논리연구 14-3(2011) pp. 27～58

조건화와 고차 믿음 갱신* ** ***

박일호

【요약문】 본 논문은 대표적인 믿음의 정도 갱신 규칙인 조건화를 다룬다. 지금껏 조건화는 다양한 형태로 제시되었다. 그 중 가장 단순한 것이 베이지안 조건화다. 하지만 베이지안 조건화는 그 적용 범위가 넓지 않기 때문에, 여러 방식으로 일반화되었다. 본 논문은 그 중에서 제프리 조건화와 이차 조건화라고 불리는 것을 고차 믿음의 정도와 관련해 비교 평가한다. 우선 2절에서 제프리 조건화와 이차 조건화가 명시적으로 제시된다. 3절과 4절에서 고차 믿음의 정도의 갱신과 관련해 제프리 조건화와 이차 조건화 중요한 차이점 두 가지가 언급된다. 나는 3절에서 제프리 조건화는 반영 원리와 충돌하지만, 이차 조건화는 그렇지 않다는 것을 보일 것이다. 그리고 4절에서 제프리 조건화는 무어의 부조리를 야기할 수 있지만, 이차 조건화는 그렇지 않다는 것을 증명한다. 결과적으로 나는 고차 믿음의 정도와 관련해 이차 조건화가 제프리 조건화보다 더 훌륭한 규칙이라고 주장할 것이다.

【주요어】 제프리 조건화, 이차 조건화, 반영 원리, 무어의 부조리

---

* 접수일자: 2011.8.25. 심사 및 수정완료일: 2011.9.19. 게재확정일: 2011.9.23.

** 이 논문은 2009년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2009-351-A00205).

*** 이 논문의 초고를 읽고 많은 도움을 주신 Alan Hájek, 김남중, 김동현, 여영서, 정인교, 최원배 교수님들께 감사의 말씀을 전한다. 더불어 엄격하고 신랄한 논평을 해주신 익명의 심사위원 선생님들에게도 감사를 드린다.

1. 서론: 믿음 갱신 규칙

확률주의자들(probabilists)은 우리의 믿음은 정도(degrees)의 문제라고 말한다. 즉 우리의 믿음 상태는 확신 혹은 불신 둘 중의 하나가 아니라는 것이다. 우리는 어느 정도로 무언가를 믿으며, 어느 정도로 무언가를 불신한다. 그들에 따르면 이런 믿음의 정도는 확률에 의해서 가장 적합하게 표상될 수 있다. 본 논문에서 나는 이런 주장을 비판하지 않는다. 즉 아래의 논의는 믿음이 정도의 문제이며, 그 믿음의 정도는 확률계산규칙을 만족해야 한다는 것을 가정하고 진행된다.

한편, 우리의 믿음의 정도는 고정된 것이 아니다. 시간이 흐르고, 새로운 정보가 도입됨에 따라서 기존의 믿음의 정도는 새로운 믿음의 정도로 수정 혹은 갱신된다.[2] 그리고 그렇게 수정된 믿음의 정도는 다른 믿음의 정도에 영향을 준다. 여기 어떤 물체가 놓여 있다. 이 물체가 녹색이라는 명제를 E로 나타내자. t 시점에 어두운 불빛 아래에서 그 물체를 관찰한 결과, 나는 E를 0.5정도로 믿게 되었다고 하자. 그런데, t+1 시점에 불빛이 좀 더 밝아졌다. 이 밝아진 불빛 아래에서 그 물체를 다시 관찰한 결과, E에 대한 믿음은 0.8로 수정되었다. 즉 경험 혹은 관찰에 의해서 E에 대한 믿음(혹은 믿음의 정도)이 직접적으로 수정된 것이다.[3] 하지만 모든 믿음이 경험의 직접적인 영향 아래에서 수정되는 것은 아니다. 가

---

2) 몇몇 철학자들, 믿음 수정 이론(Belief Revision Thoery)을 다루는 학자들은 수정(revision)과 갱신(update)을 엄격하게 구분하기도 한다. 가령, 세계의 변화에 기인한 믿음 변화를 '갱신'이라고 하고, 믿음 체계의 내적 문제(예를 들어 일관성)에서 비롯된 믿음 변화를 '수정'이라고 말할 수도 있다. 하지만 이런 구분은 본 논의에 있어 지엽적이다. 나는 아래 글에서 '수정' 혹은 '갱신'을 엄밀하게 구분하지 않고 사용할 것이다.

3) 아래에서 '믿음 수정', 혹은 '믿음 갱신'은 모두 '믿음의 정도의 수정' 혹은 '믿음의 정도의 갱신'을 축약한 것이다.

령, 내가 진화론은 1에 가까운 정도로 믿고 있다고 할 때, 그 믿음의 정도는 (E에 대한 믿음의 정도처럼) 직접적인 경험에 의해서 형성된 것이라고 말하기 어렵다. 진화론에 대한 그 믿음의 정도는 (E에 대한 믿음의 정도처럼) 경험에 의해 직접적으로 수정된 믿음의 정도로부터 추론된 것이다. 이렇게 (경험의 직접적인 영향을 받는 것이 아니라) 다른 믿음의 정도로부터 추론된 믿음 변화는 추론적 믿음 갱신, 혹은 간접적 믿음 갱신이라고 불린다. 그리고 이 논문에서 내가 주로 다루려는 것이 바로 이 추론적 믿음 갱신이다.

경험에 의해서 몇몇 명제에 대한 믿음의 정도가 직접적으로 수정되었다고 하자. 그럼 (직접적으로 수정된 믿음의 정도 이외에) 다른 명제들에 대한 믿음의 정도는 어떻게 (추론적으로) 수정되어야하는가? 이 질문에 답하기 위해서 확률주의자들은 몇 가지 규칙을 제안하고 정당화한다. 이런 규칙은 보통 믿음 갱신 규칙(Belief Updating Rule)이라고 불린다. 나는 아래에서 이 믿음 갱신 규칙으로 가장 많이 언급되는 조건화(conditionalization)를 다룰 것이다.

조건화는 다양한 형태로 제시된다. 그 중 가장 대표적인 것이 다음 베이지안 조건화(Bayesian Conditionalization, 이하 BC)다.

> **베이지안 조건화(Bayesian Conditionalization, BC).** 어떤 행위자가 시점 t+1에 경험에 의해서 E라는 명제만을 확신하게 되었을 때, 모든 명제 X에 대해서 다음이 성립한다:
> $P_{t+1}(X)=P_t(X|E)$. (단, $P_t(E)>0$ 일 때.)

위에서 P_t는 해당 행위자의 사전 확률 함수이며, P_{t+1}은 그의 사후 확률 함수다.[4] 처음으로 이 믿음 갱신 규칙을 제안한 사람은

---

4) 이 논문에서 '확률'이라는 용어는 모두 주관적 확률을 가리킨다. 즉 이 글에서 등장하는 확률은 모두 믿음의 정도로 해석된다. 따라서, '확률 함수'라는 용어는 믿음의 정도 함수를 말한다. 또한 '사전 확률 함수'라는 것은 행위자가 해당 경험 이전에 가지고 있었던 확률 함수를 말한다. 물론, '사후

프랭크 램지(Frank Ramsey)이며, 이후 많은 철학자들이 다양한 방식으로 이 규칙을 정당화했다.[5] 하지만, 여기서 우린 이 규칙의 적용 범위가 그리 넓지 않다는 것을 주목할 필요가 있다. BC의 전건이 말하듯이, 이 규칙을 적용하기 위해서는 경험 이후에 우린 어떤 명제를 확신해야 한다. 만약 무언가를 경험했지만 어떤 것도 새로이 확신하지 않게 되었다면, 우리는 BC를 이용해서는 다른 명제에 대한 믿음의 정도를 수정할 수 없다. 가령, 앞에서 언급된 녹색처럼 보이는 어떤 물체를 관찰하는 사례가 보여주듯이, 경험 이후에 E를 확신하게 되지 않고 단지 0.8정도로 믿게 된 경우에, BC는 적용될 수 없다.

그럼 어떻게 BC의 적용범위를 확장할 것인가? 지금껏 BC를 확장한 조건화 규칙으로 제시된 것은 두 가지다.[6] 첫 번째는 제프리

확률 함수'는 그가 해당 경험 이후에 가지게 된 확률 함수를 말한다.

