

고차 조건화와 믿음 기반 약화* **

박 일 호

【국문요약】 이 논문의 목표는 표준적인 베이즈주의가 믿음 기반 약화 증거 (undermining evidence)에 의해서 촉발된 믿음 갱신을 잘 다룰 수 없다는 와이즈버그의 주장에 답변하는 것이다. 우리의 인식론적인 직관에 따르면, 믿음 기반 약화 증거는 몇몇 관련된 신념도를 감소시켜야 하는 듯하다. 하지만 와이즈버그에 따르면 그런 믿음 변화는 표준적인 믿음 갱신 규칙, 즉 (제프리) 조건화를 통해서서는 이루어질 수 없다. 그 이유는 (제프리) 조건화를 통해서서는 일부 명제들 사이에 성립하는 확률적 독립성 관계가 보존되기 때문이다. 그러나 나는 이 논문에서 그러한 반베이즈주의적인 결론은 다소 성급하다고 주장할 것이다. 특히, 나는 다른 종류의 조건화가 또 있으며, 그 조건화를 이용하면 믿음 기반 약화 증거를 통한 믿음 갱신도 충분히 베이즈주의적 이론들 속에서 다루어질 수 있다는 것을 논증할 것이다. 그러한 조건화는 종종 ‘고차 조건화’라고 불리는 것이다.

【주요어】 단순 조건화, 제프리 조건화, 고차 조건화, 믿음 기반 약화, 확률적 독립성, 고정성, 믿음 갱신

접수일자: 2015.02.23 심사 및 수정완료일: 2015.03.30 게재확정일: 2015.05.15

* 이 논문은 2014년도 전북대학교 연구기반 조성비 지원에 의하여 연구되었음.

** 엄격한 심사를 통해 몇 가지 중요한 점을 지적해주신 심사위원들에게 감사의 인사를 드린다. 비록 심사위원들의 지적이 본 논문에 완벽히 반영되지는 않았으나, 그들이 제기한 주요 논점들은 필자의 추후 연구의 중요 지점이 될 것이라 확신한다.

1. 서론

베이즈주의 인식론에 대한 철학자들의 관심은 최근 들어 더욱 늘어나고 있다. 예전에 베이즈주의 인식론에 적극적인 관심을 보인 대부분의 학자들이 과학 철학자들이었다면, 최근의 관심은 철학의 다양한 분야에 걸쳐 있다. 가령, 증언, 기적 등을 다루는 종교 철학, 인식적 정당성이나 회의주의와 같은 주제를 다루는 전통 인식론, 심지어 분석/종합 구분 등을 다루는 언어 철학까지, 철학의 다양한 영역이 베이즈주의적 철학 방법론을 이용하여 관련 문제들을 해결하고 검토하려는 시도를 하고 있다.¹⁾

많은 주목을 받게 된 대부분의 연구 프로그램들과 마찬가지로, 베이즈주의 인식론에 대한 관심의 증가는 그것의 한계와 문제점을 드러내고 있다. 그리고 그런 한계와 문제점에 직면한 베이즈주의자들은 자신들의 연구 프로그램을 점점 세련되게 수정·보완·확장하고 있다. 예를 들어, 잡자는 미녀 문제는 표준적인 베이즈주의 인식론이 소위 ‘일인칭적 믿음(belief de se)’과 그것의 변화를 다룰 수 없다는 것을 보여주고 있으며, 이에 대해 몇몇 인식론자들은 베이즈주의의 중요 논제들을 수정·보완하여 베이즈주의 이론틀 속에서 관련 믿음들을 다룰 수 있는 방안을 마련하고자 한다. 또, 베이즈주의가 전통 인식론의 한 가지 입장인 지각적 독단론과 충돌한다는 주장은 베이즈주의자들로 하여금 지각적 독단론의 문제점을 지적하거나, 혹은 지각적 독단론과 양립가능한 방식으로 베이즈주의를 수정하게끔 만들었다.

본 논문은 위와 같은 식으로 진행되는 베이즈주의 인식론의 이

1) 종교 철학과 베이즈주의에 대해서는 Chandler and Harrison (2013)을 보라. 회의주의와 베이즈주의에 대해선 Weatherson (2007), White (2006) 등을 보라. 그리고 분석/종합 구분과 베이즈주의에 대해서는 Chalmers (2011)을 보라.

론적 전개 양상의 한 사례를 보여줄 것이다. 즉 나는 베이즈주의 인식론에 가해진 의미 있는 한 가지 비판을 소개할 것이며, 이 비판에 대해서 어떻게 베이즈주의에 호의적인 방식으로 답할 수 있을지 논증할 것이다.

이 논문이 고려하고 있는 비판은 ‘믿음 기반 약화(undermining)’라는 다소 전통 인식론적 주제에서 온 것이다. 우리가 특정 명제에 대해서 가지고 있었던 믿음의 정당성은 여러 방식으로 약화될 수 있다. 가령, 그 명제를 직접적으로 반박하는 증거를 획득하였을 때, 그 명제를 믿는 것에 대한 정당성은 약화될 것이다. 물론, 직접적으로 반박하는 증거를 획득하지 못하여도 믿음의 정당성은 약화될 수 있다. 예를 들어, 믿음을 형성하게 된 과정의 신빙성(reliability)이 의심스러워진다면, 직접적으로 반박하는 증거가 없다고 하더라도 해당 믿음의 정당성은 약화된다. 직접적인 관찰을 통하여 앞에 있는 종이가 노란색이라는 믿음을 형성하였다고 하자. 하지만 얼마 뒤 주변 불빛이 왜곡되어 있다거나 혹은 관찰자의 시신경이 어떤 문제를 가지고 있다는 사실이 밝혀지면, 앞에 있는 종이가 노란색이라는 믿음의 정당성은 약화된다. 이렇게 믿음을 형성하게 된 과정의 신빙성을 약화시키는 것은 흔히 ‘믿음 기반 약화’라고 불리며, 이때 사용된 증거는 ‘믿음 기반을 약화하는 증거(undermining evidence)’라고 불린다.

