

정신분석에 기반한 Cognitive Mental Algorithm: 차세대 지능로봇의 Mental Model 설계를 위한 이론적 배경

A Cognitive Mental Algorithm based on Psychoanalysis Theory: Theoretical study for design the mental model of a next intelligent robot

박 경 숙¹ · 권 동 수²

Kyung-Sook Park¹ · Dong-Soo Kwon²

Abstract This paper presents a theoretical study for making intelligent robots with human-like mind. For the development of a cognitive mental model, we developed three algorithms based on the cognitive process for human psychoanalysis. Specifically, the concept of id, ego and superego from the theory of Sigmund Freud was adopted and the procedural algorithms were presented.

Keywords: Cognitive Robotics, Mental Model, Psychoanalysis Theory, Consciousness, Unconsciousness

1. 서 론

우리가 로봇에게 지능과 별개로 마음(mind)이라는 것을 주고자 한다면 우리는 무엇을 참조해야 하는가? Behavior-based robotics 분야에서 로봇의 움직임을 표현하기 위해 동물의 행태학을 모방하는 것처럼 로봇의 마음을 만들기 위해 동물의 마음을 모방할 수 있는가? 그러나 이것은 현재 불가하다. 동물이 마음을 가지고 있느냐 아니냐의 문제는 인지과학자, 신경과학자, 철학자들이 아직 결론을 내리지 못한 사항이고 설사 동물에게 마음이 있다 하더라도 현재의 기술로 그것을 알 수가 없기 때문이다. 그래서 현 상태에서 우리는 인간의 마음을 관찰해서 로봇에 적용할 실마리를 찾아야 한다.

인간의 본성에 관한 많은 이론들을 통해 우리는 인간의 마음에 대한 간략한 sketch를 할 수 있게 되는데 그것을 기반으로 기본적인 framework을 만들 수 있다. 그 framework 중 어디를 강조하는지 무슨 색을 칠하는지

에 따라 여러 형태의 변형된 mental structure가 구현될 수 있고 로봇 연구자들의 관점에 따라서도 응용방법이 다양하게 존재할 수 있다.

단 여기서 “로봇의 마음을 인간과 유사하게 만들 수가 있는가? 없는가?” 라는 질문은 인공지능에서 야기되었던 해묵은 논란을 또다시 가져올 것이므로 논외로 하겠다. 단지 우리는 이 논문에서 음성, 비전, 제스처, 대화인식, 상황이해, 추론기법 등 기존의 인공지능적인 방법으로 연구가 진행되고 있는 현재의 지능로봇에 기술 극치가 발생했을 때, 그 이후 염두에 둘 차세대 지능로봇을 인지기능을 가진 인지로봇(cognitive robot)이라고 가정하고 그 인지로봇의 인공 마음 개발에 관한 기초 이론 연구를 하고자 하는 것이다.

이 논문에서 우리는 향후 인지로봇의 인공 마음에 사용될 수 있을 알고리즘들을 작성하기 위해 프로이트 정신분석 기법을 쓰고 있다. 프로이트 정신분석은 여러 가지 단점¹⁾들이 지적되고 있으나 인간 정신구조를 가장 잘 묘사한 이론으로 알려져 있으므로²⁾ 그의 이론을

※ 이 논문은 산업자원부 지원으로 수행하는 21세기 프론티어 연구개발사업(인간기능 생활지원 지능로봇 기술개발사업)의 일환으로 수행되었습니다.

¹⁾ KAIST 기계공학과 초빙교수

²⁾ KAIST 기계공학과 교수

¹⁾ 의식을 말로 표현할 수 있는 것과 동일한 점, 종교적 경험을 신경증(neurosis)으로 귀착하였다는 것, 대부분의 행동 동기를 성욕에 기초한 병원학적 이론으로 분석하는 점

이용하면 인간정신구조를 상세한 알고리즘으로 나타낼 수 있고 우리는 그것을 로봇이나 인공 생명체 등의 마음을 개발할 때 사용할 수 있을 것이다.

2장에서는 최근 대두되고 있는 인지로봇의 연구방향과 인지로봇에 응용할 수 있는 기존 의식 연구 및 프로이트 이론 중 의식과 무의식에 관한 기본 개념을 설명하고, 3장은 2.3에서 설명할 의식과 무의식을 이루는 이드(id) 자아(ego) 초자아(superego)에 대해 개략적인 지도(map)를 보여준다. 4장은 마음 내부에서 발생하는 에너지 흐름을 조사하여 이드 자아 초자아의 에너지 변화 알고리즘을 제시한다.

2. 차세대 지능로봇을 위한 기반 연구들

2.1 Cognitive Robotics 연구

제한된 도메인 내에서 간략한 인지기능을 보여주는 연구를 일부 공학자들이 시작하고 있고^[34] 인지제어라는 분야는 앞으로도 지능 로봇의 사고 과정을 나타내는 기능으로 발전할 것이라 생각할 수 있다.

또한 Sara Kiesler와 Hiroyasu Miwa등이 로봇의 mental model 설계와 그 mental 내부에서의 감정 표현 등에 대해서는 연구를 지속하고 있으나^[56] 아직 mental model 전반에 대한 제시는 하지 못하고 있다. Kazuhiko Kawamura는 인지적 타스크 수행에서 출발하여 인지로봇(cognitive robot)이라는 새로운 개념을 제시하고 있다^[7,8] 인지로봇(cognitive robot)이라는 용어가 아직 정확히 정의된 단계는 아니지만 인지공학과 인간공학에 기반을 둔 연구자들을 중심으로 최근 본격적으로 인지적 제어 연구가 시작되고 있으므로 인지적 타스크를 수행하는 로봇 등을 포함하는 상위개념으로 인지로봇을 생각 한다면 그 인지로봇은 특정 도메인에서 타스크를 수행 하는 수준을 뛰어넘는 인지기능을 지닌 mental 구조를 가지고 있어야 한다고 본다. 따라서 특정한 도메인에 한정되지 않는 포괄적인 인지적 mental 구조를 생성하는 연구를 수행할 필요가 있다.

일반적으로 우리가 로봇의 행위체계를 구현할 때 그림1과 같이 외부자극이 인식되고 그에 따라 정서를 동반한 행위가 발생하는 모델을 일반적으로 많이 구성한다. 그러나 인지로봇이 진화해 나가면서 Kawamura 가 제안하는 그림2의 형태를 예상할 수 있다. 즉 단순한 반사장치(Reactive)가 아닌 사고(Deliberative)의 기능을 가진 것처럼 보여주다가 자의식(Self-awareness)을 지니고 있는 것처럼 행동할 수 있고 나아가 의식(consciousness)을 가진 것 처럼 사용자가 느낄 수 있는

Stimulus-Cognition-Behavior-Emotion

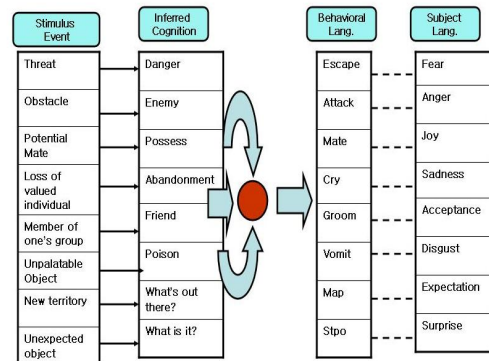


그림 1. 자극-인식-반응-행위 체계

그런 로봇의 진화를 기대할 수 있다^[7,8].

