

지향성 - 문제와 접근*

Naturalizing Intentionality

민찬홍†

ChanHong Min

요 약

자연화 과정은 기계적 유물론을 가정한다는 점을 지적하고 지향성의 문제가 무엇인지 설명한다. 이 문제에 대한 이론들로서 취칠랜드의 제거주의, 행동주의적 제거주의, 대벳의 도구주의를 설명하고 비판한다. 다음으로 지향성 실재론을 포더와 드레츠키를 중심으로 설명한 후, 드레츠키의 자연화 프로그램이 오류상이 발생하는 적어도 하나의 매커니즘을 드러내준다는 점에서 포더보다 더 설득력 있는 접근이라고 논증한다.

주제어 지향성, 자연화, 표상, 정신, 철학

ABSTRACT

Beginning with the historical background under the light of which the program of naturalization proves to presuppose the mechanistic materialism, this paper first exposes what is the problem about intentionality, and then explains and examines the approaches

† 한남대학교 철학과 부교수

대전시 대덕구 오정동 133

e-mail:chanhong@eve.hannam.ac.kr

* 이 글은 한남대학교 교비 연구비의 지원에 의해서 작성되었다. 이 글은 1996년 2월 한국 인지과학회에서 주관한 <제1회 인지과학 학술 강연회>에서 읽었으며, 5월은 <철학적 자연주의>(한국분석철학회 편, 1995)에 실린 줄고 "자연주의와 심적 표상"의 일부이기도 하다.

to the program of naturalization of intentionality. Churchland's eliminativism and Dennett's instrumentalism are examined critically. Intentional realism is argued for; however, Dretske's theory is preferred to than Foder's because the former but not the latter reveals at least one of the various possible mechanisms of misrepresentations.

Keywords mind, intentionality, representation, naturalization, philosophy

1. 기계적 유물론의 재등장

일찍이 데모크리토스의 기계론에 대해서 “목적(telos)에 호소하지 않고서는 지능적인 동물들의 행동, 특히 사람의 행동을 설명할 수 없다”고 지적하면서 비판을 가했던 것은 아리스토텔레스였다. 나무를 베어서 깎고, 돌을 나르는 등 분주히 오가는 사람들이 “왜 그렇게 움직이는가?” 하는 물음에 대해서 그들의 다리가 움직인 것은 그 다리의 운동 신경의 어떤 전기-화학적 변화가 다리의 어떤 근육을 수축시켰기 때문이고 또 그들의 운동 신경이 그러한 전기-화학적 변화를 겪은 것은 중추 신경의 어떤 어떠한 변화 때문이고 하면서 그들의 행동의 원인, 동작의 메커니즘을 아무리 정교하게 추적해 간다고 하더라도 우리는 그들이 왜

그렇게 움직이는지 이해한 것이 아니다. 그들의 행동은 “집을 지으려고”라는 대답을 들어야 이해된다. 아리스토텔레스는 “완성된 집”이라는 행동의 끝점(telos)이 주어져야 그 행동들이 이해될 수 있다는 점에서 행동이란 인과적 질서를 넘어서는 목적론적 질서에 의해서만 제대로 파악된다고 생각하였던 것이다.

이 때에 완성된 집은 인간을 포함한 세계의 사물들에 영향을 가하며 그것들을 바꾼다. 그런데 실재하는 사물들에 작용력을 갖는 것은 실재한다고 해야 하지 않겠는가? 그러므로 목적과 수단, 끝점과 거기에 이르는 과정이라는 목적론적 질서는 세계 속에 실재하는 질서라고 해야 한다. 물론 원인과 결과라는 인과적 질서 또한 실재하는 질서이다. 아리스토텔레스는 세계를 이루는 두 가지 질서를 보았던 셈

이다.

근세에 뉴우튼과 라플라스에 의해서 기계론이 부활했을 때 사람들은 뉴우튼의 그 놀라운 세계관과 정교한 설명 장치에 매혹되었다. 그들은 뉴우튼이 그려준 세계의 모습에 매료된 나머지 아리스토텔레스를 (어쩌면 신까지도) 기꺼이 버렸고 사람의 지적인 행동에 관한한 문제가 그리 간단치 않다는 그의 충고를 잊었다. “나에게 초기 조건을 주면 우주의 전 과정을 보여주겠노라”던 라플라스의 말은 아마도 맞을 것이다. 인간도 그 우주의 작은 일부인 다음에야 제가 어찌됐는가? 아마도 인간의 사고, 인간의 행동이란 것도 완성된 물리학, 목적론을 배제한 이 새로운 기계론의 정교한 이론에 의해서 결국 그 비밀을 모두 벗어 버리리라. 제까짓 게 조금 복잡하기밖에 더 하겠는가 말이다.

실제로 뉴턴-라플라스의 기계론은 매력있는 것이기는 했어도 그들의 이론은 물리적 대상들의 물리적 속성을 다룰 뿐이었다. 물리학과 천문학, 화학의 영역에서 한 걸음 벗어나서, 생명 현상에 주목하면 목적론적-기능적 설명에 호소하지 않고서는 설명이라는 게 이루어지지 않는 듯이 보였다. 나방의 애벌레는 “연한 나뭇잎을 먹으려고” 나무의 꼭대기로 기어 올라가며, 나무의 뿌리는 “물을 찾아서” 물가로 뻗고, 심장은 “피를 순환시키기 위해서” 박동한다. 생명 현상의 설명을 목적에의 호소로

부터 구제한 것은 다아윈이었다. 다아윈은 생물학에서의 목적론적 설명이란 사실은 가장된 (disguised) 역사적 설명임을 보여 주었다. 목적에 대한 호소란 것은 알고 보면 그 생물 종의 기나긴 적응의 역사를 간략하게 줄인 약식 화법이라는 것이다.

다아윈 덕분에 생물학이 목적론으로부터 완전히 벗어났다고 쳐도 기계론의 승리가 바로 얻어지는 것은 아니다. 헤겔을 통해서 부활한 아리스토텔레스는 다시 인간은 어떻게 된 거냐고 묻는다. 인간을 비롯한 고등 동물들의 지능적 행동에 관한한 기계론은 약효가 없지 않은가 말이다. 그리하여 아리스토텔레스-헤겔의 전통을 이어받은 일군의 학자들은 설명을 목표로 하는 과학, 법칙의 체계를 만들려는 과학과 이해를 목표로 하는 과학, 사고의 기술을 목표로 하는 과학을 구별한다. 전자는 기계론적 과학이요 후자는 목적론적 과학이다. 물론 목적론적 과학의 주제는 인간이다. 인간 행동을 주요 대상으로 삼는 인간 과학, 사회 과학에 관한한 기계론 철학은 통하지 않는다고 그들은 주장한다. 그들에게 있어서 인간은 여전히 “목적의 왕국”의 주인이며, 아리스토텔레스 선생의 충고는 여전히 타당한 것이다.

20세기 중반 이후 컴퓨터의 등장과 사람을 어지럽게 하는 그 엄청난 발달은 현대 철학자들로 하여금 인간도 기계이며 그점은 물리학의 발달에 의해서 곧 밝혀질 거라는 근세 유물

론자들의 주장으로 되돌아 가게 만들었다. 최근 심리철학의 논의들은 대부분 컴퓨터에서 그 문제와 통찰을 빌어 온 것이다. 컴퓨터에 관한 그 작동의 메커니즘을 모르는 사람도 그것이 물질의 덩어리요 기계라는 점을 의심하지는 않는다. 그러나 인간 정신의 비밀이 컴퓨터의 개념 장치에 호소해서 설명된다면, 인간과 같은 지능적 행동을 보이는 기계가 만들어진다면 아리스토텔레스 선생께 드릴 말씀이 생긴다는 얘기가.

이러한 맥락에서 보면 인지과학이란 뉴우튼이 남겨둔 기계론의 마지막 미개척지에 대한 도전이요 목적론의 마지막 요새에 대한 공격이다. 인지과학의 목표는 인간 정신의 과학적 해명이다. 물론 이때의 '과학적'이란 해석학적(hermeneutic) 과학을 의미하는 것이 아니다. 인간 정신의 메커니즘을 해명하자는 것이요, 지능적(인간) 행동의 인과적 발생학(etiology)을 하자는 것이다. 이런 점에서 인지과학의 이념 내지 이데올로기는 분명하다. 기계론적 유물론인 것이다. 물론 아직까지 인지과학은 기계적 유물론을 증명한다기 보다는 그것이 증명될 수 있으리라고 약속하는 것에 불과하지만 말이다.

그동안 남한의 이념적으로 편향된 정치와 교육은 '유물론'이라는 말이 듣는 사람까지 두드러기를 일으키게 만들었는데 이제는 '유물론'과 '동무'라는 말을 되찾을 때가 되지 않았을

까. 이제는 총을 들고 위협하는 아저씨 앞에서 "아저씨 살려주세요"라고 울지 않고 "나는 공산당이 싫어요"라고 부르짖었던 이승복 어린이는 왜곡된 이념에 의해 세뇌된 불행한 아이였다고 말할 때가 되지 않았을까. 나는 심신 문제에 관한 맑스주의의 이른바 변증법적 유물론이 진짜 유물론인지가 의심스럽다. 변증법적 유물론자들은 기계적 유물론이 매우 소박한 유물론이어서 극복되어야 할 사고 방식이라고 공격하였었는데, 자신들의 유물론은 실상은 발생적 기원에 있어서 유물론일 뿐 정신이 물질과 질적으로 다른 것이라고 인정하고 있는 듯 하기 때문이다. 정신이라는 물질 아닌 어떤 것이 물질의 진화 과정을 통해서 생겨났으되, 량의 차이가 질의 차이를 낳는다는 것이니 정신의 발생은 말하자면 진화 상의 비약이다. 일단 물질의 진화 과정을 통해서 탄생한 정신은 이미 물질이 아닌 것이다.

나는 이러한 발생적 유물론에 대해서 유감이 있기 때문에 이점에 대해서 조금 더 얘기 하겠다. 최근에 우리나라 분석철학계에서 창발론(emergentism)이라는 심신 이론이 자주 논의되었다. 창발론의 최초의 형태는 화합물들의 화학적 속성이 그것을 이루는 물질의 물리적 속성에서 발생하였으되 그 물리적 속성들 가지고서는 설명되지 않으며 따라서 화학적 변화들은 물리적 속성에 대한 지식을 가지고서는 예측되지 않는다는, 말하자면 화학적 속성

의 설명적 독립성의 주장이었다고 알려지고 있다. 물론 이런 형태의 창발론은 양자역학의 등장으로 해서 폐기되었다. 고체의 딱딱함(solidity)이 구성 입자들의 격자 구조(lattice structure)에 의해서 설명되자 화학의 창발론은 설 자리를 잃었던 것이다. 그 다음 "생명 현상에 대한 창발론"은 분자 생물학의 등장으로 폐기되었다고 여겨진다. 이제 과학 사상가들이나 철학자들은 자신이 생기론자라고 오해받는 것을 두려워한다.(현장의 과학자들은 너무 바빠서 세계관의 문제에 대해서 고민할 시간이 없기 때문인지 이런 문제에 대해서는 "네 마음대로 생각해라"는 태도를 취하는 경우가 많지만.) 이제 정신의 창발론자들을 정신 현상이야말로 물질의 진화 과정에서 발생하였으되 그것의 작동 방식은 물질을 다루는 법칙에 의해서 설명되지도 않고 예측되지도 않는다고 말한다. 정신이 물질과 질적으로 다르다면, 그래서 물질의 대사만 가지고서는 정신의 현상들이 설명될 수 없다고 여긴다면 그것은 적어도 "지금"이라는 시점에서 보면 심신 이원론과 다름없지 않은가. 그렇다면 변증법적 유물론과 창발론은 발생론적으로 볼 때에만 유물론이고 공시적(synchronic), 설명적으로는 유물론이 아니다.

