

## Putnam의 心身理論에 관하여

李 鍾 權

### I. 머 리 말

주지하는바와 같이 精神(Mind)과 身體(Body)간의 관계에 관한 종래의 心身理論은 二元論(Dualism), 唯物論(Materialism) 및 行態論(Behaviorism)으로 대별되는 경향을 보이고 있다. 이 중에도 후자의 두 이론은 보통 二元論으로 취급되고 있다. Putnam은 1960년대에 걸쳐 발표한 일련의 논문을 통해 이와 같은 재래의 心身理論을 독특한 방식으로 분석, 비판하는 한편, 자기 나름의 독자적인 이론을 제시하고 있다. 그의 심신이론은 그러나 과학전반에 관한 그의 철학과 깊숙히 관련돼 있으며 따라서 Putnam의 전반적인 과학이론에 언급하지 않고서는 心身問題에 관한 그의 견해도 그 전모가 드러나지 않을 것으로 생각된다. 그러나 본논문에서는 가능한대로 전통적인 唯物論, 특히 心身同一論(Identity Theory)에 대한 Putnam의 批駁을 따로 떼내어 그것이 지닌 몇가지 문제점 및 논리적인 함축을 지적하고자 한다.

정신과 신체간의 관계에 관한 Putnam의 이론은 건설적인 측면과 파괴적인 측면을 다 포함하고 있다. 그는 위에서 말한 기존의 여러 이론들을 모두 비판, 배격하는 동시에, 그와같은 비판을 바탕으로하여 자기나름의 독자적인 이론을 전개하고 있다. 심리철학분야에서도 Putnam의 이론은 독특한 위치를 차지하고 있는데 그의 이러한 위치는 전통적인 여러 이론을 비판하는데 있어 그가 사용한 논변의 독특함과도 무관하지 않다. 그의 논변은 이른바 機能的 同型(functional isomorphism)의 개념에 의존하고 있다. 이 점에서 그의 이론을 機能的 同型論이라고 부를수도 있을 것 같다. 이 논문의 전반에서는 우선 그의 기능적 동형론에 관한 비판에 초점을 맞출 것이다.

기능적 동형의 개념에 입각한 Putnam의 논변에서는 인간과 로보트 내지는 일종의 計算器(Computing machine) 간의 類比推理(Analogy)가 결정적인 역할을 하고 있다. 따라서 그의 방법은 類推에 依한 論證(Argument from analogy)이라고 할만하다. Putnam의 기능적 동형론에 대한 이 논문의 비판은 類推論證이 갖는 일반적인 결함을 드러내는 것을 주축으로 하고 있다. 필자의 느낌으로는 類

推論證은 Putnam의 말을 빌리면 ‘ 좋지 못한 논증(bad argument)’이며 그러한 논증에 입각한 그의 心身理論 또한 허약한것 처럼 보인다. 이 점을 이 논문 전반에서는 선명하게 밝혀 보고자 한다.

Putnam의 논문은 함축이 다양하고 또 논지가 여러 갈래로 동시에 전개되는 듯한 부분이 많아 손쉽게 이해할 수가 없게 되어 있다. 따라서 우선 비판에 앞서 전통적인 심신이론에 대한 그의 비판의 요점을 그 논리를 선명하게 드러내는 방향으로 정리하여 보았다. 여기서 드러난 점은 Putnam이 종래의 唯物論 내지는 同一論에 대한 비판에서 중요한 하나의 전제를 묵시적으로 가정하고 있는듯 하다는 것이다. 필자는 그 전제를 ‘Indexicality Thesis’라고 이름 붙이었는데, 그러한 전제는 이른바 그의 실재론적 意味論과도 연관이 있는 것으로 생각된다. 어쨌든 그러한 전제가 암시하고 있는 문제는 同一論을 둘러싼 심리철학적인 논쟁을 새로운 차원으로 이끌어 가고 있는것 같다. 이 논문에서는 그러나 그러한 묵시적 전제를 드러내는데 그쳤으며 그 전제가 안고있는 문제점을 천착하는 일은 다음 기회로 미루었다.

## II. 機能的 同型理論

Putnam은 그렇지 않아도 여러 갈래로 복잡하게 논의되어온 心身關係에 관한 문제에 로보트 혹은 계산기와 같은 인간이 만든 機械의 精神(Minds of Machines)이라는 새로운 요소를 도입한 이유를 다음과 같이 설명하고 있다.

“...기계의 정신(Minds of Machines)에 관한 문제는 적어도 얼마간은 心理哲學의 전통적문제에 접근할 수 있는 흥미로운 새로운 돌파구를 제공해 줄 것이다.”<sup>1)</sup>

Putnam에 의하면 로보트 또는 계산기와 인간을 관련시켜 생각할 때, 주요한 心身理論에 관한 결정적인 문제점이 부각되는데, 과거 心身問題(Mind Body Problem) 혹은 他人의 精神에 관한 문제(Problem of other minds)를 취급한 학자들은 모두 이를 외면해 왔다. 따라서 이제는 로보트에 관한 문제를 논의하는 것이 마땅하다는 것이다.

Putnam의 생각에 의하면 心身問題 혹은 他人의 精神에 관한 문제에 관한 한, 人間과 예컨대 Turing machine 과 같은 로보트 혹은 계산기 사이에는 상당히 긴밀한 관련이 있다. Turing machine 이란 수리논리학자인 A. M. Turing

1) “Robots: machines or artificially created life?”

Putnam의 論文集 *Mind, Language and Reality*, 1975, Cambridge, p. 387.

