논리연구 17-3(2014) pp. 425-439

수리기사의 역설, 무엇이 역설적인가?*

원 치 욱

【국문요약】 최근 김한승 교수는 그의 논문 “확률과 시간에 관한 두 가지 퍼즐”(2014)에서 수리기사의 역설을 소개하고 이에 대한 흥미로운 논의를 내놓았다. 김한승(2014)에 따르면, 수리기사 역설에서 ‘오전’이 참일 확률과 ‘오후’가 참일 확률은 동일하지만 후회라는 감정의 본성상 ‘오후’를 선택하는 것이 합리적이다. 본 논문에서 필자는 김한승 교수가 수리기사 역설을 이해, 진단하는 방식에 대해 문제 제기를 하고(1, 2절), 그 역설의 한 가지 가능한 해결 방향에 대해 간략히 논의한다(3절). 필자가 보기에 후회의 감정은 수리기사 역설을 일으키는 중요한 원천이 아니며, 오히려 문제의 상황이 정말로 역설적인 이유는 후회의 부정적인 효용을 고려하지 않는다 하더라도 여전히 퍼즐이 발생하기 때문이다.

【주요어】 수리기사 역설, 후회, 확률, 김한승

접수일자: 2014.09.17 심사 및 수정완료일: 2014.10.06 게재확정일: 2014.10.12
* 이 논문의 부족한 점을 지적해주신 익명의 심사위원 선생님들, 그리고 수리기사 역설을 알려주시고 유익한 토론을 해주신 김한승 선생님께 감사의 말씀을 드린다.

1. 수리기사의 역설

케이블 수리기사가 오늘 당신의 집을 방문할 예정이며 서비스를 받으려면 당신은 집에 있어야 한다. 기사가 오전 9시에서 오후 3시 사이에 온다고 하였기에 당신은 출근을 포기하고 집을 지키기로 한다. 심심함을 덜기 위해 친구를 부른 당신은 기사의 도착 시간을 두고 친구와 내기를 한다. 내기는 간단하다. 기사가 오전 9시에서 정오 사이에 올지 혹은 정오에서 오후 3시 사이에 올지를 놓고 내기를 걸어 이기는 사람이 만원을 받는 것이다. '오전'을 선택하는 것이 나은가, '오후'를 선택하는 것이 나은가? 기사가 9시에서 3시 사이에 온다는 것이 당신이 가진 정보의 전부인 상황에서 당신은 두 선택지 중 어느 하나를 특별히 선호할 이유가 없어 보인다. 그러나 당신은 곧 다음을 깨닫는다. 당신이 만일 '오전'을 선택했다고 하자. 그렇다면 당신은 9시 이후부터 시간이 지날수록 당신의 선택을 계속 후회할지 모른다. 9시부터 정오까지 시간이 지나면 지날수록 당신이 이길 확률은 점점 더 줄어들 것이기 때문이다. (그리고 이것을 당신은 지금 선택하는 시점에 이미 안다.) 그렇다면 당신은 '오후'에 내기를 거는 것이 더 나은 선택 아닌가?

이것이 헤이젝(Hájek 2005)이 내놓은 이른바 "수리기사의 역설"(The Cable Guy paradox)이다. 이 역설에 대한 헤이젝 자신의 진단은 우리는 오전, 오후 선택에 무관심해야 한다는 것이다. 좀더 구체적으로, 헤이젝은 오후를 선호하게끔 하는 "후회할 것이 확실한 선택은 피하라"는 취지의 원리(Avoid Certain Frustration Principle, 이하 ACF)를 거부하는 방향으로 나아간다. 큰 틀에서 필자는 ACF가 근거해있는 기반을 잘 따져보고 의심해봐야 한다고 생각한다는 점에서 헤이젝에 동의한다. 그러나 이 논문은 어떻게 역설을 해결해야 하는지에 대한 논의는 아니다. 그보다는 역설의

본성을 어떻게 이해해야 할지 – 좀더 정확히, 그것을 어떻게 이해하지 말아야 할지 – 의 논의에 초점을 둔다.

