

형이상학적 기능주의 고찰

— 기능주의 분류와 감각질(Qualia) 문제를 중심으로 —

金 泳 植

(외 대)

I

‘기능주의(functionalism)’란 용어는 인류학, 문예비평, 심리학, 심리철학 등 다방면에 걸쳐 여러 맥락에서 쓰이고 있는데 여기서 고찰하려고 하는 기능주의는 심리철학에서 정신의 본성이 무엇인가 하는 문제를 다루는 형이상학적 기능주의이다. 기능주의자들에 따르면, 정신적 상태란 간략히 말해 기능적 상태이다. 다시 말해, 정신적 상태란 그것의 인과적 역할들로써, 즉 감각적 자극들, 행동적 반응들 그리고 다른 정신상태들과의 인과적 관계들으로써 정의될 수 있다. 예를 들어, 고통이란 세포의 손상에 의해 야기되는 경향성, 그것을 없애려는 욕구를 야기하는 경향성, 신체의 손상된 부분을 그 손상을 유발한 것으로 생각되어 지는 것으로부터 분리시키려는 행위를 야기하는 경향성 등을 통해 규정될 수 있다고 형이상학적 기능주의자들은 주장한다.

II

형이상학적 기능주의는 보통 크게 두개의 계열로 나뉘고 있는데 그 하나는 퍼트남-포도-하만 계열이고 다른 하나는 스마트-암스트롱-루이스-슈우메이커 계열이다. 전자를 기능 상태 동일론(Functional State Identity Theory), 튜링 기계 기능주의(Turing Machine Functionalism), 혹은 심리기능주의(Psychofunctionalism)등으로 부를 수 있으며, 이에 상응하여, 후자를 기능 명기론(Functional Specification Theory), 비튜어링 기계 기능주의(Non-Machine Functionalism), 혹은 상식 기능주의(Common Sense Functionalism)등으로 부를 수 있을 것이다. 네드 블록은 *Readings in Philosophy of Psychology*[11]의 제 3부 기능주의 서문에서 이 두 계열간의 가장 두드러진 차이점으로 앞의 계열은 기능주의가 물리주의의 허위성을 보여주고 있음에 반해 뒤의 계열은 기능주의가 물리주의의 진리성을 보여주고 있다는 점을 짚고 있다. 그러나 필자의 생각으로는 블록의 이러한 양 계열 구분 기준은 잘못된 것인 듯싶다.

주 1) Harvard University Press : Cambridge, 1980

우선 물리주의는 크게 보편적 유형 동일론(Universal Type Identity Theory), 종제한적 유형 동일론(Species-specific or Domain-specific Type Identity Theory), 그리고 개별자 동일론(Token Identity Theory) 등의 세가지로 나눌 수 있는데 퍼트남-포도 계열의 기능주의자들이 논박하고 있는 물리주의는 보편적 유형 동일론임에 반해, 암스트롱-루이스 계열의 기능주의자들이 옹호하고 있는 물리주의는 종제한적 유형 동일론내지 개별자 동일론이기 때문이다. 암스트롱은 그들의 물리주의가 보편적 유형 동일론을 주장하고 있지 않으며, 또한 그러한 동일론을 거부하여도 아무런 문제거리를 야기하지 않는다고 최근의 한 논문에서 다음과 같이 분명히 말하고 있다.²⁾

포도의 논증은 물리주의(물질주의)의 핵심에 아무런 영향도 미치지 않는 것처럼 보인다. 그 논증이 정신적 유형이 물리적 유형과 동일시될 수 없다는 것을 보여주나, 각기 정신적 개별자 - 각 개별적 감각, 사고 등등 - 가 순수히 물리적 두뇌 개별자라는 견해에는 의문을 제기하지 않는다. 개별적 통증은 두뇌의 물리적 상태일 수 있는 것이다. ... 심리학의 상대적인 자율성은 물리주의자들의 꿈과 필연적인 갈등관계에 있는 것이 아니다. ... 심리학이 생리학으로 환원될 수 있다는 것을 부정하며 또한 동시에 물리주의를 고수하는 것은 완전히 합리적인 것처럼 보인다. 이러한 입장을 물리주의라는 이름이 외에 무엇이라고 부를 수 있겠는가?