이 1936년 제안한 일종의 이론적인 계산기이다. 그러나 우리는 Turing machine이 구체적으로 어떠한 기계인지 그 세부적인 내용을 알 필요는 없다. 단지 그것이 어떤 식으로 움직이는지 그 作動(operation) 방식을 결정하는 작동표(machine table)와 그것의 물리화학적 구조에 관한 진술만으로 그것이 완전히 규정된다는 것은 알기만 하면 된다.<sup>2)</sup> 이 계산기가 가질 수 있는 모든 기능적 혹은 논리적 상태(state)는 물론 작동표에 의해 주어진다. 일반적으로 작동표가 복잡해질 수록 그러한 작동표를 갖는 계산기가 가질 수 있는 논리적 상태의 수도 늘어날 것이다. 반면 그것이 어떤 종류의 물리화학적 상태를 가질 수 있는가 하는 것은 그것의 물리화학적 구조, 다시말해 계산기의 실제 설계에 좌우될 것이다. Turing machine을 電磁氣的으로 설계했다면 그것은 오직 전자기적인 상태만을 가질 수 있을 것이며, 力學的으로 설계되었다면 오직 기계적인 상태만을 가질 수 있게 될 것이다.

여기서 하나 주목할 것은 어떤 한 Turing machine의 작동표와 그것의 물리화학적 구조, 따라서 작동표에 의해 기술되는 그것의 논리적, 기능적상태와 실제 설계에 의해 결정되는 그것의 구조적 상태 간에는 아무 필연적인 관련이 없다는 사실이다. 왜냐하면 동일한 작동표를 갖는 Turing machine을 우리는 얼마든지 다른 방식으로 설계할 수 있을 것이기 때문이다. 전자기적으로 설계된 Turing machine과 역학적으로 설계된 Turing machine을 비교할 경우, 전자는 그 기능적 상태가 전자적인 상태와 관련을 가질 것이나 후자의 경우에는 역학적인 상태와 관련을 맺게 될 것이다. 그러나 동일한 작동표를 갖는 Turing machine의 경우, 그들의 기능적 상태들간의 관계가 그대로 유지된다는 의미에서 기능적으로 同型(functionally isomorphic)이라고 할 수 있다. 예를 들어 기능적 상태들간의 관계가 순서관계라고 할 경우, Turing machine 1이 항상 기능적상태 A를 가진다음 반드시 상태 B를 갖는다고 할때, Turing machine 2도 반드시 A에 대응하는 상태 A'를 가진다음 B에 대응하는 상태 B'를 갖게 된다는 것이다. 이것을 보다 일반화시켜 말한다면 어떤 假說 혹은 理論 H가 Turing machine 1의 기능에 관한 올바른 이론이라고 할 경우, 이론 H에 등장하는, Turing machine 1의 기능적상태 혹은 그것들간의 관계를 나타내는 용어들을 모두 그에 대응하는 Turing machine 2의 기능적상태 혹은 그것들간의 관계를

2) Turing이 실제로 제안한 계산기는 순전히 이론적인 계산기이기 때문에 작동표만으로 완전히 규정된다. 그러나 우리의 논의에서는 계산기의 기능뿐만 아니라 구조적인 측면도 문제가 되므로 이와같은 계산기를 고찰하려는 것이다.

나타내는 용어로 바꾸었을때 결과되는 이론 T'도 옳다는 것이다.<sup>3)</sup>

그런데 Putnam 에 依하면 작동표가 충분히 복잡한 계산기의 경우, 그 작동표에 의해 주어지는 계산기의 기능적상태와 그것의 실제구조에 의존하는 구조적상태간의 관계는 엇비슷하게 인간의 精神과 身體간의 관계에 비유될 수 있다는 것이다. 어떤 의미에서 비유될 수 있다는 말인지를 알기 위해서는 Putnam 의 이야기를 직접 들어보는 것이 가장 좋을듯 하다. 그는 논문 “Minds and machine”(1960)에서 다음과 같이 말하고 있다.

“앞서 제시한 Turing machine 의 논리적 상태와 인간의 정신상태, 그리고 Turing machine 의 구조적상태와 人間の 물리적상태간의 類比的 關係(Analogy)는 매우 시사하는 바가 많다고 생각한다.”<sup>4)</sup>

또한 다음과 같은 Putnam의 말도 매우 시사적이라고 생각된다.

“...唯物論과 二元論, 그리고 論理的 行態論의 類比的 理論(Analog)들이 Turing machine 의 경우, 왜 틀리는지를 알게되면 上記한 이론들이 人間の 경우에 왜 옳지 않은지 그 이유를 알 수 있을 것이다...”<sup>5)</sup>

이상과 같은 Putnam의 논문에서의 인용문과 실제로 그가 人間과 계산기 사이의 類比的 關係를 사용하고 있는 방식(modus operandi)을 살펴볼때 다음과 같은 것을 의미하고 있음이 분명하다. Turing machine 의 작동표를 충분히 복잡하게 할 경우, 동일한 작동표를 갖되 설계는 다른 계산기들간에 성립하는 기능적 동형관계가 Turing machine 과 人間간에도 성립할 수 있다는 것이다. 이것을 Functional Isomorphism Thesis (以下 FI 로 略)라고 부르기로 하자. FI 는 다음과 같이 定式化할 수 있다.

人間은 어떤 충분히 복잡한 작동표를 가진 Turing machine 과 機能的으로  
同型이다. (FI)

Putnam이 볼때 Turing machine 과 관련된 문제가 상당히 시사적일 수 있는 이유는 FI 가 성립하기 때문이다. 따라서 앞서 소개한 機能的同型的 개념에 입각하여 FI 가 의미하는 바를 살펴보는 것이 Putnam의 논점을 분명히 이해하는데 유용할 듯하다. 여기에 人間과 기능적으로 同型인 Turing machine 을 T라고 하자. 그리고 인간의 精神(mind)과 身體(body)간의 관계에 관한 어떤

3) “Philosophy and our mental life” 前掲 논문집 p. 292 참조.

4) 上掲논문집 p. 373.

5) “The mental life of some machines” 上掲 논문집.

중요한 이론 H가 있다고 하자. 이 이론H에 대해 우리는 그것을 대응하는, T에 관한 이론 H'를 다음과 같이 하여 얻을 수 있다. 이론 H에는 인간의 정신상태를 나타내는 용어와 신체적상태를 용어들이 포함될 것이다. 그러한 정신상태를 나타내는 용어들을 인간과 T 간에 성립하는 동형관계에 의해 그러한 용어들에 대응되는, T의(작동표에 따라 결정되는) 논리적상태를 나타내는 용어들로 바꾸기로 하자. 또한 인간의 신체적상태들, 예컨대 뇌의 상태를 지칭하는 용어들도 그것들에 대응되는, M의 구조적 상태를 지시하는 용어들로 바꾸기로 한다. 그 결과 얻어지는 M에 관한 이론을 H'라고 하자.

