

쌍둥이 지구로 간 늪지 인간

- 의미의 규범성과 자연주의 -

신상규

개요

표상, 의미, 혹은 지향성의 개념을 자연화하기 위한 시도 중의 하나인 목적론적 이론의 난점을 지적하기 위한 논변으로 “늪지 인간(*swampman*)” 혹은 “우연적 복제물(*accidental double*)” 논변이란 것이 있다. 늪지인간은 우연적인 사건에 의해 생겨난 어떤 사람의 완전한 물리적 복제자이다. 목적론적 의미론에 따르면 우리는 이 복제인간(늪지인간)에게 어떠한 지향적 상태도 귀속시킬 수 없다. 많은 철학자들은 이러한 결론이 대단히 반직관적이라고 생각한다. 이 논문에서는 외재주의와 내재주의의 구분을 통하여 목적론적 의미론의 핵심적인 주장이 무엇인지를 조명하고, ‘지향적 상태’가 일종의 규범적인 현상이라는 것을 전제할 경우, 늪지인간에 대한 목적론적인 결론은 지향성의 규범적 성격에 대한 일종의 최상의 설명으로의 추리(*an inference to the best explanation*)라는 선형적 논변의 형태로 지지될 수 있음을 주장한다.

I

데이빗슨에서 예를 따온 “늪지인간”의 논변은 다음과 같이 진행된다.¹ 어느 날 갈대가 우거진 늪지의 나무 위에 벼락이 떨어졌다. 엄청난 전류가 흐르면서 그 옆에 서있던 나의 몸은 분자단위의 구성요소로 분해되고, 나무의 분자들이 절묘하게 결합 함으로서 물리적, 기능적인 면에서 나와 완전히 똑 같은 늪지인간이 생겨났다. 이 늪지 인간은 나와 수적으로(*numerically*) 구분될 뿐만 아니라, 그 구성 분자들도 사실은 수적으로 다른 요소들로 이루어져 있다. 하지만, 현재의(*current*) 인과적, 성향적 속성의 차원에서 이 늪지인간은 나와 별반 차이가 없어보인다. 사람들을 만나고, 이야기하고, 심지어 사랑도 하는 것처럼 보인다. 말하자면 이 늪지 인간은 시스템의 내부적인 물리적 구성과 그 상태, 그리고 관찰 가능한 외향적인 움직임에 있어서 나와 완전히 일치한다.

전통적인 기능주의 혹은 행동주의적 분석에 따르면 이 늪지 인간은 나와 완전히 동일한 심리적 상태를 가지게 될 것이다. 전통적인 기능주의에서는, 비록 심리적 상태가 두뇌의 상태와 동일시되지는

¹ Donald Davidson, “Knowing One’s Mind”(1987), p.19 in Davidson, *Subjective, Intersubjective, Objective* (2001)

않는다 하더라도, 기능적 상태로서의 심리적 상태는 여전히 우리의 신체나 두뇌의 상태에 수반하는 것이다. 그러므로 이 늑지 인간이 나와 완전히 똑 같은 신체적, 두뇌적인 상태를 가지고 있다면, 이것이 가지고 있는 기능적 상태 또한 나와 완전히 일치할 것이다. 어떤 대상에 심리적 혹은 지향적 상태를 귀속시키는 원리로서의 튜링 테스트를 생각하여도 동일한 결론이 도출된다. 늑지 인간은 튜링 테스트를 완벽히 통과할 것이다. 그렇다면 우리는 이 늑지 인간이 지향적인 태도로서의 심리적인 상태를 가진다고 말할 수 있는가? 가령 이 늑지 인간에게 어떤 믿음, 혹은 욕구를 귀속시킬 수 있는가? 당신의 직관은 어떠한가?²

우리가 이 논문에서 다루고자 하는 목적론적 의미론자들은 이 늑지인간에게 지향적 태도를 귀속시킬 수 없다고 주장한다. 목적론적 의미론에 따르면 믿음이나 욕구는 고유 기능(*proper function*)의 범주에 속하는 것이며, 고유 기능적 범주는 그 구성적 본질에 있어서 역사적인 것이다. 그런데, 이 늑지 인간은 어느 순간 발생했던 자연적 우연의 산물일 뿐이며, 보통의 우리네 인간이 가지고 있는 탄생등과 관련한 어떠한 적절한 역사도 가지고 있지 않다. 그러므로 이 늑지 인간이 비록 물리적 상태나 관찰 가능한 행동적 성향으로 보아서 나와 전혀 구분이 되지 않더라도 믿음이나 욕구와 같은 지향적 상태를 귀속시킬 수는 없다. 많은 철학자들은 늑지인간에게 아무런 지향적 상태도 귀속시킬 수 없다는 목적론적 의미론의 주장이 대단히 반 직관적인 것이라고 비판한다. 물론 고유기능을 통한 지향적 태도의 분석이라는 목적론적 의미론의 기본적인 주장을 유지하면서도 늑지 인간에 어떤 지향적 태도를 귀속시키려는 시도가 있을 수 있다. 그러나 대체적으로 이러한 시도는 고유기능에 대한 역사적 이해를 포기할 것을 요구한다. 하지만 밀리칸이나 니앤더(*Neander*) 같은 목적론자들은 지향적 규범성은 고유기능에 대한 역사적 이해를 통해서만 확보될 수 있다고 믿는다. 그래서 이들은 일종의 양립논제를 추구하기 보다는 늑지 인간이 아무런 지향적 상태를 갖는 않는다는 반직관적인 결론을 기꺼이 받아들인다.

이 논문의 기본적인 목표는 늑지 인간에 대한 목적론자들의 이러한 주장이 처음에 들리는 것처럼 그렇게 반직관적이기만 한 것은 아니라는 점을 보이는 데에 있다. 밀리칸(*Millikan*)같은 이는 일단 지향성이나 고유기능에 대한 자신의 이론이 믿음이나 욕구와 관련한 우리들의 일상적인 언어 사용에 대한 필요 충분 조건을 탐색하는 개념적 분석이 아니라, 일종의 이론적 정의(*theoretical*

² 어떤 측면에서 늑지 인간 논변은 서얼의 중국어 방 논변에 비교될 수 있다. 서얼의 중국어 방 논변이 보여주고자 했던 것은 구문(*syntax*)과 의미(*meaning*)을 구분하고, 구문만으로는 의미를 귀속시킬 수 없다는 것이었다. 서얼은 이러한 질문을 통하여 무엇이 의미를 구성하는 것인가에 대한 질문을 던진다. 지향적 태도란 기본적으로 의미 혹은 내용(*content*)을 갖는 상태이다. 늑지인간 논변은 서얼적인 의미에서의 구문을 포함한 더욱 포괄적인 (내적인) 상태를 늑지인간이 갖는다고 인정한 후에, 늑지인간에게 의미 혹은 내용을 갖는 지향적 태도를 귀속시킬 수 있는가 하고 묻는 것이다. 덧붙이자면, 이 논문에서 우리의 관심은 지향성 혹은 지향적 태도에 관한 해명이다. 그러므로 앞으로의 논의는 심리적 상태 일반이 아니라 지향적 태도로서의 지향적 상태에 국한하여 진행될 것이다.

definition)임을 분명히 한다. 그러므로 자신의 이론에 대한 궁극적인 옹호는 우리들의 일상적인 언어적 사용에 얼마나 부합하느냐에 의해서가 아니라, 이러한 개념을 통하여 구성되는 이론들이 얼마만큼 다양한 현상들을 잘 설명해주느냐에 따른 이론의 설명력에 의해서라고 주장한다.

우리의 일상적인 직관과 어떤 이론의 귀결이 불일치하는 것은 철학이나 과학에서 매우 흔한 사례이다. 멀리 갈 것도 없이 지구가 태양의 주위를 돌고 있다는 것은 얼마나 반직관적인 주장인가? 많은 경우에 있어서 정교하고 설명력이 있는 이론과 우리의 일상적인 직관이 충돌할 때에, 포기해야 할 쪽은 결국 우리의 일상적인 직관이었다. 어떤 좋은 이론의 뒷받침이 없는 직관은 말 그대로 직관일 뿐이다. 우리의 직관은 흔히 어떤 것의 겉으로 드러난 징표에 의존한다. 그러나 어떤 것이 모종의 성질 자체를 갖는다는 것과, 그 성질을 갖는 가를 판단하기 위해 사용하는 여러 가지 징표를 보여주고 있다는 것은 엄연히 다른 문제이다. 금처럼 보이는 모든 것이 금은 아닌 것이다. 목적론자들이 생각할 때, 늑지 인간의 예가 보여주고 있는 것은 단지 늑지 인간이 지향적 상태에 관련된 여러 가지 징표를 가지고 있다는 것뿐이다.