최근의 논문에서 김한승(2014)은 문제의 상황에서 "후회라는 감정의 본성상" 오전을 선택하는 것이 합리적이라고 주장한다. 어떤 의미에서 필자는 김한승 교수의 이러한 결론에 전적으로 동의한다. 그러나 필자는 이러한 결론이 수리기사 역설에 대한 진정한 해법이라고 생각하지 않는다. 그러한 결론은 수리기사 역설이 왜 역설인지 그 진가를 온전히 음미하지 못하는 방식의 이해에서 비롯된 것이라고 생각한다. 그 이유는 후회라는 감정이 수리기사 역설을 일으키는 중요한 원천이 아니기 때문이다. 아래에서 필자는 수리기사 역설의 발생은 오히려 후회의 부정적인 효용을 무시하는 데에서 비로소 출발한다고 논변한다.

2. 무엇이 역설적인가?

좀더 단순한 상황에서 출발해보자. 누군가 철수에게 동전 던지기 내기를 제안하자, 철수는 앞면에 만원을 건다. 불행히도 동전은 뒷면이 나오고 철수는 돈을 잃는다. 철수는 뒷면에 걸었어야 했다고 후회를 한다. 이러한 상황에서 철수가 비합리적이었다고 생각할 이유는 없다. 철수는 그저 운이 없었을 뿐이다. 그러나 만약 철수가 동전이 뒷면으로 나올 것을 미리 알았다고, 혹은 확신했었다고 해보자. 그렇다면 철수는 무언가 잘못된 것처럼 보인다. 철수는 분명 돈 잃기를 원할 사람이 아니라고 가정했을 때, 우리는 철수의 합리성을 의심해야 할 것 같다. 철수의 행위는 표준적인 의사결정 이론(Decision Theory)의 틀 내에서 이해하기 힘들어 보인다.

결정 이론에서의 합리성 혹은 기대 효용(Expected Utility)을 최대화하라는 원리는 우리에게 후회가 확실시되는 선택은 하지 말라

고 요구하는 것 같다. 좀더 정확히, 기대 효용 원리가 “확실한 후회는 피하라”는 취지의 ACF 원리를 함축하는지는 분명치 않더라도, 적어도 그 두 원리 사이에는 일견 아무런 모순이 없어 보인다. 그러나 수리기사 역설이 보여주는 바는 기대 효용 최대화 원리와 ACF가 서로 충돌하는 것처럼 보인다는 것이다. 다시 말해, ACF가 어떻게 결정 이론의 틀 내에서 수용될 수 있을지 의문시된다는 것이다. 수리기사 역설의 상황에서 기대 효용 원리는 당신으로 하여금 오전, 오후의 선택에 무관심할 것을 요구하는 반면, ACF는 오후를 선택할 것을 요구한다. 이 충돌을 어떻게 볼 것인가?

이러한 역설적 상황을 이해함에 있어 김한승 교수는 후회라는 감정의 역할을 핵심적인 것으로 본다. 김한승(2014)은 헤이젝을 비판하면서 다음과 같이 말한다.

> 헤이젝은 선택으로부터 생겨나는 후회가 주는 고통과 같은 부정적인 효용은 고려하지 않는다. 아마도 그는 케이블기사의 역설에서 선택을 할 때 고려해야 할 것은 각 선택지에 대한 확률뿐이라고 여기는 듯하다. 하지만 필자는 그렇지 않다고 생각한다. 자신이 특정한 선택을 하면 그로부터 특정한 감정이 생겨날 것이 분명하고 그 감정을 원치 않는다면, 그 선택을 피하는 것이 합리적이다. (p. 11)

문제의 상황에서 “후회가 주는 고통과 같은 부정적인 효용”을 고려했을 때, 오전을 선택하는 것이 더 합리적이라는 주장에 필자는 아무런 이의가 없다. 그러나 필자가 보기에, 후회가 주는 고통 혹은 그것의 부정적인 효용은 수리기사 역설을 발생시키는 본질적인 요소가 아니다. 다음의 세 가지를 고려해보자.