그럼 이런 믿음 기반 약화는 베이즈주의 인식론에 어떤 문제를 야기하는가? 이 질문에 대한 답은 베이즈주의 인식론의 핵심적인 요소인 ‘조건화(conditionalization)’라는 믿음 갱신 규칙과 관련 있다. 베이즈주의 인식론자들은 우리의 믿음은 정도의 문제이며, 믿음의 정도는 확률 계산 규칙을 만족해야 한다고 말한다. 그들은 확률 계산 규칙을 만족하는 믿음의 정도를 신념도(credence)라고 부른다. 이와 더불어 베이즈주의 인식론은 신념도의 합리적 변화를 규정하

는 규칙도 가지고 있다. 조건화가 바로 그 규칙이다. 다음은 조건화의 두 가지 버전이다:²⁾

단순 조건화(Simple Conditionalization, SC).

경험에 의해서 B 를 확신하게 되었을 때, 즉 $p'(B) = 1$ 일 때, 모든 A 에 대해서 다음이 성립한다:

$$p'(A) = p(A|B). \quad (\text{단 } p(B) > 0 \text{일 때.})$$

제프리 조건화(Jeffrey Conditionalization, JC).

경험에 의해서 B 에 대한 신념도가 직접적으로 수정되었을 때, 즉 $p(B) \neq p'(B) = b$ 일 때, 모든 A 에 대해서 다음이 성립한다:

$$p'(A) = p'(B)p(A|B) + p'(\neg B)p(A|\neg B).$$

(단 $0 < p(B) < 1$ 일 때.)

여기서 p 는 B 에 대한 신념도를 직접적으로 바꾸게 만드는 경험을 하기 전의 신념도 함수이며, p' 는 그 경험 이후의 신념도 함수이다. 잘 알려져 있듯이, SC는 JC의 특수 사례이며 JC는 SC의 일반화이다. SC는 확실한 증거만을 다루지만 JC는 확실한 증거는 물론 불확실한 증거, 즉 해당 증거가 참이라는 것에 대한 신념도가 1보다 작은 증거도 다룰 수 있다.

‘믿음 기반 약화’의 베이즈주의에 대한 공격은 이 조건화 규칙을 표적으로 한 것이다. 즉 믿음의 기반을 약화하는 증거가 야기하는 신념도의 변화를 조건화가 제대로 다룰 수 없다는 것이다. 이 논변을 처음으로 제시하고 발전시킨 학자는 조나단 와이즈버그

2) ‘조건화’를 서술할 때, ‘그 밖에 경험에 의해서 직접적으로 수정된 신념도는 없다’는 조건이 덧붙여진다. 그러나 아래 SC와 JC의 정식화에서는 서술의 단순성을 위해서 그런 조건을 생략하였다. 특별한 이야기가 없다면, 나는 이 조건이 성립한다는 것을 채 논의를 진행할 것이다.

(Jonathan Weisberg, 2009, 2015)이며, 몇몇 학자들(Gallow 2014; Wagner 2013)이 이에 관련된 논쟁을 진행하고 있다.

본 논문을 통해, 나는 와이즈버그의 주장과 달리 베이즈주의자들은 믿음 기반 약화를 충분히 잘 다룰 수 있으며, 그 방법은 베이즈주의 인식론의 핵심에서 크게 벗어나지 않는다는 것을 보일 것이다. 특히, 나는 믿음 기반 약화를 다룰 수 있도록 조건화 규칙을 재정식화할 것이며, 이 재정식화가 기존 SC나 JC로부터의 극단적인 이탈이 아니라는 것을 설명할 것이다. 이를 위해서 본 논문은 다음과 같이 구성된다. 우선 2절에서 베이즈주의가 믿음 기반 약화를 다룰 수 없다는 와이즈버그의 논증이 설명된다. 이때 나는 이런 비판에 대해 답하기 위해서 필요한 것이 무엇인지 분명히 할 것이다. 3절은 새로운 형태의 베이즈주의 믿음 갱신 규칙을 제시하고 설명한다. 나는 이 새로운 규칙을 고차 조건화(Higher Order Conditionalization)라고 부를 것이다. 그리고 4절에서 고차 조건화가 믿음 기반 약화를 어떻게 다루는지 서술할 것이다.

2. 와이즈버그의 논증: 조건화는 믿음 기반 약화를 다룰 수 없다

평소에 붉은 양말을 즐겨 신지 않는 철수는 서랍장 안에 있는 양말이 붉은 색이라는 것을 약하게 믿고 있었다. 그런데, 서랍장을 열어 양말을 관찰한 뒤 그 양말이 붉은 색이라는 것을 꽤 강하게 믿게 되었다. 즉 양말이 붉은 것처럼 보이는 철수의 시각 경험이 양말이 붉은 색이라는 명제에 대한 철수의 믿음의 정도를 1에 가까운 값으로 증가시킨 것이다. 하지만 잠시 뒤 철수는 주변의 불빛이 잘못된 시각 경험을 야기할 정도로 왜곡되었다는 것을 알게 되었다. 이 경우, 그 양말이 붉은 색이라는 명제에 대한 철수의 신념

도는 어떻게 되어야 하는가? 우리의 자연스러운 인식론적 직관은 그 양말이 붉은 색이라는 것에 대한 철수의 신념도는 감소해야 한다고 말하는 듯하다. 즉 불빛이 왜곡되었다는 것은 양말의 색에 대해 철수가 가지고 있었던 믿음의 기반을 흔드는 것이며, 따라서 양말이 붉은 색이라는 것에 대한 철수의 믿음은 약화되어야 한다.

위 ‘철수의 사례’는 전형적인 믿음 기반 약화 사례를 묘사하고 있다. 그럼 과연, 베이즈주의 인식론은 이 사례 속 믿음의 변화를 제대로 포착할 수 있을까? 와이즈버그는 그렇지 않다고 논증한다. 즉 베이즈주의의 조건화 규칙대로 믿음을 갱신한다면 위 사례 속 믿음의 변화와 다른 결과가 나온다는 것이다. 이를 확인하기 위해, 철수의 사례가 묘사하는 믿음의 변화를 다음과 같이 형식적으로 나타내보자.³⁾

$$(2.1) \quad p(E) = a \approx 0;$$

$$(2.2) \quad p'(E) = b \approx 1;$$

$$(2.3) \quad p''(F) = 1;$$

$$(2.4) \quad p''(E) < b;$$

$$(2.5) \quad p(F|E) = p(F).$$

여기서 p 는 철수가 서랍 안에 있는 양말의 색에 대한 어떤 경험도 하기 전에 가지고 있었던 신념도 함수이다. p' 은 철수가 양말의 색을 직접 경험한 뒤에 가지게 된 신념도 함수이다. 마지막으로 p'' 은 철수가 기반 약화 증거, 즉 주변 불빛이 왜곡되었다는 증거를 획득한 이후의 신념도 함수이다. E 는 ‘양말은 붉은 색이다.’는 명제, F 는 ‘주변의 불빛이 왜곡된 시각 경험을 야기할 정도로 잘못되었다.’

