

잠자는 미녀의 시공간[†]

김 명 석[‡]

잠자는 미녀 문제의 교훈으로서, 제프리의 조건화 규칙을 수정해야 한다는 견해부터 루이스의 주요 원리를 수정해야 한다는 견해까지 다양한 논제들이 여태 제안되었다. 이러한 제안들을 거부하면서 이 논문은 다음과 같은 점들을 주장한다. 첫째, 확률 또는 믿음직함을 매기는 주체는 자기 니름의 시공간을 재설정한다. 통상의 경우 주체의 시공간은 경험을 통해 상호주관 시공간 또는 공공 시공간과 손쉽게 동기화되지만, 잠자는 미녀의 경우는 그렇지 않다. 둘째, 잠자는 미녀가 깨어난 후 정보를 얻는 것이 아니라 정보를 잃는데, 그는 그 자신의 시공간과 공공 시공간을 동기화하는 데 필요한 정보를 잃어버린다. 셋째, 기억을 잊거나 정보를 잃는 경우, 조건화 규칙을 사용해서는 안 된다. 조건화 규칙은 기준 정보가 유지되는 한에서만 사용해야 한다. 따라서 일요일에서 월요일을 거칠 때 미녀의 믿음직함이 어떻게 변하는지를 알기 위해 조건화 규칙을 적용하는 것은 옳지 않다. 넷째, ‘나’, ‘지금’, ‘여기’ 등과 같은 지표가 포함된 문장을 다루기 위해, 별도의 조건화 규칙을 고안할 필요가 없고, 대신 지표의 사용을 최소화하여 기존 조건화 규칙을 그대로 사용해도 무방하다. 다섯째, 미녀의 답변이 1/3이어야 한다는 견해는 거의 일어날 수 없는 기적 같은 일이 늘 있는 하찮은 증거에 의해 입증된다는 귀결을 품고 있다.

【주요어】 잠자는 미녀, 자기 위치 정보, 조건화 규칙, 정보, 확률, 입증

[†]이 논문은 2013년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구이다. NRF-2013S1A5A8024144. 이 논문의 심사자들께서 논문의 개선을 위한 합당한 논평과 향후 연구를 위한 유익한 조언을 주신 데 깊이 감사드린다. 몇몇 각주에 그들의 논평 일부에 대해 답변해 놓았다.

[‡]국민대학교 교양대학, myeongseok@gmail.com.

1. 잠자는 미녀

잠자는 미녀 문제¹⁾에 대해 엘가(Elga 2000)와 루이스(Lewis 2001)가 각각 다른 답변을 내놓은 이후 여러 철학자들이 다양한 의견을 쏟아내었다. 미녀의 답변이 1/3이어야 한다는 것(송하석 2011)이 다수 견해인 가운데, 정확히 얼마인지는 모르겠지만 1/2보다 작아야 한다는 견해(Kim 2009), 1/2이어야 한다는 견해(Jenkins 2005; White 2006; 김명석 2012), 1/2과 1/3이 모두 허용된다는 견해(Bostrom 2007; Groisman 2008; 김한승 2009) 등이 제안되었다.

미녀가 깨어났을 때 그가 새로운 정보 또는 새로운 증거를 얻게 된다고 주장하는 사람(Weintraub 2004; 김한승 2009; Karlander & Spectre 2010)이 있고, 새로운 정보를 전혀 얻지 못한다고 주장하는 사람(Elga 2000; Lewis 2001)도 있다. 새로운 정보를 얻지 못했는데 일요일의 확률이 1/2이던 것이 월요일에 1/3로 바뀐 것에 대해, 일부 1/3주의자들은 반 프라센의 반성 원리와 제프리의 조건화 규칙을 포기하거나 제한해야 한다고 주장하기도 한다. 많은 이들이 조건화 규칙을 ‘나’, ‘여기’, ‘지금’, ‘오늘’, ‘어제’ 등과 같은 지표가 있는 명제에 적용해서는 안 된다고 주장한다(Meacham 2008; Kim 2009; Lewis 2010; Schulz 2010). 새로운 정보를 얻지 못했기 때문에 일요일의 확률 1/2이 월요일에도 1/2로 유지되지만, 미녀가 자신이 깨어난 날이 월요일이라는 것을 알게 되면 그 때 확률이 2/3로 바뀌게 된다는 1/2주의의 귀결을 토대로, 루이스의 주요 원리를 포기하거나 제한해야 한다고 주장하는 이들도 있다(Lewis 2001).

1) 일요일에 동전을 던지고 미녀를 재운 다음, 동전이 앞면이 나오면 월요일에 한 번 깨우고, 뒷면이 나오면 월요일과 화요일에 깨운다. 미녀를 깨운 후 “던진 동전이 앞면일 확률은?”이라는 물음을 던지고 답변을 마치면 물음을 받았다는 기억을 지운 후 다시 재운다. 미녀가 깨어났을 때 그가 아는 것은 일요일에 들은 문제의 설정과 지금이 일요일 이후라는 사실뿐이다. 월요일인지 화요일인지 알 수 없는 어느 날 미녀가 “던진 동전이 앞면일 확률은?”이라는 물음을 듣게 된다면, 그는 얼마라고 답해야 하는가?

잠자는 미녀 문제를 둘러싼 이러한 논쟁들에 참여하면서 이 논문은 다음과 같은 논제들을 주장한다. 첫째, 확률을 계산하는 주체 또는 믿음직함을 매기는 주체는 자기 나름의 시공간을 재설정한다. 이 논제는 제2절의 주제인데 ‘믿음직함의 칸트주의’라고 이름 지었다. 통상의 경우 주체의 시공간은 경험을 통해 상호주관 시공간 또는 공공 시공간과 손쉽게 동기화된다. 이러한 동기화가 지표가 포함된 명제를 통해 이루어진다는 점이 제3절에서 제안된다. 둘째, 먼턴이 이미 주장했듯이 (Monton 2002), 잠자는 미녀는 월요일에 정보를 얻는 것이 아니라 정보를 잃는다. 다시 말해 그는 알고 있던 것을 잊어버린다. 일요일 밤 이후 실험하는 사람은 미녀의 주관 시공간을 공공 시공간과 동기화하는 데 필요한 정보를 미녀에게서 강제로 지운다. 셋째, 기억을 잊거나 정보를 잃는 경우, 그 정보가 확률 변화에 관련되는 한, 조건화 규칙을 사용해서는 안 된다. 조건화 규칙은 기존 관련 정보가 유지되는 한에서만 사용할 수 있다. 넷째 논제와 셋째 논제는 제4절에서 다룬다. 이 논문은 이들 논제들에 대해 새로운 것을 제안한 것이 아니며, 다만 잠자는 미녀 문제에 이 논제들을 적용하고 있을 뿐이다. 다섯째, ‘나’, ‘지금’, ‘여기’ 등과 같은 지표가 포함된 문장을 다루기 위해, 별도의 조건화 규칙을 고안할 필요가 없고, 대신 지표의 사용을 최소화하여 기존 조건화 규칙을 그대로 사용해도 무방하다. 제5절에서 김남중의 해법을 비판하면서 넷째 논제를 주장한다. 다섯째, 미녀의 답변이 $1/3$ 이어야 한다는 견해에는 설명하기 어려운 귀결이 숨어 있다. $1/3$ 주의의 계산 방식을 따를 경우, 거의 일어날 수 없는 기적 같은 일이 언제나 목격할 수 있는 하찮은 증거에 의해 입증되는 사례를 만들어낼 수 있다. 이러한 귀결을 피하고자 한다면, 동전을 일요일에 던지는 경우와 월요일에 미녀를 재운 뒤에 던지는 경우가 같다는 가정을 포기해야 한다. 이런 것들은 제6절에서 다룬 내용이다.

2. 의식과 시공간

우리가 ‘나’, ‘지금’, ‘여기’ 등과 같은 개념을 갖지 않은 채 믿음을 갖는 것은 거의 불가능해 보인다. 믿음을 갖는 주체 또는 의식은 ‘나’, ‘지금’, ‘여기’ 등과 같은 개념을 이미 갖고 있다. 그래서 우리는 다음을 가정할 수 있겠다.

믿음을 갖는 이는 시공간 좌표 또는 지표를 이미 갖고 있다. 그 시공간 좌표의 중심에 자기 자신이 위치해 있다. 의식이 작용하는 시점이 현재이고, 기억으로 남아 있는 것은 과거의 일로, 기대하는 일은 미래의 일로 간주된다. 이런 식으로 의식은 ‘나’, ‘여기저기’, ‘지금’ 등과 같은 지표 개념을 이미 갖고 있다.

