

는 편

형식적 학습이론과 발견의 논리

이 종 권 (중앙대)

인류의 학문의 역사는 갓 태어난 이론에 대해 쏟아지는 허다한 비판적인 견해를 극복하고 중요한 학문으로 성장한 많은 예들을 보여 주고 있다. 수학에서의 비유클리드 기하학과 집합론이나 물리학에서의 양자론이 그러하다. 그러한 예들은 학문 세계에서 “나무도 보기 전에 떡잎부터 자르려는” 완고한 태도가 학문의 발전에 얼마나 해로운가 하는 것을 잘 보여 주고 있다. 이 말은 아직 유아기에 있는 것으로 보이는 형식적인 학습이론에 대해서도 그대로 적용될 것이다. 그런 만큼 그 이론에 대한 파괴적인 비판은 바람직하지 않을 것이다. 또한 논평자로서는 형식적 학습 이론에 대해 별반 아는 것이 없는 만큼 그러한 종류의 비판은 할 생각이 없다. 단지 이와 같은 종류의 이론에 처음 접한 사람으로서 해 볼 수 있는 몇 가지 질문만을 제기해 보고자 한다.

발표된 논문의 제목은 ‘형식적 학습이론’이 ‘발견의 논리’와 모종의 관련이 있음을 암시하고 있다. 그렇다면 ‘형식적인 학습이론’은 아마도 ‘논리화적인 이론’임을 의미하는 것으로 해석해야 할 것이다. 다시 말해 그 이론에 등장하는 모든 원리라든가 법칙은 논리적인 법칙 혹은 논리적인 원리이다. 서양 철학에서 거론된 법칙으로서 우리가 고찰할 가치가 있는 것으로 윤리적인 법칙주의자들이 주장하는 도덕 법칙과 과학적인 인과법칙, 그리고 수학 혹은 논리 법칙을 들 수 있을 것이다. 인간이 어떤 정신적인 목적을 성공적으로 달성하려 할 경우, 그 법칙들에 따라 자신의 사고와 행동을 규제해 가지 않으면 안된다. 예를 들어 법칙주의자들의 주장이 옳다면 우리는 윤리적인 인간이 되고자 할 경우, 도덕법칙에 따라 행동하지 않으면 안된다. 또한 우리를 둘러싼 세계에 대해 올바른 인식을 얻으려 한다면 수학 혹은 논리법칙에 따라 사고하는 것이 필수적이다. 마찬가지로 자연을 효과적으로 이용하려 할 경우, 인과적인 과학법칙에 따라 자연을 조작하는 것이 중요하다.

논리법칙이나 자연에 관한 인과적인 과학법칙은 그 자체 세계의 어떤 측면을 기술함과 동시에 우리에게 규범적인 원칙으로서 기능한다. 다시 말해 우리는 논리법칙에 따라 사유를 하거나 하지 않거나 할 수 있으며 마찬가지로 과학법칙에 따라 자연의 어떤 측면을 조작하거나 조작하지 않을 수 있다. 그러한 법칙에 따라야 한다는 것은 그 법칙과 관련된, 우리가 지향하는 어떤 목표라든가 목적을 두고 하는 말이다. 즉 논리법칙이나 과학법칙을 따라야 한다는 것은 그 법칙에 따르지 않을 경우 어떤 목적을 달성하는 데 실패한다는 뜻이다. 논리법칙이 규범적인 것은 자율적으로 인식적인 사고를 하는 정신에 대해서이다. 논리법칙은 정신이 따르는 법칙은 아니지만 자유롭게 사유하는 정신에 대해서만 규범적인 법칙으로서 기능한다.

