과학철학 19권 3호 (2016) pp. 1-20

잠자는 미인에게 '아직'과 '이미'†

김 명 석‡

엘가와 루이스는, 잠자는 미인 문제에서 동전을 던지는 시점이 일요일 저녁이든 월요일 저녁이든, 이 시점이 미인의 답변을 달라지게 하지 않는다고 주장한다. 학자들이 이 주제를 이미 많이 논의했지만, 동전을 월요일 저녁에 던지는 설정에서 이 문제의 올바른 답변이 무엇인지는 놀랍게도 그다지 탐구하지 않았다. 이 논문은 이런 설정에서 미인의 올바른 답을 찾는 이른바 '잠자는 미인의 아직 문제'를 풀고자 한다. 동전을 아직 던지지 않았다는 사실을 미인이 알게 되면, 앞으로 던질 멀쩡한 동전이 앞면이 나올 것이라고 믿는 그의 믿음직함은 당연히 1/2이어야 한다. 이러한 직관을 "아직의 원리"라고 부를 수 있으며, 이것은 엘가의 직관이기도 하다. 하지만 아직의 원리로부터 따라 나오는 결과들은 루이스의 예상과 다르고, 심지어 엘가의 예상과도 다르다. 우리 셈에 따르면 미인이 어느 날 깨어났을 때 앞면이 나올 그의 믿음직함은 1/2보다 작아야 한다. 우리는 그 값이 어림잡아 2/5라고 셈했다. 잠자는 미인의 문제는 조건화 규칙, 루이스의 '주요 원칙', 반 프라센의 '반성 원리' 등을 의심하게 하는 반례가 될 수 없다. 그 대신 우리는 믿음직함 셈에서 '아직'과 '이미'가 다른 결과를 낳을 수도 있다는 오래된 교훈을 다시금 확인한다.

【주요어】 믿음직함, 시간, 신념도, 주요 원칙, 중심 명제, 지표

---

† 박일호 선생은 잠자리에 들 시간과 심지어 새벽녘까지 페이스타임으로 토론해 주셨는데, 토론의 처음에는 전혀 예상하지 못했던 열매를 얻게 되었다. 이에 선생께 큰 고마움을 전한다. 이 논문의 심사자들께서 논문의 보완을 위해 합당한 논평을 주신 것에도 감사드린다.

‡ 국민대학교 교양대학, myeongseok@gmail.com.

1. 아직의 원리

한 명제가 다른 명제보다 더 참일 것이라 믿을 까닭이 없는 한, 두 명제에 대한 나의 믿음직함(credence)은 똑같아야 한다. 우주에 진정한 우연이 있고 내가 그 우연의 일어남직함(chance)을 안다면, 그러한 일어남직함에서 벗어났다고 생각할 다른 정보를 내가 갖고 있지 않는 한, 나의 믿음직함은 그 일어남직함을 따라야 한다. 데이비드 루이스는 이러한 원칙을 “주요 원칙”이라 했다. 통상의 일어남직함에서 벗어났을지 모른다는 생각을 뒷받침해주는 정보를 “못 받아들이게 하는 증거”[1]라 한다. 못 받아들이게 하는 증거를 내가 갖고 있지 않는 한, 한 우연 사건이 일어날 것이라는 나의 믿음직함은 그 우연 사건의 일어남직함을 그대로 따라야 한다.

멀쩡한 동전을 던졌을 때 앞면이 나오거나 뒷면이 나오는 것이 엄밀한 의미에서 우연은 아니다. 우주에 진정한 우연이 있는지 없는지는 그 자체로 매우 중요한 물음이자 어려운 물음이다. 내기를 건다면 나는 없다는 쪽에 걸 것이다. 아무튼 던진 동전이 앞면이 나오는 사건의 일어남직함은 1/2이라고 말들 한다. 동전을 아직 던지지 않았다는 것을 내가 안다면, 주요 원칙에 따라, “동전이 앞면이 나온다”에 대한 나의 믿음직함은 1/2이어야 한다. 한편 윌슨은 못 받아들이게 하는 증거를 갖고 있지 않는 이에게 이미 일어난 사건조차도 “사실상 우연한”(effectively chancy) 사건으로 여겨질 수 있다고 주장했다.[2] 동전을 이미 던졌다는 것을 내가 안다 하더라도, 못 받아들이게 하는 증거를 내가 갖고 있지 않는 한, 주요 원칙에 따라, “동전이 앞면이 나왔다”에 대한 나의 믿음직함은 1/2이어야 한다.

잠자는 미인 문제에서 1/3이 올바른 답이라고 주장하는 엘가에게 가

---

1) 이 표현은 “inadmissible evidence”를 옮긴 말이다. 이 낱말은 편견, 뜬소문, 무관한 이야기, 나쁘게 얻은 증거 등처럼 법정에 제시될 수 없는 또는 받아들일 수 없는 증거를 말한다. 이 글에서 ‘못 받아들이게 하는 증거’란 기존의 믿음을 이전과 똑같이 받아들이기 어렵게 하는 증거를 말한다.

2) Wilson (2014), pp. 584-89.

장 중요한 가정은 실험하는 사람이 동전을 일요일 밤에 던지든, 미인을 재운 뒤 월요일에 던지든, 물음의 답변은 같다는 것이다.[3] 이 문제의 올바른 답변이 1/2이라고 주장하는 루이스조차도 엘가의 이 가정을 의심하지 않고 받아들인다.[4] 엘가의 가정을 받아들이면, 미인이 오늘이 월요일이라는 사실을 알게 되면 그는 동전이 아직 던져지지 않았다는 것을 알게 된다. 윌슨이 표현했듯이, “잠자는 미인이 앞날을 내다볼 수 없고, 뒤로 거슬러 미치는 인과작용 같은 것이 전혀 없다면”,[5] 곧 던질 동전이 앞면이 나올 것이라는 미인의 믿음직함은, 주요 원칙에 따라서, 1/2이어야 한다. 그래서 오늘이 월요일이라는 사실을 미인이 알게 된 다음 동전이 앞면이 나올 그의 믿음직함 C(H|월)은 1/2이다.[6] 여기서 H는 “던진 또는 던질 동전이 앞면이 나온다”이고 “월”은 “오늘은 월요일이다”이다. 하지만 루이스는 C(H|월)이 2/3라면서, 이 사례가 자신의 주요 원칙이 적용되지 않는 사례라고 생각했다.[7] 브래들리는 오늘이 월요일이라는 정보를 들을 경우 미인은 못 받아들이게 하는 증거를 갖게 되는 것이고 이 경우 미인은 C(H|월)이 1/2이라고 계속 생각할 이유가 없다고 주장한다.[8]

