

1.1. 축적의 이미지

...

1.1.2. 공작 기계

사람다워지기 위한 노력이 조금씩 쌓이다 보면 사람다움에 가까워지는 과정은 물리적으로는 너무 당연하다. 미미한 변화량이라도 여러 횟수에 걸쳐 누적되면 큰 규모를 이룬다. 파인만(Richard Feynman)은 동일한 작업의 반복에 불과하지만 누적하다 보면 놀라운 결과를 내는 아주 작은 로봇에 대한 강연을 한 적이 있다. 그는 핵실험실에서 위험한 물건을 다룰 때 차폐 유리 밖에 있는 사람의 손동작을 그대로 모방하는 로봇 손을 설명한다. 이 로봇 손으로 같은 기능에 부피만 절반인 로봇 손을 만든다면 절반 짜리 로봇 손이 만든 결과물은 원래 로봇 손이 만든 것의 절반이다. 다시 사람이 원래 로봇 손을 조작하고 원래 로봇 손이 절반 크기의 로봇 손을 조작해서 절반 짜리 로봇 손을 만든다면 그 부피는 원래 크기의 반의반이다. 이 과정을 여러 차례 반복하면 분자 하나를 조작할 수 있는 로봇 손을 가지게 된다. 물론 로봇을 작게 만드는 데에는 여러 공학적 문제가 있다. 그렇지만 그에 상응하는 이점도 있다. 재료가 적게 들고 재료의 강성이 더 낮아도 되니 재료비가 낮아진다. 시간도 크게 문제 되지 않는데 사람 손을 따라 하는 로봇 손을 한 개가 아니라 여러 개 만들면 되기 때문이다.¹⁾

원자 단위를 다루는 로봇 손 제작이 공장같지만 절반 짜리 로봇 손 만들기를 이어 나가는 과정은 논리적으로 타당하다. 문제는 이 계획을 현실화할 때 발생한다. 로봇 손을 제작할 때 들어가는 물질의 특성은 선형적이지 않다. 로봇 손의 뼈대를 구성하는 물질의 강성은 부피에 비례하지 않는다. 또 물질이 작아지면 일상에서는 신경 쓰지 않아도 되는 물리 법칙이 극복하기 어려운 제약으로 작용한다. 로봇 손의 오차도 문제다. 첫 번째 로봇 손이 경미한 오차를 지니고 있어도 마지막 로봇 손까지 이어지는 과정에 로봇 손의 개수가 많아질수록 오차는 누적되기에 마지막 로봇 손이 정상 작동하기 위해서는 개별 로봇 손은 아주 높은 수준의 정밀도를 지녀야 한다. 원자 단위를 다루는 로봇 손을 만들기 위한 도면을 아무리 철저히 그린다고 해도 현실화 과정에서는 늘 문제가 생길 것이다.

원자 크기를 조작하는 로봇 손이 아니라 그저 내 의지를 대행하는 처음 크기의 로봇 손의 개수를 늘려서 내 능력을 최대화할 때는 재료의 물리적 특성과 제작의 공학적 특성에서 오는 한계를 마주하는 문제는 발생하지 않는다. 이 로봇 손을 많이 만들수록 나는 내 의지를 더 많이 구현할 수 있다. 하루 동안 하나의 산출을 낼 수 있는 상황에서 로봇 손을 도입하면 산출물의 개수는 로봇 손의 개수와 비례한다. 그런데 로봇 손으로 내 능력을 증가시키는 작업에는 한계가 있어 보인다. 제작에 들어가는 비용 문제와 구현에 필요한 물질적 제한이 있기 때문이다. 아무리 많은 로봇 손을 만들고 싶어도 내 수명 안에 만들 수 있는 로봇 손은 제한되고 그마저도 자본이 충분할 때의 문제다. 심지어 자본이 충분하고 로봇 손으로 로봇 손을 만들어서 생산되는 로봇 손의 수가 기하급수적으로 증가해도 로봇 손 제작에 필요한 재료가 부족하거나 설치할 수 있는 공간이 부족한 상황에 도달할 수 있다.

원자를 조작하는 로봇 손 만들기의 문제는 처음 크기의 로봇 손 많이 만들기에 있는 문제와 그 특성이 다르다. 전자는 로봇을 작게 만들수록 발생하는 문제의 난도가 급격하게 높아진다.

1) 리처드 파인만, 『발견하는 즐거움』, p. 179-184

단순히 로봇 손의 오차가 중첩되는 문제만 생각해도 반의반 크기의 로봇을 만들기 위해 들일 노력이 절반 크기의 로봇을 만들기 위해 들일 노력과 같을 수 없다. 로봇이 작아지면 작아질수록 더 큰 노력이 들어간다. 분자 단위를 조작하는 로봇을 만들기 위한 처음 절반의 여정보다 남은 여정의 절반이 더 어렵고 또 이 여정보다 남은 여정의 절반이 더 어렵다면 목표 근처에 왔을 때 들어가는 여정은 거의 무한에 가까워서 현실적으로 목적지에 도달하지 못할지도 모른다. 그렇지만 후자는 이미 이를 수 있는 한 과정을 계속 반복하면 되는 일이다. 전자에 있는 자료 자체의 물리적 특성이나 제작의 공학적 특성에 대한 문제는 후자로 오면 사실상 경제적 문제로 전환된다. 반복의 횟수는 내게 주어진 제화와 접근 가능한 공간에 따른다. 전자와 후자 모두 점진적 축적을 통해 거대한 결실로 향한다. 그렇지만 후자보다는 전자가 사람다워짐을 만드는 산출을 평생 조금씩 쌓아가는 과정에 더 가깝다.

1.1.3. 인류

한편으로는 같은 크기의 로봇 손을 만들 수 있는 상한선이 로봇 손의 개수가 늘어남에 따라서 상향된다고 믿어 볼 수 있다. 로봇 손이 많아질수록 내 능력은 더욱 향상된다. 같은 시간을 들여도 내 의지를 따르는 로봇 손이 많으면 많을수록 더 많은 일을 할 수 있다. 이는 더 많은 로봇을 만들 수 있고 또 기술 개발에 더 많은 여력을 사용할 수 있다는 말이 된다. 기술 개발로 로봇의 속도를 높이면 같은 수의 로봇으로 더 많은 일을 할 수 있다. 그뿐만 아니라 우주 개척에 관련된 기술을 확보한다면 물질적 제한이 지구에 한정되지 않을 수도 있고 의료 기술이 향상되어 나의 수명이 늘어난다면 전체 기계 손이 일한 양 자체도 증가한다. 그렇지만 아무리 기술 개발로 개인의 수명이 늘어나도 결국 로봇 손이 살아있는 내 의지를 대행하기를 멈추는 순간은 온다. 로봇 손의 최대 개수는 내가 세상에 있는 동안 만들 수 있는 수로 정해진다.

부류의 존속 기간은 개체보다 훨씬 길다. 인류에 비하면 개인이 누릴 수 있는 수명과 자원은 제한적이다. 로봇 손을 축적해 나가는 방법을 개인이 아닌 인류가 해나간다면 그 결과는 달라질 수 있다. 인류가 충분히 존속하기만 하면 우주의 끝까지 로봇 손으로 가득 채우는 세계를 상상하는 것이 불가능하지 않다. 매일 조금씩 로봇 손의 개수를 늘리며 목표에 다가갈 수 있다. 그렇지만 그 과정은 쉽지 않다. 예측할 수 없는 인류 절멸의 위험을 피하면서 기술 수준을 끌어올리는 노력이 끊임없이 필요하다. 인류의 종신 과업인 것이다.

1.2. 장기주의(Longtermism)

1.2.1. "장기주의"

인류가 미래에 엄청난 수의 로봇 손을 가지고 지금은 상상도 할 수 없는 능력을 지니게 될 가능성을 가지고 있다고 믿으면 지금 당장 로봇 손 하나를 덜 만드는 일의 파급 효과가 엄청나게 크게 계산된다. 로봇 손 제작에 1년이 걸린다고 보고 앞으로 10년 동안 만들어질 로봇 손을 계산할 때 처음에 로봇 손이 1개 있다면 1,024개지만 2개 있다면 2,048개로 그 차이가 작지 않고 시간이 지날수록 이 차이는 더욱 벌어진다. 이를 고려하면 지금 당장 내 작은 기쁨을 위해 인류의 큰 잠재력을 해치게 될지도 모르는 선택을 쉽게 하지 못할 수 있다. 이처럼

지금 우리가 어떤 선택을 할 때 인류의 장기적 미래를 고려해야 한다고 보는 관점으로 장기주의(Longtermism)가 있다. 이 단어는 보스트롬(Nick Bostrom)과 벡스테드(Nicholas Beckstead)의 생각을 토대로 오드(Toby Ord)와 맥어스킬(William MacAskill)이 제안했다.²⁾ (장기주의에 대한 설명은 이 글 뒤에 첨부했다.)

위에 언급한 인물들은 미래 가치에 대해 파핏(Derek Parfit)과 같은 관점을 갖는다. 미래 가치에 대한 시간의 할인율 인정하지 않고 미래 가치를 측정할 때 한 사람이 받는 영향의 측면에서가 아닌 그 사회가 지닌 가치의 총합을 고려해야 한다고 본다. 또 이들은 지금 가지고 있는 증거를 통해 기술에 대한 낙관적 태도를 갖고 기술 발전을 통해 인류가 앞으로 가질 수 있는 모든 가치의 합에 대한 기댓값이 엄청난 규모라는 점을 강조하면서 이를 잃는 선택을 해서는 안 된다고 말한다. 장기주의자에게 핵전쟁 같은 전 지구적 재앙으로 인류가 절멸한 상황과 이런 재앙에서 가까스로 수 명의 사람이 살아남아 인류가 이어질 가능성이 조금이라도 있는 상황 사이에는 엄청나게 큰 차이가 있다. 인류의 생존은 미래에 잠재된 가능성을 실현할 기회가 남아있다는 말이고 인류의 절멸은 미래의 모든 가능성을 잃어버린 상태이기 때문이다.

