논리연구 14-1(2011) pp. 1~22

잠자는 미녀의 문제, 그의 대답은?* †

송 하 석

【요약문】 이 글은 잠자는 미녀의 문제에 대하여 엘가의 1/3주의자의 견해를 지지하고, 루이스가 주장하는 1/2주의를 비판한다. 특히 최근 1/2주의를 옹호하는 흥미로운 견해를 제시한 프란체스치의 논변을 비판한다. 나아가서 보스트롬과 김한승의 절충주의는 논점을 선취하거나 잠자는 미녀의 문제가 진정한 역설임을 입증해야 하는 부담을 지닌다고 주장한다. 이 글의 주된 목적은 잠자는 미녀의 문제에 대한 1/3주의의 주장이 직관적이라는 주장과 함께, 1/2주의자의 주장의 오류가 무엇인지를 밝히는 것이다.

【주요어】 잠자는 미녀의 문제, 확률, 엘가, 루이스, 절충주의 견해

* 접수일: 2011. 1. 14. 심사 및 수정완료일: 2011. 2. 2. 게재확정일: 2011. 2. 9.

† 매우 신랄하면서도 유익한 심사평을 해주어 이 논문을 보다 명료하게 할 수 있도록 해 주신 익명의 심사위원들께 감사를 전한다.

1. 들어가는 말

잠자는 미녀의 역설(Paradox of Sleeping Beauty)을 간단히 소개하면 다음과 같다. 일요일 저녁 미녀(이후 SB)를 재운다. 그리고 일요일 자정에 정상적인 동전을 던져서 앞면이 나오면 월요일 아침에만 SB를 깨운다. 동전이 뒷면이 나오면, 월요일에 깨우고 잠시 후 그가 깨어났다는 기억을 지우는 약을 먹이고 다시 재운 다음, 화요일 아침에 다시 깨운다. SB는 이러한 실험내용을 알고 있고, 그는 매우 탁월한 확률 계산가이다. 이런 상황에서 월요일에 SB가 깨워졌을 때, 그러나 그날이 월요일이라는 사실을 그가 모른다고 했을 때, 그가 동전을 던져서 앞면이 나올 것이라고 믿을 믿음의 정도는 얼마인가?

이에 대해서 1/2이라는 주장과 1/3이라는 주장이 그럴 듯한 근거를 가지고 제시되고 있다. 1/2이라고 주장하는 대표적인 철학자(1/2주의자, halfers)는 루이스(D. Lewis), 미첨(C. Meacham), 화이트(R. White), 그리고 프란체스치(P. Franceschi) 등이다. 반면에 1/3이라고 주장하는 사람(1/3주의자, thirders)들은 또 둘로 나눌 수 있는데, 엘가(A. Elga), 먼튼(B. Monton) 등은 SB가 월요일에 깨워졌다는 사실은 그가 믿음의 정도를 바꿀 새로운 증거는 아니지만 SB가 실험을 시작하기 전과 월요일에 깨워졌을 때 동전이 앞면일 나올 확률에 대한 믿음의 정도가 변하여 1/3이라고 주장하고, 도르(C. Dorr), 와인트로브(R. Weintraub), 호간(T. Hogan) 등은 월요일에 깨워졌다는 사실은 그의 확률에 대한 믿음의 정도를 변하게 할 새로운 정보이기 때문에 1/3이라고 주장한다.

이 글은 SB의 문제에 대한 올바른 대답은 1/3이라고 주장하고, 왜 루이스 등이 주장하는 1/2주의가 왜 옳지 않은지 논증할 것이다. 특히 최근 1/2주의에 대한 설득력 있는 논변을 제시한 것으로

인정되는 프란체스치가 제시한 논변에 대해서 비판하고 1/3의 입장을 옹호할 것이다. 또한 최근 보스트롬(N. Bostrom)의 절충주의를 지지하면서 제시한 김한승의 견해를 비판적으로 검토할 것이다. 이를 위해서 다음 절에서 SB 문제에 대한 논의를 촉발시킨 엘가와 루이스의 논변을 간단히 살펴보고, 그 차이에 대해서 설명할 것이다. 그리고 3절에서는 프란체스치가 루이스의 1/2의 입장을 옹호하기 위해서 제시한 논변을 소개하고 그에 대하여 비판하고, 마지막으로 4절에서 보스트롬의 절충주의와 김한승의 견해에 대해서 살펴볼 것이다.

2. 엘가와 루이스의 논변

먼저 엘가의 1/3주의 논변을 살펴보자. 그는 SB가 월요일에 깨워졌을 때 그가 있을 수 있는 경우를 다음 세 가지로 구별한다.

H_1: 동전이 앞면이 나오고 월요일에 깨워진 경우.
T_1: 동전이 뒷면이 나오고 월요일에 깨워진 경우.
T_2: 동전이 뒷면이 나오고 화요일에 깨워진 경우.

엘가의 논변은 두 단계로 이루어지는데, 첫 번째 단계는 SB가 T_1에 대해서 갖는 믿음의 정도와 T_2에 대해서 갖는 믿음의 정도가 같음을 증명하는 단계이고, 두 번째 단계는 SB가 H_1에 대해서 갖는 믿음의 정도와 T_1에 대해서 갖는 믿음의 정도가 같다는 것을 증명하는 단계이다. 다시 말해서 엘가의 논변은 다음을 증명하는 두 과정으로 이루어진다.

1) $P(T_1) = P(T_2)$
2) $P(H_1) = P(T_1)$

SB가 동전 던지기 결과가 뒷면임을 안다면, 그는 자신이 T_1에 있거나 T_2에 있음을 알 것이다. 이 경우 T_1에 있음과 T_2에 있음은 주관적으로 아무런 차이가 없다. 즉 SB는 자기가 T_1에 있든 T_2에 있든 정확하게 동일한 명제를 참이라고 생각할 것이다. 그러므로 다음이 성립한다.

$$P(T_1|T_1 \vee T_2) = P(T_2|T_1 \vee T_2)$$

따라서 다음 (가)가 성립한다.

