론이다. 이 모델은 현재의 과학이 제공하는 세계에 대한 다층적 인과 설명적 구조를 그대로 유지할 수 있는 이론이다. 이렇게 생각한다면 환원주의와 이원론적 모델 사이에서 우리는 이원론적 모델을 선택할 좋은 이유가 있다.

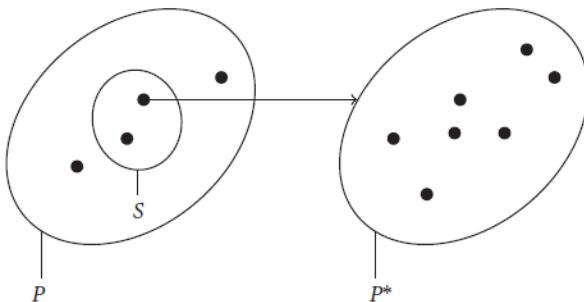
---

27) 김재권이 어떤 이유로 선언적 설명을 거부하는지를 알기 위해서는 다음을 볼 것. Kim (1999), p. 13.

28) 김재권은 정신 인과가 중요한 것은 인간의 행위성이 중요하기 때문이라고 주장한다 (Kim 2007, p. 236).

### 3.3 부분집합 실현 개념과 비환원적 물리주의<sup>29)</sup>

3.2에서 필자는 우리가 반 흡 적 형이상학적 그림을 받아들일 경우 김재권의 과잉 결정 금지 원리를 받아들여야 하고 이 경우 우리는 필자의 이원론적 모델을 선택할 좋은 이유가 생긴다고 주장했다. 그러나 일부 철학자들은 설령 우리가 반 흡 적 형이상학적 그림을 받아들인다고 하더라도 과잉 결정 금지 원리를 받아들여야 하는 것은 아니라고 주장한다. 필자는 여기서 이런 주장을 하는 대표적인 철학자인 J. 윌슨의 부분집합 실현 개념을 검토해 볼 것이다.<sup>30)</sup> 윌슨의 이론의 핵심 아이디어는 다음의 다이어그램에 잘 나타나 있다.



<그림3> (Wilson 2021, p. 70)

이 그림은 기본적으로 속성이 힘들의 집합, 또는 묶음이라는 생각에 기반해 있다.<sup>31)</sup> 여기서 P는 하위 수준의 토큰 속성(예컨대 철수의 C-섬유 활성화)를 의미하며 S는 P가 (부분집합) 실현하는 상위 수준의 토큰 속성(예컨대 철수의 고통)을 의미한다. 그리고 P\*은 P와 S에 의해서 과잉 결정되는 하위 수준의 토큰 속성(예컨대 철수의 얼굴 찡그림)을 의

29) 이 부분의 논의는 필자의 다음 논문에서의 논의에 기초해 있다. 이재호 (2023a).

30) 부분집합 실현 개념은, 비록 그 명칭은 K. 모리스가 도입했지만(Morris 2019, Ch. 3), S. 슈메이커에 의해서 제안된 것이다(Shoemaker 1980, 2001).

31) 윌슨은, 슈메이커와는 달리, 속성이 힘의 묶음이라는, 특정한 반 흡 적 형이상학적 그림을 전제할 필요가 없다고 생각하지만 윌슨의 생각은 명백히 잘못된 생각이다. 왜 이 생각이 잘못된 것인지에 대한 필자의 논증은 다음을 볼 것. 이재호 (2023a), 3절.

미한다. 이 그림에서 P를 구성하는 힘들(점으로 표시됨)의 부분집합이 S를 구성하며, 따라서 P는 S를 (부분집합) 실현한다. S(따라서 P)를 구성하는 힘들 가운데 하나가 P\*를 “야기하기(causing)”<sup>32)</sup>하기 때문에 S와 P는 모두 P\*를 야기하게 되며 이 상황은 체계적 과잉 인과 상황이 된다. 그러나 이 체계적 과잉 인과 상황은 형이상학적으로 전혀 문제될 것이 없다. 왜냐하면 인과 관계는 두 곳에서 발생하고 있지만 “야기하기”는 하나밖에 없기 때문이다. 이 그림은 반흡적 형이상학적 그림과 반흡적 인과 개념을 배제하지 않고 있지만 그럼에도 체계적 과잉 인과 상황을 오컴의 면도날 같은 형이상학적 원리와 긴장 없이 수용할 수 있게 해준다고 윌슨은 주장한다.