그런데 발생적인 유물론이 충분한 유물론인가? 모든 형이상학이 그렇듯이 유물론도 하나의 과학적 탐구의 강령 내지 프로그램이라고

이해한다면 발생적 유물론은 결코 충분하지 않다. 어떤 물리적 속성들이 있다고 하자. 우선 존재론의 관점에서 유물론은 세계의 모든 속성이 이 물리적 속성들에 의존한다고 간주한다. 그리고 이러한 존재론적 태도는 유물론자들에게 탐구의 준칙을 제시한다. 세계의 모든 속성들이 물리적 속성에 어떻게 의존하는지 설명하라는 것이 그 준칙이다. 말하자면 존재론으로서의 유물론은 설명적 유물론, 방법론적 유물론을 요구한다.

존재론적 의존만 가지고 말한다면 발생적 의존도 일종의 의존임에 틀림없다. 또 발생적 의존만을 인정하는 유물론에서 일종의 설명적 유물론을 도출할 수도 있다고 혹자는 말할지도 모르겠다. 그러나 그때에 설명이란 오직 역사적 기술(historical narratives)에 불과하게 될 것이다. 그러한 유물론은 현재 세계를 이루고 있는 속성들이 구성 요소들의 속성에 의해서 공시적으로 설명될 것을 요구하지 않을 것이다. 그러므로 발생적 유물론은 공시적으로는 존재론적 다원주의, 설명적 다원주의를 인정한다. 따라서 정신적 속성과 물리적 속성의 이원성은 발생적 유물론자들에게는 걱정거리가 아니게 된다. 그러나 심신 이원성을 걱정거리로 간주하지 않는 유물론이 정말 유물론인가?

유물론의 가장 단순하고 설득력 있는 형태는 환원적(reductive) 물리주의, 즉 환원적 유물

론에서 찾을 수 있다. 환원적 물리주의는 모든 대상의 모든 속성들이 궁극적으로 물리적 속성에 의해서 환원적으로 설명되리라고 기대한다. 수반 이론(supervenience theory)은 어떤가? 수반 이론은 모든 속성이 물리적 속성에 의존하리라고 보는 점에서 존재론적인 유물론의 요구에 들어 맞는다. 그러나 수반 이론은 설명적 환원의 전망에 대해서 회의적이다. 그러므로 수반 이론은 방법론적 함축이 없는 유물론이라고 할 수 있다. 그럼에도 불구하고 수반 이론이 유물론이라고 불리울 수 있는 것은 세계를 이루는 모든 비물리적 속성이 공시적으로 물리적 속성에 의존한다는 입론 때문이다. 창발론은 이 점이 분명치가 않다.

2. 지향성

최근 심리철학의 문제의 기원을 거슬러 올라간다면 아리스토텔레스를 만나기 훨씬 전에 데카르트와 만나게 된다. 데카르트는 방법적 회의라는 하나의 사고 실험을 통해서 인간 정신의 표상(representation)들과 그것이 표상하는(나타내는, 가리키는) 대상들과의 연관을 끊어 놓았다. 대상들이 전혀 존재하지 않는더라도 정신의 표상들은 그대로 있을지도 모른다는 것이 그의 회의의 요체였던 것이다.

데카르트가 정신과 물질의 인과적 상호작용을 믿었으면서 동시에 정신의 표상들이 물질

적 대상과 아무런 관련도 없을 수 있으리라고 생각했다는 것은 참으로 이상한 일이다. 내 마음 속의 '저 책상'이라는 생각(=표상)은 왜 저 책상에 대한 표상인가? 그것은 저 책상이 지각을 매개로 해서 나의 '저 책상'이라는 심적 표상의 원인이 되기 때문에 그런 것 아닌가? 그렇다면 '저 책상'이라는 나의 심적 표상이 바로 저 책상의 표상인 것은 그것이 저 책상에 의해서 야기된(caused) 것이기 때문이다. 표상의 대상이 표상의 동일성 규정에 들어온다. 그렇다면 내 마음 속의 표상들이 그대로 이면서 물질적 대상들이 전혀 존재하지 않게 될 수는 없다. 물질적 세계가 존재하지 않는 순간 내 표상들은 이미 동일성을 유지할 수 없고 따라서 그대로일 수 없는 것이다.

하기야 데카르트의 생각 중에는 이보다 훨씬 이상한 것도 있다. 유명한 얘기지만 그는 심신이원론을 옹호하였고, 심신의 인과적 상호작용을 인정하였다. 정신과 물질은 서로 독립된 별개의 실체들이지만 서로 인과적으로 상호작용한다는 것이다. 그런데 그는 동시에 모든 물리적 사태의 모든 원인은 또한 물리적 사태들이라는 것도 믿었다. 물리적 사건, 사태들의 원인 계열을 추적한다면 우리는 언제나 물리적 사건들을 만나리라는 것이다. 물리적 세계의 인과적 폐쇄성(causal closure)이라고 불리는 이 원칙은 세계 속의 모든 사건들은 물리학의 근본 법칙에 포섭되는 사건, 대상들이라는

물리주의(physicalism)의 대전제라고 할만한 것이며, 당연히 유물론의 세계관과 가장 잘 어울리는 것이다. 물리적 세계가 인과적으로 닫혀 있지 않다고 믿는다는 것은 물리적 대상들의 원인을 추적하다 보면 물리학의 개념이 적용될 수 없고 물리학의 법칙에 포섭시킬 수 없는 대상들을 만날 수 있다고 생각하는 것이며, 물리학을 완성하기 위해서 물리학 아닌 학문, 예컨대 심리학을 하지 않으면 안된다고 생각하는 것이다.

심신 이원론, 심신 상호 작용, 그리고 물리적 세계의 인과적 폐쇄성, 이렇게 세 가지 생각이 서로 정합적이지 않은 게 아니냐는 엘리자베스(?) 공주의 질문 편지를 받고서 데카르트가 대답하느라고 찢찢매었던 것은 당연하다. 이 세가지는 서로 정합적이지 않은 것이다. 만일 물리적 폐쇄성이 옳다면 “‘아아!’ 하고 소리지름”이라는 물리적 사건의 원인을 추적해서 만나게 되는 “통증을 느낌”이라는 사건은 당연히 물리적 사건이려야 한다. 우리가 정신의 과정, 정신의 표상이라고 부르는 모든 심리적 사건과 대상들은 물리적 사건, 물리적 대상이려야 하는 것이다. 물리 세계가 인과적으로 닫혀 있고 동시에 우리가 정신이라고 부르는 것들이 물질과 인과적으로 상호 작용한다면 그 논리적 귀결은 유물론 뿐이다.

그럼에도 불구하고 데카르트 이후 심리 철학의 역사는 유물론에 대한 반례와 반론들로 점

철되어 왔다. 심적 상태는 내성(introspection)에 의해서 직접 알려지지만 두뇌 상태는 내성에 의해서 알려지지 않는다는지, 모든 물리적 사건은 공간적 위치를 갖지만 심적 사건은 위치를 갖지 않는다는지, 심적 상태에 대해서는 의미론적 평가가 가능하지만 물리적 상태에 대한 의미론적 평가는 넌센스라든지 등, 이러한 문제들 중 어떤 것들은 관련된 개념들을 적절히 재해석함으로써 의해서 답변되는 것들도 있고 또 어떤 것들은 상당한 체계를 갖춘 이론을 제시하지 않고서는 대답되기 어려운 것들도 있다.

어쨌거나 이러한 문제들은 모두 유물론자들이 대답해야 하는 유물론의 과제로 파악될 수 있겠는데, 이러한 문제들 중에서 가장 포괄적이고 가장 심층적인 문제는 브렌타노(F. Brentano)의 “지향성(intentionality)”이라는 개념으로 집약된다. 브렌타노는 지향성을 정신적 현상과 물리적 현상을 구별해주는 특징으로 삼았다고 알려져 있다. 그에 따르면 정신 현상은 대상 지향(directedness upon an object)과 내용 연관(relatedness to content)이라는 두 가지 특징을 갖는다.

대상 지향: 원한다는 것은 무엇인가를 원하는 것이고, 상상한다는 것은 무엇인가를 상상하는 것이며, 두려워한다는 것은 무엇인가를 두려워하는 것이다. 그런데 이때에 원하고, 상상하고, 두려워하는 대상들은 만진다거나 들

어울린다거나 위에 앉는다거나 등의 물리적 동작의 대상들처럼 존재할 필요가 없다. 예컨대 누가 아내를 원한다고 해서 그로부터 그의 아내인 어떤 대상이 존재한다는 결론도 끌어낼 수 없으며 그가 결혼하고 싶어하는 어떤 특정한 여자가 있을 필요도 없다. 또 내가 빛보다 빠른 우주선을 원한다고 해서 그로부터 빛보다 빠른 우주선의 존재가 나오지는 않는다. 그러나 그가 아내를 때린다거나 내가 우주선을 탄다면 그가 때리는 아내는 있는 것이고 내가 타는 우주선도 있는 것이다. 그러므로 원함의 대상은 브렌타노가 말한대로 지향적 비존재(intentional inexistence/지향적 내재?)를 갖는다고 할 수 있다. 또 '상상함', '기억함', '바람' 등에 대해서도 마찬가지로 말할 수 있다. 나는 제우스를 상상할 수 있고 죽은 사람을 기억할 수 있고 AIDS의 치료법을 바랄 수 있지만 그 어느 것들도 존재할 필요는 없다.

내용 연관: 한편 "귀신"을 믿는 대신 "귀신이 존재한다"고 믿을 수도 있다. 마찬가지로 우리는 "...임을 상상 할 수도 있고 "...이기를 희망 할 수도 있다. 이 때에 정신 현상은 어떤 대상을 지향하고 있다기 보다는 어떤 명제와 연관되어 있다고 말해야 할 것이다. 이때의 정신 현상은 정확히 말해서 - 명제의 illocutionary force와 구별되는 - 명제적 내용(propositional content)과 연관되어 있다. 지향적 심리 상태들 중에서 이렇게 명제적 내용

과 연관되어서 규정되는 태도를 명제 태도(propositional attitude)라고 한다.

지향성이란 "향해 있음", 무엇인가에 "관한 것임(aboutness)"을 뜻한다.

나의 마음의(혹은 두뇌의) 어떤 상태 x 가 예컨대 서태지에 관한 생각이라고 하자. 이 때 (x , 서태지)라는 관계는 지향적 관계이다. 지향적 관계의 개념에 대해서 주목할 것이 두 가지가 있다. 첫째는 앞서 말하였듯이 지향적 관계를 이루는 순서쌍의 후건이 반드시 존재하는 것일 필요가 없다는 점이다. 물리적 관계에 있어서는 이럴 수가 없다. 모든 물리적 관계 $R(x,y)$ 의 두 항 x, y 는 어떤 실재하는 대상을 가리켜야 한다. 바로 이점 때문에 브렌타노는 지향적 관계가 물리적 관계로 환원될 수 없다고 했던 것이다.

