

표상의 역할과 환원적 제거주의*†

박 제 윤‡

처칠랜드 부부는 계산적 신경과학과 연결주의 시의 연구 성과로부터, 새로운 표상이론을 제안하고, 나아가 그러한 표상이론의 미래 역할에 비추어 통속심리학이 근거하고 있는 문장적 표상이론의 제거를 전망한다. 흥미롭게도 클락은 새로운 표상이론이 주는 함축적 충격을 인정함에도 불구하고 두 가지 이유, 문장적 표상이론의 유용성과 차원에 따른 이론들의 역할과 목적의 차이에 비추어 제거주의에 반대한다. 본 논문은 처칠랜드와 클락의 입장을 비교하고 어느 편이 더 설득력을 갖는지 검토한다. 필자는 그 검토를 통해 처칠랜드의 입장을 옹호하고 새로운 표상이론이 함의하는 제거적 유물론을 적극적으로 구성해 볼 것이다.

【주요어】 표상, 환원주의, 제거주의, 통속심리학, 처칠랜드

1. 서론

심리철학의 중심 쟁점들을 통하여 제거적 유물론을 주장하기란 일반적으로 무모한 것처럼 보이기 쉽다. 분명 우리는 뚜렷하게 스스로를 인지하는 자신의 사고에 대해 자각하며, 그런 정신 작용들은 물리적 현상들과 근본적으로 다르다고 생각하기 쉽다. 따라서 정신 작용을 물리적 작용으로

* 접수완료: 2009.5.20 / 심사 및 게재확정: 2009.6.2 / 수정완성본 접수: 2009.6.8

† 유익한 심사평과 오류를 지적해 주신 익명의 두 심사위원에게 감사드린다. 그리고 본 논문의 작성 과정에서 도움을 주신 고인석 교수, 이영의 교수에게, 그리고 본 논문을 읽고 조언을 주신 이선경 선생에게 감사드린다.

‡ 인하대학교 철학과 박사과정

설명하고, 나아가 전자가 후자로 대체 또는 제거될 것으로 전망하기란 매우 어려워 보인다. 그럼에도 불구하고 폴 처칠랜드(Paul Churchland)와 페트리샤 처칠랜드(Patricia Churchland)는 제거적 유물론을 지속적으로 주장해왔다.

처칠랜드 부부는 한편으로는 과학철학에 대한 반성적 논의를 통하여, 다른 한편으로는 신경과학과 인공지능의 연구 성과들에 근거하여 제거적 유물론을 주장한다. 그들은 현대 인지과학의 성과를 조망한 구체적 이론을 제안하고, 그 이론을 통하여 과학철학의 전통적 문제들을 해결하고자 한다.¹⁾ 그들의 제안이 여기에서 특별히 우리의 관심을 끄는 이유는 실험 가능한 새로운 ‘표상 이론’을 제안한다는 점에 있다. 그들의 이론은 외부세계의 특징들이 어떻게 내적으로 표현되는지 그리고 그 내적 표현들이 어떻게 계산적으로 처리되는지를 설명할 뿐만 아니라, 그 이론에 대한 실험적 성과를 보여주기도 한다. 따라서 새로운 표상 이론은 제거적 유물론의 입장에서 기대되는 인지 작용에 대한 구체적 증거를 제시하며, 궁극적으로 우리에게 제거적 유물론에 더욱 개방적 태도를 취하도록 유도한다.

그렇지만 여기에서 다음과 같은 의문이 제기될 수 있다. 처칠랜드 부부가 제시하는 표상이론이 진정으로 제거적 유물론을 함축하는가? 과연 새로운 표상 이론에 의해 통속심리학의 문장적 범주들이 환원적으로 대체 또는 제거될 것인가? 이러한 질문들에 대해 앤디 클락(Andy Clark, 1996)은 ‘그렇지 않다’고 대답한다. 흥미롭게도 그는 처칠랜드의 표상 이론과 연결주의 AI에 기반을 둔 학습 이론을 긍정적으로 평가하면서도, 제거적 유물론 그 자체에는 동의하지 않는다. 그에 따르면, 처칠랜드가 주장하는 표상 이론과 통속심리학은 양립 가능하며, 따라서 그 표상 이론이 통속심리학의 문장적 범주의 제거를 요구하지도 않는다.

본 논문의 목적은 일차적으로 클락의 관점과 처칠랜드의 관점의 차이를 알아봄으로써, 새로운 표상 이론과 통속심리학이 과연 양립가능한지 검토할 것이며, 그러한 검토를 통해 어느 주장이 더 타당한지를 알아볼 것이다. 필

1) 이런 논의를 특히 P.M. Churchland(1990, 2002a)에서 잘 알아볼 수 있다.

자는 이러한 검토 과정에서 처칠랜드의 배경 전제라 할 수 있는 과인과 쿤의 사상을 고려할 것이다. 그러한 고려가 새로운 표상 이론이 어떻게 제거적 유물론을 유도하는지 보다 원초적으로 드러내줄 것이라고 기대한다. 결론적으로 본 논문은 이러한 논의를 통해서 개념적으로 무모해 보이는 처칠랜드의 환원적 제거주의 배경 전제들을 드러내어 보여줌으로써 우리가 그 관점에 좀 더 친근하게 다가설 수 있는 계기를 제공할 것으로 기대한다.

2. 표상의 인지신경생물학적 이해

2-1 뇌의 기초 구조와 표상

처칠랜드 부부의 근본 목표는 현대 과학의 연구 성과들을 통해서 과학 철학과 인지과학의 중요 쟁점들에 대해 대답하는 것이다. 그러한 시도에서 가장 돋보이는 것은 표상과 설명에 대한 새로운 접근이다. 그들의 제안은 인지신경생물학과 연결주의 AI의 연구로부터 유래한 것으로 뇌의 표상과 계산에 관한 강력하고 새로운 이론적 접근을 가능하게 해준다.²⁾

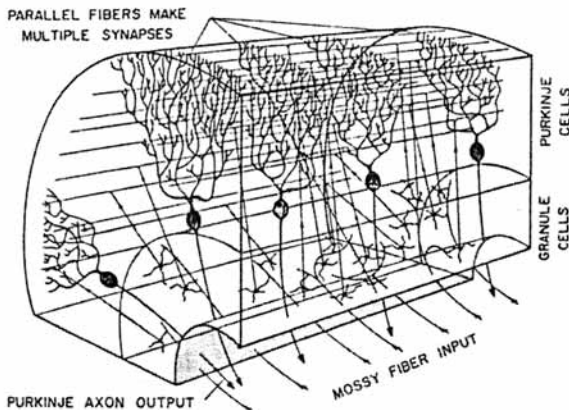
특히 『신경계산적 전망』(*A Neurocomputational Perspective*, 1989)에서 보여주는 표상과 계산에 관한 이론은 우리가 장차 마음을 환원적으로 설명해 줄 가능성과 제거적 유물론을 주장한다.³⁾ 폴 처칠랜드의 제안에 따르면, “뇌는 적절한 상태 공간의 위치에 의해 다양한 실재의 국면들을 표상하며, 뇌는 한 상태공간에서 다른 상태공간으로 일반적 좌표 전환에 의해 그 표

2) 이런 논의를 P.M. Churchland(1979, 1981, 1989), P.S. Churchland(1986), P.S. Churchland & T.J. Sejnowski(1992)에서 잘 알아볼 수 있다. 그 중에서도 가장 요약적인 설명을 P.M. Churchland(1989), 5장에서 볼 수 있다. 아래의 내용은 주로 그곳에서 가져왔다.

3) 이 주장은 ‘환원주의와 제거주의는 양립할 수 없다’는 가정에서 보면 수용되기 어려워 보인다. 그러나 처칠랜드 주장에 따르면, 환원하는 이론과 환원되는 이론 간에 그 환원의 정도는 다양하며, 그 정도에 따라서 이론들 사이에 동일화, 수정, 대체, 제거 등이 다양하게 나타날 수 있다. 이 논문의 뒤(4 개념체계의 통일), 그리고 P.S. Churchland(1986), 7-9장 참조.

상들에 대한 계산을 수행한다.”⁴⁾ 이 가설이 갖는 함축을 논의하기 앞서 우선 뇌의 표상과 계산 이론부터 구체적으로 알아보자.

대뇌의 미시 물리적 구조물인 피질은 전형적으로 아래와 같은 구성을 보여준다. 대뇌의 층판 피질(laminar cortex)은 특징적으로 6층으로 이루어져있다. 그리고 (그림1)과 같이 외부 입력 신호를 각 층의 수평적 투사 구조를 통해 받아 수직적 연결을 통해 처리하고, 출력 신호를 하부 층을 통해 내보낸다.