더구나 퍼트남은 자신의 기능주의가 암스트롱-루이스 계열의 기능주의보다 나올 수 있는 한 장점으로 자신의 이론이 정신에 관한 이원론과 양립할 수 있다고 주장하는데,³⁾ 암스트롱은 위의 논문에서 또다시 퍼트남의 이러한 견해는 오해에서 연유한 것이라고 지적하고 있다. 스마트의 이론을 이어받은 자신의 이론도 원리상 정신에 관한 이원론과 얼마든지 양립할 수 있다는 것이다. 암스트롱은 다음과 같이 말한다.⁴⁾

퍼트남은 자신이 주장하고 있는 견해의 한 장점은 물질주의 견해와는 달리 정신에 관한 이원론과 양립 가능하다는 것이라고 말한다. 퍼트남의 이러한 진술은 어떤 잘못된 이해에서 특히 스마트에 대한 잘못된 이해에서 연유한 것처럼 보인다. 기능주의의 선구적 역할을 한 것은 스마트의 감각에 대한 화제-중립적(topic-neutral) 설명이지 그의 감각적 두뇌 현상과의 동일시가 아니다. 더 나아가, 스마트는 그의 화제중립적 설명의 한 장점으로 그것이 이원론과 적어도 논리적으로 양립 가능하다는 것이라고 생각했다.

위에서 보았듯이 양 계열의 기능주의 모두 원리상 물리주의와 같은 존재적 일원론이나 데카르트식의 존재적 이원론 양자 중 어느 것이라도 수용할 수 있는 이론이다. 따라서 기능주의 이론 자체는 일원론/이원론 등에 무차별적이다. 그러나 실제적으로 기능주의자들의 일반적인 경향은 - 양 계열 기능주의자들 모두 - 종제한적 유형 동일론 내지 개별자 동일

주 2) "Recent Work on the Relation of Mind and Brain", in *Contemporary Philosophy, A new survey*, vol. 4 (Martinus Nijhoff Publishers: 1983)

주 3) "Psychological Predicates" in W.H. Capitan and D.D. Merrill(eds), *Art, Mind and Religion* (University of Pittsburgh Press, 1967)

주 4) "Recent Work on the Relation of Mind and Brain"

론은 받아 들이고 있음에 반해, 보편적 유형 동일론이나 이원론에 대해서는 거부 반응을 보이고 있다. 따라서 양 계열의 기능주의 구분 기준으로서 기능주의와 물리주의 간의 포용/배척 관계를 짚는 네드 블록의 견해는 잘못된 것이라 해야 할 것이다.

그 다음 양 계열 구분 기준으로 양 계열에 붙여진 튜어링 기계기능주의와 비튜어링 기계 기능주의라는 이름이 나타내듯 각 계열의 기능주의를 정의함에 있어 튜어링 기계 도표 상태(Turing Machine Table State)라는 개념의 사용 유무를 들 수 있다. 앞 계열은 정신상태의 한 유형은 튜어링 기계 도표에 명기된 튜어링 기계 도표상태의 한 유형과 동일하다고 주장함에 반해, 뒷 계열은 튜어링 기계 개념을 사용한 이러한 협소한 기능주의 정의를 받아들이지 않고 대신 램지 문장(Ramsey Sentence)을 사용하여 보다 포괄적으로 기능주의를 정의한다. 튜어링 기계를 이용한 기능주의 설명은 주로 퍼트남에 의해 수용되었는데 이는 블록과 포도의 "What Psychological States Are Not"⁵⁾이라는 논문에서 결정적인 비판을 받고있고 이 비판은 대체로 받아들여지고 있으므로 튜어링 기계를 이용한 기능주의 설명은 여기서는 생략하기로 하고 램지 문장을 사용한 기능주의 설명만을 살펴보자.

T를 통증, 다른 정신 상태들, 감각적 자극들(sensory inputs), 행동적 반응들(behavioral outputs) 간의 관계를 말해주는 어떤 심리 이론이라고 하자. 이러한 심리 이론 T를 이루고 있는 문장들 속에 나타나는 정신 상태 용어들을 우선 모두 명사어가 포함된 어휘들로 바꾸어 쓴다. 예를 들어 '슬프다'라는 표현은 '슬픔을 갖다' 혹은 '슬픔을 느끼다'와 같이 '슬픔'이라는 명사어가 들어간 표현으로 바꾸어 쓴다. 그다음 이러한 문장들을 연결하여 단일한 연접문으로 만든다. 이렇게 하여 얻어진 이론 T를 아래와 같이 표기하기로 하자.

