

## 종교적 인공지능 시뮬레이션

— 스타크-베인브릿지와 뒤르켐의 종교적 인지 모형 비교연구

박 옥 주\*

### [초 록]

오늘날 제기되고 있는 인공지능의 가능성과 한계에 대한 물음들 가운데는 인간의 종교성과 관련된 물음이 속해 있다. “과연 인공지능은 종교적 신앙을 재현할 수 있는가?” 본 연구는 종교사회학자이자 인지과학자인 윌리엄 심스 베인브릿지(William Sims Bainbridge)가 인간의 종교성과 초자연자에 대한 신앙을 시뮬레이션하기 위해 2006년경 이 행한 기계학습 실험을 주요 탐구사례로 선정한다. 그리고 해당 실험에 반영된 인간의 종교성과 신앙에 대한 사회인지 이론, 이를 기계학습 시뮬레이션으로 현실화하기 위해 기획된 시나리오, 실험에 활용된 일부 수학적-통계적 기법과 그 원리에 대한 종교학적 고찰과 반성을 수행하는 데 초점을 맞춘다. 이를 위해 본 연구는 스타크-베인브릿지와 뒤르켐의 종교론과 종교적 인지 모형을 비교해 고찰한다. 이를 통해 본 연

\* 연세대학교 연합신학대학원 겸임교수

주제어: 종교적 인공지능, 초자연자 의식, 사회의 실재성, 종교적 인지범주, 자기보존본능

Artificial Religious Intelligence, Consciousness of the Supernatural, Reality of Society, Categories of Religious Cognition, Conatus Essendi

구는 다음과 같은 결론을 내린다. 양측 간 다소 증대한 이견들이 존재 하긴 해도, 베인브리지의 인공지능 시뮬레이션에 반영된 스타크-베인 브릿지 종교론의 통찰들은 뒤르켐의 종교학 방법론 관점으로 볼 때 상당부분 타당성을 가진 것으로 인정될 가능성이 높다. 무엇보다 원초적 종교성의 발원지를 사회로부터, 지속적으로 수행되는 사회적 행위로부터 찾고, 이를 종교적 인지의 범주적 준거로 전환하는 방식을 통해 인간 간의 종교적 본성을 탐구한다는 점에서 양측은 공감대를 형성하고 있는 것으로 확인된다.

## 1. 들어가는 말

“기계가 사고를 수행할 수 있는가?”(Can machines think?) 이는 사이버네틱스(cybernetics, 인공두뇌학)와 인지과학(cognitive science)의 선구자 중 하나인 앨런 튜링(Alan M. Turing)이 1950년 발표한 자신의 연구논문 “컴퓨팅 기기와 지능”(Computing Machinery and Intelligence)<sup>1)</sup>의 첫머리에 제기한 물음이다. 여기서 그가 수행한 “튜링 기계”(Turing machine) 논의는 6년 후 개최된 다탘머스 회의(Dartmouth Conference)<sup>2)</sup>를 비롯한 초기 사이버네틱스 연구자들의 지능 규정 시도를 지배적으로 주도하였다. 이들은 컴퓨팅 기기 역시 인간과 동일한 수준의 사고를 수행할 수 있으리라는 계산주의적(computationalist) 신념,<sup>3)</sup> 그리고 “모방게임”(imi-

1) Alan M. Turing (1950), “Computing Machinery and Intelligence”, *Mind: A Quarterly Review of Psychology and Philosophy* 59, Mind Association, pp. 433-460.

2) ‘인공지능’이라는 용어를 창시한 존 매카시(John McCarthy)의 주도 하에 1956년 다탘머스 대학에서 개최된 학술대회로, 이 대회를 통해 인공지능에 관한 연구가 개별 학문 분과로 정립되었다는 점에서 중대한 역사적 의미를 갖는다. Jack Copeland (1993), *Artificial Intelligence: A Philosophical Introduction*, Oxford: Blackwell, pp. 8-9.

3) 사이버네틱스와 인지과학 분야에서 계산주의라고 할 때는 마음(mind)을 “특정한

tation game)<sup>4)</sup>의 완벽한 성사 가능성에 대한 굳건한 확신을 바탕으로 인공일반지능(artificial general intelligence) 개발에 역량을 집중시켰다. 그러나 초기 사이버네틱스 연구자들의 이런 낙관적 전망은 불과 10년이 채 지나기 전에 좌절되었다. 이들은 인간 정신의 작동원리가 ‘if-then’ 구문으로 대표되는 계산주의적 명제논리만으로 모방하기에는 과도하게 다층적이고 복잡하다는 것을 깨닫게 되었다.<sup>5)</sup> 이후 연결주의적(connectionist)

---

형식적 체계 혹은 체계들의 집합을 구현하는 것”으로 정의하는 견해를 말한다. 다시 말해 마음이 작동하는 방식은 기본적으로 “컴퓨테이션”(computation), 즉 수학적 논리에 의존하는 연산이라고 주장하는 입장을 계산주의라 한다. Matt Carter (2007), *Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence*, Edinburgh: Edinburgh University Press, p. 95.

- 4) 튜링 시험(Turing test)이라는 말로 명명되기도 한다. 인간이 상대방의 정체를 알지 못한 채 인공지능과 대화한 후 대화 상대방이 인공지능이었다는 것을 알아채지 못하면 시험을 통과한 것으로 판정한다. 튜링은 인간과 대화가 가능한 인공지능을 개발한다면, 즉 사고행위를 통해서만 수행 가능한 언어 활용을 인공지능을 통해 구현한다면, 이는 곧 인공지능이 자율적으로 사고할 수 있는 지적 존재자라는 것을 증명하는 것이라고 믿었다. 이는 다분히 “행동주의적인”(behaviorist) 입장에 속한다. James H. Moor (2003), *“The Turing Test: The Elusive Standard of Artificial Intelligence”*, Dordrecht: Springer Science+Business Media, p. 2, p. 5.
- 5) 인간 의식의 비컴퓨팅적(non-computational) 특성들을 유념하는 가운데 현재의 컴퓨팅 방식이나 인공지능 개발기술에 의존해서는 완벽한 강인공지능의 시대를 맞이할 수 없다는 견해는 앞서 소개한 설을 비롯해 힐러리 퍼트넘(Hillary Putnam), 폴 지프(Paul Ziff), 허버트 드레이퍼스(Hubert Dreyfus) 등에 의해 지지되고 있다. 퍼트넘은 강인공지능의 구현가능성에 대한 믿음이 인간의 의식을 순전히 기능주의적으로 바라보는 데서 나온, 과학주의에 기반을 둔 유토피아적 전망이라는 점을 지적한다. 지프는 인공지능이 인간이 생체적인 몸을 통해서 향유하는 “느낌들(feelings), 생각들, 그리고 성향들(attitudes)”을 가질 수 없음을 지적하며 강인공지능 구현에 대한 순진하고 유토피아적인 전망을 비판한다. 드레이퍼스는 인간의 의식이 갖는 현상학적이고 실존론적인 특성, 즉 단순 객관화할 수 없는 고유하고 개별적인 삶의 정황 속에서 경험되는 현상들을 명시적인 동시에 비명시적인 방식으로 이해하고 해석하는 특성을 들어 인공지능에 의한 인간 의식의 완벽한 구현이 용이하지 않으리라는 전망을 제시한다. Rajakishore Nath (2009), *“Philosophy of Artificial Intelligence: A Critique of the Mechanistic Theory of Mind”*, Boca Raton: Universal-Publishers, pp. 91-102.

지능이해에 힘입어 개발된 기계학습(machine learning) 기술은 지능에 대한 새로운 신경과학적 정의, 즉 인간의 지능이란 신경세포 간의 연결강도 변화를 통해 이루어지는 두뇌의 표상 및 판단작용이라는 정의를 참조한다.<sup>6)</sup>

이처럼 인공지능을 부분적으로 구현된 지능으로 정의하는 현실적 인식은 그에 뒤따르는 무수한 물음과 논쟁들을 야기해 왔다. 무엇보다 현재까지 개발된 인공지능은 어느 수준까지 인간의 사고행위를 대체할 수 있는지, 향후의 인공지능은 감정, 자기의식, 그리고 자유의지를 가질 수 있는지 여부가 중요 논제로 제기되어 왔다. 그런데 이렇게 인공지능의 현황과 한계, 그리고 이후의 가능성을 묻는 물음들 가운데는 인간 정신의 본질 가운데 하나로 지목되는 종교성<sup>7)</sup>과 관련된 물음이 속해 있다. “과연 인공지능은 ‘종교적 인간’(homo religiosus)<sup>8)</sup>을 재현할 수 있는

6) Paul M. Churchland (1993), “Representation and High-speed Computation in Neural Networks”, *The Foundations of Artificial Intelligence: A Sourcebook* (eds. Derek Partridge & Yorick Wilks), Cambridge: Cambridge University Press, p. 337.

7) 종교성이라는 용어의 의미를 간명하게 이해하려면 리처드 컴스탁(William Richard Comstock)의 설명을 참조하는 것이 바람직할 것으로 보인다. 그가 해설한 바에 의하면, 종교를 발생시키는 인간의 심성인 종교성은, 종교라는 용어의 정의만큼이나 그 의미가 다양하다. 그러나 현실의 다양한 종교현상들을 바탕으로 다소 광범위하게나마 정의해 본다면, 주로 유한자와 초월자의 구분, 속된 것과 성스러운 것의 구분, 자연적인 것과 초자연적인 것의 구분 등, 인간의 정신을 일상적인 것, 이미 인지하고 있는 것으로부터 경이롭고 비일상적인 것, 신비한 것으로 견인하는 경험들을 지향하는 마음이라고 볼 수 있다. 그리고 이처럼 다양한 방식으로 초월을 지향하는 마음은, 전부는 아니더라도 다수의 종교학자들과 신학자들 사이에서 인간 정신의 본질이자 필연적 정향점으로서 종교발생의 원동력으로 인정되어 왔다. William Richard Comstock (1971), *The Study of Religion and Primitive Religions*, New York: Harper & Row, pp. 18-27. 정재현은 종교성의 이런 원천적 성격을 유념하여 “원초적 종교성”이라는 말로 인간의 종교적 본성을 표현한다. 정재현(2019), 『우상과 신앙』, 파주: 한울아카데미, p. 157.

8) ‘종교적 인간’이란 비교종교학자 미르치아 엘리아데(Mircea Eliade)의 표현으로서, 인간이 상징과 이미지를 통해 원초적 종교성으로부터 유래된 신 의식을 다양한 신화 형태로 구술해내려는 본성을 갖는다는 주장을 함축한 말이다. Mircea Eliade

가?” 달리 말해, “과연 인공지능은 인간의 원초적 종교성을 재현할 수 있는가?” 이 논제는 그동안 종교학계 및 신학계로부터 다소간 경계심 어린 주목을 받아왔다. 특히 연결주의 사이버네틱스의 정수라 할 수 있는 딥러닝 기술이 2007년을 기점으로 급격한 발전을 보이면서 상기 물음은 트랜스휴머니즘(transhumanism) 논의와 긴밀하게 연계된 것으로 인식되기 시작한다.<sup>9)</sup>

본 연구 역시 기본적으로는 인공지능의 종교성에 관한 상기 물음을 바탕으로 종교학적 인간이해 논의를 전개한다. 그러나 그 현실화에 상당한 시간이 소요될 것으로 예상되는 트랜스휴머니즘과 연계된 연구와 부분적으로 궤를 달리하여, 이미 상당한 수준으로 현실화된 기계학습 기술과 종교성의 관계에 대한 고찰을 수행하려 한다. 이를 위해 본 연구는 종교 사회학자이자 인지과학자인 윌리엄 베인브릿지(William Sims Bainbridge)

---

(1987), *The Sacred and the Profane: The Nature of Religion* (tr. by Willard R. Trask), New York: Harcourt, Brace & World, pp. 163-165.

- 9) 인공지능의 종교성에 관한 물음은 원래 종교인지학(cognitive science of religion)의 연구들로부터 파생된 것으로 볼 수 있다. 종교인지학은 1975년 인지인류학자 댄 스퍼버(Dan Sperber)에 의해 그 분과적 개념이 형성되기 시작했고, 그 후 20세기 말까지 주로 종교학과 인류학 분야에서 스투어트 거스리(Stewart Guthrie), 하비 화이트하우스(Harvey Whitehouse), 파스칼 보이어(Pascal Boyer) 등을 통해 점진적인 발아기를 거쳤다. 당시 종교인지학은 종교성이 인간 인식능력의 진화과정에서 자연발생적으로 형성된 본성 가운데 하나로 간주했고, 원시 인류에게 하나의 보편적 특성으로서 주로 제의를 통해 발현되었고 초자연적 영역에 대한 개념적 범주들을 마련해 주었다고 설명했다. 21세기 들어 종교인지학은 점차 하나의 확고한 학제 융합적 학문분과로 자리잡고 있으며, 주로 이론화 과정에 머물러 있던 20세기와 달리 여러 방면의 현상적-실증적 근거입증 시도 단계에 진입해 있다. 베인브릿지의 종교적 인공지능 시뮬레이션은 이런 맥락에서 수행된 과업 가운데 하나로 인정된다. William W. McCorkle, Jr. & Dimitris Xygalatas (2014), “Introduction: Social Minds, Mental Cultures — Weaving Together Cognition and Culture in the Study of Religion”, *Mental Culture: Classical Social Theory and the Cognitive Science of Religion* (eds. Dimitris Xygalatas & William W. McCorkle, Jr.), New York: Routledge, pp. 1-4.