(그림1) 소뇌(cerebellum)의 도식적 단면: 입력신호는 이끼섬유(mossy fibers)를 통해서 피질로 올라가고, 피질의 평행섬유들을 통해 나란히 늘어선 푸르키니에 세포들(purkinje cells)에 동시에 전달되어, 그곳에서 동시 다발적으로 신호처리가 이루어진 후에, 푸르키니에 축삭들(axons)을 통해 출력되는 구조를 보여준다.⁵⁾

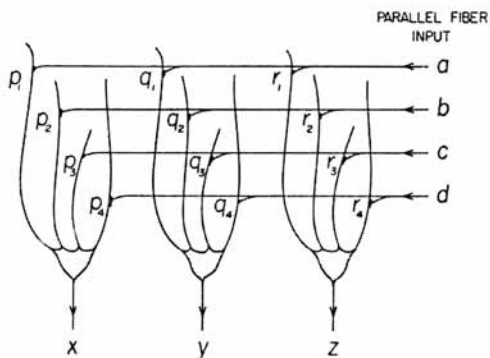
위의 구조를 갖는 세포집단들은 그 기능을 위한 일종의 국소 맵(topographic map)을 형성한다고 보고되고 있다. 예를 들어, 망막의 세포

4) 그런 가설에 대한 접근이 가능했던 것은 신경과학자들, 펠리오니즈와 리나스(Andras Pellionisz and Rodolf Llinas)의 연구를 접할 수 있었던 때문이라고 밝힌다. P.M. Churchland(1989), 78-9쪽.

5) P.M. Churchland(1989), 101, 182쪽.

들은 이웃관계를 유지하면서 일차 시각피질로 투사한다. 망막의 국소 조직을 그대로 반영하는 일차 시각피질은 따라서 국소 맵을 구성한다. 세계와 대응하는 이러한 국소 맵들은 뇌의 다양한 영역에 존재하는 것으로 밝혀졌다. 신체감각피질(브로드만 영역3)은 신체 감촉 면의 국소 맵을, 운동피질(브로드만 영역4)은 신체 근육 시스템의 국소 맵을, 청각피질(브로드만 영역 41, 42)은 주파수 공간의 국소 맵을 구성하는 것으로 밝혀졌다. 여기에서 분명한 것은 뇌의 피질이 국소 표상을 갖는다는 점이다.⁶⁾

뇌가 왜 이런 패턴의 구조를 조성하는가? 뇌 기능의 인지적 의미가 무엇인가? 처칠랜드에 따르면, 그 질문들에 대한 대답은 감각운동 조율에서 찾아볼 수 있다. (그림1)에 대한 도식적 (그림2)을 통해 그것의 기능적 역할을 이해할 수 있다. 세포 집단이 (a, b, c, d)의 입력신호를 받아 처리하고, (x, y, z)의 출력신호를 내보내는 구조를 갖는다⁷⁾고 가정했을 경우, 그러한 구조에서 세포가 하는 일은 위상공간에서 ‘체계적 좌표변환’이다.



(그림2) 신경망 행렬(neural network matrix)의 도식적 그림: 신경세포들이 어떻게 물리적인 행렬처리, 즉 벡터 대 벡터 변환(vector to vector transformations)을 효과적으로 수행하는지 알아볼 수 있다.⁸⁾

6) 같은 책, 81-2쪽.

7) 이런 신경적 표상, 즉 숫자들 조합으로 표현되는 표상이 과연 실제 세계에 대한 표상이냐는 의문과 관련한 논의를 P.S. Churchland(2002b)에서 볼 수 있다.

좌표변환을 통한 뇌의 처리과정은 매우 간단하면서도 효과적인 함수적 계산처리 방식이다. (그림2)에서 입력 신호는 두 단계의 과정을 거쳐 출력 신호를 산출한다. 입력신호의 행렬은 시냅스의 연결 가중치들(p_{nv} , q_{nv} , r_n)에 의한 변조 과정과 변조된 신호들이 세포로 유입되어 통합되는 과정을 거치며, 이러한 좌표변환 장치의 처리 과정은 벡터의 행렬 곱에 의한 변환방식으로 해석된다.⁹⁾ 그 변환방식은 실제 신경계의 병렬적 구조에서 아주 단순하게 구현되며, 그 구조가 상태 공간의 좌표변환을 수행하는 방식이다. 따라서 그 병렬 구조는 다음과 같은 능력을 발휘하며, 그것은 생존의 현장에서 탁월한 유용성을 발휘한다.

첫째, 비-선형 변환을 계산할 수 있다.

둘째, 신경 행렬은 탁월한 처리속도를 갖는다.

셋째, 큰 행렬과 잉여 세포들이 주어질 경우, 그 구조물은 일부 요소 세포들의 상실에도 기능적 내구성을 보일 것이다.¹⁰⁾

위의 능력들은 다음과 같은 특징적 행동들을 설명해준다. 첫째, 우리는 다수의 고려 목록들에 의해 평소 가지고 있었던 확신을 갑자기 등장한 하나의 작은 특정 고려 항목에 의해 바꾸어 주위 사람들을 당황하게 하는 경우가 있다. 위의 능력들은 그러한 우리의 돌발적 의사결정 변경을 설명할 수 있다. 둘째, 실제의 뇌는 범용컴퓨터의 순차처리방식과 달리 병렬적 계산방식을 가져서 범용컴퓨터가 수천 단계의 처리 과정을 거쳐 처리해야 할 과제를 단 몇 번의 과정으로 성취할 수 있는 능력을 갖는다. 따라서 실시간 문제의 해결은 물론, 일부 회로의 가중치를 바꿈으로써 그 회로에 학습을 담아낼 수 있게 한다. 결국 신경망의 구조는 처리장치인 동시에 기억장치이다. 위의 능력은 그 특성을 설명해준다. 셋째, 위의 능력들은 일부 세포의 상실이 협소한 특정 기능만을 상실시키지 않고, 전체적으로 뇌의 기

8) P.M. Churchlan(1989) 99, 182쪽.

9) 여기에 대한 보다 구체적인 세부적인 논의를 P.S. Churchland & T.J. Sejnowski(1990, 1992)에서 볼 수 있다.

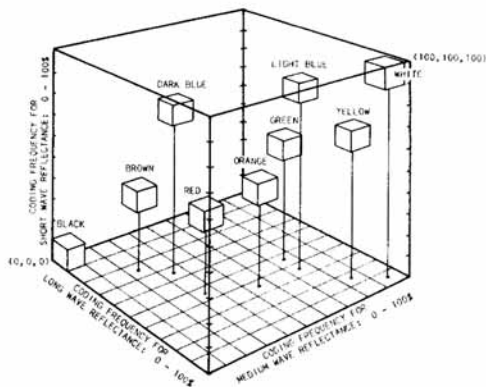
10) P.M. Churchlan(1989), 101쪽.

능을 조금 저하시키는 측면을 이해하게 한다.

2-2 기초 표상의 역할

위의 이야기가 처칠랜드가 밝히는 신경 연결망이 갖는 표상에 대한 이야기의 전부는 아니다. 이제부터 그 표상 이론이 표상의 역할과 관련하여 우리에게 신경계의 어떤 능력들을 이해시켜주는지 알아보기로 하자.

첫째, 위상공간과 관련하여 세계의 특징들에 대한 세포의 정교한 표상 능력을 생각해보자. 예를 들어, 만약 행렬망 신경장치가 입력으로 세 가지 변수들을 갖는다면, 즉 세계에 대한 어떤 특징들을 3종류로 수용한다면, 그리고 만약 (그림3)과 같이 그 3종류의 특징들이 각기 다른 축에 적어도 10개의 분별 등급만을 갖는다고 가정해도, 그 간단한 행렬망 신경장치가 자체의 위상공간에 표현 가능한 특징적 신호들은 10^3 가지에 이른다. 이런 가설적 설명을 통해 우리의 (색, 맛, 후각 등의) 지각들이 어떻게 그렇게 정교하고 섬세한 분별 능력을 갖는지 설명 가능하며, 그 분별 정도가 언어로 일일이 호칭하여 구분하고 설명할 한계를 훨씬 뛰어 넘어서는지 이유도 납득된다.



(그림 3) 색깔 상태 공간(color state space)의 도식적 그림: 세 벡터 값의 차이에 의해서 얼마나 다양한 색깔들이 위치벡터로 표현될 수 있는지 보여준다.¹¹⁾

둘째, 위와 같이 위상공간으로 설명되는 표상이론은 우리 감각의 '유사성' 인지 기능을 매우 잘 설명해준다. 우리 뇌의 행렬적 실행 장치인 신경망은 (앞서 설명했듯이) 국소 맵을 조성한다고 가정된다. 그 맵의 신경망은 위상공간의 특정 지점에 특별히 민감하게 반응하도록 장치되어있다고 가정해보자. 그러면 그 신경망은 위상공간의 대푯값을 가지는 셈이다. 그것을 치칠랜드는 "전형(prototype)"이라 부른다. 만약 (그림3)과 같이 특정 위상공간 지점에서 오렌지 색에 가깝게 반응하는 숫자들의 조합이 민감하게 반응하게 된다면, 그렇게 학습된 연결망은 그 표상 위치에 가까운 것들을 유사한 것들로 분류하게 할 것이다.

연결망의 전형들은 단지 감각들을 분별해낼 도식만을 제공하지는 않을 것이며, 동일한 방식으로 연결망의 계층적 구조를 가질 경우 상위차원의 연결망은 추상적 전형을 담아낼 수 있다고 기대할 수 있다. 그렇다면 연결망이 각종 추상적 개념들의 전형을 담아내어 우리에게 개념적 지식의 풍부함을 제공했을 것이며, 그것에 인한 세계의 인지능력을 부여했을 것이다. 다시 말해서 유사성을 분별하는 연결망의 (표상적) 전형들은 우리가 세계를 바라 볼 개념적 도식을 담아내는 역할을 담당한다고 가정된다.