마찬가지로 과학법칙은 그 법칙의 대상이 되는 사물에 대해 어떤 믿음과 욕망을 지니는 정신에 대해 규범적이다. 정신은 전 우주를 그 대상으로 할 수 있지만 그 자신이 우주의 일부분이다. 따라서 정신에 관한 과학법칙이 있을 지도 모른다. 만일 있다면 그것은 논리법칙과는 달리 정신이 따르는 법칙이라고 해야 할 것이다. 그러나 그 법칙은 그 법칙에 따르는 정신의 영역을 대상으로 사유할 수 있는 또 다른 반성적인 정신에 대해서만 규범적인 역할을 할 수 있다.¹⁾ 정신에 관한 인과적인 과학법칙은 즉, 심리학의 법칙은 우리가 어떤 목적에서 주로 타인의 절신 과정을 조작하는 데 이용된다. 반면에 논리법칙은 우리의 사고과정을 자율적으로 규제하는 데 이용된다. 예를 들어 실제로 인간이 무엇인가를 배우는 과정에서 거치게 되는 심적인 상태에 관한 심리학적 인과법칙이 있다고 하자.²⁾ 그러한 심리학적 법칙을 동원하여 우리는 다른 사람이 무엇인가를 효과적으로 학습하도록 하는 데 이용될 수 있는 “학습이론”을 구성할 수 있을 것이다.

“형식적인 학습이론”에 등장하는 법칙들이 어떤 성격의 법칙일 수 있는가? 인과적인 심리학의 법칙과 같은 종류의 것인가? 아마도 심리학적 기술적인 인과법칙은 아닐 것이다. 형식적인 학습이론은 우리가 다른 사람을 혹은 우리 스스로를 효과적

-
- 1) “모든 번뇌는 부질없는 욕망에서 비롯되며 또한 후자는 무지에서 비롯된다”는 취지의 불교의 연기설도 정신에 관한 일종의 인과적인 법칙으로 볼 수 있을지 모른다. 그 법칙은 정신의 어떤 상태들을 인과적인 연쇄 속에 넣고 있는데 그 법칙을 반성적으로 사유해서 받아들이거나 받아들이지 않으려는 정신마저도 그 연쇄 속에 들어간다고 해서는 안될 것이다. 왜냐하면 그 경우 그 법칙이 규범적인 법칙으로 기능할 수 있는 정신의 영역이 완전히 없어지기 때문이다.
 - 2) 그러한 법칙은 우리의 정신에 관한 이른바 기술적인(descriptive) 법칙으로서 규범적인 논리법칙과는 종류가 다르다는 것은 잘 알려진 사실이다.

으로 학습시키기 위한 목적에 이용할 수 있는 것은 아닐 것이기 때문이다. 효과적인 학습이 이루어지도록 어떤 조작을 하기 위해서는 심리학적인 학습이론 이외의 다른 종류의 학습이론이 필요할 것으로 생각되지 않는다. 논문의 제목은 형식적인 학습이론이 논리학적인 이론에 속할 것임을 암시하고 있다. 그렇다면 그 이론에 등장하는 법칙은 논리적인 법칙과 같은 부류의 법칙으로 해석해야 할 것이다. 다시 말해 우리의 정신에 대해 대상적인 위치에 있는 것을 조작하기 위해 이용되는 것이 아니라 우리의 자율적인 정신이 자신의 작업을 스스로 규제하기 위해 이용되는 법칙이다.