둘째는 지향적 관계를 이루는 쌍의 전건이 언제나 정신의 상태일 필요는 없다는 점이다. 예컨대 '서태지'라고 말할 때 나는 소리는 물리적 사건으로서 어떤 공기의 파동 유형(type)이다. 그러나 우리 말의 언어적 약속을 전제하면 ('서태지', 서태지)는 지향적 관계이다. 전건은 후건에 "관한 것"이다. 또 사진, 도표, 교통 신호 등도 모두 지향적인 대상들이다. 단어나 기호가 지향성을 가진다함은 그것들이 무엇을 의미함, 무엇을 지시함을 뜻한다. 어떻게 해서 <의미함>이라는 현상이 가능한가고 물었던 언어철학 내지 의미론의 물음은 실상은 어

편 물리적인 대상이 어떻게 해서 지향성을 가질 수 있는가 하는 물음이었던 것이다.

흔히 지향성을 갖는 이러한 대상들을 통틀어 표상(representation)이라고 불러 왔다. 브렌타노는 물론 언어적 표상의 지향성은 문제 삼지도 않았는데 그것은 그를 포함한 많은 사람들이 언어적 표상들은 단지 파생적으로만 지향적이라고 생각하였기 때문이다. 그것들의 지향성은 그것들을 만들고 사용하는 정신을 가진 존재에 달려 있으며 그것들 개개의 표상들은 그것들을 만든 자들의 관념, 믿음, 의도 등의 특정한 정신적 표상에 의해서 결정되는 것이다. '서태지'가 서태지를 가리키는 것은 이 단어가 서태지에 대해서 알고 있고 그에 대해서 의사 소통하기를 원하는 사람들에 의해

서 사용되는 방법에 달려 있다는 것이다. 이런 식으로 생각하면 지향성의 진정한 원천은 정신뿐이다. 이런 견해에 따르면 문장들, 그림들, 도표들과 온갖 기호들은 결국 그것을 사용하는 사람들의 정신을 확대하는 도구이며 그 자체로는 어떤 본래적 의미도 없고 단지 그것들이 사용됨으로 해서 의미를 얻어낼 뿐이다.²⁾

그렇다면 지향성의 원천은 역시 정신이며, 지향적 현상 중에 가장 중심적인 것은 믿음, 욕구 등의 심적 상태들이다. 물론 모든 심적 상태가 지향적 상태인 것은 아닐지도 모른다. 예컨대 초기 실존주의자들은 대상을 갖는가 그렇지 않은가를 기준 삼아서 공포와 불안을 구별하였다. 이 구별을 역면 그대로 받아들이면 공포란 지향적인 상태이지만 불안이란 비

2) 이런 생각은 언어적 표상의 근거에 관한 매우 그럴법한 견해이기는 하지만, 이 견해를 따라서 지향성을 유물론적으로 해명하는 것은 쉬운 일이 아니다. 심적 표상의 지향성이 원초적이라는 생각은 써얼(J.Searle)과 그라이스(P.Grice)에서 볼 수 있다. 특히 언어적 표상의 근거를 정신적 표상에서 찾으려는 이러한 시도는 그라이스의 지향적 의미론(intentional semantics)에서 제시되었고 따라서 "그라이스 프로그램(Gricean Program)"이라고 불리운다. 언어적 표상의 지향성(의미)을 정신적 표상의 지향성으로 설명하되, 심리 상태가 갖는 표상 자체는 언어적 표상의 지향성에 호소하지 않는 방식으로 설명이 이루어져야 한다는 것이 바로 그것이다. 예를 들어서 '서태지는 은퇴를 선언하였다'라는 문장의 지향성

이 이 문장을 발화하고 이해하는 사람들의 믿음과 의도에 의해서 설명되는 경우를 생각해 보면 문제가 좀더 분명해진다. 이 경우 그 사람들의 믿음이나 의도란 어떤 것이겠는가? 그 믿음이나 의도가 문장이 공적으로 갖는 의미에 호소하지 않고 어떻게 규정될 수 있겠는가? 앞서 언급하였듯이 믿음, 의도, 욕구 등의 심리적 태도들은 명제적 내용을 갖기 때문에 지향적이라고 여겨진다. 이러한 심리적 태도들은 그것이 어떤 명제를 내용으로 갖는가를 규정함으로써만 그 동일성이 확인된다는 말이다. 그렇다면 이 심리적 태도들이 내용으로 갖는 명제들을 어떻게 선결문제를 요구하지 않으면서, 즉 공적인 언어의 의미에 호소하지 않고 규정할 것인가 하는 문제가 해결이 쉽지 않은 것이다.

지향적 심적 상태라고 할 수도 있다. 물론 불안이란 무(無)를 대상으로 한다고 말하는 실존주의자들도 있지만 말이다. 감각질(qualia)의 문제를 논할 때 감초처럼 끼는 '통증'이란 것도 원인은 가지지만 대상을 갖지는 않는다고 볼 수 있으니 역시 비지향적 심적 상태라고 할 수 있겠다.

모든 심적 상태가 지향적 상태가 아니라면 지향성의 문제가 정신의 문제 전체를 포괄하는 것은 아니지 않은가? 그렇다. 그래서 현대 심리철학은 정신의 문제를 크게 감각질의 문제와 지향성의 문제로 나누곤 하였다. 그러나 그렇다고 하더라도 지향성이야말로 인간을 인간이게끔 하는 것이라는 점은 분명하다. 고대의 철학자들이 인간이 자연과 구별되는 존재임을 강조하였을 때 그들이 염두에 둔 것은 인간의 행동은 이유(reasons)라는 내적 원인에서 비롯된다는 점이었다. 인간 행동의 이유가 무엇인가? 그것은 바로 믿음과 욕구이다. 냇물은 낮은 곳으로 가기를 원해서, 그리고 바다가 낮은 곳이라고 믿어서 그리로 흘러가는 것이 아니다. 그러나 인간은 집을 짓기를 원하고 집을 지으려면 나무가 필요하다고 믿기 때문에 나무를 베다. 믿음과 욕구라는 내적인 행동 발생 장치로부터 나오지 않은 동작은 행동이라고 부르지도 않는다.

실제로 데모크리토스에게 아리스토텔레스가 "그렇게 해서는 인간 행동이 설명 안된다"고

말하면서 보여주려 했던 목적론적 질서란 지향적 질서에 다름 아니다. 또 다아윈이 목적론에 호소하는 것처럼 보이는 생물학의 설명이 도태와 적응의 진화사 얘기로 해석될 수 있다고 보여주었을 때, 그는 생명 현상에 관한 지향성에의 호소는 불가피한 것이 아님을 보여주었던 것이다. 조금 과장해서 말하면 미신으로부터의 해방의 역사로서 과학의 발달사란 지향적 관점으로부터의 해방의 역사라고 할 수도 있다. 모든 자연 현상에 지향성이 침투해 있는 것으로 보던 어린 아이가 자라면서 "거기에 지향성은 없었다"는 점을 깨달아 가듯이 말이다.

흔히 믿음과 욕구를 부여해서, 즉 지향성을 부여해서 인간 행동을 설명하는 설명 방식을 통속 심리학(folk psychology)이라고 한다. 철학자들이 통속 심리학의 과학적 건전성을 얘기할 때 그는 지향성에 대해서 말하고 있는 것이다. 혹자가 통속 심리학의 언어들이 두뇌 상태를 가리키는 언어로 환원되리라고 말한다면 그는 브렌타노의 문제가 유물론의 틀 안에서 해결되리라고 말하고 있는 것이다. 또 누군가가 통속 심리학이 조만간 발전된 인지 과학에 의해서 대체되고 과거의 미신으로 전락하리라고 말한다면 그는 지향성이란 것이 실재(reality)에 기반을 둔 것이 아니라고 말하고 있는 것이다.

어쨌거나 진정한 유물론자라면 지향성을, 지

향적 언어를 그대로 용납해서는 안된다. 지향성 내지 지향적 관계는 그 자체로는 유물론과 양립할 수 없다. 만일 유물론이 옳다면 지향성은 궁극의 실재일 수 없다. "언젠가 물리학자들이 사물들의 궁극적이고 환원 불가능한 속성들의 목록을 완성하게 되면... spin, charm, charge 같은 것들은 그 목록에 있을 것이다. 그러나 aboutness는 틀림없이 거기에 없을 것이다. 지향성은 그렇게 깊은 차원의 것이 아니다. 이런 점을 고려할 때에 어떤 식으로든 환원주의자가 되지 않고서 지향성의 실재론자일 수 있으리라고는 생각할 수 없다. 의미론적 속성과 지향적 속성이 사물의 실재하는 속성이라면 그건 지향적이지도 의미론적이지도 않은 속성들과의 동일성 (또는 어쩌면 그것에의 수반?) 덕택일 수 밖에 없다. aboutness가 실재한다면 그것은 실제로 다른 어떤 것이다.(Fodor,J.,1987)"

3. 접근 1 : 제거주의

지향적 언어를 과학의 영역에서 추방하려 했던 행태주의의 실패는 지향적 언어가 그리 만만하게 물리칠 수 있는 게 아니라는 점을 분명하게 보여준다. 행태주의는 정신에 관한 이론으로서의 실패한 것으로 평가되고 있음에도 불구하고 지향성 및 그 관련 개념들의 연관성에 대한 분명한 이해를 보이고 있을 뿐 아니라

지향적 개념군 전체에 대해서 아주 일관된 입장을 취하고 있다. 그들은 브렌타노 논제를 액면 그대로 인정하고, 바로 그점을 이유로 해서 지향적 언어를 비과학적 언어라고 낙인 찍고 과학 언어의 영역으로부터 추방하려고 했던 것이다.

방법론적 행태주의자들은 지향적 언어를 비지향적인 언어로 대치함으로써 정신주의적(mentalist) 언어 즉 지향적 언어를 과학적 심리학에서 추방하려 하였으며, 철학적 행태주의자들은 심적 상태가 행동의 경향성이라는 자연적 속성에 의해서 정의적(definitional)으로 환원될 수 있다고 믿었다. 그러나 경향성 언어(dispositional terms)는 관찰적으로 정의하려면 반사실적(counterfactual) 조건문(=가정법 조건문)에 호소하지 않으면 안된다. 그리고 반사실적 조건문은 외연적 논리학과 관찰 조건들에 의해서 설명되기 어려운 조건문으로 알려져 있다. 우리는 아직 반사실적 조건문에 대하여 그 진리 조건을 알지 못하는 것이다. 또 행태주의적 환원이라고 제시된 것들은 실상 정신 언어를 완전히 제거한 것이 아니다. 예컨대 '제주도에서 휴가 보내기를 원한다'는 심적상태가 다음과 같이 설명되었다고 하자.

- (1) 민은 제주도에서 휴가 보내기를 원한다 =
(1-1) 만일 그게 원하는 바냐고 질문 받으면 그렇다고 대답한다. 또는

(1-2) 제주도 관광 안내 책자와 실악산 관광 안내 책자를 받으면 제주도 안내 책자를 읽을 것이다. 또는

(1-3) 제주도 비행기표와 숙박권을 얻으면 제주도로 갈 것이다....

그러나 이 정의에서 민이 자신의 휴가 계획에 대해서 비밀에 "부치려고 하지 않을 때"에만 (1-1)은 참이다. 또 제주도 안내 책자에 "싫증이 나지 않았을 때"에만 (1-2)가 참이다. 또 그가 제주도행 비행기가 납치되거나 폭파되지 않으리라고 "믿을 때"에만 (1-3)은 참이다. 그런데 이런 관련된 제한 사항을 끌어들이면 정의항에 다시 지향적인 심적 상태를 가리키는 말이 끼어들게 된 결과가 되고 만다.

한편 심리학이 과학적이기 위해서는 지향적 언어를 피하는 길 뿐이라고 믿었던 방법론적 행태주의자들에 대해서 데넷(D. Dennett, 1978)은 그 난점의 원천이 지향성의 직접적 제거가 불가능하기 때문이라는 점을 분명하게 지적한다.