셋째, 위의 위상공간에 의한 상태-공간(state-space)의 분석적 접근은 뇌가 어떤 방식으로 다양한 종류의 정보들을 통합할 수 있는지 설명하기도 하다. 또한 하위차원의 위상공간들이 상위차원의 위상공간으로 통합될 수 있다면, 반대로 상위차원 위상공간은 하위차원의 위상공간들로 어떻게 분해될 수 있을지도 추론하게 한다. 예를 들어 만약 평소 훈련에 의해 신체 각 부위들의 동작에 대한 전형을 담아내는 하위차원 위상공간을 갖는다면, 외야수는 의식적 고려 없이 그 전형을 실행하는 신경망의 국소 맵에 의해 신속하고 즉각적으로 대처할 것이다. 다시 말해서 각인된 전형들은 신속하게 자동적 계산처리, 즉 좌표공간의 벡터 변환을 가능하게 하여, 표상들이 통합되거나 분해됨으로써 각 신체 부위가 자연스럽게 움직이도록 해줄 것이다. 다만 날아가는 공을 바라보는 것만으로도 공을 따라가도록 신체를

11) 같은 책, 104쪽.

무의식적으로 움직일 것이다.

넷째, 위와 같이 국소 맵을 통한 행동 조절은 단지 신체만을 조절하기 위한 것이라 제한할 필요가 없다. 고도로 복잡한 상위차원의 위상공간을 갖도록 구성될 경우, 우리는 그것을 통해 매우 추상적 예측 능력을 담아낼 것이다. 예를 들어, 공의 속도에 대한 전형들을 국소 맵에 표상으로 담아낼 경우, 야구공을 쫓는 운동에 앞서 다만 날아가는 공을 바라보는 것만으로도, 우리는 그 공의 낙하지점을 쉽게 그리고 매우 빠르고 정확히 계산하여 예측할 수 있다. 그런 상위차원의 다양한 추상적 전형들을 더욱 많이 가질 수록 우리는 예측을 위한 더욱 많은 조건들을 담아낼 것이다. 그렇게 생각해보면, 이제 맵이 (단지 사물을 분별하기 위한 개념적 도식의 수준을 넘어) 복잡한 경험들을 담아낼 추상적 전형이라고 생각해보면, 그것을 우리가 “일반화”라 말하는 것에 상응할 것이다.

지금까지 살펴본 바에 따르면, 신경세포의 단순한 표상적 도식은 우리의 감각식별과 감각운동조율 능력을 발휘하게 하는 생물학적 실재의 기초 양태라고 주장된다. 뿐만 아니라, 보다 상위차원의 언어 사용을 포함하는 인지활동에 대한 설명의 자원이라고 주장된다. 따라서 처칠랜드는 이러한 생각이 자신을 제거주의로 인도한다고 말한다.

이런 보다 핵심을 찌르는 개념체계는 아마도 간주관적 기술(intersubjective description)과 자동적(무의식적) 내성(spontaneous introspection)을 매개하는 상식적 체계를 제거(displace)할지도 모른다.¹²⁾

제거주의로 인도하는 새로운 표상이론은 인지의 기초 자원을 ‘구문론적’ 개념으로 보지 않고, ‘기하학적’ 개념으로 바라본다. 상태공간 시스템의 표상과 그것의 계산 능력은 좌표변환에 의해서 가능하며, 그 기초 인지적 개념은 신경계를 넘어 지식 일반을 설명해줄 인식론의 강력한 이론으로 보인다. 우리 뇌는 그런 물리적 시스템을 가지고 있다는 가설이다.

12) 같은 책, 109쪽.

3. 통속심리학의 실용성

3-1 비-문장적 표상의 충격

새로운 표상이론에 근거하여 주장되는 처칠랜드의 제거적 유물론에 대해, 클락은 논문, “미래를 다루기: 인지과학에서의 통속심리학과 표상의 역할”(Dealing in Future: Folk Psychology and the Role of Representations in Cognitive Science, 1996)¹³⁾에서 아래와 같은 이유로 이의를 제기한다. 그에 따르면, 근본적으로 차원간 또는 영역간 이론들 사이에 설명의 목적과 역할은 서로 다르다. 구체적으로 통속심리학에서 문장적 표상의 분석과 문장적 표상의 추론의 가치는 신경생물학적 하부-인지 구조 내의 표상의 역할과 아주 다른 기능을 갖는다. 따라서 신경생물학의 연구 결과에 따른 처칠랜드의 하부-인지 표상이론이 통속심리학의 문장적 표상 이론의 제거를 주장할 수 없으며, 오히려 그 정당화를 제공하려 노력해야한다.

처칠랜드는 신경과학 연구 성과들을 근거로 들어 제거주의를 주장하지만, 클락은 동일한 근거로부터 통속심리학의 제거를 거부한다. 클락의 이런 입장이 지지된다면, (아마도 많은 심리철학, 인지과학, 과학철학 관련 학자들이 그의 입장을 선호할 것으로 보이는데) 처칠랜드의 제거주의는 그의 표상이론에 기초한 신경연결망이 이미 실용적 활용이 되고 있음에도 불구하고 과학의 성과를 과장하는 과학주의자로 평가될 수 있다. 그 점에서 클락의 주장은 매우 흥미를 끌며, 따라서 세부적으로 그리고 비판적으로 검토해볼 가치를 갖는다.

그러한 검토를 위해 우선 처칠랜드 주장에 대한 클락의 분석과 평가를 살펴볼 필요가 있다. 클락의 분석에 따르면, 처칠랜드의 하부-인지 표상이론이 갖는 가장 중요한 의도는 아래와 같다.

의도(1): 통속심리학의 일상적인 심적 설명 체계의 용어가 갖는 가치를 낮춘다.¹⁴⁾

13) A. Clark(1996), 86-103쪽 참조.

의도(2): 앞으로 인지과학의 연구는 신경과학의 연구로부터 나오는 표상과 설명의 자원을 더 많이 다룰 것이며, 연결주의자 계산적 모델링을 더 넓게 다루게 될 것임을 논증한다.¹⁵⁾

처칠랜드의 제거적 유물론의 입장을 더 잘 드러내어 본 논문의 쟁점으로 들어가기 위해 위의 두 의도를 아래와 같이 두 전망으로 바꿔 말할 수 있다.

전망(1): 통속심리학은 미래에 제거될 것이며, 제거되어야 한다.¹⁶⁾

전망(2): 인지활동을 설명하는 이론들은 하부-인지 표상 이론이란 새로운 자원을 통해 다시 설명되어야 한다.

위와 같이 바꿔 보면, 전망(1)은 전망(2)에 의해서 구체적으로 성취될 것임을 알 수 있으며, 따라서 의도(1)은 의도(2)에 의해서 가능한 주장임을 알 수 있다. 실제로 처칠랜드는 유력한 현대 인지신경생물학 연구와 연결주의자 AI 연구결과를 통해 새로운 하부-인지 표상이론을 제안하며, 따라서 심적 이론의 기반은 통속심리학이 아닌 새로운 표상이론에 의해 다시 설명되어야 한다고 주장한다.¹⁷⁾

그럼에도 불구하고 클락은 처칠랜드의 전망에 동의하지 않는다. 그가 보기에, 처칠랜드의 통속심리학에 관한 철학적 전망(1)은 오히려 마음의 과학적 탐구 방법의 전망(2)을 흐리게 한다.¹⁸⁾ 그 이유는 두 전망이 서로 교차

14) 이러한 처칠랜드의 의도를 특히 P.M. Churchland(1979, 1981, 1989)에서 잘 알아볼 수 있다.

15) 같은 책, 86쪽 참조. 이러한 처칠랜드의 의도를 특히 P.S. Churchland(2002), P.S. Churchland & T.J. Sejnowski(1992)에서 잘 알아볼 수 있다.

16) 처칠랜드가 무엇을 제거될 대상으로 지목하는지 다음의 말에서 알아볼 수 있다. “심리학적 현상들에 대한 우리의 상식적 개념은 근본적으로 틀린 이론을 조성하며, 그 이론의 원리들과 존재론은 완성된 신경과학에 의해 마침내 부드럽게 환원되기보다 제거될 것이다. ... (1) 상식적 심리학의 원리적 요소들: 명제태도(믿음, 욕망 등)과 (2) 그러한 요소들이 형성되는 합리성의 개념” P.M. Churchland(1989), 1쪽.

17) 같은 책, 9장 참조.

하는 입장에 있기 때문이다. '미래 인지과학 연구방법과 설명의 자원과 기획에 대한 전망'은 '통속심리학의 체계가 빈곤하다는 주장'에 반대하는 많은 사람들에게 의해서 다시 평가될 것이며, 결국 전망(1)에 대한 반발로 전망(2) 자체도 그들에게 수용되기 어렵게 만들어, 처칠랜드 자신의 입장을 곤경으로 끌고 들어간다는 관측이다. 이러한 관측이 어떤 근거에서 나오는지 구체적으로 보여주기 위해, 클락은 처칠랜드가 주장하는 표상의 담론이 인지과학의 미래에 어떤 역할을 담당할지에 대해 관심을 가져야한다고 주장한다.