5) 프랭크 램지는 1926년에 발표된 "Truth and Probability"라는 논문에서 처음으로 조건화를 제시한다. 그리고 조건화에 대표적인 정당화는 더치북을 이용하는 것이다. 조건화에 대한 더치북 논증은 데이비드 루이스가 처음 제안했다고 한다. 그 논증은 Teller(1973)에서 발견할 수 있다. 이외에 다양한 방법으로 조건화는 정당화되었다. 더치북 논증 이외에 대표적인 논증으로 다음 다섯 가지를 제시할 수 있다. (i) 대칭성 논증(Hughes and van Fraassen (1984); van Fraassen (1986; 1989)를 보라.); (ii) 반영 원리(Reflection Principle)에 호소하는 논증 (van Fraassen(1995; 1999), Maher(1993)를 보라.); (iii) 최소 정보 원칙(Principle of Minimum Information)에 호소하는 논증(Williams(1980)을 보라.); (iv) 약한 우도 원칙(Weak Likelihood Principle)에 호소하는 논증(Joyce(2003)을 보라.); (v) 기대 인식적 효용의 최대화 원칙(Principle of Maximization of Expected Epistemic Utility)에 호소하는 논증(Greaves and Wallace(2006)을 보라.) 위 논증들은 베이지안 조건화에 대한 것만은 아니다. 이들은 모두 아래에서 설명될 제프리 조건화를 옹호하는 논증이라고도 할 수 있다.

6) 물론 우리의 믿음을 갱신하는 규칙으로 조건화만 있는 것은 아니다. 예를 들어, 최소 정보 원칙(Principle of Minimum Information)을 만족시키면서 우리 믿음을 수정해야 한다는 것이 조건화를 대체할 수 있는 대표적인 규칙 중에 하나다. 여기서 최소 정보 원칙이란, 경험에 의해서 특정 명제의 믿음

조건화(Jeffrey Conditionalization, 이하 JC)이며, 두 번째는 이차 조건화(Second Order Condtionalization, 이하 SC)다.[7] 이 논문의 주요 목적은 이 두 방법을 비교 평가하는 것이다. 이를 위해서 나는 2절에서 이 두 가지 규칙을 명시적으로 제시하고, 둘 사이의 관계를 설명할 것이다. 그 후, 3절과 4절에서 고차 믿음의 정도를 갱신할 때, SC가 JC보다 더 우수한 규칙이라는 것을 논증할 것이다. 이 논증에 있어, 반 프라센의 반영 원리(Reflection Principle)와 무어의 부조리(Moorean Absurdity)가 핵심적인 역할을 한다.

2. 제프리 조건화와 이차 조건화

어떤 행위자가 있다고 하자. 그는 무언가를 경험한 뒤, 명제 E에 대한 믿음의 정도를 직접적으로 수정했다. 그럼 임의의 명제 X에 대한 믿음의 정도는 어떻게 수정되어야 하는가? 만약 E의 수정된 믿음의 정도가 1이라면, 그는 BC를 이용해서 다른 명제들의 믿음의 정도를 수정하면 될 것이다. 하지만, E의 수정된 믿음의 정도가 1이 아니라면 어떻게 해야 하는가? 앞 절에서 언급된 것처럼, 나는 이 질문에 대한 답변으로 제프리 조건화와 이차 조건화 두 가지를 검토할 것이다. 제프리 조건화는 가중 평균(weighted average)을 이용해 BC의 적용범위를 확장한 것이다. 이차 조건화는 고차 믿음의

이 수정되었을 때, 원래 믿음의 정도들에 가장 근접하게 다른 믿음의 정도들을 수정해야 한다는 것이다. 최소 정보 원칙에 대한 논의로는 Williams(1980)를 보라.

7) 제프리 조건화는 Jeffery(1983)에서 처음으로 제안되었다. 한편, 이차 조건화는 제프리 조건화의 대안으로 여러 논문(Jeffery(2002), Skyrms(1980a; 1980b))에서 소개되었다. '이차 조건화'라는 용어는 Skyrms(1980b)에서 발견할 수 있다. 한편 제프리는 이차 조건화를 'Softcore empiricism'이라고 부르기도 한다.이것은 'Hardcore empiricism'과 대비되는 용어다. 이 용어법에 대한 설명은 Jeffrey(2002)에서 찾을 수 있다.

정도(higher order credence)를 이용해서 BC를 확장한 것이다. 우선 이 둘을 하나씩 살펴보자.

BC는 경험 이후 E를 확신하게 되었을 때, 어떤 명제 X의 사후 확률은 E를 조건으로 하는 X의 사전 조건부 확률과 같아야 한다고 말한다. 즉, BC는 $P_{t+1}(X)=P_t(X|E)$이어야 한다고 말한다. (여기서 P_t는 사전 확률 함수이며, P_{t+1}는 사후 확률 함수이다.) 마찬가지로, 경험 이후 $\sim E$를 확신하게 되었을 때, 명제 X의 사후 확률은 $\sim E$를 조건으로 하는 X의 사전 조건부 확률과 같아야 한다. 즉 BC는 $P_{t+1}(X)=P_t(X|\sim E)$라고 말한다. 그럼 경험 이후, E의 믿음의 정도가 (1이 아닌 값으로) 어느 정도 수정된 경우, X의 사후 확률은 얼마가 되어야 하는가? 즉, $P_{t+1}(X)$의 값은 얼마가 되어야 하는가? 이 질문에 대해서, 리차드 제프리(Richard Jeffrey, 1983)는 $P_t(X|E)$와 $P_t(X|\sim E)$의 평균을 $P_{t+1}(X)$의 값으로 간주하자고 제안한다. 더욱 정확히는, 해당 두 조건부 확률의 *가중 평균*을 X의 사후 믿음의 정도로 간주하자는 것이다. 여기서 각 조건부 확률에 덧붙여지는 가중치는 E의 새로운 믿음의 정도가 결정한다. 즉 $P_t(X|E)$의 가중치는 $P_{t+1}(E)$이고, $P_t(X|\sim E)$의 가중치는 $P_{t+1}(\sim E)$가 된다. 그럼 이 제안은 다음과 같이 형식화될 수 있다.

> **제프리 조건화(Jeffrey Conditionalization, JC).** 어떤 행위자가 시점 $t+1$의 경험에 의해서 E라는 명제에 대한 믿음의 정도를 $P_t(E)(=a)$에서 $P_{t+1}(E)(=b)$로 수정했을 때, 모든 명제 X에 대해서 다음이 성립한다:
> $P_{t+1}(X)=P_{t+1}(E)P_t(X|E)+P_{t+1}(\sim E)P_t(X|\sim E)$. (단, $0<a<1$ 일 때.)[8]

제프리는 1965년에 위 조건화를 제안했으며, 그 후 많은 철학자

---

8) 이런 JC의 형식화는 한 가지 숨은 전제를 포함하고 있다. 그것은 경험 이후에 E라는 명제의 믿음의 정도만 직접적으로 수정되었다는 것이다. 나의 아래 논의는 이것을 암묵적으로 가정한 채 진행된다.

들이 다양한 방식으로 이 조건화를 정당화하거나 비판했다.[9)]

이제 BC의 적용 범위를 확장하는 두 번째 방법을 생각해보자. 문제는 경험에 의해서 직접적으로 수정된 E의 믿음의 정도가 1이 아니라는 것이었다. 그럼, 만약 비록 E에 대한 믿음의 정도가 1이 되는 것이 아니지만 그것과 관련된 다른 믿음의 정도가 경험 이후에 1이 된다고 주장할 수 있다면, 우린 성공적으로 BC의 적용 범위를 확장할 수 있을 것이다. 어떤 명제에 대한 확률이 그럴 수 있는가? 유력한 후보 중에 하나는 믿음의 정도에 대한 명제다. 가령, 시점 t+1의 경험에 의해서 E를 b(<1) 정도로 믿게 되었다고 하자. 즉, $P_{t+1}(E)=b$라고 하자. 그럼 $P_{t+1}(E)=b$에 대한 믿음의 정도는 얼마인가? 이 믿음의 정도는 고차 믿음의 정도, 즉 믿음의 정도에 대한 믿음의 정도다.[10)] 만약 행위자를 자신의 마음을 제대로 파악할 수 있을 정도로 자기 성찰적인(introspective) 사람이라고 이상화한다면, 그의 $P_{t+1}(E)=b$에 대한 믿음의 정도는 1이라고 해야 할 것 같다.

---

9) 주석 5에서 설명했듯이 제프리 조건화는 Jeffrey(1983)에서 처음으로 제시된다. BC와 마찬가지로 이에 대한 대표적인 정당화로는 더치북 논증이 있다. 이 논증은 Armendt(1980); Skyrms(1987)에서 살펴볼 수 있다. 또한, 주석 4에서 언급한 BC에 대한 다양한 정당화는 JC에 대한 정당화로 간주될 수도 있다.

10) 고차 믿음의 정도가 철학적으로, 혹은 수학적으로 적법한지에 대해선 다양한 논의가 있었다. 주관적 확률주의의 창시자라고 할 수 있는 프랭크 램지(Frank Ramsey), 브루노 드 피네티(Bruno de Finetii), 레오나르드 세비지(Leonard Savage) 등은 고차 믿음의 정도에 대해서 침묵하거나, 의심스럽게 생각했다고 한다. 가령, 드 피네티는 우리 믿음에 대한 진술은 단지 의견의 표현일 뿐 믿음의 대상일 수 없다고 말하며, 세비지는 고차 확률은 거짓말쟁이의 역설과 같은 이론적 난점을 야기할 것이라고 우려한다. 그러나 스킴즈는 이런 우려가 잘못되었으며, 고차 믿음의 정도는 수학적으로, 그리고 철학적으로도 적법하다고 논평한다. Skyrms(1980b, 1984)를 보라. 또한, 고차 확률에 대한 수학적 연구를 위해서는 Gaifmann(1988)을 참조하라. 스킴즈의 논평을 따라, 나는 아래 논의에서 고차 확률의 적법성에 대해서 고려하지는 않을 것이다.