그러나 인간의 마음과 비슷하게 작용하는 마음을 가지는 인지로봇을 제작하고자 한다면, 복잡한 인간의 마음 메커니즘들을 인지로봇 내부에 삽입하여야 할 것이다. 인간의 마음을 움직이게 하는 요소가 의식뿐만 아니라 무의식의 영향도 강하게 받고 있으므로 우리는 인공적인 마음의 모델을 설계할 때 의식뿐 아니라 무의식 기능도 삽입하여야 한다. 그러나 아직 무의식을 로봇에 적용하려는 시도를 하는 연구자는 별로 없다.

그림2에서 Kazuhiko Kawamura는 cognitive robotics를 인식(cognition) 스펙트럼에 따라 인식과 robotics를 연결하고 있으며 body-mind를 통합하는 문제를 로봇연구의 next grand challenge라고 선언하고 있다^[9]. 그러나 로봇의 의식(consciousness)을 자의식(self-consciousness)으로만 설명하는 그의 주장은 심리학에서 다루는 의식은 범위가 더 포괄적이다 라는 점에서 조금 더 확대되어야 할 부분이다. 또한 그는 의식의 다음

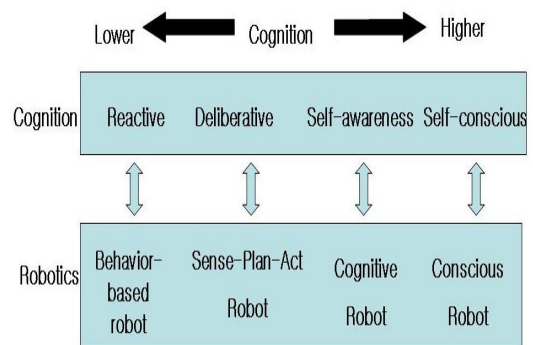


그림 2. 로봇 인식의 스펙트럼(출처:Kawamura 2006 ^[9])

단계인 무의식의 문제는 전혀 언급하지 않고 있으나 앞서도 설명했듯이 우리는 무의식 과정도 인공지능에 포함하여야 한다고 본다.

그런데 현재 로봇 mental design의 일부 중에는 부분적인 무의식 과정을 설명 없이 삽입하고 있는 경우는 있다. 예를 들어 Kawamura는 인지로봇을 구성하는 세 종류의 서브 시스템인 Reflective Process, Deliberative Process, Reactive Process가 장단기 기억장치, 행위체계, 인식체계, 주의집중, 정서발화 기능 등이 복잡하게 얽혀 작업하고 있다고 말하고 있고, 무의식에 대해서는 언급하고 있지 않으나 그가 말하는 Reflective Process는 의식과정이 반복된 후, 자동 처리 가능한 상태로 변화된 무의식으로 해석할 수 있다. 여기서의 무의식은 프로이트의 무의식을 전체 다 포함하는 것은 아니고 반복적인 행위를 통해 자동 학습된 것만을 말하는 것이다. 또한 Deliberative Process는 의도적으로 고려하는 의식의 일 부분으로 생각할 수 있고 Reactive Process는 본능(id)의 일부로 볼 수 있지만 절대적인 경계 나누기란 상당히 난해한 부분이다. 그가 아직 정확하게 분류하지 않고 있는 세 서브시스템을 프로이트가 상세히 설명한 의식과 무의식 즉 id, ego, superego로 연결할 수 있고 그러한 작업에 앞서 우리는 이 논문에서 id, ego, superego의 상세한 알고리즘을 설계하고자 한다. 그러한 알고리즘들은 그가 설명한 부분뿐 아니라 설명하지 않은 부분까지도 상세하게 표현할 수 있고 이것은 향후 인지로봇을 구현할 때 기본 mental 알고리즘으로 응용할 수 있게 될 것이다.

2.2 의식의 연구

그림2에서 Kawamura가 인지로봇의 마지막 단계로 표현한 의식(consciousness)에 대한 연구는 E. B. Titchener를 중심으로 한 코넬대학교의 연구 집단이 선봉이 되어 19세기에 발전되었다. 최초에는 내성법(introspection)을 통해 의식의 증거를 찾으려 했는데 이것은 관찰자가 자신의 개인적 의식 내용들을 분석하고 이것을 다른 사람과 비교하는 방법이였으나 관찰자의 개인차, 개인의 언어적 보고 시 오는 어려움 등을 고려하지 못한 오류로 인해 결과의 정확성에 회의를 갖게 하였다. 그리하여 내성으로 알지 못하는 것을 알 수 있다고 주장하는 행동주의(Behaviorism)가 득세를 하였고 학습과 행동의 동기유발 분야에 많은 성과를 보였다. 그러나 행동주의에서도 내성법과 유사한 문제점이 드러나기 시작했고 논리적인 대답을 구할 수 없는 문제는 문제가 되는 것조차 거부하는 논리적 실증주의가 학계에 퍼지면서 의

식은 수 년 동안 연구에서 무시되었고 의식 그 자체의 존재까지도 부인하는 사태에 이르렀다.

역시 행동주의에 불만을 가진 비엔나학파는 프로이트(Sigmund Freud)에 의해 발달된 정신분석 (Psychoanalysis)으로 몰리게 되었는데, 정신분석의 공헌은 인간 마음에 대한 심층적인 조사를 한 점과 ‘인간’을 이해하려는 학자들에게 좋은 연구 기반을 준 것 등으로 대표할 수 있다. 그 이후 60년대 전반에 나타난 인본주의 심리학 (humanistic psychology)에서는 행동주의와 정신분석 모두에서 왜곡되었던 것을 수정하려 했는데 인간적 ‘성장’, 인간관계의 진행과정(interpersonal processes) 등에 초점을 맞추어 연구했으나 과학적 유용성을 저평가한 이유로 크게 부각되지는 못했고 그 후 의식을 연구하는 학자들은 자신이 인본주의인지 과학주의인지를 밝히는 질문을 받아야 하는 분위기가 형성되었고 지금까지도 그런 질문은 계속되고 있다.

최근의 의식현상 연구는 크게 세 방향으로 나누어지고 있다. 첫 번째 물질과 정신을 분리시켜 보는 이원론 주장자들은 분자 생물학적인 방향으로서 뇌를 세포로 해석하고, 세포 보다 아래 단계의 분자론적인 입장에서 신경전달 물질을 이해하며, 생물물리학적 성질을 고려한 유전자 조작을 통한 뇌 연구를 하면서 정신적인 차원을 배제하는 입장이다. 에클스(John Eccles), 펜필드(Wilder Penfield), 셰링턴(Charle Sherrington)등이 대표적 연구자로 환원주의적 패러다임으로 신경과학(neuroscience)을 연구한 후에 이원론을 확신하게 된 경우이다. 그 외 포퍼(Karl Popper)는 다원론적 입장에서 세계를 1,2,3으로 나누어 이원론보다는 새로운 현실적 접근인 다원론적 시도를 하고 있다^{10,11}.

두 번째 이원론자들은 정신은 어떤 형태로든 일종의 뇌의 상태이라는 주장을 하며 여전히 다원론과 이원론을 공격하는 환원주의 입장에서 정신적인 것을 물질적인 것(생물적인 것, 물리적인 것)으로 설명하려 한다. 펜로즈(Roger Penrose), 크릭(Francis Crick)과 코흐(Robert Koch), 피슈바흐(Fishbach), 포조(Poggio), 프리먼(Freeman), 포터(Fodor)등을 선두주자라 볼 수 있다^{12,13}.

세 번째 그룹인 통합적으로 바라보는 연구자들은 이원론을 탈피하여 정신적인 것이 신경회로나 신경 과정을 통해서 나타난 것이지만 다른 관점으로 본다면 그것은 독립적인, 환원될 수 없는 그 자체로 동일성을 갖는 과정이라는 관점을 지닌 학자들이다. 어쩌면 앞으로 올바른 방향이 될지 모르는 이러한 제3의 연구를 선두하는 대표적 학자로는 스페리(Roger Sperry), 데넷(Denett),

에델만(Edelman), 맥긴(McGinn), 찰머스 (Chalmers)등을 들 수 있다^{14,15)}.

