2009년 2월 6일 논리학회 | 김명석 이화여대 oxys@paran.com

우리는 미녀에게 무엇을 물었을까?

만일 내가 잠자는 미녀이고 확률을 계산할 잠시의 시간이 주어진다면 나는 "동전이 앞면이 나왔을 확률은 얼마인가?"에 대해 1/3이라고 대답할 것이다. 이것은 다수의 견해이다. 여기서 이 다수 견해를 색다르게 정당화하는 묘안을 제시할 수는 없다. 다만 이 퍼즐과 관련된 몇몇 혼동들의 원인을 해명하고자 한다.[1)]

미정 사건의 예측과 기정 사건의 사후추측

미래에 벌어질 일에 대한 예측(prediction)과 이미 벌어졌지만 자신이 모르는 일에 대한 사후추측(retrodiction, postdiction, 역측)은 다르다. 앞으로 일어날 사건의 확률과 이미 일어난 사건의 확률은 대부분의 경우 같겠지만 다를 수 있다. 그래서 잠자는 미녀에게 던진 질문은 "동전이 앞면이 **나왔을** 확률은 얼마인가?"와 "동전이 앞면이 **나올** 확률은 얼마인가?" 중에서 어느 하나여야 하지 둘의 모호한 혼합이어서는 안 된다. 동전을 이미 던졌는지 (이미 앞면과 뒷면이 결정되었는지) 아직 던지지 않았는지를 미녀가 모를 경우 "동전이 앞면이 나왔는가?" 또는 "동전이 앞면이 나올 것인가?"에 적절히 답변할 수 없다. 질문에 답하기 전에 미녀는 동전을 이미 던졌는지 아직 던지지 않았는지 먼저 확인해야 한다.

왜 우리는 동전이 앞면이 **나올** 확률이 1/2이라고 생각하는가? 그것은 던져질 동전이 앞면과 뒷면 중 어느 면이 나올지 우리가 예측하는 것이 완전히 불가능하고 또한 앞면과 뒷면 중 특정 면을 선호하도록 하는 모종의 요인가 없다는 것을 우리가 알기 때문이다. 1/2이라는 확률은 바로 그런 것이다. 두 사건을 좌우할 요인에 대한 완전한 무지를 표현하는 수이다. 미래 사실에 대한 예측은 해당 사건을 야기하는 데 인과적으로 개입하는 다른 사건들에 대한 정보에 의존한다. 동전 던지기의 경우 우리는 그 정보를 전혀 가지고 있지 않다. 따라서 동전이 앞면이 **나올** 주관적 확률은 우리에게 신통력이 없는 한 언제나 1/2이다.

반면 과거 사실에 대한 사후추측은 통상적인 미래 예측뿐만 아니라 이미 일어난 사건들과 현재 사건 사이의 인과적 연결에도 의존한다. 후자에 대한 정보가 없을 때 사후추측 확률은 예측의 통상적 확률 값과 동일하다. 이미 던져진 동전이 어떤 면이 **나왔는지** 우리가 완전하게 무지하며, 동시에 통상적으로 앞면이 **나올** 확률이 1/2이라는 것을 알고 있을 때, 우리는 앞면이 **나왔을** 확률에 1/2을 주관적으로 부여하게 된다. 그러나 비록 던져질 동전이 어느 면이 **나올지** 우리가 예측하는 것이 완전히 불

---

1) 발표자는 잠자는 미녀 문제를 순진하게 접근하기 위해 다음 두 논문만 참조했다. Adam Elga 2000, "Self-locating Belief and the Sleeping Beauty Problem", *Analysis* **60**: 143~147. David Lewis 2001, "Sleeping Beauty: Reply to Elga", *Analysis* **61**: 171~176.

가능하다 하더라도 이미 던져진 동전이 어떤 면이 **나왔는지** 완전히 무지하지는 않을 수 있다. 가령 내가 동전이 앞면이 나왔던 것을 보았다면 그 동전이 앞면이 **나왔을** 확률은 나에게 주관적으로 1이다. 이러한 인과적 지각 경험처럼 만일 이미 일어난 사건들과 현재 사건 사이의 인과적 연결이 감지된다면, 두 확률은 달라질 수 있다. 더구나 아직 실현되지 않은 미래로부터 정보가 전송될 수 없지만, 이미 실현된 사건으로부터 현재 우리에게 정보가 전송될 수 있다. 따라서 통상적으로 앞면이 **나올** 확률이 1/2이라는 것을 우리가 알고 있다 하더라도, 또한 비록 우리에게 신통력이 없다 하더라도, 동전이 앞면이 **나왔을** 주관적 확률은 1/2이 아닐 수 있다.

엘가와 루이스는 동전을 던지는 시점이 "동전이 앞면이 나왔는가?"에 대한 미녀의 주관적 확률에 영향을 끼치지 않는다고 주장한다. 만일 미녀가 동전이 아직 던져지지 않았다고 생각한다면, 미녀는 동전이 앞면이 나올 확률이 1/2라고 언제나 대답할 것이다. 비록 동전을 던진 시점이 미녀를 깨우기 전이라 하더라도, 미녀가 동전의 앞면과 뒷면이 아직 결정되지 않았다고 믿는다면, 그녀의 답변은 역시 언제나 1/2이다. 그러나 만일 그녀가 동전의 앞면과 뒷면이 이미 결정되었다고 생각한다면, 그녀는 동전이 앞면이 나올 확률이 1/3라고 대답할 것이다. 비록 동전을 던지는 시점이 미녀를 깨운 후라 하더라도, 미녀가 동전의 앞면과 뒷면이 이미 결정되었다고 믿는다면, 그녀의 답변은 역시 1/3이다.

