

## 계산주의, 연결주의, 단순성

공 용 현 (한국과학기술원)

### 1. 서론 : 경쟁이론의 평가와 단순성의 기준

‘마음의 문제’에 대한 탐구 역사는 철학의 역사만큼이나 오래되었다. 이미 플라톤과 아리스토텔레스의 시대부터 철학자들은 인간의 본질을 ‘이성’으로 규정하고, 인지현상을 해명하기 위한 여러 이론들을 구성하고 발전시켜 왔다. 그러나 인간의 지적 행위에 대한 경험과학적 지식이 축적되지 못했던 20세기 전반까지의 연구는 대체로 원리적이고 사변적인 논의의 수준을 벗어나지 못했으며, 제안된 이론의 타당성이나 효용성을 검증해 볼 수 있는 수단 역시 충분치 못했다. 심적 작용의 주체와 심적 상태 그리고 심적 과정에 대한 본격적인 해명이 시작된 것은 20세기 중반 흔히 ‘마음의 과학’이라고 불리우는 인지과학(認知科學)이 태동한 이후의 일이다.

인지과학은 다학문적(多學問的), 학제적(學際的) 과학으로서 인지 및 인지와 관련된 심적 현상을 마음의 내적 구조, 정보의 수용, 저장, 검색, 변형, 전달, 지식의 활용 등을 중심으로 해명해보고자 하는 과학적 연구이다(Smith, 1990: x; Stillings & Feinstein, 1989: 1-2). 인지과학은 그 역사가 일천함에도 불구하고 다양한 방법론적 틀을 통해 심적 현상을 설명하는 여러 검증가능한 모델 및 이론들을 배출하였다. 한편 최근에 와서 급속도로 발전을 거듭하고 있는 컴퓨터의 등장은 이러한 모델이나 이론을 구체적으로 적용해 볼 수 있는 효과적인 도구이자, 그 타당성을 시험해 볼 수 있는 적절한 수단을 제공하게 되었다.

현재의 인지과학의 탐구 전략이자 방법론 중에서 가장 큰 영향력을 끼치고 있는 것은 계산주의(computationalism)와 연결주의(connectionism)라고 불리우는 접근방식이다. 이 두 접근방식은 인지과학의 영역에서 제안되는 대부분의 인지

모델과 이론이 구성되는 토대이자, 실현가능한 컴퓨터 모델이 설계되고 그 효과가 검증되는 배경이 된다는 점에서, 명실공히 인지과학의 양대 패러다임이라고 불릴 만 하다.

계산주의는 인지과학의 초기 연구를 촉발시켰던 고전적 패러다임으로서 디지털 컴퓨터로 통칭되는 대부분의 컴퓨터 모델과 규칙기반적(rule-based) 인공지능 이론이 근거하고 있는 입장이다. 계산주의자들은 인간의 마음을 계산가능한 기호 처리체제로 간주함으로써, 복잡하고 다양한 인지적 행위가 구성단위들의 유한하고 결정적인 조합으로 설명될 수 있다고 주장한다.

연결주의는 최근에 와서 크게 주목을 받고 있는 인지과학의 새 패러다임으로서 신경망(neural network) 모델과 병행분산처리적(PDP: Parallel Distributed Processing) 인공지능 이론이 근거하고 있는 입장이다. 연결주의자들은 마음을 두뇌의 신경구조와 같은 복잡한 연결망의 작동체제로 간주함으로써, 인지적 행위가 구성요소들의 형식적 계산이 아닌 다른 인과적 상호작용을 통해 설명될 수 있다고 주장한다.

계산주의와 연결주의는 서로 대립되는 패러다임이다. 계산주의에 근거한 인지이론과 연결주의에 근거한 인지이론은 전체와 논리적 구조에서 뚜렷이 구별된다. 각각의 이론들이 받아들이고 있는 존재론적 가정이 다르고, 핵심용어들과 논리적 장치들이 서로 상충하고 있다. 그럼에도 불구하고 계산주의적 인지이론과 연결주의적 인지이론은 공히 내적으로 논리적 일관성을 갖추고 있으며, 인지현상의 여러 영역에 걸쳐 경험적 대응성을 지니고 있다. 과학적 이론의 구성요건을 충족시키고 있다는 점에서 두 이론은 사이비 과학이론이 아니라 진정한 과학이론이다.

한편 계산주의적 인지이론과 연결주의적 인지이론은 서로 경쟁관계에 놓여져 있는 이론들이다. 계산주의와 연결주의의 신봉자들은 인지현상의 일부분이 아니라 인지현상 전반을 설명할 수 있는 통합이론의 구축을 목표로 하고 있다(Newell, 1990; Churchland, 1986). 또한 양측은 상대방이 근거하고 있는 패러다임이 잘못 설정된 것이라고 비판하며, 한쪽이 다른 쪽을 포괄하거나 대체할 수 있다고 주장한다(Pylyshyn, 1984; Ramsey et al., 1990).

동일한 현상을 설명하는 두 이론의 경쟁상태가 오래 지속되면 이론의 평가와 더불어 선택의 문제가 뒤따르게 마련이다. 이 두 경쟁이론 중 어느 것이 인지현상을 적절하게 설명하고 있는 이론인가? 이들 중 하나를 선택해야 한다면 어느 것을

취해야 하는가?