(가) $P(T_1) = P(T_2)$

또 SB가 오늘이 월요일임을 안다면, 그는 자신이 H_1의 상태에 있거나 T_1의 상태에 있음을 알 것이다. 이 경우 그가 H_1의 상태에 있음을 믿을 믿음의 정도는 정상적인 동전을 던져서 앞면이 나올 것이라고 믿을 믿음의 정도와 같다. ‘H’를 ‘동전이 앞면이었다.’를 나타낸다고 하자. 그러면,

$$P(H_1|H_1 \vee T_1) = P(H)$$

이다. 그런데 $P(H) = 1/2$이므로,

$P(H_1|H_1 \vee T_1) = 1/2$이다.

이로부터 다음 (나)를 증명할 수 있다.

(나) $P(H_1) = P(T_1)$

(가), (나)로부터 다음이 따라 나온다.

$$P(H_1) = P(T_1) = P(T_2)$$

그런데

$$P(H_1) + P(T_1) + P(T_2) = 1$$

이므로 다음이 증명된다.

$$P(H_1) = 1/3$$

엘가는 이상의 논변을 통해서 SB가 실험에 참여하기 전에 동전이 앞면이 나올 확률은 1/2이라고 믿지만, 그가 월요일에 깨워질 때, 그 믿음이 변해서 1/3이 된다고 주장한다. 엘가는 SB가 월요일에 깨워졌을 때 그의 믿음이 변하는 이유는 그에게 새로운 정보가 주어졌기 때문이 아니라 자신의 시간적 위치(temporal location)가 동전이 앞면임이라는 사건(이하 H)의 참과 무관한 것으로 여겨지는 상황에서 자신의 시간적 위치가 그 사건의 참과 관련된 것으로 여겨지는 상황으로 변했기 때문이라고 설명한다. 엘가에 따르면, 새로운 정보란 정보를 받아들이는 주체가 지금까지 가진 증거에 의해서 배제하지 않았던 가능세계를 배제하는 증거이다. 그리고 SB는 월요일에 깨워질 것이라는 사실에 대해서 이미 알고 있었고 따라서 그가 깨워졌다는 사실은 지금까지 배제하지 않았던 가능세계를 배제할 만한 새로운 증거가 아니라는 것이다. 그러나 SB가 깨워짐은 자신의 시간적 위치가 어떤 명제의 참과 관련된 것으로 여기지 않다가 새롭게 그 명제의 참과 관련된 것으로 여기게 하는 사건이다. 엘가에 따르면, 행위자가 자신의 시간적 위치를 어떤 명제의 참과 관련된 것으로 간주한다는 것은 그 행위자의 믿음이 자신이 어떤 시각 t에 있음과 양립가능하고, 자신이 t에 있다는 조건

하에서 그 명제에 대한 자신의 믿음의 정도가 그러한 조건이 없는 경우와 다른 그러한 시각 t가 있다는 말이다. 다시 말해서 SB가 월요일에 깨워졌을 때, 그러나 그는 그 날이 월요일인지 화요일인지 알 수 없을 때, SB가 실험이 시작되기 전에 “동전을 던져서 앞면이 나왔다(이하 ‘H’로 표기함).”에 대해서 가졌던 믿음과 다른 믿음을 갖게 되는데, 그 이유는 SB가 새로운 정보를 갖게 되어서가 아니라 자신의 시간적 위치가 H의 참과 무관한 것으로 여겨지는 상황에서 자신의 시간적 위치가 H의 참과 관련된 것으로 여겨지는 상황으로 변했기 때문이다.[1)]

이에 대해서 루이스는 엘가의 첫 번째 논변을 받아들이지만, 두 번째 논변이 옳지 않다고 주장한다. 그는 다음과 같은 몇 가지 확률 함수를 제시한다.

> P: 월요일에 깨워진 직후 SB가 갖는 믿음의 정도를 나타내는 함수
> P+: 월요일 깨워져서 오늘이 월요일임을 알려준 후, SB가 갖는 믿음의 정도를 나타내는 함수
> P-: 실험에 관한 설명을 듣고 재워지기 전에 SB가 갖는 믿음의 정도를 나타내는 함수

여기서 루이스는 엘가의 처음 논변의 주장,

(1) $P(T_1) = P(T_2)$

을 받아들인다. 그리고 다음과 같은 엘가도 동의하는 주장을 제시한다.

(2) $P\text{-}(H) = P\text{-}(T) = 1/2$
(3) $P+(H) = P(H|H_1 \vee T_1)$
$P+(T) = P(T|H_1 \vee T_1)$

---

1) Elga (2000), p. 145.

(4) $P(H|H_1 \vee T_1) = P(H_1|H_1 \vee T_1) = P(H_1) / [P(H_1) + P(T_1)]$
$P(T|H_1 \vee T_1) = P[(T_1 \vee T_2)|(H_1 \vee T_1)]$
$= P(T_1) / [P(H_1) + P(T_1)]$
(5) $P(H) = P(H_1)$
$P(T) = P(T_1 \vee T_2) = P(T_1) + P(T_2)$
(6) $P\text{-}(H) = 1/2$
$P\text{-}(T) = 1/2$

그런데 루이스는 월요일에 SB가 깨워졌다는 사실은 믿음을 바꿀 만한 새로운 정보가 아니라는 점에 주목하여 다음을 주장한다.

$$P\text{-}(H) = P(H)$$

즉 SB가 실험을 하기 전에 정상적인 동전을 던져서 앞면이 나올 것이라고 믿을 믿음의 정도와 월요일에 깨워졌을 때 정상적인 동전을 던져서 앞면이 나올 것이라고 믿을 믿음의 정도가 같다는 것이다. 그런데 $P(H) = 1/2$이고, 따라서 다음이 성립한다.

$$P(H_1) = 1/2$$

그리고

$$P(T) = 1/2$$

이므로 (5)에 의해서,

$$P(T_1) = P(T_2) = 1/4$$

이다.