2.3. 프로이트의 의식과 무의식

프로이트로부터 시작된 정신분석은 인간의 정신현상을 상세히 이해할 수 있는 이론적 바탕을 제공해주며, 의식뿐만 아니라 무의식도 함께 연구하고 변화시키는 방법으로 탁월한 기능을 발휘한다¹⁶⁾. 물론 이상상태의 정신을 분석하는 것이 원래의 정신분석 역할이라 정상적인 마음의 모델보다는 극적인 현상만을 강조하고 있어 다소 문제가 있기는 하지만 정신의 상태라는 것이 양적인 차이만 있을 뿐, 질적 차이는 별로 없는 것으로 알려져 있으므로 정신분석을 사용하여 정신병리의 구조를 분석하면 정상적인 마음의 상태도 유추할 수 있다^{16,17)}.

프로이트는 지형학적 이론 (Topographical theory)이라 불리는 초기의 이론에서 인간의 마음을 무의식 (unconsciousness), 전의식(preconsciousness), 의식(consciousness)으로 분류하였다. 그러다가 후기로 가면서 프로이트는 이드(id) 자아(ego) 초자아(superego)라는 용어를 사용하여 정신구조를 세분한 구조적 이론(Structural theory)을 발표하였다. 이 두 가지 이론과 함께 인간의 심리 현상이 우연한 것이 아니라 주로 과거 경험에 의한 무의식적인 동기가 있다는 정신적 결정론(Psychic determinism)과 최소한의 노력을 들여서 최대의 효과, 즉 불쾌를 최소로 피하고 쾌감을 최대로 하는 쪽으로 심리 작용을 한다는 정신적 경제주의(Psychic economism), 소아기의 발달이 성숙에 중요하다는 Genetic hypothesis이론 등을 프로이트는 주장하였다¹⁸⁾. 프로이트가 설명한 인간의 정신 구조는 그림3과 같은데 의식과 전의식 무의식에 대해 먼저 설명하고 id, ego, superego의 기능은 다음 장에서 설명한다.

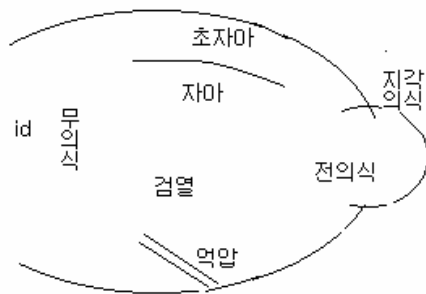


그림 3. 프로이트의 정신모델(출처 : Freud1933:[18] p.78)

2.3.1 의식(consciousness)

의식이란 우리가 사고하는 부분으로 인식하는 모든 것의 총량으로 생각할 수 있다. 프로이트는 의식을 개인의 정신활동 전체에서 조그맣고 순간적인 일부분으로만 여겼다. 의식은 주어진 순간순간마다 우리가 인식하는 모든 것들을 즉각적으로 또 끊임없이 반영하면서 변화하는 부분이다. 이때 인식의 범위는 인간의 능력에 의해 제한되며 인식의 원료는 외부세계에서 받아들이는 감각 정보의 총량과 과거로부터 기억해내는 정보에 달려있다. 우리의 관심이 과거의 기억으로 향할수록 환경에서 일어나는 현재의 변화를 인식하지 못하게 된다. 반대로 어떤 행위를 하려는 찰나에 우리는 다른 때라면 기억했을 어떤 것을 기억해내지 못하는 경우도 있다. 그러나 아무리 조그만 외부 자극이라도 과거로부터 기억할 수 있는 어떤 것과 연관되면 의식에 의미를 줄 수 있다¹⁹⁾.

2.3.2. 전의식(preconsciousness)

전의식은 평소에는 생각하지 않지만 원할 때 언제든지 생각해낼 수 있는 부분으로 해석한다. 즉 아침 식사 메뉴가 어떤 것이었는지 기억해내는 것처럼 우리가 조금만 신경을 쓰면 기억할 수 있는 것, 자발적인 회상으로 접근할 수 있는 ‘기억의 저장소’이다. 전의식은 우리가 알지 못하는 사이에 우리 행위에 영향을 미치는 모든 원시적 충동들을 무의식 영역에 집어넣는 역할을 한다. 또한 예전에는 의식 속에 존재하고 있었던 어떤 사건이 그것을 생각할 때 심각한 정서적 부담을 느끼게 되는 경우 그것을 억압시켜 무의식에 집어넣어 버린다²⁰⁾.

2.3.3. 무의식(unconsciousness)

인간이 알지 못하는 정신의 부분을 무의식으로 설명한다. 억압된 모든 것은 필연적으로 무의식 속에 남아 있으나 무의식 전체가 억압된 내용들로 구성되는 것은 아니다. 무의식은 단순히 잠재적이고 일시적으로 무의식적이지만 어떤 때에는 의식적인 행동과 전혀 구분되지 않는 행동들을 포함한다^{19,20)}.

3. 프로이트 마음의 상세구조

성격을 이루는 요소인 이드(id), 자아(ego), 초자아(superego)는 프로이트의 후기이론으로 최근까지도 사용되고 있다. 우리는 이 세 종류의 성격체계 이론에 따라 인간 마음을 이루는 요소를 분석, 각각의 지도를 작성하고 그것을 통합하여 하나의 ‘마음의 구조’를 보여준다.

3.1. 이드(id)의 발달 과정에 의한 마음의 지도

심리학적 영역 가운데 가장 오래된 이드(id)는 인류의 조상으로부터 물려받은 것으로 체질이나 신체 속에 각 인되어 있는 본능을 말한다. 본능은 이드를 통해 우리에게 미지의 형태일 수밖에 없는 심리적 표현을 드러내기 시작한다. 이러한 이드는 신체 내부에서 발생한 긴장과 외부에서 들어온 자극에 의해 활동하기 시작하는데 긴장이 고조되면 불쾌감을 느끼게 되고 이러한 불쾌감을 없애기 위해 다량의 흥분 에너지나 긴장을 표출시키는 작업을 시도하게 된다. 쾌락원칙 (Pleasure Principle)이라고 불리는 이 행위는 유기체를 항상 일정하게 유지시키려는 항상성 원칙의 한 종류이다.

이드는 최초로 모든 감각적 흥분을 반사작용에 의해 직접 표출시켜 버리려 하지만, 일부는 표출되지 않고 욕구불만으로 남게 된다. 이러한 욕구불만과 불쾌감으로 인해 이드는 발달하게 되는데 이것을 제 1 차 과정 (Primary Process)이라고 부른다^[21]. 유기체에 일어난 모든 긴장이 반사행동으로 배출된다면 원시적인 반사장치 이상으로 어떠한 정신적 발달도 필요하지 않을 것이다. 그러나 사실은 이와 달라서 반사에 의한 표출 외에도 긴장은 많이 일어나고 있다.

이러한 제 1 차 과정으로 인하여 cognitive 기능을 가지는 로봇의 구현이 가능해 질 수 있다. 욕구를 발산시키고도 남은 에너지 즉 채워지지 않는 욕구는 지각체계를 자극하게 되고 지각체계는 흥분을 받아서 감각체계에 나타나 있는 대상을 마음속에 ‘상상’ 하게 된다. 또한 기억체계를 자극하여 ‘기억’ 되어 있는 것을 생각해 내려고 하거나 상상의 결과인 심상들을 기억의 흔적으로 기억기능 속에 보존한다. 로봇 역시 욕구를 곧바로 충족하지 못하고 불만이 발생하면 그것을 해소하기 위해 상상력, 기억장치 등을 동원할 수 있게 만들 수 있다.