과학적 탐구의 대상이 될 수 있는 한 영역에 대해 서로 경쟁하는 과학이론은 참된 이론이라고 알려진 다른 영역의 과학이론들과 수평적, 수직적 정합성을 이룰 때 강한 지지를 받는다. 이런 이론이 성숙한(mature) 과학이론이고, 인지이론 역시 인접된 영역의 과학이론, 즉 물리학, 화학, 신경생리학의 이론들과 정합성을 이룰 때 성숙한 과학이론이 된다(Rohrlich & Hardin, 1983: 604). 그러나 두 이론이 모두 성숙한 과학이론으로 평가받게 되거나, 인접된 과학이론과의 정합관계가 불분명한 경우, 이론선택의 최종 판결자로서 어떤 기준을 선택할 수 밖에 없는 데, 그럴 때 곧잘 적용되는 기준이 '단순성(simplicity)의 기준'이다. 단순성의 기준이란 '두 경쟁이론 중 단순한 이론이 진리에 가깝다'는 것이다.

본고는 계산주의와 연결주의 중 어느 것이 인지과학의 연구에 있어 타당한 것 인지, 그 패러다임을 토대로 구성된 이론들의 단순성에 주목함으로써 판정해보고자 한다. 이를 위해 계산주의와 연결주의의 실상을 먼저 살펴보고 그 관계에 대해 약간의 정리를 시도할 것이다. 다음으로 올바르게 정의된 계산주의와 전형적인 연결주의에 토대를 둔 이론들을 '존재론적 단순성'과 '인식론적 단순성'의 두 측면에서 고찰하고, 계산주의적 이론이 연결주의적 이론보다 더 단순한 이론임을 드러내보일 것이다.

## 2. 계산주의와 연결주의

계산주의 이론과 연결주의 이론을 제대로 평가하기 위해서는 이론의 토대가 되고 있는 두 패러다임 사이의 상반된 특성들을 분석해 볼 필요가 있다. 왜냐하면 이론에 포함된 존재론적, 인식론적 가정과 논리적 구조가 여기서 비롯되고 있기 때문이다.

계산주의는 종종 '계산적 모델' '기호조작주의' '고전주의'라고도 불리워진다. 계산주의에 대한 정의(定義)는 학자에 따라 세부적 내용과 표현 방식에 따라 다소 차이를 보이고 있지만, 그 요점은 "심적 과정이란 정보의 유한한 계산적 처리과정이다"라는 한마디 문장으로 기술될 수 있다(Cole, 1990: 3-4; Jackendoff, 1989: 3).

여기서 정보(information)란 기호(symbol)로 표현될 수 있다. 그러나 흔히 말하듯 의미(意味)가 할당되고 해석된 기호가 아니다. 정보는 시스템 사이의 소통

을 위해 생성되고, 전달되고, 수용되는 신호(signal)로서, 그 신호를 해석하고 사용하는 지적인 주체의 바깥에 객관적으로 존재하는 어떤 것이다. 흔히 정보로 간주되는 명제나 진술은 그것들의 원천으로서의 정보와 구별되어야 한다. 이것들은 정보에 담겨져 있는 것, 혹은 정보의 처리 과정을 통해 생성된 것이며, 인지체계에 의해 해석된 것이다(Dretske, 1982: vii). 따라서 계산주의에서 말하는 정보나 기호를 단지 명백한 심적 표상에만 대응되는 것으로 간주해서는 안된다(Sloman, 1990: 245-6).

다음으로, 계산주의의 정의에서 '계산(computation)'이 의미하는 바는 전통적인 의미에서의 산술(arithmetic), 즉 '더하기' '빼기' '곱하기' '나누기' 만은 아니다. 계산에는 정수, 실수의 계산 및 미적분을 포함하여 논리학자나 컴퓨터 과학자가 사용하는 논리연산으로서의 계산, 즉 'AND' 'OR' 'NOT' 'XOR' 과 'IF THEN ELSE' 등 비교(comparison)와 분기(branch)의 계산 개념도 포함된다. 또한 물리학자가 사용하는 바와 같은 실제의 기계 속에서 일어나는 물리적 과정, 예를들어 컴퓨터 칩 속의 '게이트(gate)' 나 '플립플롭(flip-flops)'의 작동도 계산의 개념에 포함된다(Davis & hersh, 1987, 139-145). 컴퓨터의 소자들은 전압의 어떤 영역을 두 가지 특정상태로 식별하고, 이 상태의 조합으로 새로운 상태를 산출한다는 점에서 논리적이고 계산적이다.

한편 계산주의의 정의에서 주목해야할 점은 정보의 계산적 처리가 '유한한' 과정임을 명기하고 있다는 것이다. 사람의 두뇌처럼 유기체적 시스템이든, 컴퓨터와 같은 기계적 시스템이든 시공 속에서 작동되는 물리적 시스템인 이상 무한한 양의 정보를 무한한 시간동안 처리할 수는 없다. 컴퓨터의 경우 사람이 지닌 것과 같은 생물학적 제한, 즉 호흡, 온도, 영양의 조건에 구애되지는 않지만, 유한한 용량의 메모리, 유한한 경로의 회로망, 유한한 연산속도, 거기에서 전기에너지가 지속적으로 공급되어야만 정상적으로 작동될 수 있는 것이다. 따라서 계산을 흔히 수학자들이 생각하듯 무제약적인 과정으로 간주하는 것은 추상적 계산이론에서는 타당하지만 여기서 말하는 계산주의의 본질에는 어긋나는 것이다.