‘논리’라는 낱말을 목표지향적인 절차에 관한 규범적인 분석(normative analysis)으로 간주할 수 있다는 사이먼의 말에 우리는 얼마든지 동조할 수 있을 것이다. 위에서 말한 것과 같이 “규범적”이라는 말은 그 말이 적용되는 이론을 이용하려는 정신과 그 이론과 관련된 목표를 전제로 한다. 형식적인 학습이론을 이용함에 있어 우리가 지향해야 할 목표는 무엇인가? 그것은 “국한된 영역에 관한 효과적인 발견”이다. 그러나 우리가 ‘논리학의 이론’을 가지고 그러한 목표를 지향할 수 있는지 언뜻 이해가 가지 않는다. 그와 같은 목표는 오히려 ‘심리학적인 학습이론’에 어울리는 것으로 보인다. 토목공학의 이론에 비추어 볼 때, 우리는 어떻게 작업을 하는 것이 다리를 건설하는 데 “효과적인 절차”인가를 평가할 수 있다. 마찬가지로 인간의 정신적인 작용에 관한 심리학적인 (인과)법칙에 비추어 볼 때, 어떻게 하는 것이 주어진 영역에서 학습하는 데 혹은 발견을 하는 데 효과적인 절차인가를 평가할 수 있을 것이다. 그러나 어떤 측면에서 올바른 사고와 그렇지 않은 사고를 구분하기 위한 목적이 아닌 다른 목적에 ‘논리’가 사용될 수 있는지 의심스럽다. 학습이 아닌 앎(know)도 일종의 심리적인 과정이라고 할 수 있다. 앎이 심리적인 과정인 만큼 그와 관련된 심리학적인 이론, 즉 인지심리학적인 이론이 존재한다. 그러나 그에 관한 논리로서 ‘인식의 논리’도 있음을 우리는 알고 있다. 그렇기는 하지만 “주어진 영역에 관한 효과적인 앎”을 얻는데 인지심리학의 이론이 아닌 인식의 논리가 적합한지 의문이다. 인식의 논리는 주어진 영역에서 어떤 방법이 앎을 얻는데 보다 효과적인 절차인지를 평가할 수 있는 과학 이론을 구성하는 경우라면 모르지만, 실제로 효과적인 절차라든가 방법을 얻는 일 자체에 이용될 수 있는 것으로는 생각되지 않는다. 마찬가지로 형식적인 학습이론이 논리학의 이론인 한에서는 효과적인 학습을 위한 조작적인 절차를 마련해 주지 못할 것이다.

인식의 논리는 “안다”(know)는 용어가 등장하는 문장들 사이의 논리적인 관계를 해명하는 것을 목표로 한다. 따라서 그 논리로부터 우리는 그러한 일련의 인

식논리의 문장들로부터 어떤 다른 인식논리의 문장이 연역될 수 있는가를 알 수 있다. 형식적인 학습이론은 그러한 의미에서의 학습에 관한 형식논리도 아닌 것으로 보인다. 발표된 논문에서 시사하고 있는 것은 형식적인 학습이론이 실은 일종의 “귀납논리”라는 것이다. 우리에게 익숙한 귀납논리와는 차이점은 귀납논리가 ‘단기적인’ 귀납절차의 올바름을 평가하기 위한 것이라면 형식적인 학습이론은 ‘장기적인 귀납적인 절차’의 올바름을 평가하기 위한 이론이라는 점이다. 다시 말해 형식적인 학습이론가들은 학습을 일종의 “장기적인 귀납추리” 정도로 보고 있는 것이다. 학습이란 순전히 형식적인 면에서는 장기적인 귀납추리 과정이라는 말이다. 그러한 장기적인 귀납추리 절차의 하나로 들고 있는 것이 ‘모국어 학습’이다.

형식적인 학습이론가들은 학습을 ‘장기적인’ 귀납추리로 규정하는 이유는 말할 것도 없이 전통적인 귀납의 문제들을 비켜가기 위한 것이다. 왜냐하면 그들은 “굿맨의 수수께끼라든가 까마귀의 역설은 …… 모두 관찰적인 진술과 가설간의 검증관계 혹은 관찰진술에 적용된 귀납적인 절차의 단기적인 행태”에 관해 제기되는 문제인 반면 형식적인 학습이론은 “귀납적인 절차의 장기적인 적합성에 관한 평가를 시도하기 때문”이라고 생각한다. 그러나 여기서 말하는 ‘장기적인’ 귀납추리는 비유에 지나지 않는 것으로 보인다. 왜냐하면 추리라고 할 때, 필수적인 요소인 전제와 결론이 무엇인지를 ‘장기적인’ 귀납추리에서는 발견할 수가 없다. 발표 논문에서 나타난 형식적인 학습이론가들의 학습은 ‘장기적인’ 추리이기 보다는 오히려 어떤 것을 습득하기 위한 “행동 전략”이라고 말하는 편이 옳을 것 같다. 또한 학습이론은 일종의 귀납논리 혹은 발견의 논리라기 보다는 일종의 학습이라는 게임에 관한 수학적인 이론으로 보는 편이 타당할 것이다.