긴장감소를 요구하는 대상에 대한 기억 심상을 산출하는 과정이 앞에서 말한 제 1 차 과정이고 이것은 프로이트가 ‘지각의 동일화’ 라고 부른 것으로, 이드가 기억 심상을 지각 자체와 동일한 것으로 생각하여 긴장을 해소하려는 작업을 말한다. 이드가 긴장을 표출하는 두 과정은 ‘반사와 소망’ 인데 이것은 충동적인 근육운동과 심상형성 소원충족으로 나타난다. 우리는 사람이 충동적인 행위를 할 때 언제나 활동하고 있는 이드를 볼 수 있다. 즉 이드는 생각하지 않는다. 다만 소원하거나 행동할 뿐이다^[22]. 이상과 같은 인간의 이드의 과정을 간략하게 그리면 그림 4 와 같다^[1].

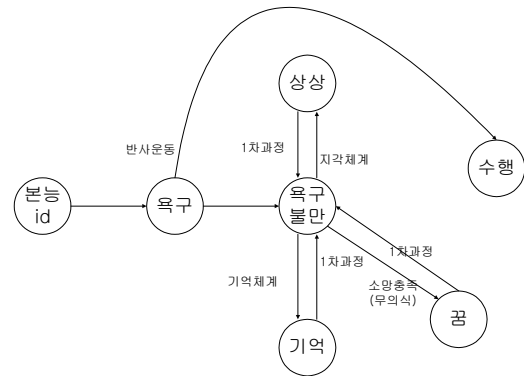


그림 4. 이드의 발달(출처: K. S. Park 2000:[1] p. 12)

3.2 자아(ego)의 발달과정에 의한 마음의 지도

이드는 소원충족이나 심상형성의 방법인 상상이나 꿈을 통해 긴장을 감소시키려 하지만 그것만으로 강렬한 욕구를 충족시킬 수 없다. 즉, 제 1 차 과정은 생존과 생식이라는 위대한 진화의 목적을 달성시켜 주지 못하므로 배우자나 음식을 구하는 등의 임무를 완수하기 위해 인간은 외부적 현실을 고려하여 현실에 순응하거나 현실을 자유롭게 지배하여 필요한 것을 세계로부터 얻으려 한다. 인간과 세계사이의 이런 거래는 새로운 정신적 기능인 자아(ego)의 형성을 요구하게 된다.

자아는 이드였던 한 부분이 주변 실제 세계의 영향을 받으며 특수한 방식으로 성장한 것이다. 원래는 자극을 받아들이는 기관과 자극을 막아내는 장치를 갖춘 피질(皮質)층이었던 것으로부터 특수한 조직이 생겨 나와 이드와 외부 세계 사이에서 매개 역할을 하게 된 것이고 이것을 자아라고 부르는 것이다.

자아는 현실원칙에 의해 행동을 연기하는데 이것은 긴장이 적절한 행동 형태로 표출될 때까지 자아가 그 긴장에 견디어 나갈 수 있음을 뜻한다. 프로이트가 제 2 차 과정이라고 부른^[21] 이 행동연기에서 행하는 첫 번째 일은 이드가 욕구불만을 일으킨 그 대상을 발견해 내거나 만들어내려는 시도이다. 제 2 차 과정 즉 문제해결의 사고는 사고력, 추리력, 인식력을 동원하여 전개된 행동 계획에 의해 현실을 발견하거나 만들어 내고 이때 하나의 행동 대안이 나온다. 그 후 이 행동 대안이 유효한지를 알기 위해 현실음미라는 과정을 거치는데 그 음미의 결과가 유효하지 않으면 즉 바라는 대상이 발견도 생산도 안될 때 다시 새로운 계획이 고안되어 대안변경이 일어나게 된다^[21].

여기서 주의할 것은 제 2 차 과정의 기능과 외부세계 참조에 의해 인간의 지각, 기억, 사고, 행동 등 정신적 과정의 발달이 완성되고 자극된다는 점이다²¹⁾.

자아에도 이드의 1 차 과정 비슷한 것이 있는데 이것은 공상(phantasy)과 백일몽(day-dreams) 이다²²⁾. 이것은 현실음미의 요구에 구속되지 않고 쾌락을 추구하는 긴장 완화의 과정이지만 자아의 이러한 기능은 이드의 제 1 차 과정과는 다른데 자아가 공상과 현실을 구별할 수 있기 때문이다. 이것을 통해 자아는 휴식한다. 자아는 대체적으로 환경과의 상호작용의 산물이지만 그 발달과정은 유전과 성숙에 의해 결정된다²²⁾. 자아의 발달을 간략히 표현하면 그림 5 와 같다²¹⁾.

3.3. 마음의 지도에 미치는 초자아(superego)의 영향

부모에 의존하여 성장하던 긴 어린 시절은 자아 속에 침전물과 같은 형태로 특수한 기능을 남겨 놓았는데 이를 통해 부모의 영향력이 계속 작용하게 된다. 이것을 초자아(superego)라고 부르는데 이것이 자아로부터 갈라져 나오거나 자아에 반대하게 될 경우 그것은 자아가 염두에 두어야 할 제 3 의 힘을 구성하게 된다. 초자아는 성격의 도덕적 판단기능을 하는데 현실이나 쾌락을 추구하는 것이 아니라 완전한 이상(자아이상)을 위해 일한다. 어릴 때 받은 상(보상)과 벌(처벌)은 육체적인 것이든 심리적인 것이든 내적 긴장을 감소시키거나 증가시키는 조건이 되어 각각 하위체계 즉 칭찬으로 형성된 자아이상과 처벌로 형성된 양심이 만들어지고 이것이 결합하여 초자아를 구성하는 것이다²²⁾.

개인의 가장 중요한 동일시 즉 부모와의 동일시에 뿌리를 두고 있는 자아이상은 부모가 선하다고 여기는 견해를 따르는 것이고, 도덕적 법전이 되는 양심

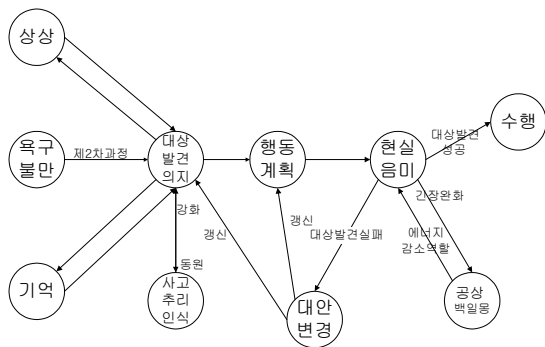


그림 5. 자아의 발달(출처: K. S. Park 2000:[1] p. 12)

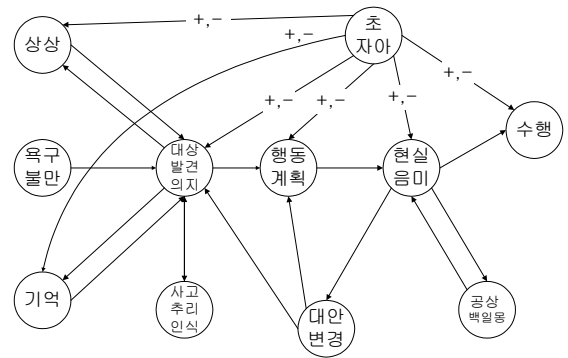


그림 6. 초자아의영향(출처 K. S. Park 2000:[1] p. 14)

이라는 것은 부모가 악하다고 생각하는 견해에 따르는 것이다. 따라서 부모 역할을 대신하는 초자아는 자아에게 상벌을 내려 자아가 지닌 긴장을 변화시킬 수 있게 된다. 자아는 행동뿐 아니라 생각만으로도 초자아의 처벌이나 상을 받게 되는데 이 상과 벌은 자아의 에너지에 영향을 주어 상의 경우는 에너지를 강화하고 벌의 경우는 에너지를 약화시키게 된다²²⁾. 그림6은 자아에 미치는 초자아의 영향을 나타낸 것으로 플러스기호(+)는 자아이상에 의한 에너지강화를 마이너스 기호(-)는 양심에 의한 에너지 약화를 나타낸 것이다²¹⁾.