'정보'와 '계산' 그리고 '유한한 절차'에 대한 위와 같은 해석을 고려해 볼 때 계산주의가 현행의 컴퓨터 모델과 상통하는 것임을 금방 알 수 있다. 디지털 컴퓨터의 모델에는 질적인 특성을 양적인 특성의 유한한 조합으로 간주하는 계산주의의 핵심이 깔려 있다. 컴퓨터 내에서 일어나는 기본적 과정은 단순한 양들의 형식적 계산과정이고 직접적으로 믿음이나 지식 또는 통속심리학의 설명에 도입되

는 고유한 심적인 범주를 필요로 하지 않는다. 이점에서 계산주의는 심리철학에서의 심신이원론(dualism)이나 부현상론(epiphenomenalism)과 대립되고 있으며, 심적 현상 전반을 원초적인 물리기호(physical symbols)로 설명한다는 점에서(Newell, 1980; Newell & Simon, 1976), 미시결정주의와 환원주의의 입장을 따르고 있다.

연결주의는 흔히 병행분산처리적 접근 혹은 신경망 모델이라고 불리기도 한다. 연결주의는 지금도 이것을 토대로 한 인지이론들이 다방면에서 분화와 변모를 거듭하고 있고, 학자들마다 계산주의와 관련하여 해석상 상당한 불일치를 보이고 있기 때문에 그 특성을 한마디로 정의하기가 쉽지 않다. 그러나 연결주의의 지지자들의 견해를 종합해 볼 때, 그 핵심을 “심적 과정이란 신경구조와 같은 대규모 연결망 속에서 활성화(activation)의 정도가 다른 구성요소들 사이의 연결통로와 연결강도의 차이에서 야기되는 정보의 변형이다.”라는 말로 요약할 수 있다(Schwartz, 1988: 123-141; Smolensky, 1990: 145-206).

연결주의는 흔히 아래와 같은 몇가지 잘 알려진 디지틀 컴퓨터와 인간의 두뇌 사이의 차이점에 주목할 때 강한 설득력을 지닌 것으로 여겨진다(Jackendoff, 1989: 33-36).

첫째, 컴퓨터의 기본적인 기능단위는 계수적(digital)이며 ‘켜짐(on)’과 ‘꺼짐(off)’의 두개의 상태만을 갖는 이진(二進) 스위치이다. 이에 반하여 두뇌의 기본적인 기능단위는 연속적인 변량(variables)이 상호작용하는 뉴런이며, 뉴런들의 활성화 정도와 뉴런들 사이의 연접(synapses)에서 일어나는 행태는 계수적, 단속적(斷續的)인 것이 아니라, 의사계수적(quasi-digital) 혹은 상사적(analogue)이다.

둘째, 현재의 계수형 컴퓨터의 표준적 구조에서는 여러 연산이 직렬적으로 처리되고 있다. 이에 비해 두뇌 속에서의 정보처리는, 뉴런들 사이의 복잡한 연결망에 비추어 볼 때 한 뉴런의 작동이 동시에 다른 뉴런들의 작동에 영향을 미치는 방식, 즉 대량의 병렬 처리의 방식으로 이루어지는 것 같다.

세째, 일반적인 컴퓨터의 경우, 컴퓨터의 각 부분은 일반 목적에 맞게 여러 용도로 사용될 수 있지만, 일단 내부 구조가 형성된 후에는 한 부분의 장애는 전체 시스템의 기능에 치명적 영향을 끼친다. 반면에 두뇌의 기능은 뇌의 좌반구와 우반구의 기능의 차이와 역할에서 알 수 있듯 정보의 성질과 처리방식에 따라 각 부분이 분화되어 있다. 두뇌는 비록 각 부분이 특화화 되어있으나, 한 부분이 손상을

입을 경우 다른 부분이 그 기능을 대신할 수 있는 생물학적인 융통성이 허용되어 있다.

네제, 컴퓨터는 하드웨어가 고정되어 있다. 컴퓨터가 다양한 작업을 수행할 수 있는 것은 물리적 구조로서의 하드웨어가 바뀌는 것이 아니라 프로그램으로서의 소프트웨어가 바뀌기 때문이다. 이에 비해서 두뇌 내에는 하드웨어와 소프트웨어의 명백한 구별이 없다. 두뇌 내에서는 끊임없이 하드웨어적인 변화가 일어나고 있다. 두뇌의 기능 변화의 원인이 된다고 생각되는 혈액과 홀몬 성분의 변화와 같은 화학적 요인에 대응하는 것이 현행의 디지털 컴퓨터 내에는 없다.

이상의 차이점에 착안하여 연결주의자들은 인지과정이 계산주의에서 말하는 분절된 정보의 계산적 처리과정이 아니라 연속된 정보가 복잡하게 상호작용하는 역동적 과정으로 파악한다. 연결적 시스템에서의 기본 요소나 변수들은 현행의 폰 노이만식의 컴퓨터 이론에서 다루어지고 있는 기호보다 낮은 단계의 것들, 즉 명제적 기호보다 더 원초적인 하위기호들(subsymbols)이며, 처리과정도 명백한 상위표상들의 형식적인 조작이 아니라 시스템 내에 광범위하게 분포된 여러 요인들의 병렬적 상호작용이다. 연결주의자들은 연결망의 이런 특징 때문에 연결적 시스템이 계산적 시스템의 직렬적 조작에서 발생하기 쉬운 오류에 잘 견딜 수 있다고 주장한다. 또한 하드웨어와 프로그램이 뚜렷이 구별되어 있고, 또 고정되어 있는 계산주의적 모델과는 달리 연결망 자체가 바로 작동 프로그램이 될 수 있다는 점에서 연결주의적 모델의 우수성을 주장하기도 한다(Smolensky, 1990: 146).