‘믿음을 갖는 이’는 ‘의식’ 또는 ‘주체’라 달리 불리곤 한다. 한 개별 의식이 설정하는 시공간을 ‘주관 시공간’이라고 부르자.

일상생활에서 우리의 주관 시공간은 다른 사람들의 주관 시공간과 조율되어 있다. 한 개인은 감각기관을 통해 공통 시공간 내에 있는 사건과 사물을 다른 개인들과 함께 경험한다. 한 공동체 내 개인들의 주관 시공간들이 조율되어 의사소통의 기반이 되는 시공간을 ‘상호주관 시공간’ 또는 ‘공공 시공간’이라 하자. 사람들은 해와 달 같은 공통 현실 세계의 사건과 사물을 함께 겪어오면서 그들의 상호주관 시공간을 객관 시공간에 동기화하는 방법을 고안해 왔다. 해나 달의 주기 운동을 통해 한 해, 한 달, 한 날, 한 시 등을 정의하고, 이 표준자를 정해서 거리 또는 공간을 객관화한다. 엄밀한 의미의 객관 시공간이 있지는 의심의 여지가 있지만, 오늘날에도 시간 측량과 공간 측량을 객관화 내지 표준화하는 일을 계속하고 있다.

오늘날 개인들은 지도와 시계 또는 달력과 같은 공공 시공간 좌표를 통해 자신의 주관 시공간을 상호주관 시공간과 동기화하고, 나아가 다른 사람들의 주관 시공간을 손쉽게 파악한다. ‘어제’, ‘오늘’, ‘내일’ 같은 낱말은 상호주관 시공간 용어로 쓰기 위해 고안한 개념 ‘하루’와

주관 시공간 개념을 결합한 것이다. 이런 낱말들은 상호주관 시공간 용어라기보다 주관 시공간 용어로 여겨야 한다. ‘지난 달’, ‘이번 달’, ‘다음 달’, ‘작년’, ‘올해’, ‘내년’ 등도 이와 같은 주관 시공간 용어이다. ‘여기서 1미터 오른쪽’, ‘저기서 뒤쪽으로 1킬로미터’ 등도 상호주관 시공간 용어를 포함하고 있지만 사실상 주관 시공간 용어이다. 사람들은 다른 사람의 이러한 주관 시공간 용어를 상호주관 시공간 용어로 번역한 뒤에야 그 말을 자신의 주관 시공간 개념을 빌려서 이해할 수 있다. “지금은 T이다.”, “여기는 X이다.” 같은 문장은 주관 시공간을 상호주관 시공간과 동기화하는 진술인데, ‘지금’과 ‘여기’는 주관 시공간 용어인 반면, ‘T’와 ‘X’ 자리에는 상호주관 시공간 용어가 나와야 한다. 가령 “지금은 2016년 5월 18일 정오이다.”나 “여기는 광주 5.18민주광장 민주의 종 앞이다.” 등처럼.

골방에 혼자 있는 사람들조차도 자기 몸이 다른 집들, 도로, 산들, 도시들, 나라들과 대략 어떻게 관계 맺고 있는지, 자신의 현재 시간이 달력의 어느 시점에 표시될 수 있는지 대략 가늠할 수 있다. 긴 악몽을 꾼 후 깨어났을 때, 의식을 잃은 상태에서 되돌아왔을 때, 자신의 주관 시공간을 공공 시공간과 동기화하지 못할 때가 간혹 있다. 이럴 때 다른 사람들의 시공간 용어를 이해하지 못하고 자기 자신의 시공간 용어도 이해받지 못하곤 한다. 자신의 주관 시공간을 상호주관 시공간과 동기화하지 못해 자신의 시공간 좌표를 공공의 시공간 좌표에 제대로 위치시키지 못할 때, 그래서 “지금은 T이다.”, “여기는 X이다.” 같은 명제를 알지 못할 때, 그는 ‘자기 자리 믿음’(self-locating belief) 또는 ‘자기 위치 정보’를 잊었다고 말할 수 있다.²⁾

깨어난다는 것, 의식이 되돌아온다는 것, 의식을 갖는다는 것은 자신의 주관 시공간을 갖는 것을 필요로 한다. 다시 말해 주관 시공간을 갖지 않으면서 의식을 가질 수는 없다. 자기 주관 시공간을 공공의 시공간과 연결하지 못한다 하더라도 여전히 의식을 가질 수 있다. 의식을 가진 후 여러 경험을 통해 차차 자기 주관 시공간을 공공의 시공간

2) 루이스는 사람의 모든 믿음들이 사실상 자기 위치 믿음이라고 주장한다.

Lewis (1979), p. 522.

과 동기화한다. 완전히 캄캄하고, 주위 어떤 사물도 만질 수 없고, 신체 운동을 할 수 없을 때조차도, 의식이 있는 한 우리는 주관 시공간을 갖는다. 이전에 갖고 있던 자기 위치 정보를 상실했다 하더라도, 그는 주관 시공간을 언제든지 재설정할 수 있다.³⁾ ‘지금’, ‘오늘’, ‘여기’, ‘나’ 등이 바로 이러한 재설정 과정 자체를 나타내는 개념이기도 하다.

잠에서 깨어난 미녀는 주관 시간이 공공 시간 내에서 어느 자리에 있는지 파악하지 못한다. 하지만 깨어난 미녀가 그의 주관 시공간조차 갖지 못한다고 주장해서는 안 된다. 이제 다음과 같은 ‘믿음직함의 칸트주의’를 제안하고자 한다. ‘믿음직함’은 통상의 저자들이 “신념도”(credence), “믿음 강도”, “주관 확률” 등이라 부르는 것과 같은 말이다.

믿음을 갖는 이는 믿음직함을 매길 때마다 주관 시공간을 설정한다.

개별 의식이 주관 시공간을 설정할 때, 주관 시공간을 공공 시공간과 동기화하는 데 필요한 정보를 잊거나 잊어버리기도 한다. 월요일인지 화요일인지 모르는 날 깨어난 미녀는 이러한 상황에 놓여 있다. 하지만 그는 여전히 자신의 주관 시공간 개념을 갖는다. 주관 시공간 좌표 중심에 ‘나’를 위치시키고 ‘여기’와 ‘지금’을 의식한다. 이런 의미에서 ‘여기’나 ‘지금’ 또는 ‘나’를 포함하고 있는 명제는 ‘중심 명제’ 또는 ‘자기 중심 명제’라고 부를 만하다.

3. 지표의 역할과 자기 위치 정보

‘나’, ‘여기’, ‘지금’ 등과 같은 주관 용어들을 다른 상호주관 또는

³⁾ 신체가 없는 상태에서도 의식만으로 개별 시공간을 재설정할 수 있다고 가정하는 것은 아니다. 또한 몸 없이도 시공간 개념을 가질 수 있는 의식을 가정하는 것도 아니다.

객관 용어들과 결합함으로써 의식은 상호주관화 또는 객관화된다. 가령 주체는 “나는 A이다.”, “지금은 T이다.”, “여기는 X이다.” 등과 같은 믿음을 통해 자신과 그의 시공간을 상호주관화 또는 객관화한다. 따라서 의사소통하고 인식하는 주체는 결국 몇몇 지표 있는 명제를 품어야 한다.

루이스는 명제 태도의 내용 또는 대상이 되는 명제를 지표 명제들로 보아야 한다고 주장한다(Lewis 1979). 지표가 있는 명제는 ‘중심 명제’라 하고, 그러지 않은 명제를 ‘비중심 명제’라고 부르곤 한다. 주관 시공간을 상호주관 시공간과 동기화하는 명제들은 지표가 들어 있을 수밖에 없다. 하지만 내 생각에, 그밖에 명제들은 원칙상 지표를 없애도 된다. 대부분의 명제는 원칙상 비중심 명제 또는 지표 없는 명제여야 한다는 이 생각은 명제에 대한 아주 오래된 직관을 반영하고 있다. 일단 한 명제가 참이라면 그 명제는 영원히 참이다. 일단 한 명제가 거짓이라면 그 명제는 영원히 거짓이다.⁴⁾

문장에 지시사가 들어 있을 경우 그 문장이 표현하는 명제도 상황에 따라 참값이 바뀐다고 생각을 한다. 하지만 문장에 지시사가 들어 있을 경우, 상황에 따라 바뀌는 것은 명제의 참값이 아니라, 그 문장이 표현하는 명제 자체이다. 문장 “오늘은 월요일이다.”는 2016년 2월 1일에는 참이지만 2016년 2월 2일에는 거짓이다. 하지만 이것은 문장 “오늘은 월요일이다.”가 표현하는 명제의 참값이 상황에 따라 바뀐다는 것을 말해주지 않는다. 오히려 2016년 2월 1일에 발화 또는 기재된

⁴⁾ 콰인은 항상 문장의 인지 의미(the cognitive meaning of an eternal sentence)로서 ‘명제’ 또는 ‘명제 같은 대상’이 언어 현상을 설명하는 데 별 도움이 되지 않는다고 줄곧 주장했다. 가장 큰 골칫거리는 그러한 대상을 개별화하는 조건을 제시하는 일이다. Quine (1969), p. 140. 이 글은, 명제에 관한 한, 콰인 대신에 데이빗슨의 견해를 은연중에 받아들이고 있다. 마음은 명제 같은 대상과 ‘실제로’ 관계 맺는 것이 아니다. 굳이 말하자면 명제 또는 의미는 해석 상황에서 출현하는 상정체이다. 우리가 명제나 의미가 파악되었다거나 교류되었다고 말하는 것은 우리가 타자의 발화나 행위를 이해하고 있다는 말일 뿐이다. Davidson (1989), p. 67.