우리는 이 논문에서 동전을 던지는 시점을 월요일 미인이 답변을 마치고 다시 잠든 때로 실험을 설정했을 때 이 문제의 올바른 답변이 무엇이어야 하는지를 탐구하고자 한다. 미인은 월요일 또는 화요일에 깨어날 것이다. 월요일에 미인이 다시 잠든 다음 화요일에 깨어났을 때 그는 월요일에 깨어났다는 사실을 전혀 기억하지 못한다. 월요일에 미

---

3) Elga (2000), pp. 144-45. 송하석(2012)은 전자의 설정을 “엘가 설정”이라 하고 후자의 설정을 “루이스 설정”이라고 불렀다.

4) Lewis (2001), p. 172.

5) Wilson (2014), p. 589, 각주 13.

6) 우리는 미인이 일요일 저녁 이후 깨어났을 때 그가 명제 X에 대해 갖는 믿음직함 함수를 C(X)라고 쓸 것이다.

7) Lewis (2001), p. 175.

8) Bradley (2015), p. 688. 한편 코지는 C(H) = 1/2을 받아들이지만, 이른바 이미지 규칙을 통해 C(H|월) = 1/2이어야 한다고 주장한다. Cozic (2011), pp. 139-40.

인이 다시 잠든 다음 동전을 던져 앞면이 나오면 미인을 다시 깨우지 않지만, 뒷면이 나오면 미인을 화요일에 다시 깨운다. 월요일인지 화요일인지 모르는 어느 날 미인이 깨어났을 때 미인은 던질 또는 던진 동전이 앞면이 나오리라고 얼마큼 믿어야 하는가? 간단히 말해 C(H)는 얼마여야 하는가? 이 문제를 원래 잠자는 미인의 문제와 구별하기 위해 “잠자는 미인의 아직 문제”라고 부르겠다.9) 엘가는 두 문제들에 모두 1/3이라고 답하고, 루이스는 두 문제들에 모두 1/2이라고 주장한다.

물론 우리는 실험하는 사람이 동전을 일요일 밤에 던지든, 미인을 재운 뒤 월요일에 던지든, 잠자는 미인 문제의 답변은 같다고 미리 가정하지 않겠다. 다만 우리가 다룰 문제는 잠자는 미인의 아직 문제이기 때문에 우리는 다음 직관을 큰 거리낌 없이 받아들인다.

(Y) C(앞으로 던질 그 동전이 앞면이 나올 것이다 | 아직 그 동전을 던지지 않았다) = 1/2

동전을 아직 던지지 않았다면, 앞으로 던질 멀쩡한 동전이 앞면이 나올 믿음직함은, 주요 원칙에 따라, 당연히 1/2이어야 한다.10) 이러한

9) 원래 문제는 “잠자는 미인의 이미 문제”라고 부르면 되겠다. 이 문제의 답이 1/3이라고 주장하는 이들을 “1/3주의자”(thirders)라고 하고, 그 답이 1/2이라고 주장하는 이들을 “1/2주의자”(halfers)라고 한다. 1/3주의자들은 C(H|월)이 1/2이라고 주장하지만 대부분의 1/2주의자는 C(H|월)이 2/3라고 주장한다. 한편 코지처럼 C(H)와 C(H|월)이 모두 1/2이라고 주장하는 이들이 있는데, 이런 견해를 “겹1/2주의자”(double-halfers)라고 부르겠다. Cozic (2011). Titelbaum(2012)은 1/3주의 관점에서 겹1/2주의를 비판한다. 그런데 세 진영 모두 내가 ‘잠자는 미인의 아직 문제’라고 부른 문제가 ‘잠자는 미인의 이미 문제’와 다를 것이 없다고 주장한다. 나는 이 논문에서 잠자는 미인의 아직 문제에 대해 세 진영 모두 올바른 답을 제안하지 못했다고 주장할 것이다.

10) 주요 원칙을 쓰지 않고 논제 (Y)를 표현할 수 있다. 아직 던지지 않은 그 동전을 곧 던질 경우 그 동전이 앞면이 나오리라고 믿을 자신의 믿음직함이 1/2이어야 한다는 것을 미인은 알고 있다. 이 경우 논제 (Y)는 다음과

직관을 "아직의 원리"라고 부르겠다. (Y)는 C(H|월) = 1/2이라는 주장과 다르지만, (Y)로부터 이것이 따라 나온다는 것을 보일 수 있다.

우리는, 아직의 원리를 써서, 다음과 같은 점을 주장하고자 한다. 첫째, 아직의 원리를 받아들일 경우, 루이스의 견해는 옳지 않다. 둘째, 아직의 원리를 받아들인다고 해서, 엘가의 견해가 옳다는 것이 따라 나오지 않는다. 셋째, 당연히 C(월|T) + C(화|T) = 1이겠지만, 그렇다고 C(월|T)와 C(화|T)가 같은 것은 아니다. 여기서 T는 "던진 또는 던질 동전이 앞면이 나온다"이고, "화"는 "오늘은 화요일이다"이다. 엘가는 너무 선불리 C(월|T) = C(화|T)를 가정했다. 1/3주의자들의 가장 큰 잘못은 C(H|월) = C(T|월) = 1/2을 주장할 때는 동전을 월요일 밤에 던진다고 가정하고, C(월|T) = C(화|T) = 1/2을 주장할 때는 동전을 일요일 밤에 던진다고 가정하는 것이다. 하지만 그들은 동전을 월요일 밤에 던질 때 왜 우리가 C(월|T) = 1/2을 받아들여야 하는지 설명하거나 논증하지 않은 채 너무 빠르게 지나쳐 버린다.[11] 우리는 잠자는 미인의 아직 문제와 이미 문제를 각각 다르게 풀이하는 것이 좋겠다. 믿음직함 셈에서는 "아직"과 "이미"를 잘 가려야 한다.