³⁾ 와이즈버그의 원래 논증(Weisberg 2015, pp. 125-126)을 약간 변형한 것이다. 이런 변형은 아래 논의에서 유의미한 차이를 만들어내지 않는다.

는 기반 약화 증거를 나타낸다. (2.1)~(2.4)는 위의 사례의 주요 신념도(믿음)와 그것들의 변화를 그대로 나타낸 것이다. (2.1)은 서랍 안에 있는 양말의 색에 대한 아무런 경험도 하기 전에는 양말이 붉은 색이라는 것에 대한 철수의 신념도가 0에 가까웠다는 사실, (2.2)는 양말의 색을 경험한 이후 양말이 붉은 색이라는 것에 대한 철수의 신념도가 1과 가까운 b 라는 값으로 수정되었다는 사실⁴⁾, (2.3)은 주변 불빛이 왜곡되었다는 정보를 획득했다는 사실을 나타낸다. 마지막으로 (2.4)는 불빛이 왜곡되었다는 정보를 획득한 이후 철수의 E 에 대한 신념도가 감소했다는 사실을 나타낸다.

와이즈버그의 논증에 있어서 중요한 역할을 하는 것은 (2.5)다. 이것은 양말의 색과 관련된 어떤 경험도 하기 전의 철수에게 있어, 양말의 색과 불빛의 상태는 서로 상관이 없다는 것을 확률적으로 나타낸 것이다. 와이즈버그는 이 가정이 그리 이상하지 않고 자연스럽다고 말한다. 사실 양말의 색과 불빛이 가진 특징 사이에는 어떤 인과적 연결도 없다. 양말의 색은 불빛이 왜곡되었다는 것에 대

4) 익명의 심사위원 중에 한 분은 경험 이후 신념도가 직접적으로 수정된 명제는 양말이 붉은 색이라는 것이 아니라, 양말이 붉은 색으로 보였다는 것(이 명제를 D 라고 하자)이라고 지적했다. 덧붙여, 그렇게 생각할 경우 믿음 갱신은 $p'(\cdot) = p'(D)p(\cdot/D) + p'(\neg D)p(\cdot/\neg D)$ 라는 식을 이용해서 이루어져야 한다고 지적하였다. 이런 논점에 대해서는 두 가지를 지적할 수 있다. 첫 번째는 양말이 붉은 색으로 보인다는 것에 대한 신념도는 1과 0 사이의 값을 가질 수 없을 것 같다는 점이다. 여기서 이와 관련된 몇몇 논의, 특히 지각적 독단론을 둘러싼 논의에서 심사위원이 말하는 종류의 명제들에 대한 신념도는 모두 1의 값이 할당된다는 사실을 주목할 필요가 있다. 가령, White (2006)을 보라. 두 번째는 D 에 1의 값이 할당되는 것이 자연스럽고, 이에 (심사위원이 제시한 것과 달리) JC가 아니라 SC를 적용해 믿음을 수정해야 한다면 그것은 결국 고차조건화를 적용하는 것과 유사하다는 점도 지적해야 한다. 뒤에서 설명하겠지만, 고차조건화는 경험에 의해서 E 에 대한 신념도가 직접적으로 수정되어 b 가 되었을 때 $q(E) = b$ 라는 새로운 명제에 대해서 SC를 적용하라는 것이다. 여기서 명제 $q(E) = b$ 는 심사위원이 말한 명제 D 와 유사한 역할을 한다고 말할 수 있다.

한 원인도 아니며, 결과도 아니다. 따라서 와이즈버그의 이런 가정은 어색하지 않고 받아들일만하다. 이 점에 있어서 필자 역시 와이즈버그의 입장에 동의한다.

그럼, 철수가 조건화 규칙을 통해 신념도를 수정한다면 (2.1)~(2.5)가 모두 만족될 수 있을까? 그렇지 않다. 와이즈버그의 논증에 따르면 베이즈주의 믿음 갱신 규칙, 즉 JC와 SC는 (2.1)~(2.5)과 함께 모순을 야기한다. 이 모순은 다음과 같이 증명될 수 있다.

증명.

우선, SC, JC, (2.2), (2.3)은 다음을 함축한다.⁵⁾

$$(2.6) \quad p''(E) = \frac{p'(E \wedge F)}{p'(F)}$$

$$= \frac{p'(E)p(E \wedge F|E) + p'(\neg E)p(E \wedge F|\neg E)}{p'(E)p(F|E) + p'(\neg E)p(F|\neg E)}$$

그럼, (2.2), (2.5), (2.6) 그리고 확률계산규칙에 의해서, 다음이 성립한다.

$$(2.7) \quad p''(E) = \frac{p'(E)p(F)}{p'(E)p(F) + p'(\neg E)p(F)} = p'(E) = b$$

그러나 이 결과는 (2.4)와 충돌한다. 따라서 SC와 JC를 포함하는

⁵⁾ JC와 (2.2)에 의해서 다음이 성립한다:

(i) 모든 X 에 대해서 $p'(X) = p'(E)p(X|E) + p'(\neg E)p(X|\neg E)$.

그리고 SC와 (2.3)에 의해서 다음이 성립한다.

(ii) 모든 X 에 대해서 $p''(X) = p'(X|F) = p'(X \wedge F)/p'(F)$.

위 (i)과 (ii)는 (2.6)을 함축한다.

베이즈주의와 (2.1)~(2.5)는 모순을 야기한다. 증명 끝.

그럼 무엇이 문제인가? 와이즈버그가 제시한 가정들 (2.1)~(2.5)는 그 자체로 일관적이며, 믿음 기반 약화와 관련된 우리의 인식론적 직관과 크게 어긋난 것처럼 보이지 않는다. 이 점을 인정한다면, 우린 모순의 원인을 베이즈주의 믿음 갱신 규칙, 즉 조건화에서 찾아야 할 것이다.