3.4. 통합된 마음의 지도

이상과 같이 id(이드)는 진화의 원인이면서 생물학적 자질의 심리적 대표자로 상속받은 요소이고, 다른 사람에게서 영향을 받아 형성된 superego(초자아)는 사회화의 소산이고 문화적 전통의 전달 수단이다. 반면 주로 그

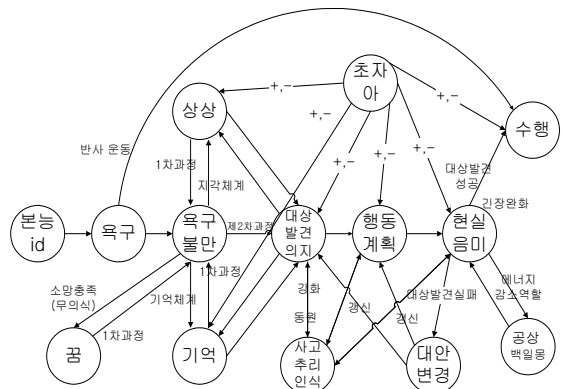


그림 7. 마음의 지도(출처:K. S. Park 2000:[1] p. 14)

개인이 직접 경험한 것으로 일시적이고 우연한 사건들에 의해 영향을 받는 ego(자아)는 개관적 현실 및 보다 높은 정신적 과정과의 상호작용 결과로 볼 수 있을 것이다. 이러한 세 종류의 체계에서 일어나는 상호 변화 과정은 그림 7 과 같이 종합할 수가 있다¹¹⁾. 통합된 이 ‘마음의 지도(mind map)’ 는 그림 1 에서 언급한 로봇의 자극-인식-반응-행위를 확대하여 기억, 사고, 계획, 자체 평가 등 인지적 제어 기능을 하는 인지로봇의 metal model 로 응용할 수 있다.

4. 심적 에너지 흐름

3장에서는 프로이트 이론에 기반한 마음의 지도를 보여주었다. 그런데 그런 마음의 지도 내에서 인간은 어떻게 의식하는 것일까? 의식의 흐름에 작용하는 무의식의 영향은 어떠한 형태일까? 이 문제에 대한 해답을 얻어야만 의식·무의식을 통합한 상세한 알고리즘이 작성된다.

따라서 이 장에서는 앞에서 보여준 마음의 지도 내에서 의식흐름을 움직이게 하는 힘을 심적 에너지라고 간주하고 마음의 활동성을 그 심적 에너지 흐름으로 분석, 상세한 알고리즘을 설명한다.

4.1. 이드(id) 에너지변화

인간은 복잡한 에너지 체계이며 그 에너지는 섭취하는 음식물로부터 들어와서 순환, 호흡, 소화, 신경전도(傳導), 근육활동, 자각, 기억, 사고 등을 위해 소비하고 있다. 인간의 신체를 유지시키는 육체적 에너지처럼 마음을 움직이는 힘을 심적 에너지라고 부르는데 이 에너지는 이드, 자아, 초자아를 움직여 마음의 변화를 일으키는 원동력이다.

4.1.1. 본능-심적 에너지 발생 근거

정신의 활동을 할 수 있게 하는 에너지의 전부는 본능에서 얻어 온다. 본능이 자리잡고 있는 곳이 이드(id)이므로 이드 에너지는 개별 유기체가 살아가는 삶의 진정한 목표를 표현하기 위해 심리적 과정에 방향을 부여하는 인간의 타고난 조건이라고 보면 될 것이다. 육체에 저장된 에너지를 방출하기 위해 신체의 조직이나 기관에서 일어나는 욕구와 충동이 이드 에너지의 주요한 원천인데 반사작용과 소망충족을 통해 이 에너지를 발산하게 된다. 반사작용에서는 에너지가 자동적으로 발산되어 운동 에너지로 변화되고 소망충족에서는 본능 대상의 심상 형성에 에너지를 사용한다.

대상의 심상에 에너지를 주거나 또는 본능을 만족시켜줄 대상물에 대한 행동을 위하여 에너지를 소비하는 것을 대상집중(cathexis) 또는 대상선택이라 하는데 이드 에너지 전부는 이 대상집중에 사용된다. 본능의 최종 목표는 신체적 욕구의 제거 즉 흥분이 있기 이전의 정지 상태로 되돌려보내는 것이므로 이드 에너지가 완전히 소멸하면 정신은 평안을 찾게 된다.

또한 이드 에너지는 한 대상에서 다른 대상으로 쉽게 이동될 수 있다는 특징이 있는데 이것을 전위(치환: displacement)라고 부른다. 특별히 어떤 비슷한 점이 있을 때 동일하게 취급하는 전위로 인해 대상들 사이에 차이가 있는데도 불구하고 동일한 것으로 취급하는 왜곡된 사고인 속성적 사고(술어사고: predicate thinking)가 만들어진다.

일정한 방향을 가진 본능 에너지의 흐름이 자아나 초자아에 의해 저지될 때 본능 에너지는 환상(공상)이나 행동에 의해 저항을 돌파하여 그 에너지를 방출하려고 한다. 이때 그 사람은 실언, 잘못 쓰기, 착각, 기억을 못하는 현상을 일으키거나, 문제해결 능력과 현실발견 능력이 충동적인 소망의 침입에 의해 감소되어 능력감소를 경험한다. 또한 혼란에 빠져서 현실과의 접촉을 하지 않으므로 사고를 당하기도 한다.

이드 에너지의 직접적인 배출구 발견이 실패하거나, 배출하고도 남은 에너지가 있으면 그 에너지는 자아와 초자아로 넘어가고 그 두 체계의 강화에 사용된다.

4.1.2. 본능의 종류와 변화

이드의 욕구가 야기한 긴장의 배후에 존재하는 힘인 본능의 종류는 여러 가지가 있지만 그 많은 본능들을 분류하면 두 가지의 근본적인 형태로 분류할 수 있다고 프로이트는 주장했다. 프로이트는 이것을 삶의 본능(Libido)과 죽음의 본능(Destrudo, Thanatos)으로 불렀는데 죽음의 본능에 대해서는 별로 알려지지 않고 있다. 드러나지 않게 작용한다는 점과 반드시 사명을 완수한다는 점만이 알려져 있는데 공격성과 파괴성이 죽음의 본능의 예라고 보면 될 것이다. 삶의 본능은 공개적으로 드러나므로 잘 알려져 있다. 이것은 모든 신체적 욕구의 정신적 대표자이므로 생존과 번식을 위해 발생하는 성욕과 식욕이 그 예이고 둘은 서로 연관되어 있다고 한다. 삶의 본능과 죽음의 본능은 그리고 그 파생물들은 서로 연합하기도 하고 중성화되기도 하며 서로 교체되기도 한다.