행태주의 심리학자들은 실험되는 동물들이 상황에 대해서 보이는 반응을 지향적 언어의 도움없이 설명하고자 하였고 따라서 그들의 이론을 지지할 실험적 상황을 고안하느라고 애써 왔다. 스키너는 순전히 관찰적으로 기술될 수 있는 물리적 자극에

대한 관찰적으로 기술될 수 있는 신체적 반응을 예측하게 해주는 여러 실험들을 고안하였다. 일단 동물들이 훈련을 받으면 비지향적으로 기술된 두 사건, 예를 들어 통제하는 자극과 막대를 누르는 반응 사이에 법칙적 관계가 발견된다는 것이다. 물론 그런 규칙성은 성립할 것이다. 그러나 그것이 비지향적으로 정의된 두 사건 사이의 관계라는 점은 피실험자의 속성 때문이 아니라 스키너의 실험 장치 때문이라고 해야한다.... 만일 실험 상자 안에서 훈련받은 생쥐가 음식을 얻기 위해서 꼭 네 걸음을 옮기면 막대를 코로 누를 수 있도록 해놓았다고 해 보자. 만일 스키너의 법칙들이 정말로 물리적 언어로 표현된 자극과 신체적 동작으로서의 반응 간에 성립한다면 막대를 한뼘 뒤로 물러나서 놓아두면 네 걸음으로는 버튼을 누를 수 없을 것이다. 그러나 이런 경우에도 스키너는 생쥐가 네 걸음가서 허공을 코로 누르리라고 예측하는 대신 다섯 걸음 가서 버튼을 누르리라고 예측할 것이다.... 스키너는 단 한가지 행동만이 지각과 믿음과 욕구에 알맞도록 환경을 극심하게 제한해 놓음으로써 비지향적 예측이 이루어진 것처럼 보이게 만들고 있으나 사실 그것은 위장된 지향적 예측인 것이다.

행태주의적 심리학자들은 지향적 언어를 피하는 방식으로 동물과 사람의 행동을 설명은 커니와 단지 그것들을 기술하는데도 어려움을 피할 수 없었던 것이다. 지향성 내지 지향적 담화는 행태주의적으로 제거될 수는 없다.

취칠랜드(Churchland, P., 1978, 1981)는 행태주의자들 못지 않은 지향성 혐오증을 보여 준다. 취칠랜드는 지향적 개념들을 핵심적 설명 장치로 갖고 있는 통속 심리학이 거짓일지 모른다는 개연성은 이미 여러 군데에서 드러나고 있다고 주장한다. 그에 따르면 첫째, 통속 심리학은 설명적으로 매우 무력하다. 통속 심리학은 정신 질환, 창조적 상상력의 기능, 지능의 개인 차이, 잠의 본성과 심리적 기능, 기억의 신비, 학습 등에 대해서 아무런 설명도 주지 못한다. 심지어는 시지각(visual perception)과 운동 간의 조정 능력, 예컨대 움직이는 차를 눈덩이로 던져서 맞히는 혼란 능력에 대해서도 통속 심리학은 할 말이 없다. 그러므로 통속 심리학은 설명적으로 실패하였다. 통속 심리학은 심층적이고 복잡한 실재에 대해서 기껏해야 부분적이고도 매우 피상적인 접근 밖에는 주지 못한다는 것이다. 둘째로 통속 심리학의 2000년 역사를 생각해 볼 때 통속 심리학은 전혀 아무런 발전도 하지 못하였다. 통속 심리학은 지독하게 정체되어 있는 이론이라는 것이다.

취칠랜드는 통속 심리학은 마치 플로지스톤 이론이나 톨레미 천문학처럼 새로운 이론에 의해서 폐기되고 말 것이라고 주장한다. 플로지스톤과 천구란 존재하지 않는다. 마찬가지로 통속 심리학이 구제될 가능성이 없는 근본적으로 잘못된 이론만큼 통속 심리학의 핵심 개념인 지향적 개념들은 모두 장차 완성될 인지 이론에 의해서 대치되고 제거될 것이라는 게 그의 주장이다.

통속 심리학이 경험적 이론인가? 인간 행동을 포함한 지능적 행동에 대해서 통속심리학이 제공하는 설명이란 합리성에 호소하는 설명, 즉 규범적(normative) 설명으로서 경험적이고 기술적(descriptive)인 경험 과학의 이론적 설명과 구별되어야 하지 않는가? 전통적으로 합리적 관계는 인과적 관계보다는 논리적인 관계에 더 가까운 것으로 여겨져 왔지 않은가? 이점에 대해서 취칠랜드는 그의 유명한 수적 태도(numerical attitudes)의 비유를 들어서 대답한다. (Churchland, 1978) 전도체에 흐르는 전류 I 는 인과적으로 두 요소에 의해서 결정된다. 전도체를 통해서 전자를 흘려 보내려는 전기적인 힘인 전압 V 에 비례해서 증가하고 그 운동에 저항하는 전기 저항 R 에 반비례해서 증가한다. ($I=V/R$) 이점을 논리학의 표현을 빌어서 정식화하면

(2) $(x)(V)(R)[(x가 V의 전압을 받고 있$

다)&(x가 R의 저항을 준다)→(∃I)(x는 I의 전류를 가진다 & (I=V/R))

여기서 <저항을 갖는다>, <전압을 받는다>, <전류를 갖는다> 등의 술어들이 이른바 수적 태도를 표현하고 있다는 것이다. 정확히 말해서 변항 V, R, I의 자리에 숫자를 대입하는 술어 함수라는 것이다. 이 점은 통속 심리학이 주는 다음과 같은 법칙

(3) (x)(P)(Q){((x가 P를 믿는다) & (x가 (P→Q)를 믿는다))→(x는 Q를 믿는다)}

와 비교될 수 있다. 여기서 <믿는다>는 술어는 명제 태도를 표현한다. 정확히 말해서 변항 P, Q의 자리에 명제를 대입하는 술어 함수이다. 취칠랜드에 따르면 명제들의 논리적 관계란 수들의 산술적 관계와 마찬가지로 추상적 대상들 간의 객관적인 관계일 뿐이다. 명제 태도와 수적 태도의 차이는 수적 태도를 가지는 경험적 이론들이 경험적 규칙성을 포착하기 위해서 “수”라는 추상적 대상들 간에 성립하는 산술적 관계를 이용하는데 비해서, 통속 심리학은 인지적 존재의 상태들과 행동들 간에 성립하는 경험적 규칙성을 포착하기 위해서 “명제”라는 추상적 대상들 간에 성립하는 논리적 관계를 이용한다는 것 뿐이라고 한다. 따라서 명제 태도가 명제들 간의 논리적 관계를 반영한다는 것은 통속 심리학

이 비경험적이라는 점을 보여주는 것이 아니라 오히려 통속 심리학이 다른 경험적 이론들과 마찬가지로 높은 추상적 단계의 이론적 틀이라는 점을 보여줄 뿐이라는 것이다.

‘P라는 믿음’이 어떤 추상적 대상과 적절한 관계를 맺는 문제라고 생각하는 것은 ‘무게가 5 Kg 나감’이 어떤 추상적 대상과 적절한 관계를 맺는 문제라고 생각하는 것과 마찬가지로 하는 것이다. ‘x weighs very little’에서 동사 뒤의 표현이 부사적 역할을 하듯 ‘x weighs 5 Kg’에서의 ‘5 Kg’도 부사적 기능을 하고 있는 것이다. 차이점이란 후자의 방식으로 표현함으로써 동사를 더 정확하고 체계적이고 유용한 방식으로 제약하고 있다는 점이다. 취칠랜드는 ‘...는 P임을 믿는다’에서 ‘P임’도 복합 술어의 한 부분으로서 부사적 기능을 하고 있는 걸로 볼 것을 제안한다.

명제 태도의 부사적 변형설(adverbial modification theory)이라고 불리는 취칠랜드의 이러한 생각에 따르면 우리가 명제 태도를 갖는다고 해서 우리(의 머리) 속에 그 명제의 형식을 반영하는 무엇인가가 있어야 한다는, 즉 그 명제의 논리적 구조를 가진 무엇인가가 두뇌에 물리적으로 실현되어 있어서 그것이 우리의 사고에서 어떤 인과적 역할을 하리라는 귀결은 나오지는 않는다. 우리가 심적 상태를 특징지을 때 사용하는 명제 태도라는 술어에 포함되는 “명제”란 심적 상태에 대한

부사적 규정을 가해주는 기능을 하는 것에 불과한 것이다. 그러한 명제적인 부사적 제약은 심적 상태를 설명하고 예측하는데 명제들 간의 형식적인 관계를 사용될 수 있게 해주는데 이것은 마치 수적 태도 술어를 사용하는 물리학 이론이 수들 간의 형식적인 관계를 이용하는 것과 다를 바 없다는 것이다.

어떤 사람들은 취칠랜드의 제거주의가 일종의 “인식적 자살”이라고 반박한다. (L.Baker, 1988) 그들은 제거주의의 충고는 실천적 맥락에서 자가당착에 빠지리라고 지적한다. 같은 맥락에서 데넷과 데이빗슨은 지향적 담화를 포기하고서는 우리는 더 이상 자신을 “합리적 인간”으로, “도덕적 행위의 주체”로 볼 수 없게 될 것이라고 말한다. 그러나 이런 점들에 호소하지 않더라도 취칠랜드의 논점에는 석연치 않은 데가 있다.

취칠랜드가 통속 심리학의 설명적 무기력함을 지적하기 위해 들고 있는 예들에 대해서 말하자면, 아난게 아니라 통속 심리학은 그런 것들을 설명하지 못한다. 그러나 그점이 왜 통속 심리학에 대한 공격의 구실이 되는가? 취칠랜드가 지적하는 점들이 넓은 의미에서 심리학의 주제들이라는 것은 분명하다. 그러나 그것이 통속 심리학의 주제는 아닌 것이다. 통속 심리학은 행위자에게 어떤 심적 내용을 부여함으로써 그들의 의도적인 행위에서 드러나는 어떤 질서를 포착하고 설명한다. 그러나 취칠

랜드가 드는 예들은 합리적 행위에 속하는 것들이 아니다. 그리고 의도적이고 합리적인 행위에 관한 우리는 통속 심리학이 제공하는 개념 장치 이외에 어떤 대안도 가지고 있지 않다.

한 건물에서 불이 났다고 하자. 그 안에 있던 길동이는 불길은 보고 출구를 향해서 달리기 시작했다. 달리는 중에 그는 커피 자판기에 동전을 넣고 있는 갑순이를 보았다. 그는 건물에 불이 났다고 말하고는 계속 달린다. 길동이의 말을 들은 갑순이는 윗층의 사무실 동료인 병수에게 전화를 걸어서 건물이 불 붙었다고 말하고는 역시 출구를 향해서 달리기 시작했다. 두 다리에 기브스를 감고 있는 병수는 곧바로 다이얼을 돌린다. 우리는 이들 모두의 행동을 ‘건물에 불이 붙었다고 “믿고” 불 붙은 건물에서 나가려고 “의도”하고 그에 따라서 움직이고 있다’고 기술할 수 있다. 그러나 이러한 기술은 전형적인 지향적 기술이다. 문제는 이 세 사람의 행동에서 보이는 패턴은 지향적 기술을 통해서만 드러나는 것 같다는 점이다. 지향적이지 않은 어떤 기술도 세 사람의 행동이 공통으로 보여주는 패턴을 포착할 수 없다. 퍼트남의 말투를 빌자면 이들이 보여주는 같은 패턴이 물리적으로 상이한 수없이 많은 방식으로 실현될 수 있기 때문이다. 세 사람은 각각 다른 입력 자극에 의해서 움직였고 각각 다른 방식으로 움직였으므로 환경적 입력과

그들의 행태만을 비지향적으로 기술해서는 이들의 행동이 공유하는 패턴은 드러낼 수 없다. 또 그들의 중추 신경계가 아무리 비슷한 구조를 가지고 있다고 해도 그들이 받은 입력 자극의 물리적 차이 때문에 이들의 〈정보 처리〉가 신경적 차원에서 동일하리라고 볼 이유도 없다. 그러므로 비지향적인 어떤 신경 생리학적 기술도 이 패턴을 잡아낼 수 없다. 그러나 이들이 보여주는 패턴이 포착되지 않는다면 우리는 세계 속에서 벌어지는 중요한 규칙성을 놓치는 것이다.