출제 정보의 유출

A. 미르 교수는 모든 문장의 진리값이 우연성에 의해 결정된다고 믿고 있다. 물론 같은 문장의 진리값은 같다. 그는 학생들에게 항상 OX 퀴즈를 내는데 학생들은 주어진 문장이 참이면 O 거짓이면 X를 해야 한다. 학생은 상자에 들어 있는 문제들 중에서 하나를 선택한다. 미녀 학생 지나는 상자에서 문제 하나를 꺼내었다. 그녀는 미르 교수의 문제 출제 방식을 전혀 모르고 있으며 자신이 선택한 문제의 정답을 결정할 (그 문장의 진리값을 결정할) 그 어떠한 지식도 없었다. 지나는 이 문제의 정답이 O일 것이라고 어느 정도로 강하게 확신할 수 있을까? 아마도 1/2. 왜냐하면 지나는 O와 X 중에서 어느 하나를 선호할 그 어떤 정보도 가지고 있지 않기 때문이다.

B. 미르 교수는 다음 같은 방식으로 OX 문제를 출제한다. 그는 시험 전날 동전을 던진다. 앞면이 1이 나오면 정답이 O인 문제를 하나 출제하고 문제 상자 속에 넣는다. 뒷면이 나오면 정답이 X인 문제 **두 개**를 출제하여 상자 속에 넣는다. **이 과정을 반복하여** 수많은 문제들의 은행을 만들어낸다. 지나는 상자 속에 **오직 한 문제**를 꺼내어 문제에 답한다. 지나는 미르 교수의 이런 출제 방식을 알게 되었다. 이 정보는 그녀의 정답률을 높이는 데 기여할 수 있을까?

지나는 주어진 문제에 O를 하는 것이 유리할까 X를 하는 것이 유리할까? 지나가 그 문장의 진리값에 대해 알고 있는 유일한 정보는 미르 교수의 그 출제 방식이다. (물론 통상적 동전 던지기 확률에 대해서는 알고 있다.) 지나는 그 정보로부터 문제 은행의 문제들 중 1/3의 문제들은 그 답이 O인 문제이고 2/3의 문제들은 그 답이 X인 문제라는 것을 추론하게 될 것이다. 자신이 문제 은행에서 선택한 그 문제가 정답 O인 문제일 확률은 1/3이다. 따라서 자신이 선택한 그 문제의 정답이 O일 확률은 1/3이며 그녀는 정답률을 높이기 위해 X라고 답할 것이다.

미르 교수가 생산한 문장의 진리값을 결정하는 것은 동전이 앞면 또는 뒷면이 나올 확률이다. 문

장의 진리값과 동전 던지기 사이의 이러한 연결은 미르 교수의 문제 제작 행위에 의해 인위적으로 생성되었다. 지나가 미르 교수의 출제 방식에 대한 정보를 입수했을 때 그녀는 미르 교수에 의해 인위적으로 생성된 이러한 인과적 연결을 감지하게 된다. 출제 방식에 대한 정보는 그녀에게 주어진 문장의 진리값에 대한 주관적 확신을 변화시켰다.

C. 미르 교수의 동료 아루치 교수는 학생들이 과거에 이미 발생한 일을 추측하게 하는 문제를 늘 출제한다. 그는 다음과 같은 방식으로 Y/N 문제를 출제한다. 시험 날 자정에 동전을 던져 앞면이 나오면 "오늘 0시에 던진 동전은 앞면이 나왔는가?"라는 문제를 출제하여 문제 상자 속에 넣는다. 이 문제의 정답은 Y이다. 이 문제를 'Y형 문제'라 한다. 자정에 동전을 던졌는데 뒷면이 나오면 역시 "오늘 0시에 던진 동전은 앞면이 나왔는가?"라는 문제 **두 개**를 출제한다. 이 문제의 정답은 N이고 'N형 문제'라 한다. 아루치 교수는 자신만 구별할 수 있는 시험지에 이 문제를 기록하여 문제 상자 속에 넣는다. **이 과정을 반복하여** 수많은 문제들의 은행을 만들어낸다.

지나는 아루치 교수의 이 출제 방식을 알고 있다. 이것은 그녀가 그 문제 문장의 진리값을 결정하는 데 동원할 수 있는 유일한 정보이다. 그는 상자 속에 **오직 한 문제**를 꺼내어 문제에 답한다. 지나는 상자에서 꺼낸 "오늘 0시에 던진 동전은 앞면이 나왔는가?"라는 문제의 답이 Y일 것이라고 어느 정도로 강하게 믿을 수 있을까? 지나는 문제 은행의 문제들 중 1/3의 문제들은 Y형 문제이고 나머지 2/3의 문제들은 N형 문제라는 것을 알고 있다. 그녀에게 주어진 그 물음이 Y형 문제일 확률은 1/3이며 따라서 그녀는 자신이 선택한 그 물음이 Y일 것이라고 1/3 정도로 믿게 될 것이다. 즉 오늘 0시에 던진 동전이 앞면이 **나왔을** 주관적 확률은 그녀에게 1/3이다.

이런 믿음을 위해 그녀가 사용한 모든 정보는 아루치 교수의 출제 방식이다. 아루치 교수의 출제 방식은 동전이 앞면 또는 뒷면이 나오는 사건이 문제 상사 속 문제 유형의 분포를 결정하는 효과를 주게 되어 있다. 다시 말해 동전 던지기 사건이 문제 은행의 성격과 인과적으로 연결되었다. 이러한 인위적 인과적 연결에 대한 지나의 정보가 자신의 추측에 동원될 경우 이 정보는 그녀의 주관적 확률을 변화시킬 것이다.