문장 “오늘은 월요일이다.”는 명제 “2016년 2월 1일은 월요일이다.”를 표현하고, 2016년 2월 2일에 발화 또는 기재된 문장 “오늘은 월요일이다.”는 명제 “2016년 2월 2일은 월요일이다.”를 표현하고 있을 뿐이다.⁵⁾ 하지만 2016년 2월 1일에 발화 또는 기재된 문장 “오늘은 2016년 2월 1일이다.”는 명제 “2016년 2월 1일은 2016년 2월 1일이다.”를 표현하는 것이 아니다. 자기 위치 정보를 표현하는 문장에서 지표를 완전히 제거하는 것은 불가능하다.⁶⁾

앞에서 믿음을 갖는 모든 주체는, 의식이 있는 한, 주관 시공간을 언제나 설정한다고 가정했다. 통상의 주체는 경험을 통해 자신의 주관 시공간을 공공 시공간과 동기화할 정보를 얻는다. 이 정보를 통해 자기 위치 정보를 알게 된다. 자신의 주관 시공간을 공공 시공간과 동기화하는 데 필요한 정보를 ‘ LI_t ’라고 하자. 여기서 아래 첨자 t 는 공공 시공간 내 주체의 좌표를 표시한다. LI_t 는 보통 “내가 있는 지금 여기

⁵⁾ 물론 화자, 청자, 인식 주체 등이 자신의 주관 시공간 용어를 객관 시공간 용어로 즉각 번역할 수 있다거나, 적어도 상호주관 시공간 용어로 즉각 번역할 수 있다고 가정할 필요는 없다. 의식을 가진 둘 이상의 주체가 공통 세계 속에서 대화를 나눌 때, 심지어 지도나 달력이 없다 하더라도, 그들은 상호주관 시공간 용어를 새로 창출하곤 한다. 그들은 의사소통 맥락에서 주관 시공간 용어 ‘여기’나 ‘그 때’를 상호주관 시공간 용어로 알아듣곤 한다. 의사소통 상황에서 ‘여기’, ‘저기’, ‘지금’, ‘그 때’, ‘거시기’ 등이 아무런 내용이 없다고 보기는 어렵다. 두 주체는 어렵잖하게나마 또는 대충이나마 공통의 시공간 좌표를 이미 그리고 있다. 실제 상황에서 사람들은 과거에 공통 세계 내에서 많은 것들을 경험하고 있었고 그것들을 기억하고 있다. 의사소통 중에 서로 확인되는 공통의 경험과 기억을 통해 그들은 하나의 시공간 지도와 달력을 대충이나마 함께 그려나간다. ‘저기’는 ‘우리가 함께 영화 「변호인」을 보았던 곳’이고, ‘그날’은 ‘내가 북촌에서 산책을 하고 네가 서촌에서 술을 마시던 날’일 수 있다. 한 개인이 자신의 주관 시공간 개념을 최초로 어떻게 획득하였는지는 별개의 문제인데, 한 개인이 제 홀로 스스로 시공간 개념을 창출할 수 있을 것 같지는 않다.

⁶⁾ 잠자는 미녀 문제에서 “오늘은 월요일이다.”는 자기 위치 정보를 표현하는 문장으로 사용된다. 이 경우 ‘오늘’은 제거할 수 없는 지표이다.

는 공공 시공간 좌표에서 t 에 있다.”라는 꿀을 띠고 있다. 보다 길게 “나는 다른 사람들이 A라고 부르는 사람이고, 내가 있는 지금은 표준 시계로는 T이고, 내가 있는 여기는 표준 지도에서는 X이다.”라고 표현 할 수 있다.

공공 시공간에서 주체의 좌표가 Δ 만큼 바뀌었다고 스스로 믿을 경우, 자신이 이미 갖고 있던 LI_t 정보를 통해, “내가 있는 지금 여기는 공공 시공간 좌표에서 $t+\Delta$ 에 있다.”를 대략 추정한다. 이것이 ‘대략’인 이유는 Δ 자체가 속셈으로 대강 파악한 것이기 때문이다. 사람의 감각 기관과 기억은 Δ 를 추정하는 데 상당히 능숙하다. 물론 잠든 후 깨어 났을 때 변화된 시간이 7시간이라고 속셈하더라도, 실제로는 한 시간 또는 이틀이 지났을 수도 있다. 우리는 시계와 달력 또는 지도를 참조 함으로써 보다 정확하게 LI_t 정보를 갱신한다.⁷⁾ 이런 장치들은 확장된 감각기관이자 확장된 기억이다.

4. 베이즈 주체는 망각하지 않는다.

베이즈 주체는 관련 정보가 새로 주어지면 이전에 갖고 있던 신념도 또는 믿음직함을 바꾼다. 가령 한 베이즈 주체가 갖고 있던 명제 X 에 대한 어제의 믿음직함을 $C_{\text{어}}(X)$ 라고 하자. 그가 오늘 정보 E 를 새로 얻었다고 하자. 그가 가져야 하는 명제 X 에 대한 오늘의 믿음직함 $C_{\text{오}}(X) = C_{\text{어}}(X|E) = C_{\text{어}}(X \& E) / C_{\text{어}}(E)$ 이다. 많은 베이즈주의자들은 이러한 조건화 규칙이 베이즈주의를 구성하는 요소라고 생각한다.⁸⁾

그들은 신념도 함수 또는 믿음직함 함수를 시간 지표 t 를 써서

⁷⁾ 오늘날 스마트폰을 통해서 실시간 및 실장소에서 얻을 수 있는 GPS 기반 위치 및 시간 정보는 자신의 주관 시공간 좌표를 공공 시공간 좌표에 동기화하는 데 거의 충분하다. 이 공공 시공간 좌표는 현생 인류 규모의 상호 주관성을 확보해주고 있으며, 심지어 상대성이론에 따른 지구 표면상의 객관 시공간 좌표를 반영하고 있다.

⁸⁾ 이영의 · 박일호 (2015), pp. 4-5.

$C_t(X)$, $P_t(X)$ 등으로 나타내는 데 익숙하다. 잠자는 미녀의 문제를 풀 때도 이와 같은 방식을 취하곤 한다. 하지만 여기에는 숨은 가정이 있다. 주체 또는 행위자가 이전의 정보를 계속 유지한다고, 즉 시간의 흐름에 따라 망각하지 않는다고 가정한다. 시간이 t 에서 $t+\Delta$ 로 흘러가면서, 행위자가 증거 E 를 새로 얻었지만, t 에 갖고 있던 다른 관련 정보를 $t+\Delta$ 에서 잊어버렸을 때, $C_{t+\Delta}(X) = C_t(X|E)$ 라고 쓰는 것은 옳지 않다.⁹⁾ 이 조건화 규칙은 주체가 망각하는 경우까지 포괄하는 규칙이 아니다. 만일 잠자는 미녀가 일요일에서 월요일 시점으로 가는 동안 잊거나 잊어버린 관련 정보가 있다면, 저 조건화 규칙을 통해 믿음직함을 바꾸어야 한다고 요구하는 것은 옳지 않다.

