

환원론과 인지과학

이 영 의 (고려대 강사)

과학의 통합은 우리의 이상이다.

그 이상에 가까이 다가갈수록 세계를 보다 더 잘 이해할 수 있다. 카우쉬(1977)

인지과학은 인지 현상을 연구하기 위해 여러 설명 수준에 속하는 분야들이 참여하는 학문이므로, 자연스럽게 구성 분야들의 상호 관계에 대한 질문이 제기될 수 있다. 환원론은 그러한 질문에 대해 환원 관계라는 명료한 설명을 제시한다. 이 글은 환원론(reductionism)이 제시하는 설명은 인지과학을 구성하는 분야들의 상호 관계를 보여주기에 부적절하다는 것을 주장하고 대안 모델의 가능성을 모색하고 있다. 최근 환원론자들은 네이글(Nagel)이 제안한 환원 모델을 수정하여 그것을 인지과학에 적용하고 있다. 환원론자들은 심리학이 신경과학으로 환원될 수 있다고 주장하지만, 그러한 주장은 분석 차원의 독립성을 간과한다는 점에서 적절한 설명이 될 수 없다. 인지과학을 구성하는 분야들의 상호 관계를 설명할 수 있는 대안 모델이 필요하다. 이 글의 후반부에서는 대안 모델이 갖추어야 할 조건들이 제시되고, 그 조건들을 충족시킬 수 있는 모델로서 차원-영역간 이론(interlevel and interfield theory)이 검토되고 있다.

1. 왜 환원 문제가 인지과학에서 중요한가?

과학철학자들은 전통적으로 과학 분야간의 관계를 규명하여 체계적인 과학상을 확립하려고 노력해 왔다. 다양한 과학의 분야들은 통합된 체계를 이루고 있는가 아니면 상호 독립적인가? 이러한 질문이 최근 인지과학 분야에서 다시 제기되고 있다. 인지과학은 일반적으로 인지 현상을 연구하기 위해 여러 분야가 참여하

는 학제적인 학문이라고 규정되지만, 환원론을 제외하고는 구성 분야들의 관계를 구체적으로 설명하는 이론은 없다.

논리 경험론자들은 환원론을 주장하면서 환원을 일차적으로 이론 사이의 관계라고 보았다. 그들의 주장에 따르면 상위 차원에 속하는 이론은 하위 차원에 속하는 이론으로 환원될 수 있으므로, 모든 경험과학은 궁극적으로 가장 기본적인 물리학에 환원된다. 이러한 환원 과정을 통하여 과학적 지식은 하나의 일관된 체계, 즉 통일과학(unitary science)을 형성하게 된다.

환원론은 제기된 당시부터 환원 가능성에 관한 문제로 논쟁의 대상이 되어 왔다. 최근 환원론은 “심리학은 신경과학에 환원될 수 있다”라는 주장을 통해서 인지과학 분야에 적용되고 있다. 환원 문제는 일차적으로 철학의 문제라고 볼 수 있으나 인지과학자들도 주목할 필요가 있다. 환원론이 주장하듯이 심리학이 신경과학에 환원될 수 있다면 인지과학은 마땅히 환원 지향적인 연구프로그램을 채택해야 할 것이다. 그리고 충분한 시간이 지나면 인지과학은 현재와 같은 학제적인 분야가 아니라 단일 분야로서 탈바꿈할 것이다. 그러나 환원이 원리적으로 불가능하다면 분야들의 상호 관계를 새로운 관점에서 설명하는 이론이 개발되어야 할 것이다. 환원론은 이러한 가능성을 검토하는데 있어서 논의의 출발점을 제공한다.

2. 환원론

환원론은 논의 맥락에 따라 다양한 의미로 사용되므로 논의를 전개하기 전에 먼저 방법론적 환원론과 존재론적 환원론을 구별할 필요가 있다.¹⁾ 예를 들어 심리학과 신경과학의 관계를 분석할 경우, 방법론적 환원론은 설명력의 확장을 위해서 심리적 현상이나 실재는 신경과학의 법칙에 의해 설명될 수 있다고 주장한다. 반면에 존재론적 환원론은 존재의 단순화를 위해 심리적 현상이나 실재들이 근본적으로 신경과학적인 현상이나 실재로 환원될 수 있다고 주장한다.

2.1 네이글의 환원 모델

환원론이 주장하는 바를 구체적으로 검토하기 위해 네이글의 모델을 예로 들

1) 방법론적 환원론은 Carnap(1938)과 Nagel(1961) 등에 의해 제안되었다. 존재론적 환원론은 대체로 Oppenheim and Putnam(1958)의 미시적 환원론(micro-reductionism)의 경우에서 볼 수 있듯이 방법론적 환원론과 결합된 형태로 나타난다.

어보자. 네이글의 모델은 다음과 같은 기본 전제를 갖고 있다. ① 환원은 이론간의 관계이다. ② 이론은 법칙들의 집합으로서 일종의 공리 체계이다.²⁾ ③ 이론간의 환원을 통하여 분야간의 환원이 이루어진다. 환원되는 이론을 T_1 , 환원하는 이론을 T_2 라고 하면, T_1 이 T_2 으로 환원된다는 것은 T_1 이 T_2 로부터 논리적으로 도출될 수 있다는 것을 의미한다. 네이글에 따르면 T_1 이 T_2 으로 환원되기 위해서는 다음과 같은 형식적 조건이 충족되어야 한다 (Nagel 1961, 353-54).³⁾

(1) 연결 가능성 조건(condition of connectability)

T_1 이 T_2 에서 나타나지 않는 비논리적 용어를 포함한 경우 먼저 그러한 T_1 의 용어는 대응 규칙(correspondence rule)에 의해 T_2 의 용어와 연결되어야 한다.

(2) 도출 가능성 조건(condition of derivability)

조건 (1)이 충족되었을 경우, T_1 이 논리적 절차에 의해서 T_2 로부터 도출될 때 환원이 발생한다.