지나가 "오늘 0시에 던진 동전은 앞면이 나왔는가?"에 답하는 것은 미래에 일어날 일에 대한 개연적 예측이 아니다. 만일 그 물음이 예측이라면 그 물음에 대해 1/2의 강도로 Y라 답해야 한다. 다시 말해 동전이 앞면이 **나올** 확률은 1/2이다. 지나에게 주어진 그 물음은 과거에 이미 벌어진 일에 대한 사후추측이다. 여기서 주의해야 할 것이다. 그녀에게 주어진 "오늘 0시에 던진 동전은 앞면이 나왔는가?"에 답하는 것은 이 물음이 Y형 문제인가 N형 문제인가를 답하는 것과 같다. 그 물음에 Y라고 답하는 것은 그 물음이 Y형 문제라고 보는 것이다. 따라서 지나의 주관적 확률은 자신에게 주어진 물음이 Y형일 확률이다. 이 확률은 문제 상자 속에서 Y형 문제와 N형 문제의 분포에 의해 결정된다. 그리고 이 분포를 결정하는 것은 두 요소이다.

(i) 멀쩡한 동전이 앞면 또는 뒷면이 **나올** 통상적 확률.

(ii) 앞면 또는 뒷면이 **나왔을** 때 벌어지는 일들.

아루치 교수의 의도적 행위에 의해 앞면은 Y형 문제 하나를 창출하고 뒷면은 N형 문제를 두 개 창출한다. 이로 인해 "오늘 0시에 던진 동전은 앞면이 나왔는가?"의 답이 Y일 확률이 1/2가 아니라 1/3이 되었다. 동전 던지기 사건과 물음 사건은 원래 자연적으로 아무런 상관이 없지만 아루치 교수는 두

사건을 인위적으로 연결시켰다. 이러한 연결은 "오늘 0시에 던진 동전은 앞면이 나왔는가?"라는 물음의 답 자체를 변화시킨다.[2]

잠자는 미녀

D. 아루치 교수는 일요일에 동전을 던져 앞면이 나오면 월요일 검사할 문제로서 "던진 동전이 앞면이 나왔는가?"라는 Y형 문제를 한 개만 만든다. 동전이 뒷면이 나오면 월요일과 화요일에 각각 검사할 문제로서 "던진 동전이 앞면이 나왔는가?"라는 N형 문제를 **두 개** 준비한다. 지나는 잠자는 미녀처럼 약을 복용한 채 시험에 응시해야 한다. 아루치 교수는 일요일에 동전을 던지고 그 결과에 따라 월요일 또는 화요일에 잠자는 지나를 깨워 묻는다. "던진 동전이 앞면이 나왔는가?"[3] 지나는 자신에게 주어진 물음에 대해 Y라고 어느 정도 강하게 믿게 될까? 물론 지나는 Y형 문제와 N형 문제를 구별할 아루치 교수만의 표식을 알지 못한다. 또한 질문을 받는 시점이 월요일인지 화요일이지 전혀 알지 못한다.

경우 D는 경우 C와 다르다는 것을 인지해야 한다. C에서 지나는 많은 문제들의 상자 중에서 한 물음에 답한다. 그 상자에 들어 있는 문제들 중 1/3은 Y형 문제이고 2/3은 N형 문제이다. 그러나 D에서 지나가 풀어야 하는 문제는 상사 속에 들어 있는 수많은 문제들 중 하나가 아니다. 아루치 교수에게 준비된 물음의 수는 하나이거나 둘이다.

E. 마루치 교수는 다음 같은 방식으로 Y/N 문제를 출제한다. 시험 날 자정에 동전을 던져 앞면이 나오면 "오늘 0시에 던진 동전은 앞면이 나왔는가?"라는 Y형 문제를 출제하여 문제 상자 속에 넣는다. 뒷면이 나오면 "오늘 0시에 던진 동전은 앞면이 나왔는가?"라는 N형 문제 **두 개**를 출제한다. 마 교수는 Y형 문제와 N형 문제를 구별하기 위해 자신만의 표식을 해두었다. 그는 이 과정은 **반복하지 않고 단 한 번만** 시행한다.

상자 속에 문제는 하나가 들어 있거나 아니면 두 개 들어 있다. 만일 동전이 앞면이 나오면 상자 속에는 Y형 문제가 하나 들어 있다. 만일 동전이 뒷면이 나오면 상자 속에 N형 문제가 두 개 들어 있다. 따라서 상자 속에 Y형 문제만 들어 있을 확률은 1/2이다. 상자 속에 N형 문제만 들어 있을 확률은 1/2이며 나머지 경우는 0이다.