조건화가 진행되는 방향이 시간 흐름과 반드시 일치할 필요는 없다. 그것이 일치할 때는 믿음직함을 매기는 주체가 시간 흐름에 따라 정보를 잊지 않거나 정보가 누적되는 경우이다. 조건화는 정보가 쌓이는 방향으로 진행되어야 한다. 가령 한 베이즈 주체가 시간 s 에 명제들의 집합 $I_s = \{A, B\}$ 를 알고 있고 시간 t 에 명제들의 집합 $I_t = \{A, B, E\}$ 를 알고 있다고 하자.¹⁰⁾ s 와 t 의 선후 관계를 알지 못하는 경우에도, $C_s(X) = C_s(X|E)$ 라고 쓸 수 있다. 시간보다는 정보 상태를 첨자에 표시하여 $C_{ls}(X) = C_{ls}(X|E)$ 라고 쓰는 것이 보다 더 적절하다. 다시 말해 믿음직함의 개선을 시간에 따라 추적하는 것이 아니라 정보의 축적을 따라 추적해야 한다. 조건화가 진행되는 방향은 시간이 흐르는 방향이

9) t 에서 $t+\Delta$ 로 시간이 흘러가면서 행위자가 t 에 갖고 있던 정보 F 를 $t+1$ 에서 잊었을 때 $C_{t+\Delta}(X) = C_t(X|\sim F)$ 라고 쓰는 것은 옳지 않다. $C_{t+\Delta}(X) = C_t(X|\sim F)$ 는 정보 F 를 잊은 것이 아니라 정보 $\sim F$ 를 얻을 때 믿음직함의 변경을 나타낼 뿐이다. 정보 F 를 잊는 것은 F 가 참 또는 거짓이라고 알고 있다가 F 가 참인지 거짓인지 모르게 되는 것이다. 정보 $\sim F$ 를 얻는 것은 F 가 참인지 거짓인지 모르고 있다가 비로소 F 가 거짓이라는 것을 아는 것이다. 따라서 행위자가 t 에서 $t+\Delta$ 로 시간이 흘러가면서, 증거 E 를 새로 얻었지만, t 에 갖고 있던 다른 관련 정보 F 를 $t+\Delta$ 에서 잊어버렸을 때, $C_{t+\Delta}(X) = C_t(X|E \& \sim F)$ 라고 쓰는 것은 옳지 않다.

10) 한 주체가 명제들의 집합을 안다는 것은 그 집합 내의 각 명제들을 안다는 것을 뜻한다.

아니라 정보가 쌓이는 방향이다. 정보가 소실될 때는 조건화 규칙을 써서는 안 된다. 제프리 조건화는 아래와 같이 정식화할 수 있다. 한 베이즈 주체가 시간 s 에 명제들의 집합 $I_s = \{A, B\}$ 를 알고 있고, 시간 t 에 명제들의 집합 $I_t = \{A, B, E_1\text{이거나 } E_2\}$ 를 알고 있다고 하자. 여기서 E_1 과 E_2 는 동시에 참일 수 없다. s 와 t 의 선후 관계를 알지 못하는 경우에도, $C_t(X) = C_s(X|E_1)C_t(E_1) + C_s(X|E_2)C_t(E_2)$ 라고 쓸 수 있다. 이것은 $C_{I_t}(X) = C_{I_s}(X|E_1)C_{I_t}(E_1) + C_{I_s}(X|E_2)C_{I_t}(E_2)$ 라고 쓰는 것이 보다 더 정확하다.

만일 자기 위치 정보 LI_t 가 믿음직함의 변화에 영향을 주는 경우에 조건화 규칙을 어떻게 해야 하는가? 한 주체가 공공 시공간 안에 놓인 한 시점 및 지점 s 에서, 자기 위치 정보 LI_s 를 알고 있고 또한 명제 A 와 B 를 알고 있다고 생각해 보자. 또 그 주체가 공공 시공간 안에 놓인 다른 한 시점 및 지점 t 에서, 자기 위치 정보 LI_t 를 알고 있고 또한 명제 A 와 B 와 E 를 알고 있다고 생각해 보자. s 든 t 든, 한 시점과 지점에서, 한 주체에게 “내가 있는 지금 여기는 공공 시공간 좌표에서 s 에 있다.”와 “내가 있는 지금 여기는 공공 시공간 좌표에서 t 에 있다.”가 동시에 참일 수는 없다. 하나가 참이면 다른 하나는 거짓이다. 따라서 s 에서 그 주체는 정보 상태 $I_s = \{A, B, LI_s, \sim LI_t\}$ 에 있고, t 에서 그는 정보 상태 $I_t = \{A, B, E, LI_t, \sim LI_s\}$ 에 있다.¹¹⁾ 결국 만일 믿음직함의 계산에서 자기 위치 정보가 관련 정보가 된다면 두 정보 상태에 대해 조건화 규칙을 적용할 수 없다.

이제 우리의 고찰을 잠자는 미녀 문제에 적용해 보자. 믿음직함을 따지는 각 순간에 주체가 가진 공통의 배경 정보를 ‘BI’라고 하자. 잠자는 미녀의 경우, 잠자는 미녀 문제 설정 자체가 BI에 해당한다. 올해 첫째 일요일 즉 CE 2016년 1월 4일 밤에 동전을 던진다.¹²⁾ 미녀는 잠에서 깨어나 “나는 지금 물음을 받고 있다.”라는 정보를 얻는다. 이 정보를 ‘ Q ’라고 하자. 정보 Q 에서 ‘지금’은 미녀의 주관 시공간 용어이

¹¹⁾ 사실 $\sim LI_t$ 와 $\sim LI_s$ 는 군더더기이다.

¹²⁾ ‘CE’는 ‘Common Era’ 또는 ‘Current Era’의 약자로서 ‘AD’ 대신에 비기독교 국가에서도 널리 사용된다.

며, LI가 주어지지 않는다면, 날짜나 요일에 관한 한, 상호주관 시공간 용어로 번역할 수 없다. 2016년 1월 4일, 5일, 6일 각각에 미녀가 갖는 정보를 나열하면 다음과 같다.

20160104: BI, \sim Q, $LI_{20160104}$

20160105: BI, Q, $\sim LI_{20160104}$

20160106: BI, Q, $\sim LI_{20160104}$

2016년 1월 4일에 미녀는 정보 $LI_{20160104}$ 를 통해 “지금은 2016년 1월 4일이다.”를 안다. 미녀는 이 날이 일요일이라는 것을 안다고 하자. 하지만 미녀는 2016년 1월 5일에 정보 $LI_{20160105}$ 또는 $LI_{20160104+\Delta}$ 를 읽었기 때문에 자신의 주관 시공간을 공공 시공간과 동기화하지 못한다. 물론 BI와 Q로부터 “지금은 2016년 1월 5일이거나 1월 6일이다.”를 추론한다. 이것은 $LI_{20160105}$ 또는 $LI_{20160106}$ 이 되기에는 충분하지 않다. 결국 미녀는 2016년 1월 5일에도, 1월 6일에도 LI를 갖고 있지 않다.

2016년 1월 4일에서 2016년 1월 5일로 넘어가면서 미녀는 자기 위치 정보 $LI_{20160104}$ 를 갖고 있다가 $LI_{20160104}$ 가 거짓이라는 것을 알게 되었기 때문에, 만일 자기 위치 정보가 믿음직함 계산에 관련이 된다면, 이 두 시점에 조건화 규칙을 적용해서는 안 된다.¹³⁾ 또한 1월 4일 주

¹³⁾ $LI_{20160104}$ 를 갖고 있다가 $LI_{20160104}$ 가 거짓이라는 것을 알게 되었을 때 미녀가 망각한 것은 정확히 무엇일까? 그것은 $LI_{20160104+\Delta}$ 이다. 미녀는 Δ 가 하루인지 이틀인지 모르고 있다. 주체가 E를 알고 있다가 $\sim E$ 를 알게 되는 경우에, E를 망각하고 $\sim E$ 를 얻었다고 가정해야 한다. 그렇게 가정하지 않을 경우 주체는 E와 $\sim E$ 를 동시에 알고 있다고 말해야 하는데 이것은 합당하지 않다. 그래서 정보 $\sim E$ 를 제공하는 것은 E를 지우는 효과도 갖는다. 또한 정보 E를 제공하는 것은 $\sim E$ 를 지우는 효과도 갖는다. 아무튼 $LI_{20160104}$, $LI_{20160105}$ 등이 미녀의 믿음직함을 바꿀 만한 정보라면, 일요일에서 월요일로 넘어갈 때 미녀는 조건화 규칙을 써서는 안 된다.

