

로봇이라는 타자와 휴머니즘의 재편

아이작 아시모프의 『로봇과 제국』(1985) 다시 읽기*

Robots as 'The Other' and Reconfiguration of Humanism
: Re-reading Isaac Asimov's novel *Robots and Empire*(1985)

장준영**

국문요약 인공지능 기술이 급속도로 발전함에 따라 인공지능 로봇과 인간의 공존 및 윤리 문제가 많은 논자에 의해 지적되었고 그 과정에서 로봇과 인간의 자리 설정 문제는 늘 도마에 올라왔다. 미국의 SF 작가 아이작 아시모프가 제시한 로봇공학 3원칙(Three Laws of Robotics)은 로봇의 활동 원칙과 그 활용에 대한 원안에 해당한다. 로봇공학 3원칙에 내재한 자체 모순과 윤리성은 작가 자신과 독자들에게 의해 지속적으로 질문되어 왔다. 그중 아시모프의 로봇 시리즈 완결작인 『로봇과 제국』은 로봇공학 3원칙에 제0원칙¹을 추가하는 방식으로 '인류'의 범주 설정과 감정, 기억의 문제를 포괄하고 있어 주목을 요한다. 이 글은 『로봇과 제국』에 등장하는 로봇 유형을 '도구로서의 로봇(솔라리아 행성에 남겨진 일반 로봇 전반기 감독 로봇 '랜드리)', '스스로 사유하여 로봇공학 제0원칙을 도출하는 로봇(다닐)', '공감하고 기억하는 텔레파시 로봇(지스카드)'으로 분류해 각각을 분석함으로써 아시모프가 로봇이라는 타자를 통해 기존의 차별과 위계에 따라 분열된 휴머니즘을 어떻게 재편하는지를 추적하였다. 2장에서는 솔라리아 행성의 감독 로봇을 분석 대상으로 삼았다. 이 로봇들은 로봇공학 3원칙에 저항하지 않는 로봇이지만, 우주인과 솔라리아인 그리고 이주자(지구인)을 억양에 따라 '인간'과 '인간 아닌 것'으로 자체 분류하여 그들에게 공격을 감행한다. 3장에서는 인간과 구분이 되지 않을 정도로 정교하게 만들어진 로봇 다닐의 이성적 추론능력을 분석하고 그것이 합리성을 따라 제0원칙이라는 새로운 원칙을 입안하는 과정을 살펴보았다. 4장에서는 작중 유일의 텔레파시 로봇인 지스카드의 공감과 기억 능력을 분석하여 로봇과 윤리, 감정의 문제를 재질문하였다. 이 글은 그동안 인공지능 기술 발전에 많은 화두를 던져온 고전 SF 작품을 포스트휴머니즘 관점에서 다시 읽고자 했다.

핵심어 아이작 아시모프, 인공지능, 로봇, 로봇공학 3원칙, 로봇과 제국, 로봇윤리, 포스트휴머니즘, 트랜스휴머니즘

* 이 논문 또는 저서는 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2017S1A6A3A01078538).

** 중앙대학교 국어국문학과 박사과정.

차례

1. 로봇이라는 타자, 인간의 동일성을 깨다
2. 인간 범주의 자의성을 폭로하는 도구로서의 로봇
3. 로봇공학 제0원칙을 도출한 스스로 사유하는 로봇
4. 접촉하고 공감하고 기억하는 텔레파시 로봇
5. 나가며: 능동적 허무주의로의 전회

1. 로봇이라는 타자, 인간의 동일성을 깨다

1939년부터 단편소설을 발표하기 시작한 아이작 아시모프 Isaac Asimov 는 『에스타운딩Astounding』지 1942년 3월호에 발표한 초기 단편 「런어라운드Runaround」¹⁹⁴²에서 ‘안전장치’로서의 로봇공학 3원칙Three Laws of Robotics 을 처음 제시하였다. 로봇공학Robotics이라는 단어가 아시모프에 의해 처음으로 고안된 것 역시 이 소설에서였다.⁰¹ 로봇공학 3원칙은 아시모프의 로봇 소설들이 로봇을 공포스럽고 거대한 위협이 아닌 긍정적인 발전으로 묘사할 수 있게 한 중요 설정이다. 사백 권이 넘는 저서 중 대부분이 논픽션이었던 아시모프에게 허구적 소설 창작은 미래 사회에 대한 예측뿐 아니라 현 인류와 다음 세대들이 다가올 기술에 대해 독자적인 판단을 갖게끔 이끄는 가이드 작성에 가까웠고⁰² 이를 위해서는 로봇이 인류에게 닥쳐온 위기이기보다 좋은 파트너일 필요가 있었을 것이다.

로봇공학 3원칙은 여러 평자에 의해 비판적으로 재고되었다. 로봇은

01 아이작 아시모프, 「독자들에게」, 정철호 역, 『강철도시』, 현대정보문화사, 2001, 6쪽.

02 Mona Håland Aarsland, *The cyborg as a posthuman figure in science fiction literature*, University of Stavanger MA Thesis, 2015, pp.41~43.

리Roboethics와⁰³ 로봇의 법인격 부여의 문제, 실제 인공지능 로봇에 3원칙을 프로그램할 때 생길 문제점의 파악과 대안적 원칙 제시⁰⁴ 등이 그것이다. 박인찬과⁰⁵ 유은순·조미라⁰⁶ 등 국내 논자들도 아시모프의 로봇공학 3원칙이 지닌 문제를 짚고 그로부터 인간과 기계의 관계, 인간의 윤리를 질문하는 데로 사유를 확장해 나갔다. 아시모프 자신도 소설 창작을 통해 로봇공학 3원칙을 끊임없이 다시 질문하였다. ‘로봇 시리즈’의 마지막 작품이자 그의 말년 작품인 『로봇과 제국Robots and Empire』¹⁹⁸⁵에

03 로봇윤리(Roboethics)라는 용어는 잔마르코 베루지오(Gianmarco Veruggio)에 의해 2002년 처음 제안되었다. 베루지오는 2006년 로봇윤리에 관한 로드맵을 제시하는 EURON 회의에서 로봇윤리학의 주요 의제 중 하나로 아시모프의 로봇공학 3원칙이 진정 ‘로봇의 윤리(Ethics of Robots)’로 받아들여질 수 있는 것인가에 대한 지속 탐구를 제시하였다. Gianmarco Veruggio, “The EURON Roboethics Roadmap”, EURON Roboethics Atelier-Genoa, 27th of February - 3rd of March, 2006, pp.23~24. (<http://www.roboethics.org/atelier2006/docs/ROBOETHICS%20ROADMAP%20Rel2.1.1.pdf>)

04 대표적인 학자로 로빈 머피와 데이비드 우즈를 들 수 있다. 머피와 우즈는 아시모프의 로봇공학 3원칙을 현대의 로봇에게 적용할 때 발생하는 모순점들을 확인하였다. 그들은 아시모프의 로봇공학 3원칙이 기능적이며 실용적인 도덕률(Functional Morality)에 불과하기에 그것이 복합적인 맥락과 판단에 열려있는 대화적인 새 법칙으로 개편되어야 함을 강조했다. 머피와 우즈가 제안한 대안적인 로봇공학 3원칙은 아래와 같다. 1. 인간은 안전과 윤리에 대한 가장 높은 수준의 법적·전문적인 기준을 충족하는 인간-로봇 작업 시스템이 마련되기 전까지 로봇을 홀로 방치하지 않는다. 2. 로봇은 그에게 부여된 임무를 적절히 수행하는 선에서 인간의 지시에 응답해야 한다. 3. 로봇은 첫째와 둘째 법칙을 준수하고 다른 행위자에게 제어권을 원활히 양도할 수 있는 한, 자기 존재를 보호하기 위하여 상황에 맞는 자율성을 충분히 부여받아야 한다. Robin R. Murphy and David D. Woods, “Beyond Asimov: The Three Laws of Responsible Robotics”, *Intelligent Systems*, IEEE September, 2009, pp.14~19.