여기서 우리는 늑지인간이 제기하고 있는 도전을, 지향적 상태를 지향적 상태로 만드는 것이 도대체 무엇인가를 묻는, 일종의 구성적(constitutive) 원리에 대한 질문으로 이해할 수 있다. 가령, 목적론자는 튜링 테스트가 지향성에 대한 구성적 원리라는 것을 부정할 것이다. 비록 우리가 많은 경우에 있어서, 겉으로 드러난 행동을 통한 일종의 튜링테스트를 통하여 지향적 상태를 귀속시키기는 하지만, 그러한 행동적 특성이 곧 지향적 태도 자체를 구성하는 것은 아니라고 주장할 수 있기 때문이다. 그렇다면, ‘늑지인간’의 논변이 목적론적 이론의 부당성에 대한 적절한 비판으로 인정 받기 위해서는, 지향적 상태를 구성하는 것이 무엇인가에 대한 좋은 이론적 뒷받침이 있어야 할 것처럼 보인다. 그러한 이론의 뒷받침이 없는 직관은 말 그대로의 직관일 뿐이며, 비판적 논점으로서의 설득력은 그만큼 떨어질 것이기 때문이다.

이 논문은 늑지인간과 관련하여 서로 상충하는 두 가지 직관 혹은 결론들이 전제하고 있는 이론적 프레임웍(framework)으로서 내재주의(internalism)와 외재주의(externalism)의 구분을 도입하고, 표상이나 지향적 태도가 본질적으로 규범적 현상이라는 것을 전제한다면, 이러한 규범성을 해명하는 데에 있어서 외재주의가 훨씬 유리한 입장에 있음을 주장하려는 것이다. 즉, 표상적 상태 혹은 지향적 내용이 그 본질에 있어서 규범적 현상이라는 것을 기본적인 전제로 가정해보자. 그 다음에 우리는 이러한 규범적 성격을 해명하기 위하여 어떠한 조건들이 지향적 상태의 구성적인 원리로서 포함되어야 하는가를 물을 수 있다. 이러한 물음에 대하여, 늑지 인간에게 지향적 태도를 귀속시키려는 입장이 가져야 할 것처럼 보이는 내재주의적인 입장보다는, 공간적인 외부 환경 뿐만 아니라 시간적인 차원의 역사적 사실 까지도 매우 중요한 구성적인 조건으로 인정하는 매우 강력한 형태의 외재주의

이론인 목적론적 의미론이 보다 설득력있는 설명을 제공하는 것처럼 보인다는 것이다. 즉, 지향성의 규범적 성격에 대하여, 외재주의의 한 형태인 목적론적 의미론이 가장 설득력 있는 자연주의적인 설명의 하나를 제공한다는 것이다. 말하자면, 최상의 설명으로의 추리(an inference to the best explanation)라는 선택적 논변의 형태를 빌어서, 내재주의에 대한 외재주의의 우월성을 지지하고자 한다.

이하에 전개될 논문의 기본적인 골격은 다음과 같다. II 절에서는 지향적 태도나 의미가 가지고 있는 규범성의 성격이 무엇이며, 이에 대한 목적론적 의미론의 기본적인 전략을 살펴 볼 것이다. III 에서는 내재주의와 외재주의의 기본적인 주장이 무엇이며, 목적론적 의미론이 깔고 있는 외재주의적인 성격을 살펴볼 것이다. 마지막으로 IV 절에서는 외재주의로서의 목적론적인 의미론을 옹호하게 하는 몇 가지의 사유 실험을 해볼 것이다.

II

목적론적 의미론의 가장 중요한 화두는 표상이나 지향적 의미 현상이 갖는 “규범성”의 해명이라고 할 수 있다. 여기서 의도된 규범성은 *Wittgenstein on Rules and Private Language* 에서 크립키가 가상의 회의론자를 등장시켜 논의한바 있다. 크립키에 따르면, 의미에 대한 이론이 해명해야만 하는 두 가지 질문이 있다.

첫째로 회의론자는 내가 퀴스(quus)가 아니라 플러스(plus)를 의미했다는, 그의 회의론적 도전에 응답할 수 있는 어떤 사실이 있는가를 묻고 있다. 둘째로, 그는 [‘68+57’의 질문에 대하여] 지금 내가 ‘5’가 아니라 ‘125’라고 대답해야만(should answer)한다고 자신 있어 하는 어떤 이유가 있는지를 묻고 있다. ... 회의론자에 대한 대답은 다음의 두 조건을 만족시켜야 한다. 첫째로, 그 대답은 내가 퀴스가 아니라 플러스를 의미한다는 것을 구성하는 (내 마음의 상태에 관한) 사실이 무엇인지에 대한 설명을 제공해야만 한다. 더 나아가 그러한 사실이 무엇이 되던 간에 그 후보가 되는 것들이 만족시켜야만 하는 또 하나의 조건이 있다. 그것은 내가 ‘68+57’에 대하여 ‘125’라는 답을 줌에 있어서 어떻게 정당화될 수 있는 지를 보여주어야만 한다.³

³ Kripke, *Wittgenstein on Rules and Private Language*, p.11

크립키의 두 가지 요구가 정확히 무엇을 의미하는 것인지에 대해서는 많은 논란이 있을 수 있다. 이 논문에서는 이 두 요구를 기술적 문제(descriptive problem)와 규범적 문제(normative problem)로 각각 이름 짓고, 대략 다음과 같이 이해하도록 하겠다. 기술적 문제는 어떤 의미현상에 대한 여러 경쟁적인 해석들 중에서 어떠한 해석이 표준적인 해석인지를 집어내는 원리적인 절차에 관한 질문이다. 즉, 심적 표상의 발화가 갖게 되는 심적내용(mental content)을 결정하는 사실이 무엇인지를 적시함으로써, 비표준적인 해석들에 대하여 우리가 표준적인 해석을 선호할 이유를 제시하는 것이다. 그런데 위와 같이 결정된 심적내용은 단순히 기술적인 차원에 머무르지 않고 일종의 올바른의 표준으로서의 규범적 지위를 누리게 된다. 즉, 이렇게 결정된 지향적 의미는 단순히 사실상 그러그러하게 따르고 있다는 것이 아니라, 마땅히 내가 그렇게 따라야만 하는(ought to, supposed to) 것이라는 규칙으로서의 지위를 누리게 된다는 것이다. 최소한적인 뜻에서, 우리가 어떤 표현을 사용할 때에 단순히 부합하는 것이 아니라 마땅히 따라야만 하는 어떤 옳음 내지 올바른의 표준이 있다는 주장정도로 의미의 규범성을 이해할 수 있을 것이다. 규범적 문제는 심적내용, 혹은 지향적 의미가 가지고 있는 이러한 규범적 성격을 해명하는 문제이다.

목적론적인 의미론은 심적 내용의 분석에 있어서 생물학적인 기능 개념을 끌어 들임으로서 이 두 가지 질문에 한꺼번에 답하려고 한다. 물론 목적론적인 의미론은 기술적 문제에 대한 대답에 있어서도 대단히 신선한 시도로 평가되고 있지만, 본 논문에서는 지향적 의미의 규범성이 어떻게 해명되고 있는가에 초점을 맞추도록 하겠다. 의미의 규범성에 대한 크립키의 요구는, 단순히 규칙과 우연히 일치하는 성향을 갖는 것과, 옳고 그름(맞음이나 틀림)이라는 규범적인(prescriptive) 술어의 적용을 받는 고유한 의미에서의 규칙을 따르는 것 사이의 구분을 어떻게 근거지울 수 있는가의 문제로 이해될 수 있다. 규칙 따르기는 언제나 규칙을 따르기에 실패할 수 있다는 것을 전제로 한다. 크립키는, 사적 언어의 경우 규칙을 따른다고 생각하는 것과 실제로 규칙을 따르고 있는 것 사이의 구분이 있을 수 없기에 오류의 가능성이 원천 봉쇄된다는 비트겐슈타인의 논변을 통하여, 의미에 대한 성향적 이론(dispositional theory of meaning)에 자세한 비판을 가하고 있다. 보고시안은 크립키의 이런 논의를 다음과 같이 정리한다.

크립키는, 어떤 표현의 외연을 정확히 포착해내는 적절하게 선택된 성향이 설령 있다고 하더라도, 그러한 성향은 의미에 대한 사실과는 여전히 동일시 될 수 없다고 생각하는 것 같다. 성향은 기술적인 개념인 반면에 의미가 기술적인 개념이 아니라는 것은 여전히 변함없이 참으로 남아있기 때문이다. 다시말해서, 크립키에 따르면, 의미술어와

동연적인(covariied with) 성향술어가 있다고 하더라도, 각각에 관련된 사실은 서로 다른 종류의 사실이기 때문에 서로 동일시 될 수 없다는 것이다.⁴

크립키의 이러한 요구가 얼마나 적절한 것인지를 여부에 대한 논의를 일단 미뤄둔다면, 목적론적 의미이론은 크립키가 요구하는 이런 구분에 답하기 위한 매우 유망한 자연주의적 전략을 제시한다. 자연주의란 말은 대단히 폭 넓고 다양한 함의를 가질 수 있는 말이다. 하지만, 현재의 맥락에 국한해서 말한다면, 지향성의 자연주의란 표상, 믿음, 욕구와 같은 지향적 현상들이 세계를 구성하는 원초적이고 기본적인 성질 혹은 관계가 아니라, 비지향적 혹은 비의미론적인 모종의 자연적(경험과학적) 성질 혹은 관계를 통하여 구성된다는 입장이다. 포더(Fodor)가 말하는데로, 지향성이 실재하는 것이라면 이는 실제로 그 본질상 비지향적인 다른 그 무엇이어서만 한다는, 일종의 환원주의적인 논제로 이해할 수 있다.⁵ 목적론적인 의미이론의 대표주자라고 할 수 있는 밀리칸(Millikan)의 기본적인 전략은, 의미의 규범성을 우리가 자연주의적으로 큰 문제가 없다고 간주할 수 있는 다른 종류의 규범성으로 환원하는 것이다. 이때 환원의 기초가 되는 규범은 진화생물학에 이야기되는 생물학적인 고유 기능의 규범이다. 물론 여기서 말해지는 환원은 규범적인 현상을 성향과 같은 어떤 기술적인 속성으로 환원하는 것이 아니라, 한 종류의 규범을 또 다른 종류의 규범으로 환원하는 것이다. 밀리칸에 따르면 생물학적인 고유 기능의 개념은 그 자체로 이미 규범적인 개념이며, 그런 의미에서 앞의 인용문에서 예로 든 규범적 영역과 기술적 영역의 간극의 문제는 애당초 제기되지 않는 것이다.