(1) 우리는 누구나 후회가 주는 고통 혹은 부정적인 느낌을 안다. 우리가 가급적 후회없는 삶을 살고자 하는 이유 중 하나도 그

러한 느낌을 피하고 싶기 때문일 것이다. 그런데 우리의 철수는 신기하게도, 후회라는 심적 상태에 보통 동반하는 그러한 부정적인 느낌을 전혀 갖고 있지 않다. 이것이 너무 비현실적인 가정이라고 생각된다면, 철수는 단지 후회라는 심적 상태에 동반하는 느낌에 대해 전적으로 무관심하다고 가정해보자. 철수가 관심있는 것은 오로지 돈, 어떻게 내기에서 돈을 딸 것인지이다. 즉, 철수에게는 금전적인 고려만이 기대 효용의 유일한 원천이다. 중요한 것은 이러한 철수에게도 수리기사 역설이 발생한다는 것이다. 철수가 지금 오전을 선택한다면, 그는 9시가 지나자마자 오전 선택보다 오후 선택이 더 좋은 선택이었다고 판단할 것이다 – 그리고 그는 이것을 선택을 하는 시점에 이미 알고 있다.[1)]

이것이 보여주는 바는 수리기사 역설이 발생하기 위해서 행위자는 표준적인 결정 이론이 요구하는 최소한의 심리학 – 즉, 믿음(주관적 확률)과 욕구(선호도) – 만 있으면 충분하다는 것이다. 가령 그 행위자는 공포나 질투와 같은 감정을 꼭 가질 필요는 없으며, 마찬가지로 후회의 감정도 꼭 가질 필요는 없다.

(2) 후회가 수리기사 역설의 원천이 아니라는 필자의 주장은 애매하고 오해의 소지가 있다. 왜냐하면 분명히 후회는 어떤 의미에서 수리기사 역설의 원천이기 때문이다. 여기서의 애매성을 보기 위해 먼저 ACF를 고려해 보자. 간단히 말해 ACF는 다음과 같이 기술될 수 있다:

---

1) 사실 Hájek(2005, p. 114)은 이와 관련하여 명시적으로 다음을 지적한다. “We have not [...] added new considerations about the disutility of the pain of regretting your choice [...] All the reasoning assumes is that money is the only source of utility in this game, and that you are an expected utility maximizer.”

> **ACF:** 양자택일의 선택에서, 당신의 합리적 미래 자아가 *후회* 할 것이 확실한 선택은 피하라.

'후회'라는 용어를 통한 이 같은 정식화는, 하지만, 애초에 헤이젝이 ACF로 의도한 바를 표현하기 위한 간편하고 직관적인 방식일 뿐이다. 좀더 정확히 ACF는 다음과 같다.

> 양자택일의 선택에서, "당신의 합리적 미래자아가 지금 당신이 선택한 대안이 아닌 다른 대안을 선호할 것이 확실하다면, 지금 선택한 것을 선택해서는 안 된다." (김한승 2014, p. 3; Hájek 2005, p. 114)

주목할 것은 이 같은 정식화에는 '후회'라는 용어는 나타나지 않고, 결정 이론적인 개념인 '선호도' 그리고 주관적 확률 개념으로 결국 분석될 '확실성' 개념만이 나타난다는 것이다. 이것이 보여주는 바는, ACF는 우리가 흔히 '후회'라고 부르는 심성 상태 혹은 그에 보통 동반하는 고통 내지는 부정적 느낌을 전혀 필요로 하지 않는다는 것이다.