예를 들어 보스트롬은 인류를 절멸에 이르게 할 존재적 위험(existential risk)에 대처해야 한다고 강조하면서 이 위험의 확률을 감소시키는 것이 우리에게 가장 중요한 과제라고 주장했고³⁾ 벡스테드는 보스트롬의 생각을 전반적으로 강화하면서 인류 존속을 위해서라면 의무론적으로 정당화할 수 없는 선택도 가능하다고 보았다.⁴⁾ 존재적 위험에 집중한 오드는 장기주의의 중요한 관심이 "우리 행동이 장기적 미래에 미치는 영향"에 있다고 설명했고⁵⁾ 인류의 존속이 최선이라는 주장을 전반적으로 보강한 맥어스킬은 장기주의를 "미래 세대의 이익을 보호하기 위해 훨씬 더 많은 일을 해야 한다고 생각하는 관점"으로 정의했다.⁶⁾

이들의 주장에 따라 인류의 잠재력을 해치지 않으려 노력하기 위해서는 우선 내 선택이 미래에 미칠 영향을 계산해야 한다. 그래야 미래 가치의 기댓값이 더 크기에 미래를 중시하는 것처럼 미래의 기댓값을 덜 해치는 선택을 할 수 있다. 그런데 이 계산을 시작하면 작은 로봇 손 만들기에서처럼 끝은 보이지만 도달하기 어려운 난관에 부딪힌다. 어떤 선택이 불러오는 아무리 작은 영향이라도 엄청난 결과로 이어질 수 있고 영향의 고리가 한 단계를 넘어갈 때마다 영향받는 사람의 수는 기하급수적으로 늘어날 수 있기 때문이다. 또 선택의 결과를 결정한다고 해도 나는 미래 세대를 위해 어디까지 노력해야 하는지에 대한 문제도 여전히 남아있다. 미래의 거대한 기댓값에 비하면 한 사람의 가치는 너무 작기 때문이다.

1.2.2. 기획력

장기주의의 설득력이 얼마나 되는지를 떠나서 장기주의가 사회에 미치는 영향력은 무시할 수 없는 수준이라는 점은 눈여겨볼 만하다. 보스트롬과 벡스테드, 오드, 맥어스킬은 보스트롬이 설립한 인류 미래 연구소(Future of Humanity Institute)에 자리를 두고 있었다. 그리고 다른 여러 기관을 설립했다. 특히 오드와 맥어스킬은 현실에서 장기주의를 적용하기 위한 여

2) 토비 오드, 『사피엔스의 멸망』, p. 401

3) Nick Bostrom, "Astronomical Waste: The Opportunity Cost of Delayed Technological Development," p. 311

4) Nicholas Beckstead, "On the Overwhelming Importance of Shaping the Far Future" (PhD Thesis), p. 11

5) 토비 오드, 『사피엔스의 멸망』, p. 65

6) 윌리엄 맥어스킬, 『우리는 미래를 가져다 쓰고 있다』, p. 8

려 기획을 했다. 이들은 효율적 이타주의 운동을 진행하면서 수입의 10%를 기부하는 기빙왓 위캔(Giving What We Can)과 직업을 통해 사회적 문제에 영향을 미치는 방법을 조언하는 8만 시간(80,000Hour)이라는 단체를 공동 설립했다.

효율적 이타주의는 합리적인 자선 활동을 할 것을 강조하는데, 윤리적 소비라는 말이 붙어 있는 옷을 사는 것에 만족하지 말고 실제로 자신이 낸 돈이 어떻게 쓰이는지 확인하고, 효율이 높은 단체에 기부해야 한다는 주장이다.⁷⁾ 이 운동에서 "장기주의"라는 말을 찾기는 어렵다. 그런데 현재 인류가 심각한 재앙을 마주하고 있다고 설명하거나 오직 행복의 증진만 목적에 두고 있다거나⁸⁾ 행복을 최대화하고 과학을 가장 효율적인 자선 단체를 찾기 위한 도구로 사용해야 한다는 설명을⁹⁾ 고려하면 장기주의의 함의를 숨기기 어려워 보인다.

기술 사업으로 부를 이룬 사업가들의 장기주의에 대한 관심도 무시할 수 없는 수준이다. 전 기자동차를 생산하는 테슬라(Tesla Motors)와 화성으로 유인 로켓을 발사하려는 목표를 갖고 있는 스페이스엑스(SpaceX)의 창업자인 머스크(Elon Musk)는 보스트롬의 작업에 대한 공개적 지지를 보냈고 150만 달러를 후원했다.(Pickles, "Elon Musk funds Oxford research into machine intelligence") 400억 달러가 넘게 약정된 후원금을 자랑하던 효율적 이타주의 지원 재단인 오픈 필란트로피(Open Philanthropy)의 재정 일부는 암호 화폐 기업으로 큰돈을 번 뱅크만(Sam Bankman-Fried)으로 부터 나오기도 했다. 장기주의자들의 정치적 영향력도 무시할 수 없다. 오드는 국제 기구나 영국과 미국 행정부의 자문 역할을 맡았고 장기주의(long-termism)를 언급한 국제연합(UN) 사무총장 명의 보고서가 나오는 데 기여하기도 했다. 뱅크만의 후원을 받은 장기주의 인사가 하원의원으로 출마한 사례도 존재한다.¹⁰⁾

1.2.3. 세대 간 문제

먼 미래에 존재할 사람들을 고려하자는 장기주의의 주장은 나 혹은 내 주변만 신경 쓰면 된다는 주장보다 이타적이다. 그렇지만 아직 존재하지 않는 앞으로 태어날 사람들을 고려하는 것은 같은 시대에 사는 모르는 사람들을 고려하는 것과는 다른 문제를 안고 있다. 만약 과거에 살았던 사람들이 미래를 위해 환경 문제에 더 관심을 갖고 대응했다면 지금 환경은 더 나아졌을 수 있다. 그렇지만 만약 과거 세대가 환경 보존을 위해 무분별한 개발을 억제하는 정책을 선택해서 인구 증가율이 더뎠다면 우리 세대 중 일부는 애초에 존재하지 못했을 것이다. 이와 같은 문제는 미래 세대와 현재 세대의 관계라고 해서 달라지지 않는다. 미래 세대는 현재 세대가 없으면 존재할 수 없고 현재 세대가 이루어 놓은 환경에 놓이게 된다.¹¹⁾

장기주의자라면 현세대는 남아있는 인류의 가능성을 훼손하는지 여부를 고려하며 행동해야 한다고 말할 것이다. 적어도 현세대에 아무리 큰 이익이 된다 해도 미래 세대의 존재 기반을

7) 윌리엄 맥어스킬, 『냉정한 이타주의자』

8) William MacAskill, "Effective Altruism," *International Encyclopedia of Ethics*, p. 1

9) William MacAskill, "Understanding Effective Altruism and Its Challenges," *The Palgrave Handbook of Philosophy and Public Policy*, p. 442

10) Émile P. Torres, *Human extinction : a history of the science and ethics of annihilation*, p. 388-389

11) 이 글에서 미래 세대는 현재 세대가 지난 다음에 올 세대를 의미하고 과거 세대는 현재 세대 이전에 있었던 세대를 의미한다. 그렇다면 내가 속한 현재 세대는 어디서부터 어디까지일까? 한 세대를 구분하는 기준은 정말 다양하다. 단순히 시기를 통해 구분하기도 하고 중요한 사건의 발생이나 사회적 특징을 기준으로 삼기도 한다. 세대에 대한 여러 정의는 다음 괄호 속 인용 부분을 참조할 수 있다. (Joerg Chet Tremmel, *A Theory of Intergenerational Justice*, p. 19-28.)

파괴하는 선택을 하지 않을 것이다. 문제는 미래 세대와 현세대는 서로 책임과 의무를 나누어서 지는 관계가 아니라는 점이다. 미래 세대는 현세대가 조성한 기반 위에 존재하게 된다. 그렇지만 아직 존재하지 않는 미래 세대는 현세대의 행동을 완전히 통제할 방법을 가지고 있지 않다. 만약 현세대가 미래 세대를 고려하지 않고 행동한다면 그 결과에 대한 책임은 미래 세대로 전가될 수 있다.

첨부: 장기주의에 관해

2.1. 구조적 지향

2.1.1. 장대한 인류의 존속

장기주의에서는 인류가 엄청난 잠재력을 지니고 있고 그 기댓값이 아직 실현되지 않은 불확실성을 압도하고도 남을 만큼 크다고 평가되기 때문에 인류의 생존에 높은 중요성은 부여된다. 이 주장에는 인류가 지속적인 기술 발전을 이룰 수 있을 것이라는 전제 위에 인류가 오랜 시간 존속하며 수많은 사람들이 엄청난 가치를 실현할 수 있다는 예측에 더해 시간에 따른 가치의 할인율 인정하지 않는 관점으로 이루어져 있다.