$$T(s_1, \dots, s_n, o_1, \dots, o_m, i_1, \dots, i_k)$$

여기서 $s_1, \dots, s_n$ 은 정신 상태들을, $o_1, \dots, o_m$ 은 행동적 산출들(반응들)을, $i_1, \dots, i_k$ 는 감각적 유입들(자극들)을 자기 나타내는 용어들이다. 여기서 명사어로 표현된 각 정신상태 용어들을 변수로 바꾸고 맨 앞에 존재양화사를 붙여 아래와 같이 램지 문장을 만든다.

$$(\exists x_1) \dots (\exists x_n) T(x_1 \dots x_n, o_1 \dots o_m, i_1 \dots i_k)$$

이제 만일 x_1 이 통증을 대체한 변수라면 통증을 아래와 같이 정의할 수 있다.

$$\text{통증} = \text{어떤 특정한 존재자 } x_1 (\exists x_2) \dots (\exists x_n) \\ T(x_1 \dots x_n, o_1 \dots o_m, i_1 \dots i_k)$$

이 정의에 따르면 통증은 이론 T가 명기하고 있는 인과적 역할의 담지자이다. 보다 구체적으로 말해, 통증이란 그것 자체가 무엇인지는 알 수 없지만, 이론 T가 명기하고 있는 다른 정신 상태들, 감각적 유입들, 그리고 행동적 산출들과 어떤 인과적 관계를 맺고있는 어

주5) in *Readings in Philosophy of Psychology*, vol.2

(金泳楨)

면 특정한 존재자이다. 예를 들어, T를 “통증이란 핀의 찔림에 의해 야기되고 근심과 신음 소리를 야기하며 근심은 또다시 이맛살의 찌푸림을 야기한다.”는 것과 같은 아주 단순한 이론이라고 가정하자. 그러면 이 이론 T의 램지 문장은

$(\exists x_1) (\exists x_2) (x_1 \text{은 핀의 찔림에 의해 야기되고 } x_2 \text{와 신음소리를 야기한다} \ \& \ x_2 \text{는 이맛살의 찌푸림을 야기한다})$

와 같이 되며 통증은 다시금 아래와 같이 표기될 수 있다.

어떤 특정한 존재자 $x_1 (\exists x_2) (x_1 \text{은 핀의 찔림에 의해 야기되고 } x_2 \text{와 신음소리를 야기한다} \ \& \ x_2 \text{는 이맛살의 찌푸림을 야기한다})$

즉 통증이란 핀의 찔림에 의해 야기되고, 이맛살의 찌푸림을 야기하는 또다른 어떤 정신상태와 신음소리를 야기하는 어떤 특정한 존재자이다.

위에서 우리는 기능주의가 정신상태의 실재를 인정하고 또 이 정신상태가 행동적 반응의 원인이나 감각적 자극의 결과가 될 수 있다는 점을 인정하며 또는 정신상태들 상호간의 인과적 관계를 인정한다는 점을 제외하고 행동주의와 매우 유사함을 알 수 있다. 실제로 후자 계열의 기능주의자들은 자신들이 행동주의자들의 후예임을 자처하고 있다. 결국 후자 계열의 기능주의자들은 그들의 기능주의를 설명함에 있어 정신상태란 튜어링 기계도 표상상태라는 것을 받아들이고 있지 않으므로, 두 계열의 기능주의를 구분함에 있어 튜어링 기계 도표상태라는 개념의 사용 여부가 그 한 기준이 될 수 있다 할 것이다. 그러나 앞 계열에서도 오로지 퍼트남만이 정신상태란 튜어링 기계 도표상태라는 주장을 하고 있고 포도, 하만 등은 기능주의를 설명함에 있어 튜어링 기계 도표상태라는 개념에 의존하고 있지 않으므로 이 구분 기준은 양 계열에 대한 엄밀한 구분기준은 아니라 할 수 있다. 더구나 블록이 앞 계열의 기능주의에 대한 포괄적인 설명을 제시하기 위하여 튜어링 기계 도표상태 개념이 아닌 램지 문장을 통한 설명 방식을 도입하고 있다는 점은 튜어링 기계 도표상태라는 개념의 사용유무가 이 양 계열을 구분하는 기준이 될 수 있다는 주장을 더욱 더 약하게 만든다.