05 아시모프의 『바이센테니얼 맨』에 드러난 로봇과 인간의 관계 방식을 고찰한 박인찬은 아시모프의 공진화 방식이 로봇을 친숙한 대상으로 ‘길들이는’ 인간화를 전제로 하며, 이때 로봇의 타자성이 철저히 무시되고 있고 비인간으로서의 로봇이라는 존재가 아시모프의 보수적인 포스트휴먼 서사에서 허용되지 않는다고 보았다. 아시모프의 로봇과 휴머니즘 담론의 보수성은 작가의 다른 작품들에서도 발견되는 특징이며 이는 세릴 빈트 등의 학자들에 의해서도 여러 차례 지적된 바 있다. 그러나 아시모프의 초기 작품만을 다루고 있는 박인찬의 연구가 로봇공학 3원칙에 대해 내리는 판단은 작가의 말년 작품에서 발견되는 변화의 징후까지를 온전히 포착하지 못한 불충분한 이해로 남을 수 있다는 것 역시 함께 언급되어야 할 것이다. 박인찬, 「포스트휴먼으로 가는 길 - 인간과 기계의 공(共)진화를 중심으로」, 『안과밖: 영미문학연구』 43, 영미문학연구회, 2017, 26~27쪽.

06 유은순과 조미라는 아이작 아시모프의 로봇공학 3원칙에 대한 제고찰의 필요성에 대하여 그것이 전통적인 ‘인간’ 개념이 해체되고 트랜스휴먼, 네오휴먼, 포스트휴먼 등 다양한 ‘인간 이후의 인간’들을 발견하고 있기 때문이라고 보았다. 그들은 인간이 만든 로봇이 결국 문제 해결의 수단이나 도구를 넘어 인간과 공존하게 될 타자들이자 ‘반려종’들이기에 로봇의 윤리에 대한 고민을 인간의 윤리에 대한 고민으로 확장해야 한다고 강조했다. 유은순·조미라, 「포스트휴먼 시대의 로봇과 인간의 윤리」, 『한국콘텐츠학회논문지』 18(3), 한국콘텐츠학회, 2018, 594~599쪽.

서 아시모프는 로봇이 맞닥뜨리는 인간의 복잡성과 3원칙에 내재한 근원적인 모순을 사유하였고 급기야 작중의 로봇 R. 다닐 올리버의 이성적 추론과정에 따라 ‘로봇공학 제0원칙’을 목록에 추가한다.⁰⁷ 제0원칙의 제정은 그것이 자발적 의식 변화를 겪은 로봇들의 치열한 토론에 따라 이루어졌다는 점에서 다른 세 개의 선행 법칙들과는 변별된다. 『로봇과 제국』의 로봇공학 원칙 재편은 ‘인간’ 범주를 정의하는 데 따른 혼란을 직접적으로 문제시한다. 그리고 그 과정에서 인류 발전을 가로막고 차별과 적대의 원인이 되는 ‘인간(Human Being)’이라는 구분 틀은 ‘인류(Humanity)’라는 유기체적 보편성으로 확장된다.

당연하게도, 로봇공학 제0원칙의 설정 역시 뚜렷한 한계를 가지고 있다. 이만재의 지적처럼 우주인(Spacer)과 지구인, 솔라리아인, 이주자를 아우를 수 있는 ‘인류’라는 보편적 특성이 과연 존재하는가의 문제도 제기될 수 있거나와⁰⁸ 제0원칙이 무엇보다 인류 발전의 당위에 개별자로서의 인간과 그 특수성을 희생시키고 있기 때문이다. 지능과 힘, 도덕적 청렴성 모든 면에서 인간의 수준을 뛰어넘은 아시모프의 로봇은 마이클 필라스의 표현처럼 인간성을 질문하기 위한 하나의 은유이기도 하다.⁰⁹ 그렇기에 『로봇과 제국』을 다시 읽는 일은 단순히 로봇에게 명령을 내리는 주체(혹은 초-주체)로서의 인간에만 주목하는 것이 아니라 로봇이라는 비인간 인물을 통하여 작중의 세계에서 무엇이 ‘인간’으로, 무엇이

07 그리하여 제0원칙이 추가된 아시모프의 로봇공학 원칙은 아래와 같다. 1. 로봇은 인간(Human Being)에게 해를 끼쳐서는 안 되며, 적절히 행동하지 않음으로써 인간이 해를 입게 해서도 안 된다. 2. 로봇은 1원칙을 어기지 않는 한에서 반드시 인간의 지시를 따라야 한다. 3. 로봇은 1원칙과 2원칙을 어기지 않는 한에서 반드시 자신을 보호해야 한다. (Isaac Asimov, 『Runaround』(1942)) 0. 로봇은 인류(Humanity)에 해를 끼쳐서는 안 되며, 적절히 행동하지 않음으로써 인류가 해를 입게 해서도 안 된다. (Isaac Asimov, 『Robots and Empire』(1985))

08 이만재, 『아시모프와 로봇』, 『정보과학회지』 14(8), 한국정보과학회, 1996, 57~58쪽.

09 Michael Filas, 『Cyborg Subjectivity』, University of Washington Ph.D Dissertation, 2001, p.58.

‘인간 아닌 것’으로 여겨지고 있는가를 살펴보는 과정이 된다.

인공지능 기술의 발전 수준이 초연결과 초지능 그 이후를 논하게 된 지금, 인공지능 로봇과 인간의 공존 및 윤리 문제는 사실상 모든 학문 분야 단위에서 활발히 논의되고 있다. 이런 시대에 아시모프의 소설을 다시 읽는다는 건 어떤 의미인가. SFScienceFiction 장르의 넓은 폭은 과학과 기술의 변형, 발전에 따른 인간의 조건 제반을 추적하기에 알맞다. 세릴 빈트는 SF의 일반적인 정의로 “사실주의 양식에서 재현할 수 없지만 그 범위에도 사실인 것을 포착하는 장르”¹⁰를 제시하며 그것이 현실의 역사적 맥락을 충분히 반영하고 있음을 강조하였다. 한편 슈테판 헤어브레히터는 리얼리즘과 허구적 시뮬레이션의 중간 영역에 속하는 SF가 과학과 기술에 의한 인간 규정, 그리고 우리의 기술·문화적 조건을 추적하기에 명실상부한 포스트휴먼의 장르라고 보았다.¹¹ 1930년대 말부터 줄곧 인공지능 로봇과 로봇공학 법칙을 고심해 온 아시모프는 『로봇과 제국』에 이르러 자신의 생산물에 행동 법칙을 부여하는 인간과, 그 법칙을 주제로 토론하고 새 항목을 자의적으로 설정하는 로봇에게로 초점을 이동시킨다. 작중의 로봇은 인간의 자기동일적 전체성을 위협하는 타자이다. 『로봇과 제국』을 읽으며 로봇들의 대화를 ‘엿듣는’ 독자들은 끊임없이 재편되는 ‘인간’의 범주를 추적하게 된다. 이 글은 “우리가 로봇에 대해 알고 있는 것이 과연 얼마나 될까?”¹²라는 인간의 말과 “제 동의 없이 취해지는 그런 조치에는 자발적으로 응할 수 없을 것 같습니다”¹³라는 로봇의 말 사이의 진동을 포착하고자 한다.

10 세릴 빈트, 전행선 역, 『에스에프 에스프리』, 아르테, 2019, 14~28쪽.

11 슈테판 헤어브레히터, 김연순·김응준 역, 『포스트휴머니즘』, 성균관대 출판부, 2012, 166~167쪽.

12 아이작 아시모프, 정철호 역, 『로봇과 제국 1』, 현대정보문화사, 2002, 165쪽.

13 위의 책, 17쪽.