그렇다면 '후회'의 애매성은 다음과 같이 설명될 수 있다. 문제의 선택 상황에서 행위자가 곧 '후회'를 한다는 것은 다음과 같은 의미에서 수리기사 역설에 핵심적이다. A, B 양자택일의 상황에서 철수가 확신하고 있는 것 중의 하나는, 그가 만일 A를 (지금) 선택한다면, 잠시 후에 철수는 자신이 애초에 A가 아니라 B를 선택했더라면 더 좋았을 것이라고 생각할 것이라는 점이다. 즉, 철수의 합리적 미래 자아는 과거의 철수가 A를 *선택했기보다는* B를 *선택했기를* 선호한다는 것이다. 물론 이러한 상황을 기술하는 간편한 방식은 '후회'라는 용어를 통해서이다. 그러나 중요한 것은 여기서의 '후회'는 후회에 보통 동반하는 부정적인 현상적 성질을 포함하지 않는다는 것이다. (사실, 바로 아래에서 설명하듯이, 그런 것을

포함시켜도 상관은 없다 – 단지 그것은 문제를 복잡하게 할 뿐이다.)

(3) (1)에서 필자는 수리기사 역설이 믿음(주관적 확률)과 욕구(선호도)의 최소한의 심리학만을 요구한다고 말했다. 그러나 또 한편 (2)에서 주장하기를, 수리기사 역설은 앞에서 말한 바와 같은 명제적 내용을 갖는 상태로서의 후회 개념을 요구한다고 말한다. 그러나 이러한 후회는 반드시 어떤 부정적인 현상적 느낌을 동반하지 않는가? 위에서 필자는 철수가 그러한 부정적인 느낌 없이도 (명제적 내용을 갖는) 후회를 할 수 있다고 가정하였다 – 혹은 그러한 부정적인 느낌에 완전히 무관심할 수 있다고 가정하였다. 그러나 이러한 가정은 수리기사 역설이 최소한의 심리학만을 전제한다는 (1)의 주장과 상충하지 않는가? 그것은 적어도 너무나 비현실적인, 혹은 개념적으로 모순적인 가정 아닌가?

이것이 우려라면 우리는 다음과 같은 수정된 수리기사의 상황을 생각할 수 있다. 이번에도 역시 당신의 선택지는 오전과 오후이다. 또한 당신이 확신하는 것 중 하나는 만일 당신이 지금 오전을 선택하면 잠시 후에 그 선택을 후회하리라는 것이다. 그러나 당신은 그러한 후회의 감정을 피하고 싶어 한다. 여기서 후회의 감정에 대한 당신의 부정적 효용이 천원이라고 가정하자. 그리고 당신에게 오전에 수리기사가 도착하면 11,000원, 오후에 도착하면 10,000원을 지급한다고 하자. 당신은 무엇을 택해야 하는가?[2] 이러한 예에서 여전히 기대 효용 원리는 오전, 오후 선택에 무관심할 것을 요구하는 반면, ACF는 오후 선택을 요구할 것이다.

---

2) 만약 당신의 후회가 시간이 지날수록 커진다면 그에 상응하는 방식으로 당신의 후회(혹은 그것의 부정적 효용)를 보상하는 방법을 생각할 수 있다. 가령, 수리기사가 9시 10분에 도착하면 10,100원, 9시 20분에 도착하면 10,200원, 9시 30분에 도착하면 10,300원, 등등을 지급하는 방식이다.

