

비환원주의, 인과적 배제, 필수 실연*

강진호

(서울대 철학과 대학원)

1. 서론

물리주의적인 세계관이 지배적인 오늘날에 있어서 물리적 영역이 인과적으로 폐쇄적이고 기초적이라는 생각은 이미 낫설지 않은 견해이다. 즉 우리는 모든 물리적 사건이 어떤 다른 물리적 사건을 자신의 완전하고 충분한 원인으로 가지고 있기 때문에 임의의 물리적 사건에 대한 원인을 찾기 위해서 물리적 영역 밖으로 나갈 필요가 없다는 것을 받아들인다(물리적 인과의 폐쇄성 원리). 우리는 또한 세계의 모든 실재들과 속성들의 존재가 물리적 실재들과 속성들에 어떤 방식으로든 의존하고 있다는 사실을 받아들인다(물리적 기초성 원리). 이 두 원리와 더불어, 물리주의적 세계관 하에서 우리는 일반적으로 세계가 실재들 및 이들을 특징짓는 속성들이 위계적으로(hierarchcially) 층을 이룬 구조로 이루어져 있다는 견해를 받아들이고 있다. 이러한 견해에 따르면 세계는 물리적 단계, 화학적 단계, 생물학적 단계, 심리적 단계...와 같은 단계(level)들의 배열로 이루어지고, 각각의 단계는 두 요소들, 즉 그 단계를 위한 개별자들의 영역을 구성하는 실재들의 집합과 그 실재들을 특징지워주는 속성들의 집합으로 이루어진다. 이러한 견해에서 각 단계의 실재들은 흔히 부분 전체론적 관계(mereological relation)를 이룬다고 여겨진다. 즉 상위 단계의 실재들은 하위 단계의 실재들이 그것들의 부분을 이룸으로써 구성된다는 것이다. 그렇다면 각 단계의 실재들을 특징짓는 속성은

* 본 논문은 석사학위 논문(강진호 1997)의 일부를 줄이고 다듬은 것이다.

인접한 단계의 실재들을 특징짓는 속성과 어떠한 관계를 맺고 있는가?

비환원주의는 위의 물음에 대한 가능한 답변들 중 하나로서, 상위 단계의 속성이 하위 단계의 속성에 어떠한 방식으로 의존하고 있기는 하지만 하위 단계의 속성으로 환원될 수 없으며 구별되는 속성이라고 주장하는 견해를 가리킨다. 그러나 더이상의 논의를 진행시키기에 앞서서, 우선 여기서 말하는 환원이 의미하는 바가 무엇인지를 명확하게 하는 것이 좋을 듯 싶다. 분석 철학에서 일반적으로 받아들여지는 견해에 따르면, 환원은 기본적으로 과학 이론들간의 관계이며, 속성들간의 환원은 그 속성들을 다루는 이론들 사이의 환원 여부에 의해 결정된다. 환원 문제가 논의될 때 가장 자주 언급되는 네이글(Nagel) 환원 모형에 의하면, 환원되는 이론의 용어들과 토대 이론의 용어들을 연결시켜주는 교량 원리(bridge principle)들이 성립하고(연결가능성 조건) 이를 통해서 환원되는 이론의 법칙들이 토대 이론의 법칙들로부터 도출된다면(도출가능성 조건), 그리고 오직 그 경우에만, 환원되는 이론은 토대 이론으로 환원된다.¹⁾ 그러나 여기서 환원 관계가 요구하는 용어들간의 연결이란 어떠한 종류의 것인가? 네이글의 최초 제안은 도출 가능성 조건이 만족되기만 한다면 용어들이 쌍조건문으로 연결되지 않고 조건문으로 연결되어도 괜찮다는 것이었다. 그러나 이후의 많은 철학자들은, 진정한 환원의 경우에 있어서 환원되는 이론의 용어들과 토대 이론의 용어들은 (법칙적 필연성을 가진) 쌍조건문으로 연결되어야 한다고 주장하였다. 이는 다음과 같은 두가지 이유에서이다. 첫번째로, 만약 그러한 쌍조건문들이 성립하면 도출가능성 조건은 자동적으로 만족될 수 있다. 왜냐하면 그 경우에 우리는 환원되는 이론의 모든 용어들을 각각 (쌍조건문이 보장하는) 그것과 동연적(coextensive)인 토대 이론의 용어들로 대치시킬 수 있으며, 이를 통해 환원되는 이론의 모든 법칙들을 오직 토대 이론의 용어들만이 나타나는 토대 이론의 법칙들로 번역할 수 있기 때문이다. 두번째로, 그러한 쌍조건문들은 환원되는 이론의 용어들과 토대 이론의

1) Nagel (1961), 11장.

용어들을 이론적으로 동일시하는 것을 가능하게 해주며, 이에 따라 우리는 형이상학적으로 볼 때 환원의 가장 중요한 목적인 존재론적 단순성을 성취할 수 있다. 이러한 두가지 이유의 타당성을 받아들여서 환원에 대한 이들의 견해를 따를 경우, 속성간의 환원은 다음과 같이 정의될 수 있을 것이다. (여기서 “ T_R ”은 환원되는 이론을, “ T_B ”는 토대 이론을 가리킨다)

(RP) T_R 의 용어들과 T_B 의 용어들간에 성립하는 쌍조건문을 사용하여 T_R 의 법칙들이 T_B 의 법칙들로부터 도출되고 그 결과로서 T_R 의 속성들이 T_B 의 속성들과 이론적으로 동일시 될 경우, 그리고 오직 그 경우에만, T_R 의 속성들은 T_B 의 속성들로 환원된다.

(RP)는 본질적으로 여전히 네이글적인 환원 개념에 속한다고 볼 수 있으며, 따라서 네이글 환원 모형에 대해서 가해진 많은 비판들, 특히 네이글 모형이 환원에 대해 지나치게 이상적인 견해를 가지고 있다는 비판에서 벗어날 수 없다. 그러나 네이글 환원 모형에 대한 비판들은 적어도 현재 우리가 논의할 주제들과 관련해서는 그리 치명적인 문제점들이 아니므로 무시될 수 있다. 더구나 네이글 모형에 대한 많은 비판에도 불구하고, 오늘날 가장 널리 받아들여지는 환원 모형들은 네이글 모형의 두가지 핵심적인 조건인 연결가능성 조건과 도출가능성 조건을 여전히 자신들의 핵심적인 조건으로 간주하고 있다.²⁾ 그러므로 나는 환원 개념을 규정짓는 틀로서 (RP)를 받아들이고, 이를 바탕으로 하여 본 논문을 전개시켜 나아가겠다.

특별히 심신 문제(mind-body problem)에 있어서 비환원주의는, 정신적 속성들이 물리적 속성들에 의존하고 있기는 하지만 물리적 속성들로 환원될 수 없는 독자적인 영역을 이룬다는 견해를 가리킨다. 비환원주의는 심리 철학에 있어서 지난 30여년간 심신 문제에 대해 가장 널리 받아들여지고 있는 견해였다. 비환원주의가 그렇게 지지를 받을 수 있었던 가장 큰 이유는 퍼트남(Putnam)이 처음으로 제시한 복수 실현 논변(multiple realization argument) 때문이라고 볼 수

2) 가령 Schaffner (1992)의 환원 모형을 참조할 것.

있다. 복수 실현 논변은 정신적 속성이 다양한 물리적 속성에 의해 실현될 수 있다는 복수 실현 입론³⁾에 근거하여, 정신적 속성이 어떤 단일한 물리적 속성으로 환원될 수 없는 독자성을 가진다고 주장하였다. 복수 실현 논변은 일반적으로 널리 받아들여졌고, 이 논변은 심리학이 자율적인 특수 과학으로 성립할 수 있는지에 대한 형이상학적 토대를 이루었다. 여기서 주목할 만한 점은, 복수 실현 논변이 정신적 속성의 독특한 성격들인 자기 의식, 감각질, 지향성 등과 전혀 무관하게 성립하는 논변이었다는 점이다. 이는 복수 실현 논변이 두 속성 영역 및 그 속성들과 관련된 과학 이론들에 대한 보다 일반적인 논변으로 확대될 수 있음을 의미한다. 실제로 포더(Fodor)는 복수 실현 논변을 과학 이론 일반에 적용하여서, 보다 상위의 과학 이론들은 보다 하위의 과학 이론들로 환원될 수 없는 자율적 영역을 이루고 있음을 논증하였다.⁴⁾

그러나 최근 몇년간 심신문제에 있어서 비환원주의는 김재권의 일련의 논문들을 통해 지속적으로 공격을 받아왔다.⁵⁾ 김재권은 비환원주의가 심신 문제에 있어서 핵심적 문제인 심성 인과(mental causation)를 설명하는 데 있어 난점을 드러내고 있으며, 그 중에서도 특히 그가 말하는 인과적 배제의 문제(the problem of causal exclusion)를 해결할 수 없다고 주장한다. 이와 더불어 김재권은 비환원주의의 형이상학적 기반이 된 복수 실현 입론이 가지는 함축을 새롭게 재해석하고 있다. 김재권에 따르면 복수 실현 가능성은 정신적 속성 및 심리학의 자율성을 보장하는 것이 아니라, 거꾸로 정신적 속성이 자율적 속성이 아니며 신경생리적 속성으로 국소적으로 환원됨을, 또한 심리학이 자율적 과학이 될 수 없으며 보다 하위의 과학인 신경생리학으로 국소적으로 환원됨을 보여준다는 것이다.