장기주의에서 인류 앞에 놓인 미래는 장구하다. 오드는 인류가 존속한 엄청나게 아득한 시간을 귀납적 근거를 바탕으로 제시한다. 화석을 통해 추정된 포유류 한 종의 평균 존속 기간은 대략 백만 년이고 포유류가 아닌 종까지 고려하면 일반적인 생물 한 종은 대략 백만 년에서 천만 년 정도 지속된다. 그렇지만 종의 존속 기간에 한계가 있는 것은 아니다. 대표적으로 앵무조개는 지금까지 5억 년이 넘게 존속하고 있다. 멸종의 위험을 피해 지금까지 존속한 능력이 있는 인류가 앞으로도 위험에 잘 대처한다면 인류의 존속 기간이 앵무조개 수준에 그치는 것이 아니라 적어도 태양의 소멸로 지구가 거주 불가능한 상태가 되는 수십억 년 후까지 이어질 수 있다는 희망적인 전망이 가능하다.¹²⁾

오드는 인류에게 남은 기나긴 시간을 직관적으로 제시하기 위해 여섯 개의 일차원 그래프를 사용한다. 첫 번째 그래프의 범위는 지난 백 년부터 앞으로 백 년이다. 중앙에는 현재가 표시되어 있다. 다음 그래프의 범위는 지난 만 년부터 앞으로 만 년인데 첫 번째 그래프의 전체 기간이 두 번째 그래프에서는 아주 작은 부분만을 차지하게 된다. 이렇게 연쇄되다 마지막 그래프는 우주가 시작한 지난 138억 년부터 앞으로 1조 년까지를 다룬다. 마지막 그래프에서 중앙의 현재부터 우주가 시작된 과거까지의 거리는 현재부터 앞으로 1조 년까지 거리에 비하면 없는 것과 마찬가지인 수준이다.¹³⁾ 이를 통해 그래프의 축척이 커질수록 인류가 보낸 시간은 남은 시간에 비해 아주 미비하다는 것이 강조되어 보여진다.

맥어스킬 또한 오드와 유사한 관점을 시각적으로 설명한다. 인류가 대략 250만 년 전에 탄생했다고 보면 농업은 2만 년 전에 그리고 산업화는 250년 전에 시작되었다. 인류 역사에서 놀라운 변화는 현재에서 가까운 시기에 이루어진 것이다. 이 설명은 그래프를 통해 더 명확히

12) Toby Ord, *The Precipice*, p. 218-221

13) 토비 오드, 『사피엔스의 멸망』, p. 300-301

드러난다. 인류가 지나온 시간을 일차원 그래프로 나타내면 중요한 사건들은 모두 현재 주변에 밀집된 형태로 나타난다. 다른 포유류와 다르게 인류는 환경을 개척하고 위험에 대처할 능력이 있다. 인류가 다가올 위험을 잘 극복하고 지구가 거주 불가능한 상태까지 존속한다고 가정하고 이 기간을 앞서 그래프와 같은 크기로 표현하면 인류의 시작 연도를 가리키는 선과 현재를 가리키는 선은 확대하지 않으면 하나로 보일 수밖에 없을 정도로 서로 아주 가까이 위치하게 된다.¹⁴⁾

인류의 긴 존속 기간을 고려하면 현재 인류가 아주 초기 단계에 있다는 점은 그래프뿐만 아니라 인류가 등장해서 사라지는 전 과정은 한 사람의 생애 주기에 비유하여 설명되기도 한다.¹⁵⁾ 일반적인 포유류의 존속 기간을 인류의 존속 기간이라고 보고 지금 인류의 위치를 한 사람의 일생에 비유해 보면 인류는 이제 막 태어나서 오 개월 된 아기에 해당한다.

장기주의에서 상정하는 인류의 미래는 지구가 거주할 수 없어지는 순간에 끝나지 않는다. 만약 인류가 지구를 벗어날 수 있다면 인류 앞에 놓인 시간은 거의 헤아릴 수 없게 된다. 지구 주변에는 태양보다 오래 유지되는 항성이 있을 것이고 해마다 새로운 항성이 생성된다. 또 어떤 행성은 수십 조년 동안 지속될 것이다. 인류의 존속은 우주적 규모의 시간 위에 놓이게 되는 것이다.¹⁶⁾ 지금 인류는 한 삶으로 보았을 때 이제 막 세상에 나와 눈을 한 번 깜빡인 직후인 것이다.¹⁷⁾

2.1.2. 존속 기간의 기댓값

인류의 존속이 적어도 지구가 주거 불가능할 때까지 이어질 것이라는 말을 어느 정도 믿을 수 있을까? 벡스테드는 인류 존속 기간에 대한 객관적 가능성을 알 수 없고, 이런 논의에서는 높은 정확도가 아니어도 합리적으로 수용할 수 있으니, 내기에서 승률을 계산하는 것 같은 주관적인 확률은 고려할 수 있다고 본다. 그는 지구가 살 수 없는 행성이 될 때까지 인류가 존속할 수 있다고 확실히 믿을 수는 없지만 반대의 상황도 단정할 수 없다고 말하고 보수적으로 보았을 때 인류가 10억 년 동안 존속할 확률을 1%로 제시한다.¹⁸⁾

인류가 우주를 개척할 확률은 어떻게 될까? 벡스테드는 여기에도 같은 입장을 취한다. 인류의 우주 식민지는 철학보다는 공상과학 소설에 어울리는 것이지만 그렇다고 인류가 우주로 나갈 수 없다고 단정하는 것도 비합리적이라고 말한다. 만약 인류가 십억 년 동안 존속한다면 그 안에 우주 개척이 가능한 기술 수준에 도달하지 못하리란 법도 없다. 마침내 우주 개척할 수 있게 된다면 인류가 100조 년 동안 존속하는 일이 불가능하다고만 볼 수는 없다. 그는 인류가 우주 개발을 할 수 있는 기술 수준에 도달할 확률을 1%로 본다. 그렇다면 인류가 지구에서 멸망하지 않고 존재하면서 우주를 개척할 확률은 0.01%(1% * 1%)이다. 이 확률에 따라서 인류에게 기대할 수 있는 남은 시간은 100억 년(0.01% * 100조 년)이 된다.¹⁹⁾

여기서 주관적 확률은 단순히 임의적인 믿음은 아니라 행위자가 의사결정 시점에 확보할 수

14) William MacAskill, *What We Owe the Future*, p. 12-14

15) 윌리엄 맥어스킬, 『우리는 미래를 가져다 쓰고 있다』, p. 18-19

16) Toby Ord, *The Precipice*, p. 223

17) 윌리엄 맥어스킬, 『우리는 미래를 가져다 쓰고 있다』, p. 19

18) Nicholas Beckstead, "On the Overwhelming Importance of Shaping the Far Future" (PhD Thesis), p. 56-57

19) Nicholas Beckstead, "On the Overwhelming Importance of Shaping the Far Future" (PhD Thesis), p. 57-58

있는 증거에 기반한 확률로 볼 수 있다. 보스트롬은 어떤 선택의 결과를 객관적으로 알 수 없을 때 이 선택은 적어도 주관적인 의미에서 위험을 가지고 있다고 말한다.²⁰⁾ 무엇이 들어있는지 알 수 없는 상자를 받았다면 얻을 수 있는 단서를 통해 이득과 위험을 따져봐야 한다. 외관이 말끔한 상자라도 안에 폭발물이 들어있을 확률을 완전히 배제할 수는 없다. 만약 상자에서 조금이라도 의심스러움이 느껴진다면 당장 상자를 열지 않는 선택을 하는 것이 합리적이다. 마찬가지로 인류가 우주를 개발한다는 전제는 당장은 어려워 보이지만 객관적으로 확정된 사실이 아니기에 그 반대의 가정도 배제할 수 없다. 그렇지만 이 가정은 행위자가 파악한 세계에 근거한 주관적 확률에 기반한다. 객관적이지 않은 정보를 통해 행위자가 올바른 판단을 할 수 있는지에 대한 의문은 남아있다.²¹⁾

2.1.3. 기술에 대한 낙관

인류에게 장대한 시간이 남아있다고 주장하기 위해서는 미래에 인류는 지금보다 발전된 기술을 가지고 있다는 전제가 필요하다. 만약 인류의 기술 수준이 현재에 머무른다면 인류에게 남은 시간은 지구가 거주 가능한 환경을 유지하는 기간으로 국한될 수밖에 없기 때문이다. 인류 존속 기간을 추정하는 방식과 유사하게 장기주의에서 인류의 기술 진보에 대한 근거는 지금까지 인류가 이룬 성과를 통해 제시된다. 지금까지 인류가 기술 발전을 도모해왔으니 앞으로도 그럴 것이라는 설명이다. 여기에 더해서 장기주의의 기저에는 지속적인 기술 발전에 대한 소박한 낙관론이 자리하고 있다. 예컨대 오드는 동력을 사용하는 비행기가 발명되고 68년 만에 우주선이 등장한 예를 들면서 인류가 우주로 나아갈 수 있는 기술 수준에 도달하는 것에 큰 문제가 없다는 낙관적인 태도를 보인다.²²⁾

장기주의에서 기술의 진보는 지구를 벗어나는 수준에서 멈추지 않고 계속될 것으로 전망된다. 인류는 기술을 통해 활동 영역을 넓힐 뿐만 아니라 능력을 증대시켜서 더 많은 자원을 활용해 더 많은 인간이 더 큰 행복을 누리는 미래를 맞이할 것이다. 장기주의자들이 생각하는 미래의 기술 수준은 보스트롬의 "유토피아에서 온 편지"라는 짧은 글에서 엿볼 수 있다. 이 글의 화자는 미래에 존재할 수 있는 인간으로 생물학적 신체의 한계를 벗어나 고도의 인지 능력을 갖추고 행복한 삶을 영위하는 모습으로 묘사된다. 그리고 이 상태는 더 이상 변화가 없는 궁극적인 도착지가 아니라 끊임없이 추구해야 할 여정으로 그려진다. 미래의 화자는 지금 인류에게 자신이 될 가능성이 있다고 말하면서 기술의 발전에 걸맞은 지혜를 기르며 나아간다면 미래의 화자 자신이 될 수 있다고 조언한다.²³⁾ 기술을 통한 신체 증강으로 인류가 새로운 상태에 도달할 수 있다는 보스트롬보다는 신중하지만, 오드 역시 인간이 기술을 통해 완전히 새로운 종으로 변화할지 모른다고 말한다.²⁴⁾