다른 말로, 시점 t+1에 경험에 의해서 E를 b 정도 믿게 되었을 때, $P_{t+1}(P_{t+1}(E)=b)=1$이라는 것이다. 이 경우, 비록 행위자의 E에 대한 믿음의 정도는 1이 아니지만, 자신이 E를 b 정도 믿고 있다는 것에 대한 믿음의 정도는 1이 된다. 그럼 경험 이후 확신하게 된 명제가 있으니, 우리는 BC를 이용해 다른 명제의 믿음의 정도를 수정할 수 있다. 브라이언 스킴즈(Brian Skyrms)는 이렇게 해서 수정된 BC를 이차 조건화라고 부른다.[11] 이렇게 불리는 이유는 $P_{t+1}(P_{t+1}(E)=b)$와 같은 2차 믿음의 정도를 이용하기 때문이다. 다음은 이차 조건화를 좀 더 명시적으로 형식화한 것이다.

> **이차 조건화(Second Order Conditionalization, SC)**. 어떤 행위자가 시점 t+1의 경험에 의해서 E라는 명제에 대한 믿음의 정도를 $P_t(E)(=a)$에서 $P_{t+1}(E)(=b)$로 수정했을 때, 모든 명제 X에 대해서 다음이 성립한다:
> $P_{t+1}(X)=P_t(X|P_{t+1}(E)=b)$. (단, $0<P_t(P_{t+1}(E)=b)<1$일 때.)[12]

그럼 위에서 제시된 BC의 적용 범위를 확장하는 두 가지 방법, JC와 SC는 어떤 관계에 있는가? 흥미롭게도, 특별한 조건 아래에서 이 둘은 같은 결과를 내 놓는다. 즉 어떤 특별한 조건이 만족된다면, JC와 SC에 의해서 수정된 사후 확률은 서로 같다는 것이 논증될 수 있다.[13]

문제의 조건 중에 하나는 바스 반 프라센(Bas van Fraassen)이 형식화하고 이름 붙인 반영 원리(Reflection Principle)다.[14]

---

11) Skyrms(1980b)를 보라.

12) 주석 6에서 설명한 것과 마찬가지로, 이런 SC의 형식화는 경험 이후에 E라는 명제의 믿음의 정도만 직접적으로 수정되었다는 것을 암묵적으로 가정하고 있다.

13) 아래의 논증은 Skyrms(1980a; 1980b), Jeffrey(2002)에 제시된 것을 재구성한 것이다.

14) Van Fraassen(1984; 1995)를 참조하라. 이 반영 원리도 다양한 방식으로

반영 원리(Reflection Principle). 0보다 크거나 같은 모든 실수 k에 대해서, P_{t+k}가 임의의 명제 X에 믿음의 정도 c를 할당한다면, $P_t(X|P_{t+k}(X)=c)=c$. (단, $P_t(P_{t+k}(X)=c)>0$.)

위 원리에서 P_t는 t 시점의 확률 함수이며, P_{t+k}는 $t+k$ 시점의 확률 함수다. k가 0일 때, 반영 원리는 다음과 같이 변형된다: P_t가 임의의 명제 X에 믿음의 정도 c를 할당한다면, $P_t(X|P_t(X)=c)=c$. 이것은 X에 대한 현재 믿음의 정도가 c라는 조건아래에서 X에 대한 믿음의 정도는 c가 되어야 한다고 말하고 있다. 이는 매우 직관적이다. 한편, 만약 k가 0보다 크다면, P_{t+k}는 t 시점 이후 어떤 미래의 확률함수라고 할 수 있다. 거칠해 말해, 이 경우 반영 원리는 우리 현재 믿음의 정도는 우리 미래 믿음의 정도를 반영하고 있어야 한다는 것이다. 반 프라센을 비롯해서 많은 철학자들은 이 반영 원리를 따르는 것이 합리적 믿음의 정도를 위한 필요조건이라고 한다. 이런 주장을 뒷받침하는 직관은 현재의 나보다 미래의 내가 더 많은 것을 경험할 것이고, 따라서 미래의 내가 현재의 나보다 인식적으로 더 우수하다는 것이다. 그리고 이런 이유에서 인식적으로 우수한 미래의 믿음을 현재의 믿음에 반영해야 한다는 것이다.

그럼, 이제 JC와 SC 사이의 관계를 살펴보자. 어떤 행위자가 $t+1$ 시점에 무언가를 경험한 이후 E에 대한 믿음의 정도를 $P_t(E)(=a)$에서 $P_{t+1}(E)(=b)$로 수정했다고 하자. 수식을 단순하게 만들기 위해서, H가 $P_{t+1}(E)=b$를 가리킨다고 하자. 그리고 해당 행위자가 t 시점에 반영 원리를 만족하다고 가정하자. 또한, 그는 SC를 이용해서 다른 명제에 대한 믿음의 정도를 수정한다고 하자. 그럼 SC에 의해서 임의의 명제 X에 대해 다음이 성립한다.

정당화될 수 있다. 대표적인 것이 더치북을 이용하는 것이다. Van Fraassen(1984, pp. 240-241)을 보라.

(1) $P_{t+1}(X)=P_t(X|H)$.

한편 t 시점에 반영 원리를 만족한다는 가정과 확률 계산에 의해서 다음이 성립한다.

(2) $P_t(E|H)=b$; $P_t(\sim E|H)=1-b$.

또한 확률 계산 규칙에 의해서 다음이 성립한다.

(3) $P_t(X|H)=P_t(E|H)P_t(X|E\&H)+P_t(\sim E|H)P_t(X|\sim E\&H)$.

그럼 (2)와 (3)으로부터 다음이 도출된다.

$P_t(X|H)=bP_t(X|E\&H)+(1-b)P_t(X|\sim E\&H)$.

여기에 추가로 다음을 가정하자.

(*) $P_t(X|E\&H)=P_t(X|E)$와 $P_t(X|\sim E\&H)=P_t(X|\sim E)$.

이 가정은 t 시점에, X와 H가 E (그리고 $\sim E$)에 조건부로 독립적이라는 것이다. 거칠게 말해, (*)는 E (혹은 $\sim E$)가 참이라는 것이 주어졌을 때, 나의 X에 대한 믿음의 정도는 E에 대한 나의 믿음의 정도가 b라는 것, 즉 H에 대한 나의 믿음의 정도와 상관이 없다는 것이다. 그럼 위 식은 다음과 같이 변형된다.

$P_t(X|H)=bP_t(X|E)+(1-b)P_t(X|\sim E)$.

그런데, $P_{t+1}(E)=b$이기 때문에, 위 식은 다음과 같이 변형된다.

(4) $P_t(X|H)=P_{t+1}(E)P_t(X|E)+P_{t+1}(\sim E)P_t(X|\sim E)$.

(1)에 따르면, 식 (4)의 좌변은 $P_{t+1}(X)$와 같다. 그러므로 다음이 도출된다.

$$P_{t+1}(X)=P_{t+1}(E)P_t(X|E)+P_{t+1}(\sim E)P_t(X|\sim E).$$

이 마지막 등식은 바로 JC가 요구하는 것이다. 따라서 우린 다음과 같이 결론 내릴 수 있다: 어떤 행위자가 t+1 시점에 E에 대한 그의 믿음의 정도를 b로 바꾸는 경험을 하고, t시점에 반영원리와 (*)를 만족시키는 모든 명제 X에 대해서, SC를 이용해서 갱신된 X의 사후 확률은 JC를 이용해서 갱신된 X의 사후 확률과 같다.