이상에서 계산주의와 연결주의의 특징을 간략하게 살펴보았다. 그러나 ‘정보’ ‘계산’ ‘연결’이라는 중요한 개념들이 모호하게 사용됨으로써, 상대방의 입장이 오해되거나, 두 입장 사이에 중복되는 부분이 있음을 발견할 수 있다. 특히 연결주의는 그것과 대비되는 계산주의를 너무 좁게 다룸으로써 상대적으로 너무 넓게 정의되고 있다.

연결주의자들이 계산적 시스템의 정보처리 방식을 단지 명제적인 기호에 대한 산술적인 조작으로 파악하는 것은 계산주의를 잘못 규정하는 것이다. 계산주의에서 말하는 기호란 반드시 명제적인 것에 국한되지는 않는다. 언어의 차원에서 보자면 문장 뿐만 아니라 ‘단어’도 ‘구(phrase)’도 그리고 각종 연결사들도 모두 기호의 단위가 될 수 있다. 그리고 계산이란 앞에서 살펴보았듯이 단순히 산술적 의미만을 지닌 것이 아니라 일반적으로 한 기호를 다른 기호로 전환시키는 과정

모두를 뜻한다. 또한 그 계산은 산술 체계나 기하학 체계에서 일어나는 것과 같은 추상적 과정이 아니라, 컴퓨터라는 기계 속에서 일어나는 구체적/물리적 과정이다.

올게 해석된 계산주의는 기호의 산술적 조작을 뜻하는 '좁은 의미의 계산주의'에 대비해서 '넓은 의미의 계산주의'라고 불러야 될 것이다. 넓은 의미의 계산주의는 연결주의의 진영으로 분류되어 왔던 모든 유형들과 상충되는 것은 아니다. 연결주의자들 중 상당수가 계산주의의 기본 요소를 공유하고 있고, 또 인지과정에 대한 그들의 표현 역시 비록 복잡하기는 해도 수학적 공식의 형태를 띠고 있음을 볼 때, 이런 유형의 연결론을 계산주의의 영역 바깥에 있는 새로운 패러다임이라고 보기는 어렵다. 이것은 계산주의의 기호처리 접근방식을 보완하는 접근방식이라고 볼 수 있거나(Cole, 1990: 4), 단지 구현(implementation) 차원에서의 상이한 이론으로 보는 것이 타당하다(Marr, 1982). 이런 연결주의는 약한 연결주의이며(Cole, 1990: 19), 기호적 계산주의와 더불어 연결적 계산주의라고 부를 수 있다. 넓은 의미의 계산주의는 기호적 계산주의와 연결적 계산주의를 포함한다.

그러나 한편으로 연결주의자의 범주에는 복잡하게 상호작용하는 연결망 속에서 하위수준의 요소로부터 새롭게 창발(emerge)되는 어떤 속성을 강조하는 사람들이 있다. 그들은 인간의 뉴런(neuron)들이 그렇게 구성되어 있으리라고 여겨지는 비선형적인(nonlinear) 연결망과, 중앙처리장치와 독립하여 자기조직화하는 지엽적인 생리현상에 주목한다(Hopfield, 1982: 2554-8). 이들은 심적 과정의 기본 요소를 뉴런보다 낮은 단위의 물리적 요소로 환원하지 않는다는 점에서, 즉 질적인 특성을 양적인 특성의 조합으로 환원하지 않는다는 점에서 계산주의와 뚜렷이 구별된다. 그리고 단위 요소들 사이의 계산과정이, 단지 복잡성의 척도에서 구분되는 한갓 구현의 차원이 아니라, 형식적 계산 이상의 과정이라고 주장하는 점에서 분명히 계산주의와 상충된다(Churchland & Sejnowski, 1989; Ramsey et al., 1990). 이러한 연결주의는 약한 연결주의와 대비해서 강한 연결주의라고 부를 수 있다(Cole, 1990: 20). 강한 연결주의는 심리철학에서의 부현상론과 상통하며, 미시결정주의나 환원주의가 아니라 목적론과 총체론(holism)의 입장을 따르고 있다.

넓은 의미의 계산주의와 강한 연결주의는 상대 패러다임과 대립되는 특징들을 확연히 드러내 보인다. 이것들을 토대로 구성된 인지이론은 기본요소와 연산과정이 아주 다르고, 바로 이러한 차이점을 통해 경쟁자들은 상대방의 이론으로서 설명하지 못하는 인지양상들을 성공적으로 모의할 수 있다고 주장하기 때문이다.

계산주의적 인지이론과 연결주의적 인지이론 중 어느 하나가 상대방이 전혀

설명하지 못하는 부분을 설명하고, 나아가서 상대방이 설명하는 부분까지 잘 설명할 수 있다면, 이론의 포괄성과 일반성의 측면에서 볼 때, 분명히 더 나은 이론으로 평가받을 것이다. 그러나 경쟁이론의 한쪽이 결정적인 우위를 차지하지 못할 경우, 상대적으로 더 간결한 구조를 지닌 이론이 선호될 것이다. 이 문제를 존재론적 단순성과 인식론적 단순성의 측면에서 논의해보자.

### 3. 계산주의와 존재론적 단순성

계산주의에서 말하는 정보의 일차적 특성은 순수하고 물리적인 양으로서 규정되며 여기에는 질적인 요소가 포함되지 않는다. 정보의 형태에 대한 전통적인 구분, 즉 계수적(digital) 정보와 유사적(analog) 정보의 차이는 질적인 차이가 아니라 정보를 부호화(encoding)하는 방식의 차이이다. 계산주의적 이론에 따르면, 정보원(information source)이 되는 가변적인 대상의 속성, 예를들어 크기, 시간, 속도, 온도, 압력 등을 연속적인 표상으로 또는 불연속적인 표상으로 부호화할 수 있다. 지각 경험에서 잘 드러나는 것처럼 사람들은 흔히 지각 내용을 연속적인 것으로 지각한다. 그러나 지각 대상을 연속적인 양으로 경험한다고 해서 대상의 속성 자체가 연속적인 것이라고 단정할 수는 없다.