그렇다면, 조건화의 어떤 특징이 위와 같은 문제를 야기했는가? 이에 대해서 와이즈버그는 다음과 같이 진단한다.

직관적으로 볼 때, 불빛의 왜곡과 양말의 색은 양말을 관찰하기 이전에는 서로 무관하며, 양말을 관찰한 이후에야 부정적인 상관관계에 있게 된다. 그러나 해당 시간 사이의 믿음 갱신이 [양말이 붉은 색이라는 명제]에 대해 고정적(rigid)이라면 이와 같은 일은 일어날 수 없다. 고정성은 E 와 F 사이의 부정적 상관관계가 도입되는 것을 막는다. (Weisberg 2015, p. 126. 괄호 []는 필자가 추가한 것이다.)

위 인용문에서 고정성(rigidity)은 경험에 의해서 어떤 명제에 대한 신념도가 직접적으로 수정되었을 때, 그 명제를 조건으로 하는 조건부 신념도는 고정된 채 변하지 않는다는 것이다. 좀 더 형식적으로 나타내자면, 다음과 같다.

고정성.

경험에 의해서 B 에 대한 신념도가 직접적으로 수정되었을 때, 즉 $p(B) \neq p'(B)$ 일 때, 모든 A 에 대해서 다음이 성립한다. $p(A|B) = p'(A|B)$ 이며 $p(A|\neg B) = p'(A|\neg B)$ 이다.

위와 같은 고정성이 성립한다는 것은 종종 ' p 에서 p' 으로의 믿음

갱신은 B 에 대해 고정적이다.’라고 표현되곤 한다. 그렇다면, 이제 우리는 위 인용문에서 ‘해당 시간 사이의 믿음 갱신이 양말이 붉은 색이라는 명제에 대해 고정적이다.’는 것이 무엇을 뜻하는지 이해할 수 있다. 이 말은 양말의 색에 대한 경험 이후 p 에서 p' 로의 믿음 갱신이 일어났을 때, 양말이 붉은 색이라는 명제를 조건으로 하는 조건부 신념도는 고정된 채 변하지 않는다는 것이다.

그렇다면, 인용문의 ‘고정성은 E 와 F 사이의 부정적 상관관계가 도입되는 것을 막는다.’는 것은 무슨 말인가? 이를 이해하기 전에 우선, 위 고정성이 JC와 동치라는 것을 지적할 필요가 있다. 고정성과 JC가 동치라는 사실은 베이즈주의 인식론자들 사이에 널리 알려져 있으며, 이런 사실은 JC의 몇몇 인식론적 특징을 규명하는데 중요한 역할을 해왔다. 그렇다면 ‘고정성은 E 와 F 사이의 부정적 상관관계가 도입되는 것을 막는다.’는 말은 ‘JC에 의한 믿음 갱신을 통해서 E 와 F 사이의 부정적 상관관계가 도입될 수 없다.’는 것과 같은 말이 된다. 이 말이 의미하는 바는 무엇인가?

이에 대해 답하기 위해서는 ‘고정성은 믿음 갱신에서 특정 명제들 사이에 성립하는 확률적 독립성을 보존한다.’, 혹은 다른 말로 ‘JC에 의한 믿음 갱신에서 특정 명제들 사이에 성립하는 확률적 독립성은 보존된다.’는 사실을 이해할 필요가 있다. 좀 더 형식적으로 말하자면 다음과 같다.

확률적 독립성의 보존성.

p 에서 p' 으로의 믿음 갱신이 B 에 대해 고정적인 경우, 임의의 A 에 대해서 다음이 성립한다.

$$(2.8) \quad \text{만약 } p(A|B) = p(A|\neg B) \text{ 라면, } p'(A|B) = p'(A|\neg B) \text{ 이다.}$$

우선 이 진술은 다음과 같이 증명된다는 것을 확인하자.

증명.

p 에서 p' 으로의 믿음 갱신이 B 에 대해 고정적이라고 가정하자. 그럼 다음이 성립한다.

$$(2.9) \quad p(A|B) = p'(A|B); \quad p(A|\neg B) = p'(A|\neg B).$$

그리고 A 에 대해서 (2.8)의 전건이 성립한다고 가정하자. 즉 다음이 성립한다.

$$(2.10) \quad p(A|B) = p(A|\neg B).$$

그럼 (2.9)와 (2.10)으로부터 다음이 도출된다.

$$(2.11) \quad p'(A|B) = p'(A|\neg B).$$

즉 (2.8)의 후건이 성립한다. 증명 끝.

위에서 (2.8)은 B 와 확률적으로 독립적인 명제들은 믿음 갱신 이후에도 B 와 확률적으로 독립적이라는 것을 뜻한다. 그럼 우린 ‘고정성은 E 와 F 사이의 부정적 상관관계가 도입되는 것을 막는다.’가 무엇을 말하는지 이해할 수 있다. (2.5)에 따르면, 양말을 관찰하기 이전에 E 와 F 는 확률적으로 독립적이었다. 그리고 양말의 색을 관찰한 뒤 JC를 통해 진행된 믿음 갱신은 E 에 대해서 고정적이었다. 따라서 ‘확률적 독립성의 보존성’에 의해서 조건화 이후에도 E 와 F 는 확률적으로 독립적일 수밖에 없다. 다른 말로, JC를 통해서 믿음이 갱신되었다는 사실은 경험 이전에 확률적으로 독립적이었던 E 와 F 가 경험 이후에도 여전히 확률적으로 독립적인 관계를 유지

하도록 하며, 이에 둘 사이에 부정적 상관관계가 도입될 수 없게 만든다. 바로 이것이 와이즈버그가 ‘고정성은 E 와 F 사이의 부정적 상관관계가 도입되는 것을 막는다.’는 말로 의도한 것이다.