루이스와 엘가의 가장 근본적인 차이는 $P(H) = 1/2$인가, $P+(H)= 1/2$인가에 있다. 엘가는 $P+(H)$가 1/2이라고 주장하고, 반면에 루이

스는 P(H)가 1/2이라고 주장한다. P+는 SB가 지금이 월요일임을 아는 경우의 확률함수를 나타낸다. 그런데 SB가 지금이 월요일임을 아는 경우에, 아직 동전 던지기가 이루어지지 않았다고 믿는다면 동전 던지기는 미래의 사건이므로 당연히 동전의 앞면이 나올 확률은 1/2이라고 믿을 것이고, 이미 동전 던지기가 이루어졌다고 해도, 자신이 지금 월요일에 깨워졌음을 안다면, 자신이 월요일에 깨워지는 것은 동전 던지기의 결과와 상관이 없다는 것을 알고 있기 때문에 여전히 H에 대한 그의 믿음의 정도는 1/2이라는 것이 엘가의 설명이다. 그러나 루이스는 SB가 자신이 깨워졌다는 것은 H에 대한 그의 원래의 믿음에 아무런 영향을 주지 못하기 때문에 P-(H)와 P(H)는 동일하다고 주장한다.

누구의 주장이 더 설득력 있는 살펴보기 위해서, SB가 월요일에 깨워져서 그가 그날이 월요일임을 알 때, 그가 H에 대해서 갖는 믿음의 정도에 대해서 생각해보자. 이에 대해서 루이스는 다음과 같이 주장한다.

$$\begin{aligned} P+(H) &= P(H|H_1 \vee T_1) \\ &= P(H_1) \ / \ [P(H_1) + P(T_1)] \\ &= (1/2)/ \ [(1/2) + (1/4)] \\ &= 2/3 \end{aligned}$$

루이스는 SB가 월요일에 깨워졌고, 그날이 월요일임을 아는 사실은 미래에 일어날 동전 던지기에서 앞면이 나올 확률에 대한 그의 믿음을 바꾸게 하는 정보임을 설명해야 한다. 루이스는 이에 대해서 “그 원리[미래의 사건에 대한 확률은 그 사건이 발생할 가능성과 같다는 원리]는 단서를 필요로 한다”고 말하면서 그런데 그 단서는 “우리가 [P-(H)=1/2을] 얻기 위해서 그 원리를 사용할 때는 만족되지만, 엘가가 [P+(H)=1/2을] 얻기 위해서 그 원리를 사용할 때는 만족되지 않는다.”고 말한다.[2] 이어서 그는 “[SB]가 월요일에

깨워져 있는 동안 그 날이 월요일임을 듣게 될 때, 그는 미래에 관한 증거, 자신이 지금 T_2의 경우에 있지 않다는 증거를 얻게 되는 셈이고,” 그것은 그날이 월요일이라고 듣기 전에는 갖지 않은 새로운 증거라고 주장한다. 그리고 루이스는 엘가도 다음이 성립한다는 것을 인정한다고 주장한다.

$$P+(H) = P(H) + 1/6$$

이로부터 루이스는 SB가 월요일에 깨워져서 그 날이 월요일임을 듣게 되는 것은 그의 믿음을 바꾸게 하는 증거라고 말한다. 그런데 루이스가 설명해야 하는 것은 SB가 월요일에 깨워졌을 때, 그 날이 월요일임을 모르는 상태에서 H에 대해서 갖는 믿음의 정도와 그 날이 월요일임을 알게 되었을 때 그가 H에 대해서 갖는 믿음의 정도가 변한다는 것이 아니다. 그가 설명해야 하는 것은 SB가 월요일에 깨워져서 그 날이 월요일임을 들음으로써 미래의 동전 던지기에서 앞면이 나올 확률을 1/2이라는 일반적인 믿음을 버리고 2/3이라는 믿음을 갖게 되는 이유이다. 다시 말해서 그는 $P-(H)=1/2$이지만 $P+(H)=2/3$인 이유가 무엇인지를 설명해야지, $P+(H)$에 대한 믿음의 정도와 $P(H)$에 대한 믿음의 정도가 다른 이유를 설명해야 하는 것은 아니다. 왜냐하면 루이스도 지적하듯이 엘가도 후자는 받아들이기 때문이다. 그런데 위에서 루이스는 $P+(H)$에 대한 믿음의 정도와 $P(H)$에 대한 믿음의 정도가 다른 이유를 설명하고 있을 뿐, $P-(H)$와 $P+(H)$에 대한 믿음의 정도가 다른 이유를 설명하고 있지는 않다.

2) Lewis (2001), p. 175.

3. 프란체스치의 3분의 1 주장에 대한 비판과 응답

프란체스치는 “잠자는 미녀와 세계 환원의 문제”라는 논문에서 SB 문제의 옳은 답이 1/2이고, 왜 1/3의 주장이 옳지 않은지를 설명한다. 이를 위해서 그는 다음과 같은 경우를 생각해보도록 제안한다.