3) 나는 본 논문에서 '복수 실현 입론'과 '복수 실현 논변'을 구분해서 쓰겠다. 전자는 정신적 속성이 다양한 물리적 속성에 의해 실현될 수 있다는 주장만을 가리킨다. 반면 후자는, 정신적 속성이 다양한 물리적 속성에 의해 실현될 수 있기 때문에 정신적 속성은 단일한 물리적 속성으로 환원될 수 없는 독자성을 가진다는 주장을 가리킨다.

4) Fodor (1974).

5) 특히 Kim (1989), Kim (1992), Kim (1993a)를 참조할 것.

본 논문의 목적은, 비환원주의에 대한 김재권의 위와 같은 비판을 고찰해보고, (어떤 일정한 제한하에서) 비환원주의를 옹호하는데에 있다. 본 논문의 구성은 다음과 같다. 먼저 2절에서 나는 김재권이 말하는 인과적 배제의 문제를 고찰하고, 비환원주의적 틀 내에서 이 문제를 해결해볼 수 있는 가능성을 제시하겠다. 동시에 나는, 복수 실현 논변이 타당하다는 전제하에서만 여기서 제시된 해결책이 성립될 수 있음을 보임으로써, 인과적 배제의 문제의 해결이 복수 실현 논변과 밀접한 관련을 맺고 있음을 보이겠다. 3절에서 나는 복수 실현 논변에 대한 김 재권의 비판을 자세히 살펴보겠다. 이어 4절에서는 복수 실현 논변에 대한 김 재권의 비판이 가지고 있는 문제점을 지적하고, 복수 실현 논변의 타당성을 옹호하겠다. 마지막으로 나는 5절에서, 앞서 옹호된 형태의 비환원주의가 일반적으로 기대되는 것보다 다소 약한 형태임을 인정하고, 그럼에도 불구하고 왜 그것이 여전히 진정한 의미에서 “비환원주의”로 불릴 수 있는지를 설명하려고 한다.⁶⁾

2. 인과적 배제의 문제

김재권은 비환원주의자라면 누구라도 받아들일 수밖에 없는 전제들로부터 비환원주의자가 인과적 배제의 문제에 직면하게 됨을 다음과 같이 논변하고 있다.⁷⁾

먼저 비환원주의자가 받아들이는 전제들을 살펴보자. 첫번째로, 그는 물리주의자이며 따라서 서론에서 밝힌 것과 같은 물리적 인과의 폐쇄성 원리와 물리적 기초성 원리를 받아들인다. 특히 기초성 원리와 관련해서, 비환원주의자는 일반적으로 모든 정신적 속성의 사례가

6) 여기서 내가 옹호하려는 것은 비환원주의 일반이며 심신 문제라는 특정한 영역에서의 비환원주의가 아님을 주의하라. 그러므로 나는 본 논문에서, 특별히 정신적 속성과 관련된 두가지 문제, 즉 의식(consciousness)과 심성 내용(mental contents)의 문제는 다루지 않겠다.

7) 이하의 논의는 주로 Kim (1993a)를 참조한 것이다. 이외에도 Kim (1989), 279-280을 참조할 것.

어떤 물리적 속성의 사례에 의해 실현된다는 물리적 실현 이론(physical realization thesis)을 받아들인다.⁸⁾ 아직 물리적 실현이 정확히 무엇을 의미하는지에 대해 일치된 견해는 없지만, 김재권은 이를 일단 다음과 같이 이해하고 있다: 물리적 속성 P가 체계 s에서 M을 “실현시킨다”고 할 때, P는 s에서 M의 구현을 위한 인과적 메커니즘을 제공하는 s의 미시물리적 속성을 규정해야 한다.⁹⁾ 나는 일단 이 규정을 받아들이고 본 논문을 진행시켜 나아가겠다.

동시에, 비환원주의자는 정신적 속성에 대한 반환원주의 이론(anti-reductionism thesis)을 받아들인다. 즉 그는 환원주의자와 달리 정신적 속성이 물리적 속성으로 환원될 수 없다고 생각한다. 여기서 김재권은, 물리적 속성으로서의 환원 불가능성 때문에 정신적 속성을 적법한 속성의 영역에서부터 추방하는 제거주의자와는 달리 비환원주의자는 심성 실재론(mental realism), 즉 정신적 속성들이 허구적 존재이거나 예측을 위한 단순한 보조 수단이 아니라 진정으로 존재한다는 것을 받아들인다는 점을 지적한다. 그리고 김재권은, 정신적 속성이 실재한다는 주장은 정신적 속성이 인과적 힘(causal power)을 가지고 있다는 주장 이외의 것이 될 수 없다는 자신의 견해를 덧붙인다. 도대체 어떤 속성이 인과적 힘을 가지고 있지 않다면 설령 그것이 있다고 하더라도 세계의 변화에 하등 영향을 끼치지 못할 것이며, 그러한 경우 우리는 그것이 있다고 말할 근거를 가지지 못하기 때문이다.

그러나 끝이어서 김재권은, 위와 같은 전제들을 모두 받아들일 경우 우리는 다음과 같은 문제점에 부딪힐 수밖에 없다고 주장한다. 어떤 물리적 속성 E의 사례(이하 “e”)가 정신적 속성 M의 사례(이하 “m”)에 의해 야기되었다고 하자. 비환원주의자는 심성 실재론에 의해서

8) 물리적 실현이외에도 정신적 속성과 물리적 속성의 의존 관계를 정식화하는 데에 널리 쓰이는 것으로는 수반(supervenience) 개념이 있다. 그러나 나는 물리적 실현에 대한 본 논문에서의 논의가 별다른 수정 없이 수반에 대해서도 적용될 수 있다고 생각하기 때문에, 수반에 대해서는 논의하지 않겠다. 수반에 대한 일반적인 논의에 대해서는 Kim (1990)와 Horgan (1993)을 참조.

9) Kim (1993a), 343쪽.

M이 E에 대해 인과적 힘을 끼쳤다는 것을 받아들인다. 그러나 또한 비환원주의자는 물리적 실현 입론에 의해, M이 어떤 물리적 속성 P에 의해 실현되었고 이 P의 사례(이하 “p”)가 m을 발생시키는데에 충분하다는 것에 동의한다. p가 m을 발생시키는 데 충분하고 m이 e를 야기시켰으므로, p 또한 e를 야기시키는데에 충분하다.¹⁰⁾ 더구나 인과적 폐쇄성 원리에 따라 e에 대한 물리적 원인이 항상 있다. 그러므로 우리는 p를 e에 대한 원인으로 간주할 수 있다.

결국 우리는 e에 대한 원인으로서 m과 p라는 두 원인을 가지고 있는 셈이다. 비환원주의 입론에 의하면 m과 p는 서로 구별되고 동일하지 않다. 그러므로 비환원주의자의 전제에 따르면 우리는 e에 대한 원인으로 서로 구별되는 두 원인을 가지게 된다. 그러나 물리주의의 전제하에서 과연 이것이 가능한가? 우리는 두 원인을 각각 e에 대한 부분원인이라고 간주할 수 없다. 왜냐하면 이는 p 단독으로는 e에 대한 원인이 될 수 없음을 의미하고 따라서 인과적 폐쇄성 원리를 위반하기 때문이다. 그렇다고 과도결정(overdetermination)의 상황, 즉 두 원인 모두가 e에 대한 충분하고 완전한 원인이 된다는 것을 인정할 수도 없다. 왜냐하면 그러한 경우 어떤 반사실적 상황에서 m은 단독으로 e에 대한 충분하고 완전한 원인이 될 수 있을 텐데, 이 또한 인과적 폐쇄성 원리에 위반되기 때문이다. 도대체 폐쇄성 원리에 의해서 p가 e에 대한 완전하고 충분한 원인임이 보장되는 한, m이 할 무슨 인과적 역할이 있단 말인가? 다시 말해서 p는 m을 e에 대한 원인으로서 배제하지 않는가? 이것이 김재권이 말하는 인과적 배제의 문제이다.