요약하자면 다음과 같다. 철수의 사례를 베이즈주의자들이 다룰 수 없는 이유는 베이즈주의 근본 원리의 중요한 특징에서 비롯된다. 그 특징은 베이즈주의 믿음 갱신 규칙인 (제프리) 조건화를 통해 믿음을 갱신하면 특정 명제들 사이에 성립하는 확률적 독립성이 보존된다는 것이다. 하지만 철수의 사례와 같은 믿음 기반 약화 증거를 획득한 경우에서, 경험 이전 확률적으로 독립적이었던 두 명제는 경험 이후 확률적으로 부정적 상관관계를 가져야 한다. 따라서 베이즈주의는 철수의 사례를 적합하게 다룰 수 없다.

그렇다면, 베이즈주의자들은 이 문제에 대해서 어떻게 답할 수 있는가? (제프리) 조건화 규칙을 훼손하지 않으면서, (2.1)~(2.5)에 맞추어 철수의 사례를 다룰 수 있는가? 한 가지 전략은 (제프리) 조건화가 잘못 적용되었다고 논증하는 것이다. 철수의 사례에서는 명시적으로 드러나지 않은 추가적인 증거, 혹은 추가적인 고려사항이 있으며, 이런 것들까지 포함하여 (제프리) 조건화를 적용해야 하고, 그런 경우 베이즈주의를 괴롭히는 모순은 발생하지 않는다는 식으로 논증하는 것이다.

하지만 이런 식의 답변은 다소 설득력이 떨어진다. 일단 ‘추가 증거가 있다’ 혹은 ‘추가 고려 사항이 있다’는 것은 와이즈버그가 원래 제시한 사례를 훼손할 수가 있다. 뿐만 아니라 철수의 사례를 일반화시킨 상황을 고려했을 때, 위와 같은 대응책이 의미가 있으려면 철수의 사례와 유사한 사례에서는 언제나 해당 추가 고려 사항이 있어야 한다는 것을 논증해야 하는데, 그런 논증은 우리 믿음의 실제 심리학적 작동으로부터 너무 많이 벗어나는 위험에 직면할 수 있다. 와이즈버그 역시 위와 같은 전략들에는 무엇이 있을지 예

상하고, 각 전략에 대해서 필자와 유사한 비판을 제시하고 있다. (Weisberg 2015, pp.140-146.)

나는 다음 절에서 위의 전략과 다른 방식으로 와이즈버그의 논증에 답변할 것이다. 필자의 답변은 특별한 새로운 정보가 있다는 것을 이용하지 않는다. 그보다 새로운 형태의 베이즈주의 믿음 갱신 규칙을 제시하여, 베이즈주의가 어떻게 철수의 사례를 다룰 수 있는지 살펴볼 것이다. 문제의 새로운 믿음 갱신 규칙은 종종 ‘고차 조건화’라고 불리는 것이다.

그럼, 고차 조건화를 이용하는 나의 전략이 설득력을 갖추기 위해서는 무엇을 보여주어야 하는가? 최소한 다음 두 가지를 보여줄 필요가 있다.

- (a) 고차 조건화에 의한 신념도의 갱신은 (2.1)~(2.5)와 양립가능하다.
- (b) 신념도를 고차 조건화를 이용해 수정한다면, 경험 이전 E 와 F 가 확률적으로 독립적이라고 하더라도 경험 이후엔 E 와 F 사이에 부정적 상관관계가 성립할 수 있다.

철수의 사례의 문제를 해결하기 위해서 (a)만을 보여주는 것으로는 충분치 못하다. (a)를 보여주는 것은 그저 고차 조건화가 철수의 사례를 다룰 수 있다는 것만을 보여줄 뿐이다. 하지만 왜 고차 조건화는 (a)를 다룰 수 있는지에 대한 설명을 위해서는 (b)에 대한 논증이 필요하다. 왜냐하면 (b)를 보여주었을 때 고차 조건화가 (a)를 다룰 수 있다는 것이 우연에 의한 것이 아님을 논증했다고 말할 수 있기 때문이다. 이런 점들은 4절에서 다룰 것이다. 그전에 고차 조건화가 무엇인지 살펴보자.

3. 고차 조건화

1절에서 소개된 단순 조건화, 즉 SC의 적용 범위는 무척 제한적이다. 그것은 획득한 증거가 확실한 경우, 즉 1의 신념도를 가진 경우에만 적용될 수 있다. 이런 적용 범위의 한계를 극복하고자 등장한 것이 제프리 조건화, 즉 JC이다. JC는 획득한 증거가 확실하지 않더라도 적용될 수 있는 규칙이다. 대략적으로 말해, 이 규칙은 명제 B 에 대한 불확실한 증거를 획득하였을 때, 임의의 명제 A 에 대한 새로운 신념도는 사전 조건부 확률 $p(A|B)$ 와 $p(A|\neg B)$ 의 평균이어야 한다고 말하고 있다. 이 평균은 B 에 대한 새로운 신념도를 가중치로 하여 계산된 것이다.

지금껏 많은 베이즈주의자들은 SC의 적용 범위를 확장시키는 데 있어 JC를 우선적으로 고려해왔으며, 그 이외의 대안들에 대해서는 JC에 비해 별 관심을 두지 않았다. 하지만 SC의 적용 범위를 확장시킨 베이즈주의 믿음 갱신 규칙으로 JC만 제안된 것은 아니다. 비록 많은 철학자들의 관심을 받은 것은 아니지만, JC를 제시한 리처드 제프리(Richard Jeffrey)는 물론, 브라이언 스킴즈(Brian Skyrms)도 JC와 다른 방식으로 SC의 적용 범위를 확장하는 방법을 고려하기도 했다. 그 방법이 바로 이 절에서 다루게 될 ‘고차 조건화’이다.⁶⁾

경험에 의해서 B 에 대한 불확실한 증거를 획득하였다고 하자. 즉 경험 이후 B 에 대한 신념도가 $b(< 1)$ 가 되었다고 하자. 고차 조건화의 기본적인 아이디어는 이런 **불확실한 증거의 획득**을 다른 종류의 **확실한 증거의 획득**으로 바꿔 생각하는 것이다. 즉 ‘경험 이후 B 에 대한 신념도가 b 가 되었다’를 ‘경험 이후 B 에 대한 신념도가 b 가 되었다는 것을 확신하게 되었다’로 간주하자는 것이다.

⁶⁾ Jeffrey (2004)와 Skyrms (1980a,b)를 보라.