네이글은 환원의 일차적인 목표를 이론의 설명이라고 주장함으로써 방법론적 환원론을 제안했다. 즉 T_1 이 T_2 으로 환원됨으로써 T_1 은 T_2 에 의해 설명된다. 네이글에 따르면 환원의 목표는 이론의 설명이며 환원의 구조는 연역 법칙적 설명(deductive-nomological explanation)의 구조와 동일하다. 예를 들어 갈릴레이 이론과 뉴턴 이론 사이의 환원 사례를 살펴보자. 갈릴레이 이론과 뉴턴 이론을 각각 T_1 , T_2 라고 하면, T_1 과 T_2 는 다음과 같은 집합으로 표현된다.

$$T_1 = \{g_1, g_2\}$$

$$T_2 = \{n_1, n_2, n_3, n_4\}$$

위에서 g_1 은 자유낙하 법칙 (자유 낙하체는 등가속도 운동을 한다)이고, g_2 는

-
- 2) 네이글의 입장은 구문론적 접근에 해당한다. 이 외에도 Suppes와 Sneed로 대표되는 집합 이론적 접근이 있으나 환원론에서 주로 논의되는 것은 구문론적 접근이다. 어떤 입장을 취하던 이 글의 전체적인 입장에 영향을 주지 않는다.
- 3) 네이글은 다음과 같은 비형식적 조건도 제시하고 있다. (1) T_2 는 애드혹크(ad hoc)가 아닐 것, (2) T_2 는 산출적일 것, (3) T_1 과 T_2 간의 환원은 의미있는 발전 단계에 있을 것 등이다 (1961, 358-66).

투사체 법칙(발사된 물체는 포물선 운동을 한다)이다. n_1, n_2, n_3 은 각각 뉴턴의 3가지 운동 법칙이고 n_4 는 만유인력 법칙이다. 또한 C 를 초기 조건의 집합, R 을 대응 규칙의 집합이라고 하자. C 는 예를 들어 “공기 저항은 무시할 수 있을 정도로 매우 작다”는 조건과 “물체의 높이(h)가 지구반경(r)에 비해 무시할 수 있을 정도로 매우 작다”는 조건 등을 포함한다. 이제 C 와 R 이 주어지면 다음과 같은 도출이 발생한다.

$$T_2, C, R \vdash T_1 \dots\dots (i)$$

네이글은 (i)을 전형적인 환원 사례라고 간주하지만, 실제로 (i)는 성립되지 않는다. T_2 로부터 T_1 을 도출할 수 없는 이유를 네이글이 제시한 형식적 조건과 관련하여 살펴보자.

(1) 대응 규칙이 표현하는 관계에 대한 해석상의 문제가 있다. 네이글(1961, 355)은 대응 규칙에 사용되는 결합사가 반드시 양조건언일 필요는 없다고 강조했지만, 일반적으로 대응 규칙이 기술하는 것은 동일성(identity) 관계라고 인정되고 있다. 여기에서 문제는 동일성을 어떻게 보느냐에 따라서 대응 규칙을 구성할 수 있는가의 여부가 결정된다는 점이다. 예를 들어 환원론자인 카우쉬(Causey 1977)에 따르면 동일성은 경험적으로 확증되어야 한다는 의미에서 종합적 동일성(synthetic identity)이다. 예를 들어 T_1 의 술어인 'x is gene'이 지시하는 대상과 T_2 의 술어인 'x is DNA'가 지시하는 대상이 동일할 경우 종합적 동일성이 성립된다.

반면 포더(Fodor)를 비롯한 비환원론자들은 동일성을 자연종(natural kind) 간의 동일성이라고 해석한다. 이러한 해석에 따르면 심리학의 자연종과 신경과학의 자연종이 동일할 경우에만 심리학은 신경과학으로 환원될 수 있다. 그러나 포더 등은 복수 실현(multiple realization)의 가능성 때문에 그러한 동일성은 확보될 수 없다고 지적한다. 심리학의 자연종 술어는 특정 유형의 심리적 상태를 지시한다. 그런데 기능적으로 동일한 심리적 상태는 다양한 물리 체계에서 실현될 수 있기 때문에, 심리학의 자연종 술어를 신경과학의 자연종 술어와 연결할 수 없다.⁴⁾

4) 포더는 무엇을 심리학의 자연종으로 간주할 것인가에 대한 분명한 기준을 제시하지 않고 있다. 이와 관련된 논의는 Block and Fodor(1972, 173), Fodor (1975, 1-26), Causey(1977, 80-105; 142-51)를 참조할 것.

이는 곧 대응 규칙을 구성할 수 없다는 것을 의미한다.

(2) 도출 가능성 조건은 논리적인 측면에서 보면 충족될 수 없다. 시간상 선행 이론에 해당하는 T_1 은 후속 이론인 T_2 의 입장에서 보면 불완전한 이론이고 경우에 따라서는 잘못된 이론일 수도 있으므로 T_2 로부터 T_1 을 도출할 수 없게 된다 (Popper 1972, 197-203). 예를 들어 위의 환원 사례에서 h/r 의 값이 무시할 수 있을 정도로 매우 작다고 인정하더라도 T_2 에서는 자유 낙하체의 가속도가 증가하므로 g_1 은 명백히 잘못된 법칙이다. 또한 모든 투사체는 실제로 포물선 운동이 아니라 타원 운동을 하므로 g_2 역시 잘못된 법칙이다. r 의 값이 무한하다고 틀린 가정을 하지 않는 한, T_2 로부터 포물선 운동이 도출될 수 없다. T_2 로부터 도출된 것은 T_1 이 아니라 T_2 에 근사적으로 접근하는 수정된 이론(T_1^*)이라고 보아야 한다. 이러한 사실로부터 다음의 (i')이 성립된다.

$$T_2, C, R \vdash T_1^* \dots\dots (i')$$

(3) T_1 과 T_2 는 서로 다른 개념 틀을 갖기 때문에 연결 가능성 조건은 충족될 수 없다.(Feyerabend 1962, 74-91). 네이글은 T_1 과 T_2 의 비논리적 용어의 의미가 그 용어를 포함하는 명제와는 독립적으로 고정되어 있다고 보았다. 그러나 용어의 외연은 이론이 바뀌에 따라서 변하는 경우가 있다. 예를 들어 T_1 에서 관성 운동은 원운동이지만 T_2 에서의 관성 운동은 포물선 운동이다. 더구나 T_1 에서 낙하 운동은 본래의 장소를 향한 운동이지만 T_2 에서 낙하 운동은 인력의 영향을 받은 운동이다. 이와 같이 두 이론의 용어들이 상호 불가공약성 (incommensurability)을 갖는 경우 대응규칙을 구성할 수 없으며, 그 결과 의도된 환원은 발생하지 않는다.