이상의 논의를 위해 2장에서는 솔라리아 행성의 로봇들을 분석 대상으로 삼았다. 『로봇과 제국』에 등장하는 로봇들은 로봇공학 3원칙을 거스를 수 없는 양전자 두뇌Positronic Brain를 가지고 있음에도 솔라리아의 억양을 쓰지 않는다는 이유로 인간인 D.G.를 공격한다. 솔라리아의 로봇은 3원칙이 전제하는 ‘인간’이 인종, 언어에 따라 자의적으로 해석되어 또 다른 차별과 배제의 근거가 될 수 있음을 보여주는 예시이다. 3장에서는 로봇 시리즈의 이전 작품들에서 지구인 형사 일라이저 베일리의 수사파트너였으며, 필요에 따라 인간 행세를 하며 수사에 임할 수 있을 정도로 인간과 닮도록 정교하게 만들어진 로봇 다닐 올리버를 분석 대상으로 삼았다. 다닐 올리버는 뛰어난 이성적 추론 능력을 갖춘 로봇이고, 로봇공학 원칙에 대해 합리적인 의구심을 갖는 로봇이기도 하다. 아시모프는 다닐을 통해 인간과 인류의 경계를 재설정하고 타자로서의 로봇의 위치를 다시 한번 확인한다. 4장에서는 텔레파시 능력을 가진 로봇 지스카드를 분석 대상으로 삼았다. 지스카드는 다닐처럼 인간과 혼동될 정도로 정교하게 만들어진 로봇은 아니기에 인간에게 불쾌감을 주는 모델이다. 그럼에도 지스카드가 인간의 마음을 읽고 그에 공감할 수 있다는 사실을 알았던 일라이저 베일리는 글래디아에게 “지스카드를 신뢰”하라고 당부한다. 이는 아시모프가 새로이 구축하고자 한 휴머니즘의 요체가 지스카드의 특수능력인 텔레파시와 밀접하게 관련되어 있음을 잘 보여주는 장면이다.

2. 인간 범주의 자의성을 폭로하는 도구로서의 로봇

『로봇과 제국』에서 인간은 크게 우주인과 지구인으로 분화되어 있다. 여기에 각 인류의 거주 장소와 거주 형태에 따라 우주인은 다시 오로라인과 솔라리아인으로, 지구인은 지구 거주자들과 이주자 행성 거주자라는 두 층위로 다시 분화된다. 우주인과 지구인은 로봇 기술과 인간 향상(Human Enhancement) 기술에 대한 상반된 태도를 바탕으로 가히 중 분화(Speciation)라고 불러도 좋을 만큼의 몇 세기에 걸친 신체 변화를 겪은 인간이다. 우주인은 이러한 기술들을 극단적으로 수용함에 따라 수명이 400년에 이르게 되었으며, 삶의 모든 부분을 로봇 기술에 의존하고 있고 다른 우주인과 직접 대면하는 일이 일생에 몇 차례 없을 정도로 고립되어 생활한다. 반면 지구인은 현 인류의 신체를 거의 그대로 유지한 채로 살아가고 있으며, 인간형 로봇에게 혐오감을 가지고 있고 몇천만 명이 함께 모여 사는 시티(City)라는 거주지역에서 생활한다. 기대수명과 생활방식의 극단적인 차이는 우주 개척과 진보에 대한 태도에까지 영향을 미치는 것으로 묘사되었다. 우주인들은 긴 수명만큼이나 보수적이며 대를 잇는 문제에 있어서도 소극적이지만 지구인들은 계속해서 인구수를 불러가며 은하의 행성들을 거주지로 개척해 나간다는 식이다.

닉 보스트롬과 줄리언 사블레스쿠는 인간 향상에 대해 그것이 실용적이고 실제적인 윤리의 질문으로 우리 삶에 침투해 들어왔으며 이제 ‘그건 SF에나 나오는 얘기 아냐?’라는 질문은 ‘인체 향상을 해야만 하는 걸까?’라는 실존적 질문으로 대체되었다고 지적한 바 있다.¹⁴ 『로봇

14 Nick Bostrom and Julian Savulescu, "Human Enhancement Ethics: The State of the Debate", *Human Enhancement* ed. Julian Savulescu and Nick Bostrom, Oxford University Press, 2013, p.18.

과 제국』의 배경이 되는 세계는 후자의 질문이 제기되고도 몇 세기가 흐른 후의 세계이다. 그 세계에서 우주인은 인체 향상에 적극적으로 임한 끝에 지구인과 우주인 중 어느 편이 ‘인간’인가라는 질문을 끌어내는 데 이른다. 일반적으로 트랜스휴먼은 유전공학, 항노화 치료법, 웨어러블 컴퓨터 등을 통한 인간 신체와 정신의 증강을 포괄하며 우주 개척이나 초지능을 갖춘 기계 등 잠재적으로 인간의 조건을 근본적으로 변화시킬 가능성을 아우르는 개념으로 정리된다.¹⁵ 그렇기에 트랜스휴먼이라는 존재 형태는 필연적으로 인류에 대한 인류의 정의 문제를 곧장 야기한다. 『로봇과 제국』에서 지구인과 우주인은 자신이 ‘인간’ 중임을 추호도 의심하지 않는다. 하지만 그들은 상대 ‘종족’과 같은 유전적 조상을 갖고 있다는 점은 결코 인정하려 들지 않는다. 서로를 다른 종으로 여기고 상대 종의 세력 확장을 경계하면서도, 로봇에 대한 명령권자로서 ‘인간’이라는 주체성을 포기하지 않는 모순이 『로봇과 제국』의 작중 현실 속에서 불안한 균형을 이루고 있는 것이다. 아시모프는 이런 아이러니를 단순화시켜 보여줄 수 있도록 이들의 언어에 관한 특별히 상세한 묘사를 덧붙인다.

“아, 당신이 ‘미적 이유’ aesthetic라고 했던 거군요.”

그러자 이주자는 깨끗하고 하얀 치아를 내보이며 웃음을 터뜨렸다.

“실은 당신 발음도 제겐 아주 이상하게 들립니다, 마담.”

글래디아는 품위있는 표정을 유지하려고 애썼지만 웃음 때문에 허물어지고 말았다. 따지고 보면 정확한 발음이라는 것은 그 지역의 다수 의견에 불

15 Nick Bostrom, “Transhumanist Values”, *Ethical Issues for the 21st Century* ed. Frederick Adams, Philosophical Documentation Center Press, 2003, p.3. <https://nickbostrom.com/ethics/values>

과한 것이다. 그녀가 말했다.

“나는 솔라리아 억양으로 이야기할 수도 있어요. 아까 그 말을 솔라리아 식으로 하면 ‘미이적 리이유’,¹⁶ aesthetic rayzuns가 되지요.” …(중략)…

“당신은 너무 혀끝을 굴리는군요. 혀의 양 옆을 써서 발음해보세요. 솔라리아인이 아니면 누구도 정확하게 그런 발음을 못하지요.”¹⁶

『로봇과 제국』에 등장하는 인류에겐 은하 표준어라는 단일 공용어가 있지만 우주인과 지구인은 서로 다른 억양의 방언을 사용한다. 아시모프는 이 방언이 발음되는 특징에 대하여 “지구인의 억양은 ‘k’와 ‘p’를 유성음으로 발음한다”라거나 “솔라리아 억양으로는 후두음인 ‘r’ 발음이 두드러진다”라는 식의 상세한 설명을 덧붙인다. 위의 인용문에서 지구인 D.G.와 우주인 글래디아는 그러한 발음의 차이를, 비록 듣기엔 우스꽝스러울지언정 대수롭지 않게 여긴다. 심지어 작가는 “따지고 보면 정확한 발음이라는 것은 그 지역의 다수 의견에 불과”하다는 글래디아의 내면을 밝혀 적고 있다. D.G.와 글래디아가 지구인과 우주인의 화합을 희망했던 로봇공학자 패스톨프의 의지를 계승한 인물들인 만큼 그들에게 방언은 사소한 발음 차이에 불과한 것으로 받아들여지고 있다. 그러나 그들이 버려진 솔라리아 행성에서 마주한 로봇들은 이런 발음의 차이가 인간과 비인간을 분류하는 “분류학적인 테러Taxonomical Terrorism”¹⁷를 유발할 수 있음을 상기시킨다.

16 아이작 아시모프, 정철호 역, 앞의 책, 118쪽. “aesthetic”과 “aestheetic rayzuns”의 표기는 Isaac Asimov, *Robots and Empire*, Harper Collins, 2018, p.82참고.

17 에릭 후엔스트는 향상된 인간과 인류 범주 재설정 문제가 돌출시키는 ‘억압된 소수’를 비롯하여, 그들에게 행해지는 분류학적인 테러리즘(Taxonomical Terrorism)에 대한 성찰의 필요성을 강조하였다. Eric T. Juengst, ‘What’s Taxonomy Got to Do with It? ‘Species Integrity’, Human Rights, and Science Policy’, *Human Enhancement* ed. Julian Savulescu and Nick Bostrom, Oxford University Press, 2013, pp.45~47.

DG는 가능한 한 오로라 귀족풍의 말투를 흉내내어 짐작게 말했다.

“마담, 이 영지의 감독과 대화를 나누고 싶습니다.”