장기주의에서 지속적인 기술의 발전은 근본적인 가정으로 여겨진다. 기술 발전을 통해 인류는 사회 문제와 생물학적 한계를 어느 정도 극복할 것이고²⁵⁾ 인류 절멸의 위험을 예측하고 대응할 수 있을 것이다.²⁶⁾ 만약 지속적인 기술의 발전을 이루지 못한다면 인류는 자신의 가능

20) Nick Bostrom, "Existential Risk Prevention as Global Priority," p. 16

21) Christian J. Tarsney, "The Epidemic Challenge to Longtermism," p. 194-195

22) Toby Ord, *The Precipice*, p. 21

23) Nick Bostrom, "Letter from Utopia"

24) 토비 오드, 『사피엔스의 멸망』, p. 258, 322

25) Nick Bostrom, "Existential Risk Prevention as Global Priority," p. 18

26) Nick Bostrom, "Existential Risk Prevention as Global Priority," p. 22

성을 실현하지 못하고 절멸할 것이다. 그래서 장기주의에서 기술 진보의 정체는 인류를 절멸에 이르게 하는 위험 중 하나로 평가된다.²⁷⁾

기술 발전은 긍정적인 결과를 만들기도 하지만 전례 없는 심각한 위험을 초래하기도 한다. 장기주의자들이 기술의 어두운 면을 무시하는 것은 아니다. 과거였다면 국가적 역량을 모아야 이룰 수 있는 결과를 지금은 기술의 발전 덕분에 한 개인이 이룰 수 있다.²⁸⁾ 예를 들어 국가만 생산하고 사용할 수 있던 대량 살상용 화학 물질을 개인이 제조해 살포할 수 있게 되었다. 오히려 앞으로 이런 추세는 가속될 것이지만 그렇다고 기술 발전을 포기하는 결정은 그 자체로 인류를 절멸에 이르게 할 재앙이 된다고 설명한다.²⁹⁾ 그는 기술이 불러오는 문제는 기술을 활용하는 지혜보다 기술 발전이 너무 앞서 나갔기 때문이라고 보기에 그가 선택할 수 있는 방법은 기술 발전의 중단이 아니라 속도 조절뿐이다.³⁰⁾

인류가 기술 발전을 잠시 보류해야 할 유일한 시기는 기술이 충분히 발전해서 인류 절멸의 위험에 대처할 수 있는 안정적인 상태가 되었을 때다. 인류는 이 시기까지 절멸의 위험과 싸워야 한다.³¹⁾ 이 시기는 숙고세(Long Reflection)라고 불리는데 인류가 우선 도달해야 할 이상적 지점이다.(윌리엄 맥어스킬, 『우리는 미래를 가져다 쓰고 있다』, p. 150) 숙고세에 도달한 인류는 열린 미래의 가능성을 탐색하면서 자신의 잠재력을 최대한 실현할 수 있는 미래 경로를 신중하게 선택해야 한다.³²⁾

2.1.4. 미래 가치의 비할인

인류에게 엄청나게 긴 시간이 주어져 있고 그 시간 동안 수많은 사람들이 행복하게 존재한다고 해서 인류의 미래에 놓인 가치가 현재보다 크다고 추산할 수는 없다. 시간에 따른 할인율이 적용된다면 아무리 큰 가치여도 먼 미래에 있다면 현재로 환산했을 때에는 미미한 가치일 수 있다. 그러나 장기주의에서는 미래 가치에 대한 할인을 인정하지 않는다. 나에게 멀리 떨어진 곳에 있는 사람의 고통이 가까이 있는 사람과 동일한 고통이듯이 먼 과거의 삶이나 먼 미래의 삶 모두 동등하다고 주장한다.³³⁾ 따라서 인류의 미래 가치를 따질 때 긴 시간에 따르는 지수적 가치 하락을 고려하지 않을 수 있게 된다.

일반적으로 아직 실현되지 않은 미래 가치는 할인되어서 현재 가치보다 낮게 평가된다. 지금 당장 100원을 받는 것과 10년 뒤에 100원을 받는 것 중 하나를 선택해야 한다면 지금 당장 100원을 받는 것이 이득이다. 10년 동안 은행에 이 100원을 저축한다면 이율에 따른 이자를 받을 수 있기 때문이다. 만약 이율이 7.2%라면 100원을 대략 10년 동안 저축한다면 200원이 된다. 이 10년 뒤의 200원을 현재 가치로 할인하면 100원이 된다. 두 선택이 현재 가치로 환산했을 때 같지만 그래도 지금 당장 100원을 받는 것이 유리한데 200원을 받을 10년 안에 어떤 일이 일어날지 모르기 때문이다. 최악의 상황에는 불의의 사고로 내가 세상에 존재하지 않아서 저 200원을 받을 기회 자체를 영영 잃어버릴 수도 있다.

그렇지만 미래 가치를 현재와 시간적 거리에 비례해서 단순히 할인하는 것은 받아들일 수

27) 윌리엄 맥어스킬, 『우리는 미래를 가져다 쓰고 있다』, p. 211

28) 토비 오드, 『사피엔스의 멸망』, p. 178

29) 토비 오드, 『사피엔스의 멸망』, p. 210

30) 토비 오드, 『사피엔스의 멸망』, p. 277

31) Bostrom, "Existential Risk Prevention as Global Priority," p. 24

32) 토비 오드, 『사피엔스의 멸망』, p. 256

33) 토비 오드, 『사피엔스의 멸망』, p. 64. William MacAskill, *What We Owe the Future*, p. 17

없는 결과를 낳는다. 엄청나게 먼 미래에 엄청나게 많은 인류 모두가 큰 행복을 누리며 산다고 해도 조금의 할인율을 인정하면 미래의 가치의 합을 현재 가치로 환산했을 때 현세대에게는 아무런 의미가 없는 숫자가 될 수 있다. 10^{16} 년 뒤의 10^{16} 단위만큼의 가치를 할인율을 7.2%로 보고 현재 가치로 환산하면 대략 4단위밖에 되지 않는다. 이렇게 단순히 할인율을 적용하면, 생명 이외의 것이 동일하다고 보았을 때, 지금 1명의 생명을 구하는 것이 100년 후 1,045명의 생명을 구하는 것과 동일하다는 결론에 이른다.³⁴⁾

그 외에도 여러 반론이 있지만 그중 하나만 소개하면 조금 적은 가치를 받더라도 가능한 이른 시간 안에 받는 선택에 대한 선호는 한 사람의 삶 안으로 한정된 상황에서만 타당하다는 것이다. 대개 사람들은 앞서 예에서 10년 뒤에 받는 보상이 210원이라도 보통 당장 100원을 받길 원할 것이다. 그렇지만 현세대가 100원을 받는 것과 10년 뒤 다음 세대가 210원을 받는 것 사이에서는 현세대가 100원을 받아야 하는 이유를 찾기는 어렵다. 한 사람의 선택이라면 210원을 받기 위해서는 10년을 기다려야 하지만 다음 세대가 210원을 받는다면 이를 기다려야 하는 사람은 없다. 이처럼 선택에 있어서 앞선 시간에 대한 선호는 세대간 문제에서는 적용되지 않는다.³⁵⁾

2.1.5. 엄청난 수의 미래 인구

아무리 긴 시간 동안 인류가 생존해도 인구가 적으면 장기주의적 입장에서 인류의 잠재력이 많이 남았다고 볼 수 없다. 한 사람이 실현할 수 있는 가치가 동일하다고 가정하면 천 년 동안 400억 명의 인구가 존재하는 것이 만 년 동안 40억 명의 인구가 존재하는 것보다 가치 있다. 맥어스킬은 앞으로 인구가 지금보다 많아질지 적어질지 모르지만 많아진다면 지금보다 훨씬 많아질 것이라고 말한다. 현재 인구는 수렵채집 시대의 인구보다 천 배는 더 많아졌기 때문에 만약 우주 개척 시대에 이르면 셀 수 없이 많은 인구가 있을 것이라 예측이다.³⁶⁾ 그는 앞으로 얼마나 많은 사람들이 존재할지 시각적으로 제시한다. 우선 사람 아이콘 하나가 인구 백억 명을 나타낸다고 설정한다. 이렇게 보면 현재 존재하는 인구는 대략잡아 아이콘 하나고 과거에 존재했던 인구는 아이콘 10개밖에 되지 않는다. 그리고 인구가 더이상 증가하지 않고 현재 상태로 인류가 지구가 황폐해질 때까지 생존할 때 존재하는 인구를 표현하기 시작한다. 같은 아이콘이 3쪽에 걸쳐 나열되고 앞으로 2.2만 쪽이 필요하다고 설명한다.³⁷⁾

인류가 우주로 나가 살 수 있게 된다면 예측할 수 있는 인구는 더 커진다. 보스트롬은 천문학적 추정치에 근거해서 미래에 개별 인류 수명의 합을 제시한다. 인류가 우주 개발을 할 수 있는 기술 수준에 도달한다면 남은 인류의 수명은 적어도 10^{34} 년이다. 만약 인간 정신이 생물학적인 신체가 아니라 컴퓨터 하드웨어 위에서 구현될 수 있다면 이 수치는 10^{54} 년까지 늘어난다. 또 인류의 기술 수준이 물리학적 한계에 도달하게 된다면 예측치는 상상할 수 없을 정도로 커질 것이다.³⁸⁾

2.1.6. 잠재 가치의 추산

34) Andreas L. Mogenden, "Moral Demands and The Far Future," p. 574

35) Tyler Cowen & Derek Parfit, "Against the social discount rate," *Justice between age groups and generations*, p. 154-155

36) William MacAskill, *What We Owe the Future*, p. 14-15

37) 윌리엄 맥어스킬, 『우리는 미래를 가져다 쓰고 있다』, p. 35-39

38) Nick Bostrom, "Existential Risk Prevention as Global Priority," p. 18

미래에 대한 할인을 인정하지 않으면서 엄청나게 긴 시간 동안 엄청나게 많은 사람이 각자의 가능성을 실현해 나간다고 가정하면 현재에 산출한 미래의 가치는 엄청나게 크다. 그렇지만 막연히 크다는 말을 계산에 걸맞은 구체적인 숫자로 바꾸는 일은 쉽지 않아 보인다. 미래에 놓인 가치를 계산하기 위해서는 앞으로 실현될 가치를 측정하고 이를 모두 더하는 난해한 작업이 필요하기 때문이다.

