

# 장기주의에 대한 짧은 설명

김유민(daseoh@gmail.com)

## 1. 장기주의라는 용어

장기주의는 벡스테드와 보스트롬의 생각을 기반으로 오드와 맥어스킬이 제안한 이름이다.<sup>1</sup> 우선 벡스테드와 보스트롬은 인류를 절멸시킬 위험을 줄이는 것이 최우선 가치라고 생각한다. 보스트롬은 공리주의 입장에서 우리에게 가장 중요한 것은 절멸 확률을 줄이는 것이라고 말하고<sup>2</sup> 벡스테드는 인류 존속을 위해서라면 의무론적으로 정당화할 수 없는 선택을 할 수밖에 없다고 주장한다.<sup>3</sup> 벡스테드와 보스트롬의 작업이 논리적 세계에서 이루어진 것이라고 보면, 오드와 맥어스킬은 장기주의를 현실에서 선택될 수 있게 각색한다. 오드는 장기주의가 “우리의 행동이 장기적 미래에 미치는 영향에 초점을 맞춘다”라고 설명하고<sup>4</sup> 맥어스킬은 장기주의를 “미래 세대의 이익을 보호하기 위해 훨씬 더 많은 일을 해야 한다고 생각하는 관점”으로 설명한다.<sup>5</sup>

## 2. 특징 하나 - 기획력

다른 말을 하기에 앞서서 이들의 기획력은 눈여겨볼 만하다. 이 네 사람은 보스트롬이 설립한 Future of Humanity Institute에 자리를 두고 있었다. 그리고 다른 여러 기관을 설립했다. 특히 오드와 맥어스킬은 현실에서 장기주의를 적용하기 위한 여러 기획을 했다. 이들은 효율적 이타주의 운동을 진행하면서 수입의 10%를 기부하는 Giving What We Can와 직업을 통해 사회적 문제에 영향을 미치는 방법을 조연하는 80,000Hours라는 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 설명이다.<sup>6</sup> 이 운동에서 “장기주의”라는 말을 찾기는 어렵다. 그런데 현재 인류가 심각한 재앙을 마주하고 있

---

1 오드, 2021, 401.

2 Bostrom, 2003, 311.

3 Beckstead, 2013, 11.

4 2021, 65.

5 2023, 8.

6 맥어스킬, 2017.

다고 설명과 오직 행복의 증진만 목적에 두고 있다는 설명<sup>7</sup> 그리고 행복을 최대화하고 과학을 가장 효율적인 자선 단체를 찾기 위한 도구로 사용해야 한다는 설명을 보면<sup>8</sup> 장기주의의 함의를 숨기기 어려워 보인다.<sup>9</sup>

### 3. 특징 하나 - 인류에 대한 비유

미래에 대한 충분한 고려가 필요함을 설득하기 위한 장기주의의 전략은 인류가 등장하고 사라지는 전 과정을 인간이 탄생하고 죽는 한 주기에 비유하는 것이다. 생의 주기에 따른 가능성을 생각해 보면, 아직 두 발로 걷지 못한 아이가, 삶이 거의 끝나가는 노인보다, 훨씬 큰 가능성을 지닌다고 할 수 있다. 인류의 존속 기간을 일반적인 포유류의 존속 기간에 놓고 보면, 인류는 개인의 생의 주기로 치면 아직 세상 구경도 하지 못한 어린 아이 단계에 있다. 인류가 지금까지 지구에 존재했던 다른 포유류와 다르게 더 큰 잠재력이 있다면 인류의 존속 기간은 훨씬 늘어날 수 있다. 이는 반대로 생각하면 지금 인류의 위치가 이제 막 세계에 나온 상태에 있다는 것이다.

(이 설명에서 한 가지 독특한 점은, 오드와 맥어스킬이 인류의 미래가 충분히 남아있다는 것을 시각적으로 제시한다는 점이다. 오드는 다섯 개의 일차원 그래프를 보여주는데, 각 그래프의 한 칸의 축적을 10년에서, 1천 년, 1천만 년, 10억 년으로 바꾸면서 인류가 정말 아주 짧은 과거를 가지고 있고 앞으로 남은 시간은 정말 길다는 것을 보여준다.<sup>10</sup> 맥어스킬은 사람 아이콘 하나에 10억 명을 할당해서 미래 인류가 얼마나 남았는지 보여주는데, 3쪽에 걸쳐 사람 아이콘을 가득 채우고 나서, 앞으로도 2만 페이지가 더 필요하다고 설명한다.<sup>11</sup>)