형식적인 학습이론가들이 ‘장기적인’ 귀납추리로서의 학습의 표본적인 예로서 드는 것은 모국어 학습이다. 모국어 학습과정에서 귀납추리의 결론에 해당하는 것은 물론 모국어의 문법규칙이라든가 낱말의 의미이다. 그 과정은 말할 것도 없이 잇달은 “시행착오”로 이루어질 것이다. 모국어를 배우는 사람들은 그 과정의 각 단계에서 자신이 입수한 증거자료를 바탕으로 모국어의 문법규칙을 추측하며 그 추측에 상반되는 증거자료가 입수된 경우에는 그 증거자료를 새로 감안하여 또 다시 새로운 문법규칙을 그려내는 작업을 반복하게 된다. 실패한 학습이 이루어지지 않는 한 그러한 추측의 결과는 궁극적으로 일정한 하나의 이론으로 수렴하게 될 것이다. 이렇게 볼 때, ‘장기적인’ 귀납추리란 결국 어떤 일정한 전략에 의거하여 단기적인 귀납추리(시행착오)를 반복하는 것에 해당한다고 말할 수 있을 것이다.

형식적인 학습이론가들이 말하는 학습이 '장기적인' 추리이기 때문에 전통적인 귀납의 문제는 회피할 수 있을지 모른다. 그러나 그 추리가 단기적인 귀납추리를 포함한다면 어떤 '장기적인' 귀납추리를 형식화할 수 있는 가능성이 단기적인 귀납추리의 형식화 가능성 여부에 영향을 받지 않을 수 없을 것이다. 그렇다면 후자의 가능성에 대해 제기되고 있는 부정적인 견해로부터는 쉽사리 빠져나갈 수 없을 것이다. 형식적인 학습이론의 창시자라고 할 만한 퍼트남(Hilary Putnam)은 1992년 저술에서³⁾ 연역논증의 분야에서는 타당한 연역논증을 만족할 만 하게 정식화하는 규칙을 발견했지만 귀납논리의 경우에는 그러한 형식적인 정식화가 가능하다는 사실이 입증되지 않았다는 사실을 지적하면서 인간 지성이 지닌 핵심적인 능력인 귀납추리를 형식적으로 정식화할 수는 없을 것임을 암시하고 있다. 이것을 바꾸어 말하면 형식적인 학습이론이 인간이 모국어 등의 실제로 하는 정신적인 과정을 모방(simulate)하기 위한 것이라면 그러한 목적은 달성될 수 없다는 것을 의미한다.

여기서 귀납논리의 형식화를 가로막는 본질적인 장애물로서 퍼트남이 지적한 것을 몇가지 언급하기로 한다. 귀납을 할 때 우리는 전제와 결론의 형태라든가 더 나아가 내용만이 아니라 결론의 "단순성"(simplicity) 혹은 "우아함"(elegance)까지 문제가 되는 경우가 있다. 그러나 단순성이라든가 우아함은 형식화를 한사코 거부한다. 또한 귀납을 포함한 모든 인식은 사물들 사이의 "유사점"(similarities)들을 인식할 수 있는 능력을 전제로 한다. 이 유사점을 인식할 수 있는 능력은 단순한 능력이 아니다. 왜냐하면 유사함이란 단순히 형태상의 닮음만을 의미하지 않기 때문이다. 예를 들어 전혀 형태를 달리하는 두 개의 대상을 인간은 "칼"이란 단일한 항목으로 분류함으로써 그 대상간의 유사성을 인식하는 경우가 있다. 그러나 그러한 인간의 능력은 형식적인 방식으로 정식화할 수가 없다. 왜냐하면 그 능력은 행위자의 목적을 식별할 수 있는 능력을 전제로 하기 때문이다. 형식적인 면에서만 생각하면 주어진 귀납추리의 전제로부터 대립된 두가지 결론 가운데 어떤 결론을 끄집어 낼 수도 있는 경우가 허다하다. 그러나 인간은 대부분 그 대립된 결론 가운데 어떤 결론이 귀납적으로 옳은 지를 잘 안다. 이것은 인간이 실제로 하는 귀납추리에는 형식적으로는 정식화할 수 없는 요소가 있다는 것을 의미하는 것이다. 형식화란 다루는 대상의 언어에서 의미를 모두 탈색시키는 절차를 포함한다. 귀납추리를 형식화할 수 없다는 것은 즉 귀납추리에 등