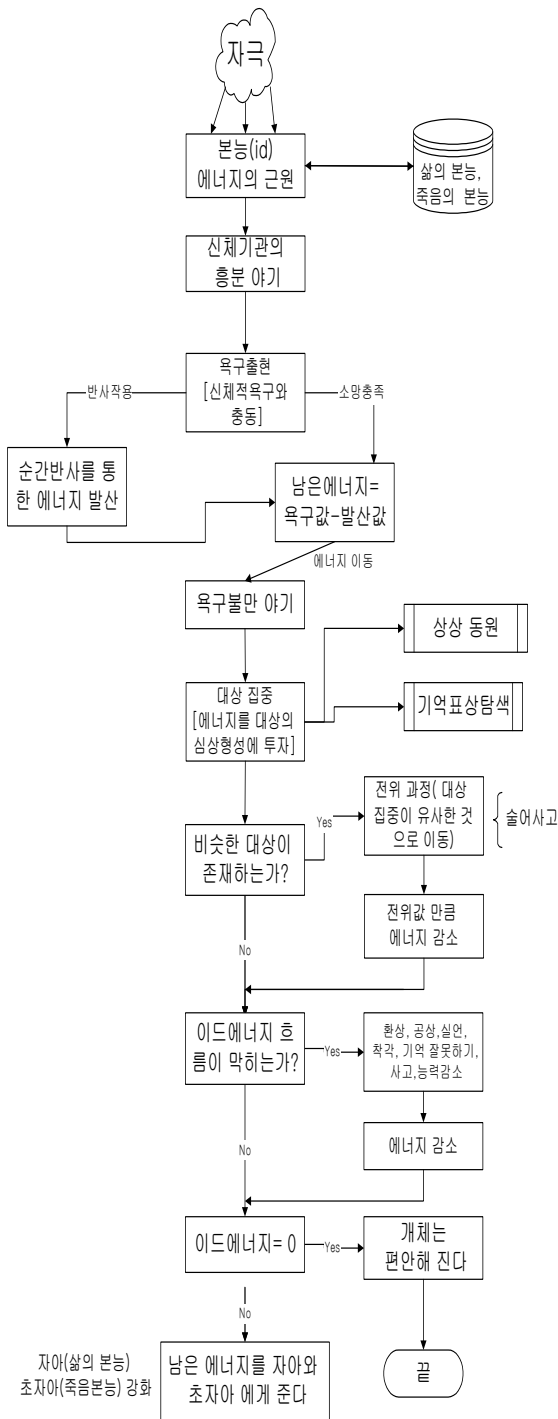


그림 8. 이드 에너지 변화과정

원래 본능은 이드에 속해 있지만 자아나 초자아의 과정을 인도하며 본질을 드러내기도 한다. 자아는 삶의 본능의 주요한 대행자인데 두 가지 방법으로 삶의 본능을 지속시킨다. 첫째, 자아는 환경과 현실적인 관계를 하며 기본적인 신체적 욕구를 만족시켜 간다.

두 번째 자아는 죽음의 본능을 죽음의 목적이 아닌 삶의 목적에 유리한 형태로 변형시켜 삶의 본능을 위해 일한다. 이에 반해 초자아를 죽음의 본능 대행자라고 하는데 그 이유는 삶의 본능 대행자인 자아를 억누르고 파괴하려고 하기 때문이다^{19,22)}. 그림 8은 이상에서 설명한 이드 에너지가 변화되는 과정을 순서대로 나타낸 것이다.

4.2. 자아(ego) 에너지와 그 변화과정

4.2.1. 자아 에너지 기본 형태

원래 자아 그 자체는 에너지를 조금도 갖고 있지 않다. 그러나 이드로부터 에너지가 넘어올 때 비로소 자아는 형성되기 시작한다. 자아가 최초로 하는 일은 심상물과 현실대상이 다르다는 인식과정이다. 이 분리의 결과로 순전히 주관적이던 이드의 내부세계가 주관적인 내부세계 마음과 객관적인 외부세계 환경으로 구분된다. 이러한 동일시(identification)로 인하여 전에는 이드에 의해 대상 자체에 쏟아지던 정신의 집중이 자아집중(ego-cathexis)으로 바뀌게 된다. 즉 동일시에 의하여 환상적인 소망실현의 1 차 과정이 현실적 사고인 2 차 과정으로 변하게 되는 것이다.

그러나 자아가 이드의 욕구를 만족시켜주지 못하는 경우에는 자아집중이 다시 본능적인 대상 집중으로 바뀌어 소아적인 소망충족이 다시 지배하게 된다. 이것은 꿈이나 현실에서 하는 자폐적 사고 혹은 소망적 사고인데 언제나 우리에게 혼란을 준다. 자아가 이드로부터 충분히 에너지를 끌어내면 자아는 이 에너지를 본능을 만족시키는 일이 아닌 다른 목적에 사용하게 된다. 즉 그 에너지는 지각, 주의, 학습, 기억, 판단, 구별, 추리, 상상 등의 정신적 발달에 사용되고 그러한 과정은 자아 에너지가 커지면서 더욱 정교해 진다²³⁾.

자아 에너지의 일부는 운동체계를 거쳐서 흥분이 노출되는 것을 방지하거나 연기하는데 사용된다. 이 지연의 목적은 자아로 하여금 행동으로 옮기기 전에 실제적인 행동 계획을 세우게 하는 것이다. 에너지가 마지막 방출의 방향으로 흘러가는 것을 방지하는 이 힘을 반집중(anti-cathexis)이라 하며 집중에 대립하는 에너지를 채우는 일이다.

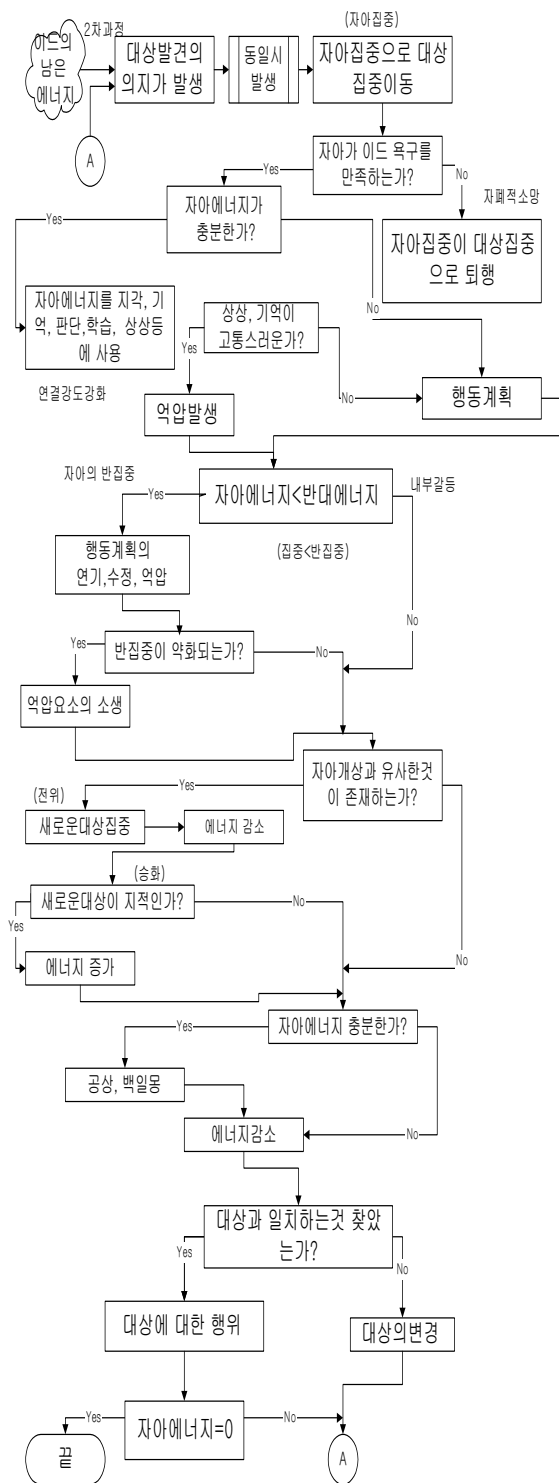


그림 9. 자아 에너지 변화과정

자아 에너지는 새로운 대상 집중에 형성하는데 사용되기도 한다. 그 대상들은 직접 유기체의 기본적인 욕구를 만족시키는 대상들과 연상의 고리로 연결되어 있지만 그것이 직접 유기체의 욕구를 만족시키지는 못한다. 마지막으로 자아는 대상과 일치하는 것을 찾게 되는데 그러한 작업이 성공하면 직접수행에 들어가지만 찾지 못했을 때는 대상변경의 의지가 발발하여 대상이 수정된다^[22]. 이상에서 설명한 자아에너지 변화에 대한 과정을 순서도로 나타내면 그림9와 같다.