벡스테드는 인류의 미래 가치를 예측하는 컴퓨터 프로그램의 구조를 제안한다.³⁹⁾ 이 프로그램은 미래를 특정 시대로 나누어서 각 시대의 가치를 측정하고 각각의 값을 합하는 방식으로 미래의 가치를 예측한다. 정확한 계산을 위해서는 한 시대를 구체적으로 기술해야 하고 또 그 안에서 발생한 사건에 대한 가치 측정 방법 명세가 필요하다. 그렇지만 벡스테드의 접근은 계산에 대한 상세한 방법을 규정하지 않아도 미래 세대의 가치를 판단할 수 있는 대략적인 수치를 제시할 수 있다.

일단 이 프로그램에서 각 기간은 가치 평가에 서로 영향을 미치지 않는다. 한 시대는 이전 시대의 가치 변동 추세와 무관하게 그 시대에서 발생한 일로만 평가된다. 한 시대와 그 이전 시대가 인과적인 관련성 부정하는 의미는 아니다. 이전 시대의 가치 추이와 관련 없이 한 시대의 평가는 그 시대 안에 있었던 일로 이루어진다는 말이다. 그리고 각 시대는 기준이 되는 좋은 일이 일어난 횟수로 평가된다. 기준이 되는 좋은 일이 동일한 횟수로 일어났다면 그 앞에 놓인 시대의 가치가 어떻게 변화했는지와 상관없이 같은 값으로 계산된다.

제안된 구조에 따라서 미래의 가치를 계산해 보자. 만약 인류에게 앞으로 주어진 시간을 100조 년이라고 가정하고 백 년을 한 시대로 놓는다면 현재 우리가 속한 시대 앞에 1조 개의 시대가 있다. 장기주의자의 희망처럼 인류가 계속해서 옳은 길로 나간다면 다음에 올 시대는 늘 앞선 시대보다 높은 가치를 갖는다. 그렇다면 미래의 가치는 현재의 가치보다 적어도 1조 배 크다. 달리 말하면 미래가 현재보다 1조 배 중요하다. 미래의 불확실성을 고려해서 미래의 가치를 평가해 볼 수도 있다. 앞서 인류가 지구가 거주 불가능이 되기 전에 우주를 개척할 확률은 0.01%였다. 그렇다면 기대할 수 있는 인류의 존속 가능 기간은 100억 년이다. 미래 가치가 현재 시대 수준으로 유지된다고 가정하면 미래 가치는 현재 가치보다 1억 배 많게 된다.⁴⁰⁾

2.1.7. 당혹스러운 결론

이렇게 결과적인 가치의 총합으로 미래의 중요성을 설명하면 '당혹스러운 결론'과 마주하게 된다. 이 결론에 따르면 가치의 총합이 가장 높은 시대가 직관적으로 가장 낮다고 말할 수 없는 상황이 언제나 존재한다.⁴¹⁾ 내가 살아가야 할 시대를 선택한다고 생각해 보자. 가치의 총합을 좋고 나쁨의 기준으로 삼는다면 가장 높은 가치를 갖는 시대를 선택해야 한다. 만약 가치의 총합은 같지만 인구가 크게 차이 나는 두 시대가 있다면 어떤 선택을 해도 차이가 없을 까? 1시대와 2시대는 모두 만 단위만큼 가치를 갖고 있다. 1시대의 인구는 만 명이고 2시

39) Nicholas Beckstead, "On the Overwhelming Importance of Shaping the Far Future" (PhD Thesis), p. 58-64

40) Nicholas Beckstead, "On the Overwhelming Importance of Shaping the Far Future" (PhD Thesis), p. 67

41) Derek Parfit, *Reasons and Persons*, p. 388-389

대의 인구는 백 명이다. 시대의 가치를 시대에 속한 사람들이 공평하게 나누어 갖는다면 1시대에 속하는 사람은 한 개의 가치를 갖고 2시대에 속하는 사람은 백 개의 가치를 갖는다. 한 사람이 겪는 좋고 나쁨을 따진다면 2시대를 선택하는 것이 더 낫다. 이번에는 1시대와 인구와 가치가 같지만 9천의 가치를 한 사람이 독점하는 2시대를 생각해 보자. 2시대에서 가치를 독점하는 그 한 사람으로 산다는 보장이 없다면 1시대를 선택하는 2시대를 선택하는 것 사이에 아무런 차이도 없다고 말하기는 힘들다.

한 사람이 겪는 좋고 나쁨을 고려하지 않으면서 당혹스러운 결론을 피하기 위해 가치를 합산하는 과정에 단서를 추가하기도 한다. 단순히 가치의 총합으로 두 기간을 비교하는 것이 문제라면 한 시대의 가치의 총합을 인구수로 나눈 평균을 비교해 볼 수 있다. 그렇지만 앞서 보았듯이 이 방법으로는 평등한 1시대와 불평등한 2시대를 구분할 수 없다. 한 시대의 가치를 합산하는 과정에 기준 수준을 도입해 볼 수도 있다. 기준을 넘지 못한 가치를 지닌 사람을 가치 합산에서 제외한다면 1시대와 2시대를 구분할 수 있다. 하지만 이 방법에서 기준을 9로 삼고 천 명의 인구가 모두 10의 가치를 갖는 2 시대는 2시대와 구분되지 않는다.⁴²⁾ 또 적절한 기준을 설정하기 어려운 문제도 있다.⁴³⁾ 만약 기준 이하는 없는 가치로 보면 기준에 겨우 미치지 못하는 사람만 존재하는 시대와 기준에 한참 미치지 못하는 사람들만 존재하는 시대가 구분되지 않는다. 기준 이하 가치를 음수로 놓는다면 기준에 한참 미치지 못하는 적은 사람이 있는 시대보다 기준에 겨우 미치지 못하는 다수의 사람이 있는 시대가 더 나은 시대로 판정될 수 있다.

당혹스러운 결론을 피하고 싶다면 한 사람이 겪는 영향을 고려하는 입장에 서면 된다. 그렇지만 이 선택은 '비동일성 문제'를 일으킨다.⁴⁴⁾ 이 문제 상황에서는 아직 태어나지 않은 사람에게 영향을 미치는 선택을 판단하는데 선택 자체가 미래 사람의 존재에 영향을 미치기에 선택에 대한 판단이 매우 어려워진다.⁴⁵⁾ 현재의 선택은 미래에 존재할 수 있는 사람에게 영향을 미친다. 열네 살인 여자가 아이를 갖기로 마음먹었다고 생각해 보자. 열네 살은 아이를 키우기에는 너무 버거운 나이다. 많은 사람들은 여자와 아이를 걱정하며 어른이 되고 안정을 이루었을 때 아이를 가지라고 만류할 것이다. 그런데 이 여자가 열네 살 때 낳은 아이와 나중에 어른이 되고 나서 낳은 아이는 동일한 아이가 아니다. 두 경우에 무엇이 더 좋은지 비교하기는 어렵다. 세상에 존재하지 않는 것보다는 불안한 환경에서 자라나는 것이 태어날 아이 입장에서는 더 나을 수도 있다. 같은 사람에게 영향을 미치는 다른 선택은 서로 비교할 수 있지만 이 여자의 선택에 따라 좋고 나쁨을 겪는 아이는 같은 사람이 아니다.

비동일성 문제에 빠지는 것도 당혹스러운 결론은 감내하는 것도 모두 받아들이기 어렵지만 맥어스킬의 입장은 당혹스러운 결론을 크게 문제 삼지 않고 덤덤히 받아들이는 것이다.⁴⁶⁾ 극단적으로 말하면 장기주의자들에게 살아갈 가치가 겨우 있는 (굳이 수치로 말하면 일의 가치를 갖는 사람 1조 명이 있는) 시대가 천의 가치를 갖는 사람 1억 명이 있는 시대보다 낫다.