이와 같은 내용을 제2절과 제3절에서 다룬 뒤 제4절에서는 앞면이 나오는 일의 일어남직함이 h인 동전을 던지고, 뒷면이 나오면 미인을 N번 깨우는, 매우 일반화된 실험 설정에서 C(H)를 어림하고자 한다. 물론 미인은 치우친 동전을 월요일 저녁 이후에 던진다는 것을 알고 있으며, 그 동전이 앞면이 나오는 일의 일어남직함이 h라는 것과, 뒷면이 나오면 자신을 N번 깨운다는 것과, 각 깨움 사이에 자기 기억을 지

같이 쓸 수 있다. C(앞으로 던질 그 동전이 앞면이 나올 것이다 | 아직 그 동전을 던지지 않았고, 아직 던지지 않은 그 동전을 곧 던질 경우 그 동전이 앞면이 나오리라고 믿을 믿음직함은 1/2이다) = 1/2.

11) C(월|T) = C(화|T)를 주장하는 근거는 무차별 원리라고 말들 한다. 동전을 일요일 저녁에 던지는 경우라면, 이미 뒷면이 나왔다는 조건에서 미인이 월요일과 화요일에 차이를 둘 까닭이 없다. 하지만 동전을 월요일 저녁에 던지는 경우라면, 동전이 이미 뒷면이 나왔다는 것과 동전이 곧 뒷면이 나올 것이라는 데 미인이 차이를 둘 까닭이 있다. 물론 한 심사자께서 지적했듯이, C(월|T)와 C(화|T)가 다르다고 미리 가정하는 것도 너무 섣부르다.

운다는 사실도 모두 알고 있다. C(H)를 어림하기 위해 우리는 몇 가지 극한 값을 가정할 것이다. 제4절의 내용은 추정 또는 어림짐작이기 때문에 논의의 엄밀성과 확실성이 다소 떨어진다.

2. 지금은 아직과 이미를 가른다

월요일인지 화요일인지 알 수 없는 어느 날 깨어난 미인에게 명제 H와 T는 다음과 같이 바꾸어야 한다.

(E1) $H = H_{\text{나왔}} \vee H_{\text{나올}}$

(E2) $T = T_{\text{나왔}} \vee T_{\text{나올}}$

여기서 "$H_{\text{나왔}}$"은 "던진 동전이 이미 앞면이 나왔다"이고 "$H_{\text{나올}}$"은 "던질 동전이 곧 앞면이 나올 것이다"이다. 또 "$T_{\text{나왔}}$"은 "던진 동전이 이미 뒷면이 나왔다"이고 "$T_{\text{나올}}$"은 "던질 동전이 곧 뒷면이 나올 것이다"이다. 이들 가운데 둘 이상이 함께 참일 수는 없다. (E1)과 (E2)가 매우 하찮게 보이지만 이 논제는 매우 중요한 역할을 한다. H와 T는 지표가 없는 명제이지만, $H_{\text{나왔}}$, $H_{\text{나올}}$, $T_{\text{나왔}}$, $T_{\text{나올}}$ 등은 지표를 갖고 있다. 이 지표가 믿음직함 셈에 차이를 낳는다.

(E1)과 (E2) 말고도 매우 미더운 논제들을 간추리면 아래와 같다.

(E3) $C(H_{\text{나왔}}|\text{월}) = C(H_{\text{나왔}}\&\text{월}) = C(H_{\text{나왔}}|\text{화}) = C(H_{\text{나왔}}\&\text{화}) = 0$

(E4) $C(T_{\text{나왔}}|\text{월}) = C(T_{\text{나왔}}\&\text{월}) = 0$

(E5) $C(T_{\text{나왔}}|\text{화}) = 1$

(E6) $C(H_{\text{나올}}|\text{월}) = C(T_{\text{나올}}|\text{월}) = 1/2$[12]

(E3), (E4), (E5)는 문제의 설정으로부터 곧바로 따라 나온다. 월요일

12) 루이스 같은 1/2주의자는 이 논제를 거부할 것이다. 각주 20에서 보겠지만 이 논제를 거부할 경우 매우 야릇한 귀결을 얻게 될 것이다.

에는 동전을 아직 던지지 않았기 때문에 동전이 이미 앞면이 나왔다는 주장은 거짓이며, 동전이 이미 뒷면이 나왔다는 주장도 거짓이다. 또 지금이 화요일이라는 것을 미인이 안 다음 그는 동전이 이미 뒷면이 나왔다는 것을 알게 된다. 또한 지금이 화요일이라는 것을 미인이 안 다음 그는 이미 던진 동전이 앞면이 나오지 않았다는 것도 안다. (E3)으로부터 곧장 $C(H_{\text{나왔}}) = 0$이 따라 나오고, 후자로부터 곧장 (E3)이 따라 나온다. (E6)은 앞에서 “아직의 원리”라고 부른 것으로부터 따라 나온다. 왜냐하면 “지금이 월요일이다”라는 정보는 “아직 동전을 던지지 않았다”는 것을 함축하고, 깨어난 미인에게 후자는 전자를 함축하기 때문이다.