이제 지나는 상자 속에서 **오직 한 문제**를 꺼내게 된다. 그 문제는 지나에게 "오늘 0시에 던진 동전은 앞면이 나왔는가?"라고 묻고 있다. 그녀는 그 물음의 답이 Y일 것이라고 어느 정도로 강하게 믿

2) 이것은 "우리가 살고 있는 이 우주가 고도로 질서 잡힌 우주일 확률은?"이라 묻는 상황과 비슷하다. 임의의 빅뱅 이후 생길 임의의 우주가 고도로 질서 잡힐 가능성은 매우 낮다. 그러나 우리가 살고 있는 우주는 이미 벌어진 우주이다. "우리가 살고 있는 이 우주가 고도로 질서 잡힌 우주일 확률은?"은 장차 벌어질 일에 대한 예측을 요구하는 것이 아니다. **이 물음을 물을 수 있는 우주**는 극도로 제한되어 있다. 다시 말해 그런 물음이 주어질 수 있는 우주는 가능한 우주들 중에 극히 일부에 해당한다. 따라서 우리에게 그 물음이 주어졌다면 그것은 이미 우리가 살고 있는 우주가 매우 특수한 유형의 우주라는 것을 함축한다. "우리가 살고 있는 이 우주가 고도로 질서 **잡힌** 우주일 확률은?"은 이 물음이 지금 해당 우주에서 벌어지고 있다는 바로 그 정보에 의해서 1에 근접한 값이다. 이 점에서 우리가 빅뱅 이후에 **생긴** 고도로 질서 **잡힌** 우주에 이미 살고 있는 것은 그다지 놀라운 일이 아니다. 이러한 '인간중심원리'는 임의의 빅뱅 이후 **생길** 임의의 우주가 고도로 질서 **잡힐** 가능성이 여전히 몹시도 매우 낮다는 사실을 잠시 잊게 한다.

3) 아루치 교수가 지나에게 던진 물음은 사실은 다음과 같다. "당신은 던진 동전이 앞면이 나왔다고 얼마만큼 믿는가?"(To what degree ought you believe that the outcome of the coin toss is Heads?) Elga 2000, p. 143.

을 수 있을까? 지나는 마루치 교수의 출제 방식을 알고 있으며 이 방식은 그 물음의 답을 결정하기 위해 그녀가 알고 있는 유일한 정보이다. 물론 지나는 상자 속에 들어 있는 문제의 수를 처음부터 끝까지 알지 못한다고 가정한다.

그녀는 자신이 문제를 뽑을 문제 상자가 오직 Y형 문제들만 들어있는 상자일 확률이 1/2이라는 것을 안다. 나머지 확률은 오직 N형 문제들만 들어 있는 상자에서 문제를 뽑게 될 확률이다. 이로부터 그는 자신에게 주어진 **그 물음이 Y형 문제일 확률**이 1/2라는 것 또한 알게 된다. 따라서 그녀는 **그 물음의 답이 Y일 확률**이 1/2이라고 생각할 것이다. 지나는 "오늘 0시에 던진 동전이 앞면이 나왔는가?"라는 물음에 대해 1/2의 확률로 Y라고 답해야 한다. 다시 말해 오늘 0시에 던진 동전이 앞면이 나왔을 그녀의 주관적 확률은 1/2이다. 1/3이 아니다! 이런 믿음을 위해 그녀가 사용한 모든 정보는 마루치 교수의 출제 방식이다.

이제 경우 E가 잠자는 미녀 경우 D와 동일한 경우인지 살펴보자. D에서는 동전이 어떤 면이 나오느냐에 따라 Y형 문제 하나 아니면 N형 문제 두 개가 생성된다. E에서도 마찬가지이다. 그런데 E에서 지나는 언제나 오직 하나의 문제만 받게 되지만 D에서는 한 번 아니면 두 번 질문을 받게 된다. 따라서 D의 경우는 E의 경우와 완전히 유사한 것은 아니다. D에서 지나는 결과적으로 아루치 교수가 만든 모든 문제를 풀게 되어 있다. 따라서 다음과 같은 경우가 D와 유사한 경우라 하겠다.

단일 문제로 위장된 중복 문제

F. 마루치 교수의 새로운 출제 방식은 다음과 같다. 시험 날 자정에 동전을 던져 앞면이 나오면 "오늘 0시에 던진 동전은 앞면이 나왔는가?"라는 Y형 문제를 하나 출제하여 문제 상자 속에 넣는다. 뒷면이 나오면 N형 문제를 두 개 출제하여 상자 속에 넣는다. 지나는 상자 속에서 "오늘 0시에 던진 동전은 앞면이 나왔는가?"라는 물음을 꺼내게 된다. 그런데 지나는 상자 속의 **모든** 문제를 풀어야 한다. 지나는 상자 속에 문제가 몇 개 있었는지, 몇 개 남았는지 알지 못한다. 상자 속에 문제가 남아 있을 경우 지나는 약을 먹고 이전에 자신이 문제를 풀었다는 사실을 망각한 채 나머지 문제를 풀어야 한다. 이제 지나가 받아본 "오늘 0시에 던진 동전은 앞면이 나왔는가?"라는 물음의 답이 Y일 것이라고 어느 정도의 개연성을 가지고 믿어야 할까?

자정이 지난 후 상자 속에는 들어 있는 문제의 수는 하나 아니면 둘이다. 지나는 마루치 교수의 출제 방식을 알고 있다. 이 정보를 '알파 정보'라 하자. 그러나 지나가 알고 있는 것은 이뿐만 아니다. 동전이 앞면이 나올 경우 시험을 한 번만 치면 되지만 뒷면이 나올 경우 재시까지 쳐야 한다는 것까지 그녀는 알고 있다. 그리고 그 재시는 초시와 동일한 문제이다. 이 정보를 '베타 정보'라 하자.

이제 지나에게 문제 상자가 주어졌다. 이 문제 상자에 오직 Y형 문제만 들어 있을 확률은 1/2이다. 즉 지나가 상자에서 꺼낸 "오늘 0시에 던진 동전은 앞면이 나왔는가?"라는 물음이 Y형 문제일 확률은 1/2이다. 따라서 이 물음의 답이 Y일 확률은 그녀에게 1/2이다. 그런데 여기서 지나는 베타 정보를 전혀 이용하지 않았다. 알파 정보는 상자 속 문제 유형의 분포를 알려준다. 반면 베타 정보는 자신이 답해야 하는 시험 응시의 빈도를 알려준다. 베타 정보는 그녀가 복수의 동일한 물음에 답할 가능성이 있음을 알려준다.