정보를 어떤 방식으로 부호화할 것인지 그리고 정보의 기본 단위를 어느 것으로 설정할 것인지는 처리 메카니즘의 효율성과 밀접한 관계를 지닌다. 현재의 정보이론과 통신이론에서는 정보량을 어떤 사태의 발생확률로 정의한다. 이것은 불확실성의 감소(reduction in uncertainty)를 뜻하고 있으며, 두가지 가능성 중 하나를 선택하는 결정의 조합으로 간주할 수 있다. 이 정보량을 표현하는 여러 방식이 있을 수 있겠지만 측정과 처리의 효율성에서 볼 때 이진화(二進化)된 계수(digit)로 표현하는 것이 큰 장점을 지닌다.

이 점에서 계산주의는 기본적인 최소 정보단위로 '비트(bit)'를 채택한다. 이것은 원래 샤논(Shannon)의 제안인데, 그에 따르면 "정보란 비트라는 최소 정보단위의 조합에 의하여 순전히 수량적으로 측정되는 것"이다(Shannon, 1949: 32). 그는 부울(Boole)의 '1'과 '0'를 이용한 이치(二值)논리를 전기통신회로에서 스위치가 갖는 '켜짐(on)'과 '꺼짐(off)'의 두 상태와 대응시킴으로써 컴퓨터를 통한 정보처리의 길을 열어 놓았다.

계산주의적 인지이론은 복잡하고 다양한 인지현상을 단지 '1'과 '0'라는 두



기본단위로서 설명하려한다는 점에서 존재론적으로 단순하다. 이에 비해 연결주의적 인지이론은 ‘뉴런의 반응시간’ ‘활성화 수치(activation value)’ ‘가중치(weight)’ ‘강도(strength)’ ‘임계함수(threshold function)’ ‘숨겨진 장치(hidden unit)’ 등 존재론적으로 복잡한 기본단위들을 지니고 있다. 연결주의자들은 혹시 이러한 요소들이 뉴런이라는 하나의 기본단위와 그것들의 연결관계로 재구성될 수 있기 때문에 계산주의적 인지이론에 비견될만한 단순성을 확보하고 있다고 주장할지 모른다. 사실 간단한 신경망 모델을 예로 들어볼 때, 거기에는 일견 입/출력 벡터와 그러한 자극치들의 변형을 결정하는 어떤 내부 장치가 있을 뿐이다.

그러나 문제는 ‘이 숨겨진 어떤 장치가 무엇인가’ 하는 점이다. 이 숨겨진 장치는, 연결주의자들이 옹호하는 수많은 기능적 장점들이 여기서부터 비롯되고 있지만, 결코 단순한 요소가 아니다. 이것의 세부과정이 향후 전개될 신경생리학적 성과에 의해 밝혀질 뉴런의 고유특성이라고 말한다면, 연결주의적 인지이론은 아직 불완전한 이론이다. 또한 이 세부과정을, 암호화하는(encoding) 연결망과 복호화하는(decoding) 연결망사이의 복잡한 텐서곱(tensor product)으로 해석하는 것은, 그 속에 하위규칙성(subregularity)과 같은 또 다른 계산규칙이 도입되기 때문에 계산주의적 인지이론보다 더 복잡한 이론이 되고 만다.

어떤 연결주의자들은 인간의 두뇌구조, 심적 표상, 그리고 처리과정이 ‘1’과 ‘0’으로 표현될 수 없을 만큼 복잡하기 때문에, 인지이론에서 만큼은 ‘단순성의 추구’가 의미없다고 반박할지 모른다. 예를들어 비명제적인 표상이나 ‘참’ ‘거짓’ 아닌 어떤 상태를 ‘1’과 ‘0’이라는 두 상태로 묘사할 수 있겠는가? 그러나 계산주의에서 말하는 ‘1’과 ‘0’이 그대로 하나의 명제에 대응하는 것은 아니다. 한 명제는 물론이거니와 ‘한 단어’ ‘한 문자’를 표시하는데도 이진숫자의 조합이 요구되는 것처럼, 비명제적인 표상도 얼마든지 비트의 조합으로 묘사할 수 있다.<sup>1)</sup>

한편 계산주의적 인지이론이 단지 이치논리만 다룰 수 있다고 보는 것도 오해이다. ‘참’ ‘거짓’ 사이에 있을 수 있는 수많은 상태를 직접 비트의 조합으로 표현할 수 있고, 그 농도를 확률분포로 다룰 수도 있다. 이점은 무한다치논리라고 할 수 있는 퍼지(fuzzy)논리를 수학적형식으로 표현할 수 있을 뿐만 아니라, 지금의 폰 노이만식 디지털 컴퓨터를 통해 잘 구사할 수 있음을 볼 때 명백하다.

1) 예를들어 ‘I am!’이라는 문장을 ASCII코드로 표시하면 ‘1001001/0100000/1000001/1001101/0100001’이라는 긴 이진수의 조합이 된다(Casti, 1989: 277).