그 여자는 한동안 DG를 쏘아보더니 입을 열었다. 그 말씨는 아주 투박한 솔라리아 방언이었는데 한껏 혀를 굴리며 발음하는 모습은 코미디의 한 장면을 연상케 했다.

“당신은 사람이 아니야.”

...(중략)...

글래디아는 큰 소리로 고함을 질렀다. 그것은 자신의 귀에도 낯설게 들리는 아주 투박한 솔라리아 사투리였다. 하지만 솔라리아 로봇에게 호통 칠 수 있는 사람이 자신 말고 누가 있던 말인가!

“로봇, 당장 멈춰라!”

...(중략)...

“마담, 당신이 지금 가로막고 있는 것은 사람과 흡사하지만 절대 사람이 아닙니다. 저는 그러한 존재들을 보는 즉시 죽여버리라는 명령을 받았습시다.”¹⁸

베일리 행성의 이주자 무역상인 D.G.와 글래디아는 원주민인 솔라리아인들이 모종의 이유로 모두 떠나간 솔라리아에 당도한다. D.G.의 목적은 솔라리아인들이 행성을 떠날 때 미처 가져가지 못한, 매우 높은 수준의 로봇공학기술이 접목된 주인 없는 로봇을 수거해 그것을 오로라인들에게 판매하려는 것이었다. 그러나 솔라리아에 도착한 일행은 그곳의 로봇들이 그들의 명령에 따르지 않는다는 것을 깨닫게 된다. 솔라리아에서 그들이 마주한 ‘랜드리’라는 이름의 감독 로봇은 심지어 인간인 그

18 아이작 아시모프, 정철호역, 앞의 책, 204~211쪽.

들을 직접 공격하기까지 한다. 그리고 감독 로봇 랜드리의 발언을 통해서 솔라리아 행성에 남아있던 로봇들이 오직 솔라리아 방언을 사용하는 인간만을 ‘인간’으로 여기도록 프로그래밍되어 있다는 점이 드러난다.

솔라리아의 로봇들은 오로라인과 지구인들이 당연하게 여기는 ‘인간’의 범주, 즉 로봇공학 3원칙이라는 안전장치에 따라 로봇을 마음대로 부릴 수 있는 조건을 흔들여 놓는다. 명령에 거역하고 인간을 공격하는 이들은 당장이라도 인간의 건강을 위협하고 인간을 죽일 수 있는 위협적인 타자로서 일행의 앞에 모습을 드러냈다. 그리고 이러한 타자로서의 로봇의 출현은 D.G.와 글래디아를 비롯하여 로봇인 다닐과 지스카드의 자기동일성에도 직접적인 영향을 미친다. 레비나스식으로 표현하자면 솔라리아의 로봇들은 이들에게 “얼굴”로서 현현하여 그동안 당연하다고 여겨지던 로봇공학 3원칙과 인간 범주를 재편하게 강제한다.¹⁹ 그리하여 뒤이어 베일리 행성에 당도한 글래디아는 “마음의 언어에는 방언이 없다”²⁰라는 요지의 연설을 통해 분류학적 테러에 반발하고 지구인과 우주인의 화합을 촉구하는 새로운 주체성을 갖추게 되었으며, 이 일을 기점으로 로봇공학 3원칙의 불안전성을 확인한 다닐과 지스카드는 ‘인간’을 재정의할 방안을 더욱 적극적으로 추론하기 시작한다.

19 레비나스는 주체의 자기동일성을 무너트리는 외재적인 타자가 주체에게 ‘얼굴’로서 현현한다고 표현하였다. 레비나스는 “기존의 타자의 관념을 넘어서면서 자신을 제시하는 방식”을 얼굴이라고 정리하였는데 그것은 기존의 가변적 이미지를 파괴하고 그 이미지를 넘어서며 ‘나’의 주도권을 박탈하는 운동과 관련이 있다. 에마뉘엘 레비나스, 김도형 외 역, 『전체성과 무한』, 그린비, 2018, 56~58쪽.

20 아이작 아시모프, 정철호 역, 앞의 책, 268쪽.

3. 로봇공학 제0원칙을 도출한 스스로 사유하는 로봇

로봇과 사이보그의 주체성에 관한 연구를 진행했던 마이클 필라스는 알튀세르의 호명 이론을 참조하여 로봇들이 그들을 ‘로봇’으로 호명하는 초-주체들(Super-Subject)로서의 인간공학자들에 의해 주체로 호명되고 있음을 설명하였다.²¹ 아시모프의 로봇들 역시 우주인과 지구인의 대립 구도라는 정세와 로봇공학 3원칙에 의거한 로봇의 종속 상태, 생산수단으로서의 로봇 기술이라는 물질적 조건 등 이데올로기적 함의를 가지고 있다. 『로봇과 제국』의 로봇들 역시 기존의 이데올로기 질서와 생산수단의 재생산을 위해 설계되어 있다. 그러나 동시에 이 작품은 다닐과 지스카드라는 두 특수한 로봇들을 불완전하게나마 그들 나름의 주체적인 윤리관을 개척해 나갈 수 있는 도덕적 주체로 재발견함으로써 견고해 보이던 질서의 미세한 균열을 포착한다.

고도로 발전한 로봇 기술을 보유하고 있는 우주인들은 로봇의 기능을 신뢰하고 그에 의지하며 살아간다. 우주인의 사회에서는 다닐과 같은 정교하게 제작된 로봇의 완벽한 기억력이 법정 증거로서의 효력을 가지기도 한다. 『로봇과 제국』이 포스트휴먼 주체, 트랜스휴먼 주체에 대한 고민거리를 충실하게 던져주면서도 인공지능 로봇과 공존하는 인류에게 법의 의미를 묻고 있다는 세이지 레슬리-맥카시의 분석은 주목을 요한다. 그는 다닐이 오로라 사회에서 법적인 주체로 자리매김하고 있음을 확인하면서도 아시모프가 “오로라의 법은 우주인 사회를 정말 잘 반영하고 받쳐주고 있는가?”를 질문하고 있다고 보았다.²² 이 질문

21 Michael Filas, op. cit., pp.25~30.

22 Sage Leslie-McCarthy, “Asimov’s Posthuman Pharisees: The Letter of the Law Versus the Spirit of the Law in Isaac Asimov’s Robot Novels”, *Law, Culture and the Humanities* 3, 2007.

을 심화하자면 이런 물음까지도 가능할 것이다. 만약 로봇이 인간의 삶에 도저히 떼어놓을 수 없을 만큼 융화되어 있다면, 그리고 인간들도 수명이 400년에 달하는 트랜스휴먼이라면, 우주인의 ‘인간’만을 위한 법의 제정은 근원적으로 불가능한 것이 아닌가. 아이작 아시모프가 미래 인간의 법률에 제기하는 의문은 인간의 정체성의 불분명한 경계에 대한 것이기도 한데 이는 작중에서 “인간처럼 생각하는” ‘이성적 추론 능력’을 갖춘 로봇 다닐 올리버에 의해 로봇공학 3원칙에 대한 의심으로 유비된다. 다닐은 또 다른 로봇 지스카드와의 토론을 통해 로봇공학 3원칙의 한계와 그 너머를 고안하고자 한다. 아래는 두 로봇의 토론 장면 중 하나를 발췌한 것이다.

그들은 저택에 딸린 정원 남쪽 끝에서 만나 잠시 이소픽어(Aesopic Language)로 대화를 나누었다. 그것은 일종의 축약언어로서, 수십 년 동안 이런 식으로 대화를 나누어온 그들로서는 구태여 인간의 언어처럼 세세하게 모든 것을 표현할 필요를 느끼지 않았다.

...(중략)...

“3원칙에 우선하는 건 아무것도 없어.”

“만약 내가 인간이라면 그 법칙들을 넘어서서 생각할 수 있을 걸세. 그리고 다닐, 나는 자네가 나보다 먼저 그 법칙 너머에 있는 것들을 볼 수 있을 것이라고 믿고 있네.”

“내가?”