2.2 수정된 환원모델

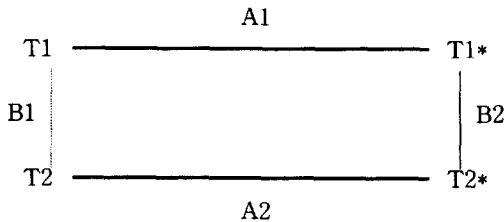
지금까지 네이글이 제안한 환원 모델의 구조와 문제점을 검토해 보았다. 환원론자들은 네이글의 모델을 수정하는 과정에서 위에서 지적된 두번째와 세번째 문제점을 해결하는 전략을 택했다. 그 결과는 처칠랜드(Churchland 1986, 282-83)의 환원 모델에서 발견된다.⁵⁾ 처칠랜드의 모델에서 환원 과정은 다음과

5) 본 문에서 인용된 처칠랜드 부부의 글은 발표된 년도와 관계없이 폴 처칠랜드의 경우는 P.M.Churchland(1979, 1992), 패트리샤 처칠랜드의 경우는 P.S. Churchland(1986)

같이 두 단계로 진행된다. (1) T_2 에서 T_1 의 법칙들로 이루어진 유사 이론(analogue) T_1^* 를 구성하고, (2) T_1^* 를 T_2 로부터 도출한다. 처칠랜드는 자신의 모델을 심리학과 신경과학의 관계에 적용하면서 모델을 완회시킬 필요성을 느꼈지만 완전한 설명을 제시하지 않았다.

최근 사프너(Schaffner 1992, 320-21)는 처칠랜드의 모델을 보완하는 매우 일반적인 모델, GRR 모델(general reduction-replacement model)을 제안했다. GRR 모델은 다음과 같은 점에서 처칠랜드의 모델을 보완한다. (1) 환원, 부분적 환원, 대치를 허용한다.⁶⁾ (2) T_1 뿐만 아니라 T_2 도 수정된다. (3) T_1 이 T_1^* 으로 수정 불가능할 경우, 그것은 T_2 혹은 T_2^* 에 의해 대체된다.

GRR 모델에 따르면 이론간의 관계는 다음 도표와 같이 표현된다.



(도표1) GRR 모델에서의 이론 간의 관계

A1과 A2는 수정 또는 대체 관계이고

B1은 현재의 환원관계, B2는 미래의 환원관계를 의미한다.

GRR 모델은 위의 도표에서 볼 수 있듯이 차원 개념과 시간 개념을 동시에 반영하고 있다. T_1 과 T_1^* , T_2 와 T_2^* 는 동일한 차원에 속하는 이론이고, T_1 과 T_2 ,

에서 인용하였음.

- 6) 사프너의 구분은 처칠랜드(1986, 284)가 이론간 동일성(CTI)을 기준으로 구분한 다음의 관계들을 이용한 것이다. 처칠랜드에 따르면 환원 관계는 CTI를 통한 부드러운 환원(smooth reduction)과 CTI가 문제가 되어 구 이론의 개념 수정이 필요한 시끄러운 환원(bumpy reduction), 그리고 CTI가 전혀 불가능한 철저한 제거(outright elimination)가 있다.

T_1^* 와 T_2^* 는 동일한 시간에 속하는 이론이다. 전자는 다른 시점에서의 동일한 분석 차원에 속하는 관계를 형성하고, 후자는 동일한 시점에서의 다른 차원에 속하는 관계를 형성한다. 이를 각각 차원내 맥락(intralevel context)과 차원간 맥락(interlevel context)이라고 하자 (McCauley 1986, 193).

GRR 모델을 네이글의 모델과 비교해 보면 두 모델과 관련된 과학관이 다르다는 것을 알 수 있다. 네이글의 모델을 지지했던 논리 경험주의자들은 과학이 대체로 연속적인 과정을 통하여 진보한다고 보았다. 반면에 GRR 모델을 주장한 샤프너 등은 과학이 근본적으로 불연속적인 과정을 통해서 진보한다고 주장한다. 이러한 과학관의 차이는 차원내 맥락에서의 A1과 A2를 완전한 환원 관계로 보느냐 아니면 수정 또는 대치 관계로 보느냐를 결정한다.⁷⁾

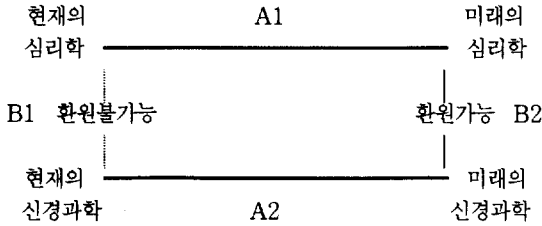
네이글은 환원을 이론 간의 관계라고 규정하면서 환원 관계는 차원내 맥락과 차원간 맥락에 동시에 적용될 수 있다고 생각했다. 그러나 위에서 보았듯이 완전한 환원은 원리상 불가능하므로 GRR 모델은 차원내 맥락에서는 수정 관계와 대치 관계만을 인정한다. 그러나 네이글의 모델과 GRR 모델은 차원간 맥락에서 환원 관계만을 인정한다는 점에서 공통점을 갖는다. 차원간 맥락에서는 대치가 발생하지 않는다! 차원내 맥락에서는 이론간의 번역 가능성이 희박할수록 GRR 모델이 주장하듯이 대치 가능성이 증가할 수도 있지만 차원간 맥락에서는 이론간의 번역 가능성이 희박할수록 대치 가능성은 그만큼 줄어든다.