4.3. 초자아(superego) 에너지의 영향

부모와의 동일시로 인해 만들어지는 초자아는 자아이상과 양심으로 구성되어 있다. 부모의 도덕 가치에 의해 만들어져 유기체의 완전을 위해 노력하는 자아이상은 이드가 상상하는 요소와 자아가 택한 대상선택이 자신과 일치되어 만족을 느끼게 될 때 유기체에게 자부심을 체험하게 한다.

양심에 의한 금지 작용은 충동적인 행동 또는 소망 실현에 직접 영향을 주거나 자아의 메커니즘에 간접적인 영향을 주어 본능 에너지의 발산을 막아내는 반집중 과정이다. 이때 죄책감과 수치심을 경험하게 된다^[22]. 그림 10은 초자아의 영향에 대한 순서도이다.

이 알고리즘들은 인간의 정신구조를 설명하는 것이므로 알고리즘 전체를 인지로봇의 마음으로 설계하는 것은 무리가 있을 것이다. 그러나 이드 알고리즘에서의 외부 자극은 각종 센서가 외부로부터 입력을 받아 들이는 것을 해석할 수 있고 내부 자극은 배터리의 전압이 일정한 값 이하로 떨어질 때의 저 전압 경보 등으로 해석될 수 있다. 우리가 이러한 알고리즘을 응용하면 자극에 의한 단순한 반응으로서의 로봇의 행위가 아닌, 외부 센서에 적의 출현이 감지된 것과 배터리가 최저로 떨어질 때 로봇은 그 둘 중 우선순위를 상황에 따라 다르게 판단할 수 있게 만들 수 있다. 그때 로봇은 사람이 고민할 때처럼 둘 중 우선순위를 랜덤이 아닌 어떤 판단에 의해 결정할 수 있게 되는 것이다. 자아 알고리즘의 행동 계획과 대안 선택의 기능은 기존 인공지능의 문제해결에서 많이 보이던 양상이지만 그 외의 2차 기능들은 인지로봇의 인지기능으로 설명될 수 있다. 주인의 심부름을 수행하는 로봇의 경우 그 심부름할 물건 즉 대상발견을 위해 자아가 집중하고 그것이 자신의 본능적 욕구(예를 들어 주인에게 충성하여 로봇 자신의 필요성과 가치를 강화시킨다)를 충족하는 단계로 구현될 수 있다.

초자아 알고리즘의 응용은 로봇이 서비스를 수행할 때마다 주인으로부터 받은 상과 벌이 학습되어 기억 속에 내재시키게 구현 가능하다. 그 기억들은 향후 유사한 행동을 재현할 때 행위 표준으로 사용될 수 있다. 즉 인간의 경우 부모와 사회로부터 형성된 초자아를 로봇의 경우는 주인과 사용자들의 영향을 받아 성격이 다르게 만들어지는 개성을 지닌 로봇으로 성장시킬 수 있다.

그러나 앞에서 제시한 세 종류의 알고리즘은 기본적인 마음의 frame 에 지나지 않는다. 마음이란 우리가

생각하는 것 보다 훨씬 복잡한 과정이고 그런 이유로 추가되어야 할 모듈이 많은 부분 남아 있다. 그리고 자아의 방어기제나 불안 등은 마음을 순식간에 변화시키는 요소로 이것들을 추가하여 통합된 알고리즘으로 확대할 필요가 있다. 이렇듯 많은 부분이 부족함에도 불구하고 우리가 이 작업을 시도한 것은 로봇 지능과 마음의 설계에 사용할 이론 연구가 미미하기 때문이다. 물론 인지과학자 철학자 일부가 앞으로 이러한 시도를 계속할 것이고 향후 우리는 human-like mind 를 탑재한 로봇의 출현을 볼 수 있을지도 모른다.

5. 결 론

우리는 때때로 무엇인가를 의식하고 있다고 생각한다. 그리고 의식이란 것에 대해 잘 알고 있다고 생각한다. 그러나 ‘어떻게 의식하는가? 내지 의식이란 무엇인가?’ 라는 질문을 받았을 때 우리는 비로소 우리가 어떤 방식으로 의식하는지 심지어 의식이 무엇인지조차 알지 못한다고 고백하고 만다. 이렇듯 의식현상의 실체에 대해서 잘 알 수 없음에도 불구하고 의식현상의 근본 원인을 밝혀려고 고민한 학자는 우리가 살펴본 바와 같이 적지 않았다.

그리고 그들의 연구를 기반으로 인간의 마음을 닮은 즉 인간처럼 의식과 무의식을 경험하는 차세대 지능 로봇 즉 인지로봇을 만들고자 하는 것이 로봇연구자의 희망이다. 그럼에도 불구하고 현재까지 인지로봇의 마음을 만들 수 있는 기본 알고리즘은 제시되지 않고 있다. 그것은 마음의 메커니즘이 너무 복잡하고 난해하며 마음이 어떤 식으로 흘러가는지에 대한 명확한 연구 결과가 없으며 로봇공학자가 기계적 기법으로 접근하기에는 마음의 요소에는 너무 기계적이지 않은 부분이 많다는 것 등을 그 이유로 볼 수 있다.

그렇다면 사람의 마음을 닮은 인지 로봇을 만들어 내는 것이 불가능한가? 이 문제에 대한 답변은 우리가 여기서 논할 성질의 것이 아니다. 단지 우리는 이 논문에서 프로이트 정신분석을 이용, 인간 정신의 심층구조를 절차적 알고리즘으로 제작하여 로봇 연구 필요 시 활용할 수 있게 하려는 것에 목표를 두고 있다.

그런데 프로이트 심리학은 1890년대 초기로부터 1930년대 말기에 걸친 그의 글들에 흠어져 있어서 전체 이론을 통합하여 알고리즘화 하는 작업은 쉽지 않다. 다만 그의 이론이 인간 마음에 대해 우리에게 보여주는 깊은 통찰은 정신분석이 인지로봇의 기초 이론 연구로 충분히 적합하다고 간주할 수 있으므로 우리는 그의 이