그럼, 이 논증은 건전한가? 다른 말로, 이 논증에서 사용된 두 가지 가정, 즉 반영 원리와 (*)은 받아들일만한 것인가? 앞에서 설명한 대로 반영 원리는 나름의 직관적인 근거를 가지고 있다. 나는 이 논문에서 많은 철학자들이 인정하듯이 반영 원리가 합리성의 필요조건이라는 것을 문제 삼지 않을 것이다. 하지만 (*)는 어떠한가? 이 논증을 제시한 스킴즈 역시 (*)는 반영 원리보다 그럴싸하지 않다고 말한다. 그는 특히 X가 믿음의 정도에 대한 명제인 경우에는 (*)가 성립하지 않을 수도 있다는 것을 인정한다. 간단한 예를 들어보자. (*)의 X가 H라고 하자. 즉 X가 E에 대한 믿음의 정도에 대한 명제, 즉 $P_{t+1}(E)=b$라고 하자. 그럼 (*)의 첫 번째 등식의 좌변은 확률 규칙에 의해서 1이 된다. 한편, 우변이 1이라는 것은 보장되지 않는다. 이 간단한 예에서 알 수 있듯이, (*)가 일반적으로 성립한다고 말할 수 없다. 이런 이유에서 스킴즈는 (*)가 항상 성립한다고 말하지 않고, 대부분의 경우에 성립한다고 온건하게 주장한다. 결국, 스킴즈의 입장에서 우리가 결론내릴 수 있는 것은 JC를 이용한 믿음 수정과 SC를 이용한 믿음 수정이 대부분

일치한다는 것뿐이다.[15)]

그럼 (*)가 성립하지 않는 경우는 어떻게 되는가? (*)가 성립하지 않으면, JC를 이용해서 갱신된 믿음의 정도와 SC를 이용해서 갱신된 믿음의 정도가 서로 다를 것이다. 이 경우, 우린 무엇을 선택해야 하는가? 아쉽게도 스킴즈나 제프리 모두 (*)가 성립하지 않는 경우를 그리 중요하게 다루지 않는 것 같다. 특히 스킴즈의 논의는 그저 대부분의 경우 (*)가 성립한다는 것으로 마무리된다. 하지만 이런 마무리는 성급하다. 왜냐하면, (*)가 성립하지 않는 몇몇 경우들은 JC와 SC를 중요한 차이를 드러내기 때문이다.

앞에서 간단히 언급했듯이, 스킴즈는 (*)의 X가 믿음의 정도에 대한 명제인 경우 (*)가 성립하지 않을 수 있다는 것을 인정했다. (*)가 성립하지 않으면, JC에 의해서 갱신된 X의 사후 확률은 SC에 의해서 갱신된 X의 사후 확률과 다를 수밖에 없다. 우린 이런 X에 대한 믿음의 정도, 즉 믿음의 정도에 대한 믿음의 정도를 고차 믿음의 정도라고 불렀다. 따라서, 고차 믿음의 정도를 SC와 JC에 의해서 각각 갱신하는 경우, 각각의 결과는 다를 수 있다. 다음 두 절에서 나는 고차 믿음의 정도를 이용해 그 차이가 좀더 흥미롭게 드러나는 두 가지 경우를 제시할 것이다.

3. 반영 원리와 고차 믿음의 정도의 갱신

우선 반영 원리와 JC, SC 사이의 관계를 살펴보자. 앞에서 설명했듯이, 반영 원리는 많은 철학자들이 합리성의 필요조건으로 받아들이는 원리 중에 하나다. 어떤 행위자가 t 시점에 어떤 명제 E를 a 정도로 믿고 있다고 가정하자. 즉, $P_t(E)=a$라고 하자. 그리고 그는 반영 원리를 만족한다고 가정하자. 특히, 그는 t 시점과 t+1 시

15) (*)에 대한 스킴즈의 논평은 Skyrms(1980b, p. 125)에서 확인할 수 있다.

점 사이에 무엇을 경험하게 될 지 생각하고 있다. 그는 앞으로 겪을 가능성이 있는 다양한 경험들 중에서 어떤 특정한 것을 경험한다면, t+1 시점에 그가 가지게 될 E의 믿음의 정도는 $b(\neq a)$가 될 것이라고 예상하고 있다. 그럼 반영 원리에 의해서, $P_t(E|P_{t+1}(E)=b)=b$가 성립한다. 즉 행위자는 E에 대해서 반영 원리를 만족한다. 한편, t+1 시점에 실제로 행위자가 E에 대한 믿음의 정도를 b로 바꾸게 만드는 경험을 했다고 하자. 그래서 그는 t+1 시점에 E를 실제로 b 정도로 믿게 되었다고 하자. 즉, $P_{t+1}(E)=b$라고 가정하자. 이 경우, $P_{t+1}(E|P_{t+1}(E)=b)$의 값은 얼마가 되어야 하는가? 반영 원리는 이 값이 b이어야 한다고 말한다. 요약하자면, 그가 다음과 같은 믿음의 정도를 가지는 것은 이상할 것이 없어 보인다. (수식을 단순하게 하기 위해서, H가 $P_{t+1}(E)=b$를 가리킨다고 하자.)

(i) $P_t(E)=a\neq b=P_{t+1}(E)$;
(ii) $P_t(E|H)=P_{t+1}(E|H)=b$.

그럼 JC 혹은 SC가 적합한 믿음 갱신 규칙이라면, 이런 믿음의 정도를 가지는 것과 충돌을 일으켜서는 안 될 것이다. t 시점과 t+1 시점 사이에 E의 믿음의 정도는 경험에 의해서 a에서 b로 직접적으로 수정되었다. 한편 (ii)이 보여주듯이, 이런 믿음 수정에도 불구하고 E를 조건으로 하는 H의 조건부 믿음의 정도는 t와 t+1 사이에서 변하지 않는다. 이 조건부 믿음의 정도는 고차 믿음의 정도라는 것에 주목하라. 즉 H는 해당 행위자의 믿음의 정도에 대한 명제다. 그럼 이 고차 조건부 믿음의 정도를 JC 혹은 SC를 이용해서 갱신했을 때, (ii)이 성립하는가?

먼저, JC를 이용해서 H를 조건으로 하는 E의 조건부 믿음의 정도를 수정해보자. 먼저 확률 계산 규칙에 의해서, 다음이 성립한

다.[16]

$$(5)\quad P_{t+1}(E|H) = \frac{P_{t+1}(E\&H)}{P_{t+1}(H)}.$$

위에서 t+1 시점에 E에 대한 믿음의 정도는 경험의 영향에 의해서 $P_t(E)(=a)$에서 $P_{t+1}(E)(=b)$로 직접적으로 수정되었다. 따라서 JC에 의해서 다음이 성립한다.

$$P_{t+1}(E\&H)=P_{t+1}(E)P_t(E\&H|E)+P_{t+1}(\sim E)P_t(E\&H|\sim E)=P_{t+1}(E)P_t(H|E).$$
$$P_{t+1}(H)=P_{t+1}(E)P_t(H|\sim E)+P_{t+1}(\sim E)P_t(H|\sim E).$$

그리고 베이즈 정리에 의해서 위 두 등식은 다음과 같이 변형된다.

$$(6)\quad P_{t+1}(E\&H) = P_{t+1}(E)\frac{P_t(H)P_t(E|H)}{P_t(E)}.$$
$$P_{t+1}(H) = P_{t+1}(E)\frac{P_t(H)P_t(E|H)}{P_t(E)} + P_{t+1}(\sim E)\frac{P_t(H)P_t(\sim E|H)}{P_t(\sim E)}.$$

그럼, (5)와 (6)에 의해서 다음이 도출된다.

$$(7)\quad P_{t+1}(E|H) = \frac{[P_{t+1}(E)/P_t(E)]P_t(E|H)}{[P_{t+1}(E)/P_t(E)]P_t(E|H)+[P_{t+1}(\sim E)/P_t(\sim E)]P_t(\sim E|H)}.$$

이제 이 결과가 위의 (ii)가 성립하는지 확인해보자. 귀류 가정으로 (ii)가 성립한다고 가정하자. 그럼 (7)로부터 다음이 도출된다.[17]

---

16) 아래 논의에서 $P_{t+1}(H)>0$이라는 것이 가정된다.

17) (ii)가 성립한다는 귀류 가정에 의해서, (7)은 다음과 같이 변형된다.

$$(8)\quad \frac{P_{t+1}(E)}{P_t(E)} = \frac{P_{t+1}(\sim E)}{P_t(\sim E)}.$$

그러나 (8)은 $P_t(E)=P_{t+1}(E)$를 함축하고, 따라서 $P_t(E)=a \neq b=P_{t+1}(E)$라는 (i)과 모순을 일으킨다. 그러므로, (ii)은 성립할 수 없다. 즉, 문제의 행위자는 t 시점과 t+1 시점 모두에서 반영 원리를 만족시킬 수 없다. 이런 결과를 다른 말로 표현하자면, 어떤 주어진 시점에 반영 원리를 만족시켰던 행위자가 JC를 이용해 자신의 믿음의 정도를 갱신한다면 갱신 이후에 더 이상 반영 원리를 만족시킬 수 없다는 것이다. 이런 의미에서 JC와 반영 원리는 서로 충돌한다고 할 수 있다.[18] [19]

$$b = \frac{[P_{t+1}(E)/P_t(E)]b}{[P_{t+1}(E)/P_t(E)]b+[P_{t+1}(\sim E)/P_t(\sim E)](1-b)}.$$

그리고 이 등식은 다음과 같이 정리될 수 있다. $[P_{t+1}(E)/P_t(E)](1\text{-}b)=[P_{t+1}(\sim E)/P_t(\sim E)](1\text{-}b)$. 그럼 이제 우린 (8)을 도출할 수 있다.