필자는 그러나 윌슨의 부분집합 실현 개념이 일원론을 옹호하는데 사용될 수 없다고 생각한다. 윌슨의 그림이 갖는 가장 큰 특징은 인과적 효력의 궁극적 담지자가 속성이 아니라 힘이게 만든다는 것이다. 어떤 속성은, 윌슨에 따르면, 그 속성이 어떤 특정한 힘을 자신의 원소로 가짐으로써 인과적 힘을 갖게 된다. 그런 점에서 이 그림은 속성이 인과적 힘의 궁극적 담지자라는 일반적인 그림과는 구별되어야 한다. 그리고 일단 인과적 힘의 궁극적 담지자를 힘으로 생각할 경우 속성의 인과적 효력에 대해서 이야기하는 것은 별로 의미가 없다. 예를 들어 위의 그림에서 물리적 속성 P가 다른 물리적 속성 P\*를 “야기하는” 것은 순전히 P에 포함되어 있는 특정한 하나의 힘 때문이다. 그 특정한 힘 외의 다른 힘들은 P\*의 야기 과정에서 어떤 역할도 수행하지 않는다. 그렇다면 우리는 그 힘이 P\*를 야기했다고 말해야지 왜 P 또는 S가 P\*를 야기했다고 말해야 하는가? 일단 힘을 인과적 효력의 궁극적 담지자로 설정하면 인과 관계의 첫번째 관계항은 힘이어야 하지 속성이 될 수 없다. 따라서 엄밀히 말할 경우 위의 그림에서 속성들, 즉 P와 S는 인과적 힘을 갖는다고 말할 수 없다. 따라서 이 그림이 표현하는 상황은 (체계적) 과잉 인과 상황으로 볼 수 없다.

아마도 윌슨은 필자의 이런 비판에 대해서 다음과 같이 말할 수 있을 것이다: 비록 이 그림에서 인과적 효력의 궁극적 담지자가 속성이

---

32) 윌슨은 토큰 속성을 구성하는 토큰 힘의 발현을 “야기하기(causing)”라고 부른다.

아니라 힘인 것은 맞지만 우리는 파생적인 의미에서의 인과를 다음과 같이 추가로 정의할 수 있다. “속성 P를 구성하는 어떤 힘이 P\*를 야기할 경우 속성 P는 P\*를 **이차적으로** 야기한다.” 그러나 이런 식으로 파생적 인과 개념을 추가로 정의한다는 것이 일원론을 정당화하는데 어떤 도움을 줄 수 있는지는 알기 어렵다. 이 그림 하에서 일차적인 의미에서의 인과 관계는 물리적 속성을 구성하는 힘들, 즉 물리적 힘들에 의해서 발생하며 단층적이다. 비록 파생적 인과 관계는 물리적 속성과 정신적 속성이 모두 가질 수 있어서 다층적인 구조를 갖고 있지만 그것은 어디까지나 파생적 인과 관계의 구조일 뿐이다. 이 그림이 단순한 눈속임에 불과하다는 것은 필자의 이원론적 모델도 비슷한 방법으로 일원론적 모델로 둔갑될 수 있다는 것을 생각해 보면 명백하다. 예를 들어 필자 역시 다음과 같이 “파생적 인과” 개념을 정의할 수 있다. “정신적 속성 S와 물리적 속성 P\* 사이에 인과 설명적 관계가 성립할 경우 S는 P\*를 **이차적으로** 야기한다.” 이 경우 필자의 모델 역시 월슨의 그림과 같은 정도로 체계적 과정 인과를 허용할 수 있으며 일원론적 그림으로 눈속임할 수 있다. 반면에 월슨이 “이차적 인과”를 필자의 인과 설명과 비슷한 것으로 생각한다면 월슨의 그림은 필자의 그림과 같은 정도로 이원론적 그림이 된다.

#### 4. 흡적 형이상학적 그림과 이원론적 모델

앞절에서 필자는 우리가 반흡적 형이상학적 그림, 즉 어떤 종류의 필연적 연결을 받아들이는 그림을 받아들일 경우 일원론적 그림을 받아들일 수 없으며 필자가 제안하는 이원론적 모델을 받아들일 필요가 있다는 것을 논증했다. 필자는 흡적 형이상학적 그림과 반흡적 형이상학적 그림을 비교할 경우 반흡적 형이상학적 그림이 여러 면에서 우월하다는 것을 다수의 논문을 통해서 논증한 바 있다(Lee 2015; 이재호 2013a, 2013b, 2013c, 2014, 2020a, 2020b, 2022). 이 논문들에서의 필자의 논증이 옳다면, 반흡적 형이상학적 그림을 받아들일 경우 우리가

이원론적 모델을 받아들여야 한다는 것은 단순히 우리가 이원론적 모델을 받아들여야 한다는 것을 의미한다. 그러나 흡적 형이상학과 반흡적 형이상학 사이의 논쟁은 거대한 논쟁이며 필자는 이것과 관련해 독자들에게 필자의 입장을 강요할 생각은 없다. 대신 필자는 이 절에서 설령 우리가 흡적 형이상학을 받아들인다고 하더라도 우리는 일원론적 모델을 정당화할 수 없다는 것을 보이고자 한다.