4. 접근 2 : 데넷의 지향 체계론

지향성이 실재하는 속성인가에 대해서는 회의적이면서도 인간의 지능적 행동이 보여주는 이러한 지향적 질서를 놓쳐서는 안된다고 생각한 사람이 바로 데넷(1978, 1987)이다. 데넷은 어떤 대상에 믿음과 욕구, 희망, 공포 등의 지향성을 부여함으로써 그 대상의 행동이 설명되고 예측될 수 있을 때 그런 대상을 지향 체계(intentional system)이라고 부른다.

그에 따르면 어떤 대상이 지향 체계인가 하는 것은 그 대상 자체의 본성에 달려 있는 문제라기보다는 그 대상의 행동을 설명하고 예측하려는 사람의 전략 내지는 자세에 달린 문제이다. 그에 따르면 우리가 어떤 대상이 보여주는 행태를 설명하고 예측할 때에 구사할 수

있는 자세에는 세 가지가 있다. 첫째는 대상의 물리적 실제 상태와 자연 법칙에 대한 지식을 이용해서 설명-예측하는 물리적 자세(physical stance)로서 특정 대상의 물리적 상태에 근거해서 예측이 이루어진다. 이때 예측은 자연 법칙에 대한 우리의 지식을 적용함으로써 이루어진다. 그러나 컴퓨터를 다룰 때 물리적 전략을 취하는 경우는 거의 없을 것이다. 한 컴퓨터의 물리적 조성에 있어서 결정적인 변항들의 수는 아무리 단순한 컴퓨터에 있어서도 엄청나게 클 것이기 때문이다.

두번째로 대상의 물리적 조성은 무시하고 대상의 설계 구조에 호소해서, 즉 그 체계가 설계된대로 행동하리라는 가정하에서 설명-예측하는 설계적 자세(design stance)가 있다. 설계적 전략에는 여러 가지 형태들이 있을 수 있지만 모두 기능(function)이라는 목적-상대적인(목적론적인) 개념에 의존하고 있다는데서 일치한다. 즉 한 체계의 설계는 크고 작은 기능적 부분들로 나뉘며, 설계적 전략은 각각의 기능적 부분들이 제대로 기능한다는 가정하에서 예측을 끌어낸다.

그리고 마지막으로 대상에 믿음, 욕구 등의 지향적 태도를 부여함으로써 대상의 행동을 설명-예측하는 지향적 자세(intentional stance)가 있다. 체계의 합리성을 가정하는 이러한 자세는 여러가지 사정에 의해 성패가 좌우된다. 무엇보다도 여기서 합리성이란 목적

들의 체계와 일련의 제약에 비추어 볼 때의 최적의 설계(optimal design) 이상을 의미하는 것은 아니다.

이러한 예측이란 매우 취약하다. 이런 예측은 대상 체계의 목적과 제약과 정보에 관한 일련의 가정에 달려 있으며, 이러한 자세를 가지고는 원칙적으로 예측될 수 없는 합선같은 사태에 대해서도 무력하다. 설계적 자세가 고장에 대해 무력한 것과 마찬가지로 (어떤 고장도 없다는 가정에 의존하므로), 이러한 예측은 설계의 약점과 잘못에 대해서 무력하다 (최적의 설계라는 가정에 의존해 있으므로). 그렇기는 하지만 지향적 자세는 너무 복잡해서 물리적 자세는 물론 설계적 자세로도 행동이 설명, 예측되지 않는 체계를 다룰 때 우리가 의존할 수 있는 유일한 전략이기도 하다.

물론 아무런 체계에 대해서나 지향적 전략이 통하는 것은 아니다. 데넛은 "적합한 설계의 가정이 보장되는 대상들이" 바로 지향적 전략이 통하리라고 기대할 수 있는 대상들이라고 생각한다. 그러면 무엇이 적합한 설계를 보장하는가? 우선 자연 도태의 과정을 거쳐 살아남은 체계라면 우리는 그것을 지향 체계로 다룰 수 있다. "만일 진화가 제 몫의 일을 해 냈다면 지향적 예측은 쓸만하고 믿을만 할 것이다." 또 데넛은 체스 컴퓨터를 예로 든다. 지금의 체스 컴퓨터는 세계 챔피언급의 프로 선수와 비등한 경기를 할 정도라고 보도되고 있다.

이런 정도의 프로그램이라면 프로그래머 자신도 설계적 자세를 취하고서 그 컴퓨터를 상대하기 어려울 것이다. 마지막으로 학습 과정이 적합한 설계를 보장한다. 진화와 설계와 학습, 이 세 가지 과정은 지향적 자세가 통할 수 있도록 "적합한 설계"를 보장하는 것들이다.

스틱취(Stich, S., 1990)는 진화 생물학의 성과들을 토대로 진화가 최적의 설계를 보장하는 것은 아니라는 논증을 하고 있다. 소우버, 스틱취, 데넛의 진화화 합리성의 연관에 대한 논쟁은 현대 진화론에서 신다윈주의(Neo-Darwinism)와 중립주의(neutralism)의 논쟁이 철학에서 재연된 것이라고 볼 수 있다. 이 논쟁에 대한 자세한 논의는 이 글의 범위를 벗어나는 얘기가 되겠다. 그러나 예컨대 유전자 부동(genetic drift)이 다윈적 적응주의(adaptationism)에 대한 반례인가 하는 점이 불분명하다는 정도는 지적해 두어야겠다. 유전자 부동이 표현형(phenotype)으로 나타나는 한 그것은 자연도태의 메커니즘의 구속을 벗어나 있는 것이 아니다. 물론 어떤 유전자 변이는 표현형으로 나타나지 않으며 게다가 표현형으로 나타나는 유전적 특질들이라고 해도 그것들이 자연 도태를 통해서 완벽하게 걸러지지 않는다. 그러나 이점이 바로 진화가 최적의 설계를 보장하지 않는다는 것을 보이는 증거인가? 이 물음에 관한 대답은 <최적의 설계>가 무엇인가를 분명히 하지 않고서는 올

바로 얻어질 수 없다.

또 스티취(1983, 1990)는 데넛 식의 노선은 정신 박약자와 언어를 갖추지 못한 지향 체계와 또 우리와 아주 먼 낯선 체계를 설명해낼 수 없을 뿐만 아니라 사람들이 흔히 범하는 인식적 (=지각과 추리에 관련된) 잘못들에 대해서도 무력하다고 생각한다. 그는 여기서 한걸음 더 나아가서 지향적 내용에 호소하는 모든 이론은 이런 운명을 피할 수 없으며 따라서 적어도 심리학의 이론으로서는 포기되어야 한다고 주장한다.

스티취는 주로 심리학의 실험들을 예로 거론하고 있는데, '믿음 지속(belief perseverance)'이라고 알려진 실험에서 피실험자들은 처음에 어떤 정보를 받고 나중에 그 정보가 잘못되었음을 설명(debriefing)받는다. 처음에 심리 검사를 하고 피실험자가 수학과 음악 분야에 아주 적성이 낮다는 결과를 통보해 놓고는 나중에 다시 불러서 이전의 검사 결과가 완전히 잘못 처리된 결과였다고 설명해주는 식이다. 그런데 피실험자들은 처음의 통보에 의해서 얻은 믿음을 그대로 가지고 있는 경향이 있다는 것이다. 또 스티취가 자주 들고 있는 예로 '선택 문제(selection task)' (Stich, 1990)라는 실험에서 피실험자는 절반이 가려져 있고 나머지 반은 비어있거나 동그라미가 그려져 있는 네개의 답지를 받고, "왼편에 동그라미가 있으면 오른편에도 있다"는 문장이 참인지 알

려면 어떤 답지의 가려진 부분을 보아야 하겠는가 하는 질문을 받는다. 스티취는 128명의 대학생 피실험자를 대상으로 한 실험에서 겨우 5명만이 정답을 맞추었다는 결과를 지적한다.

그러나 데넛에 대한 스티취의 이런 반론은 실험 예에 대한 그의 해석에 문제가 있지 않나 의심이 간다. 믿음 지속의 경우에 대해서는 하아만(Harman, G., 1989)으로부터 적절한 대답을 찾을 수 있다. 하아만에 따르면 변화하는 세계에 맞추어서 믿음을 수정해 갈 때 따를 법한 원리로 두 가지가 생각될 수 있다. 첫째는 <적극적 침식의 원리(principle of positive undermining)>로서 P를 믿는 이유가 좋지 않다는 것을 보여주는 적극적인 이유가 있으면 P를 믿어서는 안된다는 것이요, 둘째는 <소극적 침식의 원리(principle of negative undermining)>로서 P라고 믿을 적절한 이유가 없을 때에는 언제든 P를 믿는 것을 그만두어야 한다는 원리이다. 전자의 원리는 정합론자의 원리이며 둘째의 원리는 기초론자(foundationalist)의 원리라고 할 수 있겠는데, 하아만은 믿음 지속이라는 현상은 실제로 우리가 믿음을 수정해 가는 메커니즘에 대한 설명으로서 기초론이 틀렸음을 보여줄 뿐이라고 지적한다.

소극적 침식의 원리는 믿음을 정당화 해주는 근거를 잃으면 우리는 바로 그 믿음을 버려야

한다고 말한다. 그러나 실제로 우리는 우리가 가진 믿음 대부분에 대해서 그 정당성의 근거들을 함께 가지고 다니지는 않는다. 그렇다면 소극적 침식의 원리는 우리가 믿고 있는 것 대부분을 믿지 말고 버리라고 충고하는 셈인데 이거야말로 어리석은 충고이다. 보통의 인간이 가진 정보 저장 및 처리 능력에 비추어 볼 때 믿음 지속은 오히려 우리가 매우 합리적으로 설계되었다는 것을 보여준다.

선택 문제에 대해서는 그라이스의 화용론적 분석(Grice, 1968)이 적절한 대답을 준다. 아빠가 딸에게 “말 잘 들으면 자전거를 사주마”고 약속했다고 하자. 논리적으로 볼 때 그 딸이 말을 잘 안 들었음에도 불구하고 아빠가 자전거를 사 준다고 해도 아빠가 약속을 안지켰다고 할 수는 없다. 전건이 만족되지 않을 경우에 후건은 어떻게 되더라도 전제 조건문은 참이라는 것이 조건문에 대한 논리학의 해석이기 때문이다. 그러나 어차피 자전거를 사 줄 예정이었다면 아빠가 도대체 왜 그런 약속을 할 필요가 있었는지 알 수 없지 않은가? 이 점을 그라이스는 정보량에 관한 준칙으로 정식화하고 있다.

1. 필요한 정보를 주라.
2. 필요한 이상으로 정보를 주지 말라.

