

인지과학에서 설명의 문제

정 영 기*

1. 머리말

수십년 동안 설명의 문제는 과학철학의 중심문제이었다. 이제까지 설명에 대한 연구는 철학자에 의해 주로 진행되어 왔다. (물론 역사학이나 생물학이나 고고학에서도 설명의 문제가 논의되어 왔다.) E. Nagel은 <<과학의 구조>>에서 다음과 같이 언급하고 있다. “사실적인 증거에 의해 조직화하고 조절가능한 설명의 욕구가 과학을 만들어 낸다. 설명적 원리에 기초하여 지식을 조직화하고 분류하는 것은 과학의 구체적인 목표이다.”¹⁾

최근에 인지과학(cognitive science)의 등장으로 설명의 문제에 대한 새로운 접근이 시도되고 있다. 인지과학은 특히 인간의 행위 설명에 새로운 방식을 제공해 준다. 인지과학은 인간 행위를 설명할 때 정보처리적 관점에서 접근한다. 인지과학자들은 인간의 마음과 컴퓨터가 동일한 상징조작체계로써 환경에서 정보를 받아 이를 부호화하여 상징으로 저장하고 저장된 정보를 활용하여 출력을 내어 놓는 정보처리 체계로 본다.²⁾

인지과학은 인지(cognition) 또는 지능(intelligence)의 문제를 중심으로 여러 설명 수준에서 다양하게 접근하는 다학문적 과학이다. 인지과학은 인지 즉 앞의 문제가 철학, 심리학, 인공지능, 신경과학, 언어학, 인류학 등의 공통의 문제임을 보여주었다. 인지적(cognitive)이란 지각과 이해와 학습과 행위가 모두 사실수집과 가설형성, 추리, 문제해결의 모델에 의해 이해될 수 있다는 의미이다.

심리학자인 G.A.Miller는 설명에는 다음 세 가지 측면이 들어있다고 말한다. 첫째는 어떤 상황이나 사건의 이유에 대한 가설을 형성하는 것이다. 둘째는 다른 가설보다 한 가설을 선호하는 가설로 선택하는 것이다. 셋째는 다른 사람들이 이해하고 수

* 고려대학교 철학과 강사

1) E. Nagel(1961), *The Structure of Science: Problems in the logic of scientific explanation*, New York: Harcourt, Brace & World, p. 4.

2) 인간의 마음과 컴퓨터가 정보를 처리하는 정보처리체계라는 견해는 다음과 같은 몇 개의 가설로 구성되어 있다. 첫째, 인간의 마음은 정보를 처리하는 체계이다. 둘째, 정보처리는 계산, 기호를 조작하는 과정이다. 셋째, 컴퓨터의 프로그램은 기호를 조작하는 체계이다. 넷째, 인간의 마음은 컴퓨터의 프로그램으로 모형화할 수 있다.

락하는 방식으로 선택된 가설을 표현하는 것이다. 인지이론가들에 의하면 우리의 지식구조는 설명적 과정과 밀접하게 관련되어 있다. 어떤 사람의 설명은 그 설명이 기초하고 있는 지식구조에 대한 중요한 정보를 제공할 뿐만 아니라, 사람들이 지식구조를 개발하려는 중요한 이유 중의 하나는 지식구조가 설명을 제시하고 선택하고 옹호하는 방법을 제공하기 때문이다.

이 글에서는 인공지능, 인지심리학의 측면에서 과학적 설명의 문제를 고찰해보고자 한다. 먼저 2장에서는 P. Thagard가 인공지능에서 시도되고 있는 과학적 설명의 네 가지 접근법을 고찰해본다. 3장에서는 인지심리학자인 R. Cummins, J.Haugeland, Z.W.Pylyshyn의 설명이론에 대해 검토해본다.

2. Thagard의 네 가지 설명개념

P. Thagard에 의하면 철학에서 설명의 문제를 연구하는 데에는 여섯 가지 접근법이 있는데 이 중 네 가지는 AI 분야에서도 시도된다³⁾. P. Thagard는 이 접근법들이 대안적인 설명이론으로 간주되기보다는 설명의 다양한 측면에 대한 견해로 간주되어야 한다고 주장한다. 그러므로 필요한 것은 설명의 다양한 측면을 통합하는 설명이론을 제시하는 것이라고 그는 말한다. 그러한 이론은 인공지능에서 인지적 구조(cognitive architecture)라고 불리는 것과 관련하여 주어질 수 있다고 Thagard는 말한다. AI와 인지심리학의 연구자들은, 기본적인 표상과 과정의 이론적 상술을 구체화하고 이론적인 전산적 모델을 사용하는 인지적 구조를 세우려고 시도하였다.

2.1 연역적 설명 개념

철학에서 가장 영향력있는 설명모델은 Hempel에 의해 제시된 연역적-법칙적 설명 모델이다. Hempel에 의하면 과학적 설명은 피설명항을 설명항에서 논리적으로 이끌어내는 형식 즉 어떤 현상을 법칙 아래 포섭하는 형식을 취한다. 이 견해에 의하면 설명은 연역적 논증이다. 즉 전제는 일반법칙을 포함하고 피설명항은 전제로부터 도출되는 언명이다. 많은 철학자들이 Hempel의 법칙적-연역적 설명모델에 대해 비판하였다. 그들에 의하면 법칙적-연역적 설명모델이 설명을 위해 필요하지도 않으며 충분하지도 않다는 것이다. Hempel의 모델은 여러 가지 난점을 갖고 있지만 많은 경우 매

3)P. Thagard(1991), "Philosophical and Computational Models of Explanation", *Philosophical Studies* 64, pp. 87-104.

우 잘 들어 맞는다.

연역적 설명 개념은 AI에서도 유망하다. Machine Learning 분야의 많은 연구는 설명에 근거한 학습(explanation-based learning)에 관심을 가져왔다⁴⁾. Machine Learning의 대부분의 초기 연구는 인도되지 않은, 사례로부터의 학습--철학적인 용어로 그러한 학습은 귀납적 일반화에 대응된다--에 관심을 가졌다⁵⁾. 설명에 근거한 학습은, 한 체계가 어떤 영역에 대한 충분한 지식을 갖고 있어서 왜 그 사례가 연구 중인 개념의 사례인가를 자신에게 설명함으로써 단 하나의 사례로부터도 학습할 수 있을 때 발생한다. DeJong과 Mooney는 한 체계가, 컵들은 가볍고 손잡이를 가지고 있다는 단 하나의 사례로부터 어떻게 학습할 수 있는가를 기술한다⁶⁾. 훈련사례는 C인데 C는 빨갛고 가벼우며 손잡이를 갖고 있다. 배경지식은 다음과 같은 일반화를 포함하고 있다. 어떤 것이 뚜껑이 없는 용기이며 견고하고 들어올릴 수 있다면 그것은 컵이다. 어떤 것이 가볍고 손잡이를 갖고 있다면 그것은 들어올릴 수 있다. 문제는, 컵에 대한 수많은 사례를 모으지 않고도 훈련사례 C와 배경지식으로부터 컵에 대한 일반화를 이끌어내는 것이다. 왜 C가 컵이냐에 대한 설명의 일부분으로서 그 체계는 C가 들어올려진다는 것을 추리하기 위해 C가 가볍고 손잡이를 갖고 있다는 정보를 사용하며, C가 컵이라는 것을 추리하기 위해 C가 들어올려진다는 사실을 사용한다. 가볍고 손잡이를 갖고 있다는 것은 왜 C가 컵이냐에 대한 설명에 필요하므로 그 체계는 컵들은 가벼우며 손잡이를 갖고 있다라고 일반화한다⁷⁾.

AI 연구자들 사이에도 재미있는 견해 차이가 있다. Mitchell과 Keller와 Kedar-Cabelli에 의하면 설명은 분명히 연역적이다⁸⁾. C가 컵이라고 설명하는 것은 C가 컵이라고 증명하는 것이다. 많은 AI 연구자들은 연역적 체계에 기초하여 그런 일을 하였다. 연역이 한 체계의 중심적인 과정으로 간주된다면 설명도 본질적으로 연역적인 개념이다. Newell과 Simon에 의해 개척된 산출체계(production system)는 인지모델로서 그리고 전문가체계로서 널리 사용된다⁹⁾. 산출은 일반적인 IF-THEN 규칙인데 이것

4)P. Thagard(1991), p. 89.

5)Machine Learning의 철학적 견해에 대해서는 다음 논문 참조. P. Thagard(1990), "Philosophy and machine learning", *Canadian Journal of Philosophy* 200, pp. 261-276.

6)G. DeJong, and R. Mooney,(1986), "Explanation-based learning: An alternative view", *Machine Learning* 1, pp. 145-176.

7)P. Thagard(1991), p. 90.

8)R. Mitchell, and R. Keller, and S. Kedar-Cabelli(1986), "Explanation-based generalization: A unifying view", *Machine Learning* 1, pp. 47-80.