얼핏 보기에 이 문제는 비환원주의자들에게 그리 심각한 위협이 되지 않는 듯 싶다. 그 이유는 위의 논변이, 비환원주의 입론으로부터 m과 p가 동일하지 않음을 추론할 수 있다는 것을 전제하고 있기 때문이다. 그러나 이것이 가능하려면 비환원주의 입론은 다음과 같은

10) 여기서 김재권은 인과에 대한 법칙적 견해, 즉 m을 e에 대한 충분조건으로 만드는 적절한 인과적 법칙이 있을 경우 그리고 오직 그 경우에만 m은 e를 야기시킨다는 견해를 전제하고 있다. 나는 이러한 그의 가정이 이하의 논의에 있어서 별다른 영향을 끼치지 않는다고 생각하기 때문에, 이를 문제삼지 않겠다.

주장을 부정하는 것이어야 한다.

(A*) [개별자 물리주의] 정신적 속성의 사례는 물리적 속성의 사례와 동일하다.

그러나, 실제로 비환원주의 입론이 부정하는 것은 (A*)가 아니라 단지 다음과 같은 주장이다.

(A) [유형 물리주의] 정신적 속성은 물리적 속성과 동일하다.

그러므로 김재권의 논변이 성립하려면, (A)에 대한 부정으로부터 (A*)에 대한 부정이 도출될 수 있다는 것이 전제되어야 한다. 그러나 그럴 수 없다는 것이 비환원주의의 핵심적 주장들 중 하나가 아니었는가? 즉 비환원주의의 핵심적 주장들 중 하나가, 바로 유형 물리주의를 부정하면서도 개별자 물리주의를 인정할 수 있다는 것 아니었는가? 그러므로 비환원주의의 유형/개별자 구분을 받아들인다면, 우리는 (A)를 거부하면서 동시에 여전히 (A*)를 인정할 수 있을 것 같다. 그렇다면 우리는, (A)를 거부함으로써 속성 M과 P가 동일하지 않다고 주장하면서도 (이 경우 비환원주의의 입론은 유지된다), 동시에 (A*)를 인정함으로써 속성 사례 m과 p가 동일하다는 것은 받아들이는 방법 (이 경우 e에 대해서는 단지 하나의 원인이 있을 뿐 두가지 원인이 있는 것이 아니다)을 통해, 배제의 문제를 해결할 수 있을 것 같다. 결국, 배제의 문제에 대한 해결책은 이미 비환원주의자들 스스로에 의해서 제시된 것이 아닌가?

그러나 문제가 그렇게 간단하지는 않다. 우선, 김재권이 자신의 논변에서 인과 관계를 맺고 있는 항으로서 속성 사례들을 사용하고 있으며 따라서 속성 예화 사건 이론(Property Exemplification Theory of Event, 이하 "PE")을 전제하고 있다는 점에 주목해보자.¹¹⁾ 그렇다면 PE의 동일성 기준으로부터 한가지 문제가 발생한다: PE의 동일

11) PE에 관한 자세한 설명을 위해서는 Kim (1976)을 참조할것.

성 기준은 다음과 같다.

- (I) 개체 x 가 시간 t 에 속성 P 를 예화한 사건을 e 라고 하고 개체 y 가 시간 t' 에 속성 Q 를 예화한 사건을 e' 라 할 때, x 와 y , P 와 Q , t 와 t' 가 동일할 경우 그리고 오직 그 경우에 만 e 와 e' 는 동일하다.

(I)에 따르면, 사건 e 와 e' 의 동일성은 e 와 e' 가 예화하는 속성 P 와 Q 의 동일성을 함축한다. 그러므로 우리가 PE 를 전제한다면, m 과 p 의 동일성은 M 과 P 의 동일성을 함축할 것이다. 이는 PE 하에서는 유형 물리주의를 거부하면서 개별자 물리주의를 받아들일 수 없으며, 따라서 유형 물리주의와 개별자 물리주의간의 구별이 아무 의미가 없다는 것을 보여준다. 그러므로, 유형/개별자의 구별을 통해 배제의 문제를 해결하려는 시도는 어려움에 부딪히게 된다.

그러나 더 본질적인 문제는, 설사 비환원주의자가 위와 같은 문제점을 극복하고 유형 물리주의와 개별자 물리주의를 구별할 수 있다고 하더라도¹²⁾, 다음과 같은 난점을 극복할 수 없다는 점에 있다. 그 난점이란, 만약 M 과 P 가 동일하지 않다는 것이 어떤 식으로 확보된다고 해도 M 과 P 양자가 단일한 물리적 속성 E 에 대해 인과적으로 효력이 있어야 하는 이상, 이번에는 배제의 문제가 속성의 차원에서 다시 제기된다는 것이다. 즉, 물리적 폐쇄로부터 쉽게 추론할 수 있듯이 P 가 E 에 대해서 완전하고 충분하게 인과적으로 효력이 있다는 것을 받아들이면, M 이 해야 할 어떠한 부가적인 인과적 역할이 남아있단 말인가? P 의 인과적 힘은 E 에 대한 M 의 인과적 힘을 배제해버리지 않는가? 이 문제는 비환원주의에게 있어서 특별히 치명적인 것처럼 보인다. 비환원주의에 대한 지금까지의 많은 비판들은 주로 정신적 속성이 어떻게 인과적 힘을 가질 수 있는가라는 문제에 초점을 맞추

12) 유형/개별자의 구분을 살리는 한가지 방법은 PE 대신 다른 사건 이론을 받아들이는 방법이다. 가령 사건을 비구조적 개체로 보는 Davidson의 사건 이론 하에서는 유형/개별자의 구분이 가능하다. 그의 사건 이론에 대해서는 특히 Davidson (1969)를 참조.

었으며, 이에 따라 많은 비환원주의자들은 정신적 속성에 인과적 힘을 부여할 수 있는 대안을 마련하는 데 자신들의 노력을 기울여왔다.¹³⁾ 그러나 속성에 대한 배제의 문제는, 비환원주의자가 설사 어떤 환원될 수 없는 인과적 힘을 정신적 속성에 대해 확보하는데 성공했다고 하더라도 그것이 물리적 속성의 인과적 힘에 의해 미리 선취될 것임을 보여주고 있다. 즉 정신적 속성이 인과적 힘을 가지고 있다고 하더라도, 그것은 아무 쓸모도 없어지게 되는 것이다. 이러한 결론을 피할 수 있는 유일한 방법은 정신적 속성을 물리적 속성과 동일시하는 것 뿐인듯 싶다. 그러나 이는 바로 환원주의를 인정하는 것 아닌가? 다시 말해서 배제의 문제는, 환원주의를 받아들이지 않는 한 정신적 속성이 자신의 인과적 힘을 발휘할 가능성을 원리적으로 봉쇄하고 있지 않은가? 실제로 김재권 자신은 물리주의와 정신적 실재론의 전제를 거부하지 않으면서 인과적 배제의 문제를 해결할 수 있는 유일한 방법은 환원주의를 채택하는 길밖에 없다고 주장한다. 만약 사실이 그렇다면, 비환원주의에게 있어서 이보다 심각한 문제는 없을 것이다.

과연 우리가 이러한 난점으로부터 벗어날 수 있을까? 나는 그럴 수 있다고 생각한다. 내가 보기에 김재권의 논변은 한가지 문제점을 가지고 있다.

내가 지적하고 싶은 것은, 김재권이 자신의 논변에서 사용하고 있는 “물리적”이라는 용어의 의미와 관련되어 있다. 김재권은 자신의 논변에서, m 과 p 라는 두 원인을 가진 사건 e 가 물리적 사건이라는 것을 전제하고 있다. 그러나 e 는 과연 어떠한 의미에서 물리적 사건인가? 우리가 주의해야 할 점은 여기서 말하는 사건 e 가 행동이라는 사실이다. 그렇다면 서론에서 소개된 실재들과 속성들의 위계적 단계

13) 정신적 속성의 인과적 효력을 확보하려는 가장 잘 알려진 방법으로는 반사실문(counterfactuals)을 이용하는 방법과 조건부 법칙(*ceteris paribus law*)을 이용하는 방법이 있다. 전자에 대해서는 LePore and Lower (1987)과 Horgan (1989)를 참조할 것. 후자에 대해서는 Fodor (1989)를 참조할 것. 조건부 법칙이 과연 진정한 법칙일 수 있겠는가에 대한 논쟁으로는 Schiffer (1991)과 Fodor (1991)을 참조할 것.