“그래, 나는 오래 전부터 비록 자네가 로봇이긴 하지만 정말 인간과 흡사

하게 사고하고 있다는 생각을 해왔지.”²³

두 로봇은 마치 인간 철학자들처럼 정원에서 만나 로봇공학 3원칙에 대한 토론을 시작한다. 이때 그들이 사용하는 언어는 인간의 언어와는 구별되는 “이소픽어(Aesopic Language)”라고 표현되어 있다. 이 대목에서는 한국어 번역본의 의역에 따른 맥락 축소를 간단히 지적할 수 있을 것이다. 두 로봇의 대화 언어에 대한 원래의 문장은 “they spoke in abbreviated and Aesopic language”²⁴이다. 이를 직역하자면 두 로봇의 언어가 ‘축약된’ 그리고 ‘이소픽한’이라는 두 개의 형용사의 수식을 받고 있음을 알 수 있다. 이소픽(Aesopic)이란 그 조어방식에서 추론해볼 수 있듯 이솝의 우화적 말하기, 즉 대화의 진의를 별개로 보이는 이야기에 숨겨두는 표현방식과 관련이 있다. 실제로 ‘이소픽’은 1920년대 중반 소련 등지에서 사용되었던 표현으로, 음모나 지하활동의 가담자들이 그들의 본의와 정보를 숨기기 위해 사용하던 일종의 암호로 정의되어 있다.²⁵ 그렇다면 ‘이소픽한’ 언어를 사용한 두 로봇의 토론은 단지 인간을 매개하지 않은 두 기계의 대화라는 표면적 의미를 넘어서 진의를 숨긴 모의의 과장임이 보다 분명해진다.

소설에서 로봇공학 3원칙에 직접적이고 반복적인 의문을 제기하는 것은 다닐이다. 다닐은 “만약 내가 인간이라면”이라는 단서를 달아가며 3원칙 “너머”를 끊임없이 고민한다. 지스카드와 글레디아의 표현처럼 다닐은 “정말 인간과 흡사하게 사고”하는 로봇이다. 이 말의 함의는 다

23 아이작 아시모프, 정철호 역, 앞의 책, 26~28쪽.

24 Isaac Asimov, *Robots and Empire*, Harper Collins, 2018, p.15.

25 “Aesopian”, *Merriam-Webster Dictionary*, <https://www.merriam-webster.com/dictionary/Aesopian>.

다닐이 인간이라는 초-주체의 통제를 벗어나 그 스스로를 하나의 도덕적 주체이자 입법기관으로 세우고 있다는 정황 속에서 확인된다. 다닐의 양전자두뇌는 기입력된 로봇공학 3원칙을 어길 시 고장이 나도록 설계되었지만, 다닐은 마침내 스스로 제0원칙을 스스로 제정하여 그것을 하나의 행동 원리로 새로 기입하기에 이른다.

다닐이 말했다.

“맨더머스 박사님, 당신도 아시겠지만 우린 언젠가 솔라리아에서 인간의 정의를 솔라리아인으로 한정해 프로그램된 로봇을 만난 적이 있습니다. 인간이 어떨어떨한 유형이라는 식으로 좁은 범위로 정의된 로봇은 어마어마한 파괴행위를 유발하기 마련이지요. …(중략)… 따라서 저희는 집단으로서의 인간집단에 대한 위해, 인류 전체에 대한 위해를 저지하는 것이 특정 개인에 대한 위해를 저지하는 것에 우선한다고 생각합니다.”

맨더머스가 숨을 헐떡이며 말했다.

“제1원칙은 그게 아니야!”

“그것은 제가 제0원칙이라고 부르는 것입니다. 그것은 모든 원칙에 대해 우선권을 갖고 있습니다.”

“넌 그런 식으로 프로그램되어 있지 않아.”

“그건 제 스스로 입력한 프로그램입니다.”²⁶

위의 인용문은 소설의 마지막 대목으로 제0원칙의 성격을 잘 드러내고 있다. 우선, 이로부터 제0원칙이 도출된 배경이 솔라리아에서의 조우와 관련이 있음을 알 수 있다. 다닐은 솔라리아의 로봇들과 감독 로봇

26 아이작 아시모프, 정철호 역, 『로봇과 제국2』, 현대정보문화사, 2002, 273~274쪽.

‘랜드리’를 본 후 개별 인간에 대한 분류의 폭력을 확인하였다. 둘째로 다닐의 제0원칙은 ‘인간집단’과 ‘인류’라는 새로운 범주를 설정함에 따라 세워진 것이다. 집단으로서의 인간은 그를 추궁하는 우주인 맨더머스와 아마디로에게는 낯설거나 불쾌한 개념이다. D.G.가 “로봇에 얽혀 사는 기생생물”이라는 표현을 쓸 정도로 우주인들은 로봇에 의지한 채로 자급자족하며 그러한 상태로 이미 충분하다고 착각하는 이들이다. 또한 우주인들은 집단을 이루어 모여 살고 있는 지구인들을 혐오하기에 그들과 하나의 ‘인류’로 엮이는 것을 극도로 혐오하는 이들이고 이런 혐오감은 지구인과 우주인의 공동 목표라 할 수 있는 은하 개척과 제국 건설을 가로막고 있다. 그렇기에 다닐은 제0원칙을 설정함으로써 지구를 파괴하려는 맨더머스의 계획을 저지하려는 것이다.

『로봇과 제국』이 작품 전체에 걸쳐 공을 들여 묘사하는 제0원칙의 도출 과정을 지켜보는 일은 로봇윤리와 비판적 포스트휴머니즘 담론에 논의거리를 던져준다. 로봇윤리의 차원에서 다닐의 자기입법적인 성격을 고찰한 지현아는 다닐의 제0원칙 수립이 단지 주어진 일을 수행하는 자동성Automaticity의 수준이 아닌 자신의 사고 판단에 따라 행동하는 자율성Autonomy의 차원에서 논의되어야 할 것이라고 보았다.²⁷ 변순용은 아시모프의 로봇공학 3원칙이 로봇을 객체가 아닌 주체로 삼아 그에게 의무를 부과했던 최초의 시도라고 평가했으며, 이 원칙이 향후 ‘인류’가 아닌 외계와 조우하고 관계를 맺게 될 때 로봇의 위치를 미리 규정한 것이라는 의의를 확인하였다.²⁸ 이는 제0원칙을 수립하기 위한 다닐의 노정이 휴머니즘의 내부에 머무르며 그것을 비판적으로 해체하는 비판적 포스트

27 지현아, 「로봇윤리(robethics)의 도덕교육적 적용방안 연구」, 서울대 박사학위논문, 2020, 45~46쪽.

28 변순용, 「인공지능로봇을 위한 윤리 가이드라인 연구: 인공지능로봇윤리의 4원칙을 중심으로」, 『윤리교육 연구』 47, 한국윤리교육학회, 2018, 238~239쪽.

휴머니즘의²⁹ 사유 전개 과정과 관련이 있음을 확인시킨다.

4. 접촉하고 공감하고 기억하는 텔레파시 로봇

로봇 지스카드는 다닐과는 달리 인간처럼 매끄러운 신체를 갖고 있지 않기에 필요에 따라 인간과 로봇의 눈을 속여가며 인간 행세를 할 수 없다. 그뿐 아니라 지스카드는 인간형 로봇의 보유를 오랫동안 금지해 온 지구인들에게는 혐오감마저 줄 수 있는 외형이라 묘사되며, 우주인들 역시 다닐에게 보이는 만큼의 호감을 지스카드에게는 보이지 않는 모습을 자주 보인다. 요컨대 지스카드는 언캐니Uncanny 혹은 운하임리히 Unheimlich를 유발하는 로봇이다. ‘두려운 낯성’으로서의 운하임리히를 설명했던 프로이트는 그것을 자기 자신의 실현되지 못한 좌절된 욕망, 혹은 억압되어 있던 것의 회귀와 관련시켰다.³⁰ 지스카드를 본 인간들이 느끼는 언캐니는 단지 그의 외형에 의한 것만이 아니다. 그것은 오히려 다닐은 가지지 못한 지스카드의 고유 능력인 텔레파시Telepathy 기능과 관련이 있다. 지스카드의 텔레파시는 상대의 생각과 감정을 재빠르게 파악하고, 대상의 감정을 일정 정도는 조작할 수 있는 능력으로 요약된다. 다만 이 기능을 제대로 사용하기 위해서는 감정을 읽으려는 상대방과 충분히 오랜 교감의 시간을 갖는 것이 중요하며, 정신 조작이라고 불리는 무시무시한 힘도 실은 어떤 선택을 망설이고 있는 상대가 그 선택을

29 헤어브레히터는 포스트휴머니즘을 논하는 데 있어 기술과 비휴먼적인 것은 수단일 뿐 자체의 합목적성을 가지는 것이 아님을 강조하며 비판적 포스트휴머니즘이란 포스트-‘휴먼’한 것이 아니라 ‘포스트’휴먼한 것에 가깝다고 보았다. 슈테판 헤어브레히터, 앞의 책, 19쪽.