한편 $H \vee T$는 참이기 때문에 $1 = C(H \vee T|\text{화}) = C(H_{\text{나왔}} \vee H_{\text{나올}} \vee T_{\text{나왔}} \vee T_{\text{나올}}|\text{화}) = C(H_{\text{나왔}}|\text{화}) + C(H_{\text{나올}}|\text{화}) + C(T_{\text{나왔}}|\text{화}) + C(T_{\text{나올}}|\text{화})$인데, (E3)과 (E5)를 써서, $C(H_{\text{나올}}|\text{화}) + C(T_{\text{나올}}|\text{화}) = 0$이라는 것을 알 수 있다. 믿음직함은 0이거나 0보다 크기 때문에, $C(H_{\text{나올}}|\text{화})$와 $C(T_{\text{나올}}|\text{화})$가 둘 다 0이라는 것을 알 수 있다. 간추리면

(1) $C(H_{\text{나올}}|\text{화}) = C(H_{\text{나올}} \& \text{화}) = C(T_{\text{나올}}|\text{화}) = C(T_{\text{나올}} \& \text{화}) = 0$

달리 말해, 화요일에는 동전을 이미 던졌기 때문에, 미인이 물음을 듣고 있는 시점 이후 동전이 앞면이 나올 것이라는 주장은 거짓이며, 미인이 물음을 듣고 있는 시점 이후 동전이 뒷면이 나올 것이라는 주장도 거짓이다. 나아가 엘가의 직관도 이끌어낼 수 있다. $C(H_{\text{나왔}}|\text{월})$은 0이고, $C(H_{\text{나올}}|\text{월})$은 1/2이기 때문에,

(2) $C(H|\text{월}) = C(H_{\text{나왔}} \vee H_{\text{나올}}|\text{월}) = C(H_{\text{나왔}}|\text{월}) + C(H_{\text{나올}}|\text{월}) = 1/2$

이다. 식 (2)는 단순히 가정한 것이 아니라 (E1), (E2), (E3), (E6) 등을 써서 이끌어낸 것이다. 루이스가 식 (2)를 거부하고자 한다면 (E1), (E2), (E3), (E6) 가운데 적어도 하나를 버려야 한다. 그가 버릴 만한 것은 (E6)밖에 없다. 이를 버린다는 것은 ‘아직의 원리’를 거부한다는

말인데, 이를 거부하기 위해서 그는 1/2을 '못 받아들이게 하는 증거'를 갖고 있어야 한다. 그는 멀쩡한 동전을 던지면 앞면이 나올 일어남직함이 1/2이라는 것을 알고 있고, 동전을 아직 던지지 않았다는 것을 지금 알고 있는데도, 동전이 앞면이 나올 믿음직함이 왜 1/2이 아닐 수 있다고 생각할까? 그럴 만한 증거가 전혀 없는 것처럼 보인다.

3. 이미와 아직은 반반이 아니다

잠자는 미인의 아직 문제에서 C(H)가 1/2보다 작다는 것을 증명할 수 있다. 하지만 C(H)가 1/3이라는 것은 증명되지 않는다. C(월∨화|H) = 1이기 때문에, C(월|H) + C(화|H) = 1이다. (E1), (E3), (E6)을 써서

$$1 = C(\text{월}|H) + C(\text{화}|H) = C(\text{월}\&H)/C(H) + C(\text{화}\&H)/C(H)$$
$$= C(\text{월}\&H_{\text{나올}})/C(H) = C(H_{\text{나올}}|\text{월})C(\text{월})/C(H) = C(\text{월})/2C(H).$$

이로부터

(3) $C(H) = C(\text{월})/2$

을 얻는다. C(월)은 1보다 작기 때문에 C(월)/2은 1/2보다 작다. 따라서 (3)으로부터

(4) $C(H) < 1/2$

이다. 이 점은 루이스 견해와 양립할 수 없지만, 엘가 견해와 양립할 수 있다. 물론 (4)만으로는 엘가의 1/3주의가 옳다는 것을 보여줄 수 없다. 한편 C(월) + C(화) = 1이고 C(T) = 1 – C(H)이기 때문에

(5) $C(T) = C(\text{월})/2 + C(\text{화})$

이다.

$C(H)$가 1/3이라는 것을 보여주기 위해서 필요한 것은 $C(\text{월}|T) = C(\text{화}|T) = 1/2$이다. 엘가가 $C(H|\text{월}) = 1/2$을 주장할 때는 동전을 월요일 밤에 던진다고 가정하고, $C(\text{월}|T)$와 $C(\text{화}|T)$가 같다고 주장할 때는 동전을 일요일 밤에 던진다고 가정한다. 잠자는 미인의 아직 문제에서 $C(\text{월}|T) = C(\text{화}|T) = 1/2$이라고 주장하는 것은 합당한 근거를 갖는가? 이를 알아보기 위해 $C(\text{월}|T)$와 $C(\text{화}|T)$를 따로 셈해 보자.

먼저 (E2), (E3), (E6), (5)로부터

$$C(\text{월}|T) = C(\text{월}\&T)/C(T) = C(\text{월}\&T_{\text{나올}})/C(T)$$
$$= C(T_{\text{나올}}|\text{월})C(\text{월})/C(T) = C(\text{월})/2C(T) = C(\text{월})/\{C(\text{월}) + 2C(\text{화})\}.$$

한편 $C(\text{화}|T)$를 계산하면 이와 같지 않다. (E2), (E3), (E5), (5)로부터

$$C(\text{화}|T) = C(\text{화}\&T)/C(T) = C(\text{화}\&T_{\text{나왔}})/C(T)$$
$$= C(T_{\text{나왔}}|\text{화})C(\text{화})/C(T) = 2C(\text{화})/\{C(\text{월}) + 2C(\text{화})\}.$$

간추리면

(6) $C(\text{월}|T) = C(\text{월})/\{C(\text{월}) + 2C(\text{화})\}$

(7) $C(\text{화}|T) = 2C(\text{화})/\{C(\text{월}) + 2C(\text{화})\}$

우리는 (6)과 (7)로부터 $C(\text{월}|T) = C(\text{화}|T)$를 이끌어낼 수 없다. 만일 엘가가 그랬듯이 $C(\text{월}|T) = C(\text{화}|T)$를 가정한다면, (6)과 (7)로부터 $2C(\text{화}) = C(\text{월})$을 얻으며, 이것은 $C(\text{월}) = 2/3$라는 것을 뜻한다. 이것과 (3)으로부터 $C(H) = 1/3$을 얻는다. 하지만 잠자는 미인의 아직 문제에서 $C(\text{월}|T) = C(\text{화}|T)$를 가정하는 것은 너무 섣부르다. 뒷면이 나올 것이거나 나왔다고 해서, 뒷면이 나올 것이라는 믿음과 뒷면이 나왔을 믿음이 대칭을 이룰 까닭은 없다.