2.2. 행동의 촉구

42) Derek Parfit, *Reasons and Persons*, p. 434

43) Derek Parfit, "Can We Avoid the Repugnant Conclusion?," p. 111

44) Derek Parfit, *Reasons and Persons*, p. 400

45) Derek Parfit, *Reasons and Persons*, p. 356-361

46) 윌리엄 맥어스킬, 『우리는 미래를 가져다 쓰고 있다』, p. 263

2.2.1. 존재적 위험

인류 앞에 놓인 엄청난 가치를 중요하게 여긴다면 가장 피하고 싶은 상황 중 하나는 인류의 절멸이다. 인류가 절멸하면 더 이상 실현할 잠재력도 없다. 그래서 오늘 갑자기 인구의 99%가 죽는 것보다 인구의 100%가 죽어서 절멸하는 것이 훨씬 나쁘다. 두 경우에 죽은 사람의 차이는 1%지만 그 나쁜 정도의 차이는 1%가 아니다. 인류가 잠재력을 이루어 갈 가능성이 있는 상태와 없는 상태 사이에는 비교할 수 없는 큰 차이가 있다.⁴⁷⁾

인류를 절멸에 이르게 하는 위험이 다가오는 경로는 다양하다. 태양의 소멸이나 거대한 운석 충돌같이 외계에서 다가올 수도 있고 핵전쟁이나 환경 오염같이 인류의 무분별한 행동이 원인이 될 수도 있다. 인류가 물리적으로 멸종하는 것뿐만 아니라 인류가 잠재력을 실현할 가능성을 잃어버리는 것 또한 장기주의자에게 좋은 일이 아니다. 만약 인류가 잠재력을 실현할 가능성을 영원히 잃어버렸다면 인류 절멸과 다를 바 없는 상황이다. 보스트롬은 지구에서 유래한 지적 생명체를 멸종시키거나 그 잠재력을 영구적으로 엄청나게 감쇄시키는 위험을 존재적 위험(existential risk)이라고 정의한다.⁴⁸⁾

존재적 위험에 대처하기 위해 인류는 지속적인 기술의 진보가 필요하다. 지구에서 거주할 수 없어지는 존재적 위험을 마주했을 때 지구를 벗어날 수 있는 기술이 없다면 인류는 절멸하게 된다. 그리고 인류의 가능성을 실현을 위해서 기술이 필요하기에 기술에 대한 발전이 없다는 말은 인류의 가능성이 실현되는 규모가 더이상 증가하지 않는다는 것이다. 이는 인류가 이를 수 있는 가치를 실현하지 못하는 상황이므로 역시 존재적 위험이다. 한편으로는 기술 발전에 걸맞은 지혜를 갖지 못한다면 존재적 위험이 증가할 것이다. 존재적 위험을 초래할 수 있는 핵무기는 다행히 재료가 비싸고 다루기 어려워져 대중이 사용할 수 없지만 앞으로 등장할 존재적 위험을 초래할 기술이 다루기 까다로울 것이라는 보장은 없다. 그렇다고 기술이 발전을 멈추는 것은 또 다른 존재적 위험에 처하는 것이다. 로켓은 발사 직후가 가장 불안하지만 멈출 수 없다. 대기권을 돌파해서 안정된 상태로 나아가는 방법뿐이다.⁴⁹⁾

지금 인류는 멈춰있는 로켓이 아니다. 운석같이 외부에서 오는 위험에 대처할 수 있는 기술은 없지만 놀라운 발전을 통해 인류 전체의 운명을 결정할 수 있는 파괴력을 지닌 무기를 만들 수 있는 기술 수준에는 이르렀다. 오드는 인류의 상황을 이제 막 몸이 크고 힘이 생기고 있지만 아직 힘을 적절히 사용할 지혜가 부족해 성인 같은 자유가 주어지지 않는 아이에 비유한다.⁵⁰⁾ 이제 막 운전석에 앉아서 페달에 발이 닿게 된 아이가 집에 있는 자동차를 몰래 타고 나온다면 끔찍한 일이 벌어질 수도 있을 것이다.

존재적 위험을 줄이는 일의 가치는 엄청나게 크다. 앞서 보수적으로 계산한 남은 인류의 합산 수명 10^{34} 년이었다. 모든 사람이 한 단위만큼 가치를 갖고 수명도 백 년으로 같다고 가정하자. 그럼 미래에 10^{32} 단위만큼 가치가 있다. 만약 지금 인류가 절멸한다면 10^{32} 의 가치를 잃는다. 존재적 위험을 백만 분의 일%p 낮추는 것은 인구 10^{24} 명만큼의 가치가 있다. 만약 합산 수명이 10^{34} 년이 될 가능성을 1%라고 보더라도 존재적 위험을 백만분의 일%p 낮추는 것은 인구 10^{22} 명 만큼의 가치가 있다.

47) Derek Parfit, *Reasons and Persons*, p. 453-454

48) Nick Bostrom, "Existential Risks: Analyzing Human Extinction Scenarios And Related Hazards," p. 4

49) Nick Bostrom, "Existential Risk Prevention as Global Priority," p. 25-26

50) Toby Ord, *The Precipice*, p. 206-207

존재적 위험은 한번 발생하면 되돌릴 수 없기에 미리 준비해야 한다. 그렇지만 존재적 위험에 대처하기 위한 지구적 동의를 이끌어내는 것은 쉽지 않은 일이다. 존재적 위험을 감소시키는 노력으로 인한 혜택은 전 인류가 구분 없이 받는다. 위험 감소를 위한 노력을 하지 않은 사람을 혜택에서 배제할 수 없다. 노력을 많이 했다고 더 큰 혜택을 받는 것도 아니다. 더 심각한 점은 이런 노력에 대한 보상이 노력한 사람이 아니라 먼 미래 후손에게 주어질 수 있다는 점이다.⁵¹⁾

2.2.2. 현세대의 중요성

존재적 위험은 늘 있었지만 인류는 존재적 위험을 관리하는 일에 큰 노력을 들이지 않고 지금까지 존속하고 있다. 과거 세대가 그랬던 것처럼 현세대도 존재적 위험에 대한 대처를 하지 않는다고 해서 큰 문제가 생기지 않을지도 모른다. 엄청나게 긴 인류의 존속 기간을 생각해 보면 현세대가 차지하는 부분은 아주 작으니 존재적 위험에 지금 대처하나 다음 혹은 다다음 세대가 대처하나 큰 차이가 없을 수도 있다.

그렇지만 현세대는 무심코 행동한 작은 사건으로도 앞으로 남은 인류의 발전 방향에 지속적인 영향을 미칠 수 있다. 존재적 위험을 줄이는 직접적인 노력만이 아니라 일상적인 일도 인류의 미래에 큰 영향을 줄 수 있다. 벡스테드는 인류의 미래를 궤적으로 생각하면 어떤 시점에 발생한 작은 사건이 그 이후 궤적의 방향에 지속적인 영향을 미칠 수 있다고 설명한다.⁵²⁾ 그리고 그는 이 점을 고려하면 인류의 궤도를 좋은 쪽으로 변화시키는 노력이 중요하다고 보는데 먼 미래에 미칠 영향을 고려하며 일상적인 일의 중요성을 놓치지 말아야 한다고 말한다.⁵³⁾

맥어스킬은 자신이 십 대 시절에 목숨을 잃을 수도 있던 위험한 행동을 분별없이 했지만 다행히 살아있다고 안도하면서 아직 미성숙한 시절의 선택이 나머지 삶 전체에 중대한 영향을 미칠 수 있다고 설명한다. 그는 인류 역시 무분별한 십 대와 비슷한 상태라고 말한다. 우리는 지금 남은 인류를 존속시키거나 멸망시킬 수 있으며 인류의 향방에 지대한 영향을 미칠 수 있다. 따라서 현세대는 인류가 파국으로 향하는 방향에서 최대한 멀어져 안정적인 성장 궤적을 그릴 수 있게 만드는 선택을 해나가야 한다고 주장한다.⁵⁴⁾

한 사람의 삶에서 어린 시절은 정말 중요하다. 어린 시절에 보고 배운 것은 남은 삶에 지속적으로 영향을 미친다. 어린 시절 우연히 마주한 경험이 삶의 궤적을 좋은 방향으로 급격히 바꾸기도 하고 어린 시절 무의식적인 행동이 평생 고쳐지지 않는 나쁜 습관이 되기도 한다. 맥어스킬은 역사적 사례를 들며 인류의 가치관도 우연한 사건들로 인해 형성되어 유지되다가 또 다른 우연한 사건들로 인해 변화한다고 설명한다. 한 번 만들어진 가치관은 관성을 유지하면서 다양성에 대한 변화를 떨어뜨리고 새로운 위험에 대한 대응력을 약화시킨다. 그리고 언젠가 형성된 가치관은 변화의 여지 없이 고착될 것이라고 설명한다.⁵⁵⁾ 그리고 그는 인공지능의 출현으로 다음에 일어날 가치관 고착이 영원히 지속될지 모르고 혹여 잘못된 가치관이 고

51) Nick Bostrom, "Existential Risk Prevention as Global Priority," p. 26-27

52) Nicholas Beckstead, "On the Overwhelming Importance of Shaping the Far Future" (PhD Thesis), p. 6-7

53) Nicholas Beckstead, "On the Overwhelming Importance of Shaping the Far Future" (PhD Thesis), p. 72

54) William MacAskill, *What We Owe the Future*, p. 34-35

55) 윌리엄 맥어스킬, 『우리는 미래를 가져다 쓰고 있다』, p. 121

착된다면 인류를 절멸에 이르게 할 수 있다는 점을 염려한다.⁵⁶⁾ 그래서 다음 가치관 고착이 일어나기 전에 가능한 많은 도덕적 진보를 이루어야 한다고 주장한다.⁵⁷⁾

현세대가 존재적 위험에 대처해야 하는 이유는 현세대가 인류의 미래를 결정하는 정말 중요한 시기에 있기 때문이다. 현세대가 중요하다는 주장에 대한 반박으로 인류의 영향력이 지금까지 증가해 온 것처럼 미래에도 증가해서 미래 세대가 더 많은 일을 할 수 있다는 점과 현세대의 중요성과 관련된 미래 세대의 잠재력에 대한 증거가 없고 확률 또한 너무 낮다는 점이 제기될 수 있다. 이에 대해 맥어스킬은 지금 같은 인류의 진보는 유례없는 것으로 앞으로 같은 것이라는 확신은 어렵다는 점과 앞으로 여러 행성에 인류가 거주하게 되면 지금처럼 모든 사람들이 한 행성에 살아서 인류 전체에 큰 영향력을 줄 수 있는 시기가 없을 것이라는 점을 들며 대응한다.⁵⁸⁾

2.2.3. 과도한 요구

장기주의에서 현세대가 인류의 존속을 고려해야 하는 가장 큰 이유는 미래의 가치와 현재의 가치를 저울질했을 때 미래 가치가 현저하게 무겁기 때문이다. 이를 한 사람의 상황으로 단순하게 보면 지금 나의 작은 즐거움이 미래의 수많은 사람에게 큰 고통으로 전가된다면 이 즐거움을 선택하지 말아야 하고 반대로 지금 나의 작은 고통이 미래의 수많은 사람에게 큰 행복으로 보상된다면 이 고통을 감수해야 한다.