#### 3.1. 인류 절멸의 위험 - 행위력 증대

갓난아이는 많은 위험에 노출되어 있다. 인류도 마찬가지다. 지금까지 인류는 운 좋게 생존한 것이다. (보스트롬은 인류 절멸에 대해 많은 연구를 발표했는데) 인류 절멸의 위험은 다양한 경로로 발생한다. 운석 충돌 같이 외계에서 발생하는 위험도 있지만, 사실 지금 마주한 위험은 인류가 잠재력을 실현하는 과정에서 따라오는 위험이 더 많다. 인류가 기술 개발 통해 행위 능력을 키워 나갈수록 인류 절멸의 위험은 더 커진다. 핵폭탄같이 인류 전체의 운명을 결정할 수 있는 도구는 이미 등장했다. 현재 인류의 상태는 몸이 크며 힘이 생기고 있지만, 이 힘을 잘 제어할 수 없는 미숙한 상태인 것이다.<sup>12</sup> 죽음보다 더 나쁜 일은 없기에 장기주의자들은 파국에서 최대한 멀어질 수 있는 선택과 안정적인 성장 궤도에 오를 수 있는 선택을 해나가야 한다고 주장한다.

---

7 MacAskill, 2020, 1.

8 MacAskill, 2018, 442.

9 Pearlman(<https://blog.apaonline.org/2021/03/29/is-effective-altruism-inherently-utilitarian/>)은 효율적 이타주의가 공리주의의 재포장이라고 지적함.

10 오드, 2021, 300-301.

11 맥어스킬, 2023, 35-39.

12 오드, 2021, 76.

### 3.2. 인류 절멸의 위험 - 가치 고착

인생의 한 주기에서 어린 시절 보고 배운 것은 남은 삶에 지속적으로 영향을 미친다. 우연히 마주한 좋은 경험으로 삶의 궤도가 변경되기도 하고, 나쁜 습관이 평생 고쳐지지 않기도 한다. 맥어스킬은 벡스테드의 개념을 차용해 이를 인류에도 적용하는데, 인류 초반인 지금은 아직 사고가 굳지 않은 어린 상태이고, 우리가 지금 한 선택이 남은 인류의 가치관에 영향을 미친다는 것이다.<sup>13</sup> 그리고 이 선택에는 우발성이 크게 작용한다. 예를 들어 노예제 폐지는 우리 기술 발전 수준에서 보면 필연적인 것이 아니라 우연한 사건들이 모여서 가능했다는 것이다.<sup>14</sup>

이런 가치 고착은 다양성을 떨어뜨리기에 위험에 대한 대응력을 약화시키고, 전세계적으로 가치 고착이 일어나면 인류의 미래는 고착된 가치 밖으로 넘어설 수 없게 된다.<sup>15</sup> 인공지능의 발현으로 다음번 일어날 가치 고착은 무한히 지속될 되지도 모르고, 이는 인류를 절멸에 이르게 할 수도 있다.<sup>16</sup> 그러니 가치가 고착되기 전에 가능한 많은 도덕적 진보를 이를 필요가 있고, 우리가 가치를 선택할 수 있는 성찰을 가질 수 있는 안정된 상태에 도달해야 한다.<sup>17</sup>

### 4. 특징 하나 - 기술 낙관적 태도

장기 주의에서 인류가 지닌 잠재력에 대한 희망은 다양한 사례로 뒷받침된다. 지금까지 인류가 이루어 온 발전에 그 근거를 찾기도 하지만, 가장 주된 것은 기술에 대한 낙관적 태도다. 보스트롬은 기술을 통한 신체 증강으로 인류가 새로운 상태에 도달할 수 있다고 믿는다.<sup>18</sup> 좀 더 신중하기는 하지만 오드는 인간이 증강을 통해 완전히 새로운 종으로 변화할지도 모른다고 말한다.<sup>19</sup> 이들에게 인류가 기술 개발을 통해 우주로 나아가는 일은 인류가 절멸하지 않는다면 사실상 예정된 것이다.<sup>20</sup>

(보스트롬은 트랜스휴머니스트이고 오드는 포스휴머니스트로 볼 수 있다. 트랜스휴머니즘은 인간을 개선하려 하는데, 인간의 진화가 끝이 없을 것이라고 이해한다. 이는 새로운 사람이 되길 요구하는 기독교의 맥락과 유사하다. 기술적 포스휴머니즘은 트랜스휴머니즘적 사고로 구분하기도 하는데 인간 자체를 극복하는

---

13 맥어스킬, 2023, 62, 70.

14 맥어스킬, 2023, 103, 110.

15 맥어스킬, 2023, 121.

16 맥어스킬, 2023, 121-122.

17 맥어스킬, 2023, 150.

18 Bostrom,

19 오드, 2021, 258, 322.

20 오드, 2021, 34.

인공적 대상을 창조하는 것을 중시한다. 트랜스휴머니즘과 기술적 포스트휴머니즘이 기술을 대상이 아닌 목표로 삼는 반면에 비판적 포스트휴머니즘은 인본주의 사회의 이분법을 문제 삼는다.)