4. 아주 많은 물음들

보스트롬은 동전이 뒷면이 나올 경우 미인에게 아주 여러 번 깨워 물음을 던지는 상황을 생각해 보았다.[13] 가령 동전이 뒷면이 나올 경우 모두 N번을 깨우고 묻고 잊게 하고 재운다고 해보자. 이 경우 C(H)는 어떤 값을 가져야 할까? "월"이나 "화" 말고 Q_1, Q_2, …, Q_n 등을 쓰는 것이 좋겠는데, Q_n이란 "나는 지금 깨어나 물음을 받고 있으며 지금은 n번째 깨어남이고 n번째 물음이다"를 뜻한다. 실험하는 사람은 미인이 첫째 물음에 답하고 그를 재운 다음에 둘째 물음을 묻기 전에 동전을 던진다. 매우 미더운 우리의 논제들은 다음과 같이 바뀐다. (E7*)은 새로 보태었다. 미인이 뒷면이 나왔다는 것을 알고 있다면, 현재의 물음이 2번째인지 3번째인지 또는 몇 번째인지를 가늠할 만한 그 어떤 증거도 그는 갖고 있지 않다.

(E3*) $C(H_{\text{나왔}}|Q_n) = C(H_{\text{나왔}}\&Q_n) = 0$
(E4*) $C(T_{\text{나왔}}|Q_1) = C(T_{\text{나왔}}\&Q_1) = 0$
(E5*) n이 1이 아닐 경우, $C(T_{\text{나왔}}|Q_n) = 1$
(E6*) $C(H_{\text{나올}}|Q_1) = h$
(E7*) n이 1이 아닐 경우, $C(Q_2) = C(Q_n)$

멀쩡한 동전의 경우 h는 1/2이다. 보다 일반화된 경우를 염두에 두고 넓게 셈하고 싶어 기호로 나타내었다. 우리는 나중에 h가 1에 가까운 동전 또는 0에 가까운 동전을 던지는 실험 설정을 고려할 것이다. 물론 $C(T_{\text{나올}}|Q_1) = 1 - h = t$이다.

이들 기본 논제들로부터 다음을 얻을 수 있다.

---

13) Bostrom (2007), p. 62. 이 물음에 1/3주의자는 0에 가까이 가는 값을 C(H)에 주고, 1/2주의자는 여전히 1/2을 준다. 보스트롬은 멀쩡한 동전을 던졌을 경우, 설사 N이 매우 크다 하더라도, C(H)가 0으로 가까이 가는 결과는 어쨌든 피해야 한다고 생각한다.

(8) n이 1이 아닐 경우, $C(H_{나올}|Q_n) = C(H_{나올}\&Q_n)$
$= C(T_{나올}|Q_n) = C(T_{나올}\&Q_n) = 0$
(9) $C(H|Q_1) = C(H_{나왔} \vee H_{나올}|Q_1) = C(H_{나올}|Q_1) = h$

(E1*)과 (E3*)과 (E6*)을 써서

$$1 = C(Q_1|H) + C(\sim Q_1|H) = C(Q_1\&H)/C(H) + C(\sim Q_1\&H)/C(H)$$
$$= C(Q_1\&H)/C(H) = C(Q_1\&H_{나올})/C(H)$$
$$= C(H_{나올}|Q_1)C(Q_1)/C(H) = hC(Q_1)/C(H)$$

를 얻는다. 이후 $C(Q_n)$을 q_n이라고 쓰겠다. 이로부터

(10) $C(H) = hC(Q_1) = hq_1$

이다. $C(T) = 1 - C(H)$이기 때문에

(11) $C(T) = 1 - hq_1$

이다.

(E7*)에 따라 n이 1이 아닌 경우 각 $C(Q_n)$들은 모두 똑같은 크기일 테니,

$$1 = C(Q_1) + C(Q_2) + \cdots + C(Q_N) = q_1 + (N - 1)q_2$$

이다. 이로부터

(12) $q_2 = (1 - q_1)/(N - 1)$

을 얻는다. 우리는 $C(Q_1|T)/C(Q_2|T)$를 q로 놓고 이 값을 추정하고자 한다. (E2), (E3*), (E6*), (11)로부터, $C(Q_1|T) = C(Q_1\&T)/C(T) = C(Q_1\&$

$T_{나올})/C(T) = C(T_{나올}|Q_1)C(Q_1)/C(T) = tq_1/C(T)$이다. 또한 $C(Q_2|T) = C(Q_2\&T)/C(T) = C(Q_2\&T_{나왔})/C(T) = C(T_{나왔}|Q_2)C(Q_2)/C(T) = q_2/C(T)$이다. 따라서

(13) $q \equiv C(Q_1|T)/C(Q_2|T) = tq_1/q_2 = t(N - 1)q_1/(1 - q_1)$

이다. q는 0보다 항상 크다. 식 (13)으로부터

(14) $q_1 = q/\{t(N - 1) + q\}$
(15) $C(H) = hq_1 = qh/\{t(N - 1) + q\}$[14)]

를 얻는다.

식 (13)에서 q 대신에 $q/(N - 1)$를 p로 놓고 우리가 구한 식을 표현하면 식이 말해주는 것이 보다 또렷해진다.[15)]

(16) $p \equiv q/(N - 1) = tq_1/(1 - q_1)$

이를 식 (14)와 (15)에 대입하면 다음을 얻는다.

(17) $C(Q_1) = q_1 = p/(p + t)$
(18) $C(T_{나왔}) = C(\sim Q_1) = 1 - q_1 = t/(p + t)$
(19) $C(H) = hq_1 = hp/(p + t) < h$[16)]

---

14) $N = 1$일 때, 이 식에 따르면, $C(H) = h$이다. 이것은 우리 예상과 일치한다. N이 1일 때는 q가 정의되기 않기 때문에, N은 2이거나 2보다 큰 자연수라고 생각하는 것이 좋겠다.