3. ACF, 어떻게 이해할 것인가?

정리하자면, 애초의 수리기사의 상황으로 돌아가서, 우리가 만약 후회라는 감정이 주는 부정적인 효용을 고려한다면, 마땅히 당신은 오후를 선택해야 하는 것 같다. 여기에는 그러나 아무런 역설도 없다. 역설은 오히려 후회의 부정적인 효용을 고려하지 않았을 때 비로소 발생한다. 그리고 여기서 역설의 원천은 기대 효용 원리와 ACF가 상충하는 것처럼 보인다는 데에 있다. 이러한 역설을 해소하는 한 가지 방식은 물론 ACF를 거부하는 것이다. 그러나 이것이 쉽지 않아 보이는 이유 중의 하나는, ACF 자체가 상당히 그럴듯해 보일뿐만 아니라, 그 원리가 궁극적으로는 기대 효용 원리에 근거하고 있는 것처럼 보이기 때문이다. (그렇다면 그것은 기대 효용 원리와 결정 이론의 내적 일관성에 대한 위협으로 간주될 수 있다.) 이것이 수리기사의 역설이 진정 '역설'로 보이는 이유이다.

이러한 상황에서 한 가지 자연스러운 접근은 ACF가 사실은 기대 효용 원리 이외에 다른 어떤 의심스런 전제에 기반해 있으며 따라서 보기와는 달리 그리 자명한, 그럴듯한 원리가 아님을 보이는 것이다. 만일 이를 보일 수 있다면, 결정 이론이 ACF를 수용하지 못한다는 것은 별 문제가 되지 않는다. 이러한 시도와 관련해서 김한승(2014, p. 14)이 구분하고 있는 ACF의 두 버전이 유용한 출발점이 될 것이다. 그 두 버전은 대략 다음과 같이 표현될 수 있다:

> **보편 ACF:** 양자 택일의 선택에서, 너의 모든 (유관한) 합리적 미래 자아가 후회할 것이 확실한 그러한 선택은 피하라.

> **존재 ACF:** 양자 택일의 선택에서, 너의 어떤 (유관한) 합리

적 미래 자아가 후회할 것이 확실한 그러한 선택은 피하라.

먼저 존재 ACF를 고려해보자. 분명히 존재 ACF는 철수에게 오전 선택을 피할 것을 요구하는 것으로 보인다. 문제의 상황에서 만약 철수가 오전을 선택한다면, 비록 수리기사가 오전에 도착한다 하더라도, 9시 이후부터 수리기사가 도착하기 전까지의 철수의 미래 자아는 자신의 과거 결정을 후회할 것이기 때문이다. (즉, 수리기사가 도착하기 전 시점의 철수의 합리적 미래 자아는 과거의 철수가 오전이 아니라 오후를 선택했기를 원할 것이다.)

다음으로 보편 ACF를 고려해보자. 보편 ACF 역시 오후 선택을 요구하는가? 즉, 철수가 오전을 선택한다면, 철수의 모든 합리적 미래 자아는 그 선택을 후회할 것인가? 그렇지 않다고 말할 수 있는 의미가 있는 것 같다. 만약 수리기사가 오전에 도착한다면, 수리기사가 도착한 이후 시점의 철수의 미래 자아는 자신의 과거 선택을 후회하지 않을 것이기 때문이다. 이는 철수가 오후를 선택하는 경우에도 마찬가지이다. 수리기사가 언제 도착하든, 수리기사가 도착하기 전까지의 철수의 미래 자아는 자신의 과거 선택을 후회하지 않을 것이기 때문이다. 그렇다면 보편 ACF는 철수로 하여금 오전 선택 혹은 오후 선택을 강요하지 않는 것으로 보인다. 이러한 점에서 보편 ACF는 존재 ACF보다 약한 원리로 생각될 수 있으며, 보편 ACF는 기대 효용 원리와 충돌하지 않는다.

그렇다면 수리기사 역설에 대한 다음과 같은 진단이 가능해 보인다: 보편 ACF는 기대 효용 원리와 일관적이다. 그러나 존재 ACF는 철수에게 오후 선택을 요구하며 따라서 기대 효용 원리와 충돌한다. 실제로 수리기사 역설에 대한 헤이젝 스스로의 진단 역시 존재 ACF를 거부하는 방향으로 나아간다. 헤이젝은 ACF가 충분히 의심해 볼만한 원리하고 생각한다. 철수가 오전을 선택하는