이미 지적한대로, 지향적 태도 혹은 의미가 갖는 의미의 규범성을 해명함에 있어서, 목적론적 의미이론의 가장 핵심적인 장치는 생물학적 개념으로서의 고유 기능(proper function)의 개념이다. 그런데, 여기서의 기능 개념은 우리가 심신문제에서 기능주의라고 말할 때의 기능과는 약간 구분될 필요가 있다. 우리가 소위 심리적 상태가 기능적 상태라고 말할 때의 기능이란 감각적인 입력과 행동적인 출력, 그리고 그 사이의 다른 기능적 상태와의 사이에 성립하는 인과적 관계를 통하여 정의된다. 이런 기능 개념을 가장 잘 보여주고 있는 것이 바로 계산 기계적인 상태로서의 기능적 상태이다. 그런데 이때의 기능 개념은 본질적으로 기술적인 개념이며 목적성이 전혀 없는 개념이다. 하지만 생물학적인 의미로의 기능 개념은 그 자체로서 이미 목적성이 들어와 있는 일종의 규범적인 개념으로서, 심장의 기능은 혈액을 온 몸에 순환시키는 것이라고 말할 때의 기능개념에 해당한다고 할 수 있다. 가령, 어떤 병을 앓고 있는 심장의 경우, 정상적인 심장이 수행하는 혈액 순환의 기능을 수행하지 못할 수 있다. 그럼에도 불구하고, 우리는 병든 심장이 수행해야 할(supposed to perform) 그

⁴ Paul Boghossian, 'The Rule-Following Considerations', Mind 98 (1989), p.532.

목적으로서의 고유기능은 여전히 혈액의 순환이고, 이 심장의 경우는 정상성으로부터 이탈되어 그 고유의 기능을 수행하지 못하고 있는 것이라고 이야기 한다. 이때 ‘정상성’이란 이미 표준과 실제의 괴리를 전제하는 규범적인 개념이다.

목적론자들에 의하면, 어떤 생물학적 기관이 그 목적으로서의 고유의 기능을 갖는다는 사실은 본질적으로 과거의 자연선택(natural selection)의 역사의 문제이다. 어떤 기관이 고유의 기능을 갖는다는 것은, 이것이 마땅히 수행해야 하는 바로서의 목적을 갖는다는 말이다. 그런데 이러한 목적성은 그 생물학적 기관이 거쳐 온 자연선택의 역사를 통하여서만 의미를 지닐 수 있다는 것이다. 즉 어떤 유기체 O가 가지고 있는 기관 X가 Y라는 고유한 기능을 갖는다는 것은, 과거 O의 조상들에서 X와 같은 종류의 것들이 Y를 수행함으로써 O가 속하는 종족의 생존 및 보존에 기여하였고, 그 결과 X가 현재의 형태로 O에 남아 있다는 것이다. 역사적으로 결정되는 이러한 기능개념의 중요한 귀결은 현재의 시점에서 어떤 기관이 정상적인 기능을 수행하지 못한다는 점을 해명할 수 있게 해준다는 것이다. 위에서 지적한 병든 기관 혹은 유전적으로 문제가 있는 기관 등이 바로 그것이다. 이는 어떤 기관이 갖는 기능을 현재의 시점에서 그 기관이 가지고 있는 어떤 인과적, 구조적, 성향적 속성이 아니라 바로 그 기관이 현재 존재하게 된 역사를 통하여 결정함으로써 가능하게 된다. 가령 심장은 어떤 유기체의 과거 자연선택의 역사에 있어서 혈액을 순환시키는 역할을 수행함으로써 그런 유기체의 종류가 현재까지 생존, 보존되는데 기여하였다. 바로 그런 이유로 심장의 기능은 혈액을 순환시키는 것이다. 하지만 그런 역사적인 사실과는 별도로, 어떤 심장은 병이나 여타의 이유로 혈액을 순환시키는 역할을 수행하지 못할 수도 있다. 그러나 그럼에도 불구하고 그 심장의 기능은 여전히 혈액을 순환시키는 것이다.

의미 혹은 지향성이 갖는 규범적 성격을 해명하기 위한 목적론자들의 핵심적인 통찰은 심적 표상이나 믿음등과 같은 지향적 개념들이 그 본질에 있어서 생물학적인 기능적 개념들과 마찬가지로 일종의 고유 기능의 개념을 통하여 분석될 수 있다는 것이다. 이들에 따르면 우리의 심리적인 지향적인 상태는 소화나 혈액의 순환과 같은 일종의 기능적 역할을 수행하고 있다는 것이다. 가령 심적 표상의 메카니즘은 어떤 유기체(종)의 과거 역사에 있어서 환경에 대한 정보를 전달하는 역할을 하였고, 이런 역할을 통하여 그 종이 지금까지 생존, 보존되는 데에 기여하였다는 것이다. 그 결과 오늘날 그 종에 속하는 유기체도 그 조상이 가졌던 것과 유사한 심적 표상의 메카니즘을 가지게 되었으며, 그러한 이유로 현재의 유기체가 갖는 표상체계의 고유한 기능은 환경에 대한 정보를 전달하는 것이 된다. 그런데, 이미 언급한 것처럼, 고유기능을 갖는다는 것은 어떤 것이 마땅히 수행해야만 하는 혹은 수행하도록 계획된 어떤 역할이 있다는 것이다. 개념적 표상이나 믿음이 갖는 의미 혹은

⁵ Fodor, *Psychosemantics*, (1987), p.97

내용(content)은, 각각의 개념이나 믿음이 수행할 것으로 기대되는(supposed to do) 혹은 수행하여야만 하는(should) 바로서의 규제적 역할을 하는 이러한 고유기능을 통하여 결정된다. 만일 어떤 개념이나 믿음상태의 의미 혹은 내용이라는 것이 이런 고유 기능에 의하여 결정되는 것이고, 그리고 이런 고유 기능을 갖는다는 것이 언제나 그런 역할을 수행할 수 있다는 것을 의미하지 않는다면, 최소한 크립키가 지적하는 오류가 가능한 규칙따르기로서의 의미의 규범성을 해명하기 위한 최소한의 단서는 마련된 셈이다.

그런데 이미 살펴본 대로 심적 표상이 그것이 수행하는 역할 때문에 선택이 되는 메카니즘은 자연선택에 의한 진화의 역사이다. 특히 밀리칸 같은 사람에게는, 어떤 대상이 고유의 기능을 갖는다는 것은 그 대상이 그렇게 기능함에 대한 실제적인(actual) 선택이 이루어졌음을 의미한다. 이는 다시 말해서 표상 시스템을 포함한 생물체의 어떤 기관이 고유의 기능을 획득하기 위해서는 그에 걸맞은 적절한 역사를 가지고 있어야 한다는 것이다. 진화의 역사에 있어서 생물체가 어떤 새로운 특징을 가지게 되는 과정은 다양하다. 그 중 대표적인 것이 유전자의 변이에 의한 것이다. 그러나 밀리칸과 같은 목적론자들에게 이렇게 생겨난 1세대의 특징들은 어떠한 고유의 기능도 갖지 못한다. 이런 특징이 세대를 거치면서 생물체를 위하여 모종의 역할을 수행하고 그 유용성을 인정 받아서 안정화될 때만이 고유의 기능을 획득하게 된다. 그런데 만일 심적 표상과 같은 우리의 지향적 상태가 그 본질에 있어서 고유 기능적 범주에 속하는 것이라면, 그래서 그것이 갖는 지향성 자체가 부분적으로는 과거의 역사에 의해 결정되는 것이라면, 우리의 지향적 상태는 단순히 우리의 두뇌가 현재 가지고 있는 인과적 혹은 성향적 속성들 만을 통하여 결정되지 않는다. 다시 말해서 지향적 상태는 우리의 두뇌의 현재적인 물리적 상태에 수반하는 것이 아니다. 목적론자는 바로 이러한 이유에 근거하여, 늑지인간이 아무런 지향적 태도를 갖지 않는다고 결론 내린다.