> 정상적인 동전을 던져서 앞면이 나오면 빨간색 공 하나를 항아리에 넣고, 뒷면이 나오면 빨간색 공 하나와 초록색 공 하나를 항아리에 넣는다. 이렇게 만들어진 항아리에서 공을 꺼냈을 경우 빨간색 공을 꺼낼 확률은 얼마인가?[3]

이에 대해서 다음과 같은 대답이 가능할 것이다.

$$P(R) = (1/2 \times 1) + (1/2 \times 1/2) = 3/4$$

즉, 만약 정상적인 동전을 던져서 앞면이 나온다면, 항아리에는 빨간색 공만 하나 들어갈 것이므로 빨간색 공을 꺼낼 확률은 1이고, 뒷면이 나온다면 항아리에 빨간색 공과 초록색 공이 하나씩 들어갈 것이므로 빨간색 공을 꺼낼 확률은 1/2이다. 따라서 이 항아리에서 빨간색 공을 꺼낼 확률은 위와 같은 계산에 의해서 3/4이다. 그러나 다른 대답도 가능해 보인다. 위와 같은 시행을 n번 반복해서 항아리를 채웠다고 하자. 그러면 항아리에 빨간색 공의 개수와 초록색 공의 개수는 다음과 같이 될 것이다.

$$N(R) = (1/2 \times 1 \times n) + (1/2 \times 1 \times n) = n$$
$$N(G) \fallingdotseq (1/2 \times 0 \times n) + (1/2 \times 1 \times n) = n/2$$

---

[3] Franceschi (2005), p. 2.

그러므로 항아리 안의 전체 공의 개수는 3n/2이고, 그 중 n개가 빨간색 공이므로, 빨간색 공을 꺼낼 확률은 2/3이다. 즉,

$$P(R) = 2/3$$

프란체스치는 이 문제에 대한 옳은 대답은 3/4이지 2/3가 아니라고 주장한다. 그는 두 번째 풀이가 옳지 않은 이유를 하나의 독립적 사건과 하나의 사건의 일부분일 뿐인 것 사이의 차이를 파악하지 못했기 때문이라고 말한다. 즉 동전을 던져서 앞면이 나오고 항아리에 빨간색 공을 하나 넣은 것은 하나의 독립적인 사건이고 동전을 던져서 뒷면이 나오고 항아리에 빨간색 공과 초록색 공을 넣는 것도 하나의 독립적 사건이다. 다시 말해서 동전의 뒷면이 나와서 항아리에 빨간색 공을 넣는 것과 동전의 뒷면이 나와서 초록색 공을 넣은 것은 하나의 사건의 일부분을 구성하는 것이지 두 개의 별도의 사건이 아니다. 이것은 분리할 수 없는(indissociable) 사건의 일부이다. 따라서 빈도를 계산할 때, 이 둘을 따로 분리가능한 별도의 사건인 것처럼 생각해서는 안 되는데, 두 번째 풀이는 이 점에서 잘못을 범하고 있다는 것이다.[4)]

프란체스치는 위의 설명을 적용하여 엘가의 1/3주의 논변을 비판한다. 그에 따르면, T_1과 T_2는 분리할 수 없는 사건의 일부, 즉 하나의 사건을 구성하는 일부분인 반면, H_1은 별도의 하나의 사건이다. 그런 점에서 H_1과 T_1(또는 T_2)를 같은 유형의 사건으로 간주하여 빈도를 계산하는 것은 잘못이다. 그러므로 $P(H_1) = P(T_1 \vee T_2) = 1/2$이라는 것이 프란체스치의 주장이다.

그렇다면 프란체스치가 제시한 문제에서 두 번째 풀이는 과연 틀린 것인가? 동전 던지기를 한 번 할 경우 프란체스치의 주장처

4) Francheschi (2005), p. 2.

럼 P(R)의 값은 3/4이다. 그러나 두 번 던질 경우는, 세 번 던질 경우는, 그리고 n 번 던질 경우는 어떤가?

> 두 번 던질 경우: P(R) = 17/24
> 세 번 던질 경우: P(R) = 101/160
> n 번 던질 경우: P(R) = $(\frac{1}{2})$n{nC0 $\frac{n}{n}$ +nC1 $\frac{n}{n+1}$ + ... + nCk $\frac{n}{n+k}$ + ... +nCn $\frac{n}{n+n}$}

결국 이러한 시행이 무한히 반복될 경우, P(R)은 2/3에 수렴한다. 그러므로 프란체스치는 자신이 제시한 문제에 대해서 2/3이라고 답하는 것이 옳지 않고, 옳은 답은 3/4이라고 하기 위해서는 그의 물음이 보다 정확해져야 한다. 다시 말해서 동전을 한 번 던져서 위와 같은 방법으로 항아리를 채울 경우, 그 항아리에서 공을 꺼냈을 때 그 공이 빨간색일 확률이 얼마인가라고 물어야 한다. 이상의 설명이 주는 교훈은 동전을 던져서 앞면이 나오고 빨간색 공을 채우는 사건은 하나의 독립적인 사건이지만, 뒷면이 나와서 빨간색 공을 채우는 사건이나 뒷면이 나와서 초록색 공을 채우는 사건은 서로 분리할 수 없는 사건의 일부이고 따라서 빈도 계산을 할 때 달리 해야 한다는 그의 생각이 옳지 않다는 것이다. 항아리에서 공을 꺼내서 그 공이 빨간색 공일 확률을 계산하기 위해서는 여전히 앞에서 어떤 사건이 벌어졌던지 그 사건의 결과 항아리 안에 채워진 공의 전체 수에 대한 빨간색 공의 수라는 것은 여전히 동일하다.

이제 다른 경우를 살펴보자.

> 정상적인 동전을 던져서 앞면이 나오면 항아리에 한 개의 초록색 공을 넣고, 뒷면이 나오면 두 개의 빨간색 공을 넣는다. 이러한 작업을 n 번 계속했다. 이 항아리에서 공을 꺼냈을 때, 그 공이 초록색일 확률은 얼마인가?