9)Newell, A., and Simon, H. A.(1972), *Human problem solving*, Englewood Cliffs, NJ:Prentice-Hall.

은, 일련의 수락된 명제들에 대해 IF 부분의 성공적인 짝이 THEN 부분의 추리에 이르게 되는 것을 말한다. 예를 들어 x 가 개라면 x 는 꼬리를 갖고 있다고 말하는 규칙은 Fido가 꼬리를 갖는다는 결론을 산출하는, Fido는 개라는 명제와 조화된다. 철학자들에게 그런 추리는 보편적 예화와 전진긍정의 논리적 규칙을 가볍게 조합시키는 것처럼 보일 것이다. 그러나 큰 체계에서 언제 규칙을 적용하느냐하는 문제를 제기할 때 수많은 재미있는 문제가 발생한다. 실질적인 전산적 체계에서 논리학자가 연역적 결론을 손쉽게 가정하는 것은 매우 좋지 않다. 추리를 조절하기 위해서는 효과적인 방법이 발견되어야 한다. 문제해결의 맥락에서는 규칙에 근거한 체계가 보통 논의된다. 일련의 선행조건이 주어지면 일련의 목적을 수행하기 위한 규칙이 적용되어야 한다. 그러나 그런 체계의 맥락에서는 설명의 자연스러운 번역이 있다. 만일 일련의 사실을 설명하기 위해 목적이 취해진다면 설명은, 알려진 정보를 사용하는 이 목적들과 짝을 이룬 규칙들을 도출하기 위해 규칙들을 사용하는 것으로 이루어진다. 만일 규칙들이 보편적 일반화라면 그 설명은 연역적이다. 그 외의 다른 곳에서 나는, 부분적으로 규칙에 근거한 문제해결이 되도록 취해진 설명의 전산적 모델을 기술한 바가 있다.

19 세기 말에 Peirce는 설명적 가설의 형성을 위해 想定論法(abduction)이라는 용어를 사용하였다. 최근에 AI 연구자들은, 설명적 가설을 산출하고 설명적 가설을 평가하는 체계를 개발하는 상정론적 추리에 점차적으로 많은 관심을 보여왔다. 대개 이 견해에는 본질적으로 연역적 체계로서의 설명개념이 함축되어 있다. O'Rourke와 그의 동료들은 설명을 하기 위해 프로그램 언어인 PROLOG 안의 연역을 사용한다.

2.2 통계적 설명개념

Hempel은 통계적 설명에서 설명항은 피설명항에 높은 정도의 확률을 제공해야 한다는 높은 확률의 요건을 제시하였다. Salmon은 Hempel이 통계적 설명에 대해 높은 확률과 예측가능성만을 요구함으로써 적절성을 무시하게 된다는 비판을 한다. Salmon은 Hempel이 절대적인 확률 개념에만 의존하고 있다고 비판한다. Salmon은 통계적 설명에서 높은 확률보다 적절성 개념이 중요하다고 말한다. 가령 구체적인 흡연가가 암에 걸린다는 것을 예측

할 수는 없지만 흡연과 암의 통계적인 적절성이 우리에게 설명을 제공한다. 최근에 Railton의 연구에 의하면 어떤 종류의 통계적 설명은 통계적 법칙으로부터의 연역적 도출에 의해 이해될 수 있다. 통계적 설명의 본질을 정확하게 특징짓기는 어렵지만 과학에서 그리고 일상생활에서 통계적 설명의 중요성을 부인하기는 어렵다. 필요한 것은 전형적인 통계적 설명을 적절하게 해명하고 또한 다른 종류의 비통계적 설명과

일관되는 이론이다. Salmon에 의하면 통계적 적절성은 세계의 인과적 패턴에 의해 현상을 설명하는 한 측면이다.

AI 연구자들도 통계적 설명에 관심을 가졌다¹⁰⁾. 의학적 진단을 위해 수많은 체계가 발전되었는데, 이 의학적 진단은 질병과 통계적으로 관련된 징후를 설명하기 위한 과정으로 이해될 수 있다. 스탠포드의 MYCIN 프로젝트는 확률적이지 않은 불확실성을 취급하기 위한 방법을 개발하였다. MYCIN 프로젝트는 여러 가지 방식으로 확률로부터 이탈하는 불확실성 요소를 사용하였다¹¹⁾. 다른 방법의 진단은 실행상 Bayesian적이었는데 Peng과 Reggia는 확률추정을 새롭게 하기 위해 연결주의적인 망(connectionist networks)을 사용한다¹²⁾.

과학과 일상생활에서 설명은 많은 경우 통계적이다. 어떤 설명이론이든 통계적 경향을 통합해야 한다. 그러나 통계적 경향과 연역적이며 다른 경향을 어떻게 결합하는지 나하는 문제는 불분명한 채 남아있다.

2.3 도식응용(schema-application)적 설명개념

연역적 설명개념과 통계적 설명개념보다 덜 형식적이지만 무시할 수 없는 설명개념이 있다. 철학과 AI의 많은 연구자들은 어떤 사실을 설명하는 것은 그 사실을 어떤 신념의 패턴이나 체계에 끼워 맞추는 것이라고 주장하였다. 최근의 강의에서 Smart는 체계에 끼워 맞추는 것으로서의 설명 개념이 어떻게 다양한 다른 개념을 포괄하는지를 기술하였다. 예를 들어, 연역적-법칙적 모델은 피설명항과 신념을 관련시키기 위해 연역적 관계를 사용한다. 또한 그는 유비(analogy)가 매우 다른 방식으로 신념을 한 체계에 어떻게 조화시키는가를 논의한다. Smart에 의해 논의되지는 않았지만, 조화시키는 다른 방법은 설명되어야 할 것을 기술한 것에 도식 즉 조직화된 패턴을 응용하는 것을 포함한다. 다양한 현상의 통합을 제공하는 설명의 중요성에 대해 Friedman이 강조한 것을 따라서 Kitcher는 이론적 설명이 문제해결 도식을 이용하여 현상의 통합을 어떻게 제공하는지를 서술한다. Kitcher는 다양한 범위의 현상에 이론을 응용하는데 Darwin과 Newton이 반복해서 사용한 설명의 패턴을 서술한다. 이 견해에 의하면 설명은 연역적이거나 통계적 이유 때문에 한 사건이 예측될 수 있음을 보여주는 문제가 아니다. 오히려 설명은 도식을 이용하여 다양한 현상을 통합하는 것을 포함한다¹³⁾.

10)P. Thagard(1991), p. 92.

11)B. Buchanan, and E. Shortliffe (eds.)(1984), *Rule-based expert systems*, Reading, Mass.: Addison Wesley.

12)Y. Peng, and J. Reggia(1989), "A connectionist model for diagnostic problem solving", *IEEE Transactions on Systems, Man, and Cybernetics* 19, pp. 285-298.

AI에서 도식응용적인 설명 개념은 Schank와 Abelson에 의하여 옹호되었다¹⁴⁾. 그들의 글은 전형적인 사건을 기술하는 복잡한 도식적인 표현이다. 사건을 설명하는 것은 적절한 글을 사건에 적용하는 문제이다. 예를 들어, *Restaurant*에 대한 글은 음식점 안에서 앉고, 주문하고, 먹고, 돈내는 등의 일련의 사건들에 대한 기술을 포함한다. 그러므로 어떤 사람이 떠나기 전에 왜 탁자 위에 돈을 놓고 갔는가에 대한 설명은 음식점이라는 글을 그 상황과 짝을 짓고 돈을 놓고 가는 사건에 의해 그 글의 부분을 예증함으로써 얻어질 것이다. 그 설명은 연역적이거나 통계적이지 않지만 전형적인 음식점 행위에 대한 도식을 이용하여 설명되는 사건을 짜맞춤으로서 얻어진다. Abelson과 Lalljee는 설명의 원형을 응용하는 것으로서 설명을 논의하는데¹⁵⁾, 설명의 원형은 Schank의 설명패턴의 일반화된 견해이다.

Churchland는 원형 응용으로서의 설명 개념을 옹호하지만, 그는 원형을 연결주의적인 체계 안의 단위로 분류된 활성화의 패턴으로 이해한다¹⁶⁾. 그에 의하면 설명적 이해는 지각과 유사하고, 잘 훈련된 망 안의 구체적인 원형 백타의 활성화로 구성되어 있다. 정상적으로 설명에 기여하는 높은 수준의 추리를 할 수 있는 연결주의자의 모델이 존재한다면 이 견해는 더욱 매력있을 것이다. 연결주의적인 체계는 어떤 종류의 지각 사례와 학습 사례에서는 매우 성공하였지만, 우리가 정상적으로 설명이라고 기술하는 경우에도 확장되지 않았다. 연결주의적인 원형의 견해는 도식응용의 사고를 제시하는 방법이지만 그것이 설명의 이해에 기여하는 것은 어떤 사람이 연결주의적인 설명프로그램을 세우기까지에는 기껏해야 은유적이다. 아마 언젠가 우리는 내가 서술한 모든 설명 개념을 포함하는 연결주의적 모델을 가질 것이다. 그러나 그런 모델은 현재의 모델의 힘에 실질적인 부가를 요구할 것이다.