지나는 자신에게 주어진 "오늘 0시에 던진 동전은 앞면이 나왔는가?"라는 물음에 대해 Y라고 답한

다고 가정해 보자. 만일 지나가 시험을 한 번만 본다면 그녀는 그 물음에서 정답을 맞힌 것이 된다. 시험을 한 번만 볼 확률이 1/2이기 때문에 1/2의 확률로 100점을 얻게 된다. (정답을 말했을 때 100점을 부여하고 틀렸을 때 0점을 부여한다.) 그녀가 동일한 시험을 두 번 본다면 그는 두 번의 물음에서 틀리게 된다. 시험을 두 번 볼 확률이 1/2이기 때문에 1/2의 확률로 두 번의 시험에서 각각 0점을 얻게 된다. 정리하면 1/2의 확률로 한 번 시험에 100점을 얻거나 1/2의 확률로 두 번의 시험에서 각각 0점을 얻는다. 각 경우에 가중치를 0.5로 주면 주어진 물음들에 **모두** Y라고 답할 때 지나는 총점 50점을 얻게 된다.[4]

지나가 "오늘 0시에 던진 동전은 앞면이 나왔는가?"라는 물음에 대해 N라고 답한다고 가정해 보자. 만일 지나가 시험을 한 번만 본다면 그는 그 시험에서 0점을 얻게 된다. 그녀가 동일한 시험을 두 번 본다면 그는 두 번의 시험에서 각각 100점을 얻는다. 그녀는 1/2의 확률로 한 번 시험에 0점을 얻고 1/2의 확률로 두 번의 시험에서 각각 100점을 얻는다. 각 경우에 가중치를 0.5로 주면, 주어진 물음들에 **모두** N이라고 답할 때 지나는 총점 100점을 얻게 된다.[5] 이처럼 지나는 주어진 **모든** 물음들에 **모두** N이라 답하면 Y라고 답한 경우보다 두 배의 높은 점수를 얻게 된다. 따라서 지나가 "오늘 0시에 던진 동전은 앞면이 나왔는가?"라는 **질문에 한 번 이상 답해야 한다면**, Y보다 N이라고 재차 답하는 것이 두 배 더 유리하다.

지나가 **통틀어서 오직 하나의 문제만 풀어야 한다면**, "오늘 0시에 던진 동전은 앞면이 나왔는가?"의 답이 Y일 확률은 1/2이다. 그런데 지나는 1/2의 확률로 한 개의 문제를 풀고, 1/2의 확률로 두 개의 동일한 문제를 풀어야 한다. 지나가 풀어야 하는 문제는 평균적으로 1.5개이다. 경우 F의 문제 설정은 지나에게 하나의 물음이 던져진 것이 아니다. 경우 F는 지나가 150점짜리 문제를 푸는 상황과 같다. 그녀가 어떻게 답하느냐에 따라 50점을 얻을 수 있고 100점을 얻을 수 있다. Y라 답할 경우 총점의 1/3만큼 점수를 얻게 되고 N이라 답할 경우 총점의 2/3만큼 점수를 얻게 된다. 망각제 덕택에 "오늘 0시에 던진 동전은 앞면이 나왔는가?"와 "오늘 0시에 던진 동전은 앞면이 나왔는가?"가 동시에 물어졌다는 것을 모른 채 이 둘을 **단일 물음으로 알아듣고** Y 또는 N이라 답한다. 그러나 미녀가 충분히 현명하다면 자신이 답변이 그 복수 물음에 동일하게 답변하는 것이라는 것을 자각할 것이다.

잠자는 미녀 경우는 경우 F와 유사하다. 잠자는 미녀에게 던진 "동전이 앞면이 나왔다고 얼마만큼 믿는가?"라는 질문은 사실 단수의 질문이 아니다. 잠자는 미녀가 이 질문에 답변을 한다는 것은 복수의 물음들에 동일한 답변을 여러 번 하는 것에 해당된다. 그녀는 복수의 질문들에 대해 어떤 대답으로 줄곧 답변하는 것이 더 적절한지 가늠해야 한다.

총확률 1.5

기묘한 실험 상황에서 미녀에게 던져진 "일요일에 던진 동전은 앞면이 나왔는가?"라는 물음의 총점은 150점(1.5개짜리 문제)이다. 일반적으로 어떤 문장이 옳을 확률은 최대 1이며 우리는 그것이 옳을 때 100점을 부여한다. "일요일에 던진 동전이 앞면이 나왔다"가 옳을 때 부여하는 점수가 150점이라는 사실은 이것의 최대 확률이 1.5라는 것을 뜻한다. 어떻게 이것이 가능한가?

---

4) 0.5*100 + 0.5*(0 + 0) = 50.
5) 0.5*0 + 0.5*(100 + 100) = 100.