양 계열에 대한 또 다른 구별 기준은 기능주의를 경험적(empirical) 이론으로 간주하는가 선험적(a priori) 이론으로 간주하는가의 여부이다. 앞 계열에 속하는 경험적 기능주의자들은 정신상태에 대한 기능적 분석을 실질적인 과학적 가설로 간주하고 있음에 반해 뒷 계열에 속하는 선험적 기능주의자들은 정신상태에 대한 기능적 분석을 정신적 용어의 의미 분석으로 간주하고 있다.

선험적 기능주의(상식 기능주의)와 경험적 기능주의(심리기능주의)의 구분은 램지 문장을 통한 기능주의 설명에서 잘 나타날 수 있다.

우리는 앞서 통증을 정의하기 위해서는 이 통증과 인과적 관계를 맺고있는 다른 정신상태들, 감각적 자극들, 행동적 반응들을 명기해주는 어떤 심리이론 T가 필요함을 보았다. 이 때 T가 상식에 입각한 일상적 심리 용어들에 대한 선험적 분석을 기반으로 형성된 이론일 경우 선험적 상식기능주의가 되고, T가 경험과학에 입각한 이론일 경우 경험적 심리기능주의가 된다. 따라서 상식 기능주의자들은 감각적 자극과 행동적 반응들을 상식적 지

식의 범위 안에서 명기하고 있음에 반해, 심리기능주의자들은 이러한 제약을 받지 않고 자극과 반응을 뉴론과 같이 과학을 통해 얻어진 내부적 매개물을 통하여 명기할 수 있다. 암스트롱은 두 계열 간에 위와 같은 상이점이 현실적으로 존재하고 있음을 인정하고 있으나 이러한 외견상의 상이점이 기능주의를 두 계열로 가르는 궁극적인 구분 기준은 될 수 없다고 주장한다. 암스트롱은 다음과 같이 말한다.⁶⁾

퍼트남의 기능주의를 진정으로 스마트-루이스-암스트롱 계열의 기능주의와 구별케 하고, 포도의 기능주의와 같은 계열에 속하게 해주는 점이 있는 것 같다. 그것은 퍼트남이 기능주의를 개념적 이론이라기 보다는 심리적 이론으로 간주하고 있다는 점이다. 그러나 여기에서조차도 그 외견상의 상이성은 대부분 해소될 수 있는 것 같다. 우리는 스마트, 루이스 그리고 암스트롱이 정신 상태에 관한 첫 단계의 예비적인 정지(땅고르기)작업에 대한 설명을 제공하고 있는 것으로 생각할 수 있다. 우리가 개념적 분석이라는 개념을 인정하든 안하든 우리가 심리상태에 관한 이야기를 이해할 수 있기 위해서는 그러한 첫 단계의 정지 작업이 꼭 있어야만 하는 것처럼 보인다. 이러한 예비적 이해가 바로 기능적(인 과정) 측면에 대한 일상적인 이해라고 스마트-루이스-암스트롱 계열의 기능주의자들은 주장한다. 이제 이러한 이론이 비록 기능적인 차원에 계속 머무름에도 불구하고, 이러한 첫 단계의 기능적 규정이 심리학 이론의 진보에 따라 매우 발전되고, 세련되고 혹은 수정되기조차 하여야만 한다는 것을 쉽게 상상할 수 있다.

암스트롱의 견해가 옳다면 상식적 기능주의는 심리기능주의를 위한 예비적 단계이지 결코 양자가 배척적 관계에 있는 것이 아니므로, 심리 이론 T가 상식적 이론인가 과학적 이론인가의 여부를 근거로 형이상학적 기능주의에 근본적으로 서로 다른 두 계열이 있다고 주장하는 것은 합당하지 못한 것 같다.

결국 필자의 견해로는 형이상학적 기능주의를 두 가지 계열로 가르는 것은 별 의미없는 작업인듯 싶다. 비록 튜어링 기계 기능주의와 비 튜어링 기계 기능주의의 구분은 어느 정도 타당성을 가지고 있으나, 튜어링 기계 기능주의가 앞 계열을 모두 포괄할 수 있는 포괄적인 정의가 아니라, 퍼트남의 기능주의만을 특징지을 수 있는 기능주의에 대한 매우 협소한 정의에 불과하므로 엄밀한 구분이라 할 수 없다. 더구나 앞서 지적한대로 튜어링 기계 기능주의는 블록과 포도에 의해 결정적으로 비판받고 있으므로 우리가 그 건전성을 따져 보아야 할 형이상학적 기능주의는 오로지 한 종류의 기능주의 밖에 없다. 필자도 튜어링 기계 기능주의에 대한 블록과 포도의 비판이 합당하다고 생각하므로 이 논문에서는 생략하기로 하고 형이상학적 기능주의를 포괄적으로 정의하고 있는 비튜어링 기계 기능주의에 대한 비판만을 고찰하여 보겠다.