분명히, 도식응용적인 설명모델의 일부분으로서 채워질 필요가 있는 많은 부분들이 있다. 우리는 도식이 무엇으로 구성되는가, 도식이 전체적인 개념 체계에서 어떻게 조직되는가, 특정한 사건에 대한 설명을 제공하는데 그 도식이 어떻게 응용되는가를 서술할 필요가 있다. 그러나 과학과 일상생활에서 설명이 도식응용이라는 견해의 매력을 알아 보기 위해 이런 문제들을 탐구할 필요는 없다. 우리는 설명이 연역이라기

13) P. Thagard(1991), p. 93.

14) R. Schank, and R. P. Abelson(1977), *Script, plans, goals, and understanding: An inquiry into human knowledge structures*, Hillsdale, NJ: Erlbaum.

15) R. Abelson, and M. Lalljee(1988), "Knowledge structure and causal explanation", in D. Hilton(ed.) *Contemporary Science and natural explanation*, New York: New York University Press, pp. 175-203.

16) P. Churchland(1989), "On the nature of explanation: A PDP approach", in P. Churchland, *A neurocomputational perspective*, Cambridge, Mass.: MIT Press, in press.

보다 도식응용이라는 견해를 옹호하지 않을 것이지만 대신 두 기능이 동일한 인지적 구조 내에서 어떻게 수행되는가를 알아볼 것이다¹⁷⁾.

2.4 유비적 설명개념

일상생활에서 우리는 가끔 사건을 설명하기 위해 유비(analogy)를 사용한다. 아이들의 병에 관한 문제를 부모에게 물어보라. 그들은 일반적인 정보보다 아픈 그들 자신의 아이들과의 경험에 기초하여 대답을 이끌어 낼 것이다. Dolnick는 팬더에 대한 최근의 논문에서 유비를 사용하였다. 팬더의 소화기관과 생식기관의 부적절성을 서술한 후, 팬더가 수백만 년 동안 어떻게 살아남았는가라는 문제를 제기하였다. 그는 다음과 같이 말했다. “TV와 진화에서 좋을 필요는 없다. 당신은 그저 경쟁자보다 더 좋으면 된다.” 과학철학에서 Campbell과 Hesse는 설명에서 유비의 사용을 옹호하였다¹⁸⁾. 이론을 논의하면서 Campbell은, 공기와 한 줌의 탄력입자 사이의 유비가 역동적인 공기 이론에 부수적이라는 제안을 거부한다. 그는 다음과 같이 주장한다. 유비는 이론을 정립하는데 도움이 되는 것이 아니다. 유비는 이론을 정립하는데 매우 본질적인 부분이다. 유비가 없으면 이론은 완전히 가치가 없으며 이름 값을 못할 것이다. 설명에서 유비는 생소한 현상을 더욱 친숙한 현상과 관련시키는데 본질적이다. Campbell은 설명이 친숙한 것으로의 환원이라는 것을 매우 강조하며, 설명에서 유비사용의 일반성을 과장한다. 그러나 분명히 그의 목표는 가끔 유비는 과학적 설명에서 매우 중요하다는 것을 아는 것이다. 예를 들어, Darwin은 자신의 설명에서 자연선택과 인위선택 사이의 유비를 종종 사용하였다. 따라서 충분한 설명이론은 분명히 어떤 설명의 유비적 요소를 포함해야 할 것이다¹⁹⁾.

AI에서 Schank와 그의 동료들은 설명에서 유비의 사용을 중요시하였다²⁰⁾. 당신이 한 사건을 설명할 때 당신은 유비적 사건을 설명하기 위해 사용될 수 있는 설명패턴으로서의 설명을 기억 속에 저장한다. Schank는 다른 사건에의 유비에 의해 레이스 말 Swale이 왜 죽었느냐에 대한 설명을 산출하는 프로그램을 서술한다. 유비적 설명은 더 일반적인 구조에 의해 제공된 도식에 기초한 설명에의 부가로서 간주된다. 유비적 설명은 도식응용을 야기할 수 있다. 왜냐하면 두 경우 사이의 유비는 두 경우를 일반화하는 도식을 산출하는데 사용될 수 있기 때문이다.

17)P. Thagard(1991), p. 94.

18)N. Campbell(1957), *Foundations of science*, New York: Dover. H. Hesse(1966), *Models and analogies in science*, Notre Dame, IN: Notre Dame University Press.

19)P. Thagard(1991), p. 95.

20)R. Schank(1986), *Explanation patterns*, Hillsdale, NJ: Erlbaum.

AI는 사례에 기초한 추리에 관해 일련의 연구를 하는데, 그 연구의 많은 부분은 유비적 설명보다 유비적 문제해결에 관련되어 있다. 유비에 관한 다른 AI의 연구는 두 유사물이 어떻게 상호간에 일치할 수 있는가와 관련되어 있다. 태양계와 같은 친숙한 현상에 의해 원자 구조와 같은 친숙한 현상을 설명하는데 두 유사물 사이의 체계적인 일치--예를 들어 원자 구조 중앙의 양자의 위치는 태양계 중앙의 태양의 위치와 일치--를 연구할 수 있는 것은 중요하다. 유비는 화학의 설명에서 가끔 사용된다. 그리고 나의 동료와 나는 화학적인 교육으로부터 끌어온 수많은 사례에 관한 유비적 대응 프로그램을 수행하였다.

다른 설명적 기술의 보조물로서 유비적 설명에 관한 AI의 견해는 이론적 설명이 항상 유비적이라는 Campbell의 일반화보다 더 합리적이다. 그러나 유비와 다른 경향의 설명을 통합하는 체계를 세우는 일이 남아 있다²¹⁾.

3. 세 가지 설명이론

3.1 Cummins의 설명이론

R. Cummins는 <<심리학적 설명의 본질>>²²⁾에서 인지심리학의 설명모델을 발전시키고 인지과학을 둘러싼 개념적인 문제들을 해명하는데 그 모델을 적용하였다. Cummins의 심리학적 설명의 기초는 "인과적 포섭"의 전략과 "분석"의 전략을 구분하는 일이다. 인과적 포섭의 전략은 Hempel의 포괄법칙형 설명이다²³⁾.

심리학적 능력²⁴⁾을 설명하면서 우리가 사용할 수 있는 전략은 두 가지이다. 그 두 가지는 인과적 포섭의 전략과 분석의 전략이다. Cummins에 의하면 심리학적 현상들은 인과법칙들 아래 포섭됨으로써 설명되는 것이 아니라 분석에 의해 설명된다²⁵⁾.

21) P. Thagard(1991), p. 96.

22) R. Cummins(1983), *The Nature of Psychological Explanation*, Cambridge : The MIT Press. A Bradford Book.

23) Cummins는 인과적 포섭에서 "포섭"이 단순히 법칙 하에 포섭만을 의미하는 것으로 이해된다면 오해의 소지가 있다는 점을 지적한다. 그의 주장은 다음과 같다. 우리는 성질 Φ 와 성향 D를 상상할 수 있으며 이 경우 'D를 가진 모든 사물만이 Φ 를 가진다'는 참이며 준법칙적이다. 그러나 Φ 의 존재가 D와 관련한 규칙성을 설명하지는 않는다. R. Cummins(1977), "Programs in the Explanation of Behaviour," *Philosophy of Science*, 44. p. 272.

24) 인간의 심리학적 능력은 일종의 성향이다. 어떤 것에 성향을 부여하는 것은 그 성향이 어떤 조건 하에서 유지된다고 말하는 것이다.

25) R. Cummins(1983), p. 1.

심리학적 설명을 포괄법칙적 틀에 무리하게 맞추려는 것은 그 설명을 사소하게 혹은 무의미하게 만드는 결과가 된다고 Cummins는 주장한다.

Cummins는 위의 책 첫머리에서 다음과 같이 말한다.

“이 연구의 중심되는 주장은 다음과 같다. 즉 심리적 현상들은 전형적으로 인과 법칙 아래 포섭됨으로써 설명되지 않고, 분석에 의해 설명되는 능력의 표현으로 취급됨으로써 설명된다. 따라서 두 설명적 전략의 구별--포섭과 분석--은 이하에서 중요하다.”²⁶⁾

그리고 1 장 끝에서 다음과 같이 말한다.