이것은 미녀에게 주어진 물음이 통상적인 확률 물음이 아니라는 것을 뜻한다. 이 물음은 동전이 앞면이 나왔을 확률을 묻는 것이 아니다. 다시 말해 이것은 "일요일에 던진 동전은 앞면이 나왔다"라는 문장의 진리값을 개연적으로 사후추측하는 것이 아니다. 이 물음은 미녀 자신에게 던져진 "일요일에 던진 동전은 앞면이 나왔는가?"라는 물음이 Y형 문제일 확률을 묻는 것이다. 즉 "일요일에 던진 동전은 앞면이 나왔는가?"라는 구체적 발화 또는 구체적 기재가 미녀 자신의 세계에 출현할 가능성을 사후추측하는 것이다. 개별자로서 그 발화 또는 기재는 Y형과 N형이 있으며 이들은 제각기 특정 시공간에서 출현한다. Y형 사례와 N형 사례를 구별할 수 있는 표지가 실제로 있지만 미녀를 이를 구별하지 못한다.

미녀에게 출현할 가능한 물음 사건은 모두 3가지이다.

(H1) Y형 물음이 월요일에 주어진다.

(T1) N형 물음이 월요일에 주어진다.

(T2) N형 물음이 화요일에 주어진다.

그런데 H1이 출현할 확률은 1/3이 아니라 1/2이다. T1이 출현할 확률 또한 1/3이 아니라 1/2이다. T2가 출현할 확률도 1/2이다. 이것은 엘가와 루이스와 나의 차이이다.

	엘가	루이스	나
P(H1)	1/3	1/2	1/2
P(T1)	1/3	1/4	1/2
P(T2)	1/3	1/4	1/2
P(H1)+P(T1)+P(T2)	1	1	1.5

여기서 P는 월요일에 미녀가 깨어났을 때 그녀가 가지는 주관적 확률이다.

원칙적으로 총 확률은 다음과 같이 계산해야 한다. 먼저 특정 시점에 나올 만한(나왔을 만한) 모든 사건들의 목록을 작성한다. 총 확률이 1이라는 것은 그 사건들 각각의 확률을 더했을 때 1이 나온다는 것이다. 왜냐하면 특정 시점에 실제로 등장하는 것은 그 가능한 사건들 목록 중에서 하나여야 하기 때문이다. 그런데 특정 시점에 그 사건들 중에서 2개가 실제로 등장한다고 해보자. 이 경우 총 확률은 1을 넘어설 것이다. 우리는 주사위를 던졌을 때 오직 하나의 눈만 나온다고 가정한다. 그리하여 각 눈이 나올 확률은 1/6이다. 그러나 주사위를 던졌을 때 두 개의 눈이 나온다고 가정해 보자. 그러면 각 눈이 나올 확률은 1/3이 될 것이다. 그리고 각 눈이 나올 확률의 총합은 2이다.

확률의 총합은 공시적이어야 한다. 확률을 통시적으로 더했을 때는 1이 아닐 수 있다. H1이 발생하는 시점은 월요일 어느 시점이다. T1이 발생한 시점도 월요일 어느 시점이다. H1이 발생하면 T1이 발생하지 않고, H1이 발생하지 않으면 T1이 발생한다. 그리하여 월요일 마치기 전 어느 특정 시점에서 H1이 발생할 확률과 T1이 발생할 확률은 더하면 1이 되어야 한다. 한편 T2가 발생하는 시점은 화요일 어느 시점이다. T2가 발생하면 H1이 발생하지 않고 **또한** T1이 발생하지 않는가?[6] 아니다. 이것은 P(H1)+P(T1)+P(T2)가 1이 아님을 의미한다. T2가 발생하면 발생하지 않는 사건이 따로 있다. 그것은

6) 표본공간에 오직 e1, e2, e3만이 존재하고 이들이 발생할 확률들의 총합이 1이 되기 위해서는 다음을 만족해야 한다. 만일 e1이 발생하면 e2도 발생하지 않고 e3도 발생하지 않아야 한다. 그리고 만일 e2가 발생하면 e1도 발생하지 않고 e3도 발생하지 않아야 한다. 그리고 만일 e3이 발생하면 e1도 발생하지 않고 e2도 발생하지 않아야 한다.

화요일에 질문이 없는 사건이다. 그것을 H2라고 할 수 있다. T2가 발생하면 H2가 발생하지 않고, T2가 발생하지 않으면 H2가 발생한다. 따라서 총 확률은 다음과 같이 계산해야 한다.

P(H1) + P(T1) = 1

P(H2) + P(T2) = 1

두 합산은 각각 공시적이다. 반면 엘가와 루이스는 확률들을 시간을 걸쳐서 합산했다.

우리가 미녀에게 실제로 물은 것이 일요일에 던진 동전이 앞면이 나왔을 확률이 아니다. 그녀에게 물은 것은 "일요일에 던진 동전은 앞면이 나왔는가"라는 물음이 Y형 문제였을 확률이다. 다시 말해 동전 던지기 사건의 확률이 아니라 물음 묻기 사건의 확률이다. 동전 던지기의 결과가 **한 시점에** 여러 가능성 중에서 하나로 귀착되듯이 물음 묻기 사건의 결과도 **한 시점에** 여러 가능성 중에서 하나로 귀착되어야 한다. 일요일 어느 시점에는 앞면 아니면 뒷면이 출현한다. 월요일 어느 한 시점에는 Y형 문제 아니면 N형 문제가 제시된다. 각 사건이 출현할 확률은 1/2이다. 화요일 어느 한 시점에는 N형 문제가 제시되든지 아니면 아무 문제도 제시되지 않는다. 각 사건이 출현할 확률은 1/2이다. 따라서 월요일에 N형 문제가 제시될 확률은 1/2이고 화요일에 N형 문제가 제시될 확률은 1/2이다. 이 둘을 합치면 1이 된다.