예컨대 우리는 인간의 心的體驗이 私密的(private)임을 주장하는 心理哲學의 이론 H에 대해 T의 논리적 상태에 대한 T 자신의 확인과정이 '私密的'임을 주장하는 이론 H'를 대응시킬 수 있다. 또한 인간의 精神狀態와 身體的狀態가 어떤 R 이란 관계에 있음을 주장하는

“mental state S는 어떤 physical state B와 R의 관계에 있다.”(I)

와 같은 진술에 대해서도 우리는 S에 대응하는 T의 논리적상태 S'와 B에 대응하는 T의 구조적상태 B'에 관한 다음과 같은 주장을 대응시킬 수 있다.

“Logical state S'는 어떤 Structural state B'과 R의 관계에 있다.”(I')

만약 R을 定義可能性(Being definable)의 관계라고 할 경우(I)은 論理的 行態論에 속하는 명제가 될 것이다. 또한 同一關係(Identity Relation)라고 할 경우에는 心身同一論(Identity Theory)의 명제가 얻어질 것이다. 이와같은 이론에 대응하여 우리는 Turing machine T의 논리적상태와 구조적상태간의 定義可能性을 주장하는 '行態論'과 두상태가 同一함을 주장하는 '同一論'을 얻을 수 있다.

그런데 Putnam이 주장하는 FI가 의미하는 바는 이와같이 서로 대응하는 두 이론들을 동시에 받아들이거나 동시에 거부할 수 있다는 것이다.<sup>6)</sup> 즉 대응하는 두 이론들은 각기 서로 논리적으로 同値(logically equivalent), 다시말해 H를 입증(prove) 또는 反證(refute) 할 경우, 그에 따라 H'도 입증 또는 반증되며 그 逆도 성립된다는 것이다. 이것을 보다 구체적으로 설명해 보자. 지금 이론 H를 입증 또는 반증하는 어떤 論證(Argument) A가 있다고 하자. 그 경우 그에 대응되는 논증 A'에 의해 H'도 입증 또는 반증될 것이며 그 반대도 마찬가지로 성립한다. 논증 A'은 A에서 인간의 정신적상태 혹은 신체적상태를 지시하

6) 上掲논문집 p. 369.

는 용어들은 Turing machine T의 적당한 논리적상태 혹은 구조적상태를 지칭하는 용어들로 바꾸면 얻어질 것이다.

위와같은 논의가 타당하다면 心身理論은 인간의 精神과 身體에 대해서만 고유하게 주장할 수 있는 이론이 아니다. 인간이 인위적으로 만들 수 있는 로봇트 혹은 복잡한 계산기에 대해서도 그와 논리적으로 대등한 이론을 제기할 수 있으며 또 대등한 논증에 의해 해결될 것이다. 이것은 人間の 정신적상태를 가리키는 용어와 신체적상태를 가리키는 용어간의 의미론적 차이와 Turing machine 같은 기계의 기능적상태와 구조적상태를 지시하는 용어간의 의미론적 차이와 대등하기 때문이다.

여기서 心身同一論, 즉 위의 진술(I)에서 R이 同一關係를 지칭하는 경우를 생각해 보기로 하자. 예컨대 ‘苦痛(pain)’이라는 心的狀態 P와 어떤 ‘腦의 狀態’ B 간의 同一함을 주장하는

$$\text{'Mental state P = Brain state B} \quad (\text{II})$$

와 같은 진술이 그에 해당할 것이다. (II)에서 말하는 同一關係가 구체적으로 어떤 의미의 관계인지 의문을 제기할 수 있을 것이다. 그러나 일단 “熱은 에너지의 일종이다(Heat is a form of energy)” 혹은 “光은 電磁波이다(Light is an electromagnetic wave)”와 같은 진술에서와 같이 理論的 同一化(Theoretical Identification)를 나타내는 同一關係라고 보는 것이 가장 무난할 듯하다.

그러나 Putnam의 주장에 의하면 (II)에서 말하는 동일관계가 이론적인 동일성을 나타낸다고 할 경우, 그것은 적어도 物理的인 의미에서 必然的인(physically necessary) 동일성이야 한다는 것이다.<sup>7)</sup> 즉 현실세계의 물리화학적 법칙이 유지되는 한 항상 성립해야 한다는 말이다. 그러나 그것은 물리적인 동일관계가 될 수가 없다. 그 이유는 (II)를 Turing machine의 경우로 번역해 보면 분명하다. 만약 (II)가 心的狀態와 身體的狀態간의 이론적 동일성을 주장하는 명제라면 (II)에서 심적상태 P를 Turing machine의 어떤 논리적상태 P'로 바꾸고 腦의 狀態 B를 T의 구조적상태 B'로 대체하여 얻은 진술

$$\text{'Logical state B' = Structural state B'} \quad (\text{II}')$$

도 역시 이론적 동일성을 주장한다고 말할 수 있다. 따라서 Turing machine의 논리적상태와 구조적상태간에도 물리적으로 필연적인 동일관계가 성립한다는 결론이 된다. 그러나 위에서도 언급한 바와 마찬가지로 한가지 작동표를 갖는 계

7) 上掲논문집 p. 418.

산기를 우리는 여러가지 다른 방식으로 설계하는 것이 가능하므로 그것의 논리적상태 P'과 구조적상태 B' 간에는 아무 필연적 관련이 없다고 할 수 있다. 이것은 곧 P'과 B'간의 물리적 필연적 동일성을 주장하는 명제 (II') 이 사실이 아님을 의미한다. 그렇다면 Functional Isomorphism Thesis FI 에 의해 인간의 심적상태와 물리적상태간의 동일성을 주장하는 명제 (II) 에 대해서도 같은 말을 할 수 있다. 즉 인간의 경우에도 같은 心的狀態에 있으면서 다른 신체적상태에 놓이는 것이 실제의 물리화학적 법칙에 비추어 불가능하지 않으므로<sup>8)</sup> (II) 도 사실이 아니라고 말할 수 있다는 것이다.