를 고려해 볼 때, e 는 엄격한 의미에서의 물리적 사건, 즉 자신의 예화 속성인 E 가 속성들의 단계중 물리적 단계에 속한다는 의미에서 “물리적” “사건”일 수 없다. E 는 행동적 속성이고 따라서 심리적 단계에 속하기 때문이다. “행동적” 속성이 “심리적” 단계에 속한다는 주장이 매우 불합리하게 느껴질지도 모르겠다. 그러나 그러한 느낌이야말로 심신 관계를 특징지어주는 우리의 용어들이 데카르트적 이원론에 얼마나 뿌리깊게 몰들어있는지를 잘 보여주고 있다. 우리가 위와 같은 주장을 불합리하다고 느끼는 이면에는 심리학에 대한 다음과 같은 잘못된 생각, 즉 비록 우리가 마음이나 영혼과 같은 정신적 실체들을 인정할 수는 없다고 하더라도 여전히 심리학의 진정한 대상은 전통적인 의미에서의 정신적 속성들, 다시 말해서 우리의 의식과 관련된 속성들이어야 한다는 생각이 숨어 있다. 우리가 이러한 이원론적인 생각을 버리고 심리적 속성들을 그것들의 인과적 역할에 의해 규정짓는다면, 그리고 인과적 견지에서 볼 때 행동적 속성들은 엄격한 의미에서의 물리적 속성들에서 찾아볼 수 없는 복잡성과 예측불가능성을 가지고 있다는 점에 주목해본다면, 행동적 속성이 속해야 될 단계는 물리적 단계가 아니라 마땅히 심리적 단계이다. 만약 우리가 이 점을 인정할 수 있다면, 우리는 김재권이 자신의 논변에서 “물리적”이라는 용어를, 물리적 단계에 속하는 속성들에 대해서만 적용할 수 있는 것이 아니라 보다 상위의 단계들의 속성들에 대해서도 적용할 수 있는 넓은 의미로 사용한 것이라고 볼 수 있다. 그러나, 그렇다면 E 가 넓은 의미에서 물리적이라는 것은 도대체 무엇을 뜻하는가? (이제부터 나는 좁은 의미에서 물리적이라는 것을 “물리적”으로, 그리고 넓은 의미에서 물리적이라는 것을 “물리적”으로 표현하겠다.) 이는 E 가 물리적 단계의 어떤 속성으로 환원될 수 있음을 의미하는가? 그런 의미가 될 수는 없다. 바로 우리가 지금 다루고 있는 문제가 서로 다른 단계에 있는 속성들간에 과연 환원이 성립할 수 있느냐는 것인데, 여기서 E 가 환원가능하다고 가정해버리는 것은, 선결 문제 요구의 오류를 범하는 것이기 때문이다. 내가 보기에 물리적인 개념을 살려줄 수 있는 유일한 방법은 다음과 같이 물리적 실현의 개념에 호소하는 길 밖에 없는 것 같다. 즉, 어떤 속성은 그것이 물리적으로

로 실현가능하거나 또는 그 자신이 물리적 속성일 때 물리적이다. 나의 생각로는, 만약 우리가 이러한 해석을 받아들이지 않는다면 왜 E를 물리적 속성이라고 규정할 수 있는지를 설명할 길이 없어진다고 여겨진다.

그러나 김재권의 논변에서 쓰인 “물리적”이란 용어의 의미가 만약 위와 같은 물리적인 개념으로 해석될 수 있다면, 애초에 배제의 문제를 야기시켰던 근원인, 단일한 사건 (또는 속성)에 대해 두가지 충분한 원인 (또는 인과적으로 효력있는 속성)이 있다는 문제는 사라져버리고 만다. 왜냐하면, 위의 해석에 따를 경우에는 E를 실현시키는 어떤 물리적 속성(이를 “P*”라 하자)이 있을 것이고, 원래 E에 대해 인과적으로 효력있는 속성이라고 간주되었던 두가지 속성 M과 P는 실제로 각각 E와 P*에 대해 인과적으로 효력이 있을 것이기 때문이다. 즉 M과 P*간에 인과적 경쟁이 있을 필요가 없어지는 것이다. 결국 김재권의 논변에서 사용된 “물리적”이라는 용어가 가지고 있는 이중적 의미를 파악해내서 물리적인 것과 물리적인 것을 구분하면, 우리는 물리적 속성 E를 그것의 물리적 실현 토대 P*로부터 구분할 수 있고, 따라서 비환원주의의 틀 내에서 배제의 문제를 해결할 수 있다.

그러나 환원주의자는 이러한 답변에 대해 다음과 같이 반박할지도 모르겠다: 행동적 속성 E가 물리적 속성이 아니라 물리적 속성이라는 것을 인정한다고 하자. 그렇다고 해서 E가 자신의 물리적 실현 토대인 P*로 환원될 수 없다는 어떤 구체적 근거가 있는가? 단지 E가 P*로 환원가능하다는 환원주의의 가정이 선결 문제 요구의 오류를 범하고 있다는 사실만으로부터는, E가 P*로 환원되는 것이 불가능하다는 결론이 도출되지 않는다. 그러므로 비환원주의자가 배제의 문제를 뿌리뽑기 위해서는, E가 환원불가능하다는 것을 지지해줄 만한 어떤 적극적인 논변이 있어야 한다.

복수 실현 논변이 필요한 것은 바로 위와 같은 반박에 답하기 위해서이다. 왜냐하면 서론에서 언급한 것처럼, 복수 실현 이론은 정신적 속성에만 적용될 수 있는 것이 아니라 속성 일반에 적용될 수 있으며 따라서 행동적 속성에도 적용될 수 있기 때문이다. 가령 아픔의 예를 들어보자. 만약 (퍼트남이 자신의 논문에서 주장한 것처럼¹⁴⁾)

아픔이 포유류, 파충류, 연체동물 등에서 복수적으로 실현될 수 있다면, 아픔-행동 (가령 “짖그림”) 또한 포유류, 파충류, 연체동물 등에서 복수적으로 실현될 수 있을 것이다! (이러한 생물들이 보여주는 아픔-행동의 실현 토대인 근육들과 힘줄들은, 각각 그 구성에 있어서 그것들의 아픔의 실현 토대인 뉴우런들이 다양한 만큼이나 다양하다는 점을 기억하라). 그러므로 만약 복수 실현 논변이 타당하고 그래서 정신적 속성이 일반적으로 자신의 실현 토대 속성으로 환원되지 않는다면, 행동적 속성 또한 자신의 실현 토대로 환원되지 않을 것이다. 그렇다면 우리는 E가 P*로 환원가능하지 않다고 결론지을 수 있고 따라서 배제의 문제를 해결할 수 있다.

그러나 아직 문제가 완전히 해결된 것은 아니다. 그 이유는, 복수 실현 논변이 틀렸다고 김재권이 주장하고 있기 때문이다. 만약 그의 주장이 옳다면 행동적 속성을 그것의 실현 토대 속성과 구분함으로써 배제의 문제를 해결하려는 나의 시도는 정당화될 수 없을 것이다. 그러한 구분의 근거가 바로 복수 실현 논변이 타당하다는 것이었기 때문이다. 그러므로 김재권의 주장을 반박할 수 없다면, 인과적 배제의 문제는 여전히 비환원주의에게 남아있을 것이다. 그렇다면 왜 김재권은 복수 실현 논변이 틀렸다고 주장하는가? 나는 다음 절에서 그의 주장을 살펴보겠다.

3. 복수 실현 입론에 대한 김재권의 재해석

복수 실현 입론에 대한 김재권의 재해석은 다음과 같은 방식으로 이루어지고 있다.¹⁵⁾ 복수 실현 입론에 따르면 단일한 정신적 속성 사례 M이 (무한할 수도 있는) 다양한 신경생리적 속성들 N_1, N_2, N_3, \dots 에 의해 실현될 수 있다. 우선 김재권은 왜 M을 N_1, N_2, N_3, \dots 의 선

14) Putnam (1967), 228쪽.

15) 이절의 논의는 주로 Kim (1992)를 참조한 것이다. 김재권은 이밖에도 Kim (1994)에서 이 문제를 집중적으로 논의하고 있다.