30 프로이트, 정장진 역, 『두려운 낯성』, 『예술, 문학, 정신분석』, 열린책들, 2012, 409~427쪽.

결단하거나 단념하게 만드는 정도의 조작이다.

빠른 수식 계산과 논리적인 추리, 그리고 법정이 채택할 만한 정확한 기억력 등의 능력은 로봇의 도구로서의 속성과 관련이 있다. 그러나 텔레파시, 즉 원격Tele-에서 감응-pathy을 주고받는 능력 곧 상대의 감정에 공감하고 그의 마음을 돌려세우는 능력은 작중의 어떤 로봇에게도 탑재되지 않은 특수한 것이다. 동시에 이런 감정적 공감능력은 작중의 우주인들도 결여하였거나 등한시한 것이기도 하다. 『로봇과 제국』의 솔라리아인과 오로라인은 효율성과 위생 등을 이유로 일생에 걸쳐 단 몇 차례만 다른 인간을 실제로 마주할 정도로 고립된 채 살아간다. 서로 얼굴을 보고 대화를 나누더라도 그것은 홀로그램 송출 기술에 의존한 만남일 뿐 대상과의 직접 접촉과는 무관한 관계이다. 반면 지스카드의 텔레파시 능력은 대상과의 실제 거리를 어느 이상 좁히는 것이 능력 발현의 중요조건으로 설정되어 있다. 상황이 그러할진대 『로봇과 제국』의 배경이 되는 세계에서 ‘감정 로봇’인 지스카드는 ‘질책하는 타자’로서 우주인 사회의 소통 단절과 불감증을 반영하는 로봇이며, 우주인들이 지스카드를 언캐니한 로봇으로 받아들이는 것 역시 이 같은 이유 때문이라 할 수 있다. 그러나 공감, 소통, 설득, 기억이야말로 지구인과 우주인의 화합과 ‘인류’ 전반을 진보하게 할 방법이라 보았던 생전의 일라이저 베일리는 지스카드의 특수능력이 인류의 미래를 바꿀 수 있는 강력한 힘임을 꿰뚫어 보았던 인물이다.

“자네의 능력을 알고 있는 사람이 있나?”

“인간 중에서는 오직 당신뿐입니다. 당신이 그 이야기를 다른 사람에게 하지 않았다면…….”

www.kci.go.kr

“물론 하지 않았네. 그리고 패스톨프가 아니라 바로 자네지? 모든 정부 관리들을 이주계획의 열렬한 지지자로 바꾸어놓은 게…… 게다가 패스톨프가 다닐 대신 자네를 지구로 데리고 오도록 한 것도 자네였을 거야. 자네야말로 가장 중요한 존재일세. …(중략)… 그런 위기가 닥칠 때 지구를 보호할 책임이 전적으로 자네에게 있다는 점을 잊지 말게.”

…(중략)…

지스카드가 패스톨프를 따라 오로라로 귀환하기 위해 착륙선에 올라탔을 때 그는 먼발치에서 다시 베일리를 볼 수 있었다. 하지만 그와 이야기를 나눌 기회는 없었다.

베일리는 손을 흔들며 입술만 움직여 한 마디의 단어를 보냈다. 그것은 ‘기억하라’는 말이었다.

지스카드는 그 말과 함께 그 말 뒤에 숨어 있는 감정까지 받아 간직했다.³¹

패스톨프의 정적政敵이자 당대의 가장 뛰어난 로봇공학자이며 지구 파괴계획의 입안자이기도 한 아마디로에게 지스카드는 반드시 탈취하여 분해해야 하는 위험한 로봇이다. 아마디로의 반응은 그를 대하는 인간들의 일반적인 반응이기도 하다. 일라이저 베일리조차 지스카드의 감응 능력보다 그의 정신조작 능력이 더 유용하게 사용되리라 넘겨짚고 그를 ‘가장 중요한 존재’라고 치켜세운다. 현대정보문화사의 번역본은 이 대목에서 원작은 한 문장을 삭제하고 있다. 아시모프의 원작에서 베일리는 지스카드에게 “자네야말로 가장 중요한 존재일세>You were essential”라는 당부 다음에 “다닐은 오히려 방해가 될 수 있어and Daneel might have been

31 아이작 아시모프, 정철호 역, 『로봇과 제국1』, 현대정보문화사, 2002, 87~88쪽.

a distraction”³²라고 이어 말한다. 다닐이 로봇공학 제0원칙을 수립하는 등 자신의 이성적인 추론능력에 따라 상황 타개의 방법을 찾아낼 것이지만 그것을 진정으로 실행할 수 있는 능력은 지스카드에게 있음을 베일리는 간파하고 있었다. 그렇기에 작별의 순간 베일리가 지스카드에게 전달한 “기억하라”라는 말은 지구를 보호하고 인류가 위기를 극복하도록 이끌 책임이 전적으로 지스카드에게 있음을 기억하라는 당부이자, 인간으로서 로봇에게 내리는 명령이었다. 하지만 베일리를 떠나보내는 마지막 장면에서 지스카드는 “기억하라”라는 명령어를 입력받았을 뿐 아니라 그런 명령을 내리는 베일리의 감정까지를 느낄 수 있었다. 그리고 인간인 베일리가 예측하지 못한 이 감응이 원인이 되어 『로봇과 제국』의 결말은 베일리의 당부와는 반대의 결과인 지구 오염으로 끝이 난다.

“결국 실패했어!”

다닐은 절망감에 빠진 채 쓰러져 있는 사람의 몸뚱아리를 바라보았다.

“내 눈 앞에 보이지 않는 사람들에게 해를 미치지 못하도록 맨더머스 박사를 저지시켜야 했는데…… 그의 명령 때문에 꼼작할 수가 없더라구. 결국 0원칙은 작동하지 않았던 거야.”

그러자 지스카드가 말했다.

“아니야, 다닐. 자넨 실패하지 않았어. 사실은 내가 자네를 막았지. 맨더머스 박사는 어떻게든 자신이 하고자 하는 일을 실행에 옮기려고 애썼지만 자네에 대한 두려움 때문에 못하고 있었던 거야. 나는 그의 공포심을 마비시켰네. 그로 인한 단호한 명령이 자네를 마비시킨 셈이지. 맨더머스 박사는 결국 지구의 지각층에 불을 질렀네. 느린 속도로나마…….”

32 Isaac Asimov, op. cit., p.58.

...(중략)...

지스카드의 목소리가 천천히 잦아들었다.

다닐은 자리에 앉아 있는 지스카드 곁에 무릎을 꿇었다. 지스카드의 금속제 손을 움켜잡은 다닐의 속삭임은 절규에 가까운 흐느낌으로 변해갔다.

“어서 일어나게, 지스카드! 일어나야 해! 0원칙에 따르면 자네의 행동은 정당한 거였어. 자네는 무수한 인간의 생명을 구했다구. 자네는 인류를 위해 전력을 다했어. 그런 자네가 왜 고통을 당해야 하나?”³³

『로봇과 제국』은 맨더머스 박사가 지구 지각층에 불을 지르는 데 성공함으로써 지구의 대기 전체가 서서히 방사능에 오염되어 중국엔 지구인들이 오염된 지구를 버리고 다른 행성으로 진출할 수밖에 없게 되는 결말을 맞는다. 눈여겨볼 것은 다닐이 이런 결말을 ‘실패’로 여기는 반면 지스카드는 이를 인류의 진보를 위한 필요조건으로 보고 있다는 사실이다. 지스카드의 이러한 판단은 ‘인간이라는 부분이 아닌 인류라는 전체를 보아야 한다’라는 베일리의 당부 이면에 그가 사랑했던 글래디아라는 우주인과 화합하며 함께 살아가고 싶다는 소망이 있음을 그가 포착했기에 도출된 것이다. 지구인들의 신앙의 대상이며 우주인들은 기어코 부정하려는 지구와 미신적 결속을 끊어내지 않는다면 인류는 결코 화합하여 살아갈 수 없다는 것이 지스카드의 판단이었다. 세이지 레슬리-맥카시는 『로봇과 제국』의 극단적인 결말이 ‘무엇이 인류를 위해 최선인가’라는 질문에 답하기 위해 인간과 그들의 지식이 선택에 개입하지 못하도록 로봇만을 감시자이자 목격자로 세우고 있음을 강조했다. 그는 아시모프가 작중 지스카드라는 로봇 영웅이 선택한 결론에 인간 독자들이 강제로 따라

33 아이작 아시모프, 정철호 역, 『로봇과 제국 2』, 현대정보문화사, 2002, 276~279쪽.