클락의 분석에 따르면, 처칠랜드의 새로운 표상 이론은 내적 표상상태를 설명할 새로운 방안을 제시한다. 그 표상이론에 따르면 세계의 여러 다양한 특징들이 신경망 내의 내적 벡터공간의 위치로 표현된다. 따라서 '표상'이란 '세계에 대한 신경세포의 반응'이며, 그것은 상태공간에서 특정 위치, 위상 공간 내의 특정 벡터 그리고 행렬의 특정 숫자 조합 등으로 표현된다. 따라서 그 표현 방식은 위상공간에서 특정 표상, 즉 특정 벡터가 서로 어떻게 통합되고 다시 분해될 수 있을지 계산적으로 설명해줄 수단이라 할 수 있다. 그것은 생물학적으로 극히 사실적이며, 그 표상의 기초 자원은 우리를 포함한 모든 동물들의 신경계에서 인지활동의 기반을 제공한다. 그러나 그 새로운 표상의 기반은 문장의 구조와 전혀 무관하다. 따라서 문장에 의해 탐구되는 통속적 인지심리학은 올바른 접근법이 아닌 셈이다. 이러한 논의는 과학철학과 심리철학 그리고 인지과학의 방법론과 관련된다. 따라서 새로운 표상이론이 미래 인지과학의 탐구에 대해 다음과 같은 직접적 함축이나 역할을 갖는다.

- 첫째, 추론을 내적 표상의 데이터 구조를 넘어 규정되는 증명이론으로 모델화하려는 시도를 의심하게 한다.
- 둘째, 지각과 인지과정을 명확히 구분하는 경계의 개념을 의심하게 한다.
- 셋째, 기억, 데이터, 계산처리, 계산규칙 등의 친숙한 개념적 구분의 유용성에 제동을 건다.

넷째, 전통적으로 인지에 대해 기호처리 이론으로 접근했던 방법론이 근본적으로 사실적이 아닌 가정들에 근거하고 있음을 보게 한다.¹⁹⁾

이러한 함축들은, 클락에 따르면, 새로운 표상이론이 제공하는 매우 중요한 것이다. 이런 함축적 의미를 인정하기 때문에 그는 처칠랜드의 전망(2)에 대해서 긍정적으로 지지한다. 클락의 분석에 따르면, 처칠랜드의 새로운 과학적 모델이 개념적 '학습'에 주는 잠재적 충격은 다음과 같다. 처칠랜드의 학습 이론은 연결주의 학습알고리즘을 활용하며, 학습에 의한 표상 산출의 일반적 가능성을 명쾌하게 논증적으로 보여주며, 그 표상적 시스템이 어떤 내적 표상의 자원들에 대해서도 적용될 가능성을 보여준다. 그렇게 학습에 대한 연결주의자 접근은 표상 자체의 기원을 설명해줄 것으로 기대되며, 따라서 앞으로 인지활동에 관한 연구에서 중심에 놓일 전망이다. 결국 전통적 표상이론들은 뇌의 성숙된 표상적 자원이 무엇과 같은 모습일지 발견하거나 추측하는 일에 초점을 맞춘 것으로 파악된다. 나아가서 새로운 표상이론에 의해서, 우리에게 친숙한 통속심리학에 의해 규정되는 인지과학 연구는 이제 종말을 맞는 것으로 보인다고 말할 수 있다.²⁰⁾ 위와 같이 클락은 대체적으로 새로운 표상 시스템이 가져다 줄 학습 모델의 이론과 관련한 성과, 즉 전망(2)에 동의한다.

3-2 미래 인지과학에서 표상의 역할

그러나 클락이 전망하는 새로운 표상이론의 역할은 여기까지이다. 그것으로부터 그는 전망(1)을 수용할 수 없다고 아래와 같이 밝힌다.

통속심리학을 위한 그 함축(보다 정확히 말해서, 통속심리학 체계 전체에 대한 우리의 신중한 평가를 위한 함축)은 그다지 명확하지 않다. 통속은 다른 사람의 심적 상태를 규정하기 위한 수단으로 정말로 문장적

19) 같은 책, 88쪽 참조. 클락은 첫째를 Oaksford and Chater(1991)에서, 둘째를 P.M. Churchland(1989, ch. 10)에서, 셋째를 P.S. & Sejnowski(1990), A. Clark 1990에서, 넷째를 A. Clark, 1990에서 가져왔음을 밝힌다.

20) 같은 책, 92-3쪽 참조.

형식화를 이용한다.²¹⁾

위와 같이 클락이 처첼랜드 입장에 반대하는 동기는 통속심리학이 근본적으로 우리 일상의 사고와 행동에 대해 풍부하고 유익한 설명의 이해를 제공한다는 점을 무시할 수 없다는 것이다. 클락은 그러한 고려에서 과연 새로운 표상 이론의 자원이 통속심리학의 운명을 바꿔놓을 것인지를 다시 검토하고, 그 검토로부터 통속심리학의 문장적 탐구방법과 관련하여 새로운 표상이론이 미래 인지과학의 발달 전망에서 어떤 역할을 담당할지 묻는다. 즉 과연 통속심리학이 새로운 내적 표상이론으로 설명될 수 없다는 점이 문장적 접근이 제거될 것임을 함의하는가? 클락은 이러한 질문에 대해 부정적인 대답을 기대하게 되는 근거를 아래와 같이 제시한다.

첫째, 통속심리학은 인간의 심적 상태를 규정하기 위해 문장의 형식화를 이용한다. 그렇게 문장으로 표현한다는 것이 곧 '유사-문장 구조의 존재'를 위임하는 것은 아니다. 단지 문장으로 표현된 특정 심리상태를 지지하는 '상태'가 존재함을 말할 뿐이다. 그렇다면 실재론의 관점에서 연결주의자가 주장하는 내적 표상이 순수하게 비-문장적이라고 해서, 그것이 반드시 통속의 '언어적 논'을 금지시켜야 할 이유는 없어 보인다는 것이다.

둘째, 연결망의 활동 공간 내에 하부-용량(sub-volume)이 통속심리학적 내용을 부호화할 것으로 기대하는 것은 비현실적인데, 왜냐하면 통속적 논의에서 보여주는 내용은 연결망 시스템의 '전체 속성'이라고 보는 것이 옳기 때문이다. 그런 증거를 뇌 손상 환자들에 대한 연구에서 찾아볼 수 있다. 특정 뇌 영역의 손상은 빈번히 전혀 예측하지 못한 결핍과 분열을 보여준다. 한 마디로, 문장적 방법이 연결주의자 연구와 모순되지 않는다. 그러므로 그 연구결과가 통속심리학의 잘못을 지적할 수 없을 것 같다는 것이다.

셋째, 같은 맥락에서 주장되는 바, '통속심리학적 항목들, 즉 인간의 지식, 얼굴 재인, 일상적 사물의 확인 등이 신경망의 내적 영역에서 분해된다'는, 즉 '추상적 내용을 나타내는 위상공간의 특정 벡터가 여러 하부 위

21) 같은 책, 89쪽.

상공간의 요소 벡터들로 분해된다'는 처칠랜드의 기대가 심적 상태의 통속 존재론을 무너뜨리는데 실패한다. 사실상 기뢰를 파악하는 시스템이든 사물을 인지하는 복잡한 인간 신경계이든, 그것을 '수행하는 것'은 하부-시스템이 아니며 '전체 시스템(global system)'의 기능이다. 심리학적 '이행의 차원'에서 볼 때 결코 내적 세분화가능성은 어디에도 존재하지 않는다. 더구나 통속적 이야기가 해석하려는 것은 시스템 전체의 복잡한 능력이며, 파편화된 내적 자원에 의해 지원되는 능력이 아니다. 따라서 뇌의 내적 경계가 통속심리학과 상관없는 비-문장적이며 파편적이라는 점이 결코 통속심리학을 제거시킬 근거가 되기 어려워 보인다.²²⁾ 이상의 근거들로부터 클락이 다음과 같이 반론(1)을 주장한다고 볼 수 있다.

반론(1): 통속심리학이 새로운 내적 표상이론으로 설명될 수 없다고 해서, 그것이 문장적 접근이 제거될 것임을 함의하지 않는다.

나아가서 그의 주장에 따르면, "일단 우리가 통속적 실천을 재능과 지식 전체에 대한 주석으로 보기만 한다면, 우리의 일상적 행동은 심지어 후자(연결주의자)의 발견이 통속적 기술의 진리와 양립가능할 것임을 드러낸다."²³⁾ 그 양립가능성에서 보았을 때, 그는 자연스럽게 다음의 의문, 통속심리학은 영성하게 획득된 것으로, 그것이 내적표상이론에 어떤 역할도 제공하지 못하는지 의문해본다. 역시 그 의문에도 부정적 대답을 하게 되는 근거를 아래와 같이 제시한다. 통속심리학이 영성하게 획득되었다지만, 그럼에도 일상적으로 잘 작동한다. 뿐만 아니라 둔감한 통속심리학은 내적 표상들이 작동하기 위한 기반이다. 통속심리학이 비-문장적 표상들을 자극하여, 결국 통속의 믿음과 욕망 등이 잘 작동한 것이라 생각해보면, 무엇을 제거시켜야 한다는 것인지 의문된다는 것이다. 예를 들어, 우리가 주위 사람들과 또는 문서를 통해 '언어로' 정보를 주고받는 것이 서로의 뇌의 내적 표상에 영향을 미칠 것이며, 그 내적 표상에 의해 우리는 일상적 사고에서

22) 같은 책, 89-91쪽 참조.

23) 같은 책, 91쪽.