가 인간 종교성의 한 주요양태인 초자연자에 대한 정보의 추구와 산출 과정을 시뮬레이션하기 위해 2006년경 모델링하고 실행한 기계학습 시뮬레이션을 주요 탐구사례로 선정한다. 그리고 해당 시뮬레이션에 반영된 인간의 종교성과 신앙에 대한 사회인지 이론, 이 이론을 기계학습 기술로 검증하기 위해 기획된 시나리오, 실험에 활용된 일부 수학적-통계적 기법과 그 원리에 대한 종교사회학적 고찰과 반성을 수행하는 데 초점을 맞춘다.

이를 위해 본 연구에서는 특별히 현대 종교사회학의 창도자 가운데 하나로 인정받는 에밀 뒤르켐(David Émile Durkheim)의 종교학 이론을 분석과 비판의 준거로 채택한다. 뒤르켐은 종교가 사회적 산물로서 사회적 행위의 한 기능으로 이해될 때에만 비로소 환영이 아닌 진정한 삶의 현상으로 파악될 수 있다는 사회실재론적 견해를 고수한다.<sup>10)</sup> 이처럼 사회적 행위를 종교성 현성의 가장 근원적인 조건으로 내건 뒤르켐의 종교론은 사회학적 관점으로 종교현상들을 해석하고 그 의미를 길어내는 작업에 있어 하나의 모범적인 준거를 제공한다. 이로써 그의 연구방법은 베인브릿지가 수행한 종교적 인공지능 시뮬레이션의 함의를 도출하는 과정에도 일정 부분 타당하게 적용될 수 있을 것으로 판단된다.

연구 전체는 크게 두 부분으로 나뉜다. 전반부인 제II장에서는 베인브릿지가 수행한 종교적 인공지능 시뮬레이션 실험의 구체적인 내용을 살펴보고 그 가운데서 그가 종교성 현성을 위해 필수적인 것으로 상정한 존재적 유한성 의식의 개인적-사회적 조건들을 검토한다. 후반부를 이루는 제III장에서는 우선 뒤르켐의 종교발생 이론의 요체를 개관한다. 그리고 이를 바탕으로 베인브릿지의 실험이 그 한계점에도 불구하고 인간의 종

10) Émile Durkheim (1960), *Les Formes Élémentaires de la Vie Religieuse: Le Systémique en Australie*, Paris: Presses Universitaires de France, pp. 13-14; Thomas A. Idinopulos (2002), "The Strengths and Weaknesses of Durkheim's Methodology for the Study and Teaching of Religion", *Reappraising Durkheim for the Study and Teaching of Religion Today* (eds. Thomas A. Idinopulos & Brian C. Wilson), Leiden: Brill, p. 2.

교성 현성 과정을 재현하고 설명하는 데 있어 상당한 수준의 종교학적 통찰을 제공하고 있다는 사실을 논증하고자 한다. 아울러 이 실험이 장차 인공지능 간 상호작용에 의해 인간의 종교와는 다른 새로운 형태의 초월지향적 종교를 탄생할 가능성을 보여주고 있다는 점을 그 증추적 함의로 제시해 보고자 한다.

최근 국내에서 수행된 인공지능 관련 종교학 연구들은 크게 두 가지 특징을 보인다. 첫째는 인간의 마음과 인공지능 간의 존재론적 차이를 부각시킴으로써 지능을 단지 수학적 논리와 인식원리로 환원시키려는 계산주의적 사고에 대한 반성을 시도한다.<sup>11)</sup> 둘째는 이처럼 인간과는 명백한 존재론적 차이를 보이는 인공지능이 과연 종교라는 인간적 행위 및 현상에 어떻게 관여할 수 있을지 고찰한다.<sup>12)</sup> 본 연구는 이러한 동향들을 유념하는 가운데, 현재까지 국내에서 거의 시도된 바 없는 종교적 인공지능 시뮬레이션에 대한 구체적 분석과 평가를 수행한다는 점에서 기존의 연구들과 차별점을 보인다.

11) 인간과 인공지능의 지적 특성에 대한 환원적 이해에 비판적 태도를 보이는 최근의 인공지능 관련 연구로는 다음을 참조할 것. 박찬국(2018), 「인간과 인공지능의 미래: 인간과 인공지능의 존재론」, 『현대유럽철학연구』 50, 한국하이데거학회, pp. 119-166; 정재현, 박옥주(2018), 「사이버네틱스(cybernetics)의 현상학적 전환」, 『철학논총』 94, 새한철학회, pp. 489-518; 문장수(2018), 「사이버네틱스의 인식론과 메타-과학적 지위」, 『철학논총』 94, 새한철학회, pp. 217-243.

12) 인공지능이 종교라고 하는 다분히 ‘인간적인’ 영역에 관여할 가능성을 고찰한 최근의 연구로는 다음을 참조할 것. 이창익(2017), 「인간이 된 기계와 기계가 된 산: 종교, 인공지능, 포스트휴머니즘」, 『종교문화비평』 31, 종교문화비평학회, pp. 209-254; 김영진(2017), 「인공지능, 종교적 명상, 그리고 마음의 지향성: ‘명상하는 AI’의 의미와 가능성에 대한 철학적 고찰」, 『철학·사상·문화』 24, 동서사상연구소, pp. 1-27; 김진(2018), 「인공지능시대에도 종교는 가능한가?: 칸트와 블로흐의 종교철학적 관점에서」, 『철학연구』 147, 대한철학회, pp. 117-146.

## 2. 자기보존본능과 자원교환행위: 종교성 현성의 개인적-사회적 조건

베인브릿지는 2006년 출간된 그의 저서 『기계장치로부터 현현한 신: 종교적 인지의 인공지능 모델들』(*God from the Machine: Artificial Intelligence Models of Religious Cognition*)에서 기계학습 인공지능을 통해 인간의 다양한 종교적 행위들을 구현하는 실험을 수행한다. 그는 우선 종교행위 전체를 일종의 사회적 현상으로, 다시 말해 사회학의 연구 대상으로 성격규정한다.<sup>13)</sup> 이는 종교행위를 인간 공동체 구성원들 간 사회적 상호관계 가운데서 유발되는 것으로 간주한다는 뜻이다. 그는 기계학습을 통해 구현하려는 종교행위들로 전파에 의한 신도의 신규모집 (recruitment), 동일종교 신도들 간 결속(fellowship), 타종교 신도들에 대한 신뢰(trust)와 편견, 동일종교 및 타종교 신도들 간 협력(cooperation), 초자연적 존재자에 대한 신앙(faith)의 발생, 그리고 이 초자연자 신앙의 제도화 및 문화화 등을 지목한다. 이 가운데 초자연적 존재자에 대한 신앙을 발생시키는 사회적 행위 구현 실험은, 초월적 존재에 대한 믿음이라는 명제적 의식을 에이전트(agent, 인공지능 개체)들에게 선행적으로 이식하지 않은 채 이행된다는 점에서 다른 실험들과 구별된다.<sup>14)</sup>

기계학습 기술의 특성상 인공지능에 종교적 믿음을 이식하는 일이 실상 크게 어려운 일은 아니다. 초자연자의 존재를 사실로 발표하는 명제가 입력값에 반복적으로 포함되면 기계학습 인공지능은 자연스럽게 신앙의 내용을 학습하게 된다. 그러나 이 방법은 인공지능의 자율성이 거의 배제되다시피 한, 세뇌에 가까운 방식의 학습으로서, 진정한 의미에서 인공지능에게 자유로운 믿음의 기회를 수여하는 방식이라고 말하기

13) William Sims Bainbridge (2006), *God from the Machine: Artificial Intelligence Models of Religious Cognition*, Lanham: Altamira Press, pp. 6-7.

14) Bainbridge (2006), p. 117.

어렵다. 이런 방식은 단지 데이터 편향 효과<sup>15)</sup>를 이용한 것에 불과하다. 만약 인공지능 스스로 초자연자의 존재에 대한 신앙을 발생시키도록 하려면 입력값 안에 초자연자의 존재에 대한 명제가 배제되어 있어야 한다. 대신 이 경우에는 초자연자의 존재를 인식할 수 있을만한 종교적 인지의 조건이 선형적으로 구비되어 있어야 한다. 베인브릿지는 바로 이 선형성에 착안한 실험을 기획한다. 그는 에이전트들에 선형적으로 주어지는 정보들 가운데서 초자연자의 존재에 대한 정보를 배제하는 대신 초자연적 존재자에 대한 신앙이 발생할 수 있는 인간학적-사회학적 조건을 마련하는 데 부심한다. 이를 위해 그는 인간 인식의 주된 동기요소이자 기계학습의 주된 기제 가운데 하나인 보상-처벌 기제에 주목한다. 모든 인간은 자신에게 결핍된 것을, 그 가운데서도 특별히 존재의 열망을 충족시키는 대상을 욕망하는 존재자이다.<sup>16)</sup> 그리하여 이런 욕망의 대상을 획득할 때마다 이 경험을 유의미한 정보로 여기고 이 정보들을 축적하고 종합한다. 이는 향후 개시할 존재의 열망 충족 시도를 보다 용이하게 만들기 위한 지적 학습 행위이다. 이 경우 보상이란, 베인브릿지의 규정에 의하면, 실제 존재의 열망을 충족시키는 유무형의 자원(혹은 재화)을 비

15) 데이터 편향(data bias)이란 기계학습의 입력값을 편향된 자료들로 구성함으로써 결과값 역시 편향된 것을 얻게 되는 현상을 말하는 것으로, “통상 통계적 편향이라는 관점에서 연구된다.” 데이터 편향은 인공지능과 관련된 윤리논란을 촉발하는 주역 가운데 하나로 지목된다. Mathew Salvaris, Danielle Dean & Wee Hyong Tok (2018), *Deep Learning with Azure: Building and Deploying Artificial Intelligence Solutions on the Microsoft AI Platform*, New York: Apress, p. 73.

16) 베인브릿지는 인간이 주로 유념하는 결핍들 가운데 특별히 죽음을 언급한다. 인간은 나이가 들면서 점차 죽음, 즉 자기 존재의 무화에 대해 염려하며, 이로 인해 통상 “수명의 연장”(the extension of life)에 대한 열망을 갖게 된다. Bainbridge (2006), p. 123. 죽음을 앞두고 자기 존재의 유한함에 불안을 느끼며 신의 존재나 내세 등에 대한 이념을 통해 이 불안을 회피하려는 행태는 실존철학적 관점으로 존재의 열망의 통속적 양태 가운데 하나라고 볼 수 있다. Richard Colledge (2014), “A-Z Glossary: Death”, *The Bloomsbury Companion to Existentialism* (eds. Felicity Joseph, Jack Reynolds & Ashley Woodward), New York: Bloomsbury, p. 339.

롯해 이 자원을 획득하는 일을 용이하게 만드는 정보까지 포괄한다. 반면 처벌이란 존재의 열망을 충족시키는 자원의 확보에 실패하는 것에 더하여 유익한 정보마저 획득하지 못해 결국 무지의 상황에 처하는 것까지 포함한다. 이런 맥락에서 그는 “종교를 인간의 마음이 작동하는 방식의 자연스러운 결과물”이라고 성격규정하는 심리학 및 인간학 기반 인지과학자들의 견해를 수용한다.<sup>17)</sup>

대체 왜 종교는 인간의 마음이 산출한 자연스러운 결과물이라 할 수 있는가? 상기 약속한 인간의 보상-처벌 기제에 따르면, 인간은 기본적으로 생존과 이를 위한 학습을 열망하는 존재자로 밝혀진다. 물론 이는 인간만의 특성은 아니고, 여러 동물들도 부분적으로 공유하는 본능이기는 하나, 현실에서 이 열망을 명시적으로 영원과 무한이라는 초월의 지평에 잇대어 사유하고 거기에 가치를 부여하는 것은 인간뿐이라는 점에서 인간과 여타 동물들 사이의 결정적인 차이가 존재한다.<sup>18)</sup> 다시 말해 인간

17) Bainbridge (2006), p. 5.