경우에, 분명히 자신의 애초의 결정을 후회하는 철수의 합리적 미래 자아 – 즉, 수리기사가 도착하지 전 시점의 철수의 미래 자아 – 가 존재하며, 또한 우리는 철수가 이러한 사실을 (그가 선택을 하는 그 시점에 이미) 알고 있다고 가정하고 있다. 그러나 한편, 헤이젝은 지적하기를, 철수는 자신의 어떤 특정한 (시점의) 미래 자아가 후회할지 알지 못한다. 철수는 가령 9시 7분의 철수의 미래 자아 혹은 11시 30분의 철수의 미래 자아가 후회하리라고 확신할 수 없는데, 이는 수리기사가 9시 7분 이전에 혹은 11시 30분 이전에 도착할 수 있기 때문이다. 그럼에도 불구하는 헤이젝은 철수의 어떤 합리적 자아가 후회하리라는 것이 확실하다면("knowing full well that there will be *some* future self of yours") 역설은 발생한다고 주장하며 (더 상세한 논의 없이) 논문을 마무리짓는다 (p. 118).

이러한 헤이젝의 관찰은 존재 ACF가 다음의 두 가지 방식으로 읽힐 수 있음을 보여준다.

> **약한 ACF:** 양자택일의 선택지 중 어느 하나에 대해서, 너의 어떤 특정한 (시점의) 합리적 미래 자아가 그 선택을 후회할 것이 확실하다면, 그러한 선택은 피하라.
>
> **강한 ACF:** 양자택일의 선택지 중 어느 하나에 대해서, 그 선택을 후회할 합리적 미래 자아가 존재하는 것이 확실하다면, 그러한 선택은 피하라.

그렇다면 수리기사 역설에 대한 헤이젝의 진단은 약한 ACF는 기대 효용 이론과 일관적이지만, 강한 ACF는 그것과 충돌한다는 것이다. 위에서 보았듯 강한 ACF가 기대 효용 원리와 충돌한다는 것

은 분명해 보인다. 한편 약한 ACF는 어떻게 기대 효용 원리와 일관적인가? 헤이젝의 추론은 다음과 같은 것이었다. 가령 철수가 오전을 선택한다고 하자. 오전 9시와 12시 사이의 어떤 특정한 임의의 시점 t에 대하여, 우리는 (혹은 철수는) t에서의 철수의 합리적 미래 자아가 자신의 애초의 선택을 후회할 것임을 알지 못한다. 왜냐하면 t 이전에 수리기사가 도착할 수도 있기 때문이다. 이는 12시와 3시 사이의 시점에 대해서도 마찬가지이다. 즉 약한 ACF는 철수에게 오후 선택을 요구하지 않는다. 또한 마찬가지 방식으로 약한 ACF는 철수에게 오전 선택을 요구하지도 않음을 쉽게 볼 수 있으며, 따라서 기대 효용 원리와 충돌하지 않는다.

이상의 논의가 대체로 옳다면 수리기사 역설에 대한 한 가지 가능한 해결 방향은 다음과 같다. 그것은 강한 ACF가 사실, 보편 ACF와 약한 ACF와 마찬가지로, 기대 효용 원리와 충돌하지 않음을 보이거나, 혹은 보기와는 다르게 강한 ACF가 그리 그럴듯한 원리가 아님을 보이는 것이다; 강한 ACF는 어쩌면 너무 강한 원리일지 모른다. 이러한 방향에 대한 자세한 논의는 이 글의 범위를 넘어선다. 여기서는 단지 ACF 원리 – 특히 '유관한 합리적 미래 자아' – 를 어떻게 정식화하고 이해할 것인지의 문제와 관련하여 한 가지 필자가 보기에 흥미로운, 기본적인 관찰을 제시한다.