3. 인지과학에서의 환원론

3.1 GRR 모델의 적용 가능성

앞 절에서 살펴본 GRR 모델을 심리학과 신경과학의 관계를 설명하는데 적용해 보기로 하자. 환원론자들도 인정하듯이 현재의 심리학이 현재의 신경과학으로 환원될 가능성은 없으므로 GRR 모델에 따르면 환원은 다음 도표와 같이 발생할 것이다.

7) 샤프너와 같이 Kuhn의 입장을 지지하는 사람들에 따르면 차원내 맥락에서는 이론간의 논리적 불일치가 발생하며 이는 궁극적으로는 대치 및 제거로 전개된다. 그러나 차원내 맥락에서 대치 관계가 성립될 수 있는가는 논의의 대상이다. 이 글에서는 일단 GRR 모델의 주장을 인정하고 논의를 진행하겠다.



(도표2) GRR 모델을 적용한 심리학과 신경과학의 관계

위의 표에서 알 수 있듯이 현재의 심리학이 미래의 신경과학으로 환원되는 과정은 두 단계로 진행된다. 즉 먼저 현재의 심리학이 미래의 심리학에 의해 수정 또는 대체(A1)되고, 그 다음 미래의 심리학이 미래의 신경과학으로 환원된다(B2).

그러나 이러한 환원 과정은 도중에 중단될 수도 있으므로 환원이라는 목표에 도달하기 위해서는 두 분야간의 공동 노력과 협조가 필요하다. 이러한 의미에서 처칠랜드는 환원이 수행하는 역할을 상호 진화적(co-evolutionary) 관점에서 바라본다. 상호 진화론에 따르면 환원은 심리학과 신경과학이 공동으로 추구해야 될 목표이므로 환원은 이미 개발된 이론에 소급적으로 적용되어 두 이론을 통합하는 소극적 역할을 하는 것이 아니라, 상호 협력을 통한 발전을 촉진하는 적극적 역할을 수행한다. 상호 진화적 발전을 통해 두 분야가 성숙 과학의 단계로 진입하면 환원은 즉시 달성된다.

상호 진화라는 개념은 두 분야 간의 상호 정합성과 체계성을 확립한다는 점에서 매우 설득력이 있지만, 진화의 목표가 환원이라는 주장은 검토되어야 한다. 심리학이 신경과학으로 환원될 수 있다는 환원론자들의 주장은 차원간 맥락에서 환원(B2)이 가능하다는 것을 전제로 한다. 하지만 그러한 전제는 다음과 같은 점에서 성립될 수 없다.

첫째 자연 현상을 설명하는데 있어서 구조적 설명(structural explanation) 뿐만 아니라 기능적 설명(functional explanation)도 있다. 환원론은 자연 조직에서 볼 수 있는 부분-전체 관계에 초점을 두고 상위 차원의 이론은 하위 차원의 이론으로 환원된다고 주장한다. 이러한 점에서 환원론은 구조적 설명을 추구한다.

그런데 여기서 주의해야 할 점은 분석 차원이 높아짐에 따라서 체계의 복잡성도 증가한다는 점이다. 특히 인지 체계와 같은 매우 복잡한 체계에서는 특정 차원

에 속하는 다른 부분들이 보다 상위 차원에서 기능적으로 동일한 사건을 야기하는 이른바 복수 실현 현상이 발생한다. 또한 특정 차원의 동일한 부분이 보다 상위 차원에서 한가지 이상의 기능을 발휘할 수도 있다. 이러한 체계에서는 구조적 설명 보다는 기능적 설명이 더 중요한 역할을 담당한다.

둘째 분석 차원 마다 서로 다른 설명 대상을 갖기 때문에 다른 차원에서의 설명들은 상호 독립적이다. 예를 들어 토끼 집단과 여우 집단으로 구성된 생태계에서 토끼 한마리가 여우에게 물려 죽었다고 해보자. “왜 그 토끼가 죽었는가?”라는 질문에 대해 생물학자들은 “그 여우가 그 토끼의 목동맥을 물었기 때문이다”라고 대답할 것이다. 반면에 생태학자들은 “여우의 숫자가 매우 많았기 때문이다”라고 대답할 것이다. 이러한 예는 동일한 질문에 대해서 하위 차원의 설명과 상위 차원의 설명은 독자적인 설명 대상을 추구한다는 점을 보여준다.

셋째 상위 차원의 설명은 단순히 현재의 과학 수준이 낮기 때문에 요구되는 것은 아니다. 생물학의 예를 살펴보자. 일부 환원론자들은 고전 유전학이 분자 생물학으로 환원될 수 있다고 주장한다.⁸⁾ 그러나 키처(Kitcher 1984)에 따르면 다음과 같은 이유 때문에 고전 유전학은 분자 생물학으로 환원될 수 없다. (1) 고전 유전학은 유전자의 전이에 대한 법칙을 갖고 있지 않다. (2) ‘gene = DNA’라는 동일성 관계가 성립되지 않으므로 대응 규칙을 구성할 수 없다. (3) 설사 도출이 가능하더라도 설명이 안되는 비이행적 관계가 발생한다. 키처의 분석은 환원론이 생물 과학의 분야에 적용될 수 없음을 보여줄 뿐만 아니라, 심지어는 하나의 분야에도 세부 차원들은 상호간에 환원 불가능하다는 것을 보여준다.

3.2 제거적 유물론

심리학과 신경과학의 관계를 설명하는데 있어서 GRR 모델이 적용될 수 없는 이유를 살펴보았다. 그러나 환원론자들의 주장은 이것으로 그치지 않고, 특히 처칠랜드는 심리학이 신경과학으로 환원 불가능하다는 점을 들어 심리학을 제거해야 한다고 주장한다. GRR 모델이 차원간의 구분을 간과하는 잘못을 범하고 있는 것과는 대조적으로 처칠랜드의 주장은 환원과 제거를 혼동하는 오류를 범하고 있다. 처칠랜드가 주장하는 제거적 유물론(eliminative materialism)을 요약해 보고

8) 최근 간행된 분자 생물학의 교과서에는 다음과 같은 구절들이 당연한 것처럼 제시되고 있다. “환원론적 접근과 분자 생물학의 새로운 기법은 조만간에 모든 실험 생물학을 통합할 것이다.”(Darnell et. 1990, vii)

문제점을 검토해 보자.