III

램지 문장을 통해 포괄적으로 정의된 기능주의에 대한 비판은 감각질(qualia)과 관련된

주 6) "Recent Work on the Relation of Mind and Brain"

것으로, ‘전도된 감각질 논증(Inverted Qualia Argument)’과 ‘결여된 감각질 논증(Absent Qualia Argument)’ 두 가지가 있는데 이 둘은 서로 엇물려 돌아가고 있다. 이 비판도 역시 “What Psychological States Are Not”이라는 논문에서 블록과 포도에 의해 처음으로 제기되었다. 전도된 감각질 논증은 우리에게 익히 알려진 전도된 스펙트럼 가설에 주의를 기울임으로써 쉽게 이해될 수 있다. 그 전도된 스펙트럼 가설에 따르면, A라는 사람과 B라는 사람이 비록 완전히 동일한 기능적 구성을 가지고 있다 할지라도, A라는 사람이 붉은 물체를 볼 때 갖는 감각이 현상적으로 B라는 사람이 녹색의 물체를 볼 때 갖는 감각과 동일할 수 있다는 것이다. 만일 이 가설이 옳다면 A의 정신상태와 B의 정신상태가 감각질적으로(qualitatively) 혹은 현상적으로(phenomenally) 비록 다르다 할지라도 이 두 정신상태들은 기능적으로 동일할 수 있다는 것이다. 만일 이 가설이 옳다면 A의 정신상태와 B의 정신상태가 감각질적으로(qualitatively) 혹은 현상적으로(phenomenally) 비록 다르다 할지라도 이 두 정신상태들은 기능적으로 동일할 수 있다. 따라서 정신상태에 대한 기능적 규정은 이 상태의 고유한 감각질적(qualitative) 특성을 반영하지 못할 수 있다. 그러나 정신상태의 감각질적 특성은 정신상태에 있어서 빠뜨릴 수 없는 본질적 특성이다. 따라서 기능주의는 그르다.

결여된 감각질 논증도 이와 유사하게 전개된다. 단지 다른 점은 A의 정신상태는 감각질적 특성(qualitative character)을 가지고 있음에 반해 B의 정신상태는 어떠한 감각질적 특성을 가지고 있지 않음에도 불구하고, A와 B의 정신 상태가 기능적으로 동일할 수 있다는 가설에서 출발하는 것이다.

블록과 포도의 위와 같은 기능주의에 대한 반박에 대해 슈우메이커는 “Functionalism and Qualia”⁷⁾라는 논문에서 재반박을 가한다. 슈우메이커는 기능주의는 전도된 감각질의 가능성과 잘 화합할 수 있으며, 결여된 감각질의 가능성은 논리적으로 불가능하다고 주장한다. 슈우메이커의 결여된 감각질의 가능성에 대한 반박은 다음과 같다. 만일 결여된 감각질이 가능하다면, 통증의 감각질적 특성의 존재 여부는 그것의 인과적 결과에 아무런 영향을 미치지 못할 것이다. 그리고 지식의 인과설에 따르면, 만일 감각질적 특성의 존재 여부가 통증의 인과적 결과적 아무런 영향을 미치지 못한다면, 우리는 통증의 감각질적 특성에 관한 아무런 지식도 갖지 못할 것이다. 그러나 우리는 통증의 감각질적 특성에 관한 지식을 갖고 있다. 그러므로 결여된 감각질은 가능하지 않다. 이 논증을 좀 더 자세히 분석하면,

1. 만일 결여된 감각질이 가능하다면, 통증의 감각질적 특성의 존재 여부는 통증의 인과적 결과에 아무런 영향도 미치지 못한다. (결여된 감각질 전제)
- (2a) 만일 감각질적 특성의 존재 여부가 통증의 인과적 결과에 아무런 영향도 미치지 못한다면, 우리는 통증의 감각질적 특성에 관한 아무런 지식도 가지지 못할 것이다. (지식의 인과설 전제)
- (2b) 우리는 통증의 감각질적 특성에 관한 지식을 갖고 있다.