문제는 존재적 위험을 감소시키는 선택을 어디까지 해야 하는지에 있다. 최대 가치를 선택하는 규칙을 충실히 따르면 고대하던 열대의 섬으로 휴가를 포기하고 이 비용을 존재적 위험 감소에 도움이 되는 방향으로 사용해야 하는 상황에 이르게 될 것으로 보인다. 결과를 놓고 판단하는 규칙에 과도한 요구가 있다는 비판에 대해 과도한 요구는 특정 환경에서 발생하는 것이지 규칙 자체에서 발생하는 것이 아니라고 반론할 수 있다.⁵⁹⁾ 그렇지만 장기주의에서 미래 세대의 가치가 현세대와 동등하거나 낮아서 현세대의 미래 세대에 대한 노력이 필요치 않은 상황은 우주가 소멸하기 직전의, 우주의 시간으로 보았을 때는 찰나인, 시기밖에 없다.⁶⁰⁾

만약 행위자가 가치의 비중을 조절할 수 있다면 과도한 요구를 줄여지지 않을 수 있다. 그렇지만 행위자 중심으로 가치를 판단한다고 해도 장기주의에서 미래의 가치는 너무나 크다. 행위자에게 직접적으로 관련 있는 일이 갖는 가치의 비중을 높이고 미래 가치의 비중을 낮춰도 가까운 일의 가치는 먼 가치의 합과 비교하면 너무나도 작다. 그렇다고 미래 가치의 가중치를 1보다 낮게 두어 미래 가치에 할인을 인정할 수는 없는 일이기에 행위자에게 관련된 가치와 미래 가치 사이에 적절한 가중치를 찾는 것은 쉽지 않은 일이다.⁶¹⁾

2.2.4. 광신적 행동

56) 윌리엄 맥어스킬, 『우리는 미래를 가져다 쓰고 있다』, p. 121-122

57) 윌리엄 맥어스킬, 『우리는 미래를 가져다 쓰고 있다』, p. 150

58) William MacAskill, "Are We Living at the Hinge of History?"

59) 카타르지나 드 라자리-라텍 & 피터 싱어, 『공리주의 입문』, p. 138

60) 물리적 제약으로 인류가 멸종하는 것은 장기주의 전략에서 실패가 아니라 성공이다.(Nick Bostrom, "Existential Risk Prevention as Global Priority," p. 22) 만약 우주의 소멸로 인해 인류 존속 기간의 기댓값이 얼마 남지 않은 상황이라면 현세대보다 미래 세대의 가치가 가벼운 순간이 올 것이다.

61) Andreas L. Mogendern, "Moral Demands and The Far Future," p. 575-577

미래 가치의 기댓값은 극소 확률과 극댓값을 곱하는 형식으로 계산된다. 미미한 확률은 일상에서 거의 고려되지 않는다. 길을 건다가 운석에 맞을 확률이 없지는 않지만 이 일을 걱정하지 않는 것은 그 확률이 거의 0에 가까울 정도로 작기 때문이다. 인류의 미래 가치를 구할 때 사용되는 확률도 아주 작은 확률이다. 이렇게 작은 확률일지라도 0이 아닌 이상 엄청나게 큰 값에 곱한다면 유의미한 값이 나온다.

이때 작은 확률은 기댓값을 무의미하게 만드는 것이 아니라 보수적 가정이라는 표현처럼 계산 결과의 신뢰도를 높이는 방향으로 사용된다. 예측보다 낮은 확률로 계산했다는 것은 계산이 조금 틀려도 인정할 수 있는 범위 안에 들어올 여지를 만든다. 어떤 일이 발생할 확률을 낮게 잡을수록 더 보수적으로 그 일에 대처할 수 있다. 비가 오지 않을 확률을 50%로 잡는 쪽보다 10%로 잡는 쪽이 비가 왔을 때 우산을 가지고 있을 확률이 더 높을 것이다.

그런데 극소 확률과 극댓값을 곱해 나오는 값이 유의미하다고 인정한다면 납득하기 어려운 상황이 발생한다. 기댓값을 최대화하는 선택을 하는 수학자에게 복면을 쓴 강도가 다가와서 자신이 다음날 일어날 인류 절멸을 막을 수 있는 마법의 힘이 있다는 것을 어느 정도의 확률로 믿냐고 물었다. 수학자는 강도의 제안이 극히 희박하지만 객관적으로 배제할 수 없는 일이기에 천만경분의 일이라고 답했다. 그러자 강도는 수학자에게 지갑을 주지 않으면 인류 멸망을 막기 위한 마법을 시전하지 않을 것이라고 위협했다. 대부분의 사람은 이를 무시하겠지만 존재적 위험을 막는 일의 가치가 너무 커서 강도의 말을 무시할 수 없다. 앞으로 남은 인류의 합산 수명의 기댓값을 10^{32} 년으로 보면 강도가 마법을 부려 구할 수 있는 가치의 기댓값은 $10^{10}(10^{32} * 1/10^{23})$ 만큼이다. 평균 수명을 100년으로 가정하면 10^8 명만큼의 가치다. 수학자의 지갑이 10^8 명의 가치보다 높기는 어렵기에 수학자는 강도에게 지갑을 넘길 수밖에 없다.⁶²⁾

실제로 보스트롬은 가치의 총량을 최대화하는 규칙을 따르기로 했다면 결국 존재적 위험을 감소시키는 선택에 다른 어떤 일보다 가장 높은 우선순위를 부여할 수밖에 없다고 주장한다.⁶³⁾ 그래서 보스트롬은 개인이 손쉽게 대량살상 무기를 운용할 수 있는 시대에는 존재적 위험에 대처하기 위해 전 국민을 대상으로 하는 세밀한 감시 시스템을 구축하는 것을 진지하게 고민해야 한다고 주장한다.⁶⁴⁾ 심지어 벡스테드는 다른 모든 것이 동일하다는 조건을 달기는 하지만 미래 궤적에 미치는 영향이 더 크기에 부유한 나라의 사람을 구하는 것이 가난한 나라의 사람을 구하는 것보다 확연히 더 중요하다고 주장한다.⁶⁵⁾

아주 작은 확률에 기대는 광신적 행동은 선뜻 납득하기 어려워 보인다. 이에 대한 장기주의자의 대응은 다양하다. 그중 한 전략은 확률이 높은 결과를 추구한다면 불합리한 상황에 빠질 수 있음을 보이는 것이다. 예를 들어 광신적 행동을 받아들이지 않으면 이익이 되는 거래를 거부하거나 이익이 되는 거래가 더 해로울 수 있다는 점을 인정해야 한다. 1의 가치를 100%의 확률로 얻을 수 있는 상황을 생각해 보자. 이 상황에서 이 선택을 포기하면 1,000의 가치를 99%의 확률로 얻을 수 있다는 제안은 기댓값이 훨씬 높은 거절하기 어려운 제안이다. 그런데 이런 제안이 연속되면 결국 광신적 행동을 선택해야 하는 상황으로 이어진다.⁶⁶⁾

62) 이 이야기는 Nick Bostrom의 "Pascal's mugging"을 재구성했다.

63) Nick Bostrom, "Astronomical Waste: The Opportunity Cost of Delayed Technological Development," p. 311

64) Nick Bostrom, "The Vulnerable World Hypothesis," p. 465

65) Nicholas Beckstead, "On the Overwhelming Importance of Shaping the Far Future" (PhD Thesis), p. 11

66) Hayden Wilkinson, "In Defense of Fanaticism," p. 451-452

2.2.5. 모호한 미래

결과의 기댓값을 산출하고 비교하기 전에 지금 선택이 미래에 주는 영향을 파악하는 일부부터 쉬운 일이 아니다. 미래가 어떻게 변할지 알기 어렵고 현재의 선택이 어떤 결과로 나타날지 파악하기도 쉽지 않다. 먼 미래의 인간과 사회는 지금은 상상할 수 없을 정도로 달라져 있을 수 있다. 지금 높은 가치를 지닌 것이 먼 미래에 아무 가치도 없게 되어도 이상하지 않다. 그렇다면 어떤 근거에 기반해서 먼 미래에 실현될 가치를 높이는 행동을 선택할 수 있을까?⁶⁷⁾

먼 미래에 대한 단서가 없기에 먼 미래를 위한 행동을 할 수 없다는 주장에 대한 여러 반론이 있다. 그중 하나는 먼 미래가 어떤 상태일지에 대한 모호성은 미래를 위하는 행동을 한다고 해서 증가한다고 보기 어렵다는 것이다. 예를 들어 인류가 가까운 시일 안에 절멸할지 여부는 극단적으로 모호한데 절멸의 원인이 될 수 있는 감염병 예방에 대해 노력하는 것이 이 모호성을 변동시킨다고 보기 힘들다.⁶⁸⁾