우선, 흡적 형이상학을 받아들일 경우 우리는 속성이 인과적 힘을 필연적으로 가질 수 없다는 것, 즉 속성의 개별화가 그 속성이 갖는 인과적 프로파일에 의존할 수 없다는 것을 받아들여야 한다. 왜냐하면 이 형이상학적 그림에서 인과 관계는 개별 흡적 사건들의 배열에 수반해야 하는데 이는 인과 관계가 (개별 흡적 사건들을 구성하는) 속성들의 정체성에 형이상학적으로 후행한다는 것을 의미하기 때문이다. 이런 입장은 흔히 “정언주의(categoricalism)”라고 불리며, 정언주의에 따를 경우 속성은 그 속성의 인과적 프로파일이 아니라 원초적인 속성의 개별화 원리인 큐디티(quiddity)에 의해 개별화된다. 그러나 우리가 일단 정언주의를 받아들일 경우 관찰 불가능한 이론적 속성을 지시하는 용어의 정의와 관련해 패러독스가 발생한다.

이론적 용어를 정의하는데 사용하는 표준적인 방법은 흔히 “램지-루이스 방법”이라고 불리는 방법이다. 이 방법의 핵심적인 아이디어는 관찰 불가능한 속성을 지시하는 용어들을 이들이 관찰 가능한 속성들에 미치는 영향을 포함하는 전제적인 인과적 네트워크를 통해서 기능적으로 정의하자는 것이다. 이 방법에 따를 경우 “전하”나 “스핀”과 같은 근본 물리학적 용어들은 ‘전하 역할’, ‘스핀 역할’과 같은 인과적 프로파일에 의해서 정의되게 된다. 예컨대, 전하는 ‘전하의 인과적 역할을 하는 속성’으로 스핀은 ‘스핀의 인과적 역할을 하는 속성’으로 정의되게 된다. 문제는 흡적 형이상학을 받아들일 경우 인과 기능적 역할이 특정한 이론적 용어를 정의하는데 사용될 수 없다는 것이다. 왜냐하면 우리가 정언주의를 받아들일 경우 두 개 이상의 상이한 속성이 동일한 인과적 프로파일을 갖는 것을 배제할 수 없기 때문이다. 예를 들어 어떤 세계에 F, G, H라는 세개의 속성만 존재하고  $(\forall x)(Fx \rightarrow Hx)$ 와  $(\forall x)(Gx \rightarrow Hx)$ 라는 두

개의 일반화된 규칙성만 존재한다고 가정해 보자. 이 경우, 흄적 형이상학적 그림의 전제 하에, F의 인과적 역할은 H를 야기하는 것이고 G의 인과적 역할 또한 H를 야기하는 것이기 때문에 F와 G는 인과적 역할을 통해서 구분될 수 없게 된다. 이제 우리가 어떤 속성을 “H를 야기하는 속성”으로 정의했다고 가정해 보자. 이제 이 정의는 이 세계에서 단 하나의 속성을 집어낼 수 없기 때문에 이렇게 정의된 속성은 어떤 속성도 지시할 수 없게 되거나 아니면 F 또는 G라는 선언적 속성을 지시해야 한다. 전자의 경우 이 정의는 지시하는데 실패하는 것을 통해서 실패하게 된다. 후자의 경우 이 정의는 모순을 결과한다. 왜냐하면 H를 야기하는 속성이 F 또는 G라는 선언적 속성이라고 밀한다면, 선언적 속성은 인과적 효력을 가질 수 없기 때문에 이 속성이 H를 야기할 수 없다고 말해야 할 것이기 때문이다. 결국 흄적 형이상학을 받아들일 경우 우리는, 특정한 인과적 프로파일이 단 하나의 속성에 의해서 점유된다는 정당화될 수 없는 형이상학적 가정 없이<sup>33)</sup>, 이론적 용어를 램지-루이스 방법을 통해서 정의할 수 없게 된다. 반면 이 그림 하에서 속성의 정체성을 결정해주는 쿠디티는, S. 슈메이커가 지적하고(Shoemaker 1980) D. 루이스가 인정하듯이(Lewis 2009), 애초부터 인식적으로 접근 불가능한 것이기 때문에 이론적 속성을 그것을 통해서 정의하는 것도 불가능하다. 결국 흄적 형이상학을 받아들일 경우 근원적 속성들을 지시하기 위해서 사용되는 용어들은 정의 불가능하게 된다.