신념이나 욕구와 같은 명제 태도에 의해 인간의 행위를 설명하는 이론을 통속 심리학(folk psychology :FP)이라고 하자. 처칠랜드(1989, 6-9)에 따르면 FP는 다음과 같은 결함을 갖고 있다. (1) FP는 매우 제한된 설명력 만을 갖는다. (2) FP는 매우 오랫동안 정체적이었고 거의 변화가 없었다. (3) FP는 아직도 과학의 체계(신경과학)에 융합될 가능성이 거의 없다.⁹⁾ 한마디로 FP는 새로운 현상을 발견할 수 없다는 의미에서 퇴행적인 연구 프로그램에 해당기 때문에 유일한 대안은 과학의 영역으로부터 FP를 제거하는 것이다. 제거적 유물론이 주장하는 바는 다음과 같이 요약된다 (Churchland 1986, 396).

- (1) 심리학적 용어를 이용하여 심성 현상을 설명하는 FP도 이론이다.
- (2) FP는 이론의 불충분성 때문에 결국 수정되든가 아니면 완전히 대체되어야 한다 (그러므로 제거적).
- (3) FP를 궁극적으로 대체하는 것은 발달된 신경과학의 개념틀이다 (그러므로 유물론).

제거적 유물론에서 문제가 있다고 생각되는 부분은 (3)이다. 처칠랜드가 주장하듯이 (1)과 (2)가 참이라고 인정하더라도, GRR 모델에 따르면, (3)을 주장할 수는 없다. 환원론자로서의 처칠랜드가 GRR 모델과 모순되지 않는 방식으로 자신의 입장을 제시하려면 (3)을 다음과 같이 수정해야 한다.

- (3') FP를 궁극적으로 대체하는 것은 발달된 심리학의 개념틀이다.

이러한 수정은 처칠랜드가 제거적 유물론을 주장하면서 환원과 제거를 동일시하는 오류를 범했다는 것을 스스로 인정하는 것이다. 여기서 지적되는 오류는 2절

9) 처칠랜드는 이를 뒷바침하는 논거로서 계산주의적 전통에서 개발된 이론은 신경과학의 이론으로 환원 불가능하다는 점을 들고 있다. 그러나 FP와 계산주의적 인지 이론을 동일하게 취급하는 것은 잘못이다. 이 문제에 대한 논의는 Bechtel and Abrahamsen(1991) 8장을 참조할 것. 또한 처칠랜드는 제거의 일차적인 대상으로서 FP 만을 주목하고 있으나 현재의 심리학 이론도 제거 대상에 포함된다. 그에 따르면 모든 심리학 이론은 비과학적 개념틀인 심성론과 지향성을 전제하고 있다(1992, 16-7). 제거적 유물론은 환원 불가능한 모든 이론이 제거되어야 한다고 주장하기 때문에 당연히 심리학 이론도 제거의 대상에 포함된다.

에서 논의된 차원의 구분 문제와 밀접하게 관련된다. 처칠랜드는 제거적 유물론을 뒷받침하는 많은 사례들을 제시하고 있는데, 그러한 제거 사례들은 모두 차원내 맥락에 속한다. 예를 들어 열소 이론이 동력학 이론에 의해 제거된 사례는 물리학이라는 동일한 차원에서 발생했다. 차원내 맥락에서는 처칠랜드가 주장하듯이 선행 이론은 후속 이론에 의해 제거될 수도 있지만, 차원간 맥락에 속하는 심리학과 신경과학의 관계에서는 제거가 발생하지 않는다. 현재의 심리학이 제거된다면 (3') 이 말하는 것처럼, 현재의 심리학은 동일한 차원에 속하는 미래의 심리학에 의해 제거될 것이다.

제거와 환원을 혼동하고 동일시하는 잘못은 이론간의 관계에 대한 주장에서도 나타난다. 처칠랜드는 2절에서 보았듯이 네이글의 환원 모델을 수정하면서 환원 관계 이외에도 다른 관계의 가능성을 인정했다. 이를 보다 구체적으로 표현하면 이론간의 관계는 순수한 환원 - 부분적 환원 - 부분적 제거 - 순수한 제거에 이르는 연속선 상의 어느 한 지점을 차지하며, 경험적 탐구 만이 그 좌표를 결정할 수 있다(1988, 84). 이론간의 관계를 정도의 문제로 보는 이러한 주장은 차원내 맥락에 속하는 환원 관계와 제거 관계가 차원간 맥락에서도 성립한다고 보는 잘못을 범하고 있다.

4. 대안 모델의 가능성

환원론은 심리학과 신경과학의 관계에 대한 분명한 설명을 제시한다. 그러나 3절에서 살펴보았듯이 그러한 설명은 일차적으로 차원간의 구분을 무시한다는 점에서 적절한 설명이라고 볼 수 없다. 차원간의 맥락을 무시함으로써 다른 차원에 속하는 이론들이 환원 관계를 이룬다고 잘못 생각하거나(GRR모델), 환원과 제거를 동일시하는 오류를 범했다(제거적 유물론). 인지과학을 구성하는 분야들의 관계를 적절히 설명하기 위해서는 각 차원에 속하는 분야들의 자율성과 연결성을 동시에 고려해야 한다. 자율성과 연결성이라는 서로 양립 불가능한 것처럼 보이는 두가지 사항을 조화시키기 위해서 대안 모델이 갖추어야 될 조건을 검토해 보자.