문제는 기술을 개발하며 발생하는 절멸의 위험이다. 기술이 개발되면서 인간의 행위 능력 또한 증대된다. 예전에는 국가적 역량을 모아야 보일 수 있었던 산출이 지금은 한 개인의 손에서 이루어진다.<sup>21</sup> 앞으로 이런 추세는 가속될 것이지만 장기주의자들은 기술 발전을 막는 것은 그 자체로 재앙이라고 말한다.<sup>22</sup> 기술 발전을 통해 절멸 위험에 대처할 수 있기 때문이다. 그리고 인류의 시간에 비하면 기술 발전 속도를 늦추는 것은 크게 의미 있는 일이 아니다.<sup>23</sup>

인류가 기술 발전을 잠시 보류해야 할 유일한 시기는 기술이 충분히 발전해서 어떤 위험으로부터 안정적인 상태가 되었을 때다. 인류는 이 시기까지 절멸의 위험과 싸워야 한다.<sup>24</sup> 맥어스킬은 이 시기를 숙고(Long Reflection)라고 이름 붙이고 인류가 도달해야 할 이상적 지점으로 여긴다.<sup>25</sup> 이 시기에 도달하면 인류는 열린 미래의 가능성을 탐색하면서 잠재력을 최대한 실현할 수 있는 경로를 선택해야 할 것이다.<sup>26</sup>

## 5. 공리주의적 뒷받침

미래 세대가 지금보다 월등히 진보한 기술로 지금보다 나은 세상을 만들 수 있다고 해서, 지금 내가 미래 세대를 고려한 가치 판단을 해야 한다고 주장할 수는 없다. 이 부족을 채우는 것은 공리주의다. 장기 주의에서 공리주의라는 단어는 거의 등장하지 않지만, 장기주의를 공리주의가 아니라고 말하기는 더욱 어려워 보인다. 우리가 미래를 위한 행동을 해야 하는 가장 큰 이유는, 미래에는 지금까지 있었던 사람의 수와 비교도 할 수 없을 정도로 엄청나게 많은 사람이 존재할 것이고 또 이 많은 수의 사람이 지금보다 훨씬 행복하게 살 가능성이 있기 때문이다. 좋음의 총량을 높이는 선택을 해야 하는 공리주의 입장에서 지금 나의 작은 행복을 위해 가능성 있는 거대한 미래의 행복을 포기하는 것은 나쁜 것이다.

### 5.1. 공리주의적 입장의 결점 - 파스칼의 강도

이렇게 미래의 가능성을 계산하는 것은 파스칼의 강도라고 불리는 문제를 낳는다.<sup>27</sup> 인류가 절멸하지 않고 미래에 엄청난 행복이 있다고 해도, 이는 아직 발생하지 않은 사건이다. 일상에서 선택의 가치를 고려할 때 작은 확률의 사건을 거의 고려되지 않는다. 길을 걷다가 누군가 달려들어서 자신에게 십만 원을 주지 않으면

---

21 오드, 2021, 178.

22 오드, 2021, 210.

23 오드, 2021, 277.

24 Bostrom, 2013, 24.

25 맥어스킬, 2023, 150.

26 오드, 2021, 256.

27 Bostrom, 2009.

지구가 멸망한다고 말한다면 대부분은 무시하고 넘어갈 것이다. 그런데 장기 주의에서 미래 가치를 계산하는 방법을 따르면 우리는 이 사람의 말을 무시할 수 없다. 지구 멸망을 막는 것은 너무나도 큰 가치이고, 그 사람이 지구를 멸망시킬 가능성이 극히 희박하다고 해도 완전히 없는 것은 아니기 때문에, 이 작은 확률에 엄청난 게 큰 가치가 곱해지면서 유의미한 선택지가 된다.

## 5.2. 공리주의적 입장의 결점 - 당혹스러운 결론

미래 세대를 고려한 공리주의적 입장의 또 다른 문제는 “당혹스러운 결론”이다.

우선, 파핏이 제시한 비 동일성 문제(the non-identity problem)를 살펴야 한다.<sup>28</sup> 현재의 선택은 미래를 변화시킨다. 아주 사소한 사건이라도 미래에 존재할 사람에게 영향을 미칠 수 있다. 만약 어떤 사람이 편두통을 아이에게 물려 줄 수 있는 치료 가능한 질병을 갖고 있다고 생각해 보자. 보통의 선택은 이 병을 치료한 뒤에 편두통이 없는 아이를 낳는 것이다. 그런데 치료한 뒤에 낳은 아이와 그렇지 않은 아이는 동일한 아이가 아니다. 약간의 고통을 갖고 살아갈지라도 치료하지 않은 아이가 세상에 존재하는 것이 더 나은 것 아닌가?