“...나의 주제는 심리학적 설명이다. 내가 생각하기에 대부분의 심리학적 설명은 인과적 포섭에 의해 해석되었을 때 무의미하고 분석에 의해 해석되었을 때 중요한 의미를 지니기 때문에 이 장이 포함되었다. 그러므로 분석적 전략에 대한 이해는 심리학적 설명에 대한 이해에 본질적이다. 그러나 동시에 중요한 것은 분석이 중요하고도 일반적으로 적용가능한 설명적 전략이며, 이 전략은 심리학과 생명과학 이외에서도 공통적이며 기본적으로 중요한 전략이라는 것을 깨닫는 것이다. 이 점을 아는 것은 심리학적 설명이 나머지 과학에서 이례적이지--특별한 경우--않고 연 계되어 있다는 점을 아는 것이다. 심리학적 설명을 포섭 모형에 무리하게 맞추는 것은 그 설명을 사소하게 혹은 무의미하게 만드는 댓가를 치른 후에야 나머지 과학과 연계되도록 만드는 것이다.”²⁷⁾

인과적 포섭의 전략은 그가 “변화이론”(transition theory)이라고 부르는 것에 적절하고 분석이론은 그가 “속성이론”(property theory)이라고 부르는 것에 적절하다. Cummins는 과학이론을 두 가지로 구분한다²⁸⁾. 변화이론은 어떤 체계의 상태변화를 이전 원인의 결과로서 설명한다. 물리학이 변화이론을 대표한다. 변화이론은 무엇이 어떤 조건 하에서 발생하는가를 강조한다. 속성이론은 어떤 체계의 속성을 설명한다. 생물학이나 생리학이 속성이론을 대표한다. 따라서 Hempel의 포괄법칙형 설명이론은 물리적 현상을 설명할 때에는 적절하지만 의식적이고 지향적인 인간의 행위를 설명하는 데에는 부적절하다는 결론이 된다.

26)R. Cummins(1983), p. 1.

27)R. Cummins(1983), p. 27.

28)변화이론은 체계 S가 왜 S₁의 상태에서 S₂의 상태로 변화하는가 하는 물음을 제기하고 속성이론은 체계 S가 속성 P를 갖는 이유는 무엇인가 하는 물음을 제기한다.

포섭의 전략(Subsumption Strategy)은 화학과 물리학에서 흔히 볼 수 있다. 예를 들어 보자. 떠오름(elevancy)이라고 부르는 성향을 생각해 보자. 이것은 물체가 물 위에서 자체적으로 떠오르는 성향이다. 떠오름은 부력과 다르다. 뚝단배는 부력을 갖고 있지만 떠오르지는 않는다. 떠오름을 설명하기 위해서는 물에 잠긴 떠오르는 물체를 놓아주면 그 물체가 올라가는 이유를 설명해야 한다. 모든 경우에, 떠오르는 물체의 질량과 그 물체의 비침투적인 부피의 비는 물의 비중보다 작다. 아르키메데스 원리에 의하면 물은 대체된 물의 무게와 동일한 정도의 상승하는 힘을 잠겨있는 물체에 가한다. 떠오르는 물체의 경우에 이 힘은 물체의 무게보다 f 정도를 초과한다. 물체를 놓아주면 물체에 가해지는 순수 힘은 표면의 방향으로 0에서 f 정도의 순수 힘으로 변화한다. 따라서 그 물체는 떠오른다. 여기에서 우리는 놓아줌과 떠오름의 관계를 순수 힘의 변화와 운동의 변화의 관계를 표현하는 법칙 하에 포섭한다. 이러한 작업은 떠오르는 물체의 특성을 언급함으로써 가능해진다.

Cummins에 의하면 포섭의 전략은 심리학적 능력을 설명하는데에는 별 소용이 없다²⁹⁾. 심리학적인 작업의 대부분은 유기체의 복잡한 행위가 어떻게 얻어지고 어떻게 수행되는가를 설명하는 일이다³⁰⁾. 두 가지 목적은 문제의 능력을 분석함으로써 진행된다. 유기체의 능력은 그 유기체를 수많은 체계--이 체계 각각은 특징적인 능력을 갖고 있다--로 분석함으로써 설명된다. 이 능력들은 다시 구성 기관과 구조의 능력으로 분석된다.

분석의 전략(Analytical Strategy)은 a 의 성향 d 를 수많은 다른 성향 d_1, d_2, \dots, d_n 으로 분석함으로써 진행한다. 즉, 어떤 성향을 비교적 덜 문제있는 수많은 성향으로 분석함으로써 진행한다. 분석하는 성향들의 조직적인 표현은 분석된 성향의 표현에 해당한다. 전자공학에서 도식적인 도해는 물리학적 맥락에서 이런 종류의 분석의 친숙한 예를 제공한다. 각각의 기호는 어떤 능력을 가진 물리적 대상을 표상하기 때문에 복잡한 장치의 도식적인 도해는 그 장치의 전자적인 능력을 구성요소의 능력으로의 분석을 구성한다. 그런 분석에 의해 우리는 그 장치가 어떻게 분석된 능력을 수행하는가를 설명할 수 있다. 왜냐하면 분석된 능력의 수행을 분석하는 능력의 프로그램적인 수행으로 간주할 수 있기 때문이다.

예를 들어 우리의 신경체계가 어떻게 의식과 지향성 같은 속성을 소유할 수 있는가를 설명하기 위해서는 속성이론이 필요하다. 먼저 의식적이거나 지향성을 소유한다는 것이 무엇인가를 분석하고 우리의 신경체계가 그런 분석을 어떻게 실례화하는가를 보여주어야 한다. Cummins의 관심은 욕구와 믿음 같은 명제태도를 설명하는데 중요한

29)R. Cummins(1977), p. 272.

30)R. Cummins(1975), "Functional Analysis," *Journal of Philosophy* 72, p. 761.

지향적 현상을 적절하게 설명하는데 있다.

Cummins에 의하면 인지심리학이 주로 제기하는 물음은 “어떻게 S는 심리학적 능력 C를 나타낼 수 있는가?”이다. 어떻게 S가 C를 나타낼 수 있는가를 설명하기 위해서는 능력 C를 갖고 있음이 무엇인가를 이해할 필요가 있다. 그러므로 분석의 전략을 사용하는 속성이론이 요구된다. 따라서 S가 C를 갖고 있음을 설명하기 위해서는 기능적이며 해석적인 C의 분석을 제공해야 하고 특정한 경우에 C의 나타냄을 이끌어 내는 조건과 그 나타냄의 성격간의 관계를 기초지우는 이유를 제시해야 한다.

가령 영어의 문법적인 문장을 인지하는 능력은 성향적 능력이다. 어떤 사람이 그러한 복합적인 성향을 갖고 있음을 설명하기 위해서는 그 능력에 대한 기능적 분석을 제공할 필요가 있다. 문법적 문장은 상호관련된 하위능력의 집합으로 분석된다. 각각의 하위능력에 대해서도 기능적 분석이 요구된다. 문법적 능력에 대한 기능적 분석이 해석적이라고 말하는 것은 그 분석이 의미론적 해석에 의해 적절한 입력과 출력을 상술함을 말한다. 의미론적으로 해석된 대상 즉, 기호를 처리하는 모든 장치를 Cummins는 정보처리체계(information-processing system: IPS)라고 부른다. 우리는 문제의 입력을 상술하는데 도움되는 추론규칙에 따라, 주어진 입력으로부터 도출된 것을 출력으로 간주할 수 있다. 이렇게 심리학적 능력에서 추론개념이 중요하기 때문에 Cummins는 심리학적 능력을 추론적으로 특징화된 능력(inferentially characterized capacities: ICC)으로 간주한다³¹⁾. 심리학적 능력은 추론적으로 특징화할 수 있는 성향적 성질이다.

Cummins에 의하면 성향에 대한 설명이 체계적인 기능적 분석을 포함할 때마다, 그 분석은 프로그램으로 표현될 수 있다. 앞의 예에서 처럼 언어적 능력 같은 심리학적 능력을 분석할 때 우리가 얻고자 하는 것은 자의적인 입력에 대한 옳은 결정을 보장하는 프로그램이다. 한 장치가 프로그램의 지시를 수행한다면 그 장치는 특정한 입력에 대한 옳은 출력을 내보낸다.

31) Cummins의 설명을 들어 보자. 어떤 능력(capacity)은 입력-출력 조건으로 규정된다. 그리고 한 능력을 인지적으로 만드는 것은 출력이 인지라는 것이다. 따라서 특정한 인지적 능력을 규정하는 방법은 입력과 출력의 인식적 관계를 규정하는 것이다. 출력을 인지적으로 만드는 것은 출력이 입력과 적절한 관계에 있는 것이다. 인식론적으로 적절한 관계는 진리담지자들 간에만 성립하기 때문에 입력과 출력은 진리담지자로 해석되어야 한다. 한 명제가 다른 명제들과 적절한 관계에 있다고 말하는 것은 한 명제가 다른 명제로부터 타당하게 추론될 수 있다고 말하는 것이다. 따라서 한 능력이 인지적이기 위해서는 입력과 출력이 명제 토근으로 해석되어야 하며, 출력이 입력으로부터 추론될 수 있어야 한다. 결론적으로 인지적 능력은 추론적으로 특징화될 수 있는 능력이다. R. Cummins(1983), p. 53.