"미녀가 깨어났을 때"

H1, T1, T2를 독립적 사건으로 보고 이들 확률의 총합이 1이라고 생각하는 것은 이 사건이 발생 시점을 제거했기 때문이다. 다시 말해 진술 속에 시간을 포함시키고 그 진술에 시간적 요소가 없는 것으로 간주하는 것이다. 엘가와 루이스는 확률을 합산하는 공시적 시점을 "미녀가 깨어났을 때"를 잡았다. 그리고 "월요일"이나 "화요일"이라는 시점은 H1, T1, T2 속에 삽입시켰다. 그런데 "미녀가 깨어났을 때"(u)라는 시점은 "월요일"(m)과 동일할 수 있고 "화요일"(w)과 동일할 수 있다.

경우	시점 u의 확률 분포	표본공간 내 모든 사건들	이 시점에서 각 사건의 확률
u = m	P(u=m) = 1	T1, H1	P^m(T1) = 1/2 P^m(H1) = 1/2
u = w	P(u=w) = 1/2	H2	P^w(H2) = 1
u=m ∨ u=w	P(u=m ∨ u=w) = 1.5	T1, H1, H2	P^u(T1)+P^u(H1) ∨ P^u(H3) = 1.5

공시적 시점 u에서 사건 H1, T1, T2은 일어날 만한 가능한 사건들이다. 그리고 이 시점에서 각 사건들은 서로 배타적이다.

그런데 문제는 명제 "u=m"가 참이 될 확률은 1이고 명제 "u=w"가 참이 될 확률은 1/2이다. 이것은 명제 "u=m"와 "u=w"가 서로 배타적이지 않다는 것을 의미한다. 나아가 이것은 u의 시점이 실제 세계의 한 시점과 동일시될 수 없음을 의미한다. 이 시점은 미녀의 주관적 시공간에서 특정 시점이다. "미녀가 깨어났을 때"(u)라는 시간은 우리의 시공간 내 특정 시점이 아니다. 미녀와 우리는 시간을 공유하지 않는다. "월요일"(m)이나 "화요일"(w)이라는 시간은 우리의 시간이며 미녀는 이 시간 속에서 생각하지/살지 않는다. 그녀에게 m과 w는 시간적 지표가 아니라 그 외 다른 지표이다. 따라서 "u = m"라는 명제와 "u = w"라는 명제는 사이비 명제이다.

미녀의 시간 프레임에서 볼 때 T1과 H1과 H2는 u에서 일어날 만한 세 사건이다. 다시 말해 그녀의 시간 프레임에서 각 사건은 공시적 시점에 일어날 만한 사건이며 서로 배타적이다.7) 따라서 각 사건이 발생할 확률의 총합은 1이 되어야 한다. 각 사건의 확률을 어떻게 부여할 수 있을까?

프레임	시점	T1	H1	H2	총합
우리의 시공간	m	$P^m(T1) = 1/2$	$P^m(H1) = 1/2$	$P^m(H2) = 0$	1
	w	$P^w(T1) = 0$	$P^w(H1) = 0$	$P^w(H2) = 1/2$	0.5
미녀의 시공간	u	$P^u(T1) = ?$	$P^u(H1) = ?$	$P^u(H2) = ?$	1

미녀는 "Y형 물음이 월요일에 주어진다", "N형 물음이 월요일에 주어진다", "N형 물음이 화요일에 주어진다"를 예측하거나 사후추측할 만한 그 어떤 정보도 가지고 있지 않다. 그녀는 주어진/주어질 물음이 Y형 물음인지 N형 물음인지 전혀 감지할 수 없다. 그녀는 주어진/주어질 물음에 붙은 이상한 지표 "월요일"과 "화요일"의 차이를 분간할 수도 없다. 다시 말해 미녀가 가진 정보를 모두 동원해도, 자신의 시공간 내 특정 시점 u에서 T1, H1, H2 중 특정 사건이 선호될 만한 그 어떤 편향성도 찾을 수 없다. 따라서 세 사건의 발생 확률은 똑같다. $P^u(T1) = P^u(H1) = P^u(H2) = 1/3$

그런데 그녀가 알고 있는 정보에 따르면 H1이 발생하면 반드시 H2가 발생하고 H2가 발생하면 반드시 H1이 발생한다. H1와 H2가 동일한 시간에 일어날 만한 사건이고 하나의 발생이 다른 하나의 발생을 필연적으로 함축한다면, 두 사건은 구별불가능하게 접힌(중첩된) 단일 사건으로 간주될 것이다. 이 사건을 H라고 해보자. 미녀의 프레임에서 시점 u에 실제로 벌어지는 사건은 T1이거나 H이다. 시점 u에 T1이 발생하면 H가 발생하지 않고, H가 발생하면 T1이 발생하지 않는다. H가 발생할 확률 $P^u(H) = 1 - P^u(T1) = 2/3$. 시점 u에 H가 발생한다는 것은 미녀에게는 N형 물음이 주어진다는 것을 의미한다. 따라서 미녀의 시공간에서 자신이 깨어나는 시점(u)에서 N형 문제가 주어질/주어졌을 확률은 2/3이다. 반면 그 시점에 Y형 문제가 주어질/주어졌을 확률은 1/3이다.

몇 가지 교훈

엘가나 루이스는 다음 요점들 중 일부를 파악하는 데 실패한 듯이 보인다. 첫째, 미녀의 주관적 확률이 관계하는 개연적 추측은 미래에 벌어질 일에 대한 예측이 아니라 이미 벌어졌지만 자신이 모르는 일에 대한 사후 추측이다. 동전이 앞면이 "나올" 확률은 미녀에게 언제나 1/2이지만 앞면이 "나왔을" 확률은 경우에 따라 1/2가 아닐 수 있다.