2016년 1월 5일에서 거꾸로 1월 4일로 거슬러갈 경우에도 조건화 규칙을 쓸 수 없다. 두 시점은 $LI_{20160104}$ 뿐만 아니라, 정보 Q와 정보 $\sim Q$ 를 새로 얻거나 읽는다. 물론 BI 안에 일요일에는 $\sim Q$ 가 참이고, 월요일과 화요

체는 $\sim Q$ 를 알고 있다가 1월 5일 Q 를 증거로서 얻었기 때문에, 만일 Q 가 믿음직함 계산에 관련되어 있다면, 두 시점에서 조건화 규칙을 사용할 수 없다. 결국 우리는 다음 등식들을 수용해서는 안 된다.¹⁴⁾

$$C_{20160105}(X) = C_{20160104}(X | \sim LI_{20160104})$$

$$C_{20160105}(X) = C_{20160104}(X | Q)$$

$$C_{20160105}(X) = C_{20160104}(X | Q \& \sim LI_{20160104})$$

Q 든 $LI_{20160104}$ 나 $\sim LI_{20160104}$ 든 이것들이 두 공공 시점의 믿음직함 계산에 관련이 없다고 가정할 경우에 조건화 규칙을 사용할 수 있다. 이런 경우는 단순히 $C_{20160105}(X) = C_{20160104}(X)$ 이다.¹⁵⁾

일에는 Q 가 참이라는 것이 이미 포함되어 있기 때문에, 미녀에게 $LI_{20160104}$ 를 주는 것은 $LI_{20160104}$ 뿐만 아니라 $\sim Q$ 까지 제공하는 셈이다. 그리고 미녀에게 정보 Q 를 주면, 그는 이로부터 $\sim LI_{20160104}$ 를 추론할 수 있다.

14) 김명석 (2012a), p. 22에는 개념상 혼란스러운 이러한 조건화가 넘지시 시도되고 있다. 조건화 규칙이 적용되는 두 상황 s 와 o 에 대해 서로 어울릴 수 없는 정의들이 다소 부주의하게 섞여 있다. 송하석 (2012), p. 7에서 이 점이 이미 지적되었다. 김명석 (2012b), p. 79 각주 9를 보라.

15) 믿음직함 함수 $C_t(X)$ 를 정의할 때, 주체가 반드시 LI_t 를 안다고 가정할 필요는 없다. 잠자는 미녀의 경우 일요일에 미녀는 LI_t 를 알다가 다음 날부터는 LI_t 를 모르고 있다. 이 점에서 우리는 $C_t(LI_t) = 1$ 또는 $C_t(\sim LI_t) = 0$ 을 무작정 가정해서는 안 된다. 잠자는 미녀의 경우 $C_{20160104}(LI_{20160104}) = 1$ 이고, $C_{20160104}(\sim LI_{20160104}) = 0$ 이다. 하지만 $C_{20160105}(LI_{20160105}) = 1$ 이 아니며, $C_{20160105}(\sim LI_{20160105}) = 0$ 이 아니다.

이 논문의 한 심사자께서는 모든 명제 X 에 대해 무작정 $C_{20160105}(X) = C_{20160104}(X)$ 가 성립하는 것이 아닌가 하고 물었다. $C_{20160104}(LI_{20160104}) = 1$ 이고 $C_{20160105}(LI_{20160104}) = 0$ 이기 때문에, 언제나 다음이 성립한다.

$$C_{20160104}(X \& LI_{20160104}) = C_{20160104}(X)C_{20160104}(LI_{20160104})$$

$$C_{20160105}(X \& LI_{20160104}) = C_{20160105}(X)C_{20160105}(LI_{20160104})$$

이 두 식은 “참인 명제는 다른 모든 명제와 독립이다.”나 “거짓인 명제는 다른 모든 명제와 독립이다.”와 같은 논란이 많은 논제를 표현하고 있다. 분명히 말할 수 있는 것은 아무 명제 X 에 대해 무작정 $C_{20160105}(X) =$

하지만 2016년 1월 5일 어느 시점 $20160105+$ 에 미녀에게 “지금은 2016년 1월 5일이다.”라는 사실을 알려주었을 경우에는 다음과 같은 조건화 규칙을 써도 된다. 기존 정보는 그대로 유지되면서 새로운 정보 $LI_{20160105}$ 를 획득했기 때문이다.

$$C_{20160105+}(X) = C_{20160105}(X|LI_{20160105})^{16}$$

미녀에게 “지금은 2016년 1월 5일이다.” 대신에 “당신이 잠든 지 하루가 지났다.” 또는 그냥 “오늘은 월요일이다.”라고 말해도 된다.

5. 김남중의 해법

많은 저자들이 지표가 포함된 명제의 믿음직함 변화를 추적하려면 조건화 규칙을 바꾸어야 한다고 제안했다. 김남중은 기존 조건화 규칙을 수정해야 하며, 수정된 규칙에 따라 확률 계산을 할 경우 미녀의 답변이 1/2이라는 주장이 거부되어야 한다고 주장한다. 그는 자신의 조건화 규칙을 통해 다음을 이끌어낸다.¹⁷⁾ 아래에서 ‘H’는 “던진 동전이 앞면이 나온다.”이고 ‘ H_1 ’은 “H이고, 오늘은 월요일이다.”이다. ‘T’는 “던진 동전이 뒷면이 나온다.”이고, ‘ T_1 ’은 “T이고, 오늘은 월요일이다.”이며, ‘ T_2 ’는 “T이고, 오늘은 화요일이다.”이다.

$C_{20160104}(X)$ 가 성립하지는 않는다는 것이다. 예컨대 X 자리에 $LI_{20160104}$ 나 $LI_{20160105}$ 를 넣을 경우 이 등식은 성립하지 않는다. 하지만 Q나 자기 위치 정보와는 무관한 아주 많은 명제들이 있는데, 이런 명제 X에 대해 우리가 $C_{20160105}(X) = C_{20160104}(X)$ 라고 표현하는 것은 합당하다.

16) 미녀에게 $C_{20160105}(X) = C_{20160104}(X|LI_{20160105})$ 는 성립하지 않는다. 조건화가 이행되는 두 정보 상태를 사이에 정보가 누적되지 않았기 때문이다. 마찬가지 까닭에서 $C_{20160105+}(X) = C_{20160104}(X|LI_{20160105})$ 도 성립하지 않는다.

17) Kim (2009), p. 308. 그는 자신의 조건화 규칙을 “shifted Jeffrey conditionalization”이라 불렀다.

$$C_{\text{월}}(H) = C_{\text{월}}(\text{월요일에 } H\text{가 참이다} \mid \text{월요일에 깨어난다}) * C_{\text{월}}(\text{지금 깨어났고 지금은 월요일이다}) + C_{\text{월}}(\text{화요일에 } H\text{가 참이다} \mid \text{화요일에 깨어난다}) * C_{\text{월}}(\text{지금 깨어났고 지금은 화요일이다})$$

여기서 $C_{\text{월}}(\text{화요일에 } H\text{가 참이다} \mid \text{화요일에 깨어난다})$ 는 확실히 0이다. “월요일에 깨어난다.”는 확실히 참이기 때문에, $C_{\text{월}}(\text{월요일에 } H\text{가 참이다})$ 가 1/2이라면, $C_{\text{월}}(\text{월요일에 } H\text{가 참이다} \mid \text{월요일에 깨어난다})$ 도 1/2이다. 그리하여 김남중은 $C_{\text{월}}(H)$ 가 0에서 1/2 사이가 될 것이라고 결론 내린다.

나아가 김남중은 자신의 조건화 규칙을 통해 $C_{\text{월}}(H_1) = C_{\text{월}}(T_1)$ 을 증명할 수 있다고 주장한다. 이 증명을 위해 $C_{\text{월}}(\text{월요일에 } H\text{가 참이다} \mid \text{월요일에 깨어난다}) = C_{\text{월}}(\text{월요일에 } T\text{가 참이다} \mid \text{월요일에 깨어난다}) = 1/2$ 이라고 가정하고 있다. 그가 실제로 가정하고 있는 것은 $C_{\text{월}}(\text{월요일에 } H\text{가 참이다}) = C_{\text{월}}(\text{월요일에 } T\text{가 참이다}) = 1/2$ 이다. 여기서 $C_{\text{월}}(\text{월요일에 } H\text{가 참이다})$ 는 미녀가 일요일에 갖는 믿음직함이다. 그런데 “월요일에 H 가 참이다.”가 뜻하는 바가 무엇인지 잘 따져 보아야 한다. 만일 “월요일에 H 가 참이다.”가 “지금은 월요일이고 H 는 참이다.”를 뜻한다면, 미녀는 일요일에 자기 위치 정보를 알고 있기 때문에, $C_{\text{월}}(\text{월요일에 } H\text{가 참이다})$ 는 0이다. 김남중이 이를 의도한 것은 분명 아니다. 만일 “월요일에 H 가 참이다.”가 “ H 의 참값이 월요일에 참이다.”를 뜻한다면, $C_{\text{월}}(H\text{의 참값이 월요일에 참이다})$ 가 뜻하는 것이 무엇인지 잘 가늠되지 않는다. 그냥 $C(H|\text{지금은 월요일이다})$ 를 의미하는 것처럼 보인다. 여기서 아래 첨자 ‘일’을 지운 것은, 미녀가 일요일에 “지금은 월요일이다.”가 거짓이라는 것을 알고 있으며, 그래서 $C_{\text{월}}(\text{지금은 월요일이다}) = 0$ 이기 때문이다. 이 경우 $C_{\text{월}}(H|\text{지금은 월요일이다})$ 는 정의되지 않거나 1이다.