A. 버드는 정연주의가 갖는 이런 문제가 정연주의를 거부할 강력한 이유가 된다고 생각한다.<sup>34)</sup> 필자는 버드의 생각에 동의하지 않는다. 흄주의자들에게는, 그들이 선언적 설명에 대한 필자의 아이디어를 수용하

33) 흄자는 현실 세계에서의 근원 물리학적 속성들의 인과 설명적 역할이 두 개 이상의 속성에 의해서 점유될 가능성이 매우 작으므로 우리가 이런 가정을 하는 것이 크게 문제 될 수 없다고 주장할 수 있다. 그러나 이 주장을 하려면 이 가능성이 작다는 것을 정당화할 수 있는 방법이 있어야 하는데 필자가 아는 한 그런 방법은 존재하지 않는다. 특히 현실 세계 안에서 그 설명적 역할이 매우 비슷해서 전문가가 아니면 구별하기 어려운 많은 속성이 있다는 것이 주어질 경우 현실 세계 안에 그 설명적 역할이 동일한 속성이 없을 확률이 높다는 것을 보여줄 수 있는 방법을 찾기는 쉽지 않을 것이다.

34) Bird (2007), pp. 62-3.

는 한, 램지-루이스 방법의 기본 아이디어를 크게 훼손시키지 않으면서 이 문제를 회피할 수 있는 명확한 방법이 있기 때문이다. 우리는 램지-루이스 방법의 내용을 약간 수정해서 이론적 용어를 **인과적** 프로파일을 통해서 정의하는 대신 **인과 설명적** 프로파일을 통해서 정의할 수 있다. 예를 들어, 우리는 전하를 ‘전하의 설명적 역할’을 수행할 수 있는 속성들의 최대 선언 속성’으로 스플을 ‘스핀의 설명적 역할’을 수행할 수 있는 속성들의 최대 선언 속성’으로 정의할 수 있다. 우리 세계에서 전하의 인과적 역할을 수행하는 속성이 F와 G 두 개가 있다고 가정해 보자. 이런 경우 전하의 인과 설명적 역할을 수행할 수 있는 속성은 F와 G, 그리고  $F \vee G$  세 개가 있게 된다. 앞서 설명된 이유로  $F \vee G$ 라는 선언적 속성은 전하의 인과적 역할을 수행할 수는 없지만 인과 설명적 역할을 수행할 수는 있다. 이제 이 세 속성을 모두 선언으로 연결해 최대 선언 속성을 만들어 보자. 우리가 갖게 될 속성은  $F \vee G \vee (F \vee G)$ 가 될 것이며, 이는  $F \vee G$ 와 동치가 된다. 따라서 가정된 세계에서 “전하”는  $F \vee G$ 라는 (하나의) 선언적 속성을 지시하게 된다.<sup>35)</sup> 이 선언적 속성도 인과 설명적 역할을 수행할 수 있기 때문에 이 경우 패리독스는 발생하지 않는다. 반대로, 우리의 원래의 가정과는 달리 우리 세계에서 전하의 인과적 역할을 하는 속성이 F 하나라고 가정해 보자. 이 경우 전하의 설명적 역할을 수행하는 속성들의 최대 선언 속성은 그냥 F가 될 것이다. 종합하면, 우리가 이론적 용어를 위의 수정된 램지-루이스 방법을 통해 정의할 경우 우리의 이론적 용어들은 하나의 속성을 지시할 수도 있고 여러 속성들의 선언을 지시할 수도 있다. 물론, 흡 주의자들의 가정에 따를 경우, 우리가 이 방법에 의해서 정의한

---

35) 이 상황에서 “전하”가 하나의 선언적 속성을 지시하게 만드는 또 다른 방법은 “전하”를 ‘전하의 **인과적** 역할을 수행하는 속성들의 최대 선언 속성’으로 정의하는 것이다. 이렇게 정의해도 그렇게 정의된 전하가 인과적 힘을 갖는다는 것을 정당화할 수 있는 방법은 없기 때문에 결국 전하에 관한 근본 물리학의 이야기는 일차적으로 인과적 구조가 아니라 설명적 구조를 드러낸다고 봐야 한다. 따라서 이런 정의를 사용해도 결론은 달라지지 않는다. 필자가 굳이 “설명적 역할”을 변형된 램지-루이스 방법의 정식화에 사용하는 것은 이런 개정을 받아들일 경우 우리는 근본 물리학이 일차적으로 설명적 구조를 제공해준다고 봐야 한다는 것을 강조하기 위해서이다.