인 속성 $N_1 \vee N_2 \vee N_3, \dots$ (이를 “N”이라고 하자)과 (이론적으로) 동일시할 수 없는지 물어보고 있다. 이 물음에 대해 포더는, 정신적 속성과 동일시되는 물리적 선언 속성 N의 각 선언지들이 일반적으로 서로 ‘터무니없이 이질적이고’(wildly heterogeneous), 이질적인 속성들의 선언은 진정한 속성, 다시 말해서 자연종(natural kind, 이하 ‘종’)이 될 수 없기 때문에 우리는 N을 속성으로 받아들일 수 없다고 주장한 바 있다.¹⁶⁾ 왜 이질적인 속성들의 선언은 종이 될 수 없는가? 포더에 따르면 그 이유는, 이질적인 속성들의 선언은 과학 이론의 법칙에서 나타날 수 있는 항이 될 수 없기 때문이다. 물론 포더의 주장이 옳다면 속성간의 환원은 성립하지 않는다. 서론에서 살펴본 것처럼 속성들간의 환원은 관련된 과학 이론들간의 환원을 전제로 하고 있는데, 만약 이질적인 속성들의 선언이 법칙의 항으로 나타날 수 없다면 한 이론의 법칙들로부터 다른 이론의 법칙들의 도출은 가능하지 않을 것이고, 따라서 속성들간의 환원의 선결 조건인 이론간의 환원은 성립할 수 없을 것이기 때문이다. 그러나 그렇다면 왜 이질적인 속성들의 선언은 법칙에서 나타날 수 없는가? 포더는 이 질문에 대해서는 별다른 대답을 하고 있지 않다.

여기서 김재권은, 마치 비환원주의의 입장을 옹호하는 것처럼, 위의 질문에 대해 다음과 같은 답변을 한다: 이질적인 속성들의 선언을 항으로 갖는 일반화는 자연 법칙이 갖추어야 할 조건 중 하나인 (굿맨(Goodman)적 의미에서의) 투사가능성(projectibility)을 갖추고 있지 않기 때문에 법칙으로 인정될 수 없다. 김재권은 옥(jade)의 예를 들어 이를 설명한다. 우리가 하나의 단일한 종으로 알고 있어왔던 옥은 실제로 경옥(jadeite)과 연옥(nephrite)이라는 화학적으로 이질적인 두 종으로 이루어져 있다는 사실이 밝혀졌다. 김재권은 그러한 경우, 가령 우리가 법칙이라고 여겨 왔던 (L) “모든 옥은 파랗다”라는 법칙은 투사불가능한 법칙이며, 따라서 더 이상 법칙이 아니라고 주장한다. 그 이유는 다음과 같다. 옥이 경옥과 연옥의 선언 종이라면 이때 (L)은 실제로 (L*) “모든 경옥이나 연옥은 파랗다”와 동일하고,

16) Fodor (1974).

(L*)는 다시, (L₁) “모든 경옥은 파랗다”와 (L₂) “모든 연옥은 파랗다”는 두 법칙의 연언과 동일하다. 그런데 만약 우리가 지금까지 관찰한 옥의 모든 사례들이 실제로는 모두 경옥이었다고 하자. (그럴 가능성을 배제할 수는 없다) 그런 경우 이 사례들은 명백히 (L₁)은 확증해주지만 (L₂)는 확증해주지 못한다. 그렇다면 (L)은 단순히 (L₁)이라는 잘 확증된 법칙과 (L₂)라는 (법칙으로서의 지위가 의심스러운) 보편문의 연언에 불과한 것이며, 그 자체가 법칙이 될 수는 없는 것이다.

그렇다면 무엇때문에 (L)은 투사가능하지 않게 되었는가? 여기서 김재권은 선언종을 진정한 속성으로 여길 수 없는 근본적인 이유를 제시한다. 김재권에 따르면 (L)이 투사가능하지 않게 된 이유는, (L)의 항인 옥을 구성하는 선언지들인 연옥과 경옥의 인과적 힘이 유사하지 않기 때문이다. 속성들이 가진 인과적 힘은 그 속성들을 향으로 하는 인과 법칙을 함의한다. 그리고 인과 법칙이라는 것은 투사가능한 규칙성이다. 그러므로 이질적인 선언 속성이 투사가능성 조건을 만족시키지 못하는 속성이라는 것은, 곧 그 속성의 선언지를 이루는 속성들의 인과적 힘이 서로 이질적이라는 것을 뜻한다. 결국 속성의 이질성이란 그 속성의 인과적 이질성 이외의 것이 될 수 없다. 과학에 있어서의 종은 곧 인과적 종인 것이다.

그렇다면 김재권 또한 이질적인 속성들의 선언은 그 자체 진정한 속성이 될 수 없다는 포더의 주장을 수용하는 것처럼 보이고, 따라서 포더의 비환원주의적 전략은 김재권의 환원주의적 전략에 대해 승리를 거둔 것처럼 보인다. 그런데 여기서 김재권은 포더에 대해 다음과 같이 역습을 편다.

자, 이제 우리는 포더 및 그와 같은 생각을 가진 철학자들에게 다음과 같은 질문을 하겠다. 만약 아픔이 N과 법칙적으로 동연적이고, N이 터무니 없이 선언적이며 명백히 비법칙적인 속성이라고 주장된다면, 왜 아픔 그 자체도 또한 하나의 종으로서 이질적이고 비법칙적이지 않은가? . . .만약 옥이 두가지 서로 다른 미시적 구조들에 의해 ‘실현’되기 때문에 비법칙적인 것으로 드러난다

면, 왜 똑같은 운명이 아픔에게는 닥치지 않는가?¹⁷⁾

결국 김재권은 포더와 같은 비환원주의자들을 다음과 같은 딜레마에 빠뜨리고 있다: 복수 실현 입론에 따라서 동일한 정신적 속성 M을 실현시키는 다양한 신경생리적 속성들 N_1, N_2, N_3, \dots 이 있다고 하자. M은 N_1, N_2, N_3, \dots 의 선언 N와 동일시된다. 이 때, (1) 만약 N_1, N_2, N_3, \dots 이 (인과적으로) 동질적이라면 N은 그 자체 적법한 속성이 될 수 있고, 이 경우 M은 N으로 환원된다. (2) 그렇지 않고 만약 N_1, N_2, N_3, \dots 이 (인과적으로) 이질적이라면 N은 진정한 속성이 될 수 없고, 따라서 M은 N으로 환원될 수 없다. 그러나 이 경우에는 N이 진정한 속성이 아니기 때문에 N과 동일시되는 M도 진정한 속성이 아니다! 그러므로 어떠한 경우라도 M은 비환원주의자들이 바라는 것처럼 환원되지 않으면서 동시에 진정한 속성으로 남을 수는 없는 것이다.

김재권 자신은 딜레마의 두 뿔 중 후자를 택한다. 즉 그는 M자체는 진정한 속성이 될 수 없으며 따라서 용어 “M”은 어떤 단일한 속성을 지시하는 것일 수 없다고 주장한다. “M”은 오직 어떤 언어적 공동체에 의해 공유되고 있는 단일한 개념과 연관될 수 있을 뿐이며, 그 개념은 여러가지 이질적인 속성들과 대응되고 있다는 것이다. 그렇다면 김재권은 제거주의자가 아닌가? 우리는 그의 주장으로부터, 정신적 속성들은 종이 아니며, 따라서 제거되어야 한다는 결론을 내려야 되지 않는가? 김재권은 자신의 주장이 제거주의적 색채를 띠고 있음을 인정한다. 그러나 그는 자신의 입장이 전형적인 제거주의자들의 입장과는 구분됨을 지적한다. 즉 제거주의자들과 달리, 김재권은 정신적 속성 M이 그 실현 속성 N_1, N_2, N_3, \dots 과 각각 선언적으로 동치라는 것을 인정하며, 따라서 정신적 속성이 자신의 신경생리적 실현 토대들에 의해 국소적으로 환원되는 것을 허용한다. 그러므로 정신적 속성은 영역-제한적(domain-specific)인 종들로 살아남을 수 있다. 가령 아픔 자체는 종이 될 수 없고 따라서 독립적인 종으로서

17) Kim (1992), 323쪽. 강조는 원문 그대로임.

의 위치를 상실하지만, 한편으로 물리적 실현 토대와 동일시된 영역-제한적 속성들인 포유류의 아픔, 파충류의 아픔, 연체동물의 아픔 등은 여전히 종으로서 간주된다는 것이다. 우리는 이러한 김재권의 입장을 아마도 “제거적 환원주의”(eliminative reductionism)라고 부를 수 있을 것이다.