이에 대한 일반적인 대답은 아마 1/3일 것이다. 왜냐하면 1/2의 확률로 빨간색 공이 두 개 채워지고, 또 1/2의 확률로 초록색 공이 하나 채워지므로, 항아리 안에는 빨간색 공은 초록색 공의 2배가 들어있을 것이기 때문이다. 그러나 프란체스치의 위의 계산 방법을 따르면 그 대답은 1/2이어야 한다. 가능한 사건은 "동전을 던져서 앞면이 나오고 항아리에 한 개의 초록색 공이 채워짐"과 "동전을 던져서 뒷면이 나오고 항아리에 두 개의 빨간색 공이 채워짐"이라는 두 개의 사건밖에 없을 것이고 각각의 확률은 1/2이기 때문이다. 정확하게 말해서 이 문제도 동전 던지기의 횟수에 따라서 그 답은 달라진다. 즉,

동전 던지기를 한 번 한 경우: P(G) = 1/2
동전 던지기를 두 번 한 경우: P(G) = 5/12
동전 던지기를 n 번 할 경우:

$$P(G) = (\frac{1}{2})n\{nC0\frac{0}{2n} + nC1\frac{1}{2n-1} + ... + nCk\frac{k}{2n-k} + ... + nCn\frac{n}{2n-n}\}$$

이러한 시행을 무한히 계속할 경우 P(G)는 1/3에 수렴한다. 이를 통해서 우리는 프란체스치가 자신이 제시한 문제에 대해서 그 답은 3/4이고 하나의 독립된 사건과 하나의 사건을 구성하는 분리할 수 없는 부분을 구별해야 한다는 그의 설명은 설득력이 없음을 알 수 있다. 오히려 그가 제시한 문제는 그 자체로 명확하지 않고 그 문제에 대한 답이 3/4이라고 하기 위해서는 동전 던지기를 한 번했다는 조건이 필요하고, 위 질문에 대한 일반적인 대답은 1/3이라고 해야 할 것이다.[5)]

---

5) 심사위원 중 한 분은 프란체스치에 대한 이러한 해석과 비판은 공정하지 않다고 주장하면서, 프란체스치에게 중요한 것은 "항아리에서 공을 한 번

프란체스치의 잘못을 다른 측면에서 설명할 수 있고, 그 설명은 위의 공꺼내기 사례와 SB의 사례 사이의 옳은 유비를 인식할 수 있게 해 줄 것이다. 위의 사례에서 프란체스치는 항아리에 공을 채우는 사건과 항아리에서 공을 꺼내는 사건 사이에 일종의 1:1 대응이 있고, 그 사건의 확률은 같다고 생각하는 것 같다.[6] 즉 '항아리에서 초록색 공을 넣는 사건'과 '항아리에서 초록색 공을 꺼내는 사건'의 발생 확률은 같다는 것이다. 즉 위의 사례에서 항아리에 초록색 공을 넣을 사건이 발생할 확률은 1/2이다. 왜냐하면 정상적인 동전을 던져서 뒷면이 나오면 초록색 공을 하나 넣는데, 정상적인 동전을 던져서 뒷면이 나올 확률은 1/2이기 때문이다. 프란체스치는 항아리에서 초록색 공을 꺼낼 확률도 마찬가지로 1/2이라고 주장하는 데 그것은 암암리에 두 사건이 분리될 수 없는 사건으로 보기 때문인 것 같다. 그러나 초록색 공을 항아리에 넣는 사건과 초록색 공을 항아리에서 꺼내는 사건은 두 개의 구별되는 사건이고, 따라서 그 사건의 확률은 반드시 같아야 하는 것이 아니다.

앞에서 설명한 것처럼, 위 사례에서 초록색 공을 꺼낼 확률은 1/3이고, 초록색 공을 넣을 확률은 1/2이다. 루이스나 프란체스치와

꺼내는가, 아니면 여러 번 꺼내는가의 문제"라고 말한다. 그리고 프란체스치는 자신이 제안한 사례에서 한 번 꺼내는 경우를 염두에 두고 있다고 해석해야 한다고 주장한다. SB의 경우에서 SB도 한 번 깨어나기 때문에 프란체스치의 사례에서 한 번 꺼내는 경우가 유비적이라는 것이다. 그러나 프란체스치는 항아리에서 공을 몇 번 꺼내는가에 대해서는 언급하지 않는다. 그가 주장하는 것은 동전을 던져서 뒷면이 나와서 빨간색 공과 초록색 공을 넣은 사건을 분리할 수 없는 하나의 사건으로 보느냐, 그렇지 않느냐이다. SB의 경우에서도 SB가 몇 번 깨워지는가가 논의의 초점이 아니라, 엘가처럼 $P(H_1) = P(T_1) = P(T_2) = 1/3$라고 보는지, 루이스처럼 $P(H_1) = P(T_1 \vee T_2) = 1/2$이라고 보느냐이다.

6) "동전을 던져서 앞면이 나오고 항아리에 한 개의 초록색 공이 채워짐"과 "동전을 던져서 뒷면이 나오고 항아리에 두 개의 빨간색 공이 채워짐"이라는 두 개의 사건밖에 없다는 그의 주장이 이러한 생각의 근거이다.