이 두 가지 준칙은 기호논리학을 배우지 않은

대부분의 사람들이 조건문을 이해하는 방식을 잘 보여준다. 진리 함수로 표현된 조건문이란 실제로 일상적으로 사용되는 여러가지 조건문들의 다양한 의미의 어떤 (핵심적이라고 여겨지는) 한 부분만을 포착하고 있을 뿐 아니라 조건문들이 쓰여지는 맥락적 변수들로 부터도 추상되어 있다. 바로 이런 이유로 기호 논리학 입문 과정을 가르치는 선생들은 학생들에게 단순 조건(material conditional)이 왜 세대로 된 조건문인지 납득시키느라 애를 먹게 되는 것이다. 스티치의 선택 과제에서 피실험자에게 주어지는 문제는 조건문이 사용되는 화용론적 맥락을 무시하고 조건문의 진리 조건에만 주목하는 사람만이 정답을 낼 수 있도록 되어 있다. (정답을 맞힌 5명의 대학생은 아마도 기호 논리학을 들은 학생일 것이다.) 논리학 시간이 아니라면 하나의 조건문에 관계된 진리 조건과 화용론적 조건들을 분리해서 생각할 이유가 없는 것이다. 그러므로 스티치가 보통 사람이 늘 불합리하게 추리한다고 여길 증거의 하나로 제시하고 있는 선택 과제는 사람들이 불합리하게 추론한다는 것을 보여주지 못한다. 그것은 오히려 사람들이 쓸데없는 구별은 하지 않는다는 것을 보여줄 뿐이다.

지금까지 데넛의 지향 체계론을 소개하고 몇 가지 반론들로부터 그것을 옹호하는 투로 말하였다. 지향 체계론은 지향적 개념들의 연관에 대해서, 그리고 그것이 인간 행동의 설명에

서 하는 역할에 대해서 좋은 분석을 준다. 그러나 지향성에 대한 완성된 답은 아니다. 이 점은 데넷이 지향성이라는 것을 설명적 전략의 문제로 본다는 데서 금방 드러난다. 우리는 컴퓨터 프로그램이나, 진화 과정을 거치면서 살아남은 (인간을 포함한) 동물들의 행동을 설명하고 예측하는 데 지향적 전략을 택할 수 있다. 그런데 데넷은 우리가 지향적 전략을 택해서 설명하는 그 대상이 “본래” 지향성을 가지고 있는가 하는 문제는 아예 잘못 제기된 문제라고 본다. 지향성이란 그러한 대상들에 실재하는 것이 아니라 설명자인 우리의 눈에 붙어 있는 것이다. 그렇다면 대상들에 대하여 지향적 전략을 택하는 능력은 어디서 오는가? 우리들이 갖고 있는 지향성이란 것은 무엇이며 그 원천은 무엇인가?

5. 접근 3 : 지향성 실재론

지금까지 지향성 문제에 대한 취칠랜드와 데넷의 접근을 간략히 보았는데, 이 두 사람은 각각 지향성 제거주의와 지향성 도구주의를 대표하는 것으로 지향성의 심리적 실재성을 부인하는 편, 즉 반실재론의 편에서 있다고 할 수 있다.³⁾ 믿음과 욕구 등의 지향적 심적

상태에 대응하는 객관적으로 실재하는 물리적 사태가 있을 것 같지 않다는 생각들이다.

사실 우리의 언어는 아주 복잡하게 얽혀 있는 지향적 개념들을 가지고 있어서 예컨대 제인 오스틴의 소설에 나타나는 섬세한 심리 묘사들이 전부 실재하는 상태들을 가리킨다고 믿기는 어렵다. 그렇게까지 가지 않더라도 “비통함”과 “슬픔”, “참담함”, “애통함” 등등 비슷한 심적 상태를 가리키는 일상은 매우 다양하며 또 그 의미들이 조금씩 다른데 이런 일상이 각각에 대응하는, 객관적으로 구별되는 내적 상태들이 있다고 보기는 어렵다. 그러나 무엇을 욕구함, 믿음, 기대함, 두려워 함 등의 굵직한 구별들에 대해서도 그것들이 구별되는 객관적인 상태들과 대응하지 않는다고 보기도 또한 어렵지 않을까? 지향적 개념들 중 어떤 것들은 체계가 환경에 반응하는, 잘 알려진 특정한 방식들과 대응한다. 이런 경우에 환경과 반응을 매개하는 내적 상태들의 실재성을 부정하기는 힘들 것 같다.

지능적인 행동을 보이는 체계들이란 환경의 입력을 지각하는데, 지각한다는 것은 입력에 대응하는 내적 상태를 갖는다는 것이다. 그 체계들은 내적 상태를 매개로 해서 환경의 입력을 처리한다. 따라서 적어도 그 체계가 환경의 자극에 대해서 보이는 반응 패턴의 다양성은 그 체계가 누리는 내적 상태의 다양성을 반영한다고 보는 것이 자연스럽다. 그것이 자극에

3) 데넷은 이점에 있어서 분명치 않다. D. Dennett, 1991 참조.

대해서 보이는 차별적 반응의 종류만큼 구별되는 내적 상태들을 가질 거라는 말이다. 예를 들어 움직이는 작고 검은 조각과 움직이는 날벌레에 대한 개구리의 반응은 구별되지 않는다. 따라서 개구리가 이 두 가지 자극 각각에 대응하는 구별되는 내적 상태를 가진다고 생각할 수는 없다. 그러나 움직이는 작은 벌레와 자신을 덮치는 커다란 그림자에 대해서 개구리가 다른 반응을 보인다면 개구리는 이 두 자극 각각에 대응하는 구별된 내적 상태를 가져야 한다. 인간과 지능적인 동물의 경우에도 그들이 환경에 대해서 보이는 차별적인 반응들과 최소한 같은 정도의 다양한 내적 상태들이 실재할 것이다.

이것은 매우 자연스럽거나와 이렇게 생각하지 않고서는 어떻게 지능적 동물들의 행동들이 보이는 지향적 패턴을 포착할 수 있을지 상상하기 어렵다. "미래의 완성된 인지 과학"에 호소하는 취칠랜드의 경우 지금 우리에게 보이는 지향적 질서, 그 패턴이 미래에 어떤 모습으로 파악될지 상상할 수 없다는 것이 그에 대한 반론으로 먹혀들지 알 수 없으나, 지향적 질서라는 것을 인정하고 대상에 지향성을 부여하는 것이 설명적, 예측적으로 유용하다고 인정하는 데넛의 경우에는 이런 점이 꽤 강력한 반론을 만든다. 데넛은 지향성의 부여가 유용하다고 말하면서도 지향적 설명이 왜 통하는지 말할 수 없는 것이다.

지향적 심적 상태의 실재성을 인정하는 노선에서 지향성이 어떻게 유물론과 정합적일 수 있는지 설명하려는 시도를 "자연화(naturalizing)"라고 부르는데, 대표적인 자연화 이론가로 제리 포더, 드레츠키(Fred Dretske)를 꼽을 수 있다. 이들은 인과 개념 내지 정보 이론에 호소해서 지향적 관계, 특히 심적 표상이 어떤 대상을 의미하는 관계를 해명하려고 한다.

어떤 사건은 다른 사건과 법칙적으로 공변하는(lawfully covary) 관계에 있을 수 있다. 이 경우에 그 사건은 다른 사건의 정보를 담고 있다. 강이 북쪽으로 흐르는 것은 북쪽의 지형이 낮다는 정보를 담고 있으며, 동쪽에 그림자는 해가 서쪽에 있다는 정보를, 몸에 나타나는 어떠한 반점들은 홍역을 그가 홍역에 걸렸다는 정보를 담고 있다. 이때에 북쪽으로 흐르는 강, 동쪽의 그림자, 몸의 반점들은 각각 자연적 신호(natural signs)들이다. 자연적 신호들이 어떤 것에 관하여 어떤 정보를 담고 있는가 하는 것은 우리가 그것들을 보고 해석하는 일과 독립적이다. 아무도 관찰하지 않거나 또는 관찰하는 그 누구도 알지 못하고 믿지 않는다고 해도 북쪽으로 흐르는 강은 북쪽의 지형이 낮다는 정보를 담고 있다.

공변하는 관계에 있는 두 사건 중 하나가 다른 하나에 대한 정보를 담고 있다는 말을 "의미한다"고 말할 수도 있다. 물론 이 때에 의미는 언어적 규약에 의한 의미가 아니라 그리이

스(1957)가 말한 “자연적 의미(natural meaning)”라는 것이다. 자연적 의미와 정보적 내용(informational content)은 같은 말이다. 이 두 개념은 모두 두 사태 사이의 법칙적 의존 관계(또는 인과적 의존 관계)에 의해서 정의될 수 있으며 법칙적 의존관계란 다시 1의 확률의 개념과 동등하다고 할 수 있다. 어떤 자연적 기호 r 이 ‘ s 는 F ’라는 정보를 담고 있다 = 어떤 사건 내지 사태 r 이 ‘ s 는 F 임’을 (자연적으로) 의미한다 = r 은 ‘ s 의 F 임’과 법칙적으로 공변한다 = r 이 주어지면 (s 가 F) 일 조건적 확률은 1이다.(Dretske, F., 1981) 그러므로 ‘정보론적 의미론’, ‘인과적 의미론’, ‘의미의 공변이론’ 등은 같은 이론을 가리키는 말로서 모두 포더와 드레츠키에 의해서 대표되는 자연화된 의미론을 가리킨다.

물론 자연적 의미 곧 정보적 내용이 의미에 대한 얘기의 전부일 수는 없다. 자연적 신호들 이란 맥락에 따라서 무엇인가를 의미한다고 말해질 수 있긴 하지만 무언가를 다른 무엇으로 오표상(misrepresent)할 수는 없다. 민의 몸의 반점이 있지만 그가 홍역에 걸린 것은 아니고 단 것을 너무 많이 먹어서 그렇다면 민이 몸의 반점은 홍역을 (자연적으로) 의미하지 않는다. 합선되었을 때 벨 소리는 문밖의 초인종이 눌러졌다는 정보를 담고 있는 것이 아니다. 벨 소리는 초인종이 눌러졌다는 정보를 전달하도록 설계된 것이기는 하지만 이 경우 벨

소리는 전류가 흐른다는 정보를 담고 있을 뿐 초인종이 눌러졌다는 정보를 담고 있는 것은 아니다.

아까 지향성의 특징을 지향적 비존재라고 말했는데, 지향성을 가지는 것은 존재하지 않는 것을 가리킬 수도 있고 성립하지 않는 사실을 묘사할 수도 있다. 오표상할 수 없는 기호는 지향성을 가진 기호가 아니며 엄밀하게 말하면 기호도 아니다. 표상함이라는 관계, ‘의미한다’거나 ‘가리킨다’는 말로 표현되는 관계는 전형적인 지향적 관계이다. 그러므로 지향적 내재가 없으면 표상도 없다.

이점을 염두에 두고서 지향적 관계를 자연적 관계에 의해서 정의하는 일을 생각해 보자. 편 의상 포더가 말하는 사고의 언어라는 내적인 표상 체계가 있다고 가정하고 이제부터 자연 언어 대신 사고의 언어를 가지고 얘기하겠다. 그러면 내가 소를 생각한다는 것은 소를 의미하는 내적 표상 <소>와 어떤 관계를 맺는다는 뜻이 된다. 내가 소를 볼 때, 그 소는 복잡한 인과적 과정을 거쳐서 내가 <소> 표상(token)을 갖게 한다. 나의 내적 표상 <소>가 소를 의미하는 것은 그것이 소라는 대상과 공변하기 때문, 또는 소라는 대상에 의해서 인과적으로 야기되었기 때문, 또는 소가 있다는 정보를 담고 있기 때문이다. (앞으로 이 세 가지 말을 혼용하겠다.) 그러면 내적 표상 <소>가 소를 ‘의미한다’는 것은 소가 <소>를 야기하였다는 뜻

이 되지 않겠는가. 그래서 다음의 정의가 얻어진다.