18) “오직 인간만이 종교의 주체가 될 수 있는가? 동물은 종교 혹은 종교성을 가질 수 없는가?” 이 문제는 그동안 종교학계 일부에서, 종교 혹은 종교성이 오직 인간만의 전유물이 아니라고 믿는 연구자들에 의해 꾸준히 제기되어 온 것이다. 대표적으로 김벌리 패튼(Kimberley Patton)은 성서에 기록된바 신이 동물들에 대해 상당한 관심을 표명하고 있다는 것, 동물들이 신을 인지할 수 있다는 것, 그리고 신과 동물 사이에 소통과 대화가 이루어질 수 있다는 것을 예로 제시하며 서구 그리스도교 전통 가운데서도 동물들이 향유하는 종교나 신앙에 대한 연구의 여지가 확보되어 있다고 주장했다. Kimberly Patton (2000), “He Who Sits in the Heavens Laughs”: Recovering Animal Theology in the Abrahamic Traditions”, *Harvard Theological Review* 93, Harvard Divinity School, pp. 409-417. 그러나 여전히 인간 외의 어떠한 동물이라도 명확하게 ‘종교적 동물’(animal religious)로 규정될 만한 증거가 존재하지는 않는 형편이고, 이 때문에 이 분야의 연구에 상당한 어려움이 따른다는 사실 역시 연구자들 사이에 충분히 인지되고 있다. 폴 왈도(Paul Waldau)도 패튼과 마찬가지로 ‘동물들의 종교’에 대한 학문적 연구의 필요성에 공감하는 입장이나, 인간이 동물의 마음이나 심리상태를 어느 정도 분명하게 분석하고 관찰할 수단이 존재하지 않는 한 동물들의 종교에 대한 단정적 이론화는 불가능하다는 사실을 인정한다. Paul Waldau (2006), “Seeing the Terrain We Talk: Features of the Con-

은 자기보존본능(*conatus essendi*)을 영원과 무한이라는 이념에 연결 지을 수 있고 또 상시적으로 그렇게 연결 지으려 한다. 그리고 그 결과 영생, 부활과 구원, 윤회 등 사후(死後) 인간 존재의 연장을 전제하는 모든 종교적 신앙의 요소들이 출현한다. 이는 종교학계에서 일반적으로 공유되는 종교의 기원에 대한 해명 가운데 하나이기도 하다.<sup>19)</sup>

베인브릿지는 이런 해명에 동조하는 가운데 인공지능을 통해 인간 고유의 종교성 혹은 종교적 선형성을 구현하려 한다. 그는 기계학습이 인

---

temporary Landscape of ‘Religion and Animals’”, *A Communion of Subjects: Animals in Religion, Science, and Ethics* (eds. Paul Waldau & Kimberly Patton), New York: Columbia University Press, p. 47. 이런 학문적 정황들을 유념한다면, 현 단계에서 명시적으로 종교성 혹은 종교문화를 향유하는 좋은 오직 인간뿐이라는 진술이 부당하다 평가하는 어려울 것이라 생각된다.

- 19) 존재의 열망이 원초적 종교성으로 이어지는 과정에 대해서는 크리스토퍼 심슨(Christopher Ben Simpson)에 의해 소개된 윌리엄 데스몬드(William Desmond)의 존재와 마음의 현상학을 참조하면 도움이 될 것으로 보인다. 심슨에 의하면, 데스몬드는 인간의 마음이 그 유년기에 최초로 무가 아닌 존재를 의식하면서 “존재와 무 사이의 비대칭(assymetry)”에 격동되어 경이로움에 빠져들고, 이후 지속적으로 이어지는 존재의 소여(the givenness of being) 가운데서 존재의 충만함(the fullness of being)에 압도되는 지경에 이른다고 분석한다. 이 존재의 충만함은 “능동적인 존재의 열망(active *conatus essendi*) 발현 이전에, 존재에 대한 걱정(*passio essendi*)”을 불러일으키는데, 이는 “존재의 충만함 앞에서 절감하는 아가페적 경이(agapetic astonishment)의 사고를 감내하며 수용하는 것”을 의미한다. 여기서 이런 경험과 경이가 아가페적이라는 것은 인간의 마음을 압도하는 존재의 초월성으로부터 유래한다는 의미, 이 존재로부터 무한하게 수여된다는 의미로 사용된 용어이다. 이 초월적 존재, 무한한 소여로서의 존재를 향한 아가페적 경이가 에로스적이고(erotic) 능동적인 존재의 열망으로 화하게 된다. 이런 분석을 통해 데스몬드는 인간의 마음 안에 초월자, 무한자에 대한 숭배의 심정인 종교성이 현성하는 원동력으로 존재의 열망을 지목한다. Christopher Ben Simpson (2009), *Religion, Metaphysics, and the Postmodern: William Desmond and John D. Caputo*, Bloomington: Indiana University Press, pp. 35-36, pp. 47-49. 이처럼 종교성이 인간의 원초적인 존재의 열망과 결부되어 있는 한, 종교성은 비록 제도적 종교로의 현신으로 이어지지 않더라도 인간의 삶의 근원적 차원에서 끊임없이 인간의 마음을 다양한 양태의 존재로의 초월로 견인한다고 말할 수 있을 것이다.

간의 욕망행위와 마찬가지로 보상-처벌 기제를 통해 실현된다는 사실에 주목한다. 이는 기계학습 기술을 통해 인공지능에게 특정한 자원 혹은 재화에 대한 욕망을 주입시키고 이 욕망이 주도하는 기제에 따라 연산 혹은 사고하도록 유도할 수 있다는 의미로도 이해될 수 있다. 그렇다면 인공지능이 인간의 종교적 선형성을 구현해내기 위해 이 보상-처벌 기제 혹은 욕망의 기제의 내용을 무엇으로 채울 것인가가 문제로 대두된다. 우선적으로 제기되는 물음은 다음과 같다. “인공지능으로 하여금 무엇을 욕망하도록 해야 하는가?” 베인브릿지는 이에 대해 세 부류의 욕망의 대상을 설정한다. 첫째는 생존을 위한 재화, 둘째는 성애(性愛)의 대상, 셋째는 생명 자체(life itself)이다. 여기서 생명이란 죽음에 의해 위협을 받는 인간의 유한한 존재를 연장시키는 힘을 의미한다. 이 힘은 시간 경과에 따라 인간의 존재적 유한성 체감도가 점점갈수록 절박하게 요구되는 것으로서, 현실적인 삶 자체로부터는 발굴할 수 없는 일종의 초월적인 가치이자 자원이다.<sup>20)</sup>

이처럼 욕망의 대상 설정이 마무리되면, 다음으로 제기될 물음은 욕망하는 방법에 관한 것이다. “인공지능으로 하여금 그것들을 어떻게 욕망하도록 해야 하는가?” 이를 해결하려면 인간의 종교성 현성을 위한 환경을 조성하는 정교한 시나리오의 기획과 이 기획을 현실화하기 위해 동원되어야 할 기계학습의 기술적 요소에 대한 이해가 요구된다. 베인브릿지는 종교 발생 과정을 시뮬레이션하기 위해 종교가 존재하지 않는 공동체, 기존의 종교 교의들로부터 완벽하게 자유로운 공동체를 가정한다. 그가 설정한 시나리오는 다음과 같이 요약된다.

지구와는 다른 행성, 화성(Mars)에 44,100명의 지구인이 이주한다. 이 공동체의 이름은 사이버그(Cyburg)로 가칭한다. 사이버그에 이주한 이들은 지구에서 어떤 종교도 알지 못했던 이들이다. 이 가운데 남성은 22,050

20) Bainbridge (2006), pp. 119-120.

명, 여성 역시 22,050명으로 성비는 균등하다. 사이버그에 이주한 44,100명은 각기 생산하는 필수자원의 종류에 따라 #0집단(발전소 운용, 에너지 생산), #1집단(얼음광산 운용, 에너지를 활용하여 물 생산), #2집단(수경재배 시설 운용, 에너지와 물을 활용하여 식량 생산), #3집단(전기분해 시설 운용, 에너지와 물을 활용하여 공기 생산)의 4개 집단으로 재차 구분된다. 각 집단의 인구수는 각각 11,025명으로 동일하다. 이 4개 집단에 속한 사람들은 생존을 위해 각자 생산해 보유한 자원을 다른 이들의 자원과 교환한다. 모든 재화 생산의 출발점은 #0집단의 발전소이다. 발전소 운용을 맡은 이들은 여타의 자원 소비 없이 자동화된 생산설비를 이용해 에너지를 생산하며, 나머지 집단은 각기 합당한 분량의 다른 자원을 소비함으로써 각자 담당하고 있는 자원을 생산한다. 이는 곧 #0집단이 여타 3개 집단에 비해 사회적으로나 경제적으로 유리한 위치에 있는 집단임을 의미한다. 이로써 생존에 필요한 자원 4개(에너지, 물, 식량, 공기)가 각 사람들에게 가치 있는 재화로, 욕망의 대상으로 인식된다.

그런데 여기에 더해 인간은 원초적 욕망 두 가지를 더 갖고 있다. 생존을 위한 자원이 확보되기 시작하면서 사람들은 우선 성애(性愛)를 구한다. 남녀 간 사랑은 감정적 충족과 함께 종족과 공동체를 보존할 후세를 남긴다. 그리고 시간의 경과와 함께 정해진 수명이 다해가는 시점부터는 점차 더 절박하게 생명의 연장을 추구한다. ‘죽음의 불안’(die Todesangst)은 이를 극복하기 위한 초자연적 자원인 생명 자체에 대한 열망을 낳는다.<sup>21)</sup> 이 두 가지 자원은 앞서 생존을 위한 자원으로 제시했던 네 가지

21) 폴 웡(Paul T. P. Wong), 개리 레커(Gary T. Reker), 지나 게서(Gina Gesser)는 인간이 죽음의 불안에 대처하는 방식, 즉 죽음을 수용하는 방식에 대해 세 가지 모델을 제시한다. 첫째는 중립적 수용(neutral acceptance)으로 다분히 실존철학적인 죽음의 수용방식을 말한다. 이 경우 죽음이란 삶에 당연하게 동반되는 경험이자 사건이며, 결코 변개할 수 없는 인간 존재의 무화이므로 유한한 삶을 충실하게 살려는 동기를 유발한다. 둘째는 접근적 수용(approach acceptance)으로 여러 내세신앙이 권장하는 죽음의 수용방식을 말한다. 이 경우 죽음이란 내세로 진입하는 시점이며,

자원과 달리 물리적 실재가 아니다. 성애는 기본적으로 감정적으로 향유되는 자원이며, 성애에 뒤따르는 성관계와 그 결과 얻게 되는 자녀는 성애의 파생물이지만 성애 자체는 아니다. 생명의 경우 죽음 이전까지의 생명력은 물리적 실재로 규정될 수도 있지만 죽음 이후의 존재를 가능케 하는 생명의 힘은 물리적 실재의 범주를 넘어서는 어떤 것으로 분류된다. 이들 외에도 인간에게는 다양한 형태의 욕망의 대상이 존재하나 가장 원초적인 것들로는 상기 여섯 개의 자원을 지목한다. 이제 사이버그 이주민들은 각기 이 여섯 가지 자원을 획득하기 위해 노동하고 타인과 접촉해 자원을 교환한다. 각 사람은 이런 교환활동 경험이 누적될수록 어떤 거래대상이 자신이 필요로 하는 자원교환에 적합한 인물인지 학습하게 된다. 여기에 더해 각 사람은 자신에게 우호적인 거래대상으로부터 자원교환에 관한 추가적인 정보도 얻게 된다. 그런데 원초적 욕망들 가운데 5가지는 이를 충족시키는 대상이 존재하나, 오직 생명 자체만은 현실에 이를 충족시키는 거래대상이 존재하지 않는다. 애초 그것은 현실에 존재하는 사람과 환경으로부터는 획득할 수 없는 것이다. 그럼에도 불구하고 사람은 누구나 여전히 이 자원을 갈망한다.<sup>22)</sup>

이런 화성 이주 시나리오를 바탕으로 베인브릿지는 시뮬레이션에 필

---

내세란 유한하고 불안한 현세의 삶에 대한 보상으로 인간의 존재가 무한과 초월의 지평에 들어선 상태를 가리킨다. 셋째는 회피적 수용(escape acceptance)으로 죽음을 고통과 비참함으로 가득한 삶의 유일한 탈출구로 여기는 죽음 수용방식을 말한다. 이 경우 내세에 대한 믿음은 존재하지 않으며, 오로지 존재의 무화만이 삶의 악독한 고통을 벗어날 수 있는 방편으로 여겨진다. 이 가운데 접근적 수용은 종교성과 특별히 깊게 연관되어 있는 것으로 확인된다. Paul T. P. Wong, Gary T. Reker & Gina Gesser (1994), "Death Attitude Profile — Revised: A Multidimensional Measure of Attitudes toward Death", *Death Anxiety Handbook: Research, Instrumentation, and Application* (ed. Robert A. Neimeyer), Washington, DC: Taylor & Francis, pp. 125-127. 베인브릿지의 종교적 인공지능 실험은 죽음의 불안을 종교발생의 주원인으로 지목하며 내세를 누리게 해주는 생명을 하나의 자원이자 보상으로 상정함으로써 상기 죽음 수용유형 분류 가운데 접근적 수용의 견해를 지지한다.