주 7) in *Readings in Philosophy of Psychology*, vol. 2

2. (2c) 감각질적 특성의 존재 여부가 통증의 인과적 결과에 영향을 미친다.
[2a와 2b로 부터 MTT에 의해]

3. 결여된 감각질은 가능하지 않다. [1 과 2로부터 MTT에 의해]

우리는 쉽게 이 논증이 타당한 논증임을 알 수 있다. 그러나 이 논증의 전진성은 쉽게 파악되지 않는다. 왜냐하면 전제 1의 진리성이 확실하지 않기 때문이다. 실제로 블록은 "Are Absent Qualia Impossible"⁸⁾이라는 논문에서 전제 1의 진리성에 의심을 표하고 있다.

결여된 감각질은 가능하지 않다는 슈우메이커의 주장에 대한 블록의 논박의 핵심은 '대체적 통증(ersatz pain)'이란 개념이다. 대체적 통증은 감각질적 특성을 결여하고 있으니 기능적으로 진정한 통증과 동일한 통증으로, 이 때 대체적 통증에 의해 어떤 방식으로 야기된 결과와 똑 같은 결과를 야기하는데 진정한 통증의 감각질적 특성이 핵심적인 역할을 하여도 상관 없다는 점이 중요하다. 따라서 대체적 통증이 가능한 한, 비록 결여된 감각질이 가능하더라도 통증의 감각질적 특성의 존재 여부가 통증의 인과적 결과에 아무런 영향도 미치지 못한다는 결론은 나오지 않으므로 슈우메이커가 참이라고 주장하는 전제 1은 거짓이 된다.

블록은 대체적 통증의 가능성을 보이기 위해 난장이에 의해 작동되는 머리체제(homunculus-headed system)라는 가상적 체제를 제시한다. 이 난장이 머리 체제는 튜링 기계를 의인화 시켜 설명한 것으로 그 체제는 다음과 같이 기술될 수 있다.

외부적으로는 인간의 육체와 동일하고 내부적으로는 매우 다른 한 신체를 상상해 보자. 감각 기관으로부터 나온 뉴론들은 머리속에 있는 움푹 파진 동굴 속에 있는 일련의 불빛들과 연결되어 있다. 반응을 유발하는 뉴론들은 동굴의 다른 부분에 있는 계기판의 단추에 의해 작동된다. 동굴의 한 벽에는 A라는 어떤 인간을 묘사하는 매우 큰 기계 도표(machine table)가 걸려 있고 동굴의 한 구석에는 상태를 기록하는 칠판이 있다. 또 그 동굴에는 기능을 수행하는 난장이가 살고 있다. 우리는 이 난장이에게 A라는 사람의 현재 정신 상태인 S_{197} 이라는 도표 상태에서 출발하라고 말한다. 그 난장이가 수행해야 할 구실은 다음과 같다. 그는 A의 상태 번호인 "197"을 칠판에다가 쓴다. 그리고 현재 어떠한 유입(자극)이 주어지는지 알아 보기 위해 일련의 유입 불빛들을 관찰한다. 그는 그 불빛의 유형이 유입 I_{34} 를 나타내고 있음을 알아낸다. 그는 벽에 걸린 기계 도표에서 34 번째 종열과 197 번째 횡열이 만나는 칸을 본다. 이 칸에는 상태를 S_{765} 로 바꾸고 산출 O_{1098} 을 만들어 내라고 쓰여 있다. 난장이는 칠판에서 197을 지우고 765를 쓴다. 그리고 그는 산출 계기판에 가서 반응 O_{1908} 을 산출할 수 있는 단추를 누른다. 그리고 그는 다음 유입이 무엇인지 알기 위해 일련의 유입 불빛들을 다시 관찰한다. 그다음 똑 같은 작업을 반복한다.