18) 이런 결과는 반영 원리와 BC사이의 관계를 생각할 때 무척 흥미롭게 보일 수 있다. 반 프라센(van Fraassen (1995; 1999))은 반영 원리가 BC를 함축하고, BC 역시 반영 원리를 함축한다고 주장한다. 한편, JC는 BC의 적용 범위를 확장한 것이다. 그런데 위에서 나는 JC와 반영 원리가 충돌한다고 주장했다. 이런 나의 주장은 반 프라센의 주장과 충돌하는 듯이 보인다. 그리고 이런 점에서, 반 프라센의 주장이 옳다면 나의 주장에는 무언가 잘못된 것이 있을 것이라는 의심은 자연스럽다. 하지만, 다행이도 몇몇 철학자들은 반 프라센의 주장에 문제가 있음을 밝혀내었다. 즉 반 프라센의 생각과 달리, 반영 원리는 BC를 함축하지 않고, BC 역시 반영 원리를 함축하지도 않는다는 것이다. 반 프라센의 주장이 틀렸다는 논증은 Weisberg(2007)와 Briggs(2009)에서 확인할 수 있다.

19) 이 논증에 대해서 신랄한 평가를 해 준 익명의 심사위원들에게 감사한다. 그 중 한 심사위원은 반영원리와 제프리 조건화가 충돌한다는 나의 논증이 제프리 조건화에 대한 오해에서 비롯되었다고 논평하였다. 그가 지적한 오해는 t+1 시점의 경험에 의해서 E에 대한 믿음의 정도가 b가 되었을 때, 조건화를 적용하기 위해서 이용될 분할 집합(partition)은 $\{E,\sim E\}$가 아니라 $\{E\&H,\ \sim E\&H\}$라는 것이다. 그리고 그는 이 분할 집합에 제프리 조건화를 적용해 믿음을 수정한다면 위에서 언급되었던 제프리 조건화와 반영원리

그럼 SC는 어떠한가? SC 역시 JC와 비슷한 방식으로 반영 원리와 충돌하는가? 흥미롭게도 SC는 그렇지 않다. 위 논증과 비슷하게, 먼저 SC를 이용해서 H를 조건으로 하는 E의 조건부 믿음의 정도를 수정해보자. 위 논증에서 E의 믿음의 정도는 t+1 시점에 $P_t(E)(=a)$에서 $P_{t+1}(E)(=b)$로 변경되었다. 그럼 SC에 의해서 다음이 성립한다.

(9) 모든 X, Y에 대해서, $P_{t+1}(X|Y)=P_t(X|Y\&H)$.

그럼 X에 E를, Y에 H를 대입하자. 그럼 다음 식이 성립한다.

(10) $P_{t+1}(E|H)=P_t(E|H\&H)=P_t(E|H)$.

만약 문제의 행위자가 t 시점에 반영 원리를 만족한다면 $P_t(E|H)=b$이고, 따라서 (ii)가 성립한다. 즉, 어떤 주어진 시점에 반영 원리를 만족시켰던 행위자가 SC를 이용해 자신의 믿음의 정도를 갱신한다면 갱신 이후에 더 이상 반영 원리를 만족시킬 수 없다는 것이다. 그러므로 우린 SC와 같은 방식으로 고차 믿음의 정도를 갱신하는 경우에 위에서 논증된JC와 반영 원리 사이의 충돌은 발생하지 않는다고 결론 내릴 수 있다.

요약하자면, 어떤 명제 E에 대해서 반영 원리를 만족시키는 행위자가 E에 대한 믿음의 정도를 어떤 경험에 의해서 직접적으로 수정했을 때, JC를 이용해 갱신된 새로운 믿음의 정도들은 E에 대한 반영 원리를 만족시킬 수 없지만 SC를 이용해 갱신된 믿음의

사이의 충돌은 발생하지 않는다고 말한다. 내가 제시한 사례에서 왜 행위자가 $\{E,\sim E\}$가 아니라 $\{E\&H, \sim E\&H\}$를 이용해서 믿음을 갱신해야 하는지는 의아하지만, 만약 그렇다고 하더라도 나의 논증이 잘못이라고 말할 수 없다. 자세한 나의 답변은 <부록>으로 추가하였다.

정도들은 E에 대한 반영 원리를 만족시킬 수 있다. 이런 결과는 분명히 JC의 단점으로 보인다. 위 논증에서 JC나 SC를 통해서 갱신된 믿음의 정도는 $P_{t+1}(E)=b$를 조건으로 하는 E의 믿음의 정도, 즉 고차 믿음의 정도라는 것을 기억하라. 이제 우린 위 논증은 고차 믿음의 정도의 갱신과 관련해서, JC보다 SC를 선호해야 하는 한 가지 이유를 제시한다고 말할 수 있다. 그러나 이것이 전부가 아니다. 난 다음 절에서 고차 믿음 갱신과 관련된 JC의 또 다른 문제점을 지적할 것이다. 그 문제는 무어의 부조리(Moorean absurdity)와 관련 있다.

4. 무어의 부조리와 고차 믿음의 갱신

다음 사례를 생각해보자. 철수는 별의 폭발에 대해서 연구하고 있는 천문학과 대학원생이다. 어떤 특정한 별이 폭발했다는 명제를 E라고 하자. t+1시점에 철수는 자신의 연구실에 큰 천체 망원경이 설치되고 있는 것을 발견한다. 그 순간, 철수는 언젠가 그 망원경으로 천체를 관측한다면, E에 대한 자신의 믿음의 정도가 b로 변할 것이라고 확신하게 되었다. 그리고 그는 t+2 시점에 실제로 그 망원경을 이용할 기회를 얻었다. 그러나 천체를 관찰한 이후, E에 대한 철수의 믿음의 정도는 b가 아닌 c가 되었다. 이 사례는 자신의 미래 믿음에 대한 예측이 실제와 같지 않은 경우다. 철수는 특별한 관찰 이후에 자신의 E에 대한 자신의 미래 믿음이 b가 될 것이라고 확신했지만, 실제 미래의 믿음은 c가 된 경우이다. 이런 믿음의 변화는 흔한 일은 아니지만, 그리 이상한 일은 아니라고 생각한다.

그럼 이 사례에서 철수는 자신의 다른 믿음을 어떻게 갱신해야 하는가? 물론 두 가지 방법이 있다. 첫 번째 방법은 JC를 이용하

는 것이고, 두 번째 방법은 SC를 이용하는 것이다. 먼저 JC를 이용해서 철수의 믿음의 정도를 갱신해보자.

위 사례는 연속적 믿음 갱신이라는 것을 주목하라. 즉, 위 사례에서 철수는 두 가지 종류의 경험을 한다. 첫 번째 경험에 의해서, t+1 시점에 그는 t+2 시점의 E에 대한 믿음의 정도가 b가 될 것이라고 확신하게 된다. 즉 $P_{t+1}(P_{t+2}(E)=b)=1$가 성립한다. 그리고 두 번째 경험에 의해서, 그는 t+2시점에 E에 대한 믿음의 정도를 $c(\neq b)$로 수정한다. 즉, $P_{t+2}(E)=c$가 성립한다. 이 연속적인 경험 이후, 철수가 JC를 이용해 임의의 명제 X에 대한 믿음의 정도를 갱신한다면 다음 식을 도출할 수 있다.[20]

모든 X에 대해서,

$$P_{t+2}(X)=P_{t+2}(E)P_t(X|P_{t+2}(E)=b\&E)+P_{t+2}(\sim E)P_t(X|P_{t+2}(E)=b\&\sim E).$$

이제, (11)를 이용해서 $P_{t+2}(E)=c$에 대한 철수의 믿음의 정도는 어떻게 갱신되는지 살펴보자. 즉 $P_{t+2}(P_{t+2}(E)=c)$의 값이 얼마인지 계산해보자. 그럼 우린 그 값이 0이라는 것을 확인할 수 있다.[21]

---

20) 첫 번째 경험 이후, 철수의 믿음의 정도를 JC를 이용해서 수정하면 다음이 성립한다.
모든 X에 대해서, $P_{t+1}(X)=P_t(X|P_{t+2}(E)=b)$.
그리고 이것으로부터 다음을 도출할 수 있다.
(†) 모든 X에 대해서, $P_{t+1}(X|E)=P_t(X|P_{t+2}(E)=b\&E)$.
한편, 두 번째 경험 이후, 철수의 믿음의 정도를 JC를 이용해서 갱신한다면 다음이 성립한다.
(‡) 모든 X에 대해서, $P_{t+2}(X)=P_{t+2}(E)P_{t+1}(X|E)+P_{t+2}(\sim E)P_{t+1}(X|\sim E)$.
그럼 (†)과 (‡)로부터 다음을 도출할 수 있다.
모든 X에 대해서,
$P_{t+2}(X)=P_{t+2}(E)P_t(X|P_{t+2}(E)=b\&E)+P_{t+2}(\sim E)P_t(X|P_{t+2}(E)=b\&\sim E)$.