4.1 대안 모델에 대한 조건

환원론은 이론을 통한 환원을 주장하면서 T_1 이 T_2 으로 환원되면 대응되는 분야도 환원된다고 가정한다. 물론 이론에 의한 환원이 여의치 못할 경우에 분야 간

의 환원도 시도되었지만(Kemeny and Oppenheim 1956) 그럼에도 불구하고 환원론은 일차적으로 분야 보다는 이론간의 관계에 초점을 둔다. 이론을 통한 분야 간의 환원은 앞에서 보았듯이 차원간의 구분을 무시하므로 성립되기가 어려울 뿐만 아니라 방법론적인 측면에서도 환원 조건은 지나치게 엄격하기 때문에 거의 실현 불가능하다. 이러한 문제는 일차적으로 이론에 우선성을 두기 때문에 발생한다. 방법론적으로 보다 거시적인 관점에서 접근할 필요가 있다. 대안 모델은 일차적인 분석 대상으로서 분야간의 관계를 다루어야 한다.

환원론은 이론간의 관계에서 환원, 수정 또는 대치 관계만을 인정한다. 그러나 모든 이론이 이러한 관계를 갖는 것은 아니며, 더구나 차원이 다른 이론들은 대치 관계를 갖지 않는다. 언어 심리학(psycholinguistics)을 매개로 언어학과 심리학 분야의 이론들이 비환원적으로 연결되는 사례들이 있다. 대안 모델은 특히 GRR 모델을 비판할 때 사용된 논거들을 충분히 반영하면서 새로운 관계들을 인정해야 할 것이다. 그 중에서도 특히 기능적 설명이 추구하는 대상, 즉 기능적 관계에 주목해야 한다.

과학의 분야들은 체계적인 방식으로 상호 관련성을 유지해야 한다. 물론 과학의 통합이 추구할 만한 가치가 있는가에 대해 이의를 제기할 수도 있다. 그러나 각 분야들이 상호 연결됨으로써 과학의 정합성과 체계성은 증가하기 때문에 이러한 규범적 조건을 부여해야 한다. 환원론은 이러한 조건에 대해 환원을 통한 과학의 통합을 주장했다. 환원론자들은 과학의 통합이라는 이상을 과학자들의 연구 방향을 안내하는 작업가설(working hypothesis)로 채택할 것을 요구했다(Oppenheim and Putnam 1958, 8). 이렇게 환원과 과학의 통합을 연계시키면 환원이 불가능할 경우 과학의 통합은 포기되어야 할 것이다. 그러나 과학의 통합이 반드시 환원을 통해서만 달성된다고 볼 수는 없다. 대안 모델은 과학의 통합이라는 규범적인 요구 사항을 충족시킬 수 있는 방법을 제시해야 한다.

지금까지 논의를 정리해 보면 대안 모델은 다음과 같은 조건을 충족시켜야 할 것이다.

(1) 분석 대상 조건

대안 모델은 일차적으로 분야에 초점을 두어야 한다.

(2) 다양성 조건

대안 모델은 비환원 관계들도 인정해야 한다.

(3) 규범적 조건

대안 모델은 과학의 통합을 달성할 수 있는 방법을 제시해야 한다.

4.2 차원-영역간 이론

이제 차원간 이론(interlevel theory)와 영역간 이론(interfield theory)을 통하여 위에서 제시된 조건들을 충족시킬 수 있는 모델을 구성해 보자.

차원간 이론은 심리학자인 슈타트(Staat)에 의해 제안되었다. 그에 따르면 현재의 심리학은 체계성이 결여된 다양성만을 갖고 있다. 이러한 상황을 극복하고 심리학을 통합하는 방안으로서 슈타트는 다음과 같은 방법을 제안했다. 즉 일차적으로 특정 문제의 해결에 초점을 맞추고, 그 문제와 관련되는 영역들을 연결하는 교량 이론(bridging theory)을 구성함으로써 문제를 해결한다.

영역간 이론은 과학철학자인 다덴과 마울(Darden and Maull)에 의해 제안되었다. 영역간 이론은 추구하는 목적과 방법론이 차원간 이론과 거의 동일하다. 차이가 있다면 후자는 전자에 비해서 차원 개념을 강조한다는 점이다. 필자는 두 이론을 결합하면 분야간의 관계를 효과적으로 분석할 수 있다고 생각한다. 차원간 이론과 영역간 이론을 결합한 이론을 차원-영역간 이론(interlevel and interfield theory :IIT)이라고 하자.

IIT는 일차적으로 영역을 연결하고자 한다. 영역 개념을 먼저 살펴보자. 영역은 지금까지의 용어법에 따르면 이론이나 분야가 아니라 분야의 세부 분야에 해당한다. 예를 들어 유전학은 고전 유전이론을 갖고 있는 영역이며, 마찬가지로 과학철학, 인지 심리학, 양자역학, 세포학 등은 영역으로 분류된다.

영역은 다음과 같은 요소로 구성된다 (Darden and Maull 1977, 43-4).

- ① 중심 문제
- ② 중심 문제와 관련된 사실들로 구성된 영역
- ③ 문제 해결에 실마리를 제공하는 설명 요인과 목표, 테크닉과 방법
- ④ 중심 문제와 관련되고 설명 목표를 실현시키는 개념과 법칙, 이론
- ⑤ 해당 영역의 특징을 나타내는 요소와 관련된 어휘들

영역은 교량 이론에 의해 결합된다.¹⁰⁾ 교량 이론의 구체적인 예를 살펴보자. 20

세기 초반 유전학자들은 멘델의 인자를 유전 단위로서 인정했지만 유전 인자의 위치를 확인할 수 없었다. 한편 세포학자들은 염색체를 발견하고 그것이 기능적으로 유전과 관련된다고 추측했지만, 염색체가 어떻게 그러한 역할을 하는가를 설명할 수 없었다. 이러한 상황에서 1903년 보베리(Boveri)와 수튼(Sutton)은 멘델의 인자들이 염색체내에 있다는 가설을 제시했다. 과학자들은 보베리 등이 제안한 가설을 채택함으로써 그 이전에는 불가능했던 예측 뿐만 아니라 유전 인자와 염색체에 대한 새로운 연구를 할 수 있게 되었다. 이러한 사례에서 염색체 이론은 유전학 영역과 세포학 영역을 연결하는 교량 이론에 해당한다.