ACF를 정확히 어떻게 정식화해야 하는지의 쟁점과 관련해서 가장 중요한 문제 중의 하나는 '유관한 합리적 미래 자아'를 어떻게 이해해야 하는지의 문제일 것이다. 앞에서 필자는 보편 ACF가 철수에게 오후 선택을 강요하지 않는다고 주장하면서 다음을 지적한 바 있다. 철수가 오전을 선택하는 경우에 만약 수리기사가 오전에 도착한다면, 수리기사가 도착한 이후 시점의 철수의 미래 자아는 자신의 과거 선택을 후회하지 않을 것이다; 그러므로 철수의 모든 유관한 합리적 미래 자아가 자신의 애초의 선택을 후회하지는 않으

며, 따라서 ACF는 철수에게 오후 선택을 강요하지 않는다. 그러나 이러한 주장에는 문제가 있다. 그것은 수리기사 사례의 경우, 수리기사가 도착한 이후 시점의 철수의 미래 자아 – 좀더 정확히, 수리기사의 도착 시점을 아는 미래 자아 – 는 유관한 합리적 미래 자아가 아니라는 것이다.

이를 보기 위해 철수가 동전 던지기 내기를 하는 단순한 사례로 돌아가보자. 앞 혹은 뒷면이 나올 확률이 같은 상황에서 철수는 앞, 뒷면의 내기 중에 특별한 선호가 없는 것이 합리적이다. 그러나 이 경우 다음과 같은 사실이 분명 성립할 것이다. 철수가 앞면에 내기를 거는 경우, 동전이 만일 뒷면이 나온다면 그 이후 시점의 철수의 합리적 미래 자아는 자신의 애초의 선택을 후회할 것이다. 마찬가지로, 철수가 뒷면에 내기를 거는 경우, 동전이 만일 앞면이 나온다면 그 이후 시점의 철수의 합리적 미래 자아는 자신의 애초의 선택을 후회할 것이다. 하지만 이러한 미래 자아들의 후회는 철수의 애초의 선택과 관련하여 어떠한 중요한 것도 함축하지 않는다. 만일 동전이 던져진 시점 이후의 철수의 미래 자아를 보편 ACF의 평가와 관련한 철수의 유관한 합리적 미래 자아로 간주한다면, 보편 ACF는 철수에게 앞면 선택도 피할 것을 요구하고 뒷면 선택도 피할 것을 요구할 것이다. 이러한 이해 방식은 ACF를 사소하고 쓸모없는 것으로 만드는 잘못된 이해 방식이다.

정리하자면, 단순한 동전 던지기 사례의 경우, 동전이 앞면 혹은 뒷면이 나온 특정한 가능성 하에서의 동전이 던져진 시점 이후의 미래 자아는 철수의 유관한 미래 자아가 아니라고 해야 한다. 그렇다면 마찬가지로 수리기사 사례의 경우, 수리기사가 어떤 특정한 시점에 도착한 가능성 하에서의 도착 시점 이후의 미래 자아는 철수의 유관한 미래 자아가 아니다. 수리기사 사례가 역설적으로 보이는 이유 중의 하나는 수리기사의 도착 시점을 모르는 철수의 어

떤 합리적 미래 자아가 (도착 시점을 알지 못함에도 불구하고) 자신의 과거 결정을 후회하리라는 것을 철수가 선택의 시점에 이미 알고 있다는 것이다.[3] 이러한 관찰은 수리기사 사례의 경우 수리기사가 도착한 시점 이전의 철수의 미래 자아만을 유관한 미래 자아로 간주해야 할 것이라는 생각을 부추긴다. 한 가지 지적할 것은, 만일 이러한 식으로 유관한 합리적 미래 자아를 이해한다면 수리기사 사례의 경우 존재 ACF 뿐만 아니라 보편 ACF 역시 철수로 하여금 오후 선택을 요구하는 것으로 보인다는 점이다. 그 이유는, 만약 수리기사가 도착하기 이전의 미래 자아만이 유관한 미래 자아라면, 철수가 오전을 선택하는 경우 철수의 모든 유관한 합리적 미래 자아가 자신의 애초 선택을 후회할 것이기 때문이다. (이러한 고려는 수리기사 사례의 경우, 보편 ACF와 존재 ACF 사이의 구분보다 약한 ACF와 강한 ACF 사이의 구분이 더 중요한 구분임을 시사한다.)