그럼 불확실한 증거가 확실한 증거로 대체되었으니, SC를 적용하여 신념도를 수정할 수 있는 길이 열리게 된다. 이런 아이디어를 바탕으로, 고차 조건화는 다음과 같이 나타낼 수 있다:

고차 조건화 (Higher Order Conditionalization, HC).

경험에 의해서 B 에 대한 신념도가 직접적으로 수정되었을 때, 즉 $p(B) \neq p'(B) = b$ 일 때, 모든 A 에 대해서 다음이 성립한다.
 $p'(A) = p(A|p'(B) = b)$. (단 $0 < p(p'(B) = b) < 1$ 일 때.)

위에서 잘 드러나듯이, 고차 조건화는, 경험에 의해서 $p'(B) = b$ 가 성립하게 되었을 때 $p'(p'(B) = b) = 1$ 이 성립한 것으로 간주하고, 이 증거에 SC를 적용하여 신념도를 갱신하자는 것이다. 앞에서 말했듯이 제프리파와 스킴즈 모두 이 고차 조건화에 대해 언급하였으나, 상세한 논의를 진행시키지는 않았다. 그들이 세부적인 논의를 진행시키지 않은 이유는, 아마도 많은 경우 JC가 HC와 같은 결과를 도출하기 때문인 듯하다. 그럼 왜 JC와 HC는 ‘많은 경우’ 같은 결과를 도출하는가?

위 질문에 대한 답을 제시하기 위해 다음 두 명제를 고려해보자.

$$H1. \quad p(B|p'(B) = b) = b;$$

$$H2. \quad p(A|B \wedge p'(B) = b) = p(A|B) \text{이고}$$

$$p(A|\neg B \wedge p'(B) = b) = p(A|\neg B).$$

이 중 H1은 밀러 원리(Miller's Principle) 혹은 반영 원리(Reflection Principle)라고 불리는 것이다. 바스 반 프라센(Bas van Fraassen), 패트릭 마허(Patrick Maher), 콜린 하우스(Colin Howson)과 같은 철학자들은 이 원리가 직관적 호소력을 갖춘 적합한 인식

적 원리라고 주장한다. 심지어 이 중 몇몇은 조건화보다 H1이 더 기본적인 인식적 원리라고 말하기도 한다.⁷⁾ 이들이 말하는 H1의 직관적 호소력은 p' 을 가진 행위자가 p 를 가지고 있는 행위자보다 인식적으로 우월하다는 것에 근거를 두고 있다. (망각, 인지적 오작동 등과 같은 것을 논외로 한다면) 우리는 시간이 지날수록 더 많은 경험을 하게 되고, 따라서 미래의 나는 과거의 나보다 더 많은 정보를 가지고 있다는 것이다. 그리고 현재의 나보다 정보를 더 많이 가지고 있는 행위자를 인식적으로 따르는 것은 바람직하다. 따라서 H1과 같은 방식으로 명제 B 에 대한 미래의 의견을 따르는 것은 인식론적으로 정당하다는 것이 방금 언급한 철학자들이 주장이다.

한편, H2는 B (혹은 $\neg B$)라는 조건 아래에서 A 와 $p'(B) = b$ 가 확률적으로 독립적이라고 말하고 있다. 다르게 말하자면, B (혹은 $\neg B$)라는 정보는 A 와 $p'(B) = b$ 사이에 성립할 수 있는 확률적 관련성을 차단한다는 것이다. 사실, 세 명제 사이의 성립하는 H2와 같은 확률적 관계는 베이즈주의 입증 이론의 맥락에서 종종 등장한다. 가령, 몇몇 베이즈주의 입증 이론가들은 ‘증거적 독립성 (Evidential Independence)’을 다음과 같이 정의한다.⁸⁾

증거적 독립성:

$p(A|B \wedge C) = p(A|B)$ 이며 $p(A|\neg B \wedge C) = p(A|\neg B)$ 인 경우, 그리고 그런 경우에만 A 와 C 는 B 와 관련해 증거적으로 독립적이

⁷⁾ Howson (2014); Maher (1993); van Fraassen (1984)를 보라.

⁸⁾ 가령, Bovens and Hartmann (2003); Earman (2000); Olsson (2005); Shogenji (2007), Wheeler (2009)를 보라. Wheeler (2009, p. 88)가 제시하는 증거적 독립성에 대한 정의의 원문은 다음과 같다:

A is evidentially independent of B with respect to h if and only if $\Pr(A|B \wedge h) = \Pr(A|h)$, and $\Pr(A|\neg B \wedge h) = \Pr(A|h)$.

다.

H2는 위의 증거적 독립성의 정의항과 동일한 구조를 가지고 있다. 그렇다면 베이즈주의 입증 이론의 맥락에서 우리는 H2는 ‘A와 $p'(B) = b$ 는 B와 관련해 증거적으로 독립적이다.’라고 말하고 있는 것으로 이해할 수 있다. 즉 H2는 ‘B와 관련해선, A는 B를 미래에 얼마나 믿게 될지와 증거적으로 무관하다.’는 것이다.

흥미롭게도 위의 H1과 H2가 성립할 때 JC와 HC는 같은 결과를 도출한다.⁹⁾ 그럼 언제 H1과 H2가 성립하는가? 특히 H2는 언제 성립하는가? 이에 대해서 스킴즈(Skyrms, 1980b, p. 125)는 ‘많은 경우’ H2가 성립할 것이며, 이에 많은 경우 JC와 HC는 같은 결과를 내놓을 것이라고 추측한다. 여기서 ‘많은 경우’란 A가 신념도에 대한 명제가 아닌 경우들을 의도한 것이다. 가령 ‘내일 비가 온다’와 같은 명제에 대해서는 H2가 성립하지만 ‘철수는 내일 비가 온다는 것을 90% 정도로 믿고 있다.’는 것과 같은 명제들에 대해선 H2가 성립하지 않을 수 있다는 것이다. 사실 우리 믿음의 대상이 되는 것들은 대부분 ‘내일 비가 온다.’는 것과 같은 유형의 명제들이다. 그리고 이런 명제들에 대해선 (아마도) H2가 성립할 것이다.¹⁰⁾ 이

9) 경험에 의해서 B에 대한 신념도가 직접적으로 수정되었다고, 즉 $p(B) \neq p'(B) = b$ 가 성립한다고 가정하자. 그럼 H1과 H2가 성립할 때 JC와 HC가 같은 결과를 낳는다는 것은 H1과 H2가 다음을 함축한다는 것을 보이는 것으로 충분하다:

$$(a) p(A|p'(B) = b) = bp(A|B) + (1-b)p(A|\neg B);$$

한편, 확률 계산 규칙에 의해서 다음이 성립한다.

$$(b) p(A|p'(B) = b) = \frac{p(B|p'(B) = b)p(A|B \wedge p'(B) = b) + p(\neg B|p'(B) = b)p(A|\neg B \wedge p'(B) = b)}{p(B|p'(B) = b) + p(\neg B|p'(B) = b)}.$$

이제 H1과 H2가 성립한다고 가정하자. 그럼, (b)는 다음과 같이 변형된다:

$$p(A|p'(B) = b) = bp(A|B) + (1-b)p(A|\neg B).$$

즉 (a)가 성립한다. 증명 끝.