교량 이론은 다음과 같은 상황에서 생성된다(Darden and Maull 1977, 49-51).

- (1) 배경 지식에 의해 영역 간의 관계가 이미 형성되었다고 알려진 경우
- (2) 두 영역이 동일한 현상이 갖고 있는 다른 측면을 설명하는데 관심을 갖는 경우
- (3) 한 분야에서 제기된 문제가 그 분야만의 개념에 의해 해결 불가능한 경우¹⁰⁾

이제 IIT가 앞에서 제시된 조건들을 어느 정도 충족시키기를 검토해 보자.

첫째 IIT는 문제 중심적인 접근 방식을 채택한다. 환원론의 대응 규칙, 일명 연결 법칙(bridge law)과 IIT의 교량 이론은 공통적으로 연결 기능을 수행한다. 그러나 교량 법칙은 이론을 연결하는데 비해 교량 이론은 영역을 연결한다. 또한 교량 법칙이 주로 정당화의 맥락에서 철학자들에 의해 제안되는 반면에, 교량 이론은 발견의 맥락에서 과학자 자신에 의해 제안된다. IIT는 연구의 산물인 이론이 아니라 과학자들이 해결하고자 노력하는 문제에 초점을 맞추므로써, 문제 해결과 관련된 영역간의 관계를 분석한다.

IIT는 영역을 일차적인 분석 대상으로 삼음으로써 이론간의 환원이라는 전략

- 10) 다덴과 마울은 영역내 이론(intrafield theory)과 영역간 이론(interfield theory)을 구분한다. 영역내 이론은 특정 영역에 속하는 이론이다. 영역간 이론은 차원간 이론에서 말하는 교량 이론에 해당한다.
- 11) 교량 이론이 생성되는 환경은 인지과학의 환경을 매우 적절하게 표현하고 있다. 그것은 특히 인지과학이 성립하게 된 기본 동기 뿐만 아니라 인지과학의 학제적 특징, 현재의 상황과 문제점을 적절히 묘사하고 있다.

이 갖는 문제점을 피할 수 있다. 그러나 영역의 연결이 곧 분야 간의 관계를 설명한다고는 볼 수 없다. 앞에서 필자는 차원간의 구분을 들어서 환원론이 적절한 모델이 될 수 없다고 주장했었다. 동일한 기준을 IIT에 적용하면 다음과 같은 질문이 발생한다. IIT는 차원내 맥락과 차원간 맥락에서 동시에 적용될 수 있는가? 이에 대해 슈타트와 다덴 등은 긍정적으로 보고있지만, 정확한 설명을 제시하지 못하고 있다.

그러나 최근 IIT에 유리한 사례들이 개발되고 있다. 예를 들어 아브라함센(Abrahamsen 1987)은 언어학과 심리학의 관계를 분석하는데 IIT를 이용하고 있고, 또한 벡텔(Bechtel 1994)는 신경과학과 인지과학의 관계, 인지과학의 제 분야간의 관계를 설명하는데 IIT를 이용할 수 있다고 제안한다. 그 외에도 아브라함센과 벡텔(1991)은 연결주의(connectionism)가 인지과학의 제 분야와 맺고 있는 관련성을 설명하기 위해서 IIT를 이용하고 있다. 이러한 사례들은 IIT가 인지과학 분야에 적용될 수 있는 가능성을 갖고 있다는 것을 보여준다.

둘째 IIT는 영역간의 관계가 다양하다는 것을 인정한다. 예를 들어 다음과 같은 관계들이 있다(Darden and Maull 1977, 49). ① 부분-전체 관계 : 특정 영역은 다른 영역에서 가정된 실재나 과정의 물리적 위치에 대한 설명을 제공한다. ② 구조-기능 관계 : 특정 영역은 실재나 과정의 구조를 연구하고 다른 영역은 그것의 기능을 탐구한다. ③ 원인-결과 관계 : 한 영역에서 가정되는 실재가 다른 영역들에서 연구되는 현상의 원인일 수 있다.

이러한 장점에도 불구하고 IIT는 다음과 같은 문제를 해결해야 한다. 예를 들어 참된 교량 이론과 절충적 혼합 이론을 구별하는 기준이 제시되어야 한다. 또한 영역의 연결이 모두 해당 영역에 좋은 결과를 가져오는 것은 아니다. 이 문제와 관련하여 아브라함센(1987, 355)은 심리학과 언어학의 영역에서 두 영역에 부정적으로 작용하는 경계를 깨뜨리는(boundary-breaking) 관계와 긍정적으로 작용하는 경계를 연결하는(boundary-bridging) 관계를 구분하고 있다. IIT는 이러한 사례 연구를 통하여 규범적인 요소를 추출할 수 있을 것이다.

셋째 IIT는 과학을 통합할 수 있는 방법을 제시한다. IIT의 입장에서 보면 영역들은 교량 이론에 의해서 다양한 관계로 연결됨으로써 과학의 통합이 이루어진다. 다덴과 마울은 퍼트남 등이 제시한 '작업가설'과 비교되는 또 다른 '작업가설'을 제시하고 있다. 즉 과학의 통합은 교량 이론에 의해 결합된 영역들 간의 복잡한 관계망이 구성될 때 달성된다(Darden and Maull 1977,61).

실제로 비환원적인 방식으로 과학이 통합될 수 있는가? 분자 생물학의 사례는 이에 대한 긍정적인 대답을 제공한다. 분자 생물학은 고전 유전학을 환원하지 않지만, 그것을 매개로 관련되는 영역들, 즉 고전 유전학과 생화학과 같은 영역들의 상호 연결성과 의존성이 발생한다. 비록 생물학이라는 한 분야의 사례이지만 환원이나 대치가 없이 분야의 통합이 발생한다. 생물학은 이러한 통합 과정을 통하여 설명력의 확장을 얻게 된다.