III

I절의 마지막에서 나는 늑지인간의 지향적 상태와 관련된 두 상충되는 직관을 적절히 평가하기 위해서는, 이 상반되는 직관을 서로 뒷받침하는 이론적 토대에 대한 이해가 선행되어야 함을 지적한바 있다. 그렇다면, 늑지인간이 아무런 지향적 상태를 갖지 않는다는 목적론적 의미론의 이론적 성격을 어떻게 정리할 수 있을까? 목적론적 의미론의 어떠한 부분이 늑지인간에게 지향적 상태를 귀속시킬 수 없다는 결론으로 이르게 하며, 이와 상반되는 직관의 이론적 배경은 또 어떻게 정리할 수 있을까? 나는 목적론적 의미론이 지향적 상태에 대한 매우 강한 의미의 외재주의를 채택하고 있으며, 이에 상반되는 직관은 지향적 상태에 대한 일종의 내재주의를 전제해야 한다고 생각한다. 그렇다면 늑지인간과 관련된 문제는 결국 외재주의와 내재주의 중 어떤 이론이 규범적 현상으로서의 지향적 상태를

해명하는 데 보다 적절한 이론이나의 문제로 귀착될 것이다. 그리고 지향성의 구성원리로서의 내재주의가 가지고 있는 한계점을 지적함으로써, 우리는 늑지인간이 지향적 태도를 갖지 않는다는 목적론적 의미론의 결론을 간접적으로 옹호할 수 있을 것이다.

내재주의와 외재주의가 각기 어떤 입장이며 이들의 대립 관계를 어떻게 규정하는가에 대해서는 여러 가지의 다양한 관점이 있을 수 있겠지만, 이 논문에서는 맥긴(McGinn)의 정의를 따를 것이다.⁶ 먼저 내재주의는 지향적 상태란 것이 환경과는 분리되어 있는 주체에 관한 내부적인 사실들을 통하여 결정된다는 입장이다. 다시 말해서 내재주의는 우리의 몸이나 두뇌 속의 내부적인 상태 혹은 사실들만으로 어떤 심적 상태를 규정하는데 충분하다는 입장이다. 물론 심적 상태와 관련한 이런 내적 사실들은 겉으로 관찰가능한 행동적 성향(behavioral disposition)이나 우리의 감각에 주어지는 직접적인 자극(proximate stimulation)을 포함한다. 여기서 중요한 점은 우리가 심적 상태를 규정하기 위하여 주체 바깥에 있는 외부적인 사실을 언급할 필요가 없다는 것이다. 흔히 이런 내부적 사실의 경계로 우리의 피부(skin)를 언급하기도 한다. 이런 내재주의에 기본적으로 깔려있는 생각은 심적 상태의 동일성에 대한 실체적인(substantial) 관점이다. 말하자면 우리가 믿음이나 욕구와 관련한 심리적, 지향적 상태를 주체에 귀속시킬 때에 우리는 이것들을 어떤 모종의 속성(property)을 실체로서의 주체에 귀속시키는 것으로 이해한다. 어떤 의미에서 이런 내재적인 접근 방식은 어떤 사물의 동일성을 판단하는 데 있어서 우리가 디폴트(default) 값으로 채택하고 있는 조건이다. 우리가 기본적으로 사용하고 있는 언어는 그 성격에 있어서 일종의 사물중심 언어이다.

이에 반하여 외재주의는 내가 가지고 있는 심리적 상태 혹은 지향적 상태를 규정하는데 있어서 주체가 놓여있는 환경과 같은 외부적인 사실이 본질적인 구성원리로 작용한다는 입장이다. 다시 말해서 심리적 상태는 우리의 두뇌나 신체의 상태에 수반하지 않으며, 심리적 상태에 대한 수반의 기초 성질로서 우리로부터 독립적인 외부의 환경과 관련한 사실을 추가하여야 한다는 논제로 이해할 수 있을 것이다. 이러한 외재주의의 입장에서 지향적 상태는 외부적 환경과 관련된 일종의 관계적인 사실로 파악된다. 다시 말해서 마음은 세계에 대하여 근본적인 의미에서 자율성을 유지하고 있는 독립적인 영역이 아니다. 이런 의미에서 외재주의는 무엇이 나와 관련한 심리적 사실을 결정하는가와 관련된 일종의 형이상학적 논제이며, 그 기본적인 주장은 외부의 세계가 나의 심리적인 상태를 결정하는 혹은 구성하는 일부분이라는 것이다. 그리고 기본적으로 이러한 결정성의 방향은 외부에서 내부로 향하여 있다. 그런 의미에서 외재주의는 상당히 강건한 외부세계에 대한 실재론을 전제로 한다.

늑지 인간이 아무런 지향적 상태를 갖지 않는다는 것이 만일 반직관적으로 들린다면 그 이유는,

⁶ Colin McGinn, *Mental Content*, chapter. 1, pp.1-9.

지향적 상태라는 것이 늑지인간이 가지고 있는 물리적인 상태나 행동적인 성향에 의하여 결정되는 것이라고 우리가 암암리에 전제하고 있기 때문이다. 늑지인간의 예에서 우리가 인정하고 있는 것은 그것이 물리적인 측면에서 나와 완전히 동일하고 행동적인 측면에서 지향적 상태를 갖는 존재와 전혀 분간이 안 된다는 것이었다. 여기서 우리가 인정하고 있는 것은, 위의 규정에 따를 때 늑지인간과 관련된 내부적 사실이 나와 관련된 내부적 사실과 완전히 동일하다는 점이다. 그러므로 이런 전제로부터 지향적 상태라는 측면에서 보아서 나와 늑지인간이 크게 다른지 않다는 결론을 도출하기 위해서는 모종의 내재주의적 입장을 전제하여야만 한다. 늑지인간이 지향적 태도를 갖지 않는다는 목적론적 주장에 대한 거부감은 지향적 태도의 구성원리로서의 외재주의에 대한 반발이라고 볼 수 있다.

사실 이런 거부감은 지향적 심리학에 있어서 소위 좁은 내용(narrow content)이라는 개념을 통한 개별주의적인 심리학을 주장하는 사람들의 시도와 맥이 닿아있다. 그렇다면 이런 거부감을 뒷받침하고 있는 내재주의적 직관의 기본적인 논점은 무엇인가? 그 첫 번째 이유로 우리는 앞서 언급한 바 있는 실체적 사고를 들 수 있을 것이다. 물론 지향적 상태가 곧 실체적인 것은 아니다. 하지만 여기서 지향적 상태는 어디까지나 외부로부터 단절된 하나의 폐쇄된 체계로서의 실체적 대상이 갖는 속성적인 상태로서 이해되고 있다. 두 번째로 지향적 상태가 가지고 있는 인과적 효용성(causal efficacy)의 문제를 들 수 있다. 가령 전통적인 기능주의의 입장에 따를 때 하나의 심리적 상태는 그 심리상태가 갖는 인과적 역할을 통하여 정의된다. 그런데 이 기능주의에서 말하는 인과적 역할이라는 것은 결국 따지고 보면 우리의 피부 안으로 제한되어 있는 인과적 역할이다. 그리고 그것은 앞서 언급한 바 있는 모종의 실체가 지니는 속성적인 의미에서의 인과적 역할이다.

김재권이 지적하듯이 어떤 것이 존재한다는 주장은 그것이 인과적 영향력을 갖는다는 것이다. 아무런 인과적 영향력이 없는 무엇이 존재한다는 주장은 별로 흥미로운 주장이 되지 못한다. 우리가 어떤 것을 그 어떤 것으로 만들어 주는 구성적인 원리가 여차여차한 것이라고 주장하고 있다고 가정해 보자. 이 때에, 그 구성적인 원리는 그 어떤 것이 가지고 있는 인과적 영향력을 해명해주어야 하는 원리여야 한다고 요구하는 것은 대단히 적절한 요구인 것처럼 보인다. 그런데, 우리가 인과적 영향력을 운운할 경우 그때의 인과성이란 개념은 대개의 경우 국소적인(local)인 문제이다. 다시 말해서 인과적 영향력의 주체로서 생각되는 대상이나 사건도 시공간적인 측면에서 국소적으로 제한되며, 그 인과적인 영향력이란 것도 대상이나 사건이 가지는 내재적(intrinsic)인 속성을 통하여 규명된다. 만일 인과적인 영향력이 이러한 것이라면, 외재주의적으로 규정된 지향적 상태의 경우 어떻게 그 인과적 영향력을 규명할 수 있을지가 대단히 모호해진다. 우리는 앞에서 이미 외재주의적으로 규정된 지향적 상태는 우리의 두뇌나 신체의 물리적 상태에 수반하지 않음을 지적하였다. 그런데 행위의 주체로서의

우리가 행사하고 있는 인과적 영향력이 우리의 두뇌나 신체에 수반할 수밖에 없는 것이고, 우리가 가지고 있는 하나의 상태로서의 지향적 상태가 가지고 있는 인과적 영향력이란 것도 결국에는 우리의 신체나 두뇌의 일부가 갖는 인과적 영향력에 수반할 수밖에 없는 것이다. 그런데 지향적 상태라는 것이 외재적으로 규정된다면, 외부의 대상이나 역사적인 과거가 나의 행동을 일으키는 데에 모종의 별도의 인과적 영향력을 행사한다는 말인가? 만일 그렇다면 이는 인과적 영향력에 대한 대단히 이상한 이론으로 귀착된다.⁷