15) 박일호 선생은 이 점을 나에게 깨우쳐 주었다. p는 $C(Q_1|T)/C(\sim Q_1|T)$ 또는 $C(T_{나올})/C(T_{나왔})$으로 정의된다.

16) 참고로 $C(T) = 1 - C(H) = t(p + 1)/(p + t) > t$. $C(T_{나올}) = C(T) - C(T_{나왔}) = tp/(p + t) < t$.

여기서 $p/(p + t)$는 언제나 1보다 작다. $h = t = 1/2$인 경우, C(H)가 언제나 1/2보다 작다는 결론을 또 다시 얻을 수 있다.

엘가와 루이스에게 $q = 1$이다.[17] $N = 2$이고, $h = 1/2$인 경우 $C(H) = 1/3$이고 이것은 엘가의 계산과 일치한다. 하지만 앞에서 말했듯이 $q = 1$이라는 것은 섣부른 가정에 지나지 않는다. 식 (15)에 $q = 1$을 대입하면

$$C(H) = h/(tN + h)$$
$$C(T) = tN/(tN + h)$$

를 얻는다. 하지만 $q = 1$을 가정하면, N이 매우 크고 h가 거의 1인 실험 실정에서도, C(H)는 거의 0일 수 있다. 이처럼 임의의 N과 임의의 h에 대해, q가 1이라고 가정하는 것을 야릇한 귀결을 갖는다.

5. 매우 치우친 동전

C(H)를 추정하려면 합당하게 q를 가늠해야 한다. q는 $C(Q_1|T)/C(Q_2|T)$ 또는 tq_1/q_2인데, N이 클수록 q_2는 그만큼 작아질 것이다. 그래서 q는 N에 따라 언제나 증가할 것 같다. 우리는 q가 다음과 같은 함수 꼴이 될 것이라 추정한다.

$$(20)\quad q = aN + b$$

여기서 a와 b는 실수이다. 큰 N에 대해서도 q는 양수이기 때문에 a는 양수가 되어야 할 것이다. 우리가 식 (20)을 추정하는 까닭은 다음과 같다. 첫째, q는 N보다 더 느리게 커져서는 안 된다. 둘째, q는 N보다 더 빨리 커져서는 안 된다.

---

17) 이들에게 N이 2보다 클 경우에도, $q = 1$이다.

우리는 q의 함수 꼴을 추정하기 위해 다음과 같은 4개의 제약 조건을 요구한다.

(F1) 뒷면이 거의 나올 수 없는 동전을 던지는 실험에서, N이 아무리 크다 하더라도, 잠에서 깨어난 미인은 동전이 뒷면이 나오거나 나왔을 것이라 확신해서는 안 된다.
(F2) 앞면이 거의 나올 수 없는 동전을 던지는 실험에서, N이 아무리 크다 하더라도, 잠에서 깨어난 미인은 아직 동전이 던져지지 않았을 것이라고 확신해서는 안 된다.
(F3) 거의 뒷면만 나오는 동전을 던지는 실험에서, N이 2인 경우, $C(Q_1|T) \simeq C(Q_2|T)$ 또는 $C(T_{\text{나올}}) \simeq C(T_{\text{나왔}})$이어야 한다.[18)]
(F4) 거의 대부분 뒷면이 나오는 동전을 던지는 실험에서, N이 무한대인 경우, $C(T_{\text{나왔}})$은 거의 t에 가까이 간다.[19)]

(F1)과 (F2)와 (F3)에 반영된 우리 직관들을 거부하기란 매우 어려울 것이다. 반면에 (F4)는 다소 의심의 여지가 있다.

q가 N보다 더 느리게 커진다고 가정해 보자. 이 경우 식 (14)와 (15)에서 가장 빨리 커지는 것은 분모에 있는 tN이다. 따라서 N이 매우 큰 실험 설정에서 C(H)는 거의 0의 값을 갖는다. 설사 우리가 h를 거의 1로 보내는 실험 설정에서도, 다시 말해 뒷면이 거의 나올 수 없

18) 이것은 t가 거의 1이고 N이 2인 경우, p나 q가 거의 1이라는 것을 뜻한다.
19) $C(T_{\text{나왔}})$은 C(T)보다 작다. 각주 16에서 보였듯이 C(T)는 $1 - C(H) = t(p + 1)/(p + t) > t$이다. 하지만 만일 t가 1에 가까이 간다면, $(p + 1)/(p + t)$이 1에 가까이 가고, 이것은 C(T)도 t에 가까이 간다는 것을 뜻한다. N이 무한대로 가면, $C(T_{\text{나올}})$이 점차 줄어들 것이기 때문에, C(T)에서 $C(T_{\text{나올}})$을 뺀 나머지 $C(T_{\text{나왔}})$이 t에 가까이 간다고 말해도 될 것 같다.

각 제약조건을 다르게 쓰면 다음과 같다. (F1) t가 거의 0인 경우, N이 아무리 크다 하더라도, C(T)는 1에 가까이 가서는 안 된다. (F2) h가 거의 0인 경우, N이 아무리 크다 하더라도, q_1은 1에 가까이 가서는 안 된다. (F3) N은 2이고 t가 거의 1인 경우, $C(Q_1|T) \simeq C(Q_2|T)$ 또는 $C(T_{\text{나올}}) \simeq C(T_{\text{나왔}})$. (F4) t가 1에 상당히 가깝고 N이 무한대인 경우, $C(T_{\text{나왔}}) \simeq t$.

는 실험 실정에서도, N이 아주 크다는 이유로, 잠에서 깨어난 미인은 이미 뒷면이 나왔다고 매우 굳게 믿어야 한다. 이것은 제약조건 (F1)에 어긋난다. 따라서 q는 N보다 더 느리게 커져서는 안 된다.