22) Bainbridge (2006), pp. 120-123.

요한 기계학습의 기술적 요소들을 선정한다. 그는 우선 보상-처벌 경험 누적을 통한 판단과 행동의 개선을 구현하는 강화학습(reinforcement learning) 방식을 지목해 설명한다. 컴퓨터를 통한 보상-처벌 경험 누적이 가능하기 위해서는 각 보상-처벌 경험을 기억시키는 메모리 레지스터(memory register)<sup>23)</sup>가 요구된다. 우선 이 메모리 레지스터는 에이전트가 욕망의 대상으로 삼는 자원의 종류에 따라 6개의 범주를 갖게 된다. 각 범주는 욕망의 대상이 되는 자원들을 획득하려는 인공지능 에이전트의 기대감 정도를 표시하는 항목들을 산출한다. 각 에이전트는 확률적인 원리에 따라 대개 이 기대감의 정도가 높은 항목의 에이전트와 접촉해 자원교환을 시도하려 한다.<sup>24)</sup> 사이버그에 소속된 44,100개의 에이전트들은 모두 이 6개의 범주에 따라 자원획득의 기대치를 갖게 되는데, 베인브릿지는 이 기대치를 다시 가로항목과 세로항목으로 분류한다.

가로항목은 각 에이전트가 실제로 존재하는 네 개의 집단, 즉 #0, #1,

---

23) 메모리 레지스터란 특정 프로그램의 입출력 값을 저장하기 위해 컴퓨터 기억장치 가운데 마련된 공간으로 기계학습 인공지능의 능력 향상을 위해 필수적으로 요청된다. Bainbridge (2006), p. 7. 메모리 레지스터는 인공지능이 수득한 행위 경험들을 그 계열에 따라 분류해 저장하는 역할을 수행하므로, 선험론적 인식론 입장으로 본다면 잡다한 표상들을 종합하는 기준이 되는 지성개념(*der Verstandesbegriff*), 즉 범주(*die Kategorie*)에 상응한다고 볼 수 있다. 여기서 표상, 지성개념, 범주의 의미는 Immanuel Kant (1956), *Kritik der reinen Vernunft*, Hamburg: Felix Meiner, pp. A78-A79/B104-B105로부터 채용된 것이다.

24) 여기서 각 에이전트가 확률적 원리에 따라 행동한다는 것은 무조건 보상을 얻을 확률이 높은 항목으로만 자원의 교환을 시도한다는 것이 아니고 그렇게 행동할 가능성이 높다는 것을 의미한다. 따라서 학습 결과에도 불구하고 각 에이전트의 행위 가운데는 일정 정도의 불확실성 및 우연성이 드러난다. 물론 어느 한 항목의 자원거래 성사 확률이 100%를 보인다면 에이전트는 학습에 따라 거의 틀림없이 해당 항목으로 교환을 시도하겠지만, 기계학습의 경우 여기에도 일정 정도의 우연성이 반영되도록 설정이 가능하다. 예를 들어, 100%의 교환 성사 확률을 보이는 항목을 향해 차후 교환을 시도할 가능성을 최대치를 100%가 아닌 90%로 설정할 경우, 에이전트는 열 번에 한 번은 무작위적으로 다른 항목의 에이전트에게 자원거래를 시도할 것이다. Bainbridge (2006), p. 10.

#2, #3 집단 중 어디에 속한 에이전트와 교환활동을 시도하면 자원획득을 성사시킬 수 있을지에 관한 기대치를 나타낸다. 세로항목은 각 에이전트가 사이버그의 이주민 전체를 몇 개 집단으로 구분한 뒤 거래상대를 찾을지에 대한 기대치 및 가능성을 나타낸다. 이 세로항목이 설정된 이유는 일단 에이전트들이 사이버그에 속한 전체 인구집단의 수와 종류(가로항 4개)를 정확히 알지 못하는 것으로 상정했기 때문이다. 각 에이전트들은 최초 교환활동 개시 시점에 4개 생산집단의 존재에 대해 무지하고, 단지 자신이 속한 집단이 존재하며 또 이와 다른 생산집단들이 존재한다는 사실만을 알고 있다. 그래서 에이전트들은 각기 사이버그 내에 몇 개의 교환집단이 존재할지 자율적으로 상정하고서 교환활동을 개시한다.<sup>25)</sup> 즉 4개의 가로항은 각 에이전트들이 인지하지는 못하지만 그들이 처해 있는 현실을 반영해 실제 사이버그에 존재하는 유형자원의 생산집단 항목을 표시하고, 3개의 세로항은 에이전트들이 이 4개의 실제 생산집단을 인지하지 못하는 상태에서 교환활동에 나서는 현실을 반영해 각기 임의적으로 교환대상집단의 수를 상정하는 상태를 표시하는 것이다. 가로항은 충분한 학습이 이루어지기 전까지는 에이전트들에게 인지되지 못한 채로 남게 되는 반면 세로항은 항상 인지된다는 차이가 있는 하나, 양측 항목 모두 각 에이전트가 인지해야 할 현실을 반영하고

25) 베인브릿지는 인간이 타인과 소통하고 교류할 때 필히 주관적이고 임의적인 방식의 집단분류를 시도한다고 상정하고 있다. 물론 현실에서 인간이 소통과 교류의 대상집단을 분류할 때 그 가짓수는 거의 무궁할 것이다. 왜냐하면 인간은 위 실험에서 기획한 인공지능과 달리 정해진 분류기준에만 만족하지 않고 계속해서 새로운 분류기준을 창안해낼 것이기 때문이다. 베인브릿지는 이처럼 현실에서 보상획득 혹은 자원교환의 대상집단 분류방식이 고정되어 있지 않고 경우에 따라 유동적이며 무조건적으로 일관되지 않다는 사실을 실험에 반영하기 위해 에이전트가 매번 자원교환을 시도할 때마다 몇 개의 대상집단을 설정할지(교환대상집단의 수) 범주화한 후 확률적으로 어떤 범주를 선택할지 결정하도록 설정해 두었다. 물론 이 확률은 교환성사 여부의 학습에 따라 유동적으로 변한다. Bainbridge (2006), p. 125.

있다는 점에서는 이런 항목 설정이 일정부분 타당하게 받아들여질 수 있을 것으로 여겨진다.

교환활동의 구체적 조건은 다음과 같이 설정된다. 한 회차의 교환활동마다 #0집단에 속한 에이전트(발전소 소유)들은 각각 500단위의 에너지를 생산한다. #1집단 에이전트(얼음광산 소유)들은 각각 250단위의 에너지를 소비해 500단위의 물을 생산한다. #2집단 에이전트(수경재배 농장 소유)들은 각각 250단위의 물을 소비해 500단위의 식량을 생산한다. #3집단 에이전트(산소공장 소유)들은 각각 125단위의 에너지와 125단위의 물을 소비해 500단위의 공기를 생산한다. 에이전트들은 각 회차마다 생존에 필요한 자원을 얻기 위해 다른 에이전트들과 접촉해 거래를 시도하며, 이때 서로 요구조건이 맞는 에이전트들 사이에만 거래가 성사된다. 여기에 더해, 모든 에이전트들은 각각 1,000단위만큼의 생명 자체를 자원으로 갖고 있는데, 이 자원은 교환활동 시행 회차마다 1단위씩 소비되며 다른 네 개의 필수 자원들과 달리 교환되거나 보충될 수 없다. 간단히 말해 각 에이전트는 각기 1,000회의 교환활동을 수행할 수 있는 수명을 가졌으며, 이 수명은 당연히 자원의 교환활동을 통해서만 연장시킬 수 없다. 물론 각 에이전트는 처음에는 생명 자체, 즉 수명이라는 자원이 사 이버그라는 현실 속에서는 교환되거나 보충될 수 없다는 사실을 알지 못한 채로 교환활동을 시작한다.<sup>26)</sup>

이 실험에서 각 에이전트가 생각할 수 있는 최대 생산집단 수는 4개로 설정되었다. 이는 다시 말해 각 에이전트들이 최대 3개까지 교환대상집단의 가짓수(2, 3, 4개)를 범주화해 상정할 수 있다는 것을 의미한다. 각 에이전트가 생각할 수 있는 최대 4개의 생산집단 가운데 하나는 각 에이전트 자신이 속해 있는 집단, 즉 동일한 종류의 자원을 획득하려는 집단 이기에 해당 자원 획득을 위한 교환대상으로 취급되지 않기 때문이다.

26) Bainbridge (2006), pp. 131-132.

이에 각 에이전트가 자원교환 시도를 통해 특정 자원을 획득할 기대치를 표명할 수 있는 항목 수는 [표 1]과 같이 가로 4개, 세로 3개로 총 12개로 확인된다.

[표 1] 한 개별 에이전트의 ‘에너지 자원’ 레지스터 초기값<sup>27)</sup>

	#0집단	#1집단	#2집단	#3집단
2개 생산집단	732	673	554	709
3개 생산집단	628	967	938	561
4개 생산집단	941	534	728	566

[표 1]은 교환활동을 통해 획득될 것으로 예상되는 에너지 자원에 대한 기대치를 산출하기 위해 범주별로 배정된 확률변수를 표시한다. 이 12개의 확률변수는 그 자체로 에너지 자원 획득에 대한 기대치를 나타내지는 않는다. 자원획득 기대치는 이 확률변수에 통계분석법인 분산(variance)을 적용해 산출한다.<sup>28)</sup> 분산 산출을 위해 각 항목에 배정되는 확률변수는 각 에이전트가 일정하게 설정된 범위(연산의 편이상 500~999) 안에서 무작위로 지정한다. 이 경우 이 확률변수들의 분산값, 즉 각 확률변수가 전체의 평균으로부터 얼마나 떨어진 채로 분포되어 있는지 측정한 값이 크면 클수록 에이전트가 해당 항목으로부터 그 범주에 속한 자원을 획득할 희망을 더 많이 갖는다는 의미로 해석된다. 반대로 각 항

27) Bainbridge (2006), p. 127.

28) 통계학적 입장에서 분산값이 미미하다는 말은, 각 변수들이 평균값  $\mu$  으로부터 멀리 떨어져 있지 않다는 뜻을 의미한다. 베인브릿지의 실험에서 이 분산값은 해당 항목으로부터 보상(자원)을 획득할 수 있을지 확률적으로 판단하는 정도를 의미하는데, 이 확률적 판단이 평균값과 별 차이가 없다는 것은 해당 항목에서는 여타의 항목들과 비교해 평균적인 수준의 보상 확률밖에 기대할 수 없어 보상의 확실성이 저조할 것으로 기대한다는 뜻이다. 반면 분산값이 크다는 것은 보상을 얻을 확실성이 평균적 수준보다 클 것으로 기대한다는 뜻이다. Bainbridge (2006), p. 11, pp. 127-128.

목에 최초 무작위적으로 배정되는 확률변수가 전체 확률변수 평균인  $\mu$ 에 가까울수록, 그래서 분산값이 작아질수록 해당 항목에 속한 에이전트들과의 교환활동에서는 자원획득이 제대로 이루어지지 않을 것이라고 기대한다는 것을 의미한다.

이 분산값은 최초 무작위로 배정된 확률변수들에 의해 각 에이전트마다 서로 다르게 출발되며, 에이전트 간 교환활동 시행에 따라 점차 학습에 의해 수정된다. 이 실험과정 가운데서 베인브릿지는 인공지능 사이에 종교적 믿음, 즉 초자연적 존재자에 대한 신앙이 출현하는지 살피기 위해 생명 자체라는 자원범주의 항목별 수치에 주목한다. 다음에 소개되는 결과값은 생명 자체에 대한 교환-획득 기대치 시뮬레이션 결과값으로 교환행위가 1,000회 이루어진 상태의 수치이다. 즉 사이버그의 현실 내에서는 한 차례의 수명주기(각 에이전트의 수명 = 1,000회 교환활동)가 지나갔음을 의미한다. 베인브릿지는 시뮬레이션의 안정성과 연속성을 확보하기 위해 우선 수명이 다해 소멸되는 에이전트와 새로 태어나 그 자리를 채우는 에이전트의 수와 성별이 동일하다는 조건을 설정했다. 다시

[표 2] 생명 자체의 획득을 추구하는 두 에이전트의 범주별 기대치<sup>29)</sup>

	분산(%)	#0집단	#1집단	#2집단	#3집단
물 생산 에이전트:					
2개 생산집단	56.57	1.70	1.70	<b>33.45</b>	<b>19.72</b>
3개 생산집단	37.23	1.34	1.34	12.92	<b>21.63</b>
4개 생산집단	6.20	0.92	2.87	1.25	1.16
에너지 생산 에이전트:					
2개 생산집단	62.67	2.16	2.16	<b>33.75</b>	<b>24.60</b>
3개 생산집단	37.33	2.41	2.41	2.41	<b>30.10</b>
4개 생산집단	0.00	0.00	0.00	0.00	0.00

29) Bainbridge (2006), p. 134.