행동패턴과 관련하여 유용한 지식과 기대를 이끌어낼 것이다. 그 유용하고 풍부한 기능은 내적 표상에 의해 작동된다고 하지만, 그것에 영향을 주는 것이 바로 일상 언어적 구조에 의한 정보들이다. 그렇게 통속심리학적 문장들은 지식을 풍성하게 하는 일에 진정으로 기여한다. 그 점에서 클락은 내적 표상들에 의하여 통속이 제거될 이유는 없으며, 오히려 통속은 내적 표상들을 작동하게 해주는 기반이라고 주장한다.

나아가 위 주장을 확장하면 다음을 주장할 수 있다. 처칠랜드도 인정했듯이, 통속심리학의 언어적 표현인 “D-N설명모델은 정보의 교환에 아주 적절하다.”²⁴⁾ 그리고 그것이 내적 표상이론을 자극하여, 우리가 새로운 사고를 얻을 수 있는, 지식의 풍부함을 열어놓는 것이다. 그 정보의 교환에 의해 우리는 새로운 유형의 행동패턴 목록들을 확장시킬 것이다.²⁵⁾ 바로 그렇게 통속의 논의는 우리 설명의 이해를 긍정하고 확장하게 해준다.²⁶⁾ 이상의 근거들로부터 클락이 아래와 같이 반론(2)을 주장한다고 볼 수 있다.

반론(2): 통속심리학은 영성하게 획득된 것이지만, 내적표상이론이 해야 할 일을 풍성하게 도와줄 것이다. 따라서 제거되지 않는다.

이상과 같이 통속심리학의 문장적 분석에 의한 인지과학적 탐구의 비-제거주의 견해는 매우 설득력 있는 논리를 보여주는 것처럼 보인다. 우선 반론(1)은, 통속심리학의 차원과 신경과학의 차원이 각기 설명의 목적과 기능을 달리한다는, 관점에서의 방어이다. 사실 세계와 대화하는 것은 유기체 전체의 기능이지 그 부속들의 기능이라고 말하기 어렵다. 이 논증의 힘은 콰인의 전체론적 인식론과 관련시켜볼 때 더욱 돋보인다. 처칠랜드(1979)의 기본 입장은 콰인의 인식론적 관점에서 출발하고 있다. 콰인(1969)은 토대주의적 환원주의에 강력히 제동을 걸었으며, 그 배경에서 자연주의

24) P.M. Churchland(1989), 224쪽 참조.

25) 이것과 동일 관점에서 나온 쟁점에 관한 연구로 이영의(2004)를 참조.

26) A. Clark(1996), 91-100쪽 참조.

(naturalism)를 지지했다. 따라서 콰인의 인식론과 자연주의 관점을 지지하는 처칠랜드가 반론(1)을 넘어서기 쉽지 않아 보인다.

다음으로 반론(2)는 하위차원의 새로운 표상이론의 자원이 상위차원의 통속심리학의 문장적 탐구방법을 제거하기보다, 설명해야 한다는 관점에서 반론이다. 이 반론 역시 앞의 것과 마찬가지로, 차원간 관계의 분석에 기초한 주장이다. 사실상 심리 현상과 생리적 현상을 설명하는 이론들, 즉 상·하 차원의 이론들은 각기 다른 목적과 기능을 가지며, 따라서 하위 차원의 설명이 상위차원의 설명을 제거한다고 주장하기보다, 오히려 풍부하게 지원할 것을 기대하는 것이 더 자연스런 생각이라는 주장을 수용하지 않을 수 없을 것 같다. 그런 시각에서 보면, 처칠랜드의 하부-인지 표상 자원을 밝혀주는 이론이 상위-인지 문장적 표상 자원에 근거하는 통속이론의 제거, 즉 제거적 유물론을 유도한다고 믿기 어려워 보인다. 이상과 같은 전망으로 볼 때, 처칠랜드의 전망(2)로부터 전망(1)의 정당화는 불가능해 보이며, 클락의 말대로 처칠랜드의 유물론은 오히려 통속 논의를 풍부하게 확장시켜줄 것을 기대하게 하는 측면이 있다.

그러나 처칠랜드의 입장이 클락의 주장에 의해 쉽게 무너지지는 않을 것이다. 폴 처칠랜드는 클락의 반박에 정면의 반박을 시도한다. 우선 그러한 시도가 얼마나 설득력을 갖는지, 클락의 반박을 넘어설 수 있는지를 알아보기로 하자.

4. 통속심리학의 빈약함

4-1 클락 논의의 근거들

처칠랜드는 “통속심리학에 대한 클락의 연결주의적 변론”(Clark's Connectionist Defense of Folk Psychology, 1996)에서 앞의 클락의 주장을 아래와 같이 요약한다. 클락 주장의 가장 핵심은, 첫째, ‘통속심리학과 과학적 심리학이 동일한 목적을 갖지 않는다’는 차별성에 대한 고려이다. 둘째,

그 차별성에 근거하여 ‘현대 과학적 이론들의 기준을 통속심리학에 대해 엄격히 적용하지 말아야 한다’는 추론이다. 셋째, 통속심리학은 사회적 접촉물로서 연속적 형식을 가지며, 그러면서도 그것의 합법적 생존이 연결주의 과학적 함축과 매우 일관성을 가질 가능성이 있다. 그리고 그러한 클락의 주장들에 대한 세부적 근거를 아래와 같이 네 가지로 분석한다.

- 근거(1) 엄격히 거짓이지만, 그럼에도 매우 유용한 인과적/설명적 이론의 실용적 관용
- 근거(2) (환원적) 동일성 이론의 관대한 해석
- 근거(3) 데닛주의 지향적-자세 도구주의의 한 형태
- 근거(4) 통속심리학을, 교정할 수 있는 사실적인 지지의 체계보다 순수한 사회적 실천의 존재론적으로 순결한 형태로의, 대안적 초상화²⁷⁾

위의 (클락의) 네 근거들은 제각기 나름의 호소력을 갖는 것으로 비쳐질지 모르지만, 처칠랜드의 관점에 따르면, 사실 서로 양립 불가능한 것들이다. 그 점에서 클락은 문제를 명쾌하게 파악하지도 못하고 그것들에 근거해서 확대된 호소를 이끌어내려 시도한다고 아래와 같이 지적된다.

근거(1)은 역사적 사례로도 호소력을 갖는데, 뉴턴 역학을 아인슈타인 역학에 비추어 보면, 그것이 엄격히 참은 아니지만 실용적 측면에서 충분히 유용한 인과적 설명을 제공한다고 말할 수 있다. 뿐만 아니라 형식적으로도 그것을 논증해보일 수도 있다. 아인슈타인 역학의 체계 내에서도, 물질의 속도가 광속과 아주 크게 차이 날 경우 뉴턴 역학과 등식으로 설명될 수 있다. 더구나 아인슈타인 역학의 체계는 일상의 생활 영역에서 어떤 실천적 유리함도 제공하지 못한다. 그것을 통속심리학에 유비적으로 끌어들이 제거주의 방어에 적극 활용하는 것이 호소력을 줄 수 있다. 그러나 처칠랜드는 그것은 좋지 않은 유비라고 지적한다. 실제로 아인슈타인 역학이 뉴턴 역학보다 일상생활에 유용성을 제공하지 않지만, 새로운 인지신경심리학은 그렇지 않기 때문이다. 이미 우리는 신경학적으로 그리고 정신병리

27) P.M. Churchland(1996), 250-1쪽.

학적으로 환자의 뇌에 미묘한 손상과 관련하여 화학적이며 생리학적인 설명을 유용하게 활용하기 시작했다. 그 점에서 클락의 근거(1)은 자신의 반-제거주의를 주장하기 위한 힘을 상실한다는 것이다.

근거(2)도 호소력이 전혀 없지 않아 보이는 것은 계산적 신경과학의 자원들이 (적어도 특정 차원들에서) 명제태도에 어찌면 대담해질 가능성이 있기 때문이다. 그런데 문제는 클락은 그것을 기대하면서도 동시에 명제태도의 인과적 설명 역할 역시 그대로 보존되길 기대한다는 점에 있다. 만약 계산적 하부-신경 구조에 의한 통속심리학의 일부가 설명된다고 하더라도, 미래의 표상이론과 학습이론의 체계는 명제태도의 고전적 이론들의 체계와는 아주 어울리지 않는 것들이다. 그 서로 다른 체계들이 어울릴 수 있을 것으로 기대하는 클락은 동일성 이론에 너무 관대한 전망을 기대하는 측면이 있다. 이러한 처칠랜드의 방어를 아래와 같이 이해할 수 있다. 만약 통속심리학적 현상들에 대해 현재 또는 미래의 신경과학적 이론들이 설명해주는 경우가 발생하더라도, 그것을 통속심리학의 ‘생존’이라 말하기보다 ‘대체’라고 말해야 할 것이다. 우리의 행동에 대한 설명과 예측에서 통속심리학이 하던 역할을 신경인지과학과 연결주의 인공지능에서 나온 새로운 이론들이 설명을 대신한다면, 그것을 통속심리학의 생존이라 주장하기는 어렵기 때문이다. 이러한 고려에서 클락의 근거(2)는 자신의 입장을 옹호하기에 도움이 되지 못한다.