계산주의적 이론과 연결주의적 이론 중 어느 것이 존재론적으로 더 단순한 이론인가 하는 문제는 이론 내에 존재론적으로 개입하고 있는 용어들의 갯수로만 결정되는 것은 아니다. 이론적 용어라고도 할 수 있는 이런 기본요소들이 무엇인가를 알아야 하고, 이것은 최종적으로 두뇌구조와 사고과정이 실제로 어떤 요소로 이루어지고 또 작동되고 있는지가 규명되어야만 판정될 성질의 것이다. 그러나 우리는 아직 두뇌의 구조를 물리화학적 차원에서 분석하지 못하고 있고, 두뇌의 신경생리학적 과정 또한 계수적인지 상사적인지 잘 모르고 있다. 이점에서 계산주의와 연결주의는 아직 증명되지 않은 존재론적인 가정을 전제하고 있다. 하지만 동일한 인지현상에 대해 비슷한 설명력을 갖추고 있는 경쟁이론들이라면, '필요없이 실재(entity)를 증가시키지 말라'라는 오컴(Ockham)의 경구가 적용될 수 있고, 이것은 기본요소가 상대적으로 적은 계산주의적 이론을 선호할 근거가 될 수 있다.

#### 4. 계산주의와 인식론적 단순성

단순성의 기준에 있어서 인식론적 측면은 '단순한 이론이 복잡한 이론보다 잘 확인되거나 반증될 수 있기 때문에 선호된다'는 것이다(공용현, 1985: 37). 계산주의적 인지이론은 인지과정을 주로 알고리즘(algorithm)으로 기술함으로써 이런 인식론적인 단순성의 기준을 충족시키고 있다. '일반문제해결(GPS) 시스템'의 예를 들어보자. 알고리즘 방식은 규정된 출발점과 종료점을 가지고 있으며, 문제의 해답에 도달하기 위해 순차적인 처리과정을 밟아가는 방식으로서 특정한 문제의 해답이(또는 해답이 없다는 것이) 보증되어 있다. 따라서 계산주의를 토대로 구성된 인지이론이나 모델은, 그것의 성공여부를 프로그램의 처리 결과를 통해 쉽게 판별할 수 있고, 실패했을 경우 어느 부분이 잘못되었는지를 추적할 수 있다. 이에 반해서 연결주의적 인지이론이나 모델은 성공과 실패의 여러 정도(degree)가 있을 뿐, 이것이 참된 것인지 아닌지를 확인하거나 반증하기 어렵다.

연결주의에는 하위수준의 입력상태와 상위수준의 기호적 처리과정사이에 뚜렷한 구별이 없다. 연결주의자들은 이러한 모호성이 바로 인지과정의 특징이라고 주장하고, 상식적 추리, 패턴인식, 자연언어이해 분야에서의 계산주의의 잘못된 접근 방식을 지적한다. 그러나 상식적 추리, 패턴인식의 일부 분야, 의미처리의 특정 영역에서 연결주의가 효과를 거두고 있는 것이 사실이지만, 인지현상 전체를 놓고

볼 때 연결주의적 인지이론은 아직도 실패의 확율이 높은 이론이다. 특히 고수준의 지식표상, 자료검색, 의사결정과 같은 인지양상에 대해서는 계산주의적 인지이론이 갖는 장점을 거의 대체 못하고 있다. 더구나 이런 인지양상에 대한 컴퓨터 모의의 실패에 대해 어느 전제가 잘못되었고, 어느 과정에서 문제가 발생했는지를 쉽게 가려내지 못하기도 한다. 이 이유는 연결주의적 이론들이 하드웨어적인 계산 과정과 소프트웨어적인 인지과정 사이를 연결하는 절차를 명백히 기술할 수 없기 때문이고, 인지과정의 세부내용보다는 총체적인 인지기능의 발현에 초점을 맞추고 있기 때문이다. 연결주의의 이런 성격은 인식론적 단순성의 관점에서 볼 때, 계산주의에 비해 상대적으로 복잡한 이론이나 모델을 배출하게 마련이다.

한편 연결주의적 이론이 더 잘 설명하고 있다고 주장하는 분야에 대해 계산주의적 이론이 전혀 속수무책인 것은 아니다. 예를들어 '기억'과 '학습'에 있어서 연결주의에 입각한 '역전파(backpropagation) 모델'<sup>2)</sup>이나 '볼츠만(Boltzmann) 학습 모델'<sup>3)</sup>이 상당한 효능이 있음이 사실이지만, 기호처리방식을 채택하고 있는 'SOAR 시스템'<sup>4)</sup> 역시 학습에 있어서 괄목할만한 효과를 거두고 있다(Newell, 1990 : 38-39). SOAR와 같은 계산주의적 학습 프로그램이 완전한 성공을 거두기 위해서는 연결적 시스템보다는 대용량의 메모리와 고속의 처리속도를 필요로 한다. 이점은 인간의 두뇌가 제한된 용량을 가지고 있고, 그 처리 속도가 컴퓨터에 비해 아주 느리다는 점에서 곧잘 연결주의자들의 반박의 표적이 된다.