3.2 Haugeland의 설명이론³²⁾

철학과 심리학에서 인지론의 입장은 지적인 행위는 내적인 인지적 과정에 호소하여 설명될 수 있다는 입장이다³³⁾. Haugeland에 의하면 인지론은 합리론적인 전통의 상속자이다. 합리론자들은 인지를 인간의 본질로 간주한다. 지적이라는 것은 유사-언어적인 표상을 처리할 수 있음을 말한다³⁴⁾. 일반적으로 인지론은 다음과 같은 두 가지 가정을 한다. 첫째, 사물을 지적으로 다루는 우리의 능력은 사물을 합리적으로 생각하는 우리의 능력 때문이다. 둘째, 사물을 합리적으로 생각하는 우리의 능력은 내적인 자동적 기호처리 능력과 같다³⁵⁾.

인지론은 기본적으로 통속심리학적인 설명방식의 틀을 받아들이는 입장이라고 할 수 있다. 인지론이나 통속심리학이나 모두 행위의 규칙성과 행위에 관한 일반화를 얻기 위해서 내적인 믿음이나 바램의 내용을 나타낼 필요가 있다고 보기 때문이다. 인지론은 통속심리학의 경험적인 일반화가 체계적으로 될 수 있고 엄밀해 질 수 있다는 희망에서 출발하고 있다. 인지론이 통속심리학과는 달리 엄밀한 경험적인 설명을 줄 수 있었던 것은 컴퓨터의 작동에 관한 이론을 참조한 결과이다.

일반적으로 과학은 세계에서 발생하는 것들을 이해하기 위한 노력이다. 그러므로 이해와 더불어 설명은 중요한 의의를 가진다. 특히 과학적 설명은 상식적 설명보다 더 정확하고, 더 자세하며, 더 일반적이며, 설명들이 상호 통합되어 있다는 점에서 다르다. 과학적 설명은 다양하고 구체적인 상황에서 드러나는 일련의 특징을 상술함으로써 이루어지는데, 그 특징들이 드러나는 상황에서 항상 얻어지는 규칙성이나 관계를 가정하거나 앞으로써 성립한다. 전제된 규칙성이 법칙에 의해 정식화되는 설명을 철학자들은 연역적-법칙적 설명(deductive-nomological explanation)이라고 부른다. 그러나 Haugeland는 다른 형태의 설명이 두 가지 더 있다고 주장한다.

그는 그 설명을 위해 섬유광학(fiber optics)의 예를 든다³⁶⁾. 섬유광학의 한 묶음은 어떤 상(image)이든 가질 수 있는데 그 상은 한쪽 끝에서 나와 다른 쪽 끝으로 전달될

32) Margolis는 인지론을 가장 간결하고 잘 정리하여 설명한 사람은 Haugeland이라고 말하고 있다. J. Margolis(1984), *Philosophy of Psychology*, Prentice-Hall, Inc., Englewood Cliffs. p. 70.

33) J. Haugeland(1978), "The Nature and Plausability of Cognitivism", *The Behavioral and Brain Sciences* 1, in *Mind Design*. Haugeland, J.(ed.)(1981), Cambridge: The Mit Press. A Bradford Book, p. 243.

34) J. Haugeland(1981), p. 276.

35) J. Haugeland.(1985), *Artificial Intelligence, The Very Idea*, Cambridge : MIT Press/Bradford Books, p. 113.

36) J. Haugeland(1981), p. 246.

수 있다. 이 예가 이해되기 위해서는 다음과 같은 네 가지가 전제되어야 한다³⁷⁾. 첫째, 그 묶음들은 길고 가느다란 많은 섬유로 이루어져 있는데, 이 섬유들은 하나씩 하나씩 밀접하게 담겨 있으며 그 묶음의 전 길이를 따라 다른 섬유와 관련하여 같은 방향으로 정렬되어 있다. 둘째, 각 섬유는 빛을 그대로 전달하는 관이다. 섬유의 한쪽 끝으로 들어온 빛은 모두 같은 섬유의 다른 쪽 끝으로 나간다. 셋째, 투사된 상은 정밀하게 포장된 점들로 이루어진 빛의 연쇄로 생각할 수 있으며, 밝기와 색깔은 다르다. 넷째, 각 섬유의 끝은 점과 같으므로 그 묶음의 한 쪽 끝에 상을 보내면 다른 쪽 끝은 동일한 지점에서 동일한 색깔과 밝기의 점으로 빛나게 된다. 이것은 다음과 같은 세 가지 이유 때문에 연역적-법칙적 설명이 아니라고 Haugeland는 말한다. 첫째, 여기에서 설명되는 것은 어떤 대상의 성향이나 능력이다. 둘째, 그 설명은 두 가지 기본적인 전제에 의존한다. 하나는 문제의 대상이 어떤 형식이나 구조를 갖고 있다는 것이며, 다른 하나는 그런 형식이나 구조를 가진 것은 모두 어떤 성향이나 능력을 가졌다는 것이다. 셋째, 전제된 방식으로 구조화된 대상은 모두 그 능력이 설명되도록 할 것이다. Haugeland는 이러한 설명을 형태학적 설명(morphological explanation)이라고 부른다. 이 설명의 특징은, 한 능력이 특정한 구조에 호소하고 그렇게 구조화된 것의 특정한 능력에 호소함으로써 설명된다는 것이다. 과학에서 형태학적 설명은 가끔 “모델”이라고 불려지지만 Haugeland는 그 용어가 너무 넓은 의미를 지닌다고 거부한다.

다른 예로 자동차 엔진이 어떻게 작용하는가를 설명하는 경우를 생각해보자. 형태학적 설명을 시도할 경우 그 설명은 특정한 체계와 그렇게 체계화된 것의 특정한 성향이나 능력에 호소한다. 그러나 그 외에 그 설명을 더 잘하기 위해서는 서로 의존적인 상호작용의 복잡하게 조직화된 패턴을 자세하게 언급할 필요가 있다. 엔진의 다양한 부분들은, 다른 부분들이 단독으로 할 수 있었던 일과는 매우 다른 결과를 얻기 위해서 많은 다양한 일들을 수행한다. 상호작용을 자세하게 언급하는 것은 설명에 중요하고 상호작용은 별개의 상호작용자(interactor)를 요구하기 때문에 체계적인 설명을 얻는 능력을 가진 대상은 별개의 부분들로 구성되어 있다. Haugeland는 이러한 설명을 체계적인 설명(systematic explanation)이라고 부른다. 여기에서 체계³⁸⁾는 체계적으로 설명되는 능력을 가진 대상이며, 기능적 구성요소(functional components)는 설명에서 그 상호작용이 인용되는 별개의 부분이라고 하자. 한 체계에서 상술된 구조는 본질적으로 기능적인 구성요소의 배열이기 때문에 그 구성요소들은 상술된대로 상호작용할 것이다. 형태학적 설명과 체계적인 설명의 차이는 체계화된 협동적인 상호작용의 부가

37) J. Haugeland(1981), p. 246.

38) Haugeland는 체계를 회사에 비유한다. 그 회사의 구성원들은 서로 다른 역할들을 수행하는데 대부분은 조직화된 상호작용을 통해 간접적으로나마 최종 결과를 산출하는데 기여한다. J. Haugeland(1981), p. 248.

적인 요소이다. Haugeland는 체계적인 설명만 인지심리학에 직접적으로 관계된다고 주장한다³⁹⁾. 인지론에서는 인간의 인지적 행동들을 구조지어진 내적 체계로부터 나타난 것으로 본다. 다시 말하면 인지론자의 설명은 체계적이다.

Haugeland에 의하면 모든 설명의 공통적인 측면은 무엇인가를 전제한다는 것이다. 구체적으로 말하면 설명은 일반적인 규칙성을 전제한다. 그러나 그 규칙성은 자체적으로 설명되지 않기 때문에 더 기본적인 것에 호소하여 설명될 수 있다. Haugeland은 그러한 설명을 환원(reduction)이라고 말한다⁴⁰⁾. 형태학적 설명과 체계적인 설명이 전제하는 규칙성은 주로 구체적인 방식으로 구조화된 것의 성향이나 능력이다. 그러므로 형태학적 환원과 체계적인 환원은 그 능력에 대한 설명이다. 체계를 설명하면서 전제된 모든 규칙성은 인접한 구성요소와 상호작용하는 개별적인 구성요소의 능력이다. 복잡하고 상호 의존적인 조직화는 체계의 핵심이기 때문에 개별적인 구성요소에게 요구되는 능력 자체는 매우 세련되고 전문화되어 있다. (체계적인 설명에서) 그 능력은 다양하고 구별된 기능적 구성요소의 조직화된 상호작용으로부터 결과하는 것으로 이해할 수 있다.

Haugeland에 의하면 심리학은 정량적인 법칙에 관심 갖지 않으며, 심리학적 이론은 물리학적 이론과 같지 않다. 따라서 심리학적 설명은 체계적이어야 하며 법칙적·연역적이어서는 안된다.