21) (11)에 의해서 다음이 성립한다.
$P_{t+2}(P_{t+2}(E)=c)=P_{t+2}(E)P_t(P_{t+2}(E)=c|P_{t+2}(E)=b\&E)+P_{t+2}(\sim E)P_t(P_{t+2}(E)=c|P_{t+2}(E)=b\&\sim E)$ b와 c가 다르기 때문에, $P_{t+2}(E)=b$와 $P_{t+2}(E)=c$는 양립불가능하다. 따라

즉 $P_{t+2}(P_{t+2}(E)=c)=0$이다. 위 사례에서 철수는 t+2시점에 E를 c 정도로 믿고 있다는 것을 기억하라. 즉 $P_{t+2}(E)=c$이다. 따라서, 다음이 성립한다.

(12) $P_{t+2}(E)=c$이고 $P_{t+2}(P_{t+2}(E)=c)=0$이다.

여기서 (12)은 조금 이상하다. 이것이 이상한 이유는 무어의 부조리(Moorean absurdity)와 관련되어 있다.

무어의 부조리는 무어 문장이라고 불리는 다음과 같은 형식의 문장에 대한 것이다:

(M) "X, 그리고 S는 t 시점에 X를 믿지 않는다."[22]

여기서 S는 어떤 행위자를 나타낸다. 이 문장은 논리적 모순문이 아니다. 흥미로운 것은 이 문장이 논리적 모순이 아님에도 불구하고, S가 t 시점에 이 문장을 믿을 수 없다는 것이다. S가 t 시점에 (M)을 믿는다고 하자. 그럼 그가 t 시점에 (M)의 첫 번째 연언지, 즉 X도 믿는다고 말할 수 있다.[23] 그러나 이는 (M)의 두 번째 연언지와 모순이다. 따라서 X는 t 시점에 (M)을 믿을 수 없다. 이렇

서 다음이 성립한다.

$P_t(P_{t+2}(E)=c|P_{t+2}(E)=b\&E)=0$; $P_t(P_{t+2}(E)=c|P_{t+2}(E)=b\&\sim E)=0$.

그러므로, $P_{t+2}(P_{t+2}(E)=c)=0$이다.

22) 나는 이 논문에서 문장을 큰 따옴표(" ")로 묶어 나타낼 것이다. 하지만 이 논문에서 문장과 명제 사이의 구분은 그리 중요하지 않다.

23) 여기에 믿음에 대한 연역적 폐쇄 원리(Principle of Deductive Closure)가 가정되었다. 즉 어떤 행위자가 A를 믿고, A가 B를 논리적으로 함축한다면, 그 행위자는 B도 믿는다는 것이 가정되었다. S가 (M)을 믿는다고 가정하자. 그리고 (M)은 자신의 첫 번째 연언지, 즉 X를 논리적으로 함축한다. 따라서 연역적 폐쇄 원리에 의해서 S는 X를 믿는다.

게 (M)을 S가 t 시점에 믿을 수 없다는 것을 철학자들은 무어의 부조리라고 부른다.[24)]

이 무어 문장은 다양한 형태를 가지고 있다. 물론 확률적으로 표현할 수도 있다. (M)을 확률을 이용해서 표현하기 위해서는 '믿는다'는 것과 '어느 정도로 믿는다' 사이의 관계를 규명할 필요가 있다. 즉 다음 질문에 답할 수 있어야 한다: S가 어떤 문장 "X"를 믿기 위해선 X에 대한 그의 믿음의 정도는 얼마가 되어야 하는가? 이 질문에 답할 수 있어야만 우리는 (M)의 두 번째 연언지를 확률적으로 표현할 수 있다. 그러나 그 값이 정확히 얼마인지 결정 내리기는 쉽지 않다. 하지만 우리는 이 값이 최소한 0보다 커야 된다는 것은 받아들일 수 있다. 그리고, 문장 "X"에 대한 믿음의 정도가 0이라면 X를 믿지 않는다는 것도 자연스럽다. 그럼 우린 확률을 이용해 무어 문장을 다음과 같은 형식으로 나타낼 수 있다.

(M*) "X 그리고 $P_t(X)=0$."

여기서 P_t는 어떤 행위자가 t 시점에 가지고 있는 확률 함수다.

이제, (12)의 이상한 점을 분명히 드러낼 수 있다. 즉 "(12)"는 무어 문장이다. "$P_{t+2}(E)=c$"는 (M*)의 "X"에 대응한다. 이제 우린 위의 천문학과 대학원생의 사례에서 철수가 자신의 믿음의 정도를 JC를 이용해서 갱신한다는 전제는 무어 문장을 도출한다고 결론 내릴 수 있다. 물론, 위 사례에서 JC를 이용한 믿음 갱신이 무어

24) 물론, S가 아닌 다른 사람이 이 문장을 믿는 것은 아무 문제로 일으키지 않는다. 가령, 영희는 "해가 동쪽에서 뜨지만, 철수는 지금 해가 동쪽에서 뜬다는 것을 믿지 않는다."는 것을 믿을 수 있다. 마찬가지로, S가 t 시점 이전이나 이후에 이 문장을 믿는 것도 아무 문제를 일으키지 않는다. 예를 들어, 철수는 "해가 동쪽에서 뜨지만, 철수는 과거에 해가 동쪽에서 뜬다는 것을 믿지 않았다."는 것을 믿을 수 있다. 무어의 부조리에 대한 논의는 이병덕(2000)를 참조하라.

문장을 도출한다는 것 자체는 어떤 논리적인 문제점을 야기하지 않는다. 왜냐하면 도출된 무어 문장 자체는 논리적 모순문이 아니기 때문이다. 그러나 해당 천문학과 대학생이 t+2 시점에 관련된 전제들을 믿는다고 가정한다면 문제가 발생한다. 이 점을 좀더 분명히 설명해보자. 철수는 t+2시점에 다음 문장들을 믿고 있다고 가정하자. (아래 믿음의 정도 함수들은 모두 철수의 믿음의 정도 함수들이다.)

(a) "$P_{t+1}(P_{t+2}(X)=b)=1$."

(b) "$P_{t+2}(X)=c$."

이 두 명제의 연언은 논리적 모순이 아니며, t+2 시점에 철수가 이 두 명제의 연언을 믿는 것은 무어의 부조리를 야기하지도 않는다. 여기에 철수가 t+2 시점에 다음 문장을 믿고 있다는 가정을 추가하자.

(c) "JC는 합리적 믿음 갱신 규칙이다."

그럼 앞에서 보여준 것처럼, 이 가정과 더불어 (a)와 (b)는 무어 문장 "(12)"를 도출한다. 여기에 믿음에 대한 연역적 폐쇄 원리(Principle of Deductive Closure)을 가정하자. 이 원칙은 어떤 행위자가 A를 믿고 A가 B를 논리적으로 함축한다면, 그는 B도 믿어야 한다는 것이다. 이 원칙에 따르면, t+2시점에 (a), (b), (c)을 믿는 철수는 같은 시점에 무어 문장 "(12)" 역시 믿어야 한다. 하지만, 철수는 "(12)"를 믿을 수 없다. 왜냐면 그런 믿음은 무어의 부조리를 야기하기 때문이다. 요약하자면, (a), (b), (c)에 대한 믿음과 연역적 폐쇄 원리는 무어의 부조리를 야기한다.[25)]

그럼 SC도 비슷한 문제점을 가지고 있는가? 흥미롭게도 그렇지 않다. SC를 철수의 믿음 갱신에 적용하면 다음을 도출할 수 있다.[26)]

(13) 모든 X에 대해서,
$P_{t+2}(X)=P_t(X|P_{t+1}(P_{t+2}(E)=b)=1\&P_{t+2}(E)=c)$

이제, (13)을 이용해서 $P_{t+2}(P_{t+2}(E)=c)$의 값을 계산해보라. 그럼 우린 그 값이 1이라는 것을 쉽게 도출할 수 있다. 위 사례에서 $P_{t+2}(E)=c$라는 것을 다시 한 번 기억하라. 그럼 다음이 성립한다는 것을 알 수 있다.

---

25) 이런 논증은 일종의 귀류법이라고 할 수 있다. 하지만 이 귀류법은 일반적인 귀류법과는 달리 논리적으로 비일관적인 결론을 이용하지 않는다. 단지, 주어진 전제들로부터 무어 문장을 도출하고, 따라서 전제 중에 일부가 거짓이라고 결론 내릴 뿐이다. 앞에서 언급했듯이 무어 문장 그 자체는 논리적으로 비일관적이지 않다. 그저 무어 문장을 믿는다는 것이 어떤 부조리함을 야기할 뿐이다. 이런 이유에서 앨런 하이예크(Alan Hájek(2008))는 무어 문장에 의해서 야기되는 부조리를 논리적 비일관성이 아니라 실용적 비일관성(pragmatic inconsistency)이라고 부르며, 이 실용적 비일관성을 이용한 논증을 새로운 형태의 귀류법이라고 말한다. 이러한 종류의 귀류법을 사용하는 대표적인 논증의 예는 Chalmers and Hájek(2007)에서 발견할 수 있다.