(정의 1) 어떤 내적 표상 $\langle r \rangle$ 이 R을 의미한다 =df. R이 하나의 $\langle r \rangle$ 표상(token)을 (어떤 인과 법칙에 따라서) 야기한다.⁴⁾

문제는 이 정의가 그대로는 의미라는 지향적 관계에 대한 제대로 된 정의가 못 된다는 데에 있다. 내가 어두운 곳에서 말을 보고서 그것이 소라고 생각했다고 하자. 이 때 내가 가진 것은 $\langle 소 \rangle$ 표상이지만 그것을 야기한 것은 소가 아니라 말이다. 물론 소가 $\langle 소 \rangle$ 표상을 야기하기도 한다. 이런 경우에 위의 정의에 따르면 나의 $\langle 소 \rangle$ 표상은 소를 의미하지 않는다. 그것은 “소 또는 말”을 의미한다. 또 날아다니는 작은 벌레를 지각하는 보통의 개구리가 $\langle r \rangle$ 이라는 내적 표상을 가진다고 하자. 그런데 실험실의 개구리에게 작은 나뭇 조각을 빨리 움직여 보이면 개구리는 (그 반응으로 보아) $\langle r \rangle$ 이라는 표상을 갖는다. 그렇다면 개구리의 내적 표상 $\langle r \rangle$ 은 더 이상 벌레를 가리키는 것이 아니다. 정의에 따라서 $\langle r \rangle$

은 “벌레 또는 나뭇 조각”을 가리킨다. 요컨대 어떤 심적 표상이 그것이 의미하는 대상 아닌 다른 것에 의해서 인과적으로 야기되는 경우는 얼마든지 있을 수 있는데 위의 정의는 그런 경우들을 제거하지 못한다는 얘기다. 이것이 이른바 선언 문제(disjunction problem)이다.

선언 문제에 대한 제일 간단한 대답은 의미된 대상이 아닌 다른 것들이 그 표상을 야기하는 경우들이란 모두 정상적인 상황이 아니라고 답하는 것이다. 말이 나의 $\langle 소 \rangle$ 표상을 야기한 경우란 너무 어두워서 정상적으로 사물을 볼 수 없었던 상황이고, 또 실험실이란 것이 개구리가 서식하는 정상적인 상황은 아니지 않은가 말이다. 위 정의의 정의항에 “정상적인 상황에서”라고 덧붙이기만 하면 문제는 간단히 해결될 것 같다. 실제로 드레츠키는 한 때 (1981) “기호가 학습되는 상황”이라는 가정을 붙임으로써 이 문제를 해결하려고 하였으며, 어떤 사람들은 진화 과정을 통해서 그 체계가 잘 적응되어 있는 상황을 단서로 붙임으로써 문제를 해결하려고 시도한다.

그러나 이런 방법은 잘 통하지 않는다. 어떤 특정한 상황에서는 기호가 오직 한 종류의 원인에 의해서만 야기되리라는 생각이 잘못되어 있는 것이다. 예를 들어 ‘소’라는 표상은 온갖 방식으로 야기될 수 있다. 예를 들어서 말-생각이 $\langle 소 \rangle$ 를 야기할 수도 있고, 우유가 $\langle 소 \rangle$ 를 야기할 수도 있고, 심지어 실제로 포유동물이

4) 여기서 정의항은 r이 R을 일으키는 충분 조건이라고 해석해야 된다.

아닌 소 로봇이 <소>를 야기할 수도 있고, 등등 등. 그럼에도 불구하고 <소>는 소를 의미한다.(the robustness of meaning: Fodor, J., 1990) 물론 인과적 의미론이 성립하려면 이러한 인과 관계들 중에서 의미에 관계되는 인과 관계를 가려내는 방법이 있어야 한다. 그렇지만 특별한 상황을 상정함으로써 인과적 발생의 맥락을 제약하는 방법을 가지고서는 이걸 가려줄 수 없는 것이다.

포더(1987, 1990)는 오표상은 그것이 무언가를 표상하기 위해서 표상에 의존하지만 표상은 오표상에 의존하지 않는다는 사실이 선언 문제를 해결하는 열쇠가 되리라고 생각한다. 예를 들어서 말이 야기한 <소> 표상이 있다는 것은 소가 야기한 <소> 표상이 있다는 것에 의존한다. <소> 표상의 말에 대한 인과적 의존은 <소> 표상의 소에 대한 인과적 의존에 의존한다. 그러나 그 역은 성립하지 않는다. 그러므로 <소>는 소를 의미한다. 왜냐하면 소가 아닌 것이 야기한 <소> 표상은 모두 소가 야기한 <소> 표상에 반대칭적(asymmetrical)으로 의존하기 때문이다. <소> 표상이 소에 관한 정보를 담고 있지 않다면 그것은 어떤 것에 대한 정보도 담고 있지 않았으리라는 것이요, “소가 아닌 것 — <소>”의 관계들을 모두 깨뜨리지 않고서는 “소 — <소>”의 관계를 깨뜨릴 수 없다는 것이다.

포더의 이러한 제안에 대하여 자세한 논의는

하지 않겠다. 그러나 한 가지, 뭔가 충분치 않다는 느낌은 말하고 넘어가야겠다. 마치 사과를 없애는 마술을 보여주겠다고 하고선 그걸 먹어버리는 걸 구경하는 듯한 느낌이다. 포더 자신도 이것이 표상의 의미에 대한 충분한 자연화가 못 된다는 점을 인정하고 있기는 하지만 이것은 기대했던 대답이 아니다. 심적 표상의 의미를 인과 관계라는 자연적 관계에 의해서 정의하되 어떻게 해서 오표상이 가능한가 까지도 설명하겠다는 얘기를 듣고 우리가 기대했던 것은 의미에 관계되는 인과 관계와 의미와 무관한 인과 관계를 구별하는 기준 하나를 들겠다는 것은 아니었다. 오표상이 어떻게 해서 발생하게 되는가에 대한 뭔가 구체적인 그림을 기대했던 것이다. 포더의 반대칭적 의존이라는 대답은, 그것이 의미에 관계되는 인과 관계를 잘 떼어내 준다고 해도, 심적 표상의 의미에 대한 메커니즘의 해명으로서는 전혀 충분치 않다.

메커니즘을 보여달라는 기대에 조금 더 근접한 대답을 바로 드레츠키(1986, 1988)에서 찾을 수 있다. 초인종 예에서 시작하자. 합선에 의해서 초인종이 울릴 때 초인종 소리는 문 밖에 사람이 있다는 정보를 담고 있는 것이 아니라 “문 밖에 사람이 있거나 또는 합선되었다”는 정보를 담고 있다. 그러나 벨 소리가 초인종이 눌러졌다는 것을 의미하도록 설계되었다는 점으로부터 밖에 사람이 있든 없든 초인

종이 눌러졌든 아니든 벨 소리는 초인종이 눌러졌다는 것을 “의미”한다고 할 수 있지 않을까? 드레츠키는 이런 종류의 의미에 ‘기능적 의미(functional meaning)’라는 이름을 붙이고 있다. 기능적 의미는 이렇게 정의된다.

(정의 2) d가 G임을 w가 F임을 (기능적으로) 의미한다=df. d의 기능은 w의 상태를 가리키는 것인데, 특히 d는 그것이 G임에 의해서 w가 F임을 가리킨다.

“초인종(d)이 울림(G)은 문 밖의 상태(w)가 사람이 있음(F)임을 (기능적으로) 의미한다”는 말은 “초인종의 기능은 문 밖의 상태를 가리키도록 설계되었고, 특히 초인종은 울림 소리에 의해서 문 밖에 사람이 있음을 나타내도록 설계되었다”는 뜻이다.

그러나 이 예들에서 처럼 기능이란 것이 사람들에 의해서 부여된 기능일 경우에 기능적 의미란 것은 우리가 설명하고자 하는 능력을 이미 가진 체계들의 지향성에 의해서 오염되어 있다. 어떻게 표상이 가능한가를 자연주의적으로 설명하기 위한 탐구에서 필요한 것은 부여자를 가정하지 않는 기능, 우리의 의도와 목적과 해석으로부터 독립적으로 작용하는 기능이다. 그리고 이런 기능을 찾는 일은 어렵지 않은데 생물학적인 체계, 유기체들이 바로 그런 기능을 가지고 있다. 이러한 유기체들이야

말로 자연적 의미, 즉 정보의 진짜 소비자들이다.

생물학적인 필요의 충족을 위해서 핵심적인 기능으로서 정보를 획득/사용하는 기능을 가진 체계들이란 외부의 상태에 대한 정보를 담은 내적 상태를 가질 수 있는 체계들이다. 눈의 신경 세포가 제 기능을 한다면 이 신경 세포의 여러 상태들은 그 유기체가 처한 환경의 여러 특징들에 대한 정보를 담고 있어야 한다. 그런데 이 내적 상태들이 기능적으로 무엇을 의미하는가 하는 것은 물론 이 상태들이 담고 있는 정보와 그런 상태들을 갖는 기관(우리의 예에서는 눈)이 무엇을 나타내는 기능을 하는가에 달려 있다. 마그네토솜이라는 바다의 혐기성 박테리아는 자기 몸 안에 나침반의 바늘과 같은 기능을 하는 자석을 갖고 있다고 한다. 이 신체 자석 때문에 북반구 바다에 사는 마그네토솜은 언제나 지자계의 북극을 향하게 되며 따라서 늘 북쪽으로만 움직인다. 결국 이 자석은 마그네토솜으로 하여금 혐기성 박테리아에게 치사적인 산소를 많이 함유하고 있는 바다 표면에서 멀어지도록 한다는 것이다. 북반구 바다의 마그네토솜을 남반구의 바다로 옮기면 그것들은 바다 표면을 향해 추진되어서 죽게 될 것이다. 그러니까 마그네토솜의 신체 자석의 방향은 <그쪽에 산소가 없다>는 것을 기능적으로 의미하며, 마그네토솜이 사는 정상적인 환경에서 신체 자석의 방향은 <그쪽

에 산소가 없다)는 정보를 담고 있다. 그러나 남반구에 옮겨진 마그네토솜의 경우에 신체 자석의 방향의 기능적 의미는 동일할 테지만 그것이 담고 있는 정보는 동일하지가 않다. 그러나 이것이 오표상의 경우가 아닌가? 드레츠키는 이것으로 충분치 않다고 생각한다. 이 박테리아의 신체 자석의 기능이 <산소가 없는 곳의 방향>을 가리키는 것으로 해석된다면 남반구의 바다로 옮겨진 박테리아의 신체 자석은 오표상의 경우가 된다. 그러나 <지자계 북극의 방향>을 가리키는 것이 신체 자석의 기능이라고 해석된다면 남반구의 바다로 옮겨진 박테리아의 신체 자석은 환경을 정확하게 표상하고 있는 것이다. 막대 자석을 가까이 대면 그 박테리아는 지자계의 방향을 오표상할 것이라고 생각해서 오표상의 가능성을 남기려고 할 수는 없다. 이 경우에는 신체자석의 기능이 <지자계의 북극 방향을 가리키는 것>이라고 해석하지 않고 <그 박테리아 주위의 자기장의 방향을 탐지하는 것>이라고 해석하면 문제는 그대로 남는다. 이렇게 기능을 재해석함으로써 기능적 의미는 언제까지나 오표상할 수 없는 의미로 남게 되는 것이다.