3.3 Pylyshyn의 설명이론

① Pylyshyn의 설명개념

Pylyshyn에 의하면 과학에서 설명은 일반화를 포착하여 이해를 구하는 과정이다. 이러한 일반화가 존재한다면 일반화를 포착하는 것은 과학의 중요한 목적이다⁴¹⁾. 그는 설명이 특정한 어휘(vocabulary)와 관련되어 있다고 주장한다⁴²⁾. 인지심리학에서 행위의 규칙성과 일반성은 그 행위와 선행조건이 분류하는 용어에 의해 기술될 경우에만 포착된다⁴³⁾. 설명은 설명되는 현상을 기술하는 방식과 밀접하게 관련되어 있다

39) J. Haugeland(1981), p. 249.

40) J. Haugeland(1981), p. 249.

41) 행위의 규칙성을 설명하려면 우리는 거의 알지 못하는 의식상태와 똑같이 작용하는 과정과 정신상태를 가정할 필요가 있다. Z. W. Pylyshyn(1987), "What's in A Mind?", *Synthese* 70, p. 99.

42) Z. W. Pylyshyn(1985), *Computation and Cognition, Toward a Foundation for Cognitive Science*, A Bradford Book. The MIT Press, Cambridge, p. 2.

43) Pylyshyn에 의하면 심리학에서 설명의 문제는 행위 분류방식을 선택하는 문제이

고 Pylyshyn은 말한다. 예를 들어보자⁴⁴⁾.

“당신이 낫선 거리를 걷고 있었는데 다음과 같은 일련의 사건을 목격하였다고 가정하자. 행인이 인도를 따라 걷고 있다. 갑자기 행인이 방향을 바꿔 길을 가로질러 건너기 시작한다. 동시에 차는 행인을 향해 길 아래로 빠르게 질주하고 있다. 차의 운전자가 브레이크를 밟는다. 차가 미끄러지며 길 옆으로 벗어나서 전주와 부딪친다. 행인은 머뭇거리며 건너가서는 차 안의 운전자 쪽을 바라본다. 그는 길 모퉁이의 전화 박스로 뛰어가서 9와 1 번을 돌린다.”

Pylyshyn은 우리가 행위를 설명할 때 기술의 진실성에 관심 갖기 보다는 일반화를 포착하려는 다양한 기술의 능력과 다양한 어휘 속에 표현된 설명의 적절성에 관심을 가져야 한다고 말한다. 앞의 예에서 설명이 적절하려면 무엇보다 다음과 같이 언급해야 한다⁴⁵⁾.

“그 행인은 충돌을 지각하였으며, 그것을 사건으로 분류하여 인식하였으며, 상해가 있을 것이라고 추리하였으며, 어떤 사람이 실제로 상해를 입었다고 결론을 내렸으며, 자신이 목격한 것으로부터 그 결론이 맞다고 연역하였으며, 상해입은 사람을 적절하게 처리하기 위해 자신이 갖고 있는 지식에 기초하여 도움을 구하기로 결정하였으며, 옆에 전화박스를 발견하였으며, 긴급구조 번호를 상기하고, 도움을 구하려는 의도로 번호를 눌렀다.”

우리가 아무리 과학적이고 기술적인 설명을 하더라도 그 행인이 그 사건을 긴급상황으로 지각하지 않고 무엇을 해야 하는가를 알지 못하고 전화번호를 기억하지 못한다면 전체 상황은 매우 달라지게 된다. 앞의 예에서 그가 보여주고자 하는 것은 어떤 현상은 어휘로 표현된 설명을 요구한다는 것이다. Pylyshyn은 이러한 어휘들이 그 사건의 이유와 내용과 방법을 설명하는데 실제로 필요한 것이라고 주장한다. 이것은 인지적 용어들이며 많은 철학자들은 그것을 지향적 용어(intentional term)라고 부른다. 이 지향적 용어가 심리학적 설명에서 중요한 역할을 한다는 것이 Pylyshyn의 지적이다. 이러한 용어들이 필요한 이유는 동일하게 규칙적이며 조직적인 다른 행위절차도, 그 사람에게 이용가능한 정보를 변경시킴으로써 일어날 수 있기 때문이라고 Pylyshyn은 말한다⁴⁶⁾. 반사실적 경우들--일어날 수 있는 것들--에 대한 관심이 설명의 관건이다. 더구나 반사실적 사태를 야기할 수 있는 방식은 인지적 체계의 한 특징을 구성한다.

행위 설명이 인지적 용어를 포함해야 한다고 주장하는 많은 이유가 있다. 그 중요한 이유 중의 하나는 인지적 용어를 사용하여 포착될 수 있는 규칙성과 일반화가 행

다.

44) Z. W. Pylyshyn(1985), p. 3.

45) Z. W. Pylyshyn(1985), p. 5.

46) Z. W. Pylyshyn(1985), p. 6.

위적 또는 물리적 용어를 사용한 기술로는 포착될 수 없다는 것이다⁴⁷⁾. 그리고 인지적 용어로 설명하는 것이 좀더 지적으로 만족스러운 설명을 제공하는 것만은 아니다. 인지적 용어에 의한 설명은 적절한 일반화를 포착하기 때문에 더 예측적이다.

앞의 예를 상기해 보자. 행인이 전화박스로 가서 9와 1 번을 눌렀다. 그렇다면 그는 다음에 무엇을 행하는가? 이에 대한 대답은 그 상황이 특정한 용어를 사용하여 기술될 경우에만 가능하다. 그 행인은 긴급번호가 911이라는 것을 알고 있고, 그 상황을 긴급상황으로 인지하였기 때문에 그는 분명히 1번을 눌릴 것이다. 그 상황을 이렇게 기술하는 방식은 그러한 예측을 가능케 한다. 체계적 설명이, 그 행인이 다음에 행하는 것은 1번을 누리는 것이라는 예측과 자극조건을 연결시키는 것이라면, 그 설명은 다음과 같이 언급해야 할 것이다. 즉 행인은 그 상황을 사건으로 해석하였으며, 긴급전화번호를 알거나 기억하며, 행인의 행위는 '구조요청전화'라는 범주의 사례가 언급되어야 한다⁴⁸⁾.

이것이 옳다는 것은 다른 가능성, 즉 반사실적 가능성을 고려함으로써 알 수 있다. 그 예측이 유효하지 않지만 물리적으로 동일한 상황이 있다. 예를 들어, 그 거리에서 TV 쇼의 리허설이 있다는 것을 그 사람에게 미리 알리는 경우이다. 그리고 그 예측이 유효하고 물리적으로 매우 다른 상황이 있다. 예를 들어, 그 사람이 상해입은 운전자를 보는 대신 도움을 요청하는 운전자의 외침을 듣거나 누군가가 다쳤다는 것을 통행인에게 듣는 경우이다. 중요한 것은 동일한 일반화가 실현되지만 물리적으로 매우 다른 방법이 많다는 것이다. 그 일반화를 인지하지 않는다면 우리는 중요한 규칙성을 잃어버린다.

행위를 설명할 때 인지적 용어를 사용해야 한다고 주장하는 또다른 이유는 인간의 행위가 매우 유연하다(plastic)는 점이다. 이 특색은 행위가 자극-독립적이라는 점에서 매우 분명하게 드러난다⁴⁹⁾. 우리의 행위는 물리적 또는 생리학적 조건의 변화에 전적으로 의존하는 것이 아니다. 대부분의 비행동주의자에 의하면, 인간의 행위가 자극-독립적인 이유는 다음과 같다. 인간의 행위는 그 순간에 믿고 있는 것파 그 순간에 그 상황을 어떻게 지각하느냐 하는 것파 자신들이 행위의 결과라고 생각하는 것에 의존한다. 행위가 자극-독립적이라는 것은 행위가 자의적이라는 것을 의미하는 것은 아니고, 행위가 다양한 원칙에 의해 지배된다는 것을 의미한다⁵⁰⁾.

47) J. A. Fodor, H. Putnum, D. Davidson, D. C. Dennett 등이 이러한 종류의 논의를 하였다. 결국 Pylyshyn이 주장하는 것은 인지과학에 속하는 일반화와 다른 과학에 속하는 일반화 사이에 중요한 차이가 있다는 것이다.

48) Z. W. Pylyshyn(1985), p. 7.

49) Z. W. Pylyshyn(1987), p. 108

50) Z. W. Pylyshyn(1985), p. 12.

유기체는 환경에 대해 선택적으로 반응할 수 있다. 유기체는 환경으로부터의 자극에 대해 지각을 이용하여 반응을 한다. 행위의 심리학적 규칙성을 결정하는 것은 지각이다. 지각 안의 모든 것이 추리를 포함하지는 않는다. 그러나 자극을 범주의 한 사례로 인식하거나 확인하는 것은 추리를 포함한다. 우리는 빨간불을 교통신호로 보았을 때 멈춘다. 우리가 빨간불을 교통신호로 보는 것은 우리 문화의 관습에 대한 지식에 의존한다. 이것은 자극의 물리적 속성이 행위를 결정하지 않는다는 것을 말해준다.