근거(3)은 계산적 표상이론을 처칠랜드가 “계량적 내용(calibrational content)”²⁸⁾이라 호칭한 것에 대비하여, 클락이 통속심리학적 이론을 “둔감한 도구(blunt instrument)”라고 호칭한 것의 구분에서 나온다. 전자를 정교한 과학의 영역이라면, 클락의 견해에 따르면 후자는 일상적으로 훌륭한 설명과 이해를 제공하는 통속심리학의 영역이라는 주장이다. 그리고 전자의 영역은 뇌의 국소적 탐구 대상이라면 후자는 뇌의 전체적 상태에 의해 나오는 것으로 대략적으로 다듬어진 일종의 진리라는 주장으로 이해된다. 그러나 이런 주장은 이중적이다. 다시 말해서 “엄격히 P가 아니지만,

28) P.M. Churchland(1989), 43쪽 참조.

그러나 대략적으로 여전히 P'라는 주장처럼 둘러대는 이야기에 불과하다. 이점에서 근거(3) 역시 설득력을 갖기 어렵다.

근거(4) 역시 차원간 구분으로부터 나온다. 기술적 차원의 사실적 설명은 교정될 수 있는 것이며, 따라서 새로운 이론에 의해 그 이론의 대체와 제거가 일어날 수 있다. 그렇지만, 실천적 모습을 갖는 통속심리학의 실천적 설명은 그 자체가 일상적으로 필수적인 순결성을 가져서 기술적 차원의 이론에 의해 제거나 교정되지 않는다는 주장이다.²⁹⁾ 처칠랜드가 보기에, 이러한 사회적 실천에 대한 비판적 관점은 퍼스(C.S. Peirce)의 철학적 실용주의로부터 나온 것으로, 기본적으로 통속심리학이 사회적 실천의 가치를 갖는다는 점에 동의하지 않을 사람은 없을 것이다. 그러나 그렇다고 '하나의 설명적 이론으로서 통속심리학'과 '하나의 실천적 형태로서 통속심리학'을 배타적으로 구분하는 윌키(Kathleen Wilkes, 1984)의 입장을 끌어들이는 클락의 입장은 (처칠랜드의 관점에서) 지지되지 않는다고 지적된다.

사실 이론들이란 전형적으로(typically) 약간의 복잡한 실천 또는 다른 것을 지지한다. 화학, 전자, 야금학, 음향학, 역학, 세포생물학 등등을 생각해보라. 더구나 그런 이론들의 내용은 그 지지되는 실천의 형식을 설명한다.³⁰⁾

이론들은 근본적으로 그것이 지지하는 실천의 형식을 설명한다. 더구나 하부-인지 표상이론과 학습이론들은 뇌에 어떻게 전형과 맵을 형성하며, 그것이 어떻게 행동조절을 이루어내는지 설명해줄 이론이다. 다시 말해서

29) 여기서 '사실적 이론'과 사회적 실천의 이론으로서 '통속심리학 이론'의 구분은, 명시적 지식과 암묵적 지식의 구분으로 대비된다. 사실 전통적으로 인식론을 탐구해온 학자들은 암묵적 지식과 명시적 지식을 구분하면서도, 그것들의 관계를 해명하는데 실패해왔다. 이제 처칠랜드의 신경인지과학과 연결주의자 인공지능에서 나온 이론들을 끌어들이면, 그것들이 통합적으로 설명될 수 있음을 볼 수 있다. 그 설명에 따르면 문장적 이해란 우리 인지의 본질적 속성에 기초하지 않으며, 단지 계산적 표상이 구성하는 맵에 의해 드러나는 현상일 뿐으로 이해된다.

30) P.M. Churchland(1996), 254쪽.

그 이론은 실천적 효과가 어떻게 발휘되는지 이해와 설명을 제공한다. 따라서 그 미시적 이론의 설명은 실천적 효과를 설명해줄 이론이다. 그 점에서 이론과 실천은 그렇게 명확히 구분되지 않는다. 나아가서 쿤(1962)의 관점에 따르면, 이론의 습득은 '상호 관련된 기술들의 집합'이다. 그것은 과거 경험으로부터 형성되는 패러다임(paradigms) 또는 범례(exemplars)로 실천적 능력을 함께 갖는다.³¹⁾ 결국 위키와 클락은 불필요한 (넓은) 구분법을 끌어와 자신들의 의도를 설득하려 했음이 드러난다.

정말 우리는 새로운 이론들의 발달과 함께 보다 나은 실천의 능력을 기대한다. 신경과학과 인공지능에서 나온 우리의 정서, 감동, 발달, 성격, 병리학, 다른 사람의 행동 등에 대한 이해가 새로운 이론에 의해 개선될 때, 우리는 새로운 사회적 실천의 힘을 얻는다. 이러한 검토를 통해서도 근거(4)는 클락의 목적을 달성하는데 실패한다. 처칠랜드는 자신의 방어에 끝을 다음을 반문하고 대답한다. "과학적 심리학이 우리의 현재 통속심리학과 동일한 목적을 지원하는가? ... 그렇다"³²⁾

지금까지 논의를 통해, 클락의 논의가 적절한 근거를 갖지 못한다는 것을 보여주기에 부족함이 없어 보인다. 그렇지만 그의 논박이 전망(2)로부터 전망(1)로의 주장, 즉 과학적 이론에 근거하여 형이상학적 논의로 이끌어가는 제거적 유물론의 주장이 설득되기에 충분한가? 한 마디로, 우리는 이제 안심하고 처칠랜드의 주장에 동의할 수 있겠는가? 위의 논의만으로는 아직 그렇다고 대답하기에 머뭇거리려질 것 같은데, 그것은 위의 처칠랜드의 방어는 단지 클락이 지적하는 논박의 근거들에 대한 수동적 방어만을 보여준 것일 뿐, 자신의 주장을 적극적으로 논증해보여주지 못하기 때문이다. 따라서 클락의 지적에 대해 보다 적극적으로 제거적 유물론을 설득할 논증을 처칠랜드 체계 내에서 찾아볼 필요가 있다.

31) P.M. Churchland(1989), 158쪽 참조.

32) P.M. Churchland(1996), 255쪽.

4-2 개념체계의 통일

과연 전망(1)이 전망(2)에 의해 지지되는가, 즉 전망(2)가 전망(1)을 함의하는가? 앞(3-2, 3-3)에서 드러나듯이 그렇지 않다는 클락의 반론을 필자는 다음 두 가지 관점으로 집약한다. 첫째는 차원간 관계에 비추어 논의되는 문제이고, 둘째는 통속심리학의 유용성에 비추어 논의되는 문제이다.

우선 첫째 관점에 따르면, 통속심리학과 하부-인지 표상이론은 각기 다른 차원의 설명 목적과 기능의 역할을 갖는다는 근거로부터, 새로운 표상이론이 통속심리학의 제거를 요구하지 않으며, 둘은 양립가능하다는 주장이다. 만약 그 주장이 설득력을 갖는다면 전망(2)로부터 전망(1)을 논증하지 못할 것이다. 반면에 그러한 차별성에도 불구하고 새로운 표상이론이 통속심리학의 제거를 요구할 수 있겠는가? 즉 전망(2)로부터 전망(1)을 추론할 수 있겠는가? 이 논의를 위해 이론의 차원간 관계 자체가 어떤 본성을 갖는지 비판적으로 검토될 필요가 있다.

통속심리학은, 처칠랜드에 따르면, 근본적으로 일상적인 우리의 행동을 설명하고 예측하게 해준다는 점에서 하나의 이론³³⁾이며, 그것은 우리의 외적 행동을 내적 믿음, 욕망 등의 자원에 의해 설명한다. 그렇다면 통속심리학의 제거 여부의 문제는 서로 다른 차원의 두 이론들 사이의 관계에 관한 논의로 초점된다. 그 논의의 쟁점은 우리 인지 또는 심리의 상위차원과 그 내적 표상을 설명하는 하위차원의 양 이론들 사이에 환원적 관계의 문제임을 알 수 있다.³⁴⁾ 처칠랜드의 관점에 따르면, 상위차원의 이론과 하위차원의 이론은 다양한 관계에 놓인다. (여기서 양립가능성의 문제는 처칠랜드 논의에서 쿤의 공약가능성 문제로 파악된다.) 만약 두 이론의 관계가 양립가능할 경우, 즉 공약가능할 경우 상위차원 이론은 하위차원 이론과 부드러운 환원에 의해 동일성이 확보될 수 있다. 만약 두 이론의 관계가 그렇지 못할 경우, 즉 공약불가능할 경우, 두 이론들 중에 일부가 수정되거나, 아니면 제거되는 일이 일어날 수 있다.³⁵⁾ 어떻게 그럴 수 있다는 것인가?

33) 이런 견해를 P.M. Churchland(1989), 2-6쪽, P.S. Churchland(1986), 299쪽에서 볼 수 있다.

34) 이런 쟁점의 논의를 이영의(1994)와 고인석·이중원(2007)에서 찾아볼 수 있다.