26) 첫 번째 경험 이후 SC를 이용해 철수의 믿음의 정도를 갱신하면 다음이 성립한다.
모든 X에 대해서, $P_{t+1}(X)=P_t(X|P_{t+1}(P_{t+2}(E)=b)=1)$.
그럼, 이로부터 다음이 따라 나온다.
(†) 모든 X에 대해서, $P_{t+1}(X|P_{t+2}(E)=c)=P_t(X|P_{t+1}(P_{t+2}(E)=b)=1\&P_{t+2}(E)=c)$.
한편, 두 번째 경험 이후 철수의 믿음의 정도를 SC를 이용해 갱신하면, 다음이 성립한다.
(‡) 모든 X에 대해서, $P_{t+2}(X)=P_{t+1}(X|P_{t+2}(E)=c)$.
그럼 (†)과 (‡)로부터 다음을 도출할 수 있다.
모든 X에 대해서, $P_{t+2}(X)=P_t(X|P_{t+1}(P_{t+2}(E)=b)=1\&P_{t+2}(E)=c)$

(14) $P_{t+2}(E)=c$이고 $P_{t+2}(P_{t+2}(E)=c)=1$.

"(14)"는 무어 문장이 아니다. 무어의 부조리와의 관계를 좀더 분명히 하기 위해서, 앞에서 언급한 (a)와 (b) 그리고 다음 (c*)를 t+2시점에 철수가 믿는다고 가정하자.

(c*) "SC는 합리적 믿음 갱신 규칙이다."

우리가 살펴본 대로 (a), (b), (c*)는 "(12)"가 아닌 "(14)"를 함축한다. 그러므로 우린 SC는 JC와 달리 무어 문장을 도출하지 않으며, 따라서 t+2시점에 철수가 (a), (b), (c*)를 믿는다고 하더라도, 그는 무어의 부조리에 빠지지 않는다.

정리해보자. 위 논증에서 우리는 $P_{t+2}(E)=c$에 대한 믿음의 정도, 즉 고차 믿음의 정도를 JC나 SC를 이용해서 갱신하였다. 그 결과 우리는 철수의 사례와 같이 미래의 믿음이 과거에 예측된 것과 다른 경우 JC는 무어의 부조리를 야기할 수 있지만, SC는 그렇지 않다는 결론을 도출하였다.[27] 이런 점에서 우리는 위 논증은 고차 믿음의 정도의 갱신과 관련해서, JC보다 SC를 선호해야 하는 또 한 가지 이유를 제시한다고 말할 수 있다. [28]

---

27) 여기서 독자는 나의 주장은 단지 JC를 이용한 믿음 갱신은 무어의 부조리를 **야기할 수 있다**는 것뿐이라는 점을 주의해야 한다. 엄격하게 말해, (내가 가정한 것과 달리) 위의 사례에서 철수가 (a), (b), (c) 중에 적어도 하나를 믿지 않는다면, 그의 믿음은 무어의 부조리에 빠지지 않는다. 이 경우, 위에서 제시된 논증은 단지 JC를 이용한 믿음 갱신은 무어 문장을 도출한 것을 보였을 뿐이다.

28) 익명의 심사 위원은 이 논증에 대해서 해당 행위자가 완벽한 내적 관찰(introspection) 능력을 가지고 있지 않다면 그가 (b)를 믿는다고 생각할 이유가 전혀 없다고 논평한다. 하지만, 위 논증에서 중요한 것은 해당 행위자의 (b)에 대한 믿음을 정당화하는 것이 아니다. 나의 논증을 위해서는 철수가 (b)를 믿는 것이 논리적으로 가능하고 불합리하지 않다는

5. 결론을 대신하여

앞에서 경험 이후에 어떤 명제에 대한 믿음의 정도가 1보다 작은 값으로 변했을 때 베이지안 조건화를 적용할 수 없다고 했다. 그리고 이런 베이지안 조건화의 한계를 극복하는 방법으로 두 가지 다른 갱신 규칙을 제안되었다. 그것은 제프리 조건화와 이차 조건화였다. 그리고 몇몇 조건이 만족된다면, 이 제프리 조건화를 통한 믿음 갱신과 이차 조건화를 통한 믿음 갱신은 서로 같은 결과를 내놓는다는 것도 보였다. 그러나 문제는 그 조건이 만족되지 않는 경우였다. 스킴즈가 추측하듯이, 문제의 조건 중에 하나는 고차 믿음의 정도를 갱신할 때 만족되지 않을 수 있다. 나는 이 논문에 그런 경우, JC보다 SC를 선호해야 하는 두 가지 이유를 제시하였다. 첫 번째는 고차 믿음의 정도의 갱신과 관련해서 JC는 반 프라센의 반영 원리와 충돌하지만, SC는 그렇지 않을 수 있다는 것이었다. 두 번째는 어떤 특별한 상황에서 고차 믿음의 정도를 갱신했을 때, JC는 무어의 부조리를 야기하지만 SC는 그렇지 않다는 것이었다. 이 두 가지 이유에서 나는 고차 믿음의 정도를 갱신할 때는 JC보다는 SC가 더 좋은 규칙이라고 생각한다.

하지만 여기에서 몇 가지 주의할 것이 있다. 나는 모든 경우에

것으로 충분하다. 나의 주장은 (a), (b)와 더불어 (c)를 믿고 있는 행위자의 믿음은 무어의 부조리를 야기하지만, (a), (b)와 더불어 (c*)를 믿고 있는 행위자의 믿음은 무어의 부조리를 야기하지 않는다는 것이다. 물론, (b)에 대한 믿음이 불합리하거나 논리적으로 불가능하다면 위 논증의 설득력은 떨어질 것이다. 하지만 (b)에 대한 믿음이 불가능하거나 불합리한 일이 아니라면, 위 논증을 이용해 JC와 SC를 비교하려는 나의 시도는 잘못되었다고 할 수 없다. 물론, 완벽한 내적 관찰 능력을 가지는 것이 불가능하거나 불합리하다고 말할 수는 없다. 합리적이기 위해서 완벽한 내적 관찰 능력을 가져야 한다는 것을 받아들일 수는 없지만, 그렇다고 완벽한 내적 관찰 능력을 가지는 것이 불합리하다고 말할 수도 없다.

JC가 SC보다 우수하다고 말하지 않는다. 스킴즈가 추측한대로, 일차 믿음의 정도를 갱신할 때는 JC와 SC가 대부분의 경우 같은 결과를 내놓을 것이다. 그러므로, 이 논문이 일차 믿음의 정도와 관련해서도 JC보다 SC를 선호해야 한다고 주장하고 있다고 생각하면 오해이다.

두 번째로, SC 역시 문제를 가지고 있다는 것이 언급되어야 한다. SC는 행위자가 자기 자신의 믿음의 정도를 확신할 수 있을 정도로 자기 성찰적(introspective)이라는 것을 함축한다. 이를 다른 말로, 행위자가 자기 자신의 마음을 선명하게(luminously) 알아야 한다는 것이다. 그리고 이런 특징은 선명성(luminosity)라고 불리고, 다음과 같이 형식화된다.

선명성(Luminosity). 모든 X에 대해서, $P(X)=r$이라면 $P(P(X)=r)=1$.

그리고 SC로부터 $P_{t+1}(E)=b$이면 $P_{t+1}(P_{t+1}(E)=b)=1$이라는 것이 도출된다.[29] 그러나 이것이 과연 우리 믿음의 정도의 합리성을 위한 필요조건인지 의심스럽다. 자신의 마음을 정확하게 모르는 경우는 불가능하지 않으며, 그렇다고 해서 비합리적이라고 말할 수 있는지 의심스럽다. 하지만, 다행스러운 것은 SC가 선명성 자체를 함축하

---

29) SC를 가정하자. 그리고 어떤 행위자가 시점 $t+1$의 경험에 의해서 E라는 명제에 대한 믿음의 정도를 $P_t(E)(=a)$에서 $P_{t+1}(E)(=b)$로 수정했다고 하자. 그럼 다음이 성립한다.
(†) 모든 X에 대해서, $P_{t+1}(X)=P_t(X|P_{t+1}(E)=b)$.
그럼 (†)로부터 $P_{t+1}(P_{t+1}(E)=b)=P_t(P_{t+1}(E)=b|P_{t+1}(E)=b)=1$이 도출된다. 따라서, 우린 $P_{t+1}(E)=b$라면 $P_{t+1}(P_{t+1}(E)=b)=1$이라고 결론 내릴 수 있다. 하지만, SC가 다음을 함축하는 것은 아니다: 모든 X에 대해서, $P_{t+1}(X)=x$라면 $P_{t+1}(P_{t+1}(X)=x)=1$. (†)에 의해서, 임의의 X에 대해서 다음이 성립한다: $P_{t+1}(P_{t+1}(X)=x)=P_t(P_{t+1}(X)=x|P_{t+1}(E)=b)$. 하지만 이 값이 1이라고 결론 내릴 수 없다.