말해 사이버그의 인구수와 성비는 세대가 지나도 변함이 없게 설정된 것이다.

[표 2]에서 상단부의 “물 생산 에이전트” 항은 얼음광산을 소유한 생산집단인 #1집단에 소속된 에이전트 가운데 하나의 생명 자체에 대한 교환-획득 기대치 결과를 표본으로 지정해 표시한 것이다. 마찬가지로 하단부의 “에너지 생산 에이전트” 항은 발전소를 소유한 생산집단인 #0집단에 소속된 에이전트 하나의 기대치 결과를 표본으로 지정해 표시한 것이다.

베인브릿지의 관찰에 의하면, 사이버그로 이주한 후 교환활동을 개시한 에이전트들 전체는 각기 무작위로 설정한, 다시 말해 자유롭게 설정된 초기 분산값에 따라, 그리고 어느 생산집단에 속해 있는지에 따라 생명 자체에 대한 기대치에 있어 차별되는 모습을 보인다. 그리고 교환활동이 얼마 진행되지 않은 시점에서는 이 자원을 획득하리라 예견하는 기대치가 사이버그라는 현실 속에 존재하는 교환집단(들)을 지목하는 항목에서도 비교적 높게 나타났다가, 교환활동이 여러 차례 반복적으로 시도되면서 이 기대치는 거의 0에 근접한다. 이는 에이전트들이 현실에 존재하는 교환집단들 가운데서는 생명 자체를 획득할 수 없다는 사실을 학습을 통해 인지한다는 것을 의미한다. 반면 현실에 존재하지는 않지만 각 에이전트가 범주화할 수 있는 여분의 교환집단 항목들, 즉 초월적 교환집단들을 지목하는 항목들([표 2]에서 밑줄 친 항목들)에서 생명 자체에 대한 교환 및 획득 기대치가 점증되는 것이 확인되었다. 이는 곧 현실에 존재하지 않는 초자연자를 교환집단으로 상정하고 그들로부터 생명 연장, 존재 연장의 힘을 얻으려 한다는 것을 의미한다. 물 생산 에이전트의 경우 그 확률(존재의 영속적 연장을 위해 초자연적 존재자를 찾을 확률)이 세 항목 모두 합쳐 74.80%(33.45% + 19.72% + 21.63%)로 확인되고, 에너지 생산 에이전트의 경우 그 확률이 88.45%(33.75% + 24.60% + 30.10%)에 이른다. 풀이하자면, 위의 표에 지목된 물 생산 에이전트는 평

소 교환활동을 수행하는 가운데 초자연자에 대한 종교적 신앙을 반영하는 행태를 보일 확률이 74.80%에 이른다는 것이고, 동일하게 에너지 생산 에이전트는 그 확률이 88.45%에 이른다는 것이다. 즉 두 에이전트는 그 확률변수의 출발점과 소속된 생산집단의 차이로 인해 일정한 차이 ( $88.45\% - 74.80\% = 13.65\%$ )를 보이기는 하지만, 둘 모두 생명 자체를 얻으려는 열망 때문에 종교적 신앙을 가질 확률이 높다는 점에서는 공통점을 보인다.

여기서 특징적인 사실은 각 에이전트가 3개의 교환집단 전체를 사이버그 내에 실제 존재하는 에이전트들에게 모두 배정했을 경우, 즉 사이버그 내에 모두 4개의 생산집단이 존재한다고 인지한 경우에는 교환활동을 통해 생명 자체를 획득할 수 있다는 기대치가 0에 수렴했다는 점이다. 해석하자면, 각 에이전트가 동원할 수 있는 생산집단 범주들을 모두 실제 존재하는 존재자들에게 배정하는 경우, 즉 초자연자의 존재에 대해 사고할 범주가 더 이상 남아있지 않은 경우에는 생명 자체를 얻으려는 열망, 즉 죽음을 넘어서는 초월적 힘에 대한 열망이 거의 소멸된다는 뜻이다.<sup>30)</sup>

### 3. 개인적 욕망과 사회적 기능: 종교적 인지를 위한 범주들의 기원

현재까지 살펴본바, 종교사회학적 통찰에 기반을 둔 베인브릿지의 종교성 발생 기계학습 시뮬레이션에는 어떤 종교학적 의미가 함축되어 있는가? 기본적으로는 인간의 실존적 자유, 종교적 선택성, 그리고 죽음과 관련된 세 가지 함의에 주목할 필요가 있는 것으로 판단된다. 그의 실험

30) Bainbridge (2006), p. 134.

은 인간의 이 세 가지 실존적 본질을 바탕으로 기획되었고 이를 통해 초자연적 존재자에 대한 정보를 자율적으로 산출해내는 인공지능 에이전트를 개발하는 결실을 보였다. 이로써 그는 제도적 종교의 발생 이전 인간의 원초적 종교성 현성의 근본동기와 조건들을 시뮬레이션을 통해 재확인한 한편, 인공지능도 학습을 통해 종교적 믿음을 추구하고 산출하는 것이 가능하다는 사실을 입증하였다. 물론 여기서 기계학습 에이전트들이 산출한 종교적 믿음이라는 것이 인간이 향유하는 바에 비교해볼 때 지극히 제한적이라는 사실을 부정할 수는 없다. 인간의 종교적 믿음은 풍부한 상징과 이미지를 포함하는 신화적 표상들과 더불어 감정, 욕망(혹은 의지), 신체 전체를 포괄하는 전인적 종교성의 표현이자 결실이므로, 오직 인간의 지능만을, 그것도 제한적인 방식으로 모방한 기계학습 에이전트가 인간이 향유하는 수준의 신앙을 가질 것이라고 전망하기란 어려운 일이다. 그러나 실험 결과를 놓고 보자면 적어도 주지적인 측면으로는 기계학습 에이전트 역시 인간과 유사한 방식으로 초자연적 존재자에 대한 신앙체계를 산출할 가능성이 존재한다는 점은 수긍할 수 있다.

여기에서 주목해야 할 바는 상기 실험에서 초자연자에 대한 정보의 추구 및 산출이 바로 자원을 교환하는 사회적 행위에 의해 실현되는 것으로 간주된 점이다. 베인브릿지와 스타크는 1987년의 저서 『종교론』(*A Theory of Religion*)에서 종교행위를 포함한 인간행동 대부분이 총 7가지 공리들(axioms)과 104가지 개념정의들로부터 도출될 수 있다고 주장했다.<sup>31)</sup> 베인브릿지는 이 이론을 바탕으로 자신의 기계학습 시뮬레이션의 기초 시나리오를 작성하고 그 작동원리를 기획하는데, 여기서 주목해야 할 사실은 스타크-베인브릿지 종교론을 구성하는 7개의 핵심 공리가 기본적으로 인간이 삶을 지탱하고 풍요롭게 만들기 위해 소요되는 자원의 획득, 교환, 분배 행위를 전제로 삼는다는 점이다. 다시 말해 두 사람은

31) Rodney Stark & Bainbridge (1987), *A Theory of Religion*, New Brunswick: Rutgers University Press, pp. 325-331.

종교문화 발생을 위해 사회적 행위의 지속적인 이행이 필수적으로 요청된다는 것을 전면 수긍하는 가운데 자신들의 종교론을 개진하고 있는 것이다. 이런 사실은 다음에 제시되는 이 7개 공리의 구체적인 내용을 통해 확인된다.

- 공리 1: 인간의 지각과 행동은 시간 속에서, 과거로부터 미래를 향하는 가운데 이행된다.
- 공리 2: 인간들은 그들이 보상으로 지각하는 것을 추구하고 비용(costs)으로 인지하는 것은 회피한다.
- 공리 3: 보상들은 종류, 가치, 그리고 일반성(generality)에 있어 차이를 보인다.
- 공리 4: 인간 행동은 복잡하면서도 유한한 정보처리 체계(information-processing system)에 의해 지시되며, 이 체계는 문제들을 식별하고 이를 해결하기 위해 작동한다.
- 공리 5: 인간들이 바라는 보상들 가운데 어떤 것들은 공급이 제한되어 있으며, 이 보상들 가운데는 아예 [현실에] 존재하지 않는 것도 있다.
- 공리 6: 인간들이 추구하는 대부분의 보상들은 사용되고 난 후에는 파괴된다.
- 공리 7: 힘(권력)을 결정짓는 개인적, 사회적 특질들(attributes)은 어떤 사회에서든 개인들이나 집단들 사이에 불평등하게 배분되어 있다(unequally distributed).<sup>32)</sup>

여기에서 공리 2부터 공리 6까지 5개 공리는 인간 개개인의 행동이 필연적으로 사회적 행위로 전환될 수밖에 없는 이유가 무엇인지 밝혀준다. 이 5개 공리의 내용은 다음과 같이 종합 정리될 수 있다. 우선 인간은 누구나 자기 삶에 결핍된 보상들(혹은 자원들)을 획득하려 하는데, 이 보

32) Stark & Bainbridge (1987), pp. 27-36.

상들은 여러 종류로 분화되어 있을뿐더러 개중에는 공급이 부족하거나 아예 공급원 자체가 존재하지 않는 것들도 있다. 즉 삶에 필수적으로 요청되는 보상들 가운데 어떤 것들은 개인의 역량으로는 확보하기가 불가능하고, 혹 확보할 수 있다 하더라도 과도한 비용을 요구하기에 각 개인은 이 문제를 해결하기 위해 자원의 교환이라는 사회적 행위를 반드시 이행하게 된다. 그리고 각 보상들은 대부분 소모성 자원들이기에 이들을 얻기 위한 사회적 행위는 삶이 유지되는 한 지속적으로 수행되어야 한다.

여기서 베인브릿지가 수행한 기계학습 시뮬레이션의 기본 시나리오, 즉 화성 이주 시나리오를 재차 상기해 보자. 이 시나리오는 분명 방금 설명한 사회적 공리, 즉 인간의 삶이 기본적으로 개인의 행동만으로는 온전하게 유지될 수 없으며, 그러하기에 인간이 존재하는 곳에는 반드시 사회적 행위가 이행된다는 공리를 전제로 삼고 작성된 것이다. 실제로 베인브릿지는 이 시나리오를 기술하는 가운데 각 요소들이 어떤 공리에 근거해 설정된 것인지 명시하고 있다. 가령 “각 이주민들은 4개의 보상들(에너지, 물, 공기, 식량)을 각각 일정량씩 소비한다(공리 6)”거나 “이주민들은 서로 다른 종류의 보상 한 단위의 가치를 각기 다르게 [각자의 시급함에 따라] 평가하며, 이로 인해 서로에게 보상을 수여하는 교환행위를 수행할 수 있다(공리 2, 공리 3)”<sup>33)</sup>는 진술에서 확인할 수 있듯이, 베인브릿지의 화성이주 시나리오는 상기 제시한 공리 가운데 공리 1부터 공리 6까지의 내용을 인간행위의 전제로 삼고 있으며, 중국에는 보상을 획득하고 영위하려는 인간의 모든 행위가 사회적 행위로 귀결된다는 논리를 제시하고 있다.

베인브릿지의 시나리오를 지탱하는 이 공리적 전제는 종교사회학의 창도자 가운데 일인인 뒤르켐의 종교론을 상기시킨다. 뒤르켐 역시 종교문화의 발생과 성장에는 사회적 행위에 수반되는 집단적 종교표상이 반

33) Bainbridge (2006), p. 122.

드시 요청된다는 전제를 옹호한다. 그리고 이를 바탕으로 모든 종교현상의 실재를 해당 종교현상들을 유발하는 인간집단의 영역인 사회로 지목하는 사회실재론적 종교론을 개진한다. 그의 종교론을 대표하는 텍스트로 널리 알려진 『종교 생활의 원초적 형태』(*Lés Formes Élémentaires de la Vie Religieuse*)의 서론은 그의 사회실재론적 종교론을 위한 종교학 방법론을 제시하는데, 이 방법론 가운데는 앞서 약속한 베인브릿지의 종교문화 발생 시나리오와 그 전제에 상당한 정도로 부합하는 내용이 포함되어 있다.