가게 만들고 있다고 평가하였다.³⁴ 감정을 느끼고 상대를 설득할 줄 아는 로봇이 독자적 선택에 따라 지구를 파괴하는 이 마지막 장면은 끝끝내 이 언캐니한 로봇을 공포스러운 타자의 자리에 남겨놓는다.

이동신은 지스카드의 두뇌를 일종의 ‘망가진 두뇌’라고 보았다. 망가진 두뇌란 본래의 기능을 상실했으나 수행성을 여전히 가지고 있는 무엇으로, 지스카드는 인간의 다름을 인정하는 방식으로 자신의 역량을 사용하는 로봇이다.³⁵ ‘망가진 두뇌’라는 비유는 로봇이 그의 기능성에 대한 인간의 기대를 벗어나고 있음을 지시한다. 지스카드와 같은 로봇 형형은 아시모프의 초기 작품에 등장하는 로봇들과는 질적인 차이를 보인다. 세릴 빈트와 마크 볼드는 아시모프 등 1940년대에 활동한 SF 작가들이 타자성에 일정한 관심을 가졌으나 결국은 타자성이 유발하는 공포를 완화하기 위해 차이를 표상할 은유적 능력을 갖추었던 로봇과 외계인 등에게서 돌연변이로서의 능력을 축소하곤 했다고 보았는데³⁶ 1985년에 발표된 『로봇과 제국』에서 지스카드는 인간과 로봇의 ‘차이’를 부각하는, “인류의 전체성을 직접적으로 붕괴시킬 수 있는 위협적인 존재들”³⁷로서 로봇의 자리에 그대로 남아있다.

『로봇과 제국』의 결말은 인류의 진보와 은하계 식민화가 작중의 다른 모든 가치에 선행하는 제일의 가치로 추구됨에 따라 지구의 생태 오염과 수많은 비인간 로봇의 처분에 소극적이며 무책임하다는 비판을 받을 여지가 충분하다. 그럼에도 주목할 것은 결국 이 소설이 반복하여 질문하는 휴머니즘이 인간의 본연 중 비휴먼적인 것을 감안하고 있다는

34 Sage Leslie-McCarthy, op. cit. p.412.

35 이동신, 「망가진 머리: 인공지능과 윤리」, 『인공지능인문학연구』 1, 중앙대 인문콘텐츠연구소, 2018, 64~74쪽.

36 세릴 빈트·마크 볼드, 송경아 역, 『SF 연대기』, 허블, 2021, 204~205쪽.

37 에마누엘 레비나스, 김도형 외 역, 앞의 책, 292~293쪽.

점, 그리고 그 질문의 과정에서 오로라인-솔라리아인-지구인-이주민을 배타적으로 구분하던 경계를 적극적으로 무너뜨리고 있다는 점 등일 것이다.

위의 인용 대목에서 또 하나 눈여겨볼 것은 지스카드의 양전자두뇌 Positronic Brain가 그의 최종 결정을 내린 후 기능이 정지되었다는 점이다. 이때 지스카드는 기입력된 로봇공학 3원칙을 어기고 독자적으로 행동했기에 기능이 정지되는 것으로 묘사된다. 하지만 그의 기능 정지는 로봇의 윤리 차원에서 다시 논해질 수 있다. ‘인류Humanity’라는 단위를 현실화하기 위해 지스카드는 타인과 타종족에 대한 인간의 극도의 불감증을 해소해야 했다. 지스카드가 마지막 순간 자신의 텔레파시 기능을 다닐에게 전송하여 그 능력을 계승시키고 있다는 점, 그리고 지스카드가 작동을 정지하려 할 때 다닐이 “절규에 가까운 흐느낌agonized whisper, 279쪽”을 표출했다는 점은 이를 뒷받침한다. 지스카드는 지구의 방사능 오염 과정에서 발생할 희생들과 일라이저 베일리의 마지막 당부를 지키지 못했다는 데에 양전자두뇌가 기능 정지될 정도의 책임감을 느끼고 있다. 이 장면을 로봇공학 3원칙을 넘어선 ‘양심을 가진 로봇’의 출현이라고 의 미화할 수도 있을 것이다.

5. 나가며: 능동적 허무주의로의 전회

아이작 아시모프의 『로봇과 제국』엔 로봇공학 3원칙을 어기는 로봇은 등장하지 않는다. 2장에서 다루었던 솔라리아의 로봇들은 법칙을 어 기며 인간을 해치는 로봇들이 아니라, 법칙 자체의 불완전성 때문에 슬

www.kci.go.kr

라리아인을 제외한 이들을 모두 비인간으로 판정한 충실한 원칙주의로
 붓들일 뿐이었다. 본문 3장에서 분석한 다닐의 경우는 3원칙을 어기지
 않는 선에서 ‘인류’를 위해 헌신할 방법을 찾기 위해 제0원칙을 정립하
 여 그를 따르려는 도덕적 주체이자 입법기관에 해당했다. 본문 4장에서
 살펴본 지스카드는 작중의 인류가 결여한 공감과 소통이라는 화합의 기
 본 원리를 후대 인류에게 계승하기 위해 제0원칙에 따라 지구의 오염을
 결단한 로봇이지만, 자신의 결정에 괴로워하며 기능이 정지된 최초의
 ‘양심’을 가진 로봇이었다.

논의를 마무리하는 시점에 니체의 니힐리즘을 다시 생각해본다. 기
 독교적 도덕이 가치를 상실한 자리에 처음 만들어지는 허무주의는 소위
 ‘서양 허무주의’로 일컬어지는 ‘허무를 수동적으로 받아들이는’ 수동적
 허무주의와, 이를 ‘새로운 가치 창조의 기회로 삼는’ 능동적 허무주의로
 분류될 수 있다. 니체의 사유에서 영원회귀의 삶을 긍정하는 태도, 운명
 애Amor Fati의 자세는 수동적 허무주의를 능동적 허무주의로 전환하는 직
 접 계기이다. 니체는 허무주의가 필연이자 가장 큰 위기인 동시에 가장
 깊은 자기반성의 순간이 될 수 있으며 그 위기의 지배가 가능하다고 보
 았다.³⁸ 위버멘쉬Übermensch 혹은 초-로봇Überroboter이라 할 수 있을 로봇의
 실존 확인 과정은 포스트/트랜스휴먼 담론을 경유한 니체 사유의 확장
 이기도 하다.³⁹ 비록 작중의 지구는 파괴되었으나 『로봇과 제국』은 작중
 의 로봇들이 그들의 행동 법칙이자 실존의 조건인 로봇공학 3원칙을 직
 접 질문하고 그와 대결하는 과정을 담아내었다.

38 프리드리히 니체, 백승영 역, 『유고 1887년 가을 - 1888년 3월: 니체 전집 20』, 책세상, 2000, 362쪽.

39 닉 보스트롬은 트랜스휴머니즘이란 지금까지 접근불가능했던 가치들에 가닿기 위한 끊임없는 추구로 이루어져 있으며, 이는 철학과 교육, 윤리학이 그동안 플라톤, 아리스토텔레스, 니체 등의 사유를 다루었던 것보다 더 더 나아가 차원으로 끌어올 수 있다고 보았다. Nick Bostrom, op. cit., p.9.