속성이 선언적 속성을 지시하는지 아니면 비선언적 속성을 지시하는지 알 수 있는 방법은 원리적으로 존재하지 않는다.

인과적 프로파일이 아니라 인과 설명적 프로파일에 호소하는 것으로 램지-루이스 방법이 변화될 경우 “전하”나 “스핀”과 같은 이론적 용어들에 호소하는 우리의 근본 물리학이 제공해 주는 세계의 패턴화된 의존적 구조는 일차적으로 설명적 구조라고 봐야 한다. 물론 우리가 아주 운이 좋다면 그 구조가 인과적 구조와 일치할 수도 있다. 그러나 그 구조가 인과적 구조라는 믿음은 결코 정당화될 수 없다. 그리고 근본 물리학이 제공해주는 패턴화된 의존적 구조가 설명적 구조라면 그 위에 있는 특수 과학이 제공해주는 패턴화된 의존적 구조 역시 설명적 구조라고 봐야 한다. 따라서 우리는 다층적 인과 설명적 구조가 인과적 구조와 일치한다는 일원론을 받아들일 수 없게 된다. 이 상황에서 우리는 준이원론, 즉 과학이 제공해 주는 다층적 설명 구조가 인과적 구조와 일치하는지 알 수 없다는 입장을 받아들여야 한다.

종합하자면, 우리가 반 흡적 형이상학적 그림을 받아들일 경우, 우리는 이원론을 받아들여야 한다. 반대로 우리가 흡적 형이상학적 그림을 받아들일 경우 우리는 준이원론을 받아들여야 한다. 어떤 경우에도 일원론은 정당화되지 않는다.

## 5. 결론

필자가 생각하기에, 우리는 과학이 세계의 인과적 구조를 드러내 준다는 생각을 너무 쉽게 받아들이는 경향이 있다. 지금까지 필자가 제시한 논증이 옳다면 이제 그 생각은 바꿔어야 할 것으로 보인다. 우리는 이제 과학이 제공해 주는 것이 일차적으로 **인과 설명적** 구조일 뿐이지 그것이 **인과적** 구조라고 생각해서는 안된다. 즉, 과학은 무엇이 무엇 때문에 발생하는지에 대한 체계화된, 그리고 다층화된 구조를 제시해 줄 뿐이며 이 구조는 자체로 무엇이 무엇을 **야기하는지**에 대한 지식을 제공해 준다고 볼 수 없다. 필자가 생각하기에, 설명과 인과에 대한 철

학적 논의의 역사는 원인(cause)과 때문에(because)의 혼동의 역사라고 말할 수 있을 만큼 혼동으로 가득한 역사이다. 이 역사는 아리스토텔레스의 4개의 때문에(aitia)에 관한 이론이 “4원인(cause)론”으로 번역되었을 때 이미 시작되었을 지 모른다.<sup>36)</sup> 필자의 생각에 오늘날에도 엄청난 영향력을 미치고 있는 반사실적 의존성 인과 이론들, 예컨대 D. 루이스나 J. 우드워드의 이론 역시 대표적인 원인과 때문에를 혼동하는 이론이다. 앞서 언급된 바 있듯이 반사실적 의존성은 인과보다는 설명과 직접적으로 연결되는 개념으로 봐야 한다. 마찬가지로 다양한 인과적 설명 이론들 역시 원인과 때문에를 혼동하는 이론들이다. 물론 인과 개념과 (인과) 설명 개념은 밀접하게 연결되어 있다. 그러나 이들은 엄격하게 구분될 필요가 있다. 필자가 (이전 논문들과 함께) 이 논문에서 하고자 하는 작업은 인과 개념과 설명 개념 사이에 너무 멀지도 않고 너무 가깝지도 않은 적절한 거리를 설정하는 작업이다. 이 작업은 자체로도 흥미롭지만 다양한 철학적 함축과 적용 사례를 갖는 작업이라고 필자는 믿으며 그 하나의 적용 사례를 암시하는 것으로 이 논문을 마무리하고자 한다.