같은 1/2주의자들은 SB가 월요일에 동전이 앞면이라고 믿을 믿음의 정도를 위 사례에서 “항아리에 초록색 공을 넣을 확률”과 유비적이라고 생각하는 것으로 보인다. 즉 그들은 SB의 문제에서 SB에게 주어진 질문을 “동전 던지기가 발생했고, 이제 당신이 깨어났는데, 동전이 앞면이 나왔다고 믿을 합리적인 믿음의 정도가 무엇인가?”라고 해석하고 있는 셈이다. 그러나 SB의 문제에서 SB에게 물어진 질문은 동전의 본성에 대한 것이 아니라, 이번에 깨워짐이 앞면이 나와서 깨워짐이라고 믿을 합리적인 믿음의 정도에 대한 것이다. 따라서 SB에게 던져진 질문은, SB가 월요일에 깨워진 것이 그 전날 던진 동전이 앞면이 나와서 깨워진 것이라고 믿을 믿음 정도이고, 이는 위의 사례에서 항아리에서 초록색 공을 꺼내는 것과 유비적이라고 보아야 한다.7)

4. 보스트롬과 김한승의 절충주의에 대한 비판적 평가

보스트롬은 1/2주의와 1/3주의 모두 그럴 듯한 점을 갖지만 동시에 오류를 포함하고 있다고 주장한다. 다시 말해서 그는 1/2주의자인 루이스처럼 P(H) = 1/2라고 주장하면서, 동시에 1/3주의자인 엘가처럼 P+(H) = 1/2라고 주장한다는 점에서 자신의 견해를 절충주의(hybrid view)라고 부른다.8) 김한승은 최근 보스트롬의 견해를 토대로, 흥미로운 새로운 절충주의를 제안한다.

김한승은 1/2주의와 1/3주의 모두 나름의 타당성을 갖고 있기 때문에 SB의 문제는 진정한 역설이라고 보아야 한다고 주장하면서, 이 문제를 다른 방식으로 접근할 필요가 있다고 말한다. 그는 “동전이 앞면이었을 확률은 얼마인가?”라는 SB에게 주어진 질문을

7) 이 두 사례의 유비에 대한 설명은 Groisman (2009), p. 413을 참고할 것.

8) Bostrom (2003), p. 59.

1/2주의자와 1/3주의자는 서로 다르게 해석하고 있다고 주장한다. 다시 말해서 지금까지 'H'라고 쓴 '동전을 던져서 앞면이 나왔다'는 문장 내용이 비록 같은 문장이지만, 1/2주의자와 1/3주의자에게 다른 내용이라는 것이다. 이를 김한승은 외부적 관점과 내부적 관점이라는 용어로 설명한다.

> 미녀는 '동전이 앞면이었을 확률은 얼마인가?'라는 질문에 대해서 외부적 관점을 취할 수도 있고 내부적 관점을 취할 수도 있다. 내부적 관점을 취한다는 것은 자신이 깨어나는 논리적으로 가능한 세 가지 경우를 무차별적으로 본다는 것이다. 반면, 외부적 관점을 취한다는 것은 이 세 경우를 구별하여 본다는 것이다.[9)]

결국 SB의 문제에 대한 대답으로 1/2이 옳은지, 1/3이 옳은지는 외부적 관점과 내부적 관점 중 어떤 관점을 취하느냐에 달려 있다는 것이다. 그렇다면 이제 SB의 문제는 SB가 질문을 받았을 때, 외부적 관점을 취하는 것이 옳은가, 내부적 관점을 취하는 것이 옳은가의 문제가 될 것이다. 이에 대해서 김한승은 "잠에서 깨어나는 순간 잠자는 미녀가 내부적 관점을 취하는 것은 불가피하다(자연스럽다). (…) 잠에서 깨어남은, 잠자는 미녀로 하여금 내부적 관점에 자신이 서 있게끔 만든다고 할 수 있다."[10)]고 말한다. 그러므로 김한승의 주장은, SB의 문제에서 SB가 내부적 관점을 취하는 것이 자연스럽고 따라서 SB의 자연스러운 대답은 1/3이고, 다만 잠에서 깨어나는 경험에 대해서 SB가 외부적 관점을 취하는 것도 가능하고 그럴 경우 SB의 대답은 1/2일 것이라고 요약될 수 있다.

"동전을 던져서 앞면이 나왔다."는 문장이 관점에 따라 어떻게

---

9) 김한승 (2009), p. 133.

10) 김한승 (2009), p. 136. 김한승 선생님은 필자와 교환한 서신에서 '불가피하다'는 것은 '논리적으로 필연적'이라는 뜻이 아니라, '자연스럽다'는 뜻으로 해석해야 한다고 말하였다.

달리 해석될 수 있는가? 김한승은 “월요일에 깨어났다.”는 문장에 대해서 잠에서 깨어남이라는 경험이나 그러한 경험을 스스로 상상함(imagining oneself)을 통해서 내부적 관점을 취할 수 있을 것이다. 그러나 나의 어떤 경험, 혹은 어떤 경험에 대한 상상하기가 ‘동전이 앞면이 나옴’에 대해서 내부적 관점을 취하게 하는 것일까? 김한승은 SB가 잠에서 깨어나는 순간 내부적 관점을 취하는 것이 자연스럽다고 주장하는데, SB가 잠에서 깨어남이라는 경험이 어떻게 ‘H’에 대해서 다른 내용을 갖는 것으로 해석하게 하는지 설명되어야 한다. 김한승은 그것을 비개념적 내용으로서의 지표적 내용이라고 설명하고 이를 다시 “내부적 관점으로 파악된 내용은, 한 물체를 다른 물체와 구별하고 계속 추적할 수 있는 기본적 능력이 발휘된 결과 얻어진 것”이라고 설명한다. 그럼에도 여전히 SB가 잠에서 깨어나면 왜 내부적 관점을 취하는 것인지, 또 SB의 경우 어떤 물체를 다른 물체와 구별하여 지속적으로 추적할 수 있는 기본적 능력이 어떤 의미를 갖는지 불분명하다.