주 8) *Philosophical Review*, 1980, 89(2)

난장이의 이러한 작업을 통해 이 난장이가 거주하고 있는 인공적 신체는 만일 일련의 유입들이 주어진다면 A라는 인간이 할 행동과 똑같은 반응을 보이게 될 것이다. 실제로 그 난장이가 따르고 있는 기계 도표는 바로 A라는 사람을 묘사하고 있으므로 이 난장이 머리체계가 그 사람과 기능적으로 동일하리라는 것은 쉽게 알 수 있다. 이제 그 사람이 심한 통증이 생겨 아스피린을 요구한다고 가정하자. 이 난장이 머리체계도 물론 그 사람과 동일한 기능적 상태에 있으며 아스피린을 요구하는 유사한 행동적 반응을 보일 것이다. 그러나 그 단추를 누르는 난장이는 어떠한 통증을 느낄 필요가 없으며, 따라서 이 난장이 머리 체계가 통증을 느끼고 있다고 생각해야 할 아무런 이유가 없다고 블록은 주장한다. 결국 비록 난장이 머리체계가 통증이라는 감각질을 갖지 않고도 통증이라는 감각질을 가진 A라는 인간의 기능과 동일한 기능을 할 수 있다고 해서, 우리는 통증이라는 감각질의 존재 여부가 A라는 인간의 기능적 결과에 아무런 영향도 미칠 수 없다고 말할 수 없다는 것이다.

블록은 또다시 간단한 비유로서 쥐뿔을 예로 든다. 용수철을 사용한 쥐뿔과 용수철을 사용하지 않은 쥐뿔은 기능적으로 동일할 수 있다. 그러나 용수철의 유무가 용수철을 사용한 쥐뿔 상태의 인과적 결과에 아무런 영향을 미치지 않는다고 결론짓는 것은 어리석은 것이다. 틀림없이 슈우메이커는 대체적 통증의 가능성을 고려하지 못했다. 그리고 블록이 주장하는대로 대체적 통증이 가능하다면 슈우메이커의 결여된 감각질의 가능성에 대한 비판은 좌초되는 것이다. 그러나 필자는 난장이 머리체계에서와 같은 대체적 통증이 진정으로 가능할 수 있는가에 대해 의구심을 떨쳐버릴 수가 없다. 또한 진정으로 그러한 대체적 통증이 가능하다면 왜 우리는 그 대체적 통증이 감각질적 특성을 가지고 있다고 말하는 것을 주저하여야만 할까?

쥐뿔의 예를 생각해 보자. 용수철을 사용한 쥐뿔과 용수철을 사용하지 않은 쥐뿔이 기능적으로 완전히 동일할 수 있을까? 물론 쥐를 잡는 데 사용된다는 점에서만은 기능적으로 동일할 수 있겠지만, 쥐를 잡는 방식이 꼭 같을 수는 없을 것이다. 어떻게 과연 모든 인과적(기능적) 측면에서 완전히 동일한 대체적 용수철 쥐뿔이 가능할까? 또한 만일 모든 인과적(기능적) 측면에서 용수철을 사용한 쥐뿔과 용수철을 사용하지 않은 쥐뿔이 완전히 동일할 수 있다면, 용수철을 사용한 쥐뿔에서 용수철이 하는(모든) 기능에 대응하는 똑같은 기능을 하는, 용수철을 사용하지 않은 쥐뿔에 속한 어떤 특정한 존재자 즉 대체적 존재자(ersatz entity)를 왜 우리는 용수철이라 부를 수 없을까? 퍼트남도 그의 저서 *Reason, Truth, and History*⁹⁾의 제4장 *Mind and Body*에서 유사한 논증을 펴고 있다.

전도된 감각질 논증에 대한 슈우메이커의 반박의 핵심은 다음과 같다. 기능주의자들은 전도된 감각질 논증의 결과 개별적 감각질 상태가 어떠한 기능상태와 동일할 수 없다는 것을 인정하여야만 한다. 그러나 이러한 인정이 기능주의자들로 하여금 감각질적 특성을 갖

주 9) Cambridge University Press, 1981.