D. 루이스의 기념비적인 논문(Lewis 1983) 이후로 오늘날 많은 형이상학자들은 완벽하게 자연적인 속성, 완벽하지는 않지만 상당히 자연적인 속성, 그리고 자연적이지 않은 속성의 구분을 받아들인다. 이들 가운데 완벽하게 자연적인 속성과 자연적이지 않은 속성은 비교적 쉽게 파악되는 개념이다. 우리는 종종 근원적인 물리학에서 가장 원초적인 속성으로 생각하는 속성들, 예컨대 질량, 전하, 스핀과 같은 속성들이 완벽하게 자연적인 속성의 예라고 생각한다. 이런 속성들은 근원적인 물리 법칙에 등장하는 속성들이다. 반면 우리는 초랑(grue) 역설에 등

---

36) 오늘날 아리스토텔레스 학자들이 aitia를 어떻게 이해하는지는 다음의 프리랜드의 언급에 잘 나타나 있다. “아리스토텔레스 학자들은 사실상 만장 일치로 사원인설이 설명이론이라는 의견을 갖고 있다. 예를 들어 줄리어스 모라브직은 이 이론이 ‘적절하게 이해될 경우’, ‘적절한 설명을 구성하는 것이 무엇인지에 대한 아리스토텔레스의 설명이다’라고 말하며 줄리아 안네스는 비슷하게 ‘aitia를 원인(cause)으로 생각하지 않고 설명, 즉 ‘때문에(because)’로 취급하게 된 것은 대단한 발전이다’라고 지적한다”(Freeland 1991, p. 49).

장하는 초랑과 같은, 소위 게리맨더링된 속성은 전혀 자연적이지 않은 속성이며 과학적으로 아무런 존중을 받을 가치가 없는 속성라고 생각 한다. 물론 완벽하게 자연적인 속성을 정의하는 것도 쉽지 않은 문제이기는 하지만<sup>37)</sup>, 일단 우리가 자연적인 속성과 자연적이지 않은 속성의 구분을 받아들일 준비가 되어 있다면 이 개념들은 그리 신비한 개념들은 아니다. 필자가 보기에도 오히려 정말 파악하기 어려운 개념은 완벽하게 자연적이지는 않지만 상당히 자연적인, 따라서 과학적으로 충분히 존중받을 수 있는 속성이라는 개념이다. 루이스는 자연성의 정도를 완벽하게 자연적인 속성들을 통해서 정의할 때 요구되는 복잡성으로 성격 규정한다(Lewis 1986a). 그러나 이 방법은 만족스럽지 않다.<sup>38)</sup> 예를 들어, 위에서 검토된 사례에 등장하는 “C-섬유활성화  $\vee$  D-섬유활성화  $\vee$  E-섬유활성화  $\vee \dots$ ”라는 선언적 속성(즉, 고통)을 생각해보자. 이 속성을 “C-섬유활성화  $\vee$  5kg”과 비교해 보자. 이 경우 후자가 더 단순한 정의를 갖는 것으로 보이지만 과학적으로는 전자가 후자보다 좀 더 존중받을 수 있는 속성으로 보인다.<sup>39)</sup> 따라서 우리는 속성의 자연성의 정도를 설명할 수 있는 새로운 이론이 필요할 것으로 보인다. 이런 문제와 관련해서 우리는 이 논문에서 다루어진 속성이 갖는 인과 설명적 힘을 활용할 수 있다. 즉, 우리는 게리맨더링 된, 따라서 아무런 과학적

37) 루이스의 성긴(sparse) 속성에 대한 성격 규정은 다음을 볼 것. Lewis (1986a), p. 60. 이 속성이 갖는 문제에 대해서는 다음을 볼 것. Thompson (2016).

38) 사실 루이스는 “정의의 복잡성의 정도”라는 개념에 대한 충분한 설명을 제공하지 않는다. 이런 이유로 N. 톰슨은 “단순성에 대한 명확한 설명이 없기 때문에 우리는 단순성을 통한 자연성의 분석이 성공적이라고 생각할 어떤 이유도 갖지 못한다”라고 주장 한다(Thompson 2016, p. 394, 각주 4).

39) T. 사이더 역시 기능적으로 정의된 (무한한) 선언적 속성이 꽤나 높은 자연성을 가질 수 있다는 것을 보여줄 필요가 있다고 생각하며, 그것을 위해 기능적 개념(예컨대 “캥거루 역할”과 같은 것)을 통해 비교적 복잡하지 않게 무한한 선언적 속성을 정의하는 방법을 고려한다(Sider 2011, p. 130). 그러나 “고통 역할을 수행함”과 같은 속성은 결코 완벽하게 자연적인 속성이 될 수 없으며, 이런 기능적 속성을 완벽하게 자연적인 속성들을 통해 정의하는 것은 그렇게 단순한 일로 보이지 않는다. 따라서 사이더의 (변형된 루이스적) 방식은 고통이 “C-섬유활성화  $\vee$  5kg”보다 더 단순한 정의라는 것을 보여주기에 적합해 보이지 않는다.