3) Putnam, H.(1992), St. Andrews 대학(Scotland) 가을 학기 강의 교재, 1장, The Project of "Artificial Intelligence", 참조.

장하는 낱말의 의미를 언급하지 않고는 궁극적으로 그것의 올바름을 가릴 수 없다는 것을 뜻한다. 퍼트남이 지적한 귀납추리의 형식화를 가로막는 장애들은 바로 그 사실에서 비롯된다고 생각된다. 그러나 형식적인 학습이론이 안고 있는 문제는 단지 귀납추리의 형식적인 정식화의 가능성에만 있는 것은 아니다.

형식적인 학습이론은 형식이론인 만큼 그 이론에 비추어 최선의 학습 전략이라고 생각되는 절차에 따라 가설을 구성해가는 기계를 생각할 수 있을 것이다. 그 기계가 실제로 인간과 같은 의미에서 “학습한다”고 말할 수 있는가? 튜링이 사유에 대해 제시한 행태주의적인 기준을 이 경우에도 적용한다면 아마도 긍정적인 답변을 얻을지도 모른다. 즉 두 방 가운데 하나에는 언어를 학습한 기계를, 그리고 다른 방에는 인간을 집어 넣는다. 그리고 각 방에 연결된 컴퓨터를 이용하여 유한한 수의 문장을 반복해서 넣어주고 그 문장이 속한 언어의 문법을 추리하여 그 답을 컴퓨터에 입력하도록 한다. 만일 각 방에 연결된 컴퓨터에 통해 넣어 준 각 문장들의 집합에 대해 비슷한 절차에 따라 그 언어의 문법을 추리하는 데 성공했을 경우, 두 방에 있는 대상에 대해 언어학습의 능력에 있어 차이가 없다고 말해야 한다는 것이다. 즉 그 중 한 방에 있는 사람에 대해 그가 언어를 학습할 능력이 있다고 말한다면 다른 방에 있는 기계에 대해서도 마찬가지로 언어를 학습할 능력이 있다고 말해야 한다는 것이다.

형식적인 학습이론은 “학습”의 의미에 관한 위와 같은 행태주의적인 기준이 타당성을 전제하고 있는 것으로 보인다. 위에서 제기한 문제는 그 기준을 충족시키는 형식적인 이론이 혹은 학습기계가 있을 수 있는가 하는 것이었다. 그러나 그와는 별도로 그와같은 기준이 타당한 것인가하는 질문을 제기할 수도 있다. 튜링의 기준의 타당성에 대해서는 상당한 반론이 제기되었음을 우리는 알고 있다. 그것은 요컨대 사유란 어떤 자극에 대해 어떤 반응을 보이는 것으로 환원시킬 수 있다는 것이다. 그러한 반론을 제기하는 설(Searl)과 같은 철학자들은 자극과 반응의 중간에 의미를 파악하는 단계가 개재되어 있지 않는 한 전체 과정을 사유과정으로 볼 수 없다고 주장한다. 형식적인 학습이론에 대해서도 동일한 비판이 적용될 수 있는 것 같다. 다시 말해 학습이란 단지 어떤 형식적인 이론에 비추어 최선이라고 생각되는 전략에 따라 증거로부터 이론을 단계적으로 구성해가는 과정에 불과한 것은 아니라는 것이다. 학습이란 무엇보다도 학습자들이 대상을 이해하는 절차를 (모국어 학습의 경우에는 설사 문법을 학습하는 경우에도) 본질적으로 포함하는 과정인 것이다.