처칠랜드는, 쿤(1962)의 패러다임의 개념을 끌어들여, 이론들 사이의 관계를 이론체계, 또는 개념체계의 관계로 확대해석한다.³⁶⁾ 쿤에 따르면 과학의 발전에 따라 기존 이론과 새로운 이론의 공약불가능성이 높아질 경우, 이론들의 개념체계인 패러다임이 교체되는 일이 불가피하게 일어난다. 그것은 우리의 개념체계의 통일성을 위해서이다.³⁷⁾ 이런 이야기를 새로운 표상이론과 통속심리학의 관계에 접목시켜보자. 만약 통속심리학이 새로운 표상이론과 양립불가능성, 즉 공약불가능성을 높이는 쪽으로 나아간다면, 다시 말해서 앞으로 새로운 표상이론의 개념체계와 통속심리학의 개념체계 사이에 공약불가능성이 높아지는 쪽으로 진행된다면, 어느 쪽이든 개념체계의 교체가 불가피하다고 말할 수 있다.³⁸⁾ 다시 말해서 전망(2)에 의해서 전망(1)이 불가피하다고, 즉 전망(2)는 전망(1)을 '함의한다'고 주장될 수 있다.

문제는 통속심리학과 새로운 표상이론의 관계가 양립가능한지, 그리고 공약가능한지 의문이다. 근본적으로 새로운 표상이론과 통속심리학이 양립 가능하지 않은 점은, 그 배경이론들과의 관계에서 드러난다. 새로운 표상이론은 인지신경생물학의 연구결과와 함께 연결주의자 AI에 의해 탄생된 이론이다. 그렇다면 새로운 표상이론과 통속심리학의 관계는 단지 그 이론이 통속심리학적 특정 현상을 설명해줄 가능성을 갖는지에 국한하지는 않으며, 그 관련 분야의 연구 성과들이 통속심리학 이론을 지지하는지를 검토해야 한다. 이런 생각만으로도 그 양립가능성은 불신된다.³⁹⁾

이 시점에서 다음과 같은 의문이 일어난다. 위와 같이 공약불가능할 경우 한 이론이 다른 이론을 제거하는 일이 왜 일어나야 하며 어떻게 일어나

35) P.S. Churchland(1986), 7-9장 참조.

36) P.M. Churchland and P.S. Churchland(2001), 42-23쪽 참조.

37) 보다 자세한 논의를 박제운(2008) 참조.

38) 이론간 관계에서 이런 환원적 논의를 P.S. Churchland (1982)에서 볼 수 있다.

39) 클락은 한편으로는 그 양립가능성을 주장하면서도 사실상 그 주장을 아래와 같이 스스로 부정하기도 한다. “미래 인지과학의 설명 도구가 ‘현재 상식의 문장적 범주’에 기댈 것이 거의 혹은 전혀 없다는(P.M. Churchland 1989, 177쪽), 처칠랜드의 근본적 믿음은 따라서 내가 전적으로 분명히 긍정하는 것이다.” A. Clark(1996), 97쪽. 이 점은 본 논문의 3-1에서도 드러난다.

는지 자체에 대해서도 새로운 패러다임의 이론(현대 신경과학 이론)이 설명할 수 있을까? 우리 뇌가 왜 그렇게 개념체계의 통일성과 단순성을 추구하도록 경향 되어 있는지 에델만(Gerald M. Edelman)의 책, 『신경과학과 마음의 세계』(Bright air, Brilliant Fire, On the Matter of the Mind)⁴⁰⁾에서도 찾아볼 수 있다. 그의 주장에 따르면, 뇌의 각 부위 혹은 영역들은 나름대로 특정 기능을 수행하는 또는 정보처리 하는 여러 부분의 맵들(local maps)을 가지고 있다. 그 맵들은 개념적 도구로서 범주를 담아낸다. 그 범주의 맵들을 통해 뇌는 여러 기능을 위해 상호적으로 정보처리를 수행한다. 그런데 그런 기능 중에 맵들은 상호적으로 다른 맵들을 수정하는 일을 수행하는데, 그것을 그는 “재입력(reentry)”이라 말한다. 그 재입력의 과정에서 뇌가 맵들 사이에 개념적 수정을 가하는 것으로 이해할 수 있다. 맵들 사이에 상호 전체적 대응연결을 이루며(global mapping), 그것을 통해 상호간에 행동의 레퍼토리를 수정한다. 그렇게 맵들 사이에, 다시 말해서 신경계 내에 범주로서의 레퍼토리를 담아내며, 동시에 전체적 조율을 이뤄낸다. 이러한 이야기로부터 다음이 추론된다. 뇌는 나름의 개념체계의 통일성을 요구하는 본성을 갖는다. 이러한 작용은 개별 뇌의 차원을 넘어서도 작동할 수도 있다. 나아가서 정보교환의 수단(언어)을 통해 나의 개념체계는 남의 개념체계와 일치하는 것을 요구하며, 그 요구의 목적은 바로 정보교환의 의미이며, 나아가서 사회적 생활에서 생존의 유리함이다.

뇌의 연결망의 국소 맵들은 통속심리학의 개념체계와 새로운 인지적 표상이론의 개념체계 사이에 가능한 일치점을 찾으려 할 것이지만, 만약 그렇지 못할 경우 맵 전체 체계의 일관성을 위해 통속심리학의 개념을 부정할 것이다. 그런데 부정되어야 할 것이 왜 통속심리학 쪽이어야 하는가? 그것은 각 이론체계를 지지하는 배경이론들과의 관계에 의해서라고 말할 수 있다. 새로운 표상이론을 지지하는 배경 이론들이 유력하고 발전 전망이 높은 이론이라 평가되기 때문이다. 반면에 통속심리학을 지지하는 배경이론들이란 말 그대로 통속의 것들에 불과하다. 그 점에서 우리는 (통속심

40) G. Edelman(1992), 9장 참조.

리학 이론의 대대적 수정내지) 제거적 유물론에 강요된다.⁴¹⁾

둘째 관점, 통속심리학의 문장적 접근의 유용성과 관련한 논의에 따르면, 통속심리학에 기초하는 공적 언어는 정보 소통의 수단임이 강조된다. 사실 우리는 대부분 문장적 언어를 통해서 상호 정보를 교환한다. 그러므로 우리는 통속심리학의 바탕에서 상호 정보 교환한다고 주장될 수 있다. 그러한 이유로 우리에게 통속심리학이 불필요해지는 일은 결코 발생하지 않을 것이라고 말할 수 있다. 오히려 우리는 통속심리학의 방식에 의해 보다 풍부한 정보를 교환하고 발전시켜야 한다. 그 점에서 제거적 유물론은 단편만을 바라보는 극단적 주장으로 보일 수도 있다.

그러나 이러한 논박은 처칠랜드가 제거적 유물론을 통해 주장하려는 핵심을 올바르게 바라본 지적이라고 인정되기 어렵다. 처칠랜드 관점에 따르면, 공적 언어의 문장적 접근은 (앞서 살펴보았듯이) 인지적 본성을 탐구하기에 한계를 갖는다. 따라서 공적 언어구조에 의해 우리 인지적 기초를 탐구하는 것이 옳지 않다고 주장된다. 오히려 우리의 인지적 표상의 본성에 대한 탐구는 신경망의 구조와 그 알고리즘의 본성들을 탐구하는 태도에서 새롭게 출발해야 한다는 주장이다. 분명 그 주장은 앞으로 일상생활에서 공적 언어 사용 자체를 중지해야 한다는 주장을 함의하지는 않는다. 처칠랜드 자신도 공적 언어를 통해 자신의 이론을 주장하고 설득한다. 그럼에도 우리의 인지적 능력에 대한 탐구 방향이 언어적 차원의 분석에 근거해야 한다는 주장은 옹호되지 않는다. 앞에서 살펴보았듯이, 새로운 표상이론은 우리 인지의 기초 자원을 구문론적 측면을 부정하며 기하학적 측면을 긍정한다.

나아가서 처칠랜드는 공적 언어가 우리 일상생활에서조차 필수적이지 않을 가능성을 다음과 같이 시사한다. 미래에 신경과학의 발달과 전자통신 문명의 발달에 의해서 우리 각자의 두뇌에 각자 통신 장비를 장착하여, 대화 없이 서로의 생각을 소통할 수 있을 가능성을 가정해보자. 그 경우 우리는 정보교류를 위해 반드시 문장구조의 언어를 주고받을 필요는 없어도

41) 보다 자세한 논의를 박제윤(2008) 참조.

좋은 것이다.⁴²⁾ 그 점에서 소통의 유용성에 근거하여 인지적 탐구방법에서 우리가 문장 구조에 집착해야 할 이유는 없어 보인다. 이미 현대 전자 문명 시대에 소통의 자원인 디지털 신호는 사실상 문장 이하의 수준에서 교환된다. 그러므로 공적 언어의 유용성에 근거하여 통속심리학의 문장적 접근이 우리의 인지적 탐구와 관련하여 옳은 연구방법임을 주장하거나, 우리가 언어로 소통한다는 점에 근거하여 현대 인지신경생물학과 연결주의 AI 연구 결과가 지지하는 처칠랜드의 제거주의를 무력화하려는 기도는 사실 사소한 논의가 될 뿐이다.