둘째, 미녀에게 주어질 물음의 제작 방식은 그 물음에 대한 정답을 좌우하는 정보를 포함하고 있다. 따라서 그녀에게 그 제작 방식을 알려주는 것은 그녀에게 "동전이 앞면이 나왔을 확률은 얼마인가?"의 대답을 좌우할 정보를 주는 것이다. 이 정보는 반신반의 상태(확률 1/2)를 개선시킬 것이다.

셋째, 미녀에게 주어진 물음은 엄밀히 말해 동전이 앞면이 나왔을 확률을 묻는 것이 아니다. 미녀가 요구받은 것은 세계 속에서 벌어진 구체적 사건으로서 "동전이 앞면이 나왔을 확률은 무엇인가?"

---

7) 미녀는 H1과 H2가 서로 배타적이라는 것을 어떻게 알 수 있을까? "월요일"과 "화요일"이 서로 다른 시간이라는 것을 아는 우리로서는 두 사건이 서로 배타적이라는 것을 안다. 그러나 미녀에게 "월요일"과 "화요일"이 시간이 아니라 제3의 지표라면, 그 지표에 대한 충분한 이해 없이 H1과 H2가 배타적이라는 것을 감지하는 것은 불가능해 보인다. 여하튼 그 지표가 정확히 무엇이든지 간에 미녀는 "월요일 N형 문제"와 "화요일 N형 문제"가 서로 다른 것이라는 것을 안다고 가정하겠다.

라는 발화와 동전이 앞면이 나왔을 개연성 사이의 인과적 관련성을 추적하는 것이다. 동전을 던져 놓고 일요일에 미녀에게 "동전이 앞면이 나왔을 확률은 무엇인가?"라고 묻는다면 그녀의 답은 1/2이다. 이처럼 월요일이 되기 전에 실험 설계에 대한 정보는 주관적 확률을 변화시키지 않는다. 왜냐하면 동전을 던진 결과와 실험 사이에 인과적 연결이 아직 시작되지 않았기 때문이다. 미녀를 깨우고 질문을 하는 사건이 일단 발생하면 동전 던지기 사건의 인과적 효과는 시작된다. 미녀에게 질문하는 시점이 월요일이라는 것을 알려주는 것은 실험 설계를 파괴하는 것 그리하여 그 인과적 효과를 중단시키는 것에 해당한다. 그녀의 주관적 확률은 다시 1/2로 복귀할 수 있다.

넷째, 월요일 어느 시점에서 보았을 때 동전이 이미 앞면이 나온 상태이고 그 시점에 질문하는 사건(H1)이 발생할 확률은 1/3이 아니라 1/2이다. 월요일 어느 시점에서 보았을 때 동전이 이미 뒷면이 나온 상태이고 그 시점에 질문하는 사건(T1)이 발생할 확률도 1/2이다. 또한 화요일 어느 시점에서 보았을 때 동전이 이미 뒷면이 나온 상태이고 그 시점에 질문하는 사건(T2)이 발생할 확률도 1/2이다. 따라서 세 사건의 총 확률은 1이 아니라 1.5이다. 앞의 두 사건과 뒤의 한 사건은 서로 다른 시점에 발생한 사건이다. 확률의 합산은 한 시점에 발생할 만한 사건들만을 취급해야 한다.

다섯째, 망각제는 월요일과 화요일의 시간 간격을 제거하는 효과를 준다. 따라서 망각제는 실제 세계에서 월요일 시점에 발생 가능한 두 사건과 화요일 시점에 발생 가능한 한 사건이 마치 동일한 시간에 발생 가능한 세 사건인 것으로 착각하게 만든다. 미녀에게 그 세 사건은 가상적 세계, 망각제에 의해 형성된 그녀 자신의 시공간 내 어느 한 시점에 벌어질 만한 세 사건이다. 즉 그녀에게는 그 세 사건이 동일한 표본공간에 속하는 사건으로 간주된다. 표본공간 내 사건 발생 확률의 총합은 1이 되어야 하기 때문에 이 세 사건이 발생할 확률의 합계는 1.5가 아니라 1로 재조정되어야 한다. 결국 이 경우 이 가상적 세계에서 H1, T1, T2가 발생할 확률은 1/2에서 1/3로 재조정되어야 한다. 이것은 미녀의 착각일 뿐 실제 세계 즉 우리의 시공간에서 H1, T1, T2가 발생할 확률은 여전히 1/2이다.

여섯째, 망각제는 T1과 T2의 발생 시점을 한 시점으로 변환시킨다. 그런데 T1와 T2는 실험 설계상 필연적으로 연결되어 있다. T1이 발생하면 반드시 T2가 발생한다. 또한 T2가 발생하면 반드시 T1이 발생한다. 따라서 미녀는 바로 그 시점에 한 질문이 아니라 두 개의 동일한 질문을 (동시에) 받게 되는 셈이다. (물론 두 개의 질문을 동시에 받을 확률은 우리의 시공간에서 1/2이다.) 결과적으로 미녀에게 주어진 물음은 사실 하나의 물음이 아니라 중첩된 '무거운' 물음이다. 망각제는 미녀로 하여금 자신에게 주어진 물음이 단수라고 잠시 착각하게 하는 효과를 준다. 그녀가 충분히 똑똑하다면 자신에게 복수의 물음이 주어지고 또한 그 물음의 분포가 편향되어 있다는 것을 알게 될 것이며, 물음의 개수와 그 편향만큼 주관적 확률을 조정할 것이다.