여기서 우리는 국소적 인과성, 개별주의, 실체적 사고, 내재주의, 계산 기계적 기능주의, 커밍스적 기능 개념⁸ 등과 같은 일련의 생각들이 서로 맞물린 체 견고한 하나의 틀을 형성하고 있음을

⁷ 외재주의를 받아 들일 때, 지향적 성질들이 갖는 인과적 효용성은 별도의 논의를 필요로 하는 것처럼 보인다. 참고로 이 문제에 대한 버지와 드레츠키의 응답을 소개한다. 버지는 지향적 태도에 대한 개별화의 원리로서의 외재주의와 인과성에 대한 국지주의는 서로 양립가능한 것이라고 논한다. (Burge, "Individualism and Psychology", p.16) 외재주의의 논점은 두 가지 다른 환경이 행위 주체의 신체에 가하는 인과적 영향에 있어서는 서로 동일하지만, 행위 주체에게 전달되는 지향적 내용에 있어서는 차이를 만들 수 있다는 것이다. 즉, 지향적 상태를 외재적으로 규정한다고 해서, 행위의 발생과 관련하여 외부의 환경이 주체의 신체를 통하지 않는 어떤 별도의 인과적 영향력을 행사하고 있다고 생각할 필요는 없다. 환경이 문제가 되는 것은 지향적 태도의 개별화와 관련된 것이지, 지향적 태도가 갖는 인과적 힘에 영향력을 끼친다는 것은 아니라는 것이다. 그러나 이러한 주장에 대해서는 지향적 태도의 구획과 관련된 외부적 요소를 인과적으로 무력하게 만듦으로서 일종의 부산현상론에 빠진다는 비판이 적용될 수도 있다. 드레츠키는 국지적으로 문제가 되는 인과성과 지향적 설명의 맥락에서 문제가 되는 인과성을 구분함으로써 이러한 비판에 답하고자 한다. (Dretske, *Naturalizing the Mind*, p.160) 가령 중을 울렸을 때(C) 조건화되어 있는 개는 침을 흘린다(E). 우리가 침을 흘리게 만든 원인이 무엇이나고 물었을 때, 중을 울린 것이나 이로부터 야기되는 모종의 청각적 경험 등은 국지적으로 문제가 되는 인과적인 힘으로서의 촉발적 원인에 해당하는 것이다. 드레츠키는 이러한 촉발적 원인을 통한 행위의 설명은 가령 신경생리학 등이 담당해야 할 문제라고 본다. 그런데, 개가 침을 흘리는 것이 단순히 신체적인 움직임이 아니고 행동인 한에 있어서, 우리가 지향적 설명의 문맥에서 해명하고자 하는 것은 이러한 촉발적 원인이 아니다. 지향적 설명에서 문제가 되는 것은, 중을 울렸을 때 개가 도망가 간다던가 하지 않고 왜 침을 흘리는가에 대한 이유를 묻는 것이다. 말하자면 지향적 설명에서의 인과성은 C를 가지고 E를 설명하려는 것이 아니라, 중을 울리는 사건으로부터 왜 침을 흘리는 행동의 패턴이 생겨났는가에 대한 질문, 즉 C->E 라는 구조가 생겨난 것에 대한 원인과 관련되어 있다는 것이다. 드레츠키는 이러한 원인을 구조적 원인(structuring cause)이라고 부른다. 그리고 이러한 구조적 원인을 해명함과 관련하여 외부적 환경이나 과거의 역사들이 인과적 관련성을 띄게 된다. 이러한 주장에 대한 보다 포괄적인 논의는 드레츠키의 *Explaining Behavior* 를 참조하라.

⁸ 커밍스는 어떤 상태 혹은 사물이 현재 가지고 있는 성질 혹은 능력을 통하여 기능개념을 분석하려 한다. 우리가 기능주의를 말할 때의 그 기능은 이런 커밍스의 분석에 매우 가까운 것이다. 커밍스는 자신의 이론을 '기능적 분석(functional analysis)'의 이론이라고 부르는데, 이 이론에 따르면 우리가 어떤 대상 X에 기능을 귀속시키는 진술은 X가 속해 있는 어떤 복잡한 시스템의 능력 혹은 행위를 설명하기 위한 것이다. 가령 X가 속해있는 시스템 S가 어떤 복잡한 능력 C를 갖는다고 하고, S가 어떻게 하여 그런 복잡한 능력 C를 갖게 되었는가를 설명하려고 한다고 해보자. 우리는 이때 S가 갖는 이 C를, S를 구성하고 있는 부분들이 행하는 보다 단순한 행위 혹은 능력을 통하여 설명하려고 할 것이다. 즉 보다 단순한 어떤 것들이 모여서 어떻게 복잡한 행위 혹은 능력을 갖게 되는가를 설명하려는 것이 바로 기능적 진술의 목적이란 것이다. 그렇다면 X가 어떤 기능 F를 갖는다는 것은, X가 속한 시스템 S가 복잡한 능력 C를 가짐에 있어서 X가 F를 수행함으로써 기여한다는 것을 의미한다. 그런데 여기서 한 가지 중요한 점은 X가 커밍스적인 기능 F를 갖는다는 것은 X가 실제로 어떤 인과적 혹은 성향적 속성을 가지고 있다는 것을 의미한다. 즉 X가 기능 F를 갖기 위해서는 실제로 F를 수행할 수 있어야만 한다. 커밍스 이론의 이러한 점은 기능에 대한 역사적 접근에 비추어 규범성의 해명에 있어서 중대한 차이를 가져온다. 역사적 이론은 기능이 과거의 역사에 의해 결정된다는 것을 주장함으로써 규준과 실제의 괴리를 확보할 수 있었다. 하지만 커밍스의 이론에서는 규준과 실제의 괴리에 대한 설명이 그렇게 용이하지 않다. 가령 병이든 심장의 예를

관찰할 수 있다. 그리고 늑지 인간이 지향적 상태를 갖지 않는다는 주장에 대한 거부감은 바로 이러한 틀에 근거하고 있다. 자 그러면 늑지인간에 대한 이런 주장이 결코 반 직관적이지 않다는 것을 보여주기 위해서는 무엇을 하여야 하나? 여기서 가장 먼저 지적 할 수 있는 것은, 외재주의자가 가지고 있는 기본적인 문제의식이 지향적 심리학에서 가정된 설명적 단위로서의 지향적 상태를 어떻게 구획(demarcation)짓는가의 문제에 대한 것이라는 점이다. 지향적 상태에 대한 제거주의적인 입장을 취하지 않는 한, 우리는 지향적 심리학을 모종의 실재하는 지향적 상태에 대한 경험과학으로 인정한다. 그리고 현재 받아들여지고 있는 표준적인 입장은 켈라스의 주장대로 지향적 태도를 지향적 심리학에서 가정하는 일종의 이론적 대상으로 간주하는 것이다. 그렇다면 우리가 어떤 대상에게 지향적 태도를 귀속시킨다고 할 때에, 하나의 이론적 대상으로서의 그러한 지향적 태도를 어떻게 적절히 개별화(individuate)하는나의 문제가 자연스럽게 대두된다. 외재주의는 지향적 태도를 개별화함에 있어서, 외부적 환경이 그 구성적 원리로 개입된다는 주장이다.

그런데 사실 이런 외재주의는 그 주장의 파격성에 비추어 볼 때, 어떤 의미에서 우리에게 대단히 친숙한 논제이기도 하다. 그 대표적인 논의가 퍼트남의 쌍둥이 지구 논변이다. 쌍둥이 지구 논변의 핵심은 의미라는 것이 우리의 머리 속에 있는 것이 아니며, 어떤 단어의 외연은 기본적으로 화자가 처한 환경 속의 대상이 갖는 본질을 통하여 결정된다는 주장이다. 혹자는 퍼트남의 쌍둥이 지구논변이 보여주고 있는 것은 어떤 단어가 가리키는 외연을 결정하기 위하여 우리가 외부 대상의 본질을 고려해야 한다는 주장이지, 그것 자체가 외부 세계가 우리의 심적 상태를 규정한다는 외재주의 논제인 것은 아니라고 반박할 수도 있을 것이다. 과연 그런가? 우리가 퍼트남의 쌍둥이 지구논변을 우리의 심적 표상 혹은 상태를 구분하기 위한 혹은 개별화하기(individuate) 위한 동일성의 조건으로 이해한다면, 이는 전형적인 외재주의적인 입장이 된다. 가령 내가 ‘물’이라는 개념을 통하여 가지고 있는 표상은 쌍둥이 지구의 내가 가지고 있는 ‘물’이라는 표상과 같은 것인가 다른 것인가? 퍼트남의 쌍둥이 지구 논변의 핵심은 나와 쌍둥이 지구의 내가 가지고 있는 ‘물’에 대한 심적 표상은 그 내재적인 성질에 있어서는 동일하지만, 엄연히 서로 다른 내용을 갖는 구분되는 표상임을 주장한다. 물론 이때의 구분은 단순한 수적인 동일성에서의 구분을 이야기하고 있는 것이 아니다. 그러면 이때 내가 가지고 있는 표상과 쌍둥이 내가 가지고 있는 표상을 서로 다른 것으로 만들어 주는 것은 무엇인가? 퍼트남의

다시 생각해보자. 병이 든 심장의 경우 이것은 이미 혈액을 순환시킬 수 있는 어떤 인과적 성질이나 성향을 갖지 못한다. 그러므로 우리는 이 병든 심장의 기능이 혈액을 순환시키는 것이라고 말할 수 없다. 동시에 우리는 병든 심장이 그 정상적 기능을 수행하지 못하고 있는 일탈적인 것이라고 이야기 할 수도 없다. 대신에 이 병든 심장은 여전히 어떤 복잡한 시스템의 한 부분일 수 있다. 가령 극단적으로 어떤 유기체의 죽음이라는 복잡한 과정에 기여하는 것이 그것의 기능일 수도 있다. 커밍스의 분석에 있어서 분석의 대상이 되는 시스템이 꼭 어떤 유기체의 생존에 꼭 필요하다거나 하는 중요한 행위를 수행해야 하는 것은 아니다. 커밍스 이론의 목적은, 그것이 무엇이

답은 바로 환경의 기여라는 것이다. 여기서 우리는 하나의 심리적 상태를 어떤 심리적 상태이게끔 만들어주는 하나의 구성적인 원리로서 외부적 환경이 이미 개입하고 있음을 관찰할 수가 있다.