우리가 행위할 때 갖고 있는 의도와 그 행위의 결과 사이에 분명한 관련이 있음에 분명한데, 그 관련은 사회적이고 자연적인 환경에서 얻어진다. 인간의 행위를 결정하는 요인은 물리학에 의해 기술되었다기보다는 행위주체에 의해 해석된 (또는 인식된) 환경이나 선행사건이다. 그리고 행위적 규칙성에 포함되는 것은 객관적인 자연과학에 의해 기술된 행위라기보다는 의도를 가지고 수행된 행위이다⁵¹⁾.

Pylyshyn은 행위의 규칙성을 설명하는 방식을 두 가지로 나눈다⁵²⁾. 첫째 종류의 설명(첫번째 형태의 설명)은 인지적 또는 지식의존적 설명이라고 불린다. 이 방식은 인간의 행위를 인간이 신념과 목적을 소유한 결과로 설명한다. 형태 1의 설명방식의 간단한 예를 들어보자. 유기체가 목적 G를 갖고 있고 A를 행하는 것이 G를 얻는데 도움이 된다고 믿기 때문에 유기체가 A를 행하는 경우이다. 이 설명방식은 심리학의 분야에서 흔히 볼 수 있으며 이 때 의사결정(decision-making)이 중요한 과정으로 간주된다고 Pylyshyn은 말한다. 둘째 종류의 설명(두번째 형태의 설명)⁵³⁾은 자연주의적 또는 기계적-인과적 설명이라고 불린다. 이 방식은 인간의 행위를 법칙적으로 작용하는 내적인 성질을 인간이 소유한 결과로 설명한다. 형태 2의 설명은 문제의 성질이나 규칙성을 생물학적 원리로부터 도출하지만, 우리는 그 규칙성을 기능적인 수준에서 특징짓는 것으로 만족한다.

두 형태의 설명은 행위의 규칙성을 설명할 때 차이가 난다. Pylyshyn은 인간의 의도적 행위가 목적과 신념과 추리에 의해 설명되어야 하는 이유 중의 하나를 다음과 같이 말한다. 의도적 행위는 행위 주체에게 어떤 정보를 제공함으로써 일관되고 명확한 방식으로 변경될 수 있다. 그 변경은 정보가 전달되는 물리적 형식과는 무관하다. 사람이 어떤 자극에 대해 반응하는 방식이 무엇인가를 듣거나 봄으로서 근본적으로 변화될 수 있다면, 자극과 반응의 과정을 야기하는 해석과 추리와 의사결정의 단계가 분명히 있어야 할 것이다. 그리고 그 단계는 주체가 믿고 있는 것에 기초한 결정을 포

51) Z. W. Pylyshyn(1985), p. 9.

52) Z. W. Pylyshyn(1981), "Psychological explanations and knowledge-dependent processes", *Cognition* 10, pp. 267--274.

53) Pylyshyn은 불필요한 오해를 피하기 위해 형태 2라고 명명한다. 형태 2의 설명 방식은 Hempel의 설명방식을 말한다.

함해야 하므로 형태 1의 설명을 요구한다.

형태 1과 형태 2를 구별하는 것은 인지적 현상을 이해하는데 중요하다고 Pylyshyn은 말한다⁵⁴⁾. 규칙과 표상으로부터 발생하는 전체 과정의 성질과 기초적인 매개물의 생물학적인 성질로부터 발생하는 전체 과정의 성질을 구별하지 못하면, 다음의 두 가지 현상을 결과한다. 첫째, 우리는 관찰된 입력과 출력의 행위를 흉내내는 *ad hoc* 계산적 가설을 구성하게 된다. 둘째, 우리는 행위의 규칙성에 대한 발견을 정신적 메카니즘의 성질에 대한 발견으로 취급하게 된다. 그러나 행위의 규칙성에 대한 발견은 실제 세계의 성질에 대한 주체의 암묵적인 지식에 대한 발견으로 취급되어야 한다.

② 표상의 설명적 기능

어떤 체계 안의 행위의 규칙성을 설명하는 방식은 두 가지이다. 하나는 그 체계에 본질적인 성질이나 메카니즘에 호소하거나 그 체계의 기능적 능력에 호소하는 것이다. 다른 하나는 그 체계가 어떤 규칙에 의해 조작되는 표상을 갖고 있다고 가정하는 것이다⁵⁵⁾. Pylyshyn은 우리가 한 체계 안의 행위의 규칙성⁵⁶⁾을 설명하려면 표상의 내용에 호소해야 한다고 말한다⁵⁷⁾. 체계의 행위의 경험적 규칙성을 관찰한다면 우리는 그 체계가 왜 그렇게 행위하는가에 대한 설명을 세 가지로 제시할 수 있다. 첫째, 지향적 또는 표상적 설명이다. 체계는 믿음과 목적 때문에 일정한 방식으로 행위한다. 둘째, 기능적 구조 설명이다. 체계는 정보를 코드화하고 그 코드를 처리하도록 구조지어진 방식 때문에 그렇게 행위한다. 셋째, 아날로그식 설명이다. 체계는 일반법칙을 예증하기 때문에 그런 방식으로 행위한다. 우리가 행위설명을 위해 표상을 가정해야 하는 이유는 두 가지이다⁵⁸⁾. 첫째, 의미있는 행위는 유연하고 자극에 대해 독립적이다. 둘째, 어떤 조건 하에서 우리가 관찰할 수 있는 행위의 규칙성은 행위자에게 약간의 정보를 제공함으로써 체계적이고 일관된 방식으로 변경될 수 있다.

Pylyshyn에 의하면 인지적 일반화를 표현하기 위해서는 표상에 호소해야 한다⁵⁹⁾.

54) Z. W. Pylyshyn(1981), p. 269.

55) Fodor와 Pylyshyn은 마음에 관한 이론을 두 가지로 구분한다. 하나는 표상주의자(representationist)이고 다른 하나는 제거주의자(eliminativist)이다. 표상주의자에 의하면 표상적 상태를 가정하는 것은 인지이론에서 매우 중요하다. 그들은 세계의 상태를 기호화하는 기능을 수행하는 마음상태가 있다고 주장한다. 제거주의자에 의하면 그런 표상 같은 의미론적 개념은 필요없다. J. A. Fodor, and Z. W. Pylyshyn(1988), "Conectionism and Cognitive Architecture: A Critical Analysis", *Cognition* 28, p. 7.

56) Z. W. Pylyshyn(1985), p. 212.

57) Z. W. Pylyshyn(1987), p. 102.

58) Z. W. Pylyshyn(1987), p. 108.

인지적 일반화는 표상의 의미론적 내용을 언급함으로써 포착될 수 있다⁶⁰). 예를 들어 보자. 나는 현재 나의 컴퓨터 앞에 앉아서 문장들을 타이핑하거나 다른 것들에 대해 생각하고 있다. 나는 아래로 아파트 앞등을 바라보고 있으며, 뒤로는 조그만 야산을 생각하고 있다. 또한 나는 연구를 하는 대신, 식사 후에 왜 산보를 하지 않는가라고 자문하고 있다. 나의 현재 행위는 현재의 목적과 생각에 의해 야기되었다. 목적은 이 논문을 완성시키겠다는 것이고, 생각은 내 뒤의 조그만 야산을 산보하고 있는 것이다.

나의 행위가 존재하지 않는 완성된 논문이나 조그만 야산에 의해 야기되었다는 것이 의미있다면, 나의 행위는 뇌의 어떤 상태에 의해 야기되었다고 해야 한다. 그러나 왜 그 상태가 구체적인 문장을 타이핑하도록 야기했는가를 설명하는 방법은, 상태 자체가 언급된 것들과 어떤 방식으로 관련되어 있다고 말하는 것이다. 나의 뇌의 상태는 산보나 야산과 인과적으로 관련되어 있지 않다. 그 관계는 인과관계가 아니라 의미론적 관계이며, 일종의 내용이다⁶¹). 뇌의 상태는 어떤 움직임의 야기한다. 이 움직임이 '야산을 산보하는 것에 대한 문장을 쓴다'와 같이 기술된 일련의 행위의 구성요소로 간주된다면, 뇌의 상태는 산보와 야산 같은 것에 대한 표상이나 코드를 구체화하는 것으로 취급되어야 할 것이다. 이런 식으로 행위는 어떤 코드의 표상적 내용과 규칙에 의해 합리적으로 관련되어 있는 것이라고 간주될 수 있다.