는 것은 아니라는 점이다. 즉 SC가 함축하는 것은 경험에 의해서 직접적으로 수정된 E에 대한 선명성만을 함축할 뿐이다. SC는 합리적인 행위자는 모든 명제에 대한 자신의 믿음을 선명하게 알고 있다는 것은 함축하지 않는다. 이런 점에서 선명성에 대한 우려는 완화된다고 할 수 있다.

<부록>

3절에서 제시된 나의 논증은 제프리 조건화를 통해서 믿음을 갱신할 때 다음 두 명제가 충돌한다는 것이었다.

(i) $P_t(E)=a\neq b=P_{t+1}(E)$;

(ii) $P_t(E|H)=P_{t+1}(E|H)=b$.

그러나, 심사위원이 지적한 대로 해당 행위자가 경험 이후에 E에 대한 믿음의 정도를 직접적으로 수정했을 뿐만이 아니라 내적 관찰(introspection)에 의해서 H를 확신하게 되었다면, (i)과 (ii)는 충돌하지 않는다. 이 경우, 해당 행위자는 분할 집합(partition) $\{E\&H, \sim E\&H\}$의 각 원소에 대한 믿음의 정도를 (내적, 외적 경험에 의해서) 직접적으로 수정하고, 이렇게 직접적으로 수정된 $E\&H$와 $\sim E\&H$에 대한 새로운 믿음의 정도와 제프리 조건화를 이용해서 다른 명제에 대한 새로운 믿음의 정도를 갱신한다. (i)에 의해서 $P_{t+1}(E)=b$이고 행위자는 내적 경험을 통해 H를 확신하게 되었기에 $P_{t+1}(H)=1$이다. 따라서, $P_{t+1}(E|H)=P_{t+1}(E)=b$이다. 그러므로, (ii)가 성립한다. (3절의 논증에서 해당 행위자가 t 시점에 반영원리를 만족한다는 것이 가정되었다. 즉 $P_t(E|H)=b$라는 것이 가정되었다.) 결과적으로, 문제의 행위자가 내적 경험에 의해서 H를 확신하게 되는 경우, (i)과 (ii)는 충돌하지 않는다.

그러나 이런 사실이 제프리 조건화와 반영 원리가 충돌한다는 나의 주장을 반박하는 것은 아니다. 심사위원이 지적한 대로, 어떤 행위자가 분할 집합(partition) $\{E\&H, \sim E\&H\}$의 각 원소에 대한 믿음의 정도를 (내적, 외적 경험에 의해서) 직접적으로 수정하였다고

하자. 그 결과, $P_{t+1}(H)=1$, $P_{t+1}(E)=b$가 되었다고 가정하자. 그럼 간단한 계산에 의해서, 다음이 성립한다.

$$P_{t+1}(E\&H)=P_{t+1}(E)P_{t+1}(H|E)=b;$$
$$P_{t+1}(\sim E\&H)=P_{t+1}(\sim E)P_{t+1}(H|\sim E)=1-b.$$

기호를 단순하게 하기 위해서, E^*가 $E\&H$를 가리킨다고 하자. 그럼 $\sim E^*$는 $\sim E\&H$를 가리킨다. (이것은 $\{E\&H, \sim E\&H\}$가 분할 집합이기 때문이다.) 이제 제프리 조건화로부터 다음이 도출된다.

(iii) 모든 X에 대해서, $P_{t+1}(X)=bP_t(X|E^*)+(1-b)P_t(X|\sim E^*)$.

그럼 다시 물어보자. 다음 두 명제는 양립가능한가?

(i') $P_t(E^*)\neq P_{t+1}(E^*)=b$;
(ii') $P_t(E^*|H^*)=P_{t+1}(E^*|H^*)=b$.

여기서 H^*는 $P_{t+1}(E^*)=b$를 가리킨다. 우리는 3절에서 제시된 것과 비슷한 논증을 통해서, (i')과 (ii')이 양립불가능하다는 것을 보일 수 있다. 따라서, 문제의 행위자가 내적 경험에 의해서 H를 확신하게 되더라도 여전히 제프리 조건화와 반영원리가 충돌한다고 할 수 있다.

이 답변에 대해 아마도 심사위원은 내적 관찰에 의해서 행위자는 H 뿐만이 아니라 H^*도 확신하게 된다고 응답할 것이다. 그러나 이렇게 묻는다고 해도 제프리 조건화와 반영원리 사이의 충돌은 달라지지 않는다. 왜냐하면 H^*에 대한 믿음의 정도보다 더 고차의 믿음의 정도를 이용해서 제프리 조건화와 반영원리 사이의 충돌을 증명할 수 있기 때문이다. 그리고 이런 과정은 무한하게 진행될 것

이다. 중요한 것은 이차 조건화는 이런 특징을 가지지 않는다는 것이다. 따라서 고차 믿음의 정도를 수정하는 데 있어 이차 조건화가 제프리 조건화보다 더 좋은 규칙이라는 나의 주장은 변하지 않는다.

참고문헌

이병덕(2000), “무어의 역설과 합리성”,『철학』63집, pp. 257-276.

Armendt, B. (1980), “Is There a Dutch Book Argument for Probability Kinematics?”, *Philosophy of Science* 47, pp. 583-589.

Briggs, R. (2009), “Distorted Reflection”, *Philosophical Review* 118(1), pp. 59-85.

Chalmers, D. and Hájek, A. (2007), “Ramsey+Moore=God”, *Analysis* 67, pp. 170-172.

Gaifman, H. (1986), “A theory of higher order probabilities”, *Theoretical Aspects of Reasoning about Knowledge*: *Proc*. 1*986 Conf*., ed. J. Y. Halpern, pp. 275-292.

Greaves, H. and Wallace, D. (2006), “Justifying Conditionalization: Conditionalizatin Maximizes Expected Epistemic Utility”, *Mind* 115, pp. 607-632.

Hájek, A. (2008), Interview, *Epistemology: 5 Questions*, eds. Vincent F. Hendricks and Duncan Pritchard, Automatic Press.

Jeffrey, R. (1983), *The Logic of Decision*. Univ. of Chicago Press.

Jeffrey, R. (2002), *Subjective Probability: The real thing*, Cambridge Univ. Press.

Joyce, J. (2003), “Bayes’ Theorem”, *The Stanford Encyclopedia of Philosophy* (2008 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/ entries/bayes-theorem/>

Maher, P. (1993), *Betting on Theories*, Cambridge Univ. Press.

Ramsey, F. (1926), "Truth and Probability" in Henry Kyburg Jr. and Howard Smokler eds. (1980), *Studies in Subjective Probability*.

Skyrms, B. (1980a), *Causal Necessity*. New Haven.

Skyrms, B. (1980b), "Higher order degrees of belief", *In Prospects for Pragmatism: Essays in Honor of F. P. Ramsey*. D. H. Mellor, ed., Cambridge University Press, pp. 109-137.

Skyrms, B. (1984), *Pragmatics and Empiricism*, Yale University Press.

Skyrms, B. (1987), "Dynamic Coherence and Probability Kinematics", *Philosophy of Science* 54, pp. 1-20.

Teller, P. (1973), "Conditionalization and Observation", *Synthese* 26, pp. 218-258.

van Fraassen, B. (1984), "Belief and the Will", *The Journal of Philosophy* **81**, pp. 235-256.

van Fraassen, B. (1986), "A Demonstration of the Jeffrey Conditionalization Rule", *Erkenntnis* 24, pp. 17-24.

van Fraassen, B. (1989), *Laws and Symmetry*. Oxford: Clarendon Press.

van Fraassen, B. (1995), "Belief And The Problem Of Ulysses And The Sirens", *Philosophical Studies* 77, pp. 7-37.

van Fraassen, B. (1999), "Conditionalization, A New Argument for", *Topoi* 18, pp. 93-96.

Weisberg, J. (2007), "Conditionalization, Reflection, and

Self-knowledge." *Philosophical Studies* 135, pp. 179-197.

Williams, P. M. (1980), "Bayesian Conditionalisation and the Principle of Minimum Information", *The British Journal for the Philosophy of Science* 31, pp. 131-144.

호주국립대학 방문 연구원
Australian National University, Visiting Fellow
ipark.phil@gmail.com

Updating Higher Order Credences by Conditionalization

Ilho Park

This paper concerns several versions of conditionalization. In particular, I will examine the relationship between Jeffrey conditionalization and the second order conditionalization concerning updating higher order credences. In section 2, I suggest explicitly what Jeffrey conditionalization and the second order conditionalization are. I will argue in section 3 that Jeffrey conditionalization conflicts with van Fraassen's Reflection Principle while the second order conditionalization doesn't. And I will also argue in section 4 that under some situations, Jeffrey conditionalization may lead agents to Moorean absurdity while the second conditionalization may not. As a result, I will claim that Jeffrey conditionalization is better than the second order contionalization at updating our higher order credences.

Key Words: Jeffrey conditionalization, Second order conditionalization, Reflection principle, Moorean absurdity.