가치가 없는 속성과 선언적이지만 과학적으로 가치가 있는, 따라서 상당한 정도의 자연성을 갖는 속성의 차이를 그 속성이 갖는 설명적 힘을 통해서 구분할 수 있다. 이 접근법에 따르면 (고통과 같은) 어떤 선언적 속성들은 인과적 힘을 갖지 않지만 인과 설명적 힘을 갖는 것을 통해서 상당한 수준의 자연성을 가질 수 있다. 반면 “C-섬유 활성화 √ 5Kg”과 같은 속성은 이 선언적 속성을 구성하는 선언지들이 갖는 전혀 다른 인과적 힘 때문에 인과 설명적 힘을 갖지 못한다.<sup>40)</sup> 더 나아가 우리는 속성이 갖는 설명적 힘을 설명의 통일성, 단순성, 유용성 등을 통해서 추가적으로 분석할 수 있다. 이를 통해서 우리는 자연성의 정도를 분석할 수 있다. 이 작업은 여기서 암시된 것보다 훨씬 정교한 정식화를 요구할 것이지만 필자가 보기에도 루이스의 제안보다는 자연성의 정도를 정의하는 훨씬 전망이 좋은 접근으로 보인다.

---

40) 독자들은 앞서 필자가 선언적 설명이 성립하기 위해서는 그 선언적 설명형이 긍정적 레시피가 될 수 있어야 하며 따라서 선언지들이 모두 피설명형에 인과적으로 유관해야 한다고 언급했던 것을 기억할 것이다.

## 참고문헌

- 서울대학교병원 의학정보 ‘대사 증후군’ <https://terms.naver.com/entry.naver?docId=926836&cid=51007&categoryId=51007> (검색일: 2025.08.12.)
- 선우환 (2020), 『때문에- '때문에'의 의미에 대한 철학적 연구』, 아카넷.
- 이재호 (2013a), 「귀납에 대한 설명주의적 접근과 흡적 설명이론 - 암스트롱의 논변을 중심으로」, 『과학철학』 16권 1호, pp. 57-80.
- \_\_\_\_\_ (2013b), 「비사례적 규칙성 설명과 귀납에 대한 설명주의적 접근」, 『철학적 분석』 27권, pp. 1-30.
- \_\_\_\_\_ (2013c), 「흡적 설명이론들, 그리고 이들의 형이상학적 전제」, 『철학사상』 49권, pp. 207-236.
- \_\_\_\_\_ (2014), 「아리스토텔레스, 뒤힘, 밀, 그리고 설명에 관한 네 번째 길」, 『철학사상』 53권, pp. 165-202.
- \_\_\_\_\_ (2019), 「왜 체계적 과잉 결정은 존재하지 않는가?」, 『철학적 분석』 41권, pp. 1-31.
- \_\_\_\_\_ (2020a), 「기술적 법칙 이론과 존재론적 경계성 논증」, 『철학적 분석』 43권, pp. 1-31.
- \_\_\_\_\_ (2020b), 「필연적 연결과 경험주의」, 『논리연구』 23권 3호, pp. 195-230.
- \_\_\_\_\_ (2022), 「필연적 연결에 대한 이해 불가능성 논증과 상상 가능성」, 『철학적 분석』 47권, pp. 1-28.
- \_\_\_\_\_ (2023a), 「부분집합 실현 개념은 비환원적 물리주의를 구제할 수 있는가?」, 『철학탐구』 72권, pp. 209-238.
- \_\_\_\_\_ (2023b), 「신의 레시피로서의 설명 - 새로운 조작주의적 설명 이론」, 『과학철학』 26권 3호, pp. 1-37.
- \_\_\_\_\_ (2024a), 「선택적 설명에 대해서 어떻게 생각할 것인가?」, 『과학철학』 27권 3호, pp. 1-34.
- \_\_\_\_\_ (2024b), 「인과적 설명 이론의 딜레마와 평형 설명」, 『과학철학』 27권 1호, pp. 1-37.
- Armstrong, D. M. (1983), *What is a Law of Nature?*, Cambridge: Cambridge