김남중의 증명이 옳다 하더라도, 그가 굳이 기존의 조건화 규칙을 수정해야 할 분명한 이유가 있었을까? 그는 지표가 포함된 명제에 기존 조건화 규칙을 적용할 경우 이상한 결과가 나오는 사례를 제시한

다. 지표 ‘지금’이 포함된 두 명제 P와 R을 생각해 보자. 명제 P는 “지금 비가 내린다.”이고 명제 R은 “강수 현상이 지금 일어나고 있다.”이다. 김남중은 몇 가지 확률 규칙과 조건화 규칙을 통해 다음과 같은 추론을 이끌어내었다.

P이면 R이라는 점이 앞서 참이었다.

따라서 P이면 R이라는 점이 지금도 참이다.

김남중은, 2016년 1월 1일 오후 2시에 비가 내리다가 가령 같은 날 오후 2시 10분에 비가 내리지 않는 경우 저 추론은 타당하지 않다고 주장한다.¹⁸⁾ 내 생각에, 이 추론이 부당하다면 그것은 전제에 나오는 명제 P와 R이 결론에 나오는 P와 R과 다르기 때문에 생긴 일이다. 앞서 언급된 명제는 가령 다음과 같이 표현되어야 한다.

만일 2016년 1월 1일 오후 2시에 비가 내린다면 강수 현상이

2016년 1월 1일 오후 2시에 일어나고 있다는 점이 앞서 참이었다.

따라서 만일 2016년 1월 1일 오후 2시에 비가 내린다면 강수 현

상이 2016년 1월 1일 오후 2시에 일어나고 있다는 점은 지금도

참이다.

추론에 나오는 명제들을 이와 같이 고칠 경우 이 추론은 타당하다. 기존 조건화 규칙과 확률 규칙들을 유지한 채, 지표가 포함된 명제들을 다룰 수 있는 방법이 별도로 존재할 것이다.

잠자는 미녀 문제는 조건화 규칙을 수정해야 하는 예외 사례가 아니다. 오히려 조건화 규칙을 사용할 수 없는 사례이다. 미녀에게 일요일 상황과 월요일 상황은 정보가 쌓이는 경우가 아니라 일부 정보가 사라지고 일부 정보가 생기는 경우이다. 미녀는 일요일에서 월요일로 가면서 기억이 지워져 정보를 잊는 과정을 밟는다. 만일 사라진 정보가 미녀의 믿음직함 계산에 관련된 정보라면, 월요일에 미녀는 조건화 규칙

¹⁸⁾ Kim (2009), p. 301.

을 통해 믿음직함을 갱신하는 것이 아니라, 자신이 현재 알고 있는 정보만으로 새로 믿음직함을 재설정해야 한다. 여러 저자들이 수정된 조건화 규칙을 제안했는데, 정보가 상실되는 경우에도 그런 규칙을 적용하는 것은, 내 생각에, 별도의 정당화가 요구된다.¹⁹⁾

6. 기적 같은 일의 하찮은 증거

미녀가 깨어났을 때, 문제의 설정 자체를 기술하는 배경 정보 BI, 지금 자신에게 물음이 주어지고 있다는 사실 Q, 지금이 잠들기 전 일요일이 아니라는 사실 \sim LI일 등을 알고 있다. 미녀의 이러한 정보 상태를 O라고 하자. 미녀가 정보 상태 O에 있을 때 명제 X에 대해 갖는 믿음직함 합수를 $C_o(X)$ 라고 쓰자. 1/3주의자들은 대체로 동전을 일요일에 던지는 경우와 월요일에 미녀를 재운 뒤에 던지는 경우가 같다고 가정한다.²⁰⁾ 그들은 이 가정으로부터 지금이 월요일이라는 조건에서

19) 당연히 정보가 상실되는 경우에 적용할 수 있는 조건화 규칙을 개발하는 것은 환영받을 일이다. 그것을 개발하는 것은 곧 그것의 정당화를 이미 포함하고 있다. 하지만 그 규칙은 단순히 시간 지표를 다루는 조건화 규칙이어서는 안 되며, 주체의 정보 상태가 가령 $\{A, X\}$ 에서 $\{A, Y\}$ 로 바뀔 때 사용할 수 있는 조건화 규칙이어야 한다. 이 규칙은 주체의 정보 상태가 $\{A, X\}$ 에서 $\{A, \sim X\}$ 로 바뀔 때도 사용할 수 있어야 한다. 나아가 주체의 정보 상태가 $\{A, X, \sim Y\}$ 에서 $\{A, \sim X, Y\}$ 로 바뀔 때도 사용할 수 있으면 좋겠다. 이런 조건화 규칙이 개발될 경우, X나 Y에 지표 정보 LI가 포함되는 경우를 적용시켜 볼 만하다. 물론 지표 정보는 매우 특수한 정보이기 때문에 이것을 위한 별도의 조건화 규칙이 있다 해도 놀라운 일은 아니다.

20) Elga (2000), pp. 144-45. 1/3주의에 반대하는 루이스조차도 이 가정을 받아들인다. Lewis (2001), p. 172. 이 논문의 한 심사자는 1/2주의자나 1/3주의자 모두 이 가정을 받아들인다고 주장한다. 내가 아는 한 모든 1/3주의자는 이 가정을 받아들인다. 하지만 모든 1/2주의자가 그 가정을 받아들일 필요는 없다. 나는 이 가정 자체가 잘못되었다고 생각한다. 이 가정은

동전이 앞면이 나올 믿음직함이 $1/2$ 이라고 주장한다. 월요일에 미녀를 재운 뒤에 동전을 던지는 경우를 상정해 보자. 오늘이 월요일이라는 공공 시간 정보 LI 를 줄 경우, 미녀는 배경정보 BI 로부터 동전이 아직 던져지지 않았다는 사실을 추론할 수 있고, 이로부터 $C_o(H|오늘은 월요일이다) = 1/2$ 을 이끌어낼 것이다. 그 다음 $C_o(T_1|T) = C_o(T_2|T)$ 를 가정함으로써 $C_o(H) = 1/3$ 을 도출해낸다.

동전을 일요일 밤에 던지는 경우, $C_o(T_1|T) = C_o(T_2|T)$ 는 거의 명백하다.²¹⁾ 하지만 동전을 일요일에 던지는 경우와 월요일에 미녀를 재운 뒤에 던지는 경우가 같다는 가정은 의심의 여지가 있다. 이 가정은 납득할 수 없는 결과를 낳는다는 것을 보여줄 수 있다.²²⁾ $1/3$ 주의자들이 자신의 계산 결과가 이상한 귀결을 낳는다는 것을 보지 못했던 까닭은, 앞면과 뒷면이 나올 가능성성이 똑같은 상황에서, 미녀에게 앞면이 나왔고 월요일인 상황, 뒷면이 나왔고 월요일인 상황이 동등한 개연성을 갖고 발생한다고 착각하기 매우 쉽기 때문이다. 하지만 앞면이 나오라고는 거의 믿을 수 없는 치우친 동전을 사용하여 똑같은 계산을 해보면 매우 이상한 결과가 나온다는 것을 알 수 있다.

$1/3$ 주의자들의 계산이 납득할 수 없는 귀결을 갖는다는 것을 보이기 위해 완전히 일반화된 잠자는 미녀 문제를 생각해 보자. 일요일에 던진 동전이 앞면이 나올 확률은 h 이고 뒷면이 나올 확률은 t 이다. 물론 $h + t = 1$ 이다. 앞면이 나오면 잠자는 미녀를 단 한 번만 깨운다. 하지만 뒷면이 나오면 깨우고 묻고 기억을 지우고 재우는 과정을 N 번 거듭한다. N 이 아무리 크다 하더라도 미녀는 노화를 느끼지 못한 채 그 만큼 오랫동안 살 수 있다고 가정한다. 엘가의 방식으로 계산하기 위

내가 ‘가능주의 계산 방식’이라고 부른 것과 관련이 있다. 김명석 (2016).

21) 동전을 월요일 밤에 던지는 경우에도 이것이 명백하다고 자신 있게 말하기는 어렵다. 동전을 월요일 밤에 던지는 경우에 T 를 “던질 동전이 뒷면이 나오거나 던진 동전이 뒷면이 나온다.”라고 써야 할 것 같다.

22) 김명석 (2012b). $1/3$ 주의자는 이 납득할 수 없는 결과를 받아들여야 하겠지만 나는 이 결과를 받아들이지 않는다. 나아가 동전을 일요일에 던지는 것과 첫째 물음을 마친 뒤에 던지는 것 사이에 아무 차이가 없다는 엘가와 루이스의 가정을 의심한다. 김명석 (2012a).