종래의 心身同一論은 이렇게 하여 논파되는데 論理的 行態論에 대한 Putnam의 비판도 이와 마찬가지로 방식으로 진행된다. 논리적 형태론의 주장을 Turing machine T의 경우로 바꾸어 놓으면

“ T의 논리적상태에 관한 진술은 T의 실제적 혹은 잠재적 ‘行態(behavior)’에 관한 진술과 의미론적으로 대등하다”

와같은 명제가 될 것이다. 그러나 위의 명제는 틀린 명제이다. 왜냐하면 어떤 한 종류의 실제적 혹은 잠재적 형태는 동일하면서도 그에 대응되는 논리적 상태는 상이한 Turing machine 을 얼마든지 생각할 수 있을 것이기 때문이다. 이것은 ‘非定常的인(pathological)’ 작동표를 갖는 계산기의 경우이다. 따라서 FI에 의해 人間の 경우에도 그와같은 病的인 인간을 생각하는 것이 가능하다는 결론이 된다. 또한 이와같은 결론에 의해 위의 명제에 대응되는 인간의 정신에 관한 형태론적 주장도 부인된다.

위에서 본바와 같이 FI는 종래의 논리적 형태론 또는 동일론을 부정하는데 결정적인 역할을 하고있다. 그런데 Putnam은 종래의 심신이론을 논파하기 위한 파괴적인 작업뿐 아니라 새로운 이론을 수립하려는 건설적인 작업에도 큰 몫을 하고 있다. 그것은 FI로부터 Putnam의 심신이론이 자연스럽게 나오기 때문이다. 즉 인간의 심적상태는 온전한 유기체의 기능적상태이다. 그러한 상태는 Turing machine 과 같은 계산기의 개념을 사용하여 설명될 수 있다.<sup>9)</sup> FI가

8) 예컨대 火星人과 같은 가상적 인간은 우리와는 전혀 다른 신체적 구조를 갖고 있기 때문에 심적상태 P에 있으면서 물리화학적으로는 우리와 전혀 다른 신체적상태 B에 있을 수 있다는 주장이 결코 분석적으로 僞가 되지 않는다. “Rejoinder to Alvin Plantinga” 上掲논문집 p. 427 참조.

9) 이것은 Putnam의 초기입장이다. 1973년 논문에서는 心理的狀態가 Turing machine의 상태로 설명될 수 없다고 말하고 있다. 그러나 이 문제는 우리의 논의에는 관계가 없으므로 더 이상 상세히 다루지 않기로 한다.

중요한 역할을 하고있는 만큼 FI 에 입각한 Putnam의 논문에 결함이 발견된다면 그것은 그의 心身理論 전반에 대해 중대한 결과를 존재하리라고 생각된다. 아래에서는 FI 에 의거한 그의 類比的 推理를 고찰해 보기로 하겠다.

### III. Putnam의 類推논증에 대한 비판

Putnam의 心身理論에서 가장 核心的인 역할을 하고있는 것은 위에서도 언급했던 것처럼 人間 또는 有機體와 計算機 혹은 로봇과같은 인간이 제조한 기계간의 類比推理 (Analogy)이다. 이것은 구체적으로 心身同一論을 둘러싼 논의에 적용한다면 心的狀態 P와 어떤 腦의 狀態 B의 理論的 同一性을 주장하는 명제 (Ⅲ)와 그에 대응되는, Turing machine T의 論理的狀態 P'과 構造的 狀態 B'간의 理論的 同一性을 주장하는 명제 (Ⅲ')은 논리적으로 동치관계에 있다. 그러한 동치관계를 제공하는 것이 이른바 Functional Isomorphism Thesis FI이다. 그런데 Turing machine의 경우 어떤 논리적상태 P'에 있으면서 B'과는 다른 구조적상태를 갖는 것이 적어도 현실세계의 실제의 물리화학적 법칙에 비추어 원리상으로 가능하므로 (Ⅲ)는 부정된다. 그러므로 같은 논리에 의해 인간의 경우에도 心的狀態 P에 있으면서 B와는 다른 腦의 상태를 갖는 것이 원리상으로 가능하다고 말할 수 있으며 따라서 (Ⅲ)에서 표현하고 있는 이론적인 의미의 同一關係도 부정되어야 한다.

(Ⅲ)를 논박하기 위한 위와같은 類推的論證은 (Ⅲ)이 僞라는 사실 이외에 (Ⅲ)와 (Ⅲ')이 論理的으로 同値임을 주장하는 FI에도 바탕을 두고 있다. 따라서 (Ⅲ)의 眞僞가 확실히 판명됐을 경우, (Ⅲ)를 입증 또는 반증하기 위한 논증적 부담은 곧장 FI를 입증해야 하는 부담으로 옮겨진다는 말이된다. 다시말해 (Ⅲ)의 진위가 판명됐다 하더라도 FI가 입증되지 않는한 위의 논증은 완전치 못하다는 것이다. 그러나 Putnam은 어느 곳에서도 FI를 입증할 만한 논변을 제시하고 있지 않다. 이점을 다음과 같은 물리학에서의 예와 전주어서 고찰해 보기로 하자.