어떤 의미에서 늑지인간이 지향적 태도를 갖느냐 각지 앓느냐의 문제는, 쌍둥이 지구인간이 가지고 있는 ‘물’에 대한 표상적 내용이 내가 가지고 있는 ‘물’에 대한 표상적 내용과 동일한 것인지 아닌지의 문제와 동일한 선상에 있다고 할 수 있다. 물론 쌍둥이 지구 논변과 늑지인간의 논변의 경우에 중요한 한가지 차이점이 있다. 쌍둥이 지구논변에서 외부적인 환경은 일차적으로 공간적인 외부 환경이다. 하지만 늑지인간과 관련된 측면에서, 목적론적 의미론은 공간적인 의미에서의 외부적인 환경뿐만 아니라, 유기체와 관련한 역사적 사실을 이러한 외부적인 환경의 일부로 포함시킴으로서 더욱 강한 의미의 외재주의를 채택하고 있다. 물론 쌍둥이 지구의 경우도 인과적인 역사가 결부되어야 함으로 과연 공간적인 차원의 외부적인 환경만이 결부되는 것인지에 대해서는 많은 논란이 있을 수 있을 것이다. 어찌되었던, 현재 우리가 관심을 가지고 있는 외재주의는 공간적인 차원의 외부적인 환경뿐만 아니라 시간적인 차원의 외부적인 환경도 고려해야 한다는 강력한 형태의 외재주의이다.

그렇다면 늑지인간에 대한 목적론자들의 결론은 일차적으로 다음과 같이 변호될 수 있을 것이다. 가령 믿음과 같은 지향적 상태는 그 내용(content)을 통하여 그 동일성의 조건이 규정된다. 하나의 믿음을 그러한 믿음으로, 혹은 하나의 믿음과 다른 하나의 믿음을 동일한 믿음 상태로 만들어 주는 것은 바로 그 믿음의 표상적 내용이다. 그런데 내용과 관련된 의미론적인(semantical) 단계에서 지향적 상태를 구획짓는 원리로서 우리가 가지고 있는 가장 유망한 이론은, 퍼트남이 쌍둥이 지구 논변을 통해 보여주었던 넓은 내용(wide content) 즉 외재주의적인 것이다. 그렇다면 만일 넓은 내용에 관한 퍼트남의 쌍둥이 지구에 대한 논변을 받아들일 수 있다면, 늑지 인간에게 나와 같은 지향적 태도를 귀속시킬 수 없다는 주장도 그렇게 반직관적이지 않다. (시간적인 역사적 차원에 대한 논의는 IV절에서 이루어질 것이다.)

내용을 갖는 지향적 상태를 구획짓는 원리로서의 개별주의적 혹은 내재주의적 입장의 부적절성을 지적하기 위하여 다음과 같은 고려를 추가할 수 있을 것이다. 가령 내재주의적인 입장에서 지향적 상태를 구획짓는 것으로서 채택할 수 있는 원리란 물리적이거나 인과적, 계산적 기능 혹은 성향적 속성과 같은 것들이다. 그러나 지향적 상태가 내용을 통하여 개별화 되는 한, 동일한 내용을 갖는 지향적 상태는 복수의 물리적, 기능적 상태에 의하여 실현 될 수 있다. 우리는 이미 기능주의가 복수실현 가능성이란 논제를 통하여 전통적인 심신 동일론을 논박하는 것을 목격하였다. 그런데 동일론을 대체한 기능주의란 일종의 내재적인 계산적 기계적 기능주의이다. 그러나 많은 사람들이

되었던 간에 어떤 복잡한 시스템의 행위나 능력을 그 부분들의 단순한 행위나 능력을 통하여 설명하는 데에 있기 때문이다. (Cummins, ‘Functional Analysis’ 를 참조하라.)

지적하고 있듯이 하나의 지향적 상태는, 이것을 실현하는 계산적 기능의 상태라는 측면에서도 또한 대단히 유연한(plastic) 것이다.⁹ 즉 동일한 내용을 갖는 하나의 지향적 상태는 복수의 계산적 기능 상태에 의하여 실현 가능하다. 만일 동일론이 지향적 태도의 구획에 대한 적절한 이론이 아니라면, 바로 동일한 이유에 의해서 전통적인 의미에서의 기능주의적 이론도 적절한 이론이 아니다.

지향적 태도, 혹은 단순한 물리적 움직임과 구분되는 바로서의 지향적 행동이 단순히 종 한정적인(species specific) 개념이 아니라 여러 종에 걸쳐서 귀속될 수 있는 개념이라는 것을 고려하면 이점은 더욱 분명해진다. 가령 만일 원숭이가 표현하는 ‘두려움’과 개가 표현하는 ‘두려움’을 함께 묶어 줄 수 있는 구획의 원리라는 것이 있다면 그것은 도대체 어떠한 모습을 띄고 있겠는가? 이에 대한 내재주의자의 답은 대단히 웅색할 수밖에 없다. 지향적 설명의 대상이 되는 행위의 경우도 마찬가지이다. 드레츠키가 적절히 지적하고 있듯이, 지향적 설명의 대상이 되는 것은 행위(behavior)이지 단순한 신체적인 움직임이 아니다.¹⁰ 이 말은 지향적 설명의 대상이 되는 행위가 단순히 외향적인 물리적 기계적 움직임만으로는 기술될 수 없다는 것이다. 가령 어떤 종류의 개똥벌레는 빛을 받함으로써 배우자를 유혹하지만 다른 종류의 개똥벌레는 빛을 받함으로써 먹이를 유혹한다. 이들은 외향적으로는 유사하지만 전혀 다른 범주에 속하는 행위들이다. 그런데 여기서 우리는 이 두 종류의 개똥벌레가 그 내재적인 인과적, 기능적 속성에 있어서는 완전히 동일한 가능성을 상상할 수 있다. 그렇다면 내재주의자는 이 두 가지의 행위가 서로 다른 범주에 속한다는 것을 적절히 설명할 수가 없다.

여기서 한 가지 가능한 제안이 환경과 역사의 기여를 포함하는 목적론적 개념으로서의 기능개념의 도입이다. 하나의 심리적 상태를 그러한 심리적 상태로 만드는 것은 단순한 계산적인 의미에서의 기능적 상태를 실현하고 있다는 것이 아니라, 환경에 대한 정보를 전달하고 환경의 변화에 대처하기 위한 고유의 기능을 가지고 있는 상태라는 것이다. 그리고 하나의 심리적 (혹은 지향적 상태)와 다른 하나의 심리적 상태를 동일한 범주로 묶어주는 것도 바로 이들이 이런 의미에서의 동일한 고유 기능을 가지고 있다는 것을 의미한다. 이미 지적한 데로 지향적 설명 혹은 일반화라는 것이 단순히 종 한정적인 설명이 아니라는 점을 고려한다면 고유 기능을 통한 이런 식의 접근의 장점은 더욱 선명히 부각된다. 결국 목적론적 의미론자들이 주장하고 있는 것은, 어떤 지향적 상태의 수반적 기초는 단순히 그 상태가 귀속되는 시스템 내부의 물리적이거나 계산적인 상태가 아니라, 그 시스템이 속하는 시공간적 생태학적(ecological) 차원을 포함하여야 하며, 이 경우에만 지향적 상태를 적절히 개별화 할 수 있다는 것이다. 이런 점에서 목적론적인 의미론자들이 취하고 있는 지향적 태도에 진화론적 접근

⁹ 가령 퍼트남의 *Representation and Reality*, chapter 5 를 보라.

¹⁰ Dretske, *Explaining Behavior*. chapter 1.