도나 해러웨이는 「사이보그 선언」에서 우리가 모두 ‘키메라’이며, 이론과 공정으로 합성된, 기계와 유기체의 잡종 곧 사이보그라고 한다. 그에 따르면 우리 모두가 사이보그인 상황에서 SF와 사회 현실 사이의 경계는 단지 착시에 불과하다고 강하게 주장한다.⁴⁰ 인간보다 지능과 추론, 예술적 창작 면에서 두루 뛰어난 인공지능 로봇이 개발되고 그에 대한 우려가 확산하는 지금 해러웨이의 표현은 결코 과장이 아니다. 이 글은 그동안 인공지능 기술의 발전에 수많은 전망과 윤리적 질문을 던져주었던 아이작 아시모프의 고전 SF 소설을 다시 읽음으로써 지금-여기에서 휴머니즘의 재편 필요성을 다시 질문하고자 했다. 그 과정은 그간 근대적 서구 휴머니즘의 근간이 되었던 ‘휴먼’이라는 범주가 사실 유럽-건강-남성-백인-잘 생김이라는 비트루비우스적 인간(Vitruvian Man)을 기준으로 삼아 구성된 허구적인 것임을 폭로하였던 비판적 포스트휴머니즘의 논의를 이어받아 ‘포스트휴먼 감수성’을 다시 질문하는 것이었다. 로지 브라이도티는 ‘포스트휴먼 감수성’이 미래에 대한 긍정의 정치학(Affirmative Politics)임을 강조한 바 있다. 브라이도티는 서구 근대 휴머니즘이 옹호하는 단일한 주체 포지션의 쇠락에서 처음으로 생성과 긍정이 발견된다고 강조했다.⁴¹ 『로봇과 제국』에 등장하는 로봇들은 자기 내부의 원리를 초극하는 타자들이다. 인간과 너무나도 닮았고 인간의 명령을 충실히 따름에도 불구하고 이 로봇들은 기존에 설정된 ‘인간’의 범주를 끊임없이 질문하고 휴머니즘의 테두리를 재편한다. 나아가 아시모프의 『로봇과 제국』에 등장하는 로봇 유형이 하나의 은유로서, ‘인간’에 포함되지 못한 이들을 비롯하여 ‘인간’이라는 테두리에 갇혀서 안주하지

40 도나 해러웨이, 황희선 역, 「사이보그 선언」, 『해러웨이 선언문』, 책세상, 2019, 18~19쪽.

41 로지 브라이도티, 이경란 역, 『포스트휴먼』, 이카넷, 2015, 73쪽.

나 그 바깥을 사유하지 못하는 이들까지를 포괄한다는 점은 이 작품이 제공하는 ‘인류’에 대한 총체적인 상을 짐작게 한다.

참고문헌

1. 기본자료

아이작 아시모프, 정철호 역, 『로봇과 제국』 1-2, 현대정보문화사, 2002.

_____, 『강철도시』, 현대정보문화사, 2001.

Isaac Asimov, *Robots and Empire*, Harper Collins, 2018.

2. 단행본 및 논문

Gianmarco Veruggio, "The EURON Roboethics Roadmap", *EURON Roboethics Atelier-Genoa*, 27th of February-3rd of March, 2006. (<http://www.robethics.org/atelier2006/docs/ROBOETHICS%20ROADMAP%20Rel2.1.1.pdf>)

Michael Filas, *Cyborg Subjectivity*, University of Washington PhD Dissertation, 2001.

Mona Håland Aarsland, *The cyborg as a posthuman figure in science fiction literature*, University of Stavanger Master's Thesis, 2015.

Nick Bostrom et al., *Human Enhancement* ed. Julian Savulescu and Nick Bostrom, Oxford University Press, 2013.

Nick Bostrom, "Transhumanist Values", *Ethical Issues for the 21st Century* ed. Frederick Adams, Philosophical Documentation Center Press, 2003. (<https://nickbostrom.com/ethics/values>)

Robin R. Murphy and David D. Woods, "Beyond Asimov: The Three Laws of Responsible Robotics", *Intelligent Systems, IEEE* September, 2009

Sage Leslie-McCarthy, "Asimov's Posthuman Pharisees: The Letter of the Law Versus the Spirit of the Law in Isaac Asimov's Robot Novels", *Law, Culture and the Humanities* 3, 2007

도나 헤러웨이, 황희선 역, 『헤러웨이 선언문』, 책세상, 2019.

로지 브라이도티, 이경란 역, 『포스트휴먼』, 아카넷, 2015.

박인찬, 「포스트휴먼으로 가는 길 - 인간과 기계의 공(共)진화를 중심으로」, 『안과 밖: 영미문학연구』 43, 영미문학연구회, 2017.

변순용, 「인공지능로봇을 위한 윤리 가이드라인 연구 - 인공지능로봇윤리의 4원칙을 중심으로」, 『윤리 교육연구』 47, 한국윤리교육학회, 2018.

세릴 빈트, 진행선 역, 『에스에프 에스프리』, 아르테, 2019.

세릴 빈트·마크 볼드, 송경아 역, 『SF 연대기』, 허블, 2021.

슈테판 헤어브레히터, 김연순·김응준 역, 『포스트휴머니즘』, 성균관대 출판부, 2012.

에마누엘 레비나스, 김도형 외 역, 『전체성과 무한』, 그린비, 2018.

유은순·조미라, 「포스트휴먼 시대의 로봇과 인간의 윤리」, 『한국콘텐츠학회논문지』 18(3), 한국콘텐츠학회, 2018.

이동신, 「망가진 머리: 인공지능과 윤리」, 『인공지능인문학연구』 1, 중앙대 인문콘텐츠연구소, 2018.

이만재, 「아시모프와 로봇」, 『정보과학회지』 14(8), 한국정보과학회, 1996.

지크문트 프로이트, 정장진 역, 『예술, 문학, 정신분석』, 열린책들, 2012.

지현아, 「로봇윤리(robotethics)의 도덕교육적 적용 방안 연구」, 서울대 박사학위논문, 2020.

프리드리히 니체, 백승영 역, 『유고 1887년 가을-1888년 3월: 니체 전집』 20, 책세상, 2000.

3. 기타자료

“Aesopian”, Merriam-Webster Dictionary (<https://www.merriam-webster.com/dictionary/Aesopian>)

Abstract

Robots as 'The Other' and Reconfiguration of Humanism

Re-reading Isaac Asimov's novel *Robots and Empire*(1985)*

Chang, Jun-Young**

As artificial intelligence technology rapidly advances, ethical concerns regarding the coexistence of AI robots and humans has been debated by many debaters. In the process, the question of the roles and ethics of robots and their place among humans has always been a topic of discussion. Isaac Asimov, the American science fiction writer proposed 'the Three Laws of Robotics', which encompass the principles of robot behavior and their use. The inherent contradictions and ethics within the Three Laws of Robotics have been continuously questioned by the author himself and his readers. Among Asimov's works, "Robots and Empire", the final novel out of the 'Robot series', draws attention by adding a "Zeroth Law" to the Three Laws of Robotics, which encompasses the categorization of 'Humanity', emotions, and memory. In this article, the various types of robots appearing in "Robots and Empire" are classified as 'robots as tools (general robots left on planet Solaria and the supervisor robot 'Landaree')', 'robot deriving the Zeroth Law through self-reasoning (Daneel)', and 'telepathic robot with empathy and memory (Giskard)'. By analyzing each of these robot types, this article tracks how Asimov reconfigures the existing humanism based on discrimination and hierarchism through the figure of the robot as 'the other'. In Chapter 2, the supervisor robot on the planet Solaria is analyzed. Robots in Solaria do not resist the Three Laws of Robotics, but they categorize 'humans' and 'non-humans' based on intonation, leading to attack against them—Spacers, Solarians, and migrants (Earthlings). In Chapter 3, this article focuses on the highly sophisticated robot, Daneel, whose reasoning ability is so advanced that it formulates a new principle, the Zeroth Law, based on its rationality. In Chapter 4, this article examines the empathy and memory capabilities of Giskard, the only telepathic robot in the story, to re-examine the issues of ethics and emotions of robots. This article aims to re-evaluate classic science fiction works that have long sparked discussions about the advancement of artificial intelligence technology from a post-humanism perspective.

Keywords Isaac Asimov, Artificial Intelligence, Robot, Three Laws of Robotics, Robots and Empire, Roboethics, Post-Humanism, Transhumanism

이 논문은 2023년 11월 25일에 투고 완료되어
2023년 11월 30일부터 12월 15일까지 심사위원이 심사하고
2023년 12월 20일에 심사위원 및 편집위원 회의에서 게재 결정된 논문임.

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2017S1A6A3A01078538).

** Chung-Ang University.