力學에서 탄성계수 k인 spring 끝에 달린 質量m의 물체는 다음 微分方程式으로 표시되는 운동을 한다는 것이 알려져 있다.

$$m \frac{d^2x}{dt^2} + kx = 0 \quad (a)$$

여기서 x는 spring이 본래의 위치에서 늘어난 길이를 나타낸다. 또 t는



시간을 나타내는 변수이다. 한편 전기역학에서는 전기용량 C인 축전지를 Inductance 가 L 인 Inductor와 導線으로 연결했을 경우, 다음으로 나타내어지는 電荷의 변화가 있다는 것도 알려져 있다.

$$L \frac{d^2q}{dt^2} + \frac{1}{C} q = 0 \quad (b)$$

여기서 q 는 축전지의 電荷의 양을 가리킨다. 지금 mass -spring system (以下 MS - system으로 略)과 inductor - capacitor system (以下 IC - system 으로 略)을 다음과 같은 대응규칙에 의해 관련시킨다고 하자.

$$\begin{aligned} x &\leftrightarrow q \\ m &\leftrightarrow L \\ k &\leftrightarrow \frac{1}{C} \end{aligned}$$

이때 MS - system의 舉動 (behavior)을 記述하는 微分方程式 (a) 는 정확하게 IC - system의 舉動을 나타내는 식 (b) 로 옮겨진다. 이것은 MS - system과 IC - system 사이에 Putnam이 말한 의미의 機能的 同型關係가 성립함을 보여주는 것이다. 따라서 MS - system의 質量, 물체의 變位, spring의 彈性係數 및 시간간의 관계를 진술하는 임의의 문장 S (x, m, k, t)에 대해 x를 q로, m을 L로, k를  $\frac{1}{C}$ 로 바꾼 문장 S'(q, L, C, t)=S(q, L,  $\frac{1}{C}$ , t)를 대응시킬 경우 S와 S'사이에는 논리적 동치관계가 성립한다. 그러므로 S의 眞僞를 입증하는 대신 S'의 眞僞를 입증하면 그것을 입증하는데 사용된 논증 A'에 대응되는 논증 A에 의해 S의 眞僞도 입증된다는 것이다.

위의 예에서는 그러나 두 system 간의 기능적 동형관계가 명시적으로 입증돼 있다. 위에서 제시한 微分方程式 (a) 및 (b)와 대응규칙의 존재가 그것을 입증하는 것이다. 그런데 Putnam의 논의에서는 인간과 계산기간에 기능적 동형관계가 존재한다는 사실 즉 FI를 입증할 만한 적극적인 논증이 없다. 그는 FI를 비판하는 여러 사람의 여러가지 논점을 이리저리한 방식으로 반박하고 있기는 하다. 그러나 그 어느것도 FI에 대한 적극적인 증명에까지는 이르지 못하고 있다는 느낌이다. 이것이 Putnam의 類推的 論證이 갖는 일차적인 결함이다. 그러나 보다 심각한 문제는 다른데 있는듯 하다. 이제 그 점을 지적해 보기로 하겠다.

위에서도 언급한 바와같이 Putnam의 類推的 論證에서는 어떤 심신이론 H를 입증해야 할 논증상의 부담이 Functional Isomorphism Thesis FI를 증명해

야하는 부담으로 옮겨가 있다. H에 대응되는, 계산기에 관한 심신이론 H'이 확인됐다고 할 경우, FI가 입증됐다고 하면 H도 입증된 셈이다. Putnam은 FI가 사실이라는 점을 입증하는 구체적증명을 제시하고 있지는 않지만 만약 그러한 증명을 제시하기 위해서는 그는 어떻게 해야 할 것인가? 위에서는 물리학의 예에서처럼 인간의 심신에 관한 법칙과 계산기같은 로봇트에 관한 법칙을 제시하고 그 법칙들로부터 양자간의 기능적 동형관계가 존재함을 보일수도 있을 것이다. 그럴 경우에는 FI가 적극적으로 입증됐다고 이야기할 수 있을 것이다. 그러나 FI가 만약 이러한 방식으로만 입증된다면 Putnam의 FI를 기반으로한 類推的 論證은 그렇게 좋은 논증방식이라고 할 수가 없다. 그 이유는 다음과 같다.

이제 위에서 언급한 방식에 따라 FI가 입증되었다고 하자. 즉 인간의 심신에 관한 어떤 사실적인 법칙 또는 이론 L과 어떤 복잡한 계산기 내지는 로봇트의 '심신'에 관한 사실적인 법칙 또는 이론 L'을 구한다음, 이들 이론을 비교함으로써 양자간에 기능적인 동형관계가 성립한다는 사실을 입증하였다. 그렇다면 이러한 동형관계에 의거, 인간의 심신관계에 관한 어떤 이론 H도 입증된 셈이다. 그러나 이러한 절차는 부질없이 우회적이다. 왜냐하면 위의 경우 FI 대신 이론 L로부터 곧장 H가 입증될 수 있을 것이기 때문이다. 이것을 앞서든 MS-system과 IC-system의 예와 관련하여 설명해 보기로 하자. 위의 MS-system과 IC-system의 예에서는 양자간의 '기능적 동형관계'가 명시적으로 입증돼 있다. 그러나 그것은 양 system의 擧動을 기술하는 微分方程式 (a)와 (b)를 토대로 한 것이다. 따라서 이 경우에 예컨대 MS-system에 관한 어떤 진술 S를 두 system간의 기능적 동형관계와 S에 대응하는 IC-system의 진술 S'으로부터 확인하려 한다는 것은 전혀 무의미하다. 왜냐하면 두 system간의 기능적 동형관계가 성립한다는 사실은 두 system의 擧動을 기술하는 微分方程式에 의존하고 있는바 진술 S의 진위는 그러한 기능적 동형관계로 부터가 아니라 微分方程式으로부터 곧장 확인될 수 있기 때문이다. 즉 두 system간에 동형관계가 성립한다는 사실에 대한 증명은 우리가 원하는 문제에 관해 해답을 줄수있는 법칙 내지는 이론을 바탕으로 하고 있다. 따라서 그러한 관계에 입각한 논변은 결코 좋은 논변이 될수 없는 것이다.