미래 세대에 대해 우리가 갖는 직관은 보통 중립적이다. 앞으로 태어나 아이의 삶이 좋을지 나쁠지 모른다는 것이다. 그런데 이 직관에 따르면 당장 미래 세대를 신경 쓰지 않아도 되는 문제에 직면한다. 중립에 직관을 따르면 편두통을 앓는 아이를 낳는 것과 그렇지 않은 아이를 낳는 것은 아무런 차이도 없다는 모순에 이른다.

중립의 직관을 따르지 않고 비 동일성 문제를 해결하는 방법은 세상의 행복이 클수록 좋은 것이라는 공리주의적 입장을 따르는 것이다. 그렇지만 이 입장에서는 순간 “당혹스러운 결론”을 만나게 된다. 1의 행복을 갖는 1,000명의 사람이 사는 세상과 100의 행복을 갖는 10명이 사는 세상이 같은 의미를 갖기 때문이다. 이 문제를 해결하기 위해 여러 조건을 추가하기도 하지만 당혹스러운 결론을 완전히 피하기는 어렵다. 맥어스킬의 입장은 당혹스러운 결론을 그냥 받아들이는 것이다.<sup>29</sup>

## 6.1. 공리주의적 비판의 반박 - 약자의 논리다

미래에 발생할 기댓값을 계산하는 생각하는 이상 장기주의가 공리주의적이라는 비판을 완전히 피하기는 어려워 보인다. 그러니 전체의 더 큰 행복을 위해 개인에게 불행한 선택이 강요된다는 비판도 장기주의에 그대로 적용할 수 있다. 앞에서 말했듯이 보스트롬과 벡스테드는 인류의 생존 이외의 것은 크게 중요한 것이 아니라고 말한다. 이들에게 기후 위기나 빈곤은 큰 문제가 아니다.

오드는 이 비판에서 장기주의를 옹호할 수 있는 논리를 제공한다. (나는 무슨 말인지 잘 이해가 되지는 않지만) 장기주의가 약자를 위한 윤리라는 것이다. 장기주의는 시간과 공간 같은 조건에 상관없이 모든 사람을

---

28 Parfit, 1984, 351-380.

29 맥어스킬, 2023, 263.

동등하게 여기고, 행복의 가치나 고통의 공포 또한 조건에 한정되지 않는다.<sup>30</sup> 현재 행동에 영향을 받을 미래 세대는 현재에 발언권이 없지만, 장기주의에 따르면 미래 세대 또한 동등한 위치에 놓이게 된다.

## 6.2. 공리주의 아닌 뒷받침 - 죽음은 나쁘다

장기주의를 공리주의가 아닌 방식으로 설득하는 방법도 있다. 우선 인류를 개인에 비유하는 것에 이어서 죽음은 나쁘다는 것이다. 에피쿠로스는 미래에 나쁨을 경험할 자신도 존재하지 않기 때문에 죽음이 나쁠 이유가 없다고 했지만, 이는 심리적 위안이 될지 몰라도 행동의 지침이 될 수는 없을 것이다. 삶이 짧아지기 때문에 일어나지 않을 나쁨도 있겠지만 좋은 또한 덜 생기게 된다.<sup>31</sup> (그런데 결국 가치의 총량을 비교한다는 점에서 공리주의와 맥을 달리하기는 어려워 보인다.)

## 6.3. 공리주의 아닌 뒷받침 - 세대 간 호혜

오드는 장기주의를 뒷받침하는 또 다른 방식으로 세대 간 호혜를 든다.<sup>32</sup> 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.

---

30 오드 2021, 64.

31 오드 2021, 68.

32 2021, 73-74.

---

## 참고문헌

맥어스킬, 윌리엄. (2017). <<냉정한 이타주의자>>.

\_\_\_\_\_. (2023). <<우리는 미래를 가져다 쓰고 있다>>.

오드, 토비. (2021). <<사피엔스의 멸망>>.

Beckstead, Nicholas. (2013). “On the Overwhelming Importance of Shaping the Far Future” [PhD Thesis]. Department of Philosophy, Rutgers University.

Bostrom, Nick. (2003). “Astronomical Waste: The Opportunity Cost of Delayed Technological Development.” *Utilitas* 15 (3): 308–314.

\_\_\_\_\_. (2009). “Pascal’s *mugging*.” *Analysis* 69 (3): 443–445.

\_\_\_\_\_. (2013). “Existential Risk Prevention as Global Priority,” *Global Policy* 4 (1): 15-31.

MacAskill, William. (2018). “Understanding Effective Altruism and Its Challenges,” *The Palgrave Handbook of Philosophy and Public Policy*.

\_\_\_\_\_. (2020). “Effective Altruism.” *International Encyclopedia of Ethics*.

Parfit, Derek. (1984) *Reasons and Persons*.