Pylyshyn에 의하면, 표상적 상태를 갖는 것은 기억의 한 부분에 기호적 표현을 갖는 것이다⁶²). 그 표현은 의미론적 해석을 코드화(encode)하며, 그 표현의 배합적 구조는 하위표현의 내용 사이의 관계를 코드화한다⁶³). 개개의 기호들은 한 체계가 일정한 방식으로 행위하도록 야기할 수 있다. 그런데 의미론적 내용과 대응하는 기호적 표현이 있다면, 우리는 기능적 상태와 표상적 상태가 일대일 관계에 있다고 예측할 수 있으나 반드시 그렇지는 않다. 그 이유는 다음과 같다. 첫째, 의미론적으로 동일한 내용을 가진 일련의 코드(즉 유사한 표현)가 있을 수 있다. 그 코드는 기능적으로는 구별되지만 의미론적으로는 동일할 수 있다. 내용상의 차이와 구조상의 차이의 문제이다. 둘째, 의미론적 내용을 코드화한 기호적 표현으로는 행위를 유발시키기에 불충분하다. 필요한 것은 일련의 메카니즘인데, 이 메카니즘은 체계가 작용하게 하거나 기호를 해석하도록 야기한다. Pylyshyn은 이 메카니즘을 "기능적 구조"(functional architecture)라

59) Z. W. Pylyshyn(1985), p. 23. 표상은 의미론적 관계이다. 의미론적 관계는 인과적으로 정의할 수 있는 것이 아니다.

60) Z. W. Pylyshyn(1985), p. 27.

61) Z. W. Pylyshyn(1985), p. 27.

62) Z. W. Pylyshyn(1985), p. 29.

63) Z. W. Pylyshyn(1985), p. 29. 이에 관한 자세한 부분들은 지금 논의 중에 있지만 전체적인 윤곽은 널리 받아들여지는 것 같다고 Pylyshyn은 말한다.

고 부른다⁶⁴).

여기에서 중요한 것은 기능적 구조가 다르더라도 표상은 같을 수 있다는 것이다. 같은 표상적 상태 (즉 동일한 목적, 신념과 욕구 등)를 가진 두 체계는 기능적으로 구별될 수 있다. 동일한 목적과 신념을 가진 두 사람도 기능적으로 다를 수 있다. 왜냐하면 인간의 행위는 신념과 목적 등과 같은 내용 이상의 것에 의해 결정되기 때문이다. 행위는 통제구조와 모든 유용한 메카니즘을 포함하는 기능적 구조에 의해 결정된다. 표상적으로 동일한 상태는 예측적으로 다른 행위를 야기할 수 있다. 표상 내용을 언급함으로써 포착된 일반화와 기능적 메카니즘을 언급함으로써 포착된 일반화는 다르다. 표상의 의미론적 내용에 의해 표현된 일반화는 “의미론적 수준의 일반화”이고, 기능적 구조의 성질들에 의해 표현된 일반화는 “기호수준의 일반화”이다⁶⁵).

Pylyshyn은 표상적 수준 (또는 의미론적 수준)과 기능적 수준과 물리적 수준을 구별한다. 표상은 기능적 수준의 변화에도 불구하고 동일한 것으로 남아 있을 수 있다. 기능적 상태가 의미론적으로 해석되었다고 말할 때 우리는 기능적 구조가 아니라 기능적 상태의 기호적 코드의 일부분을 언급하는 것이다. 기능적 수준의 차이는 행위에 영향을 주지만 내용의 차이와 관련있는 것으로 간주되지 않는다.

Pylyshyn에 의하면 표상의 의미론은 체계가 어떤 방식으로 행위하도록 야기할 수 없다. 오직 표상의 물질적 형식만이 인과적으로 효력을 가진다. 우리가 목적 G를 바라거나 P라고 믿고 있기 때문에 어떤 행위를 한다고 말할 때, 우리는 표상에 인과성을 부여한다. 물론 그렇게 야기하는 것은 내용이 아니다. 실제로 그렇게 야기하는 것은 표상적 상태의 물리적 성질이다.

Pylyshyn은 여기에서 정신적인 사건이 내용을 가지도록 야기하는 것은 무엇인가라고 묻는다⁶⁶). 정신적인 내용이 어떤 사물을 표상하는 것은 그 내용이 사물과 유사하기 때문이다. X의 이미지가 X를 표상하는 이유는 정신적 표상이나 이미지가 X와 유사하기 때문이다. 한 사물이 다른 사물과 유사하냐 여부는 물리적으로 정의될 수 있는 성질이 아니다. 유사성은 관찰자가 알고 있거나 믿고 있는 것에 의존한다. 가령 대부분의 새들은 서로서로 유사하다. 그러나 새를 기르는 사람에게는 새들이 오리나 토끼 처럼 다르게 보인다. 유사성은 정신적 표상의 의미론적 내용에 어떤 기초도 주지 않는다. 기본적으로 중요한 것은 표상이 의미론적 관계라는 것이다. 의미론적 관계는 인과적으로 정의할 수 없다.

64) Z. W. Pylyshyn(1985), p. 30. 인지과학의 목적은 행위의 규칙성을 기술하는 것만이 아니라 이 규칙성을 유기체의 인과적 메카니즘과 관련시키는 것이다.

65) Z. W. Pylyshyn(1985), p. 32. Newell은 “의미론적 수준의 일반화”를 “지식 수준의 일반화”라고 말한다.

66) Z. W. Pylyshyn(1985), p. 40.

4. 맺음말

이제까지 필자는 인지과학에서 논의되고 있는 과학적 설명의 문제를 검토하였다. 1948년 이후 Hempel이 Oppenheim과 같이 과학적 설명의 문제를 제기한 이후 많은 논의가 진행되어 왔다. 그 대부분의 논의는 Hempel의 설명이론에 대한 비판적 논의라고 볼 수 있다. Hempel의 주장의 핵심은 과학적 설명에는 법칙이 포함되어 있어야 한다는 점이다. 이 법칙은 일반법칙이거나 확률법칙이다. 최근의 논의는 확률법칙의 문제를 중심으로 귀납적-통계적 설명의 문제로 집중되어 있다. 그리고 확률법칙이 과연 인간의 행위설명에서도 타당하느냐는 문제가 제기된다. 이 문제는 인지과학의 등장으로 최근에 다른 국면을 맞이하고 있다. 인지과학의 발달에 의해 인간의 인지구조는 조금씩 과학적으로 밝혀지고 있는 과정에 있다. 물론 앞으로 계속 발전 중에 있으므로 결정적인 결론을 내릴 수는 없다. 그러나 인지심리학의 연구결과에 의하면 인간의 의사결정구조가 언제나 확률계산에 의해 이루어지지는 않는다. Hempel의 설명이론은 확률법칙에 의한 설명이고 인과적 설명이지만 의미론적 관계를 다루지는 못한다. 인지과학의 연구결과에 의하면 인간의 인지구조는 의미론적 과정으로 진행된다. 이런 점에서 P. Thagard와 R. Cummins, J.Haugeland, Z.W.Pylyshyn의 설명이론이 의의를 가진다고 하겠다. 과학적 설명이란 무엇인가라는 물음에 대해서 대답을 하려면 통합적인 설명구조를 제시하는 과제에 봉착하지만 인간의 인지구조가 정확하게 밝혀지지도 않은 상황에서 다양한 설명이론의 시도는 충분히 의의가 있다고 하겠다. 그렇다고 모든 설명이론이 다 타당하다고 말하고 싶지는 않다. 모든 설명이 인과적 설명이어야 한다는 점은 최소요건이라고 생각된다. 그러나 과연 인과란 무엇인가? 또 다시 어려운 철학적 문제에 봉착하게 된다.

참 고 문 헌

- 정 영기(1993), "인지론적 행위설명이론", <<동서철학연구>> 제 10집.
 -----(1996), <<과학적 설명과 비단조논리>>, 엘랜출판사.
- Cummins, R.(1975), "Functional Analysis," *Journal of Philosophy* 72.
 -----(1977), "Programs in the Explanation of Behaviour," *Philosophy of Science*, 44.
 -----(1983), *The Nature of Psychological Explanation*, Cambridge : The MIT Press. A Bradford Book.
- Haugeland, J.(1978), "The Nature and Plausability of Cognitivism", *The Behavioral and Brain Sciences* 1, in J.Haugeland(1981).
 -----(1981), *Mind Design*, Cambridge: The Mit Press. A Bradford Book.
- Margolis, J.(1984), *Philosophy of Psychology*, Prentice-Hall, Inc., Englewood Cliffs.
- Nagel, E.(1961), *The Structure of Science: Problems in the logic of scientific explanation*, New York: Harcourt, Brace & World.
- Newell, A., and H. A.Simon,(1972), *Human problem solving*, Englewood Cliffs, NJ:Prentice-Hall.
- Pylyshyn, Z.W.(1981), "Psychological explanations and knowledge-dependent proceses", *Cognition* 10.
 -----(1987), "What's in A Mind?", *Synthese* 70.
 -----(1985), *Computation and Cognition, Toward a Foundation for Cognitive Science*, A Bradford Book. The MIT Press, Cambridge.
- Schank, R.(1986), *Explanation patterns*, Hillsdale. NJ: Erlbaum.
- Thagard, P.(1991), "Philosophical and Computational Models of Explanation", *Philosophical Studies* 64.