이제 김재권이 복수 실현 입론에 대한 위와 같은 고찰을 특수 과학으로서의 심리학의 위치에 대해 어떻게 적용할 것인가 추측하는 것은 어렵지 않다. 지금까지의 논변이 보여주고 있는 결론은 오직 국소적으로 환원된 정신적 속성들만이 종으로서의 인정을 받을 수 있다는 것이었다. 그렇다면 오직 종으로서의 정신적 속성들만을 다루는 심리학도 또한 국소적으로 환원될 것이다. 김재권에 따르면, 일반적 믿음과 달리 심리학에 대한 복수 실현 입론의 함축은, 자율적 특수 과학으로서의 심리학은 거부되어야 하며 심리학 또한 자신의 복수 실현 속성들에 따라 복수로 실현되리라는 것이다.

4. 속성들의 단계와 인과적 힘들의 단계

위에서 우리는 복수 실현 입론에 대한 김재권의 재해석을 살펴보았다. 우리가 여기서 주목해야 할 점은, 위와 같은 김재권의 재해석이 (서론에서 말한 퍼트남의 논변과 마찬가지로) 정신적 속성의 독특한 성격들을 전혀 사용하지 않았으며, 오직 속성의 인과적 힘과 관련해서만 이루어졌다는 사실이다. 이는 김재권의 논변이 상위 수준에 있는 과학과 하위 수준에 있는 과학 일반에 적용될 수 있다는 것을 뜻한다. 그러므로 만약 김재권의 논변이 옳다면, 상위 수준에 있는 과학은 하위 수준에 있는 과학에 의해 국소적으로 환원됨으로써 자신의 자율성을 상실할 것이다. 그러나 문제는 그러한 국소적 환원의 과정이 원리적으로 계속 진행될 수 있다는 데에 있다. 즉, 심리학은 신경생리학에 의해 국소적으로 환원되고, 신경생리학은 물리화학에 의해 국소적으로 환원되고, 물리화학은 또 물리학에 의해 국소적으로 환원

되고, 등등. 이는 만약 김재권의 논변이 옳다면 가장 하위 수준에 있는 물리학을 제외한 모든 특수 과학들이 그 자율성을 상실한다는 것을 의미한다. 그러나 오늘날 각 특수 과학들이 자율적인 분과를 이루고 있는 현실을 볼 때, 이러한 귀결은 명백히 불합리한 것처럼 보인다.

그러나 그렇다면 김재권의 논변 중 무엇이 이러한 불합리한 귀결을 야기시켰는가? 김재권은 3절의 논의에서 핵심적인 역할을 하고 있는 두 전제를 다음과 같은 명시적인 원리로 제시하고 있다.

(P1) [인과적 계승의 원리] - 만약 정신적 속성 M이 시간 t에 어떤 체계에서 물리적 실현 토대 P에 의해 실현된다면, M의 이 사례의 인과적 힘은 P의 인과적 힘과 동일하다.

(P2) [종의 인과적 개별화 원리] - 과학의 종들은 인과적 힘을 토대로 해서 개별화된다. 즉, 대상들과 사건들은 그것들이 유사한 인과적 힘을 가졌을 경우에 한해서 어떤 종으로 분류되거나 어떤 속성을 공유한다.

(P1)과 (P2)를 모두 받아들인다면 물론 앞절에서 본 김재권의 논의는 타당하다: 복수 실현 이론에 따르면 어떤 정신적 속성 M은 인과적으로 서로 이질적인 신경생리적 실현 토대 N_1, N_2, N_3, \dots 에 의해 실현된다. (P1)에 의해, M의 사례들 각각의 인과적 힘은 N_1, N_2, N_3, \dots 중 어느 하나의 인과적 힘과 동일하다. N_1, N_2, N_3, \dots 이 인과적으로 서로 이질적이므로 M의 사례들 또한 인과적으로 서로 이질적이다. 그러므로 (P2)에 의해, M은 과학의 종이 될 수 없다.

그렇다면 (P1)과 (P2)는 받아들일만한 원리인가? 먼저 (P1)을 보자. 나는 (P1)이 대체적으로 옳은 원리라고 생각한다. (P1)을 거부한다는 것은, 우리가 하위 단계의 속성의 인과적 메커니즘으로 설명할 수 없는 상위 단계 속성 사례의 인과적 힘이 발생할 수 있다는 것을 인정하는 것이나 마찬가지이다 (2절에서의 물리적 실현 이론의 정식화를 상기해보라). 다른 말로 하면, 김재권이 표현하고 있는 것처럼 이는 상위 단계에서 어떤 마술적인 인과적 힘이 발생할 수 있다는 것

을 인정하는 것이다.¹⁸⁾ 어떠한 비환원주의자라도 물리적 실현 이론을 받아들이는 한 그러한 마술적 힘을 인정하지는 않을 것이고, 따라서 그는 (P1)을 받아들여야 할 것이다. 그러나 한편으로, 나는 (P1)이 “대체적으로” 옳다고 말했다. 그 이유는, (P1)에서는 M의 사례의 인과적 힘을 P의 인과적 힘과 동일시하고 있는 반면에 나는 M의 사례의 인과적 힘이 P의 사례의 인과적 힘과 동일시되어야 한다고 생각하기 때문이다. 물론 P의 사례의 인과적 힘이 P의 인과적 힘과 같다면 (P1)은 성립할 것이다. 그러나 P의 사례의 인과적 힘은 P의 인과적 힘과 같은 것이 아니라 단지 유사하다. 만약 그렇지 않다면, P의 인과적 힘과 유사하지만 같지는 않은 인과적 힘을 지닌 사례들은 P의 사례들이 될 수 없을 것이고, 그럴 경우 김재권은 (P2) 또한 거부해야 할 것이다. 왜냐하면 그는 분명히 (P2)에서, 대상들과 사건들이 유사한 인과적 힘을 가지는 한 그들은 어떤 속성을 공유한다고 말하고 있기 때문이다. 그러므로 (P1)에서 “P의 인과적 힘”이란 부분은 “P의 사례의 인과적 힘”으로 고쳐져야 한다고 본다. 그런 한 (P1)은 별 문제 없이 받아들일 수 있을 것 같다.

보다 근본적인 문제는 (P2)에 있다. (P2)에 따르면 과학에서의 종은 인과적 힘이 유사한가 그렇지 않은가에 따라서 개별화가 된다. 그러나 내 생각에는 여기서 더 이상의 기준이 없이 단지 “인과적 유사성”이라고만 말하는 것은 여전히 불충분하게 보인다. 우리는 이미 서론에서 오늘날 물리주의자들이 세계를 층으로 되어있는 구조로 여기고 있다는 것을 보았다. 이러한 중층구조하에서 속성들은 실제들과 마찬가지로 위계적으로 배열되어 있다. 그러나 여기서 속성들의 “위계”라는 것이 의미하는 것은 무엇인가? 그것은 각 단계의 속성들이 다른 단계들에서는 발견될 수 없는 규칙성들을 보이고 있다는 것이 외의 다른 것을 의미할 수 없다. 그렇지 않다면 우리는 단지 하나의 단계의 속성만을 얘기하지 다양한 단계의 속성들에 대해 말할 이유가 전혀 없기 때문이다. 그러나 속성들이 보이는 규칙성들은 속성들의 인과적 힘에 의해서만 발현될 수 있다. 그렇다면, 만약 각 단계의 속

18) Kim (1992), 326쪽.

성들이 서로 다른 규칙성을 보이고 있다면, 그 속성들의 인과적 힘들의 유사성은 그 속성들의 각 단계에서의 규칙성에 따라 판단되어야 하지 않는가?19) 따라서, 어떤 단계에서의 종들을 개별화하는 것은 그 단계에서의 인과적 힘의 유사성에 따라서 이루어져야 하지 않는가? 즉, (P2)는 다음과 같이 수정되어야 하지 않는가?

(P2*) 각 단계의 과학의 종들은 그 단계의 인과적 힘에 의해서 정의되어야 한다. 즉, 대상들과 사건들은 그것들이 자신의 단계에서 유사한 인과적 힘을 가졌을 경우에 한해서 어떤 종으로 분류되거나 어떤 속성을 공유한다.

(P2)를 (P2*)와 같이 수정하면, (P1)과 (P2*)에서 우리는 다음과 같은 결론을 도출할 수 있다. 복수 실현 이론에 의해서, M의 신경생리적 실현 토대들의 인과적 힘은 신경생리적 단계에서 볼 때 서로 이질적이고, 따라서 M의 사례들의 인과적 힘들 또한 신경생리적 단계에서 볼 때 서로 이질적이다. 그러나 이러한 사실로부터 M의 사례들의 인과적 힘이 심리적 단계에서 볼 때 이질적이라는 것이 따라나오지는 않는다. 따라서 김재권의 논변은 성립하지 않는다.