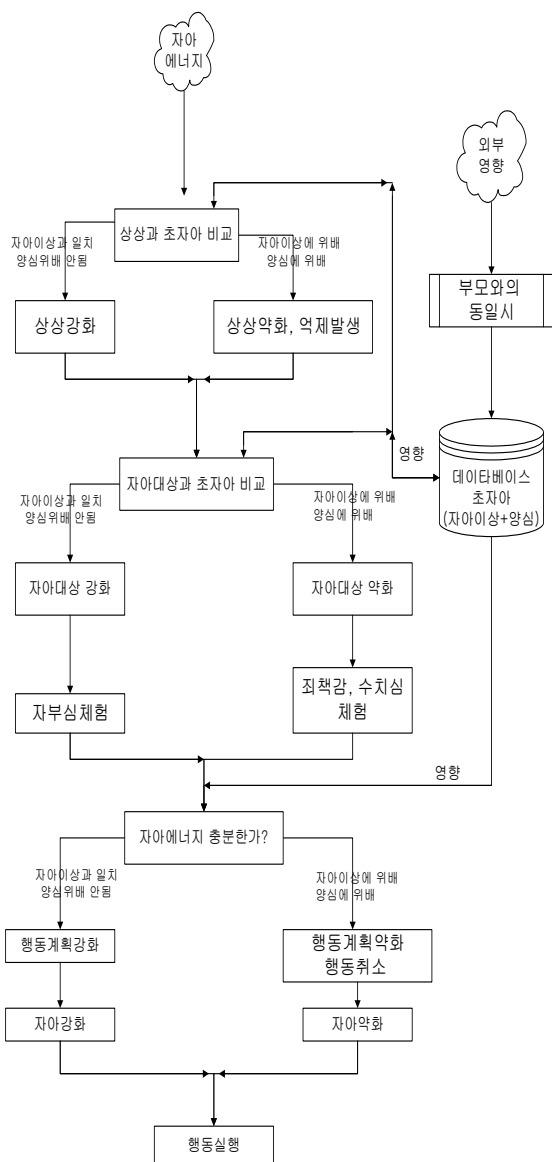


그림 10. 초자아 에너지 변화과정

론으로 출발하였다. 물론 연구가 진행되어 가면서 더 간결하고 더 훌륭한 이론들을 발견하면 알고리즘들은 교체되어야 마땅할 것이다.

이것을 출발로 우리는 인간 마음과 유사하게 본능과 의식, 무의식기능을 가진 로봇연구를 할 수 있고 인공 생명 등 또 다른 연구로도 진화될 수 있을 것임은 분명하다. 이 논문에서 우리는 심상에너지의 기본 알고리즘을 만드는 이론 연구까지만 제한하였으나 이 알고리즘들의 많은 부분이 보강될 것이고 그것들을 기반으로 각종 방어기제나 불안의 영향에 의해 심상 에너지가 변화되는 것을 나타내는 작업을 후속 진행할 예정이다. 그러나 여기서 제시된 알고리즘들을 로봇에 적용하는 것은 다음 연구주제로 남겨둘 것이다. 왜냐하면 앞에서도 언급했듯이 우리가 여기서 원하는 것은 정신분석학에 충실한 이론적 알고리즘을 만드는 것이고 이 알고리즘들을 기반으로 로봇에 응용하는 것은 아직 시기상조라 할 수도 있으며 적용 가능성이 아닌가의 논란도 불러올 수 있을 것이기 때문이다. 또한 철학과 인공지능에서 야기된 여러 문제들을 다시 상기시킬 수도 있을지 모른다.

물론 정신 분석학이 아닌 다른 심리 이론들로 mental model 을 설계해도 될 것이지만 필자는 mental model 에 의식뿐 아니라 무의식도 포함하기 위해 정신분석을 이용한 것임을 밝혀둔다. 앞으로 많은 로봇공학자, 인지과학자들이 마음의 연구 결과를 인지로봇에 응용하려고 할 것이고 현재 인간의 감정을 표현하는 메커니즘을 로봇의 표정에 적용한 감정상호작용로봇이 발표되고 있는 것처럼 조만간 인간 심리를 진정으로 이해하고 인간처럼 생각하고 느끼는 humanoid mind 를 탑재한 human-like 인지로봇이 나올 가능성을 기대할 수 있다.

참고문헌

- [1] Kyung-Sook Park, "Cognition Modeling System using the Intelligent Algorithm combined with Psychoanalysis and Emotional Theory" Ph. D. Thesis, Korea, Yonsei Univ., February. 2000.
- [2] 바니에 지음, 김연권 옮김, 정신분석의 기본원리, 솔출판사, 1999.
- [3] Palis Ratanaswasd, Stephen Gordon, "Cognitive control for robot task execution" IEEE International workshop on robots and Human Interactive Communication, pp. 440-445, 2005.
- [4] Arvin Agah, George A. Bekey. "Cognitive Architecture for Robust Adaptive Control of Robots in a team" Journal of Intelligent and Robotics systems, no. 20, pp. 251-273, 1997.
- [5] Horoyasu Miwa, "Study on the mental model for humanoid robots," Ph. D. Thesis, Japan, Waesda Univ., July 2004.
- [6] Sara Kiesler, "Human mental models of humanoid robots" International Conference on Robotics and Automation, Barcelona, SP. April. 2005
- [7] Kazuhiko Kawamura, S.M. Gordon, and P. Ratanaswasd, "Robotic Body-Mind Integration: Next Grand Challenge in Robotics", International Journal of Advanced Robotic Systems, in printed, 2006.
- [8] Kazuhiko Kawamura, "From Intelligent Control to Cognitive Control", 11th International Symposium on Robotics and Applications(ISORA) Robotics Track-WAC 2006, Budapest, Hungary, July, 2006..
- [9] Kazuhiko Kawamura, "Tutorial on Cognitive Robotics" IEEE International symposium on robots and Human Interactive Communication, Tutorial lecture note, London, UK, September, 2006.
- [10] 광형식. 인간심리의 비밀 뇌기능의 이해, 학문사, 1998.
- [11] John Eccles. 뇌의 진화, 민음사, 1998.
- [12] Roger Penrose. 황제의 새 마음, 이화여대 출판부, 1996.
- [13] Francis Crick, 놀라운 가설, 한빛출판, 1996.
- [14] Daniel C. Denett. 마음의 진화, 사이언스 북스, 2006.
- [15] Gerald Edelman, 신경과학과 마음의 세계, 범양사 출판부, 2006.
- [16] 네마이어(지음) 유범희(옮김). 정신병리학의 기초. 민음사. 1993.
- [17] 브레너(지음) 이근후 박영숙(옮김). 정신분석학. 하나의학사, 1985.
- [18] Sigmund Freud(1933), New Introductory Lectures on Psychoanalysis, Chap. 3. New York: W. W. Norton & Company, Inc., 1933
- [19] Sigmund Freud(1923), The Ego and the Id, London : The Hogarth Press, 1947.
- [20] Sigmund Freud(1915). "The Unconscious". In The Standard Edition of Complete Psychological Works of Sigmund Freud. Vol X IV. London: The Hogarth

Press, reprinted 1973.

- [21] Sigmund Freud(1913). The Interpretation of Dreams. In Translated and edited by A. A. Brill(ed), The Basic Writings of Sigmund Freud. 181-549. New York: Modern Library.
- [22] Hall, C. S. A Primer of Freudian Psychology. A Mentor Book. 1954.



박 경 속

1988, 2 중앙대학교 전자계산
(이학석사)
2000, 2 연세대학교 인지과학
(이학박사)

1988, 3 - 1995, 2 경남대학교, 진주교육대, 경상대학교 등 강사
1995, 3 - 2005, 8 해전대학 컴퓨터과 부교수
2005, 9 - 2007, 3 현재 한국과학기술원 기계공학과 초빙교수
관심분야 : Artificial Mental model, Cognitive imitation system, Human Robot Interaction



권 동 수

1982, 2 한국과학기술원 기계공학과 (공학석사)
1991, 8 Georgia Institute of Technology 기계공학과 (공학박사)

1991, 7 - 1995, 7 미국 Oak Ridge 국립 연구소 로보틱스 및 공정 시스템부 텔레로보틱스 과 선임연구원
1995, 8 - 2007, 3 현재 한국과학기술원 기계공학과교수
관심분야: Human Robot Interaction, Haptic Interaction, Medical Robotics