10) 우리는 A가 신념도에 대한 2차 진술인 경우 H2가 성립하지 않은 사례를

런 점 때문에, 스킴즈는 JC 대신 HC를 사용한다고 하더라도 많은 경우 베이즈주의 믿음 갱신 결과에 별 영향을 미치지 못한다고 말한다. 따라서 HC를 JC 대신 사용한다는 것이 베이즈주의 인식론으로부터의 급진적 이탈이라고 할 수는 없다.

그러나 직접적으로 신념도를 언급하는 것은 아니지만 신념도와 관련된 명제도 있을 수 있다. 그런 명제들 중 일부는 고차 증거라고 불릴 수 있다. 그럼 어떤 진술이 고차 증거일 수 있을까? 데이비드 크리스텐슨(David Christensen 2010, p. 186)과 같은 철학자는 믿음 기반 약화 증거를 그러한 것의 한 예로 제시한다. 그리고 그런 고차 증거는 (직접적으로 신념도를 언급하지는 않지만) H2를 위반할 수 있다. 가령 앞에 있는 종이의 색이 노랗다는 명제를 생각해 보자. 이 명제와 관련해서, 내가 그 종이가 노란 색이라는 것을 매우 높은 정도로 믿게 되리라는 것은 주변 불빛이 왜곡되었다는 것과 무관하지 않다. 실제로 노란색이 아니지만 노란색이라고 강하게 믿게 된다는 것은 내가 그런 믿음을 형성하게 된 과정의 신빙성에 의문을 제기한다. 좀 더 구체적으로, 실제로 노란색이 아니지만 노란색이라고 강하게 믿게 되리라는 것은 내 주변의 불빛이 왜곡되었다는 것에 대한 증거라고 할 수 있다. 사실 그 종이의 실제 색깔과 주변의 불빛의 상태는 서로 어떤 인과적 관련성이 없으며, 이에 어떤 확률적 관련성을 가지지 못한다. 그러나 실제 색깔과 다르게 믿고 있다는 사실은 주변 불빛의 상태에 대한 증거라고 할 수 있다. 그렇다면 다음이 성립한다는 것은 인식론적으로 자연

쉽게 생각할 수 있다. 가령, H2의 A 가 $p'(B)=b$ 와 동치인 경우를 생각해 보라. 이 경우 H2가 성립하지 않을 수 있다는 것은 분명해 보인다. 한편, 고차 조건화에 대한 국내 논문으로는 박일호 (2012)를 보라. 그는 고차 진술에 대한 믿음을 HC를 이용해 갱신하는 경우와 JC를 이용해 갱신하는 경우를 비교하고, 후자는 무어의 부조리에 빠질 수 있지만 전자는 그렇지 않다고 논증한다.

스럽다:

$$(3.1) \quad p(\text{불빛 왜곡됨 실제로 노란색 아님} \wedge \text{노란색이라고 강하게 믿게 될 것임}) \\ > p(\text{불빛 왜곡됨 실제로 노란색 아님}) = p(\text{불빛 왜곡됨}).$$

여기서 ‘실제로 노란색이 아님’은 H2의 ‘B’에, ‘노란색이라는 강한 믿음’은 H2의 ‘ $p'(B) = b$ ’에, 마지막으로 ‘불빛 왜곡’은 H2의 ‘A’에 대응한다. 위 식에서 첫 번째 부등식은 실제로 노란색이 아니지만 노란색이라고 강하게 믿게 될 것이라는 사실이 불빛 왜곡의 긍정적인 증거의 역할을 한다는 것이며, 두 번째 등식은 실제 색깔과 불빛의 상태는 확률적으로 독립적이라는 것을 나타낸다. 더 나아가 위와 유사하게 다음도 인식론적으로 자연스럽다.

$$(3.2) \quad p(\text{불빛 왜곡됨 실제로 노란색임} \wedge \text{노란색이라고 강하게 믿게 될 것임}) \\ < p(\text{불빛 왜곡됨 실제로 노란색}) = p(\text{불빛 왜곡됨}).$$

이 식은 실제 색깔과 불빛의 상태는 확률적으로 독립적이지만, 실제 색깔대로 믿고 있다는 사실이 불빛이 왜곡되었다는 것에 대한 부정적 증거의 역할을 한다는 것을 나타낸다.

위의 (3.1)과 (3.2)가 보여주듯이 믿음 기반 약화 명제에 대해서는 H2가 성립하지 않는다고 결론내리는 것이 자연스럽다. 따라서 그런 명제에 대해서는 JC와 HC가 다른 결과를 내놓을 것이다. 그렇다면, (JC를 이용한 믿음 갱신과 달리) HC를 이용한 믿음 갱신 결과가 ‘철수의 사례’와 잘 부합한다면 우린 베이즈주의적 틀에서 크게 벗어나지 않고 와이즈버그의 베이즈주의에 대한 공격에 답할 수 있게 될 것이다. 과연 그런가?