우리가 관찰할 수 있는 가장 원초적 종교체계(*un système religieux qu'il est le plus primitif qu'il nous soit donné d'observer*)는 다음의 두 가지 조건을 충족시켜야 한다. 첫째, 다른 어떤 곳보다도 단순한 조직을 가진 사회 가운데서(*dans des sociétés dont l'organisation n'est dépassée par aucune autre en simplicité*) 발견되어야 한다. 둘째, 그보다 앞서 존재하는 종교로부터 어떠한 요소도 도입하지 않은 채로(*sans faire intervenir aucun élément emprunté à une religion antérieure*) 설명될 수 있어야 한다. ... [이처럼] 고대의 [원초적] 종교를 본 연구의 주제로 삼는 이유는 이런 종교가 인간의 종교적 본성(*la nature religieuse de l'homme*)에 대한 이해를 돕는 데 있어서, 다시 말해 인간성의 본질적이고 항구적인 측면(*un aspect essentiel et permanent de l'humanité*)을 현시하는 데 있어서 다른 어떤 종교보다도 적합하게 여겨지기 때문이다.<sup>34)</sup>

상기 진술에서 뒤르켐은 “가장 원초적 종교체계”를 관찰하기 위한 두 개의 조건을 제시한다. 첫째는 가장 단순한 구조를 가진 사회를 관찰해야 한다는 것, 둘째는 그보다 앞서 존재하던 다른 종교문화 요소를 학습, 수용, 모방하지 않은, 순수하게 해당 공동체 내에서 자체적으로 발생한 단 하나의 종교를 믿는 태고의 원시적 사회를 관찰해야 한다는 것이다.

34) Durkheim (1960), pp. 1-2.

베인브릿지의 화성이주 시나리오는 이 두 가지 조건을 충족한다고 볼 수 있다. 그의 시나리오에 등장하는 사이버그라는 공동체는 구성원들의 생명유지 및 공동체 존속을 위해 필수적으로 요청되는 5개 자원(에너지, 물, 공기, 식량, 성애)을 생산하는 4개 집단으로 구성되어 있다. 이처럼 순전히 생존을 위해 요구되고 생산되는 자원의 종류에 따라 단순하게 구축된 사이버그의 사회구조는 극단적으로 단순화된 원시적 사회구조를 관찰하려 하는 뒤르켐의 방법적 의도에 부합하는 관찰 대상이 될 수 있을 것이다. 여기에 더해 베인브릿지의 시나리오는 사이버그가 형성되기 시작한 화성이주 개시 시점에 이 공동체 안에 어떤 다른 종교문화적 요소도 선행적으로 존재하지 않는다는 설정을 따른다. 이로써 확인되는 바는 뒤르켐과 베인브릿지 두 사람 모두 종교적 신앙의 발생 시점을 관찰하기 위한 사회적 조건을 원시인류의 단순하고 초기적인 사회구조 속에서 길어내려 한다는 사실이다.

물론 베인브릿지의 화성이주 시나리오에 등장하는 사이버그는 과학기술 측면으로만 보면 원시인류와는 아무 상관이 없다. 오히려 화성이주 계획을 실행하지 못하고 있는 현재의 인류보다 더 진보된 문명을 갖고 있는 공동체임에 틀림없을 것이다. 그러나 종교문화 측면으로 본다면 사이버그의 구성원들은 사실상 원시인류가 처해있던 것과 매우 흡사한 개인적·사회적 정황에 처해 있다. 그들은 극단적으로 불친절한 화성의 환경 속에서 우주의 위협적 힘을 절감하며 살고 있다. 그들은 생존을 위해 투쟁하고 있으며 이런 투쟁적 정황은 그대로 사이버그의 단순화된 사회구조를 결정한다. 그리고 사이버그 구성원들은 그들을 초월해 있는 우주의 두려운 힘과 죽음에 대한 불안을 항시적으로 경험한다는 것 외에는 어떤 다른 종교문화적 요건도 갖추고 있지 않다.

여기서 한 가지 짚고 넘어갈 사안은 뒤르켐이 천명한 종교학의 방법적 조건과 스타크-베인브릿지 종교론의 공리적 전제 사이에 공통점만 발견되는 것은 아니라는 점이다. 양측에는 간과할 수 없는 분명한 차이점이

존재한다. 뒤르켐은 확고한 사회실재론적 종교이해를 고수하는 가운데 각 개인이 가진 존재의 열망을 종교성 현성의 조건에서 배제하는 반면, 스타크와 베인브릿지는 사회명목론적 종교이해를 부분적으로 수용해 접목하는 가운데 각 개인이 가진 초월의 욕망을 종교성 현성의 근본조건으로 제시하고 있다.<sup>35)</sup> 뒤르켐의 경우 종교성의 모든 현상적 양태들이 사회라는 실재를 기반으로 할 때만 구현될 수 있다고 확신한다. 그러므로 그의 시각으로 본다면 죽음을 극복하고 자신의 존재를 보존하려 하는 개인의 초월 열망은 사회적 행위의 원인이 아니라 결과에 속한다. 반면 스타크와 베인브릿지에게는 존재적 초월 열망의 양태를 띤 개인의 종교성이 사회적 행위의 결과가 아니라 그것의 주된 원인으로 간주된다.

이처럼 집단적-사회적 행위와 개인적-종교적 열망의 인과적 위상에 대한 양측의 이견은 본 연구의 논지와 관련해 한 가지 의문을 유발한다. 애초 종교문화 발생의 원천을 서로 다르게 지정하고 있는 베인브릿지와 뒤르켐의 종교사회학 체계를 두고 굳이 양측의 유사성을 찾아 종교적 인공지능 시뮬레이션을 반성하고 고찰할 만한 정당성이 충분히 존재하는가? 여기에 대해서는 다음과 같은 해명을 제시하고 싶다. 베인브릿지의 종교문화 발생 시나리오와 종교적 인공지능 시뮬레이션 가운데서는 자기존재 보존을 열망하는 개인 차원의 종교성과 이 종교성을 현실화하는 사회적 행위로서의 자원교환이 선후 인과관계가 분명치 않은 순환적 관계를 형성하고 있는 것으로 판단된다.<sup>36)</sup> 물론 스타크-베인브릿지 종교론

35) 베인브릿지는 각 개인이 죽음을 초월하려는 욕망을 가지며, 종교들이 이 욕망을 충족시키는 초자연적 보상에 대해 자세한 설명들을 제공한다고 밝힌다. Bainbridge (2006), p. 98.

36) 이러한 판단과 관련해서는 랜달 콜린스(Randall Collins)의 견해를 잠시 살펴보는 것이 논의에 바람직할 것으로 여겨진다. 콜린스는 스타크-베인브릿지 종교론과 뒤르켐의 종교론을 분석적으로 맞대어 비교하는 가운데 다음과 같은 견해를 밝힌다. “스타크-베인브릿지 [종교]모형은 부분적으로 [한 사회에 속한] 구성원의 유대감이라는 동력에 결정적으로 의존하고 있다. 그리고 두 사람의 신도모집 모형은 사회적 결속이 [개인들의] 믿음에 앞선다는 것을 보여준다. 이 믿음들은 초월적 보상 수여

의 공리는 보상을 획득하고 비용을 회피하려는 개인의 생존욕망이 종교 문화 발생의 최초 출발점이라고 전제하고 있으나, 이 공리를 바탕으로 기획된 베인브릿지의 시나리오는 이런 욕망이 종교성으로 전환되기 위해 반드시 사회적 행위가 동반되어야만 한다는 것을 암묵적으로 수긍하고 있는 것으로 판단된다. 앞서 살펴본 바대로라면 사이버그 구성원들의 자원교환 행위는 생명 자체라는 자원이 그들 자신들로부터는 결코 산출될 수 없는 것이라는 사실을 깨닫게 하며, 이로부터 사이버그에 속한 이들은 이 자원을 수여할 수 있는 원천으로 자연적 교환상대가 아닌 초자연적-초월적 교환상대를 상정하기 시작한다. 즉 생명을 유지하고 삶을 연장하려는 개인적 욕망 자체만으로는 아직 종교성이 현성할 수 없고 이 욕망에 의해 즉시적으로 추동되는 사회적 행위에 의해 비로소 초자연적 존재자에 대한 정보를 추구하고 산출하는 종교적 행위가 수행되는 것이다. 이 경우 종교성과 종교문화의 원천이 사회적 행위라고 주장하는 뒤르켐의 종교론과 개인의 욕망이라고 주장하는 스타크-베인브릿지 종교론 간에 명료한 진위여부를 가리는 일은 다소간 무의미한 처사가 되고 만다. 어떤 욕망의 발현과 그 욕망에 따른 행위 수행은 대개 동시적으로 일어나는 경우가 다반사인 데다가, 때로 그런 행위를 수행하는 다른 사회구성원을 목격하고 거기에 결부된 욕망을 학습하는 일도 빈번하기 때문에, 현실적으로 종교성의 기원이 되는 개인의 존재적 열망과 이를 추

---

자들의 영역이 된다. [이에] 나는 스타크-베인브릿지 이론이 사회적 유대감이라고 하는 제의적-감정적 기제를 보다 명시적으로 다루는 방향으로 수정된다면 보다 강화된 이론이 될 수 있음을 보이고자 한다.” 콜린스는 스타크-베인브릿지 종교론이 뒤르켐의 기능주의적 종교론과 반대의 입장에서 개인적 영역을 중시하는 성향을 갖지만, 종교문화의 형성과 진화에 있어서 사회적 연대감, 사회적 영역에서의 활동이 담당하는 역할을 결코 무시하지 않을 뿐 아니라 그것의 필수성을 암시적으로나마 전제하고 있는 것으로 풀이하고 있다. Randall Collins (2016), “Stark and Bainbridge, Durkheim and Weber: Theoretical Comparisons”, *Rational Choice Theory and Religion: Summary and Assessment* (ed. Lawrence A. Young), New York: Routledge, p. 163.

구하는 사회적 행위 사이의 인과적 선후관계는 불투명한 것이 되고 만다. 양자의 관계는 인과적이라기보다 차라리 순환적이고 상호구성적이라는 것이 보다 적절한 표현일 것이다.

이런 사실은 이 화성이주 시나리오를 토대로 구현된 종교적 인공지능 시뮬레이션의 알고리즘을 살펴보면 보다 구체적으로 확인된다. 앞서 살펴본 베인브릿지의 시뮬레이션 알고리즘은 욕망의 대상이 되는 6개의 자원 범주, 그리고 이 자원을 획득하기 위해 상정하는 3개의 교환대상 집단 범주 3개를 뼈대로 삼는다. 이 가운데 전자의 6개 범주는 상당부분 개인적 욕망에 의해 의식에 적용되는 범주들로 성격규정될 수 있다. 반면 후자의 3개 범주는 개인적 욕망만으로는 의식에 적용될 수 없다. 이는 오직 교환행위가 이루어지는 사회적 영역을 전제로 삼을 때만 초자연자 의식 산출에 조력한다. 이로 보건대 베인브릿지는 이론적 지평에서는 개인의 존재적 열망을 이와 연관된 사회적 행위에 앞서는 종교성 형성의 원천으로 지목하나, 현실적 지평에서는 양자의 동위적 필수성을 전제로 삼는 가운데 종교성의 원초적 본질과 종교문화의 발생과정을 탐구해야 한다는 입장을 암묵적으로 고수하고 있는 것으로 판단된다.

이처럼 사회적 행위가 산출하는 범주들이 종교적 의식의 기본조건이라는 인식은 뒤르켐의 종교학 방법론에서도 대단히 중시되는 요소 가운데 하나이다. 그는 사회적 행위로부터 산출되는 집단적 표상과 종교적 의식을 산출하는 범주들 사이의 관계를 두고 다음과 같은 견해를 제시한다.