또한 (P2*)의 견지에서 보면 김재권이 든 옥의 예와 아픔의 예의 유비가 성립하지 않는다는 것을 분명히 알 수 있다. 먼저 옥의 예를 보자. 왜 옥은 종이 될 수 없는가? 김재권에 따르면 그 이유는, 옥이 경옥과 연옥의 두 이질적인 종으로 이루어져 있기 때문이다. 그러나 옥이 종이 될 수 없는 이유가 하나 더 있다. 그것은 옥이 (만약 그것이 종이려면) 자신의 선언 종과 같은 단계에 있는 종이기 때문이다. 즉 옥, 경옥, 연옥은 모두 광물의 종(mineral kind)이다. (광물학 이

19) 어떤이는 처음부터 각각의 단계의 서로 다른 규칙성들이 있다고 가정하는 것은 선결문제 요구의 오류를 범하는 것이라고 생각할지도 모르겠다. 만약 “서로 다른”의 의미가 “환원불가능한”의 의미를 함축한다면 위와 같은 가정은 정말로 선결문제 요구의 오류를 범하게 될 것이다. 그러나 어떤 두 종류의 규칙성이 서로 다르다는 것은 그 둘 사이에 환원 관계가 성립하느냐 하지 않느냐라는 문제와 관련해서 중립적이다.

외의 특수 과학으로서 경옥이나 연옥이 하지 못하는 보다 상위 단계의 인과적 역할을 옥이 하고 있는 과학이 있는가?) 그러나 아픔은 어떠한가? 아픔이 자신의 선언 속성과 같은 단계에 있는 속성인가? 물론 아니다. 아픔은 심리적 속성이고, 아픔과 동일하다고 여겨지는 선언 속성은 신경생리적 속성이다. 두 속성의 단계가 서로 다른 한, 이들이 가지고 있는 인과적 힘의 유사성은 그 단계에 맞는 기준으로 이루어져야 한다. 가령 아픔의 인과적 힘은 (기능주의자들의 주장에 따르면) 그 아픔을 야기한 환경적 자극 및 그 상태와 연관된 다른 내적 상태들, 그리고 그러한 내적 상태의 연관이 야기한 행동적 반응의 전체 연관에 의해서 규정된다. 반면 아픔을 실현시키는 신경생리적 토대의 인과적 힘은 중추 신경계에서의 뉴우런들의 활성화 및 이것이 야기한 근육의 움직임에 의해서 규정된다. 여기서 후자의 규정에 의해 유사한 것이 왜 전자의 규정에 의해서도 유사하여야만 하는가? 그러므로 김재권이 들고 있는 옥과 아픔의 유비는 성립하지 않는다. 전자는 자신의 선언 속성과 같은 단계에 있는 속성인 반면, 후자는 자신의 선언 속성과 다른 단계에 있는 속성인 것이다.

그러나 여기서 환원주의자는 다음과 같이 반박할 것이다: (P1)을 받아들이는 한, 우리는 상위 단계 속성 사례들의 인과적 힘들이 하위 단계 속성 사례들의 인과적 힘들과 동일하다는 것을 받아들여야 한다. 그렇다면 상위 단계 속성들의 인과적 힘이 하위 단계 속성들의 인과적 힘과 서로 다르다는 것이 어떻게 가능할 수 있는가? 어떻게 사례들의 인과적 힘은 동일하면서, 그 사례들을 자신의 사례들로 갖는 속성의 인과적 힘이 상위 단계와 하위 단계에서 각각 달라질 수 있던 말인가?

여기에 대해 우리는 다음과 같이 답변할 수 있다: 우선 이미 지적한 것처럼, 속성 사례들의 인과적 힘은 속성의 인과적 힘과 동일한 것이 아니라 대개는 단지 유사할 뿐이라는 점을 주목해보자. 어떤 속성의 인과적 힘은 본질적으로, 그 속성의 사례들의 인과적 속성들이 공유하는 보다 일반적인 인과적 힘이다. 그리고 상위 단계 속성의 인과적 힘은 자신의 사례들의 인과적 힘을, 하위 단계 속성이 자신의 사례들의 인과적 힘에서부터 일반화한 인과적 힘보다 더욱 추상적인

수준에서 일반화한 것이다. 상위 단계 속성 사례의 인과적 힘이 하위 단계 속성 사례의 인과적 힘과 동일하기 때문에, 상위 단계 속성의 인과적 힘이 하위 단계 속성으로부터 독립적이거나 새로운 인과적 힘이 아니라는 것은 사실이다. 그러나 상위 단계 속성의 인과적 힘은 그것의 보다 추상적인 성격 때문에, 하위 단계 속성의 인과적 힘과 동일하지도 환원되지도 않는 것이다.

상위 단계의 속성의 인과적 힘이 하위 단계의 인과적 힘으로 환원될 수 없다면, 상위 단계의 속성의 인과적 힘이 구현하는 법칙들은 하위 단계의 법칙들로 환원될 수 없고, 따라서 이들 법칙들은 자율적인 과학 이론을 이룬다. 그러므로 우리는 김재권의 주장과 달리, 복수 실현 입론이 정말로 특수 과학들의 자율성을 보장한다는 결론을 내릴 수 있다.

5. 비환원주의와 복수 현상론

이제 지금까지의 논의를 정리해보자. 우리는 김재권이 제기한 인과적 배제의 문제로부터 논의를 시작하였다. 나는 김재권의 논변에서 “물리적”이란 용어가 애매하게 사용되고 있다는 점을 지적하고, 이 용어가 적절히 이해된다면 배제의 문제는 비환원주의의 틀 내에서 해결될 수 있다고 주장하였다. 그러나 이를 위해서는 복수 실현 논변의 타당성이 전제되어야 하였고, 김재권이 복수 실현 논변의 타당성 또한 거부하고 있었기 때문에, 다시 복수 실현 논변을 옹호할 필요성이 제기되었다. 나는 복수 실현 입론에 대한 그의 재해석을 검토하고, 재해석 과정에서 그가 전제하고 있는 종의 인과적 개별화 원리를 비판하였다. 나는 개별화 원리에서 종을 나누는 기준이 되는 인과적 힘의 유사성이 속성 단계에 따라 상대적으로 정의되어야 한다고 논변하고, 이 점을 받아들이는 한 복수 실현 입론에 대한 김재권의 재해석은 성립하지 않으며 복수 실현 논변은 옹호될 수 있다고 논변하였다. 그러므로 지금까지의 나의 주장이 모두 옳다면, 우리는 비환원주의에

대한 김재권의 비판이 성공하지 못했다고 결론내릴 수 있을 것 같다.

그러나 아직 비환원주의자가 제기할 수 있는 (아마도) 마지막 반론이 남아 있다. 즉, 나는 4절에서, 정신적 속성 또는 일반적으로 말해서 상위 속성의 인과적 힘은 신경생리적 속성 또는 일반적으로 말해서 하위 속성의 인과적 힘과 다르고 환원가능하지 않지만, 그렇다고 해서 그것이 새로운 힘은 아니라는 것을 인정하였다. 그러나 어떤 속성이 진정한 속성이 되기 위해서는 그 속성이 가진 인과적 힘이 진정으로 새로운 것이어야 하지 않을까? 앞 절에서 말한 것처럼 상위 단계 속성의 인과적 힘과 하위 단계 속성의 인과적 힘이 결국 같은 인과적 힘을 추상화 수준에 따라 분류한 것에 지나지 않는다면, 어떻게 우리가 상위 단계의 속성을 진정한 속성으로 간주할 수 있단 말인가? 그렇다면 여기서 상위 단계의 “속성”이라고 불리는 것들은, 진정한 속성이 아니라, 단지 상위 단계에서 나타나는 어떤 “패턴”(pattern)에 불과한 것 아닌가? 결국, 여기서 옹호된 이른바 “비환원주의”는 부수현상론(epiphenomenalism)에 불과한 것 아닌가?

비록 내가 위에서 제기한 반박들에 대해서 일반적으로 동의하지는 않지만, 나는 이 반박들 중 마지막 것, 즉 여기서 옹호된 형태의 비환원주의가 일종의 부수현상론이라는 것에 대해서는 인정하지 않을 수 없다고 생각한다. 물론 이러한 답변은 많은 비환원주의자들에게 있어서 매우 실망스러운 답변일 것임에 틀림없다. 만약 비환원주의가 부수현상론을 받아들임으로써 “옹호된” 것이라면, 어떻게 그러한 옹호가 비환원주의에 대한 진정한 옹호가 될 수 있단 말인가? 비환원주의와 부수현상론이 서로 양립가능할 수 있는가?