이제 q가 N보다 더 빨리 커진다고 가정해 보자. 이 경우 식 (14)에서 q가 가장 빨리 커지기 때문에 N이 커질 때 q_1은 1에 가까이 갈 것이다. h를 거의 0으로 보내는 설정에서도, N이 아주 크다는 이유로, 오히려 미인은 자신이 받고 있는 현재의 물음이 첫째 물음이라고 매우 굳게 믿어야 한다. 다시 말해 앞면이 나올 가능성이 거의 없는 동전을 던지는 설정에서, 깨어난 미인은 아직 동전이 던져지지 않았다고 생각해야 한다. 하지만 이 귀결은 제약조건 (F2)에 어긋난다. 따라서 q는 N보다 더 빨리 커져서는 안 된다.

q가 N보다 더 느리게 커져서는 안 되고, N보다 더 빨리 커져서도 안 되기 때문에, 우리는 식 (20)을 받아들여야 한다. 이제 남은 것은 상수 a와 b를 찾는 것이다. 이를 위해 제약조건 (F4)를 써야 한다. N을 무한대로 놓는 실험 실정에서 p는 $q/(N-1) = (aN+b)/(N-1) \simeq a$이다. 식 (18)을 써서 $C(T_{나왔})$을 셈하면,

$$C(T_{나왔}) = t/(p+t) \simeq t/(a+t)$$

이다. 제약조건 (F4)에 따르면 $t/(a+t)$는 t와 같아야 한다. 이것은 $a+t$가 1이라는 것을 뜻하며, $a = 1 - t = h$이다. 따라서 $q = Nh + b$. 한편 제약조건 (F3)에 따르면 $N = 2$이고, h가 거의 0일 때, q는 1에 가까이 가야 하기 때문에, $b = 1$이어야 한다. 따라서

(21) $q = Nh + 1$
(22) $q_1 = (Nh+1)/(N+h)$
(23) $C(H) = hq_1 = (Nh+1)h/(N+h)$

$h = 1/2$인 경우, $C(H) = (N+2)/(4N+2)$이다. N이 매우 큰 실험 설정에서 C(H)는 1/4에 가까이 간다. $h = 1/2$이고, $N = 2$인 실험 설정에서

(24) $C(H) = 2/5$

이다. 이 값은 1/3과 1/2 사이에 있다. 식 (23)을 받아들일 경우, h가 거의 1이고 t가 거의 0인 실험 실정에서, C(H)가 0으로 무너지는 일은 일어나지 않는다. 식 (22)와 (23)은 우리의 제약 조건을 모두 만족한다.

미인이 엄청나게 많은 물음을 들어야 하는 실험 설정에서 $C(Q_1)$과 C(H)는 다음과 같다.[20)]

(25) $q_1 \simeq h$

(26) $C(H) = hq_1 \simeq h^2$

이것은 엘가 등 1/3주의자뿐만 아니라, 루이스 등 1/2주의자들에게 다소 뜻밖에 결과이다. 이런 결과가 나온 까닭을, h = 1/2이고 N이 거의 무한대인 실험 설정을 놓고 생각해 보자. h가 1/2이라는 것은, 미인이 한 번의 물음을 받게 되는 상황에 들어갈 가능성과 N번의 물음을 받게 되는 상황에 들어갈 가능성이 같다는 것을 뜻한다. 그래서 미인이 깨어났을 때 지금이 이미 첫째 날 이후 시점일 가능성은 대략 1/2일 것이다. 지금 물음이 첫째 물음이 아니리라 믿을 믿음직함은 1/2이고,

20) 우리 추정이 옳다면, h가 1/2이고 N이 거의 무한대인 경우, $C(Q_1)$은 거의 1/2이고, C(H)는 거의 1/4이다. 1/3주의자의 기존 계산은 0이고, 1/2주의자의 기존 계산은 1/2이다.

한편 우리의 직관 (E6) 또는 (E6*)을 받아들이지 않는 1/2주의는 매우 야릇한 귀결을 갖는다. 그들의 계산에 따르면, C(월) = C(월|H)C(H) + C(월|T)C(T) = (1)(1/2) + (1/N)(1/2) = (N + 1)/2N이기 때문에, C(H|월) = C(월|H)C(H)/C(월) = N/(N + 1)이다. 따라서 (E6*)을 거부할 경우, 미인이 지금이 월요일임을 안 다음, 그는 곧 던질 동전이 거의 확실히 앞면이 나올 것이라고 믿어야 한다. 넓게 말해, 1/2주의의 셈을 따를 경우, C(H|월) = Nh/(Nh + t)인데, 심지어 h가 매우 작은 값일 때도, N이 거의 무한대가 될 경우 C(H|월)은 거의 1에 다가간다. 지금이 월요일이라면 아직 동전을 던지지 않았는데, 앞면이 나오기 몹시 어려운 동전이 장차 거의 확실히 앞면이 나올 것이라고 믿어야 한다는 것은 매우 야릇한 귀결이다.

지금 물음이 첫째 물음이리라 믿을 믿음직함은 1/2이다. 따라서 N이 거의 무한대인 상황에서, 미인이 깨어났을 때 지금이 첫째 날일 믿음직함은 1/2에 가까이 가고, 곧 던질 동전이 앞면이 나올 믿음직함은 1/2이다. 이 경우 C(H)는 1/4에 가까이 가야 한다.

6. 나오는 말

엘가와 루이스를 포함하여 거의 모든 학자들이, 잠자는 미인 문제에서 동전 던지는 시점이 일요일 저녁이든 월요일 저녁이든, 이 시점이 미인의 답변을 달라지게 하지 않는다고 주장했다. 우리는 동전을 월요일 저녁에 던지는 설정에서 이 문제의 올바른 답변이 무엇인지 찾고자 했다. 그 값이 1/2보다 작아야 한다는 것은 거의 분명하다. 우리는 몇 가지 가정을 보태어 그 값이 2/5가 아닐까 추정했다.