고 있는 통증이나 붉게 보임, 혹은 다른 정신상태들의 집합을 정의하지 못하게 하는 것이 아니라고 주장한다. 결국 슈우메이커는 통증을 (a) 어떤 감각질적 특성을 갖고 있고 (b) 다른 정신 상태들, 유입들, 그리고 산출들과 어떠한 인과적 관계를 갖고 있는 상태로 정의한다. 즉 통증과 같은 정신 상태들을 경우에 따라서 다른 감각질적 특성을 가질 수 있는 상태로 간주함으로써 전도된 감각질 문제를 해결한다. 이러한 접근 방식은 감각질 상태를 포함한 모든 정신 상태를 존재 양화함으로써 구체적인 개별적 감각질에 대한 언급을 피하는 램지 문장을 이용한 기능주의 정의 방식과도 일치한다. 그러나 이러한 슈우메이커의 정의 방식에는 어떻게 ‘정신상태가 어떤 감각질적 특성을 갖고 있음’을 기능적 용어들로 정의할 수 있는가 하는 문제를 낳는다. 여기서 슈우메이커는 ‘감각질적 유사성’이라는 개념을 도입하여, 문제해결을 시도한다. 그에 따르면, 감각질적 유사성이라는 개념은 기능적 용어들로 정의가 가능하며, 또 감각질적 유사성 개념을 통해 ‘정신상태가 어떤 감각질적 특성을 가지고 있음’을 정의할 수 있다. 슈우메이커는 정신 상태가 다른 정신 상태와 감각질적으로 유사하거나 상이할 때, 그 정신 상태가 어떤 감각질적 특성을 가지고 있다고 정의한다. 따라서 남은 문제는 어떻게 감각질적 유사성을 기능적 용어들로 정의하느냐 하는 문제이다. 슈우메이커의 문장을 직접 인용해 보자.¹⁰⁾

프레게-술릭 견해는 행동주의와 기능주의에 대한 전도된 감각질 논박에 대해 대답을 제시해 준다. 감각질적 유사성이 내주관적(intrasubjective) 경우에만 잘 정의될 수 있다고 주장하는 부분적인 동기는 이 감각질적 유사성 관계를 감지하고 있다는 직접적인 행동적 증거를 내주관적 경우에만 우리가 가질 수 있기 때문이다. 이것이 바로 간주관적(intersubjective) 스펙트럼 견도는 행동으로 알아낼 수 없지만 내주관적 전도는 알아낼 수 있는 까닭이다. (프레게-술릭 견해가 주장하듯) 만일 내주관적 감각질적 유사성이 존재하는 유일한 종류의 감각질적 유사성이고 이 내주관적 감각질적 유사성이 행동적으로 감지될 수 있다면, 이 내주관적 감각질적 유사성은 행동적으로 정의 가능하다. 동일한 고려를 통해 제시된 보다 유망한 견해는 이 감각질적 유사성 관계가 기능적으로 정의 가능하다는 것이다. 기능적 설명의 핵심은 이 감각질적 유사성 관계가 한 사람의 서로 다른 시각 경험에 적용될 때, 이 유사성 관계가 그 사람에게 그가 보고 있는 것들에 객관적인 유사성이 있다는 믿음을 유발하는 경향-다시말해, 색깔 감각질의 유사성이 보여진 색깔들의 유사성에 대한 믿음을 유발하는 경향을 갖고 있다는 것이다. 그리고 그의 경험들 사이의 감각질적 유사성이 그 사람의 믿음에 영향을 줌으로 해서, 이 유사성은 다시 그의 행동, 특히 인식적 그리고 식별적 행위에 영향을 미칠 것이다.……만일 프레게-술릭 견해가 주장하는대로, 이 유사성 관계가 오로지 내주관적으로만 성립할 수 있다면, 유사성 관계가 그러한 인과적 역할을 통해 기능적으로 정의되지 못할 이유가 없는 것처럼 보인다.

주 10) “The Inverted Spectrum”, 1982 *Journal of Philosophy*

(金泳禎)

슈우메이커가 지적한대로 내주관적 감각질적 유사성이 가능한 유일한 감각질적 유사성이고 이 감각질적 유사성은 행동적으로 감지될 수 있다면, 이 감각질적 유사성이 기능적으로 정의 가능한 것은 자명한 사실인듯 싶다. 또한 감각질적 유사성이 기능적으로 정의 가능하다면, 슈우메이커의 문제거리는 모두 해결된 셈이다. 결국 슈우메이커에 따르면, 정신 상태란 다른 정신 상태들, 유입들 그리고 산출들과 어떠한 인과적 관계를 맺고 있는 상태로 이 인과적 관계 속에는 정신 상태가 어떤 감각질적 특성을 갖고 있다는 것을 반영하는 인과적 관계가 포함되어 있어야 한다. 그런데 정신상태가 어떤 감각질적 특성을 갖고 있다는 것은 항상 인과적으로 정의 가능하므로 도치된 감각질 반박은 기능주의에 아무런 결정타를 가하지 못한다.