사실 1/2주의와 1/3주의 사이의 논쟁은 자신의 주장에 대한 근거만을 제시하는 것으로 그치지 않는다. 많은 경우 자신의 주장을 옹호하기 위해서, 상대방의 주장의 논리적인 문제점을 지적하기도 한다. 김한승의 절충주의가 성공적이려면, 김한승이 주장하는 것처럼 SB의 문제가 진짜 역설이어야 한다.[11] 즉 1/2주의의 논증이나 1/3주의의 논증에 아무런 결함이 없어야 한다. 그러나 필자의 루이스와 프란체스치의 논변에 대한 비판이 옳다면, 김한승의 절충주의가 토대로 하는 전제가 거부될 수밖에 없다. 나아가서 필자가 3절과 4절에서 제시한 주장이 옳지 않다고 하더라도, SB의 문제가 여전히

11) 김한승 선생님도 이 점을 분명히 인식하고 있다. 그래서 김한승 선생님은 SB의 문제를 확률에 관한 역설로 잘 알려진 “두 아이의 역설”과 같은 것으로 설명한다. 그러나 필자는 두 아이의 문제는 확률의 역설이라고 생각하지만, SB의 문제는 그렇지 않다고 생각한다.

진정한 역설이 아니라 1/2 또는 1/3이 유일한 답이라고 주장하는 사람들을 설득하기 위해서, SB의 문제가 진정한 역설임을 입증할 책임은 절충주의자에게 있다.[12)]

---

12) 그로이스만도 1/2주의자와 1/3주의자는 주어진 질문을 다르게 해석함으로써 다른 대답을 하고 있다고 주장한다. 즉 1/2주의자들은 'H'를 "동전 던지기가 이루어진 설정에서 동전이 앞면이 나왔다(The coin landed Heads *under the setup of coin tossing*)."라고 해석하고 대답하고 있는 것이고, 반면에 1/3주의자들은 "이번의 깨워짐은 깨워짐의 설정에서 앞면으로 인한 깨워짐이다(This awakening is a Head-awakening *under the setup of wakening*)."로 해석하고 답하고 있다는 것이다. (Groisman (2009), p. 414.) 그러나 필자는 SB의 진정한 문제는 원래 엘가가 제시했던 물음, "당신은 던져진 동전이 앞면이 나왔다고 어느 정도로 믿을 것인가?(To what degree ought you to believe that the outcome of the coin toss is Heads?)"를 어떻게 해석하는 것이 옳은지에 대한 것이라고 생각한다. 그로이스만도 김한승처럼 두 가지 설정 모두 해석 가능한 것으로 제안함으로써 일종의 절충주의를 채택하고 있는 셈이다. 그러나 필자는 그로이스만, 보스트롬, 김한승 모두 절충주의를 받아들이는 한, 동일한 입증부담을 지고 있다고 생각하고, 우리가 답해야 하는 것은, SB에게 주어진 질문에 대하여 "동전 던지기 설정"에 의한 해석과 "깨워짐의 설정"에 따른 해석, 혹은 내부적 관점에 의한 해석과 외부적 관점에 의한 해석 중에서 어느 것이 더 설득력 있는 해석인가라는 문제라고 생각한다.

심사위원 중 한 분은 절충주의가 SB의 문제가 '역설적'이라고 주장하고, 그 이유는 SB에게 주어진 질문이 애매하기 때문이라고 말할 수 있고, 그렇다면 절충주의자가 SB의 문제가 역설임을 입증하는 부담을 질 필요가 없다고 주장한다. 절충주의의 주장이 그런 정도라면, 필자도 절충주의자들이 SB의 문제가 역설임을 증명해야 한다고 주장하지 않을 것이다. 그리고 절충주의의 주장이 그런 정도라면, 절충주의는 일종의 양시론—1/2주의와 1/3주의 모두 옳을 수 있다는—이고, 필자가 이 논문을 통해서 1/2주의에 대한 비판에 대해서 답해야 할 것이다.

5. 맺는 말: 1/3주의를 옹호하며

지금까지 논의를 통해서 필자는 SB 문제에 대한 보다 설득력 있는 답은 1/3이라고 주장했다. SB 문제에 대해서 1/3이라고 답하는 것이 1/2이라고 답하는 것보다 직관적이다.[13] 또한 절충주의적 견해는 SB의 문제가 진정한 역설임을 전제하고 있는데, 이것은 1/2주의자나 1/3주의자가 보기에는 논점을 선취한 것처럼 보인다. 그런 점에서 절충주의가 설득력을 가지려면 먼저 SB의 문제가 진정한 역설임을 증명하는 거의 불가능해 가까운 작업을 성공적으로 마쳐야 한다.

최근 김명석은 “우리는 미녀에게 무엇을 물었을까?”라는 글에서 1/3주의를 주장하는 흥미로운 설명을 제시한다. 김명석은 SB에게 “주어진 질문이 통상적인 확률 물음이 아니”라고 주장하며[14], 그를 설명하기 위해서 다음과 같이 SB의 문제 상황을 제시한다.

A 교수가 시험 날 자정에 동전을 던져서 앞면이 나오면, “오늘 0시에 던진 동전은 앞면이 나왔는가?”라는 문제를 하나 만들어 상자에 넣고, 뒷면이 나오면 “오늘 0시에 던진 동전은 앞면이 나왔는가?”라는 문제를 두 개 만들어 상자에 넣는다. 전자의 질문에 답은 “예”이고, 후자의 질문에 답은 “아니오”이다. 그런 뜻에서 김명석은 전자의 문제를 ‘Y형 문제’, 후자를 ‘N형 문제’라고 한다. 물론 A 교수는 Y형 문제와 N형 문제를 구별할 수 있는 자신만 알아볼 수 있는 표시를 해 두었다. 이러한 출제 상황을 이해하고 당신은 상자 속의 문제를 꺼내 풀어야 한다. 그런데 한 개의 문제를 풀고 나면

13) 심사위원들은 필자가 1/3주의를 옹호하면서 이를 위해서 단순히 직관에 호소하고 있을 뿐이라고 말한다. 필자가 1/3주의가 1/2주의가 더 직관적이라고 말한 것은 엘가의 논변이 루이스의 논변보다 설득력이 있기 때문이고, 필자는 루이스에 대한 비판을 제시함으로써 이를 논증했다고 생각한다.