방식은 지향성의 개별화와 관련한 외재주의적인 입장에 대한 보다 폭 넓은 동기를 제공하고 있는 것처럼 보인다. 진화론적 관점에서 볼 때, 믿음이나 욕구와 같은 인지 기체들은 먹이를 찾거나, 약탈자를 회피하기 위한 것과 같은 환경적 압력을 극복하기 위해 선택된 적응(adaptation)의 산물이다. 이러한 지향적 상태들이 어떠한 특징을 가지고 있는 것은 결국 이들 상태의 주체가 어떤 특정한 환경에 처해졌기 때문이며, 만일 그 주체가 다른 환경적 조건에 처해졌다면 다른 특징을 갖는 다른 상태를 발전시키지 않았겠는가? 이런 점에서, 이들 지향적 상태들이 갖는 본성은 본질적으로 관계적인 것처럼 보인다. 만일 지향적 상태의 본성이 관계적인 것이라면 심성내용의 외재성 혹은 지향적 태도의 구획에 있어서 외재주의는 당연한 귀결이다.¹¹

IV

다음으로 지향적 상태가 규범적 현상이라는 점을 중심으로 내재주의와 외재주의를 비교하여 보자. 이러한 논의를 통하여 우리는 늘지인간에게 아무런 지향적 태도를 귀속시킬 수 없다는 목적론자의 주장을 보다 적극적으로 옹호할 수 있을 것이다. 먼저 다음과 같은 예를 통하여 논의를 시작해보자. 바닷바람이 불어서 아니면 개미떼가 지나간 흔적에 의하여 정말 우연적으로 백사장에 ‘퍼트남’이라는 모양의 형체가 새겨졌다고 해보자.¹² 이 형체는 그 자체로서 퍼트남이라는 사람을 가리키는 이름인가 아니면 그냥 아무런 의미가 없는 물리적 형상일 뿐인가? 만일에 누군가가 의도적으로 모래사장에 ‘퍼트남’이라는 글자를 썼다면 이는 분명히 퍼트남이라는 사람을 가리키는 이름일 것이다. 그런데, 개미떼가 우연히 만들어 놓은 ‘퍼트남’이란 형상은 그 내재적인 물리적 성질에 있어서 누군가가 써놓은 ‘퍼트남’이란 글자와 동일함에도 불구하고 퍼트남을 가리키는 이름으로 간주될 수가 없다. 이는 아무런 의미론적 성질을 갖지 못하는 물리적 형상일 뿐이다. 이 두 가지의 경우의 차이는 발생론적인 차이이다.

만일 이 ‘퍼트남’이라는 형상에 의미를 부여할 수 없다는 주장에 동의할 수 있다면 다음의 경우는 어떠한가? 어느 날 바람이 불고 번개가 치더니, 벼락이 나무에 떨어졌다. 어떤 우주적인 우연에 의하여 주위에 널려있던 분자가 합성되어 반인반마의 모습을 한 켄타우로스 같은 존재가 생겨났다고

¹¹ 맥킨은 이런 맥락에서 진화론적인 목적론이 쌍둥이 지구논변과 관련된 지표성(indexicality)에 대하여 매우 흥미로운 설명을 제공하고 있음을 지적한다. 목이 말라서 나는 물을 마시고 싶어한다. 그런데, 내 위나 간의 기능은 H₂O를 처리하는 것이다. 이들 기관들이 갖는 특질들은 내가 처해있는 이 지구라는 환경 속에서 발견된 어떤 특정한 물질을 이용하기 위하여 진화된 것이기 때문이다. 그런 점에서, 물에 대한 나의 표상은 지구상의 바로 그 물질을 가리킨다. 한편으로 쌍둥이 지구의 나에게 속한 위나 간은 한번도 H₂O를 접한 적이 없으며 쌍둥이 지구에서 발견되는 특정한 환경적 조건 즉 XYZ를 이용하기 위해 진화된 적응물이다. 그런 점에서, 쌍둥이 지구인이 갖는 물에 대한 표상은 XYZ를 가리켜야만 한다. (McGinn, *ibid.*, p.157-8)

¹² Putnam, *Truth, Reason and History*, p.1 참조

가정하자. 그런데, 이 존재는 가령 ‘물’이라던가 하는 인간이 사용하는 말과 비슷한 몇 마디 소리를 낼 수도 있다. 덧붙여 이 존재는 물을 보면 ‘물’이라는 소리를 내기도 하고, 그런 소리를 낼 때 우리가 물을 가져다 주면 그 물을 ‘마시기’도 한다. 하지만 이 존재는 소위 퍼트남이나 크립키적인 의미이론에서 말하는 언어적 공동체에 속하지도 않으며, 우리의 ‘물’이라는 말의 사용과 어떠한 인과적 연쇄관계도 맺고 있지 않다. 이 존재는 말 그대로 갑자기 우연적으로 생겨나서 ‘물’이라는 소리를 내고 있는 것이다. 그렇다면 이 존재가 내는 ‘물’이라는 소리는 우리가 사용하는 H_2O 를 가리키기 위하여 사용하는 ‘물’이라는 말과 동일한(혹은 유사한) 표상적 내용을 갖는 언어적인 표상인가, 아니면 그것과는 구분되는 다른 의미를 갖는 언어적인 표상인가, 혹은 아무런 의미를 갖지않는 소음에 불과한가? 만일 이것이 우리가 사용하는 ‘물’이라는 표현과 비슷한 언어적인 표상이며, 이 존재는 그에 상응하는 모종의 지향적 상태를 가지고 있다고 가정해보자. 그러면 이것이 콜라를 보고 ‘물’이라는 소리를 내는 것은 거짓 믿음과 같은 오표상(misrepresentation)의 사례인가? 만일 이것이 오표상의 경우라면, 무엇이 이것을 오표상으로 만드는가?

내재주의자는 이 경우에 이 존재가 물을 보고 ‘물’이라는 소리를 낼 때에 갖게 되는 내재적으로 규정되는 일종의 물리적, 인과적, 기능적인 상태를 통하여 이 오표상을 설명하여야 한다. 그렇다면 이 존재가 가지는 내적인 기능적 상태의 어떤 부분이 이 존재가 갖는 ‘물’이라는 표상을 통하여 물 혹은 H_2O 를 가리키게 하는가? 일단 내재주의자의 입장에서는 이 표상의 내용을 고정하기 위하여 외부세계에 존재하는 대상으로서의 H_2O 에 호소할 수는 없다. 내재주의자는 물이라는 대상이 촉발하는 일종의 감각 내용으로부터 시작하여야 한다. 가령 어느 날 갑자기 이 존재가 쌍둥이 지구로 공간이동을 했다고 가정해보자. 그럼 이 존재가 내뱉는 ‘물’이라는 소리는 이제 H_2O 가 아니라 XYZ를 가리키게 되는가? XYZ에 의해 촉발되는 감각적 경험이 H_2O 의 경우와 완전히 동일할 것이므로 그렇지 않을 것이다. 그러나 어차피 환경의 변화에 따라서 의미가 결정된다는 것은 외재주의적인 논제이므로, 내재주의자에게 이것은 큰 문제가 아닐 수도 있다. 여기서 문제는 내재주의자가 호소할 수 있는 성질이 그 존재가 가지고 있는 현재의 내부적인 인과적, 성향적 속성에 국한된다는 것이다. 만일에 내재주의자가 이 존재의 표상적 상태를 기능적 상태로 정의한다면, 내재주의자가 호소할 수 있는 기능개념은 일종의 커밍스적인 기능 개념이다.(각주 8참조) 그런데 커밍스의 기능개념이란 결국 성향적 분석이며, 크립키의 논점처럼 이러한 성향적 분석을 통해서도 지향적 태도가 가지고 있는 규범성을 확립할 수가 없다. 다시 말해서 설령 이러한 존재에게 어떤 모종의 내적인 기능적인 상태를 부여할 수 있을지는 모르지만, 이것이 어떤 규범성을 지니는 지향적 태도로 간주하기는 어렵다는 것이다. 지향적 태도가 가지는 규범성을 경험과확적인 자연적 성질을 통하여 해명하는 것이 자연주의자가 가지고 있는 중요한 화두 중의 하나라고 한다면, 이런 점에서 내재주의는 결코 매력적인

입장이 아니다.