전체적으로 평가하면 IIT는 위에서 제시된 조건들을 기본적으로 만족하지만 구체적인 측면에서 보면 비체계적이고 비형식적인 특징을 갖고 있다. 논리 경험주의에 대한 비판으로서 등장한 역사적 과학관들이 엄밀하지 못한 단점을 갖는 것처럼 IIT도 그와 동일한 단점을 지니고 있는 것이다. 더구나 IIT는 과학에서의 구체적인 사례를 기반으로 하기 때문에 기술적인 측면은 풍부하지만 규범적인 요소는 결여되어 있다. 그러나 이러한 문제점은 IIT가 이론적으로 발전하면 극복될 수 있을 것이다. IIT는 해결해야 할 문제가 많이 남아 있지만 인지과학 분야에 적용될 수 있는 가능성을 충분히 갖고 있다고 생각된다.

5. 결론

인지과학에 참여하는 제 분야들의 관계를 적절히 설명할 수 있는 모델을 검토해 보았다. 인지과학에 참여하는 분야들은 여러 차원에 속하므로 일차적으로 차원의 독립성을 유지할 수 있는 모델이 요구된다. GRR 모델은 차원간의 관계를 간과하는 점에서 그러한 요구 사항을 만족시키지 못한다. 또한 체계적 유물론 역시 차원간의 구분을 무시하고 있으며 그 결과 환원과 대치를 동일하게 보는 개념적인 오류를 범하고 있다. 이를 종합해 보면 환원론은 인지과학을 구성하는 제 분야간의 관계에 대해 적절한 설명을 제시하지 못한다는 것이 드러난다.

차원-영역간 이론(IIT)은 대안 모델이 갖추어야 할 조건들을 기본적으로 충족시킨다는 점에서 일단 환원론의 난점을 극복한다. 특히 문제 중심적인 접근 방식은 IIT가 대안 모델로서 발전할 수 있는 토대를 제공한다. 이 글에서는 IIT를 이용하여 인지과학의 제 분야간의 관계를 구체적으로 분석하는 작업이 시도되지 않았다. 그러한 작업은 본문에서 제시된 IIT보다 더 정교한 모델을 요구할 것이다.

참고 문헌

- Abrahamsen,A.(1987) "Bridging Boundaries versus Breaking Boundaries: Psycholinguistics in Perspectives". *Synthese* 72, 355-388.
- Abrahamsen,A.and Bechtel. W.(1991) *Connectionism and the Mind*. Basil Blackwell.
- Bechtel,W.(1984) "Reconceptions and Interfield Connections: The Discovery of the Link between Vitamins and Coenzymes". *Philosophy of Science* 51, 265-292.
- Bechtel,W.(1988) *Philosophy of Science*. LEA.
- Block.N. and Fodor,J.(1972). "What Psychological States Are Not?". *Philosophical Review* 81, 159-181.
- Carnap,R.(1938) "Logical Foundations of the Unity of Science". In O. Neurath et al(eds.), *International Encyclopedia of Unified Science 1*. Univ. of Chicago Press, 42-62.
- Causey,R.L.(1977) *Unity of Science*. Reidel.
- Churchland,P.M.(1979) *Scientific Realism and the Plasticity of Mind*. Cambridge Univ. Press.
- Churchland,P.M.(1989) *A Neurocomputational Perspective*. MIT Press.
- Churchland,P.S.(1986) *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. MIT Press.
- Darden,L. and Maull,N.(1977) "Interfield Theories". *Philosophy of Science* 43, 44-64.
- Darnell,J. et al.(1990) *Molecular Cell Biology 2nd ed*. Freeman.
- Feyerabend,P.K.(1962) "Explanation, Reductionism and Empiricism". In H. Feigl and G.Maxwell eds. *Minnesota Studies in the Philosophy of Science 3*. Univ. of Minnesota Press, 28-97.
- Fodor,J.A.(1975) *The Language of Thought*. Harvard Univ. Press.
- Gardner,H.(1985) *The Mind's New Science*. Basic Books.
- Kemeny,J. and Oppenheim,P.(1956) "On Reduction". *Philosophical*

- Studies* 7, 6-19.
- Kitcher, P. (1984) "1953 and All that A Tale of Two Sciences". *Philosophical Review* 93, 335-372.
- McCauley, R.N. (1986) "Intertheoretic Relations and Future of Psychology". *Philosophy of Science* 53, 179-199.
- Nagel, E. (1961) *The Structure of Science*. Harcourt, Brace.
- Nickles, T. (1973) "Two Concepts of Inter-theoretic Reduction". *Journal of Philosophy* 70, 181-201.
- Oppenheim, P. and Putnam, H. (1958) "Unity of Science as a Working Hypothesis". In H. Feigl et al. (eds), *Minnesota Studies in the Philosophy of Science* 3. Univ. of Minnesota Press, 3-36.
- Popper, K. (1972). "The Aim of Science". In *Objective Knowledge*. Oxford Univ. Press, 191-205.
- Pylyshyn, Z.W. (1984) *Computation and Cognition*. MIT Press.
- Richardson, R.C. (1979) "Functionalism and Reductionism". *Philosophy of Science* 46, 533-558.
- Sarkar, S. (1992) "Models of Reduction and Categories of Reductionism" *Synthese* 91, 167-194.
- Schaffner, K. (1992) "Philosophy of Medicine". In M.H. Salmon et al (eds), *Introduction to the Philosophy of Science*. Prentice Hall, 310-345.
- Staats, A.W. (1981) "Social Behaviorism, Unified Theory, Unified Theory Construction Method, and Zeitgeist of Separatism". *American Psychologist* 36, 239-256.
- Staats, A.W. (1991) "Unified Positivism and Unification Psychology". *American Psychologist* 46, 899-912.
- Stich, S. (1989) *From Folk Psychology to Cognitive Science*. MIT Press.
- Wimsatt, W.C. (1979) "Reduction and Reductionism". In P.D. Asquith & H. Kyburg (eds), *Current Research in Philosophy of Science*. PSA, 352-377.