그러나 인지과학은 반드시 인간의 뇌와 '꼭 같은' 정보처리시스템의 구축을 목표로 하는 입장은 아니다. 새의 '나는' 기능을 모의하기 위해 비행기가 퍼덕이는 두 날개를 가지고 있어야 될 필요가 없듯, 인지기능을 모의하기 위해 컴퓨터가 반드시 두뇌의 구조와 작동방식을 복제할 필요는 없다. 오히려 제트기나 로켓이

- 
- 2) D. Rumelhart, G. Hinton, R. Williams 등이 1980년대에 제안한 학습 모델. input unit에서 hidden unit를 거쳐 output unit로 향하는 활성화의 전달과정에 있어서 오류(error)가 발생하면, 그 가중치(weight)를 조정해 결과를 역방향으로 전달하여 주어진 과제를 해결한다(Bechtel & Abrahamsen, 1991: 85-86).
  - 3) G. Hinton, T. Sejnowski가 제안한 학습 모델. feedback이 아니라 feedforward network의 특징을 지닌다. 시스템에서 원하는 결과를 산출하면 그것을 야기시킨 연결 부위의 가중치를 증가시키는 방식을 취한다(Bechtel & Abrahamsen, 1991: 97-99)..
  - 4) 1980년대 J. Laird, P. Rosenbloom, A. Newell에 의해 범용 인공지능의 구현을 목표로 개발된 시스템. '문제공간' '목표-부목표 위계' '산출(production)시스템'의 특징을 지니고 있다.

통상적인 날개의 개념을 버렸기 때문에 더 멀리 그리고 더 빨리 날아갈 수 있는 것처럼, 계산주의를 택함으로써 인간의 인지기능보다 더 탁월한 기능을 발휘할 수 있는 시스템의 제작이 가능하다고 주장할 수 있다.

컴퓨터의 하드웨어 기술이 하루가 다르게 발전하고 있다는 사실은 계산주의의 적절성을 시험해 볼 수 있는 새로운 여지를 계속 제공한다. 컴퓨터의 용량과 처리 속도에 원리적인 한계가 있지만, 그 한계에 비추어본다면 지금의 컴퓨터는 그야말로 장난감에 지나지 않는다. 계산주의적 이론이 지닌 인식론적 단순성은, 기억용량의 집적화와 처리속도의 고속화가 인지현상의 복잡성과 맺는 상관관계를 더욱 분명히 해명해줄 것이고, 이론들이 스스로를 수정 보완함으로써 점근적으로 진리에 도달할 수 있도록 촉발할 것이다.

인식론적 단순성은 경쟁이론의 평가에 충분한 기준은 아니다. 그러나 이 요건이 확보되지 않으면, 인지과학의 지식 축적과 인지이론의 점진적 발전을 기대하기 어렵다. 연결주의적 이론이나 모델들은 계산주의적 이론이나 모델들에 비해 상대적으로 산만하다. 상당수가 아직도 밝혀지지 않은 두뇌의 신경구조에 대한 직관에 의존하여 모델을 구성하고, 거듭되는 시행착오를 통해 이론을 개조해나가고 있다. 혹시 이런저런 미봉(ad-hoc)가설을 도입한 연결주의적 이론이 특정한 인지현상을 잘 설명하는 행운을 맞게될 수도 있을 것이다. 그러나 이론의 평가는 그것의 경험적 사실에 대한 대응만으로 결정되는 것은 아니다. 톨레미(Ptolemy)의 이론이 천문현상에 대해 오랜 기간 동안 설명적 성공을 거두었다해도 실상 참된 이론이 아니었던 것처럼, 이론이 인식론적으로 단순하지 않다면 결국 다른 이론에 흡수되거나 폐기될 수 밖에 없는 것이다.

## 6. 결론

현재의 시점에서 볼 때, 계산주의와 연결주의가 아직은 인지현상 전반을 남김 없이 설명하지 못하고 있는 것이 사실이다. 각각의 패러다임에 근거해서 설계된 시스템은 '추리에 있어서의 유연성과 창조성', '지각의 현상주의적 총체성', '언어이해에 있어서의 해석학적 특성', '학습능력' '자의식' 등과 같은 인지적 기능을 충분히 모의하지 못하고 있다. 특히 '정의적(emotive) 기능'에 대한 컴퓨터 모의 가능성은 이제 겨우 문제제기의 수준에 머무르고 있는 실정이다. 어떤 사람은 인간의 마음을 명백한 표상이 처리되는 상위기호의 직렬체계와 암묵적인 표상이 처

리되는 하위기호의 병렬체계가 중첩된 다층구조로 파악하고, 이 기능을 효과적으로 모의할 수 있는 '혼합모델(hybrid model)을 제안하기도 한다. 그러나 이 제안을 따르다해도 이런 기능을 발현하는 컴퓨터 설계가 불가능 할만큼 복잡한 이론을 구성해서는 안된다.

계산주의 인지이론과 연결주의 인지이론은 언어학, 심리학, 생물학의 이론이 그렇듯이 발전도상에 있는 이론이다. 이점에서 두 인지과학의 이론은 수정보완의 과정을 거치는 가운데 조만간 성숙한 과학이론이 되거나, 부적절한 것으로 판정되어 폐기될 성질의 것이지 아직은 단순성과 같은 이론평가의 기준을 적용할 상황은 아니라고 볼 수 있다. 그러나 단순성의 기준은 이론선택의 최종판결자임과 동시에 과학자로 하여금 더 좋은 이론을 구성하게 하는 동기가 될 수도 있는 것이다. 더 단순한 것을 통해 복잡한 것을 설명하려는 태도는 두뇌의 계산능력이 유한한 인간 존재가 취할 수 밖에 없는 자연스러운 경향이라고 여겨진다.