14) 김명석 (2009), p. 7.

그 문제를 풀었다는 사실을 기억에서 지우는 약을 먹고, (다음 문제가 남아 있다면) 다음 문제를 푼다. 당신이 풀어야 하는 문제는 하나이거나 두 개일 것이고, 하나라면 그것의 답은 ‘예’이고 두 개라면 그것의 답은 모두 ‘아니오’일 것이다. 이제 당신은 ‘오늘 0시에 던진 동전은 앞면이 나왔는가?’라는 문제를 받아들 것이다. 당신은 어느 정도로 이것이 Y형 문제라고 믿을 수 있는가?[15]

이 상황에서 김명석은 당신이 부딪히는 가능한 경우는 다음 3가지라고 말한다.

> (H_1) Y형 물음이 첫 질문으로 주어진다.
> (T_1) N형 물음이 첫 질문으로 주어진다.
> (T_2) N형 물음이 두 번째 질문으로 주어진다.

이에 대해서 김명석은 엘가처럼 이 세 경우가 모두 동일한 가능성을 갖는(equiprobable) 사건임을 인정하지만, 이 세 사건의 확률의 총합은 1이 아니라고 주장한다. 그 이유는 확률의 총합이 1이라는 것은 그 사건이 공시적인 경우이고, 통시적인 사건들의 확률의 총합은 1이 아닐 수 있으며, 위의 세 사건은 공시적이지 않다고 주장한다. 그러나 당신이 지금 부딪힐 수 있는 상황은 위의 세 가지 중 어느 하나임에 분명하고, 그 외에 다른 어떤 상황도 가능하지 않다. 다시 말해서 위 세 가지 상황은 서로 배타적이며(exclusive), 모든 상황을 다 포괄한다(exhaustive). 그런 의미에서 위의 세 가지 사건의 확률의 총합은 여전히 1이라고 해야 할 것이다. 김명석이 위 사건의 확률의 총합을 1.5라고 주장하지만, 그의 주장의 핵심은 위 세 가지 사건이 동일한 확률값을 갖는다는 것이다. 이 점에서 김명석은 1/3주의를 옹호하고 있는 셈이다.[16]

---

15) 김명석 (2009), p. 5.

16) 김명석은 논문의 첫 부분에서 “내가 잠자는 미녀이고 확률을 계산한 잠시의

끝으로 1/2주의보다 1/3주의가 보다 직관적이라는 주장을 설명하기 위해서 SB의 문제의 설정을 약간 수정해 보자. 일요일 자정에 동전을 던져서 앞면이 나오면 월요일 하루만 깨우고, 동전을 던져서 뒷면이 나오면 월요일부터 일요일까지 7번 깨운다고 하자. 물론 깨우고 나서 다시 재우기 전에 기억을 지우는 것은 동일하다. 이렇게 SB의 문제가 수정되었을 경우, SB가 월요일에 깨워졌을 때, 그가 동전이 앞면이 나왔을 확률에 대한 믿음의 정도는 얼마일까?

루이스와 프란체스치와 같은 1/2주의자들은 이에 대해서도 여전히 1/2이라고 답할 것이다. 그러나 이는 직관적으로 받아들이기 어렵다. 수정된 문제에서 SB가 월요일에 깨워졌을 때 그가 그 날이 월요일임을 알지 못하기 때문에, 그는 그 날이 7일 중의 어느 하루라고 생각하는 것이 자연스럽다. 그리고 SB가 깨워지는 경우는 8가지 중 어느 하나라고 판단할 것이고, 각각의 가능성 사이에 차이가 있다고 생각해야 할 아무런 근거도 없다. 그런 점에서 수정된 설정에서 SB의 합리적인 대답은 1/8이이어야 할 것이다.

시간이 주어진다면, (…) 1/3이라고 대답할 것"이라고 말한다. 김명석 (2009), p. 1.

참고 문헌

김명석 (미게재), “우리는 미녀에게 무엇을 물었을까?”, 2009년 논리학회 발표문

김한승 (2009), “비개념적 내용으로서의 지표적 내용: ‘잠자는 미녀’ 문제에 대한 관점주의적 대답”, 『철학적 분석』 20호, pp. 119-140.

Bostrom, Nick (2003), “The Mysteries of Self-locating Belief and Anthropic Reasoning”, *Harvard Review of Philosophy* 11, pp. 59-74.

Elga, Adams (2000), “Self-locating Belief and the Sleeping Beauty Problem” *Analysis* 60, pp. 143-147.

Franceschi, Paul (2005), “Sleeping Beauty and the Problem of W orld Reduction” in http://philsci-archive.pitt.edu/2175/1/sb-en.pdf

Groisman, Berry (2008), “The End of Sleeping Beauty's Nightmare,” *The British Journal for the Philosophy of Science* 59, no. 3. pp. 409-416.

Kim, Hanseung (2010), “Sleeping Beauty's Reflection: In and Out”, 『논리연구』 13집 1호, pp. 21-52.

Lewis, David (2001), “Sleeping Beauty: Reply to Elga”, *Analysis* 61, pp. 171-176.

아주대학교 기초교육대학 교수
Ajou University, University College
E-mail: song1959@hanmail.net
hasong@ajou.ac.kr

What is the Correct Answer to the Sleeping Beauty Problem?

Hasuk Song

I take the position of the thirders on the sleeping beauty problem like Elga and criticize Lewisian halfers. In particular, I attack Franceschi's recent arguments for the halfers. In addition, I claim that Bostrom's and Kim's hybrid view is not satisfactory, because it is to pre-empt or to take the burden of proof that the problem is the genuine paradox. Consequently, the purpose of this paper is to show that the thirders' argument is more intuitive than others and what the fallacies of the halfer's arguments are.

Key Words: Sleeping Beauty problem, Probability, Elga, Lewis, Hybrid view