동전을 일요일 저녁에 던지는 상황과 동전을 월요일 저녁에 던지는 상황이 같다는 기존 학자들의 가정은 엄청나게 많은 혼란을 낳았다. 1/2주의자는, 동전을 아직 던지지 않았다는 사실을 우리가 알게 되면, 앞으로 던질 멀쩡한 동전이 앞면이 나올 것이라고 믿는 우리의 믿음직함이 당연히 1/2이어야 한다는 직관을 포기한다. 1/3주의자는 자신들이 기대고 있는 두 개의 핵심 논제를 각기 다른 두 상황에서 정당화한다. 논제 C(H|월) = C(T|월)은 동전을 월요일 저녁에 던지는 상황에서 정당화한다. 하지만 논제 C(월|T) = C(화|T)는 동전을 일요일 저녁에 던지는 상황에서 정당화한다. 동전을 월요일 저녁에 던지는 설정으로 실험을 고정했을 때 이 논제를 정당화하기란 쉽지 않다.

우리 연구에서 가장 중요한 논제는 다음 둘이다.

(E1) $H = H_{\text{나왔}} \vee H_{\text{나올}}$

(E2) $T = T_{\text{나왔}} \vee T_{\text{나올}}$

H와 T는 지표가 없는 비중심 명제이지만, $H_{\text{나왔}}$, $H_{\text{나올}}$, $T_{\text{나왔}}$, $T_{\text{나올}}$ 등은

지표를 갖고 있는 중심 명제이다. 이를 통해 중심 명제가 어떻게 비중심 명제에 영향을 줄 수 있는지 잘 드러나게 되었다. 지표 정보는 지표가 있는 명제 즉 중심 명제의 믿음직함을 바꿀 수 있다. 잠자는 미인의 아직 문제에서 지표 정보는 $H_{나왔}$, $H_{나올}$, $T_{나왔}$, $T_{나올}$ 등의 믿음직함을 바꾼다. $H_{나왔}$, $H_{나올}$, $T_{나왔}$, $T_{나올}$ 등의 믿음직함이 바뀔 경우, (E1)과 (E2) 때문에, 지표가 없는 명제 즉 비중심 명제 H와 T의 믿음직함도 바뀌게 된다. 하지만 동전을 일요일 저녁에 던지는 문제 설정에서 $H_{나올}$과 $T_{나올}$은 거짓 명제이고, H와 동치인 $H_{나왔}$과 T와 동치인 $T_{나왔}$은 더 이상 중심 명제가 아니다.

우리는 미인에게 아주 많은 물음이 주어지는 상황과 엄청나게 치우친 동전을 던지는 상황을 고려하여 C(H)의 값을 가늠해 보았다. 앞면이 나올 가능성이 h인 동전을 월요일 저녁에 던지고, 뒷면이 나오면 월요일 이후 N－1번 깨우는 실험 설정에서, 언제인지 알 수 없는 때 깨어난 미인에게 C(H)는 (Nh + 1)h/(N + h)이어야 한다고 가늠했다. h가 1/2이고 N이 2인 실험 설정에서 C(H)는 2/5로 추정되었다. 또한 h가 1/2이고 N이 거의 무한대인 실험 설정에서 C(H)는 1/4로 추정되었다. 이 추정들은 확실성이 조금 떨어지지만 우리는 충분히 진실에 가까이 다가갔다고 믿는다. 우리 셈과 어림이 옳다면, 이것이 다른 학자들의 기존 논의에 대해 갖는 함축은 가늠할 수 없을 만큼 많고 넓다. 이 함축은 다른 지면으로 미루는 것이 좋겠다. 또한 동전을 일요일 저녁에 던지는 설정에서 올바른 답변이 무엇인지에 대한 우리의 논증도 다른 지면으로 미루겠다.[21]

---

21) 잠자는 미인 문제의 1/3주의에 대한 우리의 비판은 김명석(2016a)와 (2016b)를 보라.

참고문헌

김명석 (2016a), 「잠자는 미녀의 시공간」, 『과학철학』 19권 1호, pp. 1-26.

_____ (2016b), 「믿음직함과 가능 세계: 잠자는 미녀의 경우」, 『범한철학』 80집, pp. 239-59.

송하석 (2012), 「잠자는 미녀 문제, 다시 보기」, 『철학적 분석』 25호, pp. 1-21.

Bostrom, N. (2007), “Sleeping Beauty and Self-Location: A Hybrid Model”, *Synthese* 157(1): pp. 59-78.

Bradley, D. (2015), “Everettian Confirmation and Sleeping Beauty: Reply to Wilson”, *British Journal for the Philosophy of Science* 66(3): pp. 683-93.

Cozic, M. (2011), “Imagining and Sleeping Beauty: A Case for Double-Halfers”, *International Journal of Approximate Reasoning* 52(2): pp. 137-43.

Elga, A. (2000), “Self-locating Belief and the Sleeping Beauty Problem”, *Analysis* 60: pp. 143-47.

Lewis, D. (1980), “A Subjectivist’s Guide to Objective Chance”, reprinted in Eagle, A. (ed.) (2011), *Philosophy of Probability*, Routledge, pp. 458-87.

_____ (2001), “Sleeping Beauty: Reply to Elga”, *Analysis* 61: pp. 171-76.

Titelbaum, M. (2012), “An Embarrassment for Double-Halfers”, *Thought: A Journal of Philosophy* 1(2): pp. 146-51.

Wilson. A. (2014), “Everettian Confirmation and Sleeping Beauty”, *British Journal for the Philosophy of Science* 65(3): pp. 573-98.

논문 투고일	2016. 10. 14.
심사 완료일	2016. 11. 09.
게재 확정일	2016. 11. 18.

Sleeping Beauty on ‘Already’ and ‘Yet’

Myeongseok Kim

A. Elga and D. Lewis insisted that sleeping beauty’s credence upon awakening in the coin’s landing heads ought to be the same regardless of whether we use the following two methods. Method Sunday: first tossing the coin and then waking sleeping beauty up either once or twice depending on the outcome. Method Monday: first waking her up once, and then tossing the coin to determine whether to wake her up a second time. This paper pursues the right answer to the sleeping beauty problem taken Method Monday. According to our answer, sleeping beauty’s credence in the coin’s landing heads ought to be less than 1/2, approximately 2/5. However, the sleeping beauty problem cannot be a counterexample against the principal principle of Lewis.

Keywords: centered proposition, credence, principal principle, sleeping beauty problem, time