해 다음 명제들을 기호로 약칭한다.

Q_n = 미녀는 지금 물음을 받고 있으며 지금 이 물음은 n 번째 물음이다.

T_n = T 이고 Q_n

앞에서 말했듯이 미녀가 깨어났을 때 BI , Q , $\sim LI_{\text{일}}$ 등을 얻는다. 명제 X 에 대한 함수 $C_o(X)$ 는 정보 상태 $\{BI, Q, \sim LI_{\text{일}}\}$ 에 있는 미녀의 믿음직함 함수이다. $C_o(X|Q_1)$ 은 간단히 $C_+(X)$ 라고 약칭한다. 함수 $C_+(X)$ 는 정보 상태 $\{BI, Q, \sim LI_{\text{일}}, LI_+\}$ 에 있을 때 미녀의 믿음직함 함수이다. LI_+ 는 첫째 물음이 주어지는 때 미녀의 주관 시간을 상호주관 시간과 동기화할 수 있는 정보이다. 간단히 “지금은 첫째 물음이다.”라고 알려주기만 해도, 미녀는 자신의 주관 시간을 상호주관 시간과 동기화할 수 있다. 정보 상태 O 에서 $+$ 로 건너가면서 미녀가 갖고 있던 기존 정보가 사라지지 않고 새로운 정보가 쌓인다.

우리가 일요일에 던질 동전은 매우 치우친 동전이고, 이 동전이 뒷면이 나올 가능성 t 는 가령 10^{-m} 이다.²³⁾ 물론 미녀는 이 사실은 일요일에도 월요일에도 알고 있다. 우리는 나중에 m 을 무한대로 보낼 텐데, m 이 커질수록 앞면이 나올 가능성은 1에 수렴하고, 뒷면이 나올 가능성은 0에 수렴한다. 따라서 m 을 무한대로 보내는 극한의 경우 뒷면이 나올 가능성은 아예 없다. 우리는 동전을 단 한 번만 던질 것이다. 우리는 “만일 $C(X)$ 가 0에 수렴하는 값이라면, 단 한 번 시행에서 X 가 참이 될 것이라 믿는 것은 완전히 합당하지 않다.”는 것을 받아들인다.

²³⁾ 동전이 뒷면이 나올 가능성이 10^{-m} 이라는 것은 다음을 뜻한다. 1에서 10^m 까지의 숫자들이 하나씩 적힌 10^m 개 종이들 중에서 하나의 종이를 뽑는다고 했을 때, 10^m 개 숫자들 중에서 특정 한 숫자, 가령 7이 나올 가능성과 같다. 이러한 일의 믿음직함은 10^{-m} 이다. 일요일에 동전을 던진다고 말하는 대신에, 1에서 10^m 까지 숫자가 하나씩 적힌 10^m 개 종이들 중에서 한 종이를 무작위로 뽑는다고 말해도 된다. 실제 종이로 실험할 수 없다면 컴퓨터로 1에서 10^m 까지 아무 숫자를 하나 무작위로 출력해도 좋다.

1/3주의자들의 가정에 따르면, 동전을 일요일에 던지는 경우와, 첫 번째 물음을 마치고 미녀를 다시 재운 뒤에 던지는 경우가 같다. 따라서 그들은,

$$(1) C_+(H_1) = C_+(H) = h = 1 - t$$

를 받아들인다. $h = C_+(H) = C_o(H|Q_1) = C_o(H \& Q_1)/C_o(Q_1) = C_o(H_1)/C_o(Q_1)$ 이고, $C_o(Q_1) = C_o(H_1) + C_o(T_1)$ 이기 때문에, $C_o(T_1) = tC_o(H_1)/h$ 이다. 그리고 $C_o(T_1) = C_o(T_2) = \dots = C_o(T_N)$ 이고, $C_o(H_1 \vee T_1 \vee T_2 \vee \dots \vee T_N) = 1$ 이기 때문에, $C_o(H_1) + NC_o(T_1) = 1$ 이다. 두 계산 결과로부터

$$(2) C_o(T_1) = t/\{tN+h\}$$

$$(3) C_o(T) = NC_o(T_1) = tN/\{tN+h\}$$

를 얻을 수 있다.식 (2)와 (3)은 완전히 일반화된 결과이기 때문에 우리는 N 을 가령 100^m 으로 잡아도 무방하다.

미녀는 이제 깨어나 “던진 동전이 뒷면이 나왔을 당신의 믿음직함은?”이라는 물음을 받고 있다. 1/3주의자의 계산식 (3)에 따르면, 그의 믿음직함은 $10^m/(10^m + 1 - 10^m)$ 이다. m 을 매우 큰 수로 키울 경우 이 값은 1에 수렴한다. m 이 증가할 경우, 아무런 증거가 없는 상태에서, “던진 동전이 뒷면이 나왔다.”에 대한 미녀의 믿음직함은 0으로 수렴해야 한다. 하지만 자신이 지금 물음을 받고 있다는 증거 Q 가 주어질 경우, 그는 물음을 엄청나게 많이 받는 세계에 들어왔다고 믿게 되고, “던진 동전이 뒷면이 나왔다.”에 대한 그의 믿음직함은 1로 갱신된다. 나는 이것이 매우 이상한 귀결이라고 생각한다. 엄청나게 많이 여러 번 물음을 받고 있는 세계에 자신이 들어오기 위해서, 그 전에 실제 세계에서 거의 일어날 수 없는 일이 벌어져야 한다(김명석 2012b).

1/3주의자들의 가정 또는 귀결을 조금 더 따져보자. 아무 증거가 없

는 상태에서 $C(T)$ 가 \emptyset 로 수렴하는데도 $C_o(T)$ 는 오히려 1로 수렴하는 것이 어떻게 가능한가? 일어나리라고 거의 믿을 수 없는 기적 같은 일이 확실하게 일어났을 것이라고 믿기 위해서는 반드시 모종의 증거가 있어야 한다. 그 증거는 배경 정보인가, 지금 물음이 주어진다는 사실인가, 지금이 일요일이 아니라는 사실인가? 이 정보들은 T 가 확실히 일어났을 것이라고 믿도록 하는 강력한 증거가 되는가?

	일요일	어느 날 깨어났을 때
미녀의 정보 상태	$BI, LI_{\text{일}}, \sim Q$	$BI, \sim LI_{\text{일}}, Q$
T 의 믿음직함	거의 0	거의 1

일단 BI 가 주어진 상태에서는, Q 로부터 $\sim LI_{\text{일}}$ 을 알 수 있고, 거꾸로 $\sim LI_{\text{일}}$ 로부터 Q 를 알 수 있다. Q 와 $\sim LI_{\text{일}}$ 은 논리 관점에서 동치이고, 이들은 다른 두 정보가 아니라 한 정보로 간주해도 좋다. 결국 Q 라는 정보 하나만으로 T 가 벌어졌다는 믿음의 강도를 증가시킨다고 보아야 한다.

Q 라는 증거가 어떻게 T 를 매우 강하게 입증할 수 있는가? Q 가 H 를 반증하는 대신 T 를 매우 강하게 입증하는 이유는 무엇인가? 그것은 H 가 일어나면 고작 한 번의 물음만 주어지지만, T 가 일어나면 엄청나게 많이 여러 번 물음이 물어지기 때문이다. 결국 1/3주의자의 귀결에 따르면, 미녀 자신이 물음을 받고 있다는 바로 이 증거는, 물음을 많이 받고 있는 세계에 자신이 이미 들어와 있음을 매우 강하게 뒷받침하는 증거로 쓰일 수 있다. 하지만 $C_o(Q|H) = C_o(Q|T) = 1$ 이다. 다시 말해 $C_o(X)$ 를 계산하는 미녀의 시공간에서 언제 어디서나 $C_o(Q) = 1$ 이다. 그래서 $C_o(H|Q) = C_o(H)$ 이고, $C_o(T|Q) = C_o(T)$ 이다. Q 는 H 를 반증하지도 입증하지도 않으며, T 를 반증하지도 입증하지도 못한다. 만일 H 가 나왔다면, T 가 나와 물음이 주어지는 엄청나게 많은 시점에서 미녀는 물음을 받지 못한다. 하지만 바로 그런 시점들에서 미녀는 의식을 갖지 못하기 때문에 그는 믿음직함을 매길 어떤 유형의 시공간도 갖지

못한다.²⁴⁾ 다시 말해 믿음직함을 매기는 미녀의 가능한 모든 시공간에서 Q는 꽉 차게, 빈틈없이 일어나고 있다.