그러나 이러한 방식이 유관한 합리적 미래 자아에 대한 유일한 이해 방식은 아니다. 실제로 우리가 앞에서 왜 약한 ACF는 철수에게 오후 선택을 요구하지 않는지를 논의할 때 우리는 철수의 미래 자아를 수리기사가 도착하기 이전의 것으로만 제한하지 않았다. 앞에서 보았듯, 철수는 임의의 어떤 특정한 시점 t에서의 자신의 미래 자아가 자신의 과거 선택을 후회할지 후회하지 않을지 알지 못하는데, 그것은 t 이전에 수리기사가 도착할 가능성이 있기 때문이다. (특정한 시점 t의 철수의 미래 자아에 대해서 그 시점이 수리기사가 도착한 이후의 시점일 수도 있으나, 중요한 것은 선택의 시점에서 철수는 그 t 시점이 수리기사가 도착한 이후일지 이전일지 알

3) 물론 철수의 미래 자아가 수리기사의 도착 시점에 대해 아무것도 알지 못하는 것은 아니다. 철수가 오전을 선택하는 경우, 다음의 조건을 만족하는 철수의 미래 자아가 존재한다. 철수의 t 시점의 미래 자아는 t 이전에 수리기사가 도착하지 *않았음*을 안다.

지 못한다는 것이다.) 사실 필자는 이러한 방식의 이해 – 즉, 수리기사의 도착 시점에 의하지 않는 방식으로 유관한 미래 자아를 이해하는 방식 – 에 더 공감한다. 아무튼 이러한 관찰은 수리기사 사례의 경우 철수의 유관한 합리적 미래 자아를 정확히 어떻게 봐야 할 것인지, 혹은, 좀더 일반적인 수준에서, 유관한 합리적 미래 자아 개념에 대한 어떠한 일반적인 기준이나 제한 원리가 있을 수 있는지에 대한 면밀한 논의가 필요함을 보여준다.

4. 결론

이상으로 필자는 수리기사 역설의 핵심은 후회와 같은 감정에 있는 것이 아니라 표준적인 결정이론의 틀에서 ACF를 수용할 수 있는가의 문제로 파악해야 한다고 논하였다. 이 경우 역설은 기대효용 원리와 ACF 사이에 한편으로는 충돌이 있는 것처럼 보이면서도 또 한편으로는 그렇지 않은 것으로 보인다는 점에서 발생한다. 따라서 이 역설을 해결하기 위해서는 ACF 원리를 정확히 어떻게 이해하고 정식화해야 하는가에 대한 면밀한 탐구가 필요하다. 바로 이 쟁점이 수리기사 역설의 해결을 위해 먼저 충분히 논의되어야 할 이슈라고 생각한다.

참고문헌

김한승 (2014), “확률과 시간에 관한 두 가지 퍼즐”, 『철학적 분석』, 29호, pp. 1-22.

Hájek, A. (2005) “The Cable Guy Paradox”, *Analysis* 65, pp. 112-119.

고등과학원
Korea Institute for Advanced Study
cwon99@gmail.com

What is paradoxical about the Cable Guy paradox?

Chiwook Won

In his recent paper, Hanseung Kim (2014) discusses the Cable Guy paradox and argues that given the nature of regret, it is rational to bet on the cable guy's afternoon arrival, though it is equally probable for him to arrive in the morning or afternoon interval. In this paper, I argue that regret is not essential to the paradox in that the paradox still arises even if we ignore regret. I then briefly discuss a possible way to proceed to resolve the paradox.

Key Words: The Cable Guy paradox, regret, probability, Hanseung Kim