- University Press.
- Bird, A. (2007), *Nature's Metaphysics: Laws and Properties*, Oxford: Oxford University Press.
- Cohen, S. (2002), "Basic Knowledge and the Problem of Easy Knowledge", *Philosophy and Phenomenological Research* 65: pp. 309-29.
- Fodor, J. (1997), "Special Sciences: Still Autonomous After All These Years", *Philosophical Perspectives* 11: pp. 146-63.
- Foster, J. (1983), "Induction, Explanation and Natural Necessity", *Proceedings of the Aristotelian Society* 83: pp. 87-101.
- Freeland, C. A. (1991), "Accidental Causes and Real Explanations", in L. Judson (ed.), *Aristotle's Physics: A Collection of Essays*, Oxford: Clarendon Press, pp. 286.
- Hall, E. J. (2004), "Two Concepts of Causation", in J. D. Collins, E. J. Hall and L. A. Paul (eds.), *Causation and Counterfactuals*, Cambridge: MIT Press, pp. 225-76.
- Hempel, C. G. (1945), "Studies in the Logic of Confirmation", *Mind* 54: pp. 1-26.
- Kim, J. (1992), "Multiple Realization and the Metaphysics of Reduction", *Philosophy and Phenomenological Research* 52: pp. 1-26.
- \_\_\_\_\_ (1999), "Hempel, Explanation, Metaphysics", *Philosophical Studies* 94: pp. 1-20.
- \_\_\_\_\_ (2007), "Causation and Mental Causation", in B. P. McLaughlin and J. Cohen (eds.), *Contemporary Debates in Philosophy of Mind*, Oxford: Blackwell, pp. 227-42.
- Kitcher, P. (1989), "Explanatory Unification and the Causal Structure of the World", in P. Kitcher and W. C. Salmon (eds.), *Scientific Explanation*, Minneapolis: University of Minnesota Press, pp. 410-505.
- Lee, J. (2015), "Inference to the Best Explanation and Disjunctive Explanations", *Organon F* 22: pp. 143-62.

- Lewis, D. (1983), "New Work for a Theory of Universals", *Australasian Journal of Philosophy* 61: pp. 343-77.
- \_\_\_\_\_(1986a), *On the Plurality of Worlds*, Oxford: B. Blackwell.
- \_\_\_\_\_(1986b), *Philosophical Papers* 2. Oxford: Oxford University Press.
- \_\_\_\_\_(2009), "Ramseyan Humility", in D. Braddon-Mitchell and R. Nola (eds.), *Conceptual Analysis and Philosophical Naturalism*, Cambridge: MIT Press, pp. 203-22.
- Loewer, B. (2007), "Mental Causation, or Something Near Enough", in B. P. McLaughlin and J. Cohen (eds.), *Contemporary Debates in Philosophy of Mind*, Oxford: Blackwell, pp. 243-64.
- Morris, K. (2019), *Physicalism Deconstructed: Levels of Reality and the Mind-Body Problem*, Cambridge: Cambridge University Press.
- Putnam, H. (1975), "Philosophy and Our Mental Life", in Mind, Language, and Reality: *Philosophical Papers* 2. Cambridge: Cambridge University Press, pp. 291-303.
- Russell, B. (1912-13), "On the Notion of Cause", *Proceedings of the Aristotelian Society* 13: pp. 1-26.
- Salmon, W. C. (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press.
- \_\_\_\_\_(1997), "Causality and Explanation: A Reply to Two Critiques", *Philosophy of Science* 64: pp. 461-77.
- Shoemaker, S. (1980), "Causality and Properties", in R. Taylor and P. Van Inwagen (eds.), *Time and Cause: Essays Presented to Richard Taylor*, Dordrecht: Reidel Pub. Co., pp. 109-35.
- \_\_\_\_\_(2001), "Realization and Mental Causation", in C. Gillett and B. Loewer (eds.), *Physicalism and Its Discontents*, Cambridge: Cambridge University Press.
- Sider, T. (2011), *Writing the Book of the World*, Oxford: Clarendon Press.

- Skow, B. (2016), *Reasons Why*, Oxford: Oxford University Press.
- Sober, E. (1983), “Equilibrium Explanation”, *Philosophical Studies* 43: pp. 201-10.
- Thompson, N. (2016), “Is Naturalness Natural?”, *American Philosophical Quarterly* 53: pp. 381-95.
- Walter, S. (2006), “Multiple Realizability and Reduction: A Defense of the Disjunctive Move”, *Metaphysica* 7: pp. 43-66.
- Wilson, J. (2021), *Metaphysical Emergence*, Oxford: Oxford University Press.

논문 투고일	2025. 08. 12.
심사 완료일	2025. 09. 30.
게재 확정일	2025. 09. 30.

---

## The Causal Structure and the Explanatory Structure of the World

---

Jaeho Lee

---

In this paper, I examine the relationship between the causal structure and the explanatory structure of the world. First, I will argue that we must accept the existence of disjunctive explanations and that my theory of explanation best accounts for our intuitions about them. Next, I will argue that my theory of explanation makes it possible to construct a model that reconciles a multi-layered explanatory structure with a flat causal structure of the world. Finally, I will argue that given this model, we have good reason to believe that the picture science provides reveals the world's explanatory structure, not its causal structure.

Keywords: causation, explanation, causal structure, explanatory structure,  
disjunctive explanation