물론 박테리아는 지자계 북극 방향이 <지자계 북극 방향이기 때문에> 거기로 움직여 갈 필요는 없다. 마찬가지로 개구리의 눈은 움직이는 작은 물체가 <움직이는 작은 물체이기 때문에> 그것을 탐지할 필요는 없다. 다시 말해

서 마그네토솜의 신체 자석의 기능을 <지자계 북극 방향>을 가리키는 것으로 기술하는 한 그러한 기술은 그런 기능이 어떻게 해서 그 박테리아의 필요와 관련되는지 드러나지 않는다. 그러므로 유기체의 메카니즘의 기능을 기술하되 언제나 그 유기체의 궁극적인 생물학적 필요와 관련해서 기술해야 한다는 방침을 받아들인다면 이 문제에 답할 수 있는 길이 열릴 것 같다. 그러므로 마그네토솜의 신체자석의 기능은 산소가 없는 물의 방향을 가리키는 것이고, 개구리의 눈의 기능은 움직이는 벌레의 위치를 탐지하는 것이라고 말해야 한다. 따라서 만일 유기체의 생물학적 필요와 관련된 기능을 정확히 진술할 수 있고, 유기체가 가진 여러 메커니즘의 기능을 그렇게만 기술해야 한다는 방침을 인정한다면 우리는 박테리아의 경우와 개구리 눈의 경우에서 이미 오표상의 원시적인 경우를 가진 셈이다. 더 나아가서 생물학적 기능의 개념을 잘 정식화하면 자연적 의미나 법칙적 공변 같은 개념에 호소하지 않고서도 표상의 내용에 대한 자연주의적인 설명을 줄 수 있을지도 모른다. 밀리칸(Millikan, R., 1984)이 바로 이런 길을 택하고 있다.

그러나 드레츠키는 이런 방향도 택하지 않는다. 드레츠키는 원시적인 표상 체계를 통해서 오표상의 문제를 생각하는 것이 모호한 귀결을 낳는 이유는 그 체계들이 너무 단순하기 때

문이라고 생각한다. 체계의 정보 처리 능력이 어느 수준 이상 복잡할 때에 비로소 분명한 의미에서 오표상이 가능하리라는 것이다. 어떻게 또 얼마나 복잡해야 하는가? 어떤 유기체가 어떤 유해물질 F를 탐지하는 두 가지 방법을 가지고 있다고 하자. 이 유기체가 F를 분별해 내는 두 가지 감각 양태를 가졌거나 아니면 하나의 감각 양태로 F의 두가지 이상의 표지를 분별해낼 수 있다면 이런 일이 가능하다. F의 서로 다른 두 표식을 f_1 , f_2 라 하고 그 각각의 감각 입력을 s_1 , s_2 , 또 이 입력을 통하여 f_1 , f_2 와 공변하는 (즉 f_1 과 f_2 의 정보를 담고 있는) 유기체의 내적 상태를 I_1 , I_2 라고 하자.

또 이 내적 상태들을 똑같이 R이라는 제3의 내적 상태에 연결되어서 R이 유기체의 어떤 행동을 촉발한다고 하자. 이런 경우에 R은 F와 공변하고 동시에 F를 기능적으로 의미하지만 (F를 탐지하는 것이 이 메커니즘의 기능이 기 때문에) f_1 이나 f_2 를 의미하지는 않는다. F의 탐지 방법이 여러 가지라는 점 때문에 그 유기체의 내적 상태 R은 중간 단계의 표식들 ($f_1, 2$ $s_1, 2$ 들)에 대한 정보를 전달하지 않으면서 F에 대한 정보를 전할 수 있게 된다. 이제 F의 f_1 만을 갖는 어떤 가짜 F를 그 유기체에 노출시키면 일련의 인과적 사건을 거쳐서 그것은 행동을 취할 것이다. 이것이 드레츠키가 그려주는 제대로 된 오표상의 경우이다.

물론 이런 경우에도 R이 선언적(disjunctive)

인 의미를 갖는다고 인정해 버리면 오표상은 없다. R이 $f_1 \vee f_2$ ($v s_1 \vee s_2$)를 의미한다면 가짜 F의 f_1 때문에 기능을 적절히 재진술함으로써 오표상은 배제될 수 있다. 그렇다면 아무리 다양한 탐지방법을 가진 체계라고 해도 이런 식으로는 오표상의 가능성을 보여줄 수 없게 되는 것이 아닌가? 드레츠키는 학습할 수 있는 체계라는 아이디어가 이런 난관을 열어 줄 수 있다고 생각한다. F가 있을 때 조건화된 감각 자극 cs에 자꾸 노출시키면 사정이 변한다. R이 cs만에 의해서도 촉발되게 된다. 어떤 종류의 자극을 택해서 R을 촉발시키는 조건화된 자극으로 삼을 수 있는가? 답은 어떤 자극이든지 된다는 것이다. 학습가능한 체계에 있어서 조건화된 자극의 종류에는 제한이 없다. 제리 포더는 한때(1986) "구겨진 셔츠" 같은 비법칙적(non-nomological)인 속성, 투사 불가능한 속성에 반응할 수 있다는 것이 표상을 가지기 위한 조건이라고 말한 바 있거니와, 학습 가능한 체계에 있어서는 그야말로 모든 형태의 s(감각입력)가 cs로 될 수 있는 것이다. 이런 체계는 다양한 감각 입력을 하나의 내적 상태 R으로 변형해줄 수 있을 뿐 아니라 학습을 통하여 이 다대일 대응의 성격을 바꿀 수 있는 체계이다. R이 (선언적으로) 무엇과 공변하는가 하는 것은 그 개체의 학습의 역사에 달려 있다.

이쯤 되면 s들을 가지고서 R의 의미라고 말

할 수는 없게 된다. 이제 R의 의미라고 말해질 수 있는 것은 R뿐인 셈이다. 이런 수준의 복잡성을 가진 체계, 그것이 필요로 하는 정보에 접근할 수 있는 다수의 채널을 가질 뿐 아니라 그 채널을 학습을 통하여 확장해 갈 수 있는 체계에 이르면 진정한 오표상이 가능해진다고 드레츠키는 생각한다.

나는 기능의 개념과 법칙적 공변이라는 두 아이디어를 결합하고 있는 드레츠키의 견해가 지금까지 제시된 표상의 자연주의적 설명 중에서 가장 그럴듯 하다고 생각한다. 드레츠키의 설명은 우리가 오표상하는 많은 경우들 중에서 적어도 대표적인 한 가지 경우에 대한 매우 자연스러운 설명을 준다. 우리가 오표상하는 경우의 형태는 매우 다양할 것이다. 그러나 그 중에서도 가장 먼저 떠오르는 형태는 말은 소라고 오인하는 경우에 담겨 있다. 눈 앞의 대상이 소이라고 인지하는 나의 내적인 상태, 즉 <소>라는 나의 심적 표상은 매우 여러 가지 인과적 요인에 의해서 결과된 것이겠다. 그러나 인과적 표상론 내지 공변적 표상론이 옳다면 표상이 인과적으로 의존하는 것은 나의 내적인 상태의 遠因(distal cause)이어야 할 것이다. 그런데 소는 나의 내적인 상태에 직접 인과적으로 작용하지 않는다. 소의 모습이 제공하는 많은 시각적 단서를 통하여 작용할 것이다. 그래서 나의 내적인 상태를 결과하는 인과적 요인들, 즉 소의 시각적 특징들, 그 특징

들의 감각 입력들은 매우 다양하다. 나의 내적 상태는 그러한 近因들(proximal causes)을 매개로 해서, 또는 단서로 삼아서 소라는 대상과 공변한다. 말의 어떠 어떠한 시각적 특징들이 소의 시각적 특징과 동일하거나 적어도 유사하다는 점이 나로 하여금 말을 소로 오인하도록 만든 것이다. 드레츠키는 바로 이런 경우에 발생하는 오표상의 형태에 대해서 매우 설득력있는 설명을 주고 있다.

반복하거나와 오표상이 발생하는 경우들, 그 형태들은 매우 다양할 것이다. 퍼트남이 말하는 쌍둥이 지구로 옮겨진 내가 "물"이라는 내용의 심적 상태를 갖는 경우처럼 그 책임이 전적으로 환경의 탓인 경우도 있을테고, 또 시각 입력을 해석하는 나의 신경계통이 노후해서 제대로 작동하지 않는 경우도 있을테고, 등등. 오표상의 여러 경우들이 어떤 공통성을 보일 거라는 보장도 없다. 틀리는 데에 무슨 원칙이 있겠는가? 그러나 오표상이 어떻게 가능한가를 설명하려 하는 이유는 오표상할 수 있는 체계만이 표상할 수 있다는 생각 때문이었다. 표상할 수 있는 체계, 적어도 인간과 같은 정도로 표상할 수 있는 체계를 유물론적으로 설명하는데 성공한다면 그러한 설명은 인간이 범하는 모든 형태의 오표상에 대해서 그것이 어떻게 해서 발생하는가도 설명할 것이다. 불행히도 드레츠키의 설명이 과연 그런 설명인지 분명치 않다. 그러나 적어도 그는 한가지

형태의 오표상을 자연화한 것이며, 따라서 중요한 의미에서 표상할 수 있는 자연적 체계 하나를 그려낸 것이다.

참고 문헌

- (1) Baker, L. (1988) "Cognitive Suicide" in Grim, R. and Merrill, D. ed. (1988) Grim, R. and Merrill, D. (ed) (1988) *Contents of Thoughts*. The Univ. of Arizona Press
- (2) Brentano, F. (1925) "The distinction between Mental and Physical Phenomena" in Chisholm, R. (1960) *Realism and the Background of Phenomenology*. Illinois: The Free Press of Glencoe
- (3) Churchland, P. (1978) *Scientific Realism and the Plasticity of Mind*. Cambridge Univ. Press
- (4) Churchland, P. (1989) *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. The MIT Press
- (5) Dennett, D. (1978) *Brainstorms*. The MIT Press
- (6) Dennett, D. (1987) *Intentional Stance*. The MIT Press
- (7) Dennett, D. (1991) "Real Patterns" in *The Journal of Philosophy*. Vol. 88, No. 1
- (8) Dretske, F. (1981) *Knowledge and the Flow of Information*. The MIT Press
- (9) Dretske, F. (1984) "Misrepresentation" in Bogdan, R. ed. (1986) *Belief: Form, Content, and Function*. Oxford Univ. Press
- (10) Dretske, F. (1986) *Explaining Behavior*. The MIT Press
- (11) Fodor, J. (1987) *Psychosemantics: Problems of Meaning in the Philosophy of Mind*. The MIT Press
- (12) Fodor, J. (1990) *A Theory of Content and Other Essays*. The MIT Press
- (13) Grice, P. (1957) "Meaning" in Martinich, A. ed. (1990) *The Philosophy of Language*. Oxford Univ. Press
- (14) Grice, P. (1968) "Logic and Conversation" in Davis, S. ed. (1991) *Pragmatics: A Reader*. Oxford Univ. Press
- (15) Harman, G. (1989) *Change in View*. The MIT Press
- (16) Millikan, R. (1984) *Language, Thought and Other Biological*

30 민찬홍

Categories. The MIT Press

(17) Stich, S. (1983) *From Folk Psychology to Cognitive Science*. The MIT Press

(18) Stich, S. (1990) *The Fragmentation of Reason*. The MIT Press