예를들어 여기 어떤 부자가 자기집을 방문한 어떤 聖者에게 다음과 같이 질문했다고 하자: "부자도 과연 천당을 들어갈 수 있습니까?" 이에 대해 그 聖者가 다음과 같이 답변하였다고 하자. "부자가 천당에 들어가는는 낙타가 바늘구멍에 들어가기만큼 어렵다." 이 부자는 아마 그러한 聖者의 답변에 실망한 나머지 비판했을지 모를일이다. 그러나 그는 불만에 찬 나머지 왜 부자의 경우에는 천당을 들어가

기가 나타나 바늘구멍을 들어가기만큼 어렵다는 것인지 그 이유를 대라고 항의할 수도 있을 것이다. 이에 대해 聖者는 어떤식으로 답변해야 할 것인가? 그는 아마도 부자가 천당에 들어가기에 어느 정도로 어려운지 그 이유를 설명하고 또 나타나 바늘구멍에 들어가기에 얼마나 어려운 일인가를 밝힌 다음 양자의 難易度가 같다는 점을 지적해야 할 것이다. 그러나 그러한 지적에 앞서 聖者는 이미 부자의 물음에 적절한 답변을 줄 수 있는 이유를 제시하고 있다. 즉 聖者의 나타나에의 비유는 부자의 질문에 직접적인 답을 줄 수 있는 사실을 바탕으로 하고 있다. 그렇기 때문에 聖者의 비유적인 답변은 修辭的인 측면에서라면 몰라도 논리적인 측면에서는 그렇게 좋은 논변이 될 수 없는 것이다. 이와같은 상황은 연역론자들이 당하는 곤경과도 비교가 될 듯하다.

우리는 명제 ( $Q_1$ ): 모든 영웅은 短命하다와 명제 ( $Q_2$ ): Napoleon은 영웅이다라는 두개의 전제로 부터 ( $Q_3$ ): Napoleon은 短命하다라는 결론을 필연적으로 확실하게 추론할 수는 있다. 그러나 ( $Q_1$ )과 ( $Q_2$ )에 의거해서  $Q_3$ 를 설명한다는 것은 그다지 좋은 설명은 되지 못한다고 귀납론자들은 주장한다. 왜냐하면 ( $Q_1$ )이 眞이라는 것은 ( $Q_3$ )이 眞임을 함축하고 있기 때문이다. 즉 ( $Q_1$ )을 입증하는 것은 ( $Q_3$ )을 입증하는 것과 무관하지 않다. 주어진 가정 또는 이론의 입증방법과 독립적이 되지 않은 입증방법을 가진 전제에 의해 그 이론을 설명하려 한다면 좋은 설명이 될 수 없다.

위의 논의에서 드러난 점은 Putnam이 안고 있는 문제는 보다 근본적인 문제라는 사실이다. Putnam은 그의 Functional Isomorphism Thesis를 입증하고 있지 않지만 문제는 보다 원칙적인데 있다. Putnam의 방법에 의하면 인간의 심신관계에 관한 어떤 이론 H를 입증하는 절차는 인간과 기능적으로 동형인 Turing machine과 같은 기계에 착안하여 그 기계에 관한 이론 H'을 입증하는 것으로 이루어져 있다. 그러나 기능적 동형관계의 가정을 입증하기 위해서는 인간 및 Turing machine에 관한 어떤 사실적 지식 내지는 이론이 필요하다. 그런데 이러한 사실적 지식 내지는 이론은 우리가 해결하려는 문제 H 및 H'에 대해 직접 답변을 제공할 수 있다. 즉 Putnam의 FI 가정은 우리의 문제에 직접적인 답을 줄 수 있는 이론을 바탕으로 해서 입증된다. 예컨대 H는 FI를 입증하는 데 사용된 인간의 심신관계에 관한 어떤 이론으로 부터 직접 증명될 수 있다. 그렇다면 心身問題에 관한 논의에 로보트 혹은 계산기와 같은 인위적인 끌어들이는 시도에 큰 의의를 찾을 수는 없을 것 같다. 인간과 로보트간의 어떤 類比관계에 입각한 그의 類推的인 논증은 다른 모든 類推的인 논증과 마찬가지로 대단한 설명적 효과를 가질 수가 없

는 것이다. 類推的 論證이 성공했다고 할 경우, 우리는 언제나 그러한 유추에 의거하지 않은 보다 직접적인 논증을 찾아 낼수 있다. 왜냐하면 유추적인 논증은 우리의 문제에 직접적인 답변을 제공하는 이론 내지는 가설에 의존하고 있기 때문이다. 그러한 이론 내지는 가설을 명시적으로 보이고 있는 유추논증은 부질없이 우회적이다. 그러나 그러한 가정을 명시적으로 보이지 않으면 그것은 일종의 先決問題要求를 범하고 있는 셈이다. Putnam도 이러한 곤궁에서 빠져 나올 수 없는 처지에 놓여있는 것 같다.

#### IV. 心的狀態로 나타내는 用語의 Indexicality 문제

心身同一論은 인간의 心的狀態 P가 어떤 신체적상태, 예컨대 腦의 어떤 상태와 이론적으로 동일한다는 것을 주장하고 있다. 우리는 이것을 위에서 다음과 같이 정식화한 바 있다:

$$\text{"Mental state } P = \text{Brain state } B \text{"} \quad (\text{II})$$

우리는 또한 위에서 인간이 心的狀態 P에 있으면서도 B와는 다른 腦의 상태를 갖는 것이 가능하기 때문에 (II)는 부정된다고 말한바 있다. 이 문제를 좀더 천착해 보기로 하자.

Putnam에 의하면 人間이 腦의 상태 B에 놓이는 경우 또 오직 그 경우에 限해 心的狀態 P를 느끼는 것이 물리화학적으로 필연적인 사실이라 하더라도 心身同一論의 명제 (II)는 받아들일 수 없다는 것이다.<sup>10)</sup> 그 이유는 心身狀態를 나타내는 用語를 사용하는 실제 例에 비추어 볼때 다음과 같은 가상적인 경우가 전혀 비논리적인 것으로 보이지 않기 때문이다. 즉 우리는 우리와는 전혀 다른 체질을 가진 화성인과 같은 가상적인 인간을 생각해 볼수 있다. 이 화성인은 우리와 체질이 전혀 틀리므로 실제 인간과는 전혀 다른 신체적 상태를 가질 것이다. 그러면서도 우리와 똑같은 심적상태를 느낀다고 가정할 수도 있을 것이다. 그러한 가상적인 인간의 경우는 심신동일론의 주장은 성립할 수 없다.