Q가 H를 반증하고 T를 입증하기 위해서는 H가 나올 경우 Q가 거짓이 되는 경우가 존재해야 한다. 그런 경우가 전혀 존재하지 않는다면, 지금 물음이 물어지고 있다는 바로 그 사실은 뒷면이 나왔다는 사실을 뒷받침할 수 없을 것이다. 단 하나의 하찮은 증거, 모든 시점에서 흔하게 벌어지는 그런 증거 때문에, 일어날 가능성성이 \varnothing 로 수렴해서 거의 일어날 것 같지 않은 일이 이미 일어났다고 거의 확실하게 믿는 것이 과연 합당할까? 1/3주의는 어디서나 찾을 수 있는 하찮은 증거가 기적 같이 드문 일을 입증하는 증거가 될 수 있다는 귀결을 갖는다. 이 귀결을 받아들일 수 없다면 1/3주의의 기본 가정, 동전을 던지는 시점이 일요일 밤이든 월요일 밤이든 미녀의 믿음직함에 차이가 없다는 가정을 버려야 한다.

7. 나오는 말

잠자는 미녀 문제는 믿음직함의 본성에 대해 많은 함축을 갖는다. 1/3주의자들은 믿음직함의 변경을 관장하는 조건화 규칙을 바꾸어야 한다고 주장하곤 한다. 몇몇 저자들은 ‘나’, ‘지금’, ‘여기’ 등과 같은 지표가 포함된 문장을 다루는 특별한 유형의 조건화 규칙을 제안했다. 지표가 포함된 명제들 사이의 조건화를 다루기 위해서, 조건화 규칙을 수정하는 대신에, 명제에서 지표를 최소화하는 길을 선택할 수 있다. 명제에서 지표를 줄이기 위해 다음과 같은 점을 제안했다. 명제 태도를 갖는 사람은 무엇보다 자신만의 주관 시공간 좌표를 갖는다. 통상 그는 자신의 주관 시공간 좌표를 상호주관 시공간 좌표와 동기화하는

²⁴⁾ 만일 미녀가 자신이 공공의 시공간과 연결되어 있다고 스스로 가정한다면, $C(Q) = 1$ 이 아닐 수 있다. 앞면이 나와 화요일에 계속 자고 있을 때 미녀에게 물음이 물어지지 않는다. 하지만 이 때 미녀의 믿음직함 $C(Q)$ 는 미녀가 깨어났을 때 믿음직함 $C_o(Q)$ 와 구별해야 한다.

데 필요한 정보를 경험과 기억을 통해 입수한다.

잠자는 미녀의 경우, 일요일 밤 이후 깨어났을 때, 자신의 주관 시공간 좌표를 상호주관 시공간 좌표와 동기화하는 데 필요한 정보를 잃는다. 조건화 규칙은 정보가 누적되는 경우이거나 상실되지 않는 경우에 적용되기 때문에, 미녀가 일요일에서 월요일로 넘어갈 때 조건화 규칙을 적용하는 것은 옳지 않다. 일요일과 월요일 사이에 조건화 규칙을 쓰지 못한다면 언제 이 규칙을 사용할 수 있는가? 월요일이든 화요일이든, 미녀에게 상호주관 시공간 정보를 줄 때는 그 규칙을 사용할 수 있다. 왜냐하면 그 전후에 미녀는 기존 정보를 그대로 유지한 채 새로운 정보를 입수하기 때문이다.

거의 모든 1/3주의자는 동전을 일요일에 던지는 경우와 월요일에 미녀를 재운 뒤에 던지는 경우가 같다고 가정한다. 우리는 1/3주의의 이 가정이 납득할 수 없는 귀결을 갖는다는 것을 살펴보았다. 한 명제의 믿음직함이 거의 0에 육박하는 경우에도, 미녀는 자신이 물음을 받고 있다는 증거가 주어지는 것만으로 그 명제의 믿음직함에 거의 1의 값을 부여해야 한다. 기적 같은 일의 증거라는 것이 잠에서 막 깨어난 미녀의 시공간에서는 언제나 늘 반드시 일어나는 일에 불과하다. 이 야릇한 귀결을 받아들이지 않기 위해서 우리는 미녀를 깨우기 전에 동전을 미리 던지는 경우와 첫 물음을 묻고 미녀를 다시 재운 뒤에 던지는 경우가 같다는 가정을 버려야 한다. 그리하여 월요일에 미녀에게 상호주관 시공간 정보를 준 후에는 일요일에 던진 동전이 앞면이 나왔을 확률이 1/2로 바뀐다는 가정을 버려야 한다.

참고문헌

- 김명석 (2012a), 「동전을 던진 후 미녀를 깨우다」, 『논리연구』 15집 1호, pp. 17-52.
- _____ (2012b), 「세계의 문턱과 사건의 되풀이」, 『철학적 분석』 26호, pp. 75-90.
- _____ (2016), 「믿음직함과 가능 세계: 잠자는 미녀의 경우」, 『범한 철학』 80집, pp. 91-111.
- 김한승 (2009), 「비개념적 내용으로서의 지표적 내용: ‘잠자는 미녀’ 문제에 대한 관점주의적 대답」, 『철학적 분석』 20호, pp. 119-40.
- 송하석 (2011), 「잠자는 미녀의 문제, 그의 대답은」, 『논리연구』 14집 1호, pp. 1-22.
- 이영의 · 박일호 (2015), 「베이즈주의 인식론」, 『과학철학』 18권 2호, pp. 1-14.
- Bostrom, N. (2007), “Sleeping Beauty and Self-location: A Hybrid Model”, *Synthese* 157(1): pp. 59-78.
- Davidson, D. (1989), “What is Present to the Mind”, in Davidson, D. (2001), *Subjective, Intersubjective, Objective*, New York: Oxford University Press, pp. 53-67.
- Elga, A. (2000), “Self-locating Belief and the Sleeping Beauty Problem”, *Analysis* 60: pp. 143-47.
- Groisman, B. (2008), “The End of Sleeping Beauty’s Nightmare”, *The British Journal for the Philosophy of Science* 59: pp. 409-16.
- Jenkins, C. (2005), “Sleeping Beauty: A Wake-Up Call”, *Philosophia Mathematica* 13(2): pp. 194-201.
- Lewis, D. (1979), “Attitudes De Dicto and De Se”, *The Philosophical Review* 88(4): pp. 513-43.
- _____ (2001), “Sleeping Beauty: Reply to Elga”, *Analysis* 61: pp.

171-76.

- Lewis, P. (2010), “Credence and Self-Location”, *Synthese* 175(3): pp. 369-82.
- Meacham, C. (2008), “Sleeping Beauty and the Dynamics of de Se Beliefs”, *Philosophical Studies* 138(2): pp. 245-69.
- Monton, B. (2002), “Sleeping Beauty and the Forgetful Bayesian”, *Analysis* 62(1): pp. 47-53.
- Karlander, K. & Spectre, L. (2010), “Sleeping Beauty Meets Monday”, *Synthese* 174(3): pp. 397-412.
- Kim, N. (2009), “Sleeping Beauty and Shifted Jeffrey Conditionalization”, *Synthese* 168(2): pp. 295-312.
- Quine, W. V. (1969), “Propositional Objects”, *Ontological Relativity and other Essays*, Columbia University Press, pp. 139-60.
- Schulz, M. (2010), “The Dynamics of Indexical Belief”, *Erkenntnis* 72(3): pp. 337-51.
- Weintraub, R. (2004), “Sleeping Beauty: A Simple Solution”, *Analysis* 64(1): pp. 8-10.
- White, R. (2006), “The Generalized Sleeping Beauty Problem: a Challenge for Thirders”, *Analysis* 66: pp. 114-19.

논문 투고일	2016. 02. 11
심사 완료일	2016. 03. 11
제재 확정일	2016. 03. 23

A Subjective Spacetime of Sleeping Beauty

Myeongseok Kim

As lessons of the Sleeping Beauty problem various theses have been proposed, from the view that we must modify Jeffrey's conditionalization to the view that we need to restrict the principal principle of Lewis. This paper, while rejecting such proposals, asserts the followings. Firstly, subject who has the probability function or the credence function resets her own space and time. While, in common case, subject's spacetime is easily synchronized with intersubjective spacetime, or public spacetime, with the help of experiences, but not in the case of Sleeping Beauty. Secondly, when Sleeping Beauty awakes she does not obtain but loses some information, such as the information that is required to synchronize her own subjective spacetime with public one. Thirdly, if Sleeping Beauty loses information, she should not use Bayesian conditionalization. Fourthly, thirder's calculation such as Elga's has strange implication that a miraculous hypothesis can be confirmed by a very trivial evidence.

Keywords: conditionalization, confirmation, credence, information, probability, self-locating belief, Sleeping Beauty