4. 고차 조건화와 믿음 기반 약화: 와이즈버그의 논증에 대한 답변

고차 조건화, 즉 HC를 철수의 사례에 적용해보자. 본격적인 계산에 앞서 다음을 가정하자.

$$(4.1) \quad p(E|p'(E) = b) = b;$$

$$(4.2) \quad p(F|E \wedge p'(E) = b) < p(F|E) = p(F) = p(F|\neg E) \\ < p(F|\neg E \wedge p(E) = b).$$

$$(4.3) \quad p(E|p''(F) = 1 \wedge p'(E) = b) = p(E|F \wedge p'(E) = b).$$

여기서 p 는 양말의 색과 관련된 어떤 경험도 하기 전에 철수가 가지고 있었던 신념도 함수이다. 마찬가지로 p' 는 E 에 대한 신념도가 b 로 수정되었을 때, p'' 은 F 를 알게 되었을 때, 철수가 가지게 된 신념도 함수이다. 2절에서 가정한 것과 같이, b 는 1에 가까운 실수 값이다.

그럼 이 세 가정은 수용할만한가? 우선 (4.1)을 생각해보자. 이것은 반 프라센의 반영원리를 진술 E 에 적용한 것이다. 따라서 반영원리를 수용한다면 (4.1)을 거부할 이유가 없을 것 같다. 두 번째로, (4.2)은 어떤가? 3절에서 우리는 믿음 기반 약화 명제와 같은 고차 증거에 대해서는 H2가 성립하지 않는다는 것을 살펴보았다. F 가 믿음 기반 약화 진술이라는 것과 (3.1)과 (3.2)를 고려하면, 철수의 사례에서 (4.2)이 성립한다는 것은 자연스럽게 받아들일 수 있다.

마지막으로 (4.3)은 어떤가? 직관적으로 생각해볼 때, (4.3) 역시 충분히 받아들일 수 있을 것 같다. (4.3)의 좌변과 우변의 차이점은 조건부 신념도의 조건의 일부로 $p''(F) = 1$ 가 포함되느냐, 아니면

F 가 포함되느냐에 있다. (4.3)가 말하는 것은 “두 경우가 $p'(E) = b$ 가 주어졌을 때 E 의 조건부 신념도(conditional credence in E given that $p'(E) = b$)의 값을 결정하는 데 있어 고려해야 할 어떤 정보의 차이도 만들어내지 못한다.”는 것이다. 이것은 수용할 만한 가정인가? (전혀 이론의 여지가 없는 것은 아니지만) 나는 그리 어렵지 않다고 생각한다. 우선 어떤 진술을 확신한다는 것은 매우 강한 증거를 필요로 한다. 여기서 ‘강하다’는 말의 의미는 반대 증거를 제시하기란 무척 힘들며, 또한 그런 확신의 근거를 약화하는 증거를 찾기도 무척 힘들다는 말이다. 그렇다면 ‘ F 가 조건으로 삽입되었을 때 전해주는 정보와 ‘ F 를 확신한다’는 것이 조건으로 삽입되었을 때 전해주는 정보는 대동소이할 것이다. 따라서 나는 (4.3)을 받아들이는 것이 그리 어려운 일은 아니라고 생각한다. 뿐만 아니라, (4.3)은 다소 특수한 가정이라는 것도 고려해야 한다. (4.3)은 모든 신념도에 대한 것이 아니다. 그것은 단지 $p'(E) = b$ 가 주어졌을 때 E 의 조건부 신념도와 관련된 것일 뿐이다. 이런 점을 볼 때, (4.3)를 가정하는 것은 그리 부담스러운 일이 아닐 것이다.

이제 우리가 할 일은 (4.1)~(4.3)을 가정하고, 고차 조건화 HC를 이용하면, 철수의 사례가 잘 다루어질 수 있다는 것을 보이는 것이다. 여기서 잠깐, 2절에서 제시한 우리의 목표를 상기해보자. 2절 마지막 부분에서 나는 다음 두 가지 목표를 제시하였다.

- (a) 고차 조건화에 의한 신념도의 갱신은 (2.1)~(2.5)와 양립가능하다.
- (b) 신념도를 고차 조건화를 이용해 수정한다면, 경험 이전 E 와 F 가 확률적으로 독립적이라고 하더라도 경험 이후엔 E 와 F 사이에 부정적 상관관계가 성립할 수 있다.

2절에서 보인 것은 SC와 JC를 이용해서 철수의 신념도를 갱신하였

을 때, 모순이 발생한다는 것이었다. 여기서 모순이란, SC와 JC를 이용해 갱신된 철수의 사후 신념도 $p''(E)$ 가 (2.4)와 충돌한다는 것을 의미했다. 따라서 철수의 신념도를 HC를 이용해서 수정한다면 (2.4)가 도출된다는 것이 증명되기만 하면 (a)를 보인 것이라고 말할 수 있다. 그리고 그 과정에서 어느 순간 부정적 상관관계가 도입되는지 규명할 수 있다면 (b)를 증명한 것이라고도 할 수 있을 것이다.

그럼, 먼저 HC를 이용해 철수의 신념도를 수정할 때 어떤 일이 일어나는지 살펴보자. 일단 HC에 의해서 철수의 신념도가 수정되었다고 가정하자. 그럼 확률 계산 규칙에 의해서 다음이 성립한다.¹¹⁾

$$(4.4) \quad \begin{aligned} p'(E) &= p(E|p'(E) = b); \\ p'(F) &= p(E|p'(E) = b)p(F|E \wedge p'(E) = b) \\ &\quad + p(\neg E|p'(E) = b)p(F|\neg E \wedge p'(E) = b); \\ p'(E \wedge F) &= p(E|p'(E) = b)p(F|E \wedge p'(E) = b). \end{aligned}$$

이제 (4.2)과 (4.4)로부터 다음을 도출할 수 있다.¹²⁾

¹¹⁾ HC와 (2.2)에 의해서 다음이 성립한다.

(i) 모든 X 에 대해서, $p'(X) = p(X|p'(E) = b)$.

이에 다음이 성립한다.

(ii) $p'(E) = p(E|p'(E) = b)$; $p'(F) = p(F|p'(E) = b)$;

$p'(E \wedge F) = p(E \wedge F|p'(E) = b)$.

한편 확률 계산 규칙에 의해서 다음 두 식이 도출된다.

$p(F|p'(E) = b) = p(E|p'(E) = b)p(F|E \wedge p'(E) = b) \\ + p(\neg E|p'(E) = b)p(F|\neg E \wedge p'(E) = b)$

$p(E \wedge F|p'(E) = b) = p(E|p'(E) = b)p(F|E \wedge p'(E) = b)$.

이에 (ii)