2.2.6. 지난한 계산

모호한 미래를 위한 행동이 가능해도 한 행동이 초래하는 영향을 판단하는 일은 쉽지 않다. 사소한 행동 하나라도 남은 인류의 발전 궤적에 지속적인 영향을 미칠 수 있다. 어떤 행동이 연쇄되어 어떤 결과를 만들지 알 수 없다면 행동을 결심하는 것은 쉽지 않다. 이득을 얻는 것보다 손실을 피하는 것을 중요하게 여길수록 행동하는 것은 더 어려워진다.⁶⁹⁾

한 사람이 마트에 가는 상황을 생각해 보자. 이 사람이 사는 곳 근처에는 마트가 두 개 있다. 하나는 차를 타고 가야 하는 곳에 있는데 파란색 간판을 달고 있다. 다른 하나는 걸어서 갈 수 있는데 빨간색 간판을 달고 있다. 처음에는 차를 가지고 파란 마트에 가기로 마음먹었다. 자동차 운전이 지구 환경에 악영향을 미쳐서 미래 세대가 우주 개척 기술을 습득하기 이전에 절멸하는 것은 아닌지 잠시 걱정했다. 그런데 더 시급한 문제는 오늘 그 마트 근처에 외교 행사가 있다는 점이었다. 혹여 자신이 운전 미숙으로 중요한 외교관을 다치게 한다면 국제 평화에 영향을 미칠 수도 있다는 생각에 이르렀다. 운전은 포기하고 빨간 마트로 가기로 했다. 그런데 또 생각해 보니 만약 자신이 빨간 마트의 매출을 올려주어서 전국 여기저기에 빨간 마트가 더 많이 생기게 되고 어느 날 화가 잔뜩 난 시위대가 새로 생긴 빨간 간판을 보고 더 흥분해서 폭력적인 행동을 할지도 모를 일이다. 생각이 여기까지 오니 그는 자신의 행동이 어떤 영향을 미치는지 파악하기를 포기하고 아무것도 하지 않기로 했다.⁷⁰⁾

이 소심한 사람이 아무것도 하지 않는 것이 정말로 나은 결과를 낳을까? 그리고 이렇게 행동의 영향력이 엄청나다는 점에 대해 전전긍긍하면서 결국 아무 선택도 하지 않는 것이 최선의 선택일까? 아무것도 하지 않는 것이 늘 아무 영향도 미치지 않는 것은 아니다. 이 사람이 한 발짝도 집 밖으로 나서지 않던 날을 지속하고 있다가 갑자기 지진이 나서 정신을 잃고 깨어보니 폐허가 된 건물 잔해 속에 심각한 부상을 당한 누군가를 깔고 있었다. 그가 어서 움직이지 않으면 깔린 사람이 곧 죽을 상황이다. 이때에도 그는 어떤 행동을 하는 것이 불려울지 모르는 미래의 영향 때문에 아무것도 하지 말아야 할까? 이 경우에 그가 움직이지 않으면 깔

67) Hilary Greaves & William MacAskill, "The Case for Strong Longtermism," p. 19

68) Hilary Greaves & William MacAskill, "The Case for Strong Longtermism," p. 24

69) Andreas Mogensen & William MacAskill, "The Paralysis Argument," p. 1

70) Dylan Balfour의 "Pascal's Mugger Strikes Again"을 재구성했다.

려 있는 사람이 죽는다. 이는 미래에 영향을 주는 것이다. 아무것도 하지 않는 것이 아무 미래에 대한 영향도 남지 않는 것은 아니다. 손실의 회피 측면에서 무엇인가 한 것과 하지 않은 것은 그렇게 큰 차이가 없다.⁷¹⁾

2.2.7. 여타의 옹호

결과론적 입장을 취하지 않고 장기주의를 보강하려는 시도도 존재한다. 그중 하나는 장기주의가 약자를 위한 논리라는 주장이다. 오드는 장기주의가 시간과 공간 같은 조건에 상관없이 모든 사람을 동등하게 여긴다고 설명한다.⁷²⁾ 현세대에 대한 발언권이 없는 미래 세대라도 장기주의에 따르면 그 중요성이 낮아지지 않는다. 또 다른 시도는 세대간 호혜를 주장하는 것이다.⁷³⁾ 시간은 한 방향으로만 흐르기에 과거의 사람을 돕는 것보다는 미래 사람을 돕는 것이 쉬움으로 우리의 의무도 미래를 향해 있다는 설명이다. 과거 세대로부터 받은 호혜를 다시 미래 세대로 전하지 않으면 이는 과거 세대에 대한 배신행위이다. 또 미래를 고려하지 않는 것은 과거에 발생한 피해에 대해 조금이라도 보상할 기회를 포기하는 것이다.

71) Andreas Mogensen & William MacAskill, "The Paralysis Argument," p. 12

72) 토비 오드, 하인해 율김, 『사피엔스의 멸망』, p. 64.

73) 토비 오드, 『사피엔스의 멸망』, p. 73-74

참고문헌

단행본

- 리처드 파인만, 승영조·김희봉 옮김, 『발견하는 즐거움』, 도서출판 승산, 2013.
- 윌리엄 맥어스킬, 이영래 옮김, 『우리는 미래를 가져다 쓰고 있다』, 김영사, 2023.
- _____, 전미영 옮김, 『냉정한 이타주의자』, 부키, 2017.
- Derek Parfit, *Reasons and Persons*, Oxford University Press, 1984.
- Émile P. Torres, *Human extinction : a history of the science and ethics of annihilation*, Routledge, 2024.
- Joerg Chet Tremmel, *A Theory of Intergenerational Justice*, Routledge, 2009.
- Kenneth J. Arrow, "Agency Cost versus Fiduciary Duties," *Principals and Agents*: Toby Ord, *The Precipice*, Hachette Books, 2020.
- Tyler Cowen & Derek Parfit, "Against the social discount rate," *Justice between age groups and generations*, Yale University Press, 1992.
- William MacAskill, *What We Owe the Future*, Basic Books, 2022.
- _____, "Are We Living at the Hinge of History?," *Ethics and Existence: The Legacy of Derek Parfit*, Oxford University Press, 2022.
- _____, "Effective Altruism," *International Encyclopedia of Ethics*, 2020.
- _____, "Understanding Effective Altruism and Its Challenges," *The Palgrave Handbook of Philosophy and Public Policy*, Palgrave Macmillan, Cham, 2018.

논문

- Andreas L. Mogenden, "Moral Demands and The Far Future," *Philosophy and Phenomenological Research*, Vol. 103, Iss. 3, 2021.
- Andreas Mogensen & William MacAskill, "The Paralysis Argument," *Philosopher' Imprint*, Vol. 21 No. 15, 2021.
- Christian J. Tarsney, "The Epidemic Challenge to Longtermism," *Sythese*, Vol. 201 Iss. 6, 2023.
- Dylan Balfour, "Pascal's Mugger Strikes Again," *Utilitas*, Vol. 33, 2021.
- Derek Parfit, "Can We Avoid the Repugnant Conclusion?," *Theoria*, Vol. 83, 2016.
- Fei Teng, "Climate Change and Moral Responsibility toward Future Generations: A Confucian Perspective," *Philosophy East and West*, Vol. 71, No. 2, 2021.
- Hayden Wilkinson, "In Defense of Fanaticism," *Ethics*, Vol. 132 Iss. 2, 2022.
- Herbert Fingarette, "The Music of Humanity in The Conversations of Confucius," *Journal of Chinese Philosophy*, Vol. 10, 1983.
- Matthew Dennis & Elena Ziliotti, "Living Well Together Online: Digital Wellbeing from a Confucian Perspective," *Journal of Applied Philosophy*, Vol. 40, Iss. 2, 2022.
- Michael C. Jensen & William H. Meckling, "Theory of the Firm: Managerial

Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics*, Vol. 3, No. 4, 1976. (<http://papers.ssrn.com/abstract=94043>)

Nicholas Beckstead, "On the Overwhelming Importance of Shaping the Far Future" (PhD Thesis), Rutgers University, 2013.

Nick Bostrom, "The Vulnerable World Hypothesis," *Global Policy*, Vol. 10 Iss. 4, 2019.

_____, "Existential Risk Prevention as Global Priority," *Global Policy*, Vol. 4 Iss. 1, 2013.

_____, "Pascal's mugging," *Analysis*, Vol. 69 Iss. 3, 2009.

_____, "Letter from Utopia," *Studies in Ethics, Law, and Technology*, Vol. 2 Iss. 1, 2008.

_____, "Astronomical Waste: The Opportunity Cost of Delayed Technological Development," *Utilitas*, Vol. 15 Iss. 3, 2003.

_____, "Existential Risks: Analyzing Human Extinction Scenarios And Related Hazards," *Journal of Evolution and Technology*, Vol. 9, 2002. ([url: https://ora.ox.ac.uk/objects/uuid:827452c3-fcba-41b8-86b0-407293e6617c](https://ora.ox.ac.uk/objects/uuid:827452c3-fcba-41b8-86b0-407293e6617c))

William Sin, "The Demandingness of Confucianism in the Case of Long-Term Caregiving," *Asian Philosophy*, Vol. 23, No. 2, 2013.

기타

Hilary Greaves & William MacAskill, "The Case for Strong Longtermism," Global Priorities Institute, 2021.

Matt Pickles, "Elon Musk funds Oxford research into machine intelligence," <https://www.ox.ac.uk/news/arts-blog/elon-musk-funds-oxford-research-machine-intelligence>, 2015.