오히려 인지적 탐구의 목적에 비추어 볼 때, 통속 심리학의 문장적 접근과 새로운 표상이론의 벡터적 접근 중에 어느 것이 더 유용성을 가질지 이 시점에서 반문해볼 수 있다. 통속 심리학의 문장적 접근은 인지적 탐구에서 몇 가지 근본적인 중요한 결함을 가지고 있음이 지적된다. 그 접근은 인간을 포함한 동물 일반의 인지적 연구에 도움을 줄 가능성이 없으며, 인간의 경우 언어 장애를 가지거나 아직 언어를 배우지 못한 어린 아이의 인지적 능력을 설명해줄 가능성이 없어 보인다는 점이다.⁴³⁾ 그 점에서 통속 심리학 이론에서 제안되는 표상 이론의 탐구와 문장적 접근은 유용성보다 빈곤함을 드러낸다. 반면에 새로운 표상 이론은 분명 현대 신경과학 관련 여러 분야들로부터 지지되며 공약가능하다는 점에서 미래에 보다 풍부한 결실을 가져다 줄 전망이다. 그러한 고려에서, 그리고 앞서 지적했듯이 두 이론들 사이에 공약불가능성이 유지된다는 점에서, 우리는 제거주의적 전망을 하게 됨을 이해할 수 있다.

5. 결론

처칠랜드의 제거주의와 관련하여 클락은 두 가지 관점에서 반론하였다.

42) P.M. Churchland(1989), 20-1쪽 참조.

43) P.S. Churchland (1982), 1045쪽 참조.

그 반론들에 대한 방어를 아래와 같이 요약할 수 있다.

첫째, 반론(1)은 통속이론과 미시물리적 이론의 설명 차원이 다르다는 점에 주목하였는데, 특히 통속이론은 전체적 이해의 차원이라는 점에서 미시물리적 이론에 의해 설명될 수 없다는 환원불가론을 주장한다. 그러나 쿤의 “패러다임”과 “공약불가능성” 개념을 끌어들이 보면 이론들은 전체 체계의 통일성을 위해 차원을 넘어 상호 간섭하여 다른 이론의 수정과 제거를 일으킬 수 있음이 드러난다. 더구나 이러한 제거주의 논의는 지식체계의 전체적 통일성을 주장한다는 점에서 콰인의 전체론적 인식론과 일관성을 갖는다. 따라서 미시적 차원의 새로운 표상 이론과 설명에 의해서 거시적 차원의 통속 심리학 이론과 설명을 무력화하는 일이 일어날 수 있음이 드러난다.

둘째, 반론(2)은 통속심리학이 대단한 유용성을 가진다는 점에 주목하였으며, 따라서 새로운 표상이론과 학습모델 이론은 그것의 제거보다는 오히려 통속심리학을 더 풍성하게 해줄 것이라는 예측이다. 그 예측은 분명 제거주의에 반대하며, 환원적 동일성을 지지함을 함축한다. 따라서 이 주장은 환원불가론을 주장하는 반론(1)과 논리적 일관성을 잃고 있음이 드러난다. 나아가서 계량적 표상이론은 인지신경생물학과 연결주의 AI 연구로부터 탄생한 것들로, 그 개념체계는 통속의 개념체계와 전혀 어울리지 않는다. 그 점에서 통속과 새로운 표상이론은 개념체계의 통일성을 이루기 어려움, 즉 양립불가능성을 드러낸다. 구체적으로 새로운 표상이론들이 제안하는 많은 인지적 현상들은 문장의 구조에 의해 탐구될 가능성을 배제한다. 만약 우리의 일상적 언어가 벡터로 표현되는 내적 표상에 근거해서 어떻게 작용되는지 설명된다고 하더라도, 문장적 탐구 방법은 유지되기 어렵다고 여겨지며 환원적 동일성은 기대되지 않는다.

우리가 특정 이론들의 동일성, 수정, 교체 등을 일으키도록 강요하는 것은 우리의 전체 개념체계에 의해서이다. 그 점에 대해서도 현대 신경인지생물학과 연결주의 AI에서 나온 새로운 표상이론의 개념체계 내에서 설명이 가능하다. 뇌는 세계에 대한 적절한 반응을 위해 신경세포 집단의 국소 맵에 전형들을 담아낸다. 그 전형들은 동물 행동의 표준적 대푯값으로, 감

각에 있어서는 세계의 대표적 표상들이며, 동시에 운동에 있어서는 대표적 행동 출력들을 산출한다. 그러한 전형들에 의해 과거의 경험을 담아낼 경우, 그것은 미래 행동을 위한 일반적 행동 패턴이기도 하다. 만약 그 맵들의 전형들이 사물에 대한 인지작용에 의해 형성된다면, 그것은 세계를 분별하게 해줄 개념적 범주들로 간주해도 좋을 것이다. 만약 그 맵들의 전형들이 관찰된 경험에 대한 전형으로 추상적 위상공간에 형성된다면, 그것은 다음의 경험을 위한 예측과 설명을 제공해줄 이론들로 간주해도 좋을 것이다. 그 이론들이 뇌에 국소 맵들로 각인되어지며, 반대로 그 맵들이 집단적으로 개념체계를 만들어 간다고 가정된다. 우리는 그러한 개념체계를 뇌에 담아내며, 그러한 개념체계에 의해 형성된 우리의 인지 능력은 뛰어난 것이어서, 문장적 언어로 표현하기 어려울 정도로 정교하고 미묘한 판단을 가능하게 한다. 끝으로, 지금까지 얻어진 현대 과학의 성과에 의해서 형성된 개념체계는 우리가 제거적 유물론을 선택할 것인지, 아니면 이원론을 선택할 것인지도 결정할 주제라고 말할 수 있다.

참고문헌

- 고인석 · 이종원 (2007): “현대 생물학의 환원주의적 존재론에 대한 성찰”, 『과학철학』 10(1), 113-137쪽.
- 박제윤 (2008): “이론간 환원과 제거주의”, 『과학철학』 11(2), 147-171쪽.
- 이영의 (1994): “환원론과 인지과학”, 『철학연구』 34(1), 173-190쪽.
- 이영의 (2004): “무엇이 적절한 연결주의 과학철학인가?”, 『철학적 분석』 9, 33-63쪽.
- Clark Andy (1990), “Connectionism, competence and explanation.” *British Journal for the Philosophy of Science*, 41, 195-222쪽.
- Clark Andy (1996), “Dealing in Future: Folk Psychology and the Role of Representations in Cognitive Science,” in *The Churchlands and*

- their Critics*, Blackwell Publisher, 86-103쪽.
- Churchland, Patricia S. and Sejnowski, Terrence J. (1990), "Neural Representation and Neural Computation," *Philosophical Perspectives*, Vol. 4, 15-48쪽.
- Churchland, Patricia S. and Sejnowski, Terrence J. (1992), *The Computational Brain*, Cambridge, Mass.: MIT Press.
- Churchland, Paul M. and Churchland, Patricia S. (2001), "Intertheoretic Reduction: A Neuroscientist's Field Guide," *Philosophy and the Neruosciences*, Blackwell, 41-54쪽.
- Churchland Paul M. (1979), *Scientific Realism and the Plasticity of Mind*, Cambridge: Cambridge University Press.
- Churchland Paul M. (1981), "Eliminative Materialism and the Propositional Attitudes", *Journal of Philosophy* 78, No. 2, 67-90쪽.
- Churchland Paul M. (1989), *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, MA: MIT Press.
- Churchland Paul M. (1996), "Clark's Connectionist Defense of Folk Psychology," in *The Chruchlands and their Critics*, Blackwell Publisher, 250-255쪽.
- Churchland Patricia S. (1982), "Mind-Brain Reduction: New Light from the Philosophy of Science," *Neuroscience*, vol. 7, No 5, 1041-1047쪽.
- Churchland Patricia S. (1986), *Neurophilosophy*, Cambridge, Mass.: MIT Press.
- Churchland Patricia S. (1990), "Is Neruoscience Relevant to Philosophy?", *Canadian Journal of Philosophy; Supplementary Vol.*, 16, 323-341쪽.
- Churchland Patricia S. (2002a), *Brain - Wise*, Cambridge, Mass.: MIT Press.
- Churchland Patricia S. (2002b), "Neural Worlds and Real Worlds", *Natural Reviews Neuroscience* 3, (11) 903-907쪽.

- Gerald Edelman (1992), *Bright air, Prilliant Fire, On the Matter of the mind*, 황희숙 번역(1998) 「신경과학과 마음의 세계」, 범양사.
- Kuhn, Thomas (1962, 1970), *The Structure of Scientific Revolutions*. Second edition. Chicago: University of Chicago Press.
- Oaksford, M. and Chater, N.(1991), "Against logicist cognitive science." *Mind and Language*, 6, 1, 1-18쪽.
- Quine, W. V. O. (1969), "Epistemology Naturalized," in *Ontological Relativity and Other Essays*, New York: Columbia University Press.
- Wilkes Kathleen (1984), "Pragmatics in Science and Theory in Common Sense." *Inquiry*, 27, 4: 339-61쪽.

The Role of Representation and the Reductive eliminativism

Jeyoun Park

The Churchlands suggest a new theory of representation from the most recent results in computational neuroscience and connectionist AI, in which they foresee elimination of the theory of sentential representation on which folk psychology is based, in the light of the future role of the new representation theory. Interestingly, Andy Clark denies the eliminativism on the two reasons, the practical utilities, in the daily life, of the sentential representation theory and the differences of the roles and purposes of theories at levels, though he recognizes the impact of the new theory of representation. This article aims to compare the backgrounds of the two arguments, to consider which one is more reliable, and, from the consideration, to reconstruct a persuasive argument of the eliminative materialism which the new theory of representation entails.

[Key words] representation, reductionism, eliminativism, folk psychology, Churchland