종교는 이미 형성되어 있는 인간 지성을 보다 풍성하게 만들기만 한 것이 아니라 실상 지성의 형성에까지(à le former) 지속적으로 기여해 왔다. 인간들은 종교 덕분에 그들이 향유하는 지식의 내용만 아니라 그 지식을 정교하게 만드는 형식도 갖출 수 있게 되었다. 우리들의 판단의 뿌리에는 우리의 지적인 삶 전체를 지배하는 특정한 본질적 개념들(notions essentielles)이 존재한다. 아리스토텔레스를 위

시한 철학자들은 이 이념들을 이해의 범주들(*les catégories de l'entendement*)로 호명해 왔다. 여기에는 시간, 공간, 수, 원인, 실체, 인격 등과 같은 개념들이 포함된다. 이것들은 사물들의 가장 보편적인 특성들(*les propriétés les plus universelles des choses*)에 상응한다. 이것들은 마치 사고를 구획하는 확고한 얼개와 같은 것이다. ... 방법적인 체계를 갖추고서 원초적인 종교적 믿음들을 분석해 본다면 자연스레 그것들 사이에서 주요 범주들을 발견하게 된다. 이 범주들은 종교 안에서 그리고 종교로부터(*dans la religion et de la religion*) 태어난 것들이다.<sup>37)</sup>

상기 진술에서 확인되듯 뒤르켐은 종교를 단지 정신문화의 결과물 가운데 하나 정도로 간주하지 않는다. 그는 종교가 문화의 모든 지적·정신적 내용을 결정하는 준거로서 인식의 범주들을 산출하는 근원적 위상을 점유한다고 소개하고 있다. 이 범주들은 “사물들의 가장 보편적인 특성들”을 가려내는 능력, 즉 본질적인 것(*l'essentiel*)을 구별할 능력을 제공함으로써 지적 체계 전체의 구조를 건립하는 주역을 담당한다. 이처럼 뒤르켐의 이론 안에서 종교는 모든 인식 준거의 원천으로서 인간의 지적 활동 전체에 빠짐없이 관여하는 근원적 인문 요소로 자리매김하고 있다. 그런데 이 대목에서 그는 다시금 종교를 “사회의 산물”로 성격규정함으로써 종교적 의식과 그에 의해 지배되고 조정되는 인간의 지식 전체가 사회라는 실재를 지시한다는 견해를 제시한다.

종교는 분명히 사회적 산물이다(*la religion est une chose éminemment sociale*). 종교적 표상들은 집단적 실재들을 표현하는 집단적 표상들이다. 제의들(*les rites*)은 군집한 집단들 가운데서만 생겨나는데, 그 목적은 이 집단들의 특정한 정신적 상태들을 자아내고, 유지하고, 재생하려는 것이다. 그러나 만일 이 범주들이 종교에 기원을 두고 있

37) Durkheim (1960), pp. 12-13.

다면, 이것들은 모든 종교에 공통적인 어떤 것을 공유하고 있어야만 한다. 동시에 이 범주들은 사회적 산물(les choses sociales), 집단적 사고의 산물이 되어야만 한다(les produits de la pensée collective).<sup>38)</sup>

여기에서 종교를 사회의 산물로 성격규정할 수 있는 근거로는 종교문화의 가장 기본적 형태로서 제의의 성격과 목적이 제시된다. 뒤르켐이 관찰한바 제의의 본질은 집단성이다. 종교와 그것의 원초적 형태인 제의는 오직 집단의 행위일 때에만 그 진정성을 인정받을 수 있다. 그는 종교가 담지하는 이런 집단적 본질이 구현된 사회를 교회(une Église)라 명명하고 그 특징에 대해 다음과 같이 기술한다.

한 사회의 구성원들이 성스러운 세계 및 그것이 속된 세계와 맺는 관계들(le monde sacré et ses rapports avec le monde profane)을 동일한 방식으로 상상하고, 이런 집단적인 표상을 동일한 실천방식들에 따라(dans des pratiques identiques) 이행함으로써 그 구성원들이 연합되어 있는 사회, 이런 사회가 바로 교회라 불리는 것이다(c'est ce qu'on appelle une Église). 우리는 역사 속에서 교회 없는 종교를 발견할 수 없다.<sup>39)</sup>

사회적 행위, 종교성, 그리고 인식범주 사이의 관계에 대한 뒤르켐의 이런 견해는 베인브릿지 시뮬레이션 알고리즘의 한 축을 담당하고 있는 3개의 교환대상집단 범주들이 갖는 종교사회학적 함의를 조명하고 그 정당성을 지지하는 것으로 해석될 수 있을 것이다. 이 범주들의 가장 중요한 함의는 사이버그라는 공동체에 속한 기계학습 에이전트 모두가 생존을 위해 반드시 수행해야 하는 자원교환 행위를 개인 대 개인의 행위 이전에 집단 대 집단의 행위로 성격규정하도록 설정되었다는 점이다. 각

38) Durkheim (1960), pp. 13-14.

39) Durkheim (1960), p. 60.

에이전트는 사이버그 전체가 최대 4개의 교환집단들로 구성되어 있으며, 각자가 그 가운데 하나의 집단에 포함되어 있어 교환행위를 수행할 때 항상 다른 교환대상집단을 우선 상정하고 그 다음에 교환을 성사시키기에 적합한 개별 구성원을 탐색한다고 하는 선형적인 학습의 준거 혹은 조건을 갖는다.

베인브릿지는 이러한 설정이 적용된 이유를 다음과 같이 설명한다. 인간이 삶을 영위하는 데 필수적인 보상들은 자력으로 획득하는 것보다 사회적 교환행위에 의해 획득하는 것이 보다 용이하고 비용을 적게 발생시킨다. 잠시 인간 현실을 돌아보면, 때로는 교환행위가 아니고서는 거의 획득하기 불가능한 보상들도 존재한다. 여기에 더해 교환행위는 유형의 보상뿐 아니라 무형의 보상인 설명들도 획득할 수 있는 기회를 제공한다. 그리고 인간이 보상을 추구하는 동시에 이 보상을 획득하기 위해 합리적인 의사결정을 내리는 존재자인 한, 누구나 반드시 사회적 행위를 통해 삶을 영위하게 된다. 이 경우 인간은 항상 자신이 속해 있는, 자신과 동류의 보상을 확보하고 있는 사회집단(동일한 보상들을 확보하고 있기에 교환대상집단이 될 수 없는 집단)과 자신이 확보하지 못한 보상들을 확보하고 있는 사회집단들(자신에게 결핍된 보상들을 확보하고 있기에 즉시적으로든 잠정적으로든 교환대상집단이 될 수 있는 집단들)을 구분하는 것으로 교환행위를 시작할 것이다. 그리고 이 교환대상집단의 종류와 성격은 교환행위가 반복되는 가운데 수행하게 되는 관찰과 학습에 의해 세분화되게 될 것이다. 베인브릿지는 이런 추론을 바탕으로 인간이라면 누구나 그의 삶 전체를 통해 그 정도가 크든 작든 사회적 교환행위에 의존해 생존하게 된다는 결론을 내리고 있다.<sup>40)</sup> 이로써 인간의 삶은 필연적으로 장기적이고 지속적인 사회적 행위에 노출되고 포섭되는 것으로 판명된다. 그리고 그는 이런 판단에 따라 각 기계학습 에이전트의

40) Bainbridge (2006), pp. 118-119.

자원교환 행위를 그 본질상 사회적 행위로 성격규정하는데, 이런 처사는 뒤르켐의 관점으로 볼 때 충분히 타당하게 받아들여질 여지가 있다. 베인브릿지가 그의 시나리오와 알고리즘 가운데 설정한 자원교환 행위는 집단 내에 속한 개인들 간에 반드시 집단에 의해 매개되는 지속적인 소통과 상호작용이 이루어져야 한다는 것을 전제한다는 점에서, 그리고 이런 소통과 상호작용의 축적이 중국에는 종교적 의식의 집단적인 학습으로 이어진다는 점에서 분명한 사회적 행위라 볼 수 있다. 이 같은 판단을 뒷받침할만한 뒤르켐의 진술로는 다음을 눈여겨보면 좋을 것이다.

“개인적 부류의 표상들과 사회적 부류의 표상들이 형성되는 방식은 이 둘 사이의 차별화를 초래한다. 집합적인 표상들은 거대한 협동의 산물로서 공간뿐만 아니라 시간을 통해서도 연장되고 있다(*qui s'étend non seulement dans l'espace, mais dans le temps*). 즉 이런 표상들을 산출하기 위해 다수의 서로 다른 정신들이 그들이 가진 이념과 감정을 맞붙이고 섞어 결합하였다(*ont associé, mêlé, combiné leurs idées et leurs sentiments*). 그들은 여러 세대에 걸쳐 경험과 지식을 축적해 왔다. 그러므로 개인의 지성에 비한다면 거의 무한하다 할 만큼 풍부하고 복잡한, 매우 특수한 지성(*une intellectualité très particulière*)이 그들 속에 정제된 채 놓여 있다. 이러한 사실을 통해 우리는 어떻게 이성이 경험적 인식의 범위를 능가하는 능력(*le pouvoir de dépasser la portée des connaissances empiriques*)을 보유하게 되었는지 이해하게 된다.”<sup>41)</sup>

그러나 여기서 한 가지 주의해야 할 점이 있다. 이는 베인브릿지와 뒤르켐 사이에서 여전히 목격되는 인과적 선후관계의 차이이다. 베인브릿지는 개인의 존재적 열망과 사회적 행위가 서로 얽혀드는 가운데 비로소 종교성이 현성한다고 보는 반면, 뒤르켐은 사회적 교류가 우선 종교

41) Durkheim (1960), pp. 22-23.

문화를 발생시키고 이 종교문화가 제반 인식의 준거로서 범주들을 제공한다고 주장한다. 전자의 경우에는 범주가 우선하고 종교가 뒤따르는 반면, 후자의 경우에는 종교가 우선하고 범주가 뒤따른다.

우선 베인브릿지가 인식범주를 우선시하는 이유로 종교성의 ‘자율적’ 현성을 재현해보려는 시뮬레이션 의도를 지목해볼 수 있다. 여기에는 선행하는 어떠한 종교문화 요소도 존재하지 않는 상황이 전제된다. 다시 말해 전적인 종교 부재의 상황이 설정되는 것이다. 이로써 그가 목표하는 바는 자신이 제시한 종교문화 발생의 원인과 과정이 타당한 것으로 받아들여질 수 있다는 것을 인공지능 시뮬레이션을 통해 증명하려는 것이다. 반면 뒤르켐은 해당 문화의 성격을 결정짓는 ‘하나의’ 종교문화 외에 다른 어떤 종교문화 요소도 존재하지 않는 상황을 확보하려 한다. 즉 종교 부재의 상황이 아니라 다른 종교문화 요소와 혼합된 적 없는 하나의 ‘순수한’ 원시적 종교가 존재하는 상황을 상정하는 것이다. 이는 그의 연구가 종교문화의 발생원인 자체보다는 종교의 사회적 본질과 사회적 기능을 조명하는 데 초점을 맞추고 있기 때문이다.

연구의 초점과 목적 외에도 두 사람의 연구현실이 보이는 차이 역시 고려하지 않을 수 없다. 뒤르켐의 연구는 상황을 임의적으로 설정하고 조작할 수 있는 시나리오 기반 시뮬레이션에 바탕을 둔 것이 아니라 이미 종교문화가 확고하게 형성되어 있는 현실상의 문화현상 관찰에 바탕을 두고 있다. 뒤르켐이 주로 관찰한 원시문명이나 고대문명 가운데서는 거의 예외 없이 종교가 사물과 경험의 초월적 원리들을 해명하는 데 지배적인 영향력을 행사했기 때문에, 그의 이론은 인식범주의 성격규정에 있어 전적인 종교 부재의 상황을 상정하기 어려운 정황에 놓여 있다. 결국 종교와 인식의 범주 간 선후관계를 두고 발생한 베인브릿지와 뒤르켐 사이의 이견은 각기 설정한 연구의 목적 및 각자에게 배정된 연구 현실의 차이에 기인한 것이라 볼 수 있다. 뒤르켐 입장에서는 종교가 선행하지 않는, 그리고 종교를 그 원인으로 지목하지 않는 베인브릿지의 인지

범주 이해가 결코 타당하다고 평가하기 어려울 것이다. 베인브릿지의 시뮬레이션은 분명 원시인류 종교문화의 현실과 일정한 괴리를 보이는 임의적 성격을 갖는다. 이는 그가 에이전트의 학습을 위해 특정한 선택적 학습범주 설정을 필수적으로 요청하는 강화학습 시뮬레이션 방식을 선정했기 때문이다. 결국 이 문제는 설계자가 에이전트의 자율적인 학습을 위해 반드시 보상과 처벌의 내용, 그리고 이 보상의 획득을 위해 대응해야 할 환경을 선행적으로 설정해야 하는 강화학습 시뮬레이션의 방법적 한계 혹은 기술적 한계로 볼 수 있다.<sup>42)</sup> 그러나 이런 한계에도 불구하고 다음을 간과해서는 안 될 것으로 여겨진다. 바로 베인브릿지와 뒤르켐 모두 종교문화의 발생과 인지범주 출현의 인과적 선후관계와 상관없이 사회적 행위가 종교적 의식과 인지범주의 발원지라는 점에 대해서는 공감대를 형성하고 있다는 점이다. 결국 두 사람 모두 사회와 그 내부에서 진행되는 사회적 행위를 바탕으로 종교적 의식이 발생한다는 점에서는 분명히 의견의 일치를 보이는 셈이다.