계산주의는 마음의 다층적 구조를 단일한 정보처리 이론으로 이해할 수 있는 길을 열어 놓는다. 계산주의는, 특히 넓은 의미의 계산주의는 물리적 기반에서 추구되는 신경생리학적 연구성과와 정합될 수 있을뿐만 아니라, 일부에서 제기되는 창발론적 반박을 연결주의적 계산이론을 통해 봉쇄시킬 수 있다는 점에서 큰 장점을 지닌다. 그러나 무엇보다도 계산주의의 가장 큰 매력은 계산주의 이론이 지닌 지극한 단순성에 있다. 계산주의의 바탕을 이루고 있는 요소는 단지 '1'과 '0'로 표현되는 계수(digit)이다. 더 이상 단순할 수 없는 이 원초적 요소로서 복잡한 심적 현상을 설명하려고 하는 점에 계산주의적 인지이론이 지닌 간결함과 아름다움이 엿보인다. 단순한 이론은 복잡한 것에 비해서 상대적으로 무한한 변용이 가능하며, 확인가능성과 반증가능성을 동시에 지닌다는 점에서 연결주의적 인지이론보다 진리에 가깝게 접근해 있다고 볼 수 있다.

## 참 고 문 헌

- 공용현. (1985). “적 이론의 단순성에 대한 연구” 서강대학교 석사학위논문.  
 공용현. (1989). “마음, 지능, 인공지능 - AI의 정의에 관한 철학적 반성” 한국인지과학회 편. 『인지과학』 1권 2호, pp.175-192.  
 공용현. (1991). “인간-컴퓨터 상호작용에 있어서의 선험적 문제들”

*Proceedings of HCI '91 Winter Workshop.*  
HCIL/CAIR/KAIST.

- 공용현. (1992). “계산주의의 철학적 함축” 서강대학교 박사학위논문.
- 박연숙. (1993). “계산주의의 한계와 연결주의” 한국과학기술원 인공지능연구원  
HCI연구실 편, 『HCI 기초이론』 제2집, pp.45-58.
- 소홍렬. (1994). “연결주의와 원자주의” 『정보기술』 1994년 3월호.
- 최종덕. (1994). “생물학주의에서 본 연결주의” 한국정보과학회 인간과 컴퓨터상  
호작용 연구회 편, 『HCI '94 학술대회 발표논문집』, pp.49-57.
- 최 훈. (1993). “라이벌이나 구현이나” 한국과학기술원 인공지능연구원 HCI연  
구실 편, 『HCI 기초이론』 제2집, pp.29-44.
- Bechtel, W. & Abrahamsen, A. (1991). *Connectionism and the Mind.*  
Cambridge, MA: Basil Blackwell.
- Casti, John L. (1989). *Paradigms Lost.* New York: Avon Books.
- Cole, David J. (1990). “Cognitive Inquiry and the Philosophy of Mind”  
in Cole, Fetzer, & Rankin eds. *Philosophy, Mind and  
Cognitive Inquiry.* Dordrecht: Kluwer Academic Pub.
- Churchland, Patricia S. (1986). *Neurophilosophy: Toward a Unified  
Science of the Mind-Brain.* Cambridge, MA: MIT Press.
- Churchland, P. S. & Sejnowski, T. J. (1989). “Neural Representation  
and Neural Computation” in W. Lycan. ed. (1990). *Mind  
and Cognition.* Oxford: Basil & Blackwell.
- Davis, Philip J. & Hersh, Reuben. (1986). *Descartes' Dream: The  
World according to Mathematics.* San Diego: Harcourt  
Brace Javanovich.
- Dretske, Fred. (1982). *Knowledge & the Flow of Information.*  
Cambridge, MA: MIT Press.
- Hopfield, J. (1982). “Neural Networks and Physical Systems with  
Emergent Computational Abilities” in *Proceedings of the  
National Academy of Science*, 79.
- Jackendoff, Ray. (1989). *Consciousness and the Computational Mind.*  
Cambridge, MA: MIT Press.
- Marr, David. (1982). *Vision.* San Francisco: Freeman.
- Neisser, Ulric. (1966). *Cognitive Psychology.* New York: Appleton

Century.

- Newell, Alan. (1980). "Physical Symbol Systems," *Cognitive Science* 4, pp.135-83
- Newell, A. & Simon, H. (1976). "Computer Science as Empirical Inquiry," M. Boden, ed. (1990). *The Philosophy of Artificial Intelligence*. Oxford: Oxford Univ. Press.
- Newell, Alan. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard Univ. Press.
- Pylyshin, Zenon. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, MA: MIT Press.
- Ramsey, W., Stich, S. & Garon, J. (1990). "Connectionism, Eliminativism, and the Future of Folk Psychology" in Cole, Fetzer, & Rankin eds. *Philosophy, Mind and Cognitive Inquiry*. Dordrecht: Kluwer Academic Pub.
- Rohrlich, F. & Hardin, L. (1983). "Established Theories" in *Philosophy of Science*, vol. 50.
- Schwartz, Jacob. T. (1988) "The New Connectionism" in S. R. Graubard ed, *The Artificial Intelligence Debate*. Cambridge, MA: MIT Press.
- Shannon, C. & Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana: Univ. of Illinois Press.
- Slovan, Aaron. (1990). "Motives, Mechanisms, and Emotions" in M. Boden ed., *The Philosophy of Artificial Intelligence*. Oxford: Oxford Univ. Press.
- Smith, J. C. ed. (1990). *Historical Foundations of Cognitive Science* Dordrecht: Kluwer Academic Publishers.
- Smolensky, Paul. (1990). "On the Proper Treatment of Connectionism" in Cole, Fetzer, & Rankin eds. *Philosophy, Mind and Cognitive Inquiry*. Dordrecht: Kluwer Academic Pub.
- Stillings, N. & Feinstein, M. (1989). *Cognitive Science: An Introduction*. Cambridge, MA: MIT Press.