그런데 위의 화성인의 예는 현실세계가 아닌 일종의 가상세계 혹은 가능세계에서의 일이다. 따라서 Putnam의 말은 어떤 종류의 가능세계에서는 (II)가 성립하지 않는다는 뜻으로 풀이된다. 다시말해 Putnam의 주장은 일단

$$\diamond (\text{Mental state } P \neq \text{Brain state } B) \quad (\text{III})$$

10) "Rejoinder to Alvin Plantinga" 前掲논문집 p. 427.

으로 정식화할 수 있을 듯하다. 그런데 (III)은 가능성을 나타내는 논리적인 연산기호 '◇'와 필연성을 나타내는 기호 '□'간의 관계에 따라 다음과 같이도 나타낼 수 있다.

$$- \square (\text{Mental state } P = \text{Brain state } B) \quad (\text{IV})$$

그런데 Putnam의 주장은 (III)이 성립하기 때문에 (II)가 사실일 수 없다는 것이다. 그러나 (II)는 (III) 또는 (IV)가 성립한다고 해서 부정할 수 있는 것은 아니다. 가상적인 화성인의 경우에는 心身狀態 P가 腦의 狀態 B와 갖지 않을지 모르지만 현실세계(actual world)에 사는 인간의 경우에는 P가 B와 同一하다고 말하는 것이 무의미하지 않을 수도 있기 때문이다. 그렇다면 Putnam은 어떤 이유에서 (III)을 시사하는 화성인과 같은 공상과학의 예에 의해 心身同一論의 진술 (II)가 부정된다고 생각하고 있는 것인가? 위에서 언급한 것처럼 명제 (III)으로부터 직접 (II)를 부정하는 명제가 나오지 않는다. 그러나 (III)에 덧붙여서 다음과 같은 명제를 가정할 경우 우리는 (II)와 모순되는 명제를 얻을 수 있다.

$$\diamond (\text{Mental state } P \neq \text{Brain state } B) \rightarrow (\text{Mental state } P \neq \text{Brain state } B) \quad (\text{V})$$

왜냐하면 (III) & (V)  $\rightarrow$   $\neg$  (II)임을 금방 알 수 있기 때문이다. 그런데 우리는 양상 논리의 추론규칙에 의해 (V)로 부터 다음을 얻을 수 있다. 즉,

$$(\text{Mental state } P = \text{Brain state } B) \rightarrow \square (\text{Mental state } P = \text{Brain state } B) \quad (\text{VI})$$

위와같은 논의에 비추어 볼때 Putnam은 명시적으로는 말하고 있지 않지만 (VI)을 묵시적으로 가정하고 있는 듯 하다. 적어도 다음과 같이 말할 수는 있다. (VI)을 가정하지 않는 한, Putnam이 든 火星人의 예에 의해 종래의 心身同一論이 부정되지는 않는다. 그렇다면 명제 (VI)은 무엇을 의미하는가? 그것은 心的狀態를 나타내는 用語 P가, Putnam의 用語를 사용해서 말한다면, indexical하다는 것을 의미하는 것으로 해석된다. 즉 현실세계에서 예컨대 '苦痛'이라는 심적상태를 나타내는 用語가 이러저러한 신체적상태를 지시한다면 어떤 可能世界에서도 그 用語는 그와 같은 신체적상태를 지시해야 한다는 것이다. 다시말해 心身同一論에서의 同一關係는 다음과 같은 의미에서 필연적인 동일관계를 나타낸다. 心的狀態와 身體的 狀態간의 同一性이 현실세계에서 성립할 경우 그러한 관계는 모두 가능한 세계에서 다 성립해야 한다. 그러나 앞에서 말한 화성인이 등장하는 가상적인 세계에서는 그러한 관계가 성립하지 않는다. 따라서 현실세계에서도 心的狀態와 身體的狀態는 동일할 수 없다.

위와같은 논의는 心身關係를 둘러싼 논쟁을 새로운 방향으로 이끌어 가는 것같

다. 그것은 즉 心的狀態를 나타내는 용어들이 과연-고유명사 내지는 自然的인 種類(natural kinds)를 나타내는 用語등과 마찬가지로-indexical 한 단어들 인가 하는 문제이다.

위에서 지적한 것처럼 (III) 과 (V) 는 -(II)를 함축한다. 그런데 (V)는 Indexicality Thesis (VI)와 논리적으로 同値이므로 (III) 과 (VI) 도 -(II) 를 함축한다. 즉,

$$(III) \& (VI) \rightarrow -(II)$$

위의 문장을 다시 고쳐쓰면

$$-(II) \& (III) \& (VI)$$

위의 문장은 명제 (II)와 (III) 및 (VI) 이 동시에 성립할 수 없다는 것을 말하고 있다. 이 셋중에서 기껏해야 둘만이 사실일 수 있다는 것이다. Putnam 이 든 火星人의 예가 타당하다고 할 경우, 즉 현실세계의 인간과 같은 심적상태를 가지면서 신체적상태는 전혀 다른 가상적인간이 존재할 수 있다고 할 경우 (III)은 움직일수 없는 사실이 되므로 우리들이 선택할 수 있는 대안은 결국

$$(II) \& (III)$$

혹은

$$(VI) \& (III)$$

의 두가지중 하나가 된다. Putnam은 후자쪽 즉 心的狀態를 지시하는 用語들이 indexical 함을 주장하는 길을 선택하고 있는 것이라고 해석된다. Putnam 은 (VI) 을 명시적으로 주장하고 있지 않다. 心的狀態를 나타내는 용어들이 과연 indexical 한 용어인가에 대해서는 더더욱이 아무 도움이 될만한 시사도 주고 있지 않다. 단지 그가 다른 곳에서 대부분의 명사들이 indexical 혹은 rigid하다고 말하고 있는 점에 비추어 볼때<sup>11)</sup> Putnam 이 아마 후자의 길을 선택하리라는 것이 더욱 확실하게 생각될 뿐이다. 여기서 心的狀態를 나타내는 用語들의 Indexicality 를 주장하는 명제 (VI) 에 관해 상세한 논의를 할 수는 없다. 단지 Putnam 과 心身同一論者들간의 한가지 대비점을 고찰하는데 그치기로 하겠다.