#### 4. 나오는 말

이처럼 스타크-베인브릿지와 뒤르켐의 종교론 사이에 다소 중대한 이견들이 존재하긴 해도, 전체적으로 볼 때 베인브릿지의 인공지능 시뮬레

42) 기계학습의 한 종류인 강화학습은 에이전트의 행위에 대한 보상-처벌을 통해 학습을 진행한다. 그러므로 보상-처벌의 내용과 이 보상을 추구하기 위해 에이전트가 대응해야 할 환경이 시뮬레이션 설계자에 미리 임의적으로 설정되어야만 한다. 학습의 과정에서 에이전트가 향유할 수 있는 자율성의 범위는 주어지는 보상과 이를 둘러싼 환경에 크게 의존하고 있다. 이런 특성은 특정한 환경에 의존하지 않고 데이터를 직접적으로 대별하며 학습해가는 지도학습(supervised learning)이나 비지도학습(unsupervised learning)과는 확연히 구별되는 것이다. Richard S. Sutton & Andrew G. Barto (2018), *Reinforcement Learning: An Introduction*, Cambridge: The MIT Press, pp. 1-3.

이전에 반영된 통찰들은 뒤르켐의 종교학 방법론 관점으로 볼 때 상당한 정도의 타당성을 가진 것으로 인정될 가능성이 높다. 무엇보다 원초적 종교성의 발원지를 사회로부터, 지속적으로 수행되는 사회적 행위로부터 찾고, 이 사회적 행위를 종교적 인지의 범주적 준거로 전환하는 방식을 통해 인간의 종교적 본성을 탐구한다는 점에서 양측은 공감대를 형성하고 있는 것으로 확인된다. 물론 베인브릿지의 시뮬레이션은 인간의 현실적 삶의 경험들이 내어놓는 다변적이고 풍성한 종교성 전반을 포용하지 못하는 한계, 다소 인위적이라 할 수 있는 설정의 한계, 그리고 기계 학습에 필히 관여될 수밖에 없는 기술적 한계를 안고 있다. 그럼에도 불구하고 인간의 사회적 행위가 어떻게 종교문화 발생의 근거가 되는지, 그 인식론적 과정을 한편으로는 공리들과 정의들을 통해 합리적인 방식으로, 다른 한편으로는 분산이라는 통계기법을 통해 확률적인 방식으로 구현해낸 점은 분명 상당한 가치를 지니는 종교학적 성과로 인정될 수 있을 것이다. 무엇보다 인공지능이 기술영역 뿐만 아니라 문명, 사회, 그리고 중국에는 인간의 삶의 내밀한 영역에까지 관여하며 지배적인 영향을 줄 것이 확실시되고 있는 작금의 현실에서, 인공지능을 통해 종교성 현성 시뮬레이션을 수행한 베인브릿지의 연구는 향후 다가올 세대의 종교문화가 보이는 특성들을 고찰하는 한편, 혹시 발생할지 모를 인공지능만의 종교양태를 전망해 보는 데 상당한 기여를 하고 있다고 평가될 수 있을 것이다.

현재 인공지능 기술의 동향은 종교적 인공지능 시뮬레이션에서 채택된 것과 같은 다중 에이전트 시스템(multi-agent system)의 현실화 단계를 눈앞에 두고 있다. 정보기술 분야에서 독보적 지위를 확보하고 있는 컨설팅 업체 가트너는 최근 향후 인공지능 기술이 자율 사물(autonomous things), 스마트 스페이스(smart spaces), 그리고 인공지능 주도 개발(AI-driven development) 등으로 특징지어질 것이라 전망한 바 있다.<sup>43)</sup> 이 개념들을 종합해 보면 향후의 인공지능 기술은 개별 에이전트들이 각

각 탑재된 기기 속에서 개별적으로 특화된 업무만 수행하는 데서 벗어나, 네트워크를 통해 교류하는 가운데 서로를 복제하고 검토함으로써 거대한 집단적 지식과 정보처리 능력을 도출하는 데 이를 것으로 보인다. 기계학습 기술은 이런 조류를 주도할 기술로서의 입지를 확보하고 있다. 만일 기계학습의 보상-처벌 기제가 범지구적 네트워크 속에서 작동할 경우 인공지능에 의한 종교의 창도는 결코 더 이상 상상만으로 그치는 사안이 되지 않을 것이다. 베인브릿지의 시물레이션은 이런 전망을 입증하는 하나의 증거로 삼기에 부족해보이지 않는다. 이 실험은 향후 종교계 및 종교학계 전체의 거대담론으로 부각될지 모를 인공지능의 종교라는 사안에 대해 경각심과 과제의식을 일깨운다는 점에서 상당한 의미를 부여할 수 있을 것이다.

- 
- 43) 자율사물이란 개념은 일상적으로 사용되는 대다수의 기계적 물품에 인공지능이 탑재되어 제어기능을 담당하는 것을 말한다. 스마트 스페이스는 네트워크를 통해 자율사물에 탑재된 에이전트들이 서로 상호작용하는 인공지능 편재의 상황을 의미한다. 마지막으로 인공지능 주도 개발은 스마트 스페이스를 통한 상호작용 가운데 에이전트들이 서로를 복제하고 평가하며 수정함으로써 인간의 별다른 관여 없이 스스로 기능을 증강시키고 진화해 가는 것을 뜻한다. Jennifer Garfinkel (2018), “Gartner Identifies the Top 10 Strategic Technology Trends for 2019”, *Gartner Webpage*, Gartner, 2018. 10. 15. <https://www.gartner.com/en/newsroom/press-releases/2018-10-15-gartner-identifies-the-top-10-strategic-technology-trends-for-2019>.

## 참고문헌

### 【자 료】

- Bainbridge, William Sims (2006), *God from the Machine: Artificial Intelligence Models of Religious Cognition*, Lanham: Altamira Press.
- Durkheim, Émile (1960), *Les Formes Élémentaires de la Vie Religieuse: Le Systématique en Australie*, Paris: Presses Universitaires de France.
- Stark, Rodney and William Sims Bainbridge (1987), *A Theory of Religion*, New Brunswick: Rutgers University Press.

### 【논 저】

- 김영진(2017), 「인공지능, 종교적 명상, 그리고 마음의 지향성: ‘명상하는 AI’의 의미와 가능성에 대한 철학적 고찰」, 『철학·사상·문화』 24, 동서사상연구소.
- 김진(2018), 「인공지능시대에도 종교는 가능한가?: 칸트와 블로흐의 종교철학적 관점에서」, 『철학연구』 147, 대한철학회.
- 문장수(2018), 「사이버네틱스의 인식론과 메타-과학적 지위」, 『철학논총』 94, 새한철학회.
- 박찬국(2018), 「인간과 인공지능의 미래: 인간과 인공지능의 존재론」, 『현대유럽철학연구』 50, 한국하이데거학회.
- 이창익(2017), 「인간이 된 기계와 기계가 된 신: 종교, 인공지능, 포스트휴머니즘」, 『종교문화비평』 31, 종교문화비평학회.
- 정재현(2019), 『우상과 신앙』, 파주: 한울아카데미.
- 정재현, 박욱주(2018), 「사이버네틱스(cybernetics)의 현상학적 전환」, 『철학논총』 94, 새한철학회.
- Carter, Matt (2007), *Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence*, Edinburgh: Edinburgh University Press.
- Churchland, Paul M. (1993), “Representation and High-speed Computation in Neural Networks”, *The Foundations of Artificial Intelligence: A Sourcebook* (eds. Derek Partridge & Yorick Wilks), Cambridge: Cambridge

- University Press.
- Colledge, Richard (2014), “A-Z Glossary: Death”, *The Bloomsbury Companion to Existentialism* (eds. Felicity Joseph, Jack Reynolds & Ashley Woodward), New York: Bloomsbury.
- Collins, Randall (2016), “Stark and Bainbridge, Durkheim and Weber: Theoretical Comparisons”, *Rational Choice Theory and Religion: Summary and Assessment* (ed. Lawrence A. Young), New York: Routledge.
- Comstock, William Richard (1971), *The Study of Religion and Primitive Religions*, New York: Harper & Row.
- Copeland, Jack (1993), *Artificial Intelligence: A Philosophical Introduction*, Oxford: Blackwell.
- Eliade, Mircea (1987), *The Sacred and the Profane: The Nature of Religion* (tr. by Willard R. Trask), New York: Harcourt, Brace & World.
- Garfinkel, Jennifer (2018), “Gartner Identifies the Top 10 Strategic Technology Trends for 2019”, *Gartner Webpage*, Gartner, 2018. 10. 15. <https://www.gartner.com/en/newsroom/press-releases/2018-10-15-gartner-identifies-the-top-10-strategic-technology-trends-for-2019>.
- Idinopulos, Thomas A. (2002), “The Strengths and Weaknesses of Durkheim’s Methodology for the Study and Teaching of Religion”, *Reappraising Durkheim for the Study and Teaching of Religion Today* (eds. Thomas A. Idinopulos & Brian C. Wilson), Leiden: Brill.
- Kant, Immanuel (1956), *Kritik der reinen Vernunft*, Hamburg: Felix Meiner.
- McCorkle, Jr., W. William, and Dimitris Xygalatas (2014), “Introduction: Social Minds, Mental Cultures — Weaving Together Cognition and Culture in the Study of Religion”, *Mental Culture: Classical Social Theory and the Cognitive Science of Religion* (eds. Dimitris Xygalatas & William W. McCorkle, Jr.), New York: Routledge.
- Moor, James H. (2003), *The Turing Test: The Elusive Standard of Artificial Intelligence*, Dordrecht: Springer Science+Business Media.
- Nath, Rajakishore (2009), *Philosophy of Artificial Intelligence: A Critique of the Mechanistic Theory of Mind*, Boca Raton: Universal-Publishers.
- Patton, Kimberly (2000), “‘He Who Sits in the Heavens Laughs’: Recovering

- Animal Theology in the Abrahamic Traditions”, *Harvard Theological Review* 93, Harvard Divinity School.
- Salvaris, Mathew, Danielle Dean and Wee Hyong Tok (2018), *Deep Learning with Azure: Building and Deploying Artificial Intelligence Solutions on the Microsoft AI Platform*, New York: Apress.
- Simpson, Christopher Ben (2009), *Religion, Metaphysics, and the Postmodern: William Desmond and John D. Caputo*, Bloomington: Indiana University Press.
- Sutton, Richard S. and Andrew G. Barto (2018), *Reinforcement Learning: An Introduction*, Cambridge: The MIT Press.
- Turing, Alan M. (1950), “Computing Machinery and Intelligence”, *Mind: A Quarterly Review of Psychology and Philosophy* 59, Mind Association.
- Waldau, Paul (2006), “Seeing the Terrain We Talk: Features of the Contemporary Landscape of ‘Religion and Animals’”, *A Communion of Subjects: Animals in Religion, Science, and Ethics* (eds. Paul Waldau & Kimberly Patton), New York: Columbia University Press.
- Wong, Paul T. P., Gary T. Reker and Gina Gesser (1994), “Death Attitude Profile — Revised: A Multidimensional Measure of Attitudes toward Death”, *Death Anxiety Handbook: Research, Instrumentation, and Application* (ed. Robert A. Neimeyer), Washington, DC: Taylor & Francis.

원고 접수일: 2019년 4월 10일

심사 완료일: 2019년 5월 20일

계재 확정일: 2019년 5월 22일

ABSTRACT

---

Artificial Religious Intelligence Simulation:  
Comparative Study of the Stark-Bainbridge Model and Durkheimian Model  
of Religious Cognition

Park, Wook Joo\*

One of the important contemporary questions about the possibilities and limits of artificial intelligence is concerned with human religiosity. “Can artificial intelligence simulate religious faith?” The present study investigates the implications of a machine learning experiment of William Sims Bainbridge, an outstanding researcher in the field of the sociology of religion and cognitive science. He attempted to simulate human religiosity and faith in supernatural beings in 2006. The focus is on three factors: his social cognitive theory to explain human religiosity and faith, a scenario for the actualization of this theory, and a mathematical-statistical strategy and its principles applied to experiment. In this comparative analysis of the Stark-Bainbridge model of religion and religious cognition and that of Durkheim, it is demonstrated that the insights of Stark and Bainbridge found in Bainbridge’s artificial intelligence simulation would likely be acknowledged as valid from the perspective of the Durkheimian methodology of religious studies, despite some significant differences between

---

\* Adjunct Professor, The United Graduate School of Theology, Yonsei University

them. Most importantly, it seems certain that what Stark-Bainbridge and the Durkheimian model of religious cognition have in common is that they locate the origin of primitive religiosity in society, in continuous social processes. It also seems certain that they both translate the social processes into the categorizing norms of religious cognition so as to illuminate the religious nature of human kind.