나는 그럴 수 있다고 생각한다. 우선, 부수현상론이 도대체 무엇을 의미하는지에 대해서 명확히 해보자. 나는 다음과 같은 두가지 종류의 부수현상론이 있다고 본다.

(E1) [강한 부수현상론] 상위 단계의 속성들은 다른 속성들에 대해 어떠한 인과적 힘도 발휘할 수 없다.

(E2) [약한 부수현상론] 상위 단계의 속성들은 하위 단계의 인과적 힘으로부터 자신의 인과적 힘을 계승할 경우에만 다른 속성들에

대해 인과적 힘을 발휘할 수 있다.

여기서 내가 받아들이고 있는 부수현상론은 (E2)이지 (E1)이 아니다. 그러나 전통적인 부수현상론은 (E1)의 형태를 취해왔다. 그러므로 나의 입장은 전통적 의미에서의 부수현상론은 아니다. 그러나 비록 (E2)가 약한 형태의 것이긴 하지만 여전히 부수현상론의 일종이라는 것은 분명해 보인다. 그리고 나는 물리주의, 특히 물리적 실현 이론을 받아들일 경우, 결코 (E2)를 거부할 수는 없다고 생각한다.(E2)를 거부한다는 것은 하위 단계 속성 사례의 인과적 힘으로써 설명할 수 없는 상위 단계 속성 사례의 인과적 힘을 인정하겠다는 것과 마찬가지로, 이미 앞 절에서 밝혔듯이, 그러한 인과적 힘을 인정한다는 것은, 어떤 마술적인 인과적 힘이 발생할 수 있다는 것을 인정하는 것과 마찬가지로이다. 이는 어떠한 물리주의자라도 받아들이기 힘든 귀결일 것이며, 따라서 비환원주의자가 진정한 물리주의자라면 마땅히 거부해야 할 귀결이다. 그러므로 비환원주의자는 (E2)를 받아들여야 한다.

그렇다면 왜 나는 여전히 비환원주의를 옹호하는가? 이 질문에 대한 나의 대답은 다음과 같다: 위에서 언급된 이른바 환원불가능한 “패턴”들은, 비록 전적으로 새로운 속성은 아니라고 하더라도, 진정한 속성으로 간주될 자격이 충분하다. 왜 그러한지를 알아보자. 우선 이 “패턴”들은 단순히 우리의 개념들에 불과한 것이 아니라 우리에게 실제로 드러나고 있는 세계의 특징들이다. 또한 이 “패턴”들에는 물리적 속성들 이외에 우리가 일반적으로 속성이라고 생각하는 모든 속성들이 포함된다. 더구나 이 “패턴”들은 이 세계에서 반복적으로 발생하며, 자신들의 다양한 실현 토대에도 불구하고 항상적으로 보존되는 규칙성을 지닌다. 그러나 무엇보다도 가장 중요한 것은, 이 “패턴”들이 보편성, 반사실성, 투사가능성 조건을 만족하는 각 특수 과학들의 법칙들에서 나름대로의 역할을 하고 있으며 따라서 종들로 취급될 수 있다는 사실이다. 우리가 물리적 속성들의 실재성을 그렇게 확신하는 이유는, 이들 속성들의 존재에 대한 선험적 확실성이나 감각적 확실성 때문이 아니라 바로 이 속성들이 물리학의 법칙들을 구성하는 요

소들이기 때문이다. 그렇다면 왜 위에서 언급된 “패턴”들 또한 정확히 같은 이유로 해서 속성으로 취급될 수 없단 말인가? 나는 이른바 위에서 말하는 “패턴”들이 어떠한 면에서 보더라도 “속성”이라고 불리기에 충분한 자격을 갖추었다고 생각한다. 아마도 더욱 적절한 구분은 위에서 말한 속성과 패턴의 구분이 아니라 기본적 속성과 속성의 구분일 것이다.²⁰⁾

결국 지금까지의 논의의 결론은 다음과 같다. 심적 속성 또는 일반적으로 상위 단계의 속성은 신경생리적 속성 또는 일반적으로 하위 단계의 속성과 동일하지 않으며, 전자의 인과적 힘은 후자의 인과적 힘으로부터 계승된 것이긴 하지만 후자로 환원될 수는 없다. 그러므로 비환원주의는 옹호될 수 있다. 단 정신적 속성, 그리고 신경생리적 속성을 비롯한 모든 속성들이 물리적 속성의 (약한 의미에서의) 부수현상적 속성이라는 단서하에서 말이다. 어떤 비환원주의자들에게 있어서 이러한 결론은 충분히 만족스럽지 못할 것이며, 어떤 비환원주의자들에게 있어서 이것은 전혀 비환원주의가 아닐 것이다. 그러나 나는 이러한 결론이, 물리주의의 틀 내에서 이루어질 수 있는 비환원주의에 대한 최대한의 옹호라고 생각한다.

20) Dennett은 자신의 논문들에서 위에서 말한 이른바 “패턴”의 실재성에 대해 강력한 옹호 논변을 펼치고 있다 (특히 Dennett 1991 참조). “패턴”이라는 용어는 그의 논문에서 빌려온 것이다. 하지만 Dennett은 자신의 입장을 부수현상론으로 여기지 않는다는 점에서 나와 다르다. 그러나 나는 이러한 입장 차이가 본질적인 것이라기보다도, 부분적으로는 Dennett이 여기서 내가 제시한 두가지 종류의 부수현상론을 구분하지 못했다는 점 때문에, 또 부분적으로는 그가 (오늘날 “부수현상론”이라는 용어가 함축하고 있는 경멸적인 뉘앙스를 고려해서) 자신의 입장을 부수현상론으로 낙인찍고 싶지 않았기 때문에 비롯된 차이라고 생각한다.

참고 문헌

- 강진호 (1997), 「환원, 환원모형, 환원주의」, 서울대학교 철학과 석사 학위 논문.
- Block, Ned (1980) ed., *Readings in Philosophy of Psychology*, Vol.1, Cambridge: Harvard University Press.
- Davidson, Donald (1969), "The Individuation of Events," reprinted in Davidson (1982), *Essays on Action and Events*, Oxford: Oxford University Press.
- Dennett, Daniel (1991), "Real Patterns", *Journal of Philosophy* 88: 27-51.
- Fodor, Jerry (1974), "Special Sciences, or The Disunity of Sciences as a Working Hypothesis", reprinted in Block (1980).
- _____ (1989), "Making Mind Matter More," reprinted in Fodor (1990), *A Theory of Content and Other Essays*, Cambridge MA: MIT Press.
- _____ (1991), "You Can Fool Some of The People All the Time, Everything Else Being Equal: Hedged Laws and Psychological Explanations", *Mind* 100, Vol. 397: 19-34.
- Horgan, Terrence (1989), "Mental Quausation", *Philosophical Perspectives* 3: 349-76
- _____ (1993), "From Supervenience to Superdupervenience: Meeting the Demands of a Material World", *Mind*, Vol.102: 555-586.
- Kim, Jaegwon (1976), "Event as Property Exemplifications", reprinted in Kim (1993b).
- _____ (1989), "The Myth of Nonreductive Materialism",

- reprinted in Kim (1993b).
- _____ (1990), "Supervenience as a Philosophical Concept", reprinted in Kim (1993b).
- _____ (1992), "Multiple Realization and the Metaphysics of Reduction", reprinted in Kim (1993b).
- _____ (1993a), "The Nonreductivist's Troubles with Mental Causation", reprinted in Kim (1993b).
- _____ (1993b), *Supervenience and Mind*, Cambridge: Cambridge University Press.
- _____ (1994), "'Second-Order Properties' and Mental Causation", unpublished draft.
- LePore, Ernest and Loewer, Barry (1987), "Mind Matters", *Journal of Philosophy* 84: 630-641.
- Nagel, Ernest (1961), *The Structure of Science*, New York: Harcourt, Brace, and World.
- Putnam, Hilary (1967), "The Nature of Mental States", reprinted in Block (1980).
- Schaffner, Kenneth, "Philosophy of Medicine", in *Introduction to the Philosophy of Science*, Engelwood Cliffs, New Jersey: Prentice Hall.
- Schiffer, Stephen (1991), "Ceteris Paribus Laws", *Mind* 100, Vol. 397: 1-16.