다음으로 외부적 환경이 시간적 차원을 포함하여야 한다는 보다 강력한 의미의 외재주의를 지지하기 위하여, 켄타우로스와 같은 존재가 내 뱃는 ‘물’이라는 소리가 현재 그것이 처한 공간적인 환경과의 인과적인 상호작용에 의해 결정될 수 있다는 주장을 검토해보자. 커밍스적인 기능적 분석에 있어서 근접인(proximate cause) 뿐만이 아니라 distal cause나 distal effect도 포함 시킴으로서, 분석의 대상이 되는 시스템을 공간적인 환경을 포함하는 방식으로 확장하는 시도가 될 것이다. 이른바 긴팔 기능주의자(long-arm functionalist)들의 주장이 이에 해당한다. 어떤 방식이 되었던 간에, 이들의 주장은 ‘물’이라는 발화와 H₂O 사이에 성립하는 모종의 인과관계에 의하여 ‘물’이라는 발화 혹은 이와 관련된 지향적 태도의 내용이 결정된다는 것이다. 이 경우에도, 우리는 이 늑지존재가 쌍둥이 지구로 이동하는 예를 통하여 그 한계를 지적할 수 있을 것이다. 가령, 우리의 켄타로우스가 늑지에서 생겨난 지 얼마되지 않아서 쌍둥이 지구로 이동하였다고 가정해보자. 여기서 한가지 깔려져 있는 가정은 이 존재가 지구에서 머문 기간이 너무나 짧아서 H₂O와 충분한 인과적 상호작용을 갖지도 못했고, 지구상의 언어공동체에 속하지도 않기에 우리로부터의 어떠한 형태의 의미 빌림(reference borrowing)도 가능하지 않다는 것이다. 그러면 이제 쌍둥이 지구 상에서 이 존재가 내뱃는 ‘물’이라는 발화는 XYZ를 가리키게 되는가? 우리는 연속하여 다음과 같은 질문들을 계속 던질 수 있을 것이다. 설령 어떤 명시적인 명명식(baptism)이 결부되어 있지 않다고 하더라도, 이 존재가 H₂O와 갖게 되는 최초의 인과적 상호작용이 ‘물’이라는 발화의 내용을 고정하게 되는가? 아니면 그 내용을 고정하기 위하여 도대체 어느 정도의 상호작용이 필요한가? 물론 나는 그 대답을 알지 못한다. 하지만 내가 그 답을 모르고 있는 것은 내가 꼭 알아야만 하는 어떤 인과적, 기능적 사실에 무지하여서가 아니라, 정답이 없기 때문이라는 것이 바로 그 이유일 것이다. 어떤 지향적 상태가 규범적인 성격을 갖는다 함은 그 상태가 표상하도록 되어있는 (supposed to represent) 어떤 사태가 있다는 것이다. 그러나, 위의 간단한 사유실험이 보여주는 것은, 그 존재에 대한 어떤 적절한 역사가 없이, 우리는 그 존재의 지향적 상태가 무엇을 나타내도록 되어있는 것인지에 대해서 말할 수 없다는 것이다. 목적론적 의미론자들도 하여금 늑지인간에게 아무런 지향적 상태도 귀속시킬 수 없다고 주장하게 하는 것은 바로 이러한 이유 때문이다.

물론 이러한 가상의 예들은 양쪽의 날을 가진 칼 같은 것이라고 할 수 있다. 이러한 예들은 가령 테넷과 같은 (아마도 콰인을 포함하여) 사람들이 주장하듯이 지향성이나 의미라는 것은 실재하는 것이 아니라 우리가 어떤 시스템의 움직임을 예측하고 설명하기 위한 도구적인 성격만을 지닐 뿐이라는 결론을 유도하기 위하여 사용될 수도 있다. 하지만 우리가 지향성이나 의미와 관련한 모종의 실재론을 전제한다면, 이러한 경우에 외재적인 요인을 고려하지 않은 채, 단순히 그 존재의

개별주의적인 물리적, 기능적, 행동적 성향만으로 그 의미를 결정할 수 있을 것처럼 보이지는 않는다. 더군다나 그런 사실들 만에 의거하여 오표상을 논할 수 있는 규범적인 차원에서의 의미를 부여할 수 있을 것 같지는 않다. 이러한 예들에 있어서 목적론자들은, 켄타우로스 같은 이 존재가 내뱉는 ‘물’이란 소리에는 그 규범적 의미를 결정지어 줄 수 있는 아무런 의미론적 사실도 없다는 크립키의 회의주의자의 주장에 동의할 것이다.

마지막으로 밀리칸 같은 목적론적 의미론자들이 늑지인간이 아무런 지향적 태도를 가지지 않는다고 말할 때, 이들이 부정하는 것은 정확히 무엇인지를 정리하면서 이 논문을 마무리하자. 앞서도 지적하였듯이 이들이 노리고 있는 것은 인간이나 유기체등의 행위를 설명하기 위한 지향적 심리학에서 채택하고 있는 믿음이나 욕구와 같은 상태들에 대한 일종의 유망한 이론적 정의를 제시하는 것이다. 그리고, 이들 상태들은 이것들이 갖는 내용을 통하여 구획되는 상태들이며, 본질적으로 이들이 표상하는 내용에 근거한 규범적인 평가의 적용을 받는 상태들이다. 앞서 지적되었듯이, 규범적 성격을 지닌다는 것은 언제나 어떤 것이 마땅히 그러해야 한다거나 그러할 것으로 기대된다는 일종의 목적성이 전제되는 개념이다. 이들의 주장은 이러한 목적성을 해명하기 위해서는 시간적, 공간적 환경으로서의 외부적인 사실들의 개입이 필수적이라는 것이다. 결국 늑지인간에게서 부정되고 있는 것은 이러한 규범적 평가의 적용을 받는 태도로서의 지향적 상태들이다. 목적론자들도 늑지인간의 내부에 우리와 유사한 모종의 물리적인 인과적 과정이 진행되고 있으며, 그러한 과정을 통하여 어떤 외부적인 물리적 움직임이 촉발되고 있음을 부정하지는 않는다. 하지만 아무리 그러한 인과적 과정이 우리와 유사하다고 하더라도, 그런 인과적 과정에 대하여 아무런 목적성을 부여할 수 없으며, 그 결과 가령 그것이 참이라거나 충족되었다는 식으로의 규범적 평가를 적용할 수 없다는 것이다.

이런 식으로 이해된 늑지인간에 대한 목적론자들의 주장은 대단히 온건한 주장이다. 한편으로 혹자는 이런 목적론자들의 주장이 지향적 상태에 대한 너무나 협소한 규정에 기초해 있다고 비판할 수도 있을 것이다. 하지만 이들의 이론이 표상적 내용에 대한 일상적 개념의 분석이 아니라 일종의 이론적 정의로 제안되고 있음을 기억하자. 아마도 이들이 기대하는 것은 우리들이 채택하게 될 과학적 이론으로서의 지향적 심리학에서 지향적 태도들이 목적론적 용어로 통하여 정의되고, 늑지인간의 경우가 아무런 저항 없이 받아들여지는 날이 오는 것일 것이다. 지향성이란 개념자체가 어차피 일상적인 개념이 아니라 철학자들만의 전문적인 용어가 아니던가?

참고문헌

- Allen, C; Beckoff, M; Lauder, G., eds., (1998) *Nature's Purpose: Analyses of Function and Design in Biology*, Cambridge: MIT Press
- Boghossian, P., (1989) "The Rule-Following Considerations", *Mind* 98, 507-549
- Braddon-Mitchell, D. and Jackson, F., (1996) *Philosophy of Mind and Cognition*, Oxford: Basil Blackwell
- Buller, D. J., (1999) *Function, Selection and Design*, Albany: SUNY Press
- Burge, T., (1986) "Individualism and Psychology", *Philosophical Review* 95, pp.3-46
- Cummins, R., (1975) "Functional Analysis", *Journal of Philosophy* 72. 741-765. Reprinted in Buller (1999)
- Cummins, R., (1989) *Meaning and Mental Representation*, Cambridge: MIT Press
- Davidson, D., (1987) "Knowing One's Own Mind", Reprinted in Davidson (2001), *Subjective, Intersubjective, Objective*, Oxford: Clarendon Press
- Dennett, Daniel C., (1987) *The Intentional Stance*, Cambridge: MIT Press
- Dretske, F. I., (1986) "Misrepresentation" in Stich and Warfield (1994)
- Dretske, F. I., (1988) *Explaining Behavior*, Cambridge: MIT Press
- Dretske, F. I., (1997) *Naturalizing the Mind*, Cambridge: MIT Press
- Fodor, J. A., (1981) *Psychosemantics*, Cambridge: MIT Press
- Fodor, J. A., (1990) *A Theory of Content and Other Essays*. Cambridge: MIT Press
- Godfrey-Smith, P., (1996) *Complexity and the Function of Mind in Nature*, Cambridge: Cambridge University Press
- Kripke, S. A., (1982) *Wittgenstein on Rules and Private Language*, Oxford: Basil Blackwell
- McGinn, C., (1989) *Mental Content*, Oxford: Basil Blackwell
- Millikan R. G., (1984) *Language, Thought, and other Biological Categories*, Cambridge: MIT Press
- Millikan, R. G., (1993) *White Queen Psychology and Other Essays for Alice*, Cambridge: MIT Press
- Neander, K. (1991a) "Function as Selected Effects: The Conceptual Analyst's Defense", *Philosophy of Science*, 69, 454-468, reprinted in Allen, Beckoff and Lauder (1998)
- Neander, K. (1991b) "The Teleological Notion of "Function"", *Australasian Journal of Philosophy*, 69, 454-468, reprinted in Buller (1999)
- Neander, K. (1995) "Misrepresenting and Malfunctioning", *Philosophical Studies*, 79, 109-141, Excerpt in Buller (1999)
- Papineau, D., (1987) *Reality and Representation*, Oxford: Basil Blackwell
- Papineau, D., (1993) *Philosophical Naturalism*, Oxford: Basil Blackwell
- Putnam, H., (1975) "The Meaning of "Meaning"" in *Mind, Language and Reality*

Putnam, H., (1981) *Reason, Truth and History*, Cambridge: Cambridge University Press

Putnam, H., (1988) *Representation and Reality*, Cambridge: MIT Press

Rey, G., (1997) *Contemporary Philosophy of Mind; A Contentiously Classical Approach*, Oxford: Basil Blackwell

Sterelny, K., (1990) *The Representational Theory of Mind*, Oxford: Basil Blackwell

Stitch, S. P. and Warfield, T. A., eds., (1994) *Mental Representation: A Reader*, Cambridge: Blackwell