Putnam 이 (VI) & (III) 을 선택한 것은 心身同一論 (II) 를 부정함을 의미한다. 따라서 火星人의 예에도 불구하고 심신동일론자들이 자신들의 주장을 고집하려면 (II) & (III) 의 길을 선택할 수밖에 없다. (II) 와 (III) 을 선택했다는 것은 무엇을 의미하는가? 그들은 火星人에 대해 다음과 같이 주장할 수밖에 없을 것이다. 즉 화성인이 心的狀態 P에 놓일때마다 항상 현실세계에 속하는 인간의 腦의 상태 B와는 다른 腦의 상태 B<sub>1</sub>에 놓인다는 것이 발견되었다고 하자. 이경우 그들은 화성인들에 대

11) "Meaning of 'meaning'" 上掲논문집 pp. 242-244 를 참조할 것.

해서는

“ Mental state P = Brain state B ” (  $B_1 \neq B$  ) (VII)

와같이 말해야 할것이다. 명제 (III) 을 수용하면서 심신동일론을 계속 주장할 수 있는 방법은 이와같은 길밖에 없을 것이기 때문이다. 물론 ‘P’ 가 indexical 한 단어라는 입장에 선다면 (II) 와 (VII) 을 동시에 주장하는 것 자체가 불가능하다. 그러나 문제는 (II) 와 (VII) 을 동시에 受容하게끔 心身同一論을 해석하는 것이 불가능하나 하는 점이다. 만약 그러한 해석이 실제 심적상태를 지시하는 용어들의 사용방식에 의해서 가능하다고 하면 오히려 Indexicality Thesis (VI) 이 부정돼야 할 것으로 보인다. 따라서 심신동일론자들에게 가장 유리한 해석을 내려보기로 하자.

心身同一論이 서로 동일하다고 말하는 心的狀態나 腦의 상태는 어떤 단일한 특정상태를 말하는 것이 아니다. 우리는 심신동일론이 다음과 같은것을 의미한다고 생각할 수 있을 듯하다. 즉 어떤 임의의 유기체  $x$  에 대해  $x$  의 心的狀態 P는 그것의 어떤 腦의 狀態 B와 동일하다. P는 곧 B를 가리킨다. 여기서 腦의 상태 B는 유기체의 種에 따라 달라질 수 있다. 예를들어 인간의 고통의 상태와 동일한 뇌의 상태와 고양이의 고통의 상태와 동일한 고양이의 뇌의 상태는 서로 다를 수가 있다. 이점은 가능세계의 인간을 논의에 끌어넣기 이전의 이야기므로 전통적인 심신동일론이 모두 수긍하고 들어가야할 명제라고 생각된다. 이와같은 논의를 Putnam이 예로든 가상세계 내지는 가능세계에도 확대할 수도 있지 않겠는가? 사람과 동물들을 모두 포괄하는 심신동일론의 테두리내에서라면 어떤 가능세계의 인간의 심적상태 P는 현실세계의 인간의 腦의 상태 B와는 다른 腦의 상태  $B_1$  과 동일하다고 말하는 것이 의미있을지 모른다. 그러나 필자의 생각으로는 동물들의 예에서와는 달리 가능세계의 인간의 경우에는 상당한 난점이 있는 것으로 보인다. 그러나 Indexicality Thesis 를 둘러싼 심신동일론자들과 Putnam과의 대립에 관해서는 본 논문에서는 그 개괄적인 측면을 기술하는데 그치고 본격적인 논의는 다음기회로 미루기로 하겠다.

## V. 맺 음 말

본 논문에서는 우선 인간과 로봇 혹은 계산기같은 人造機械간의 기능적 동형관계의 개념에 입각한 Putnam의 유추적논증의 방법을 소개하고 유비추리가 갖는 일반적인 결함에 의거해 Putnam의 방법의 문제점을 시도해 보았다. 人間과 충분히 복잡한 작동표를 갖는 계산기간에 기능적 동형관계가 성립한다는 Functional Isomorphism Thesis 는 전통적인 이론을 논박하는데 뿐만아니라 그와 대립되는

새로운 이론을 수립하는데도 결정적인 역할을 하고 있다. 따라서 그러한 Thesis가 흔들린다면 Putnam의 심신이론 전체에 큰 손상이 갈 것이라고 생각된다.

또한 본논문 후반부에서는 心身同一論에 대한 Putnam의 비판을 분석함으로써 그가 心的狀態를 나타내는 용어들에 대한 Indexicality Thesis를 암묵적으로 승인하고 있음을 드러내 보였다. 이 Thesis에 관한 상세한 논의는 여기서는 시도하지 않았으나 이 문제는 앞으로 心身理論을 새로운 차원으로 이끌수도 있을 것으로 기대된다.

### 参 考 文 献

- H. Putnam, "Philosophy and our mental life" (1973)  
 \_\_\_\_\_, "Brains and behavior" (1963)  
 \_\_\_\_\_, "Minds and machines" (1960)  
 \_\_\_\_\_, "Robots: machines or artificially created life?" (1964)  
 \_\_\_\_\_, "The mental life of some machines" (1967)  
 \_\_\_\_\_, "The nature of mental states"  
 \_\_\_\_\_, "Logical positivism and the philosophy of mind" (1969)  
 \_\_\_\_\_, "The meaning of 'meaning'" (1975)  
 (상기 논문은 모두 H. Putnam의 논문집 *Mind, Language and Reality*, Cambridge: Cambridge University Press, 1975에 수록된 것임)
- John O'Connor ed., *Modern materialism: Readings on Mindbody Identity*,  
 New York: Harcourt, Brace & World, Inc., 1969
- Norman Malcolm, *Problems of mind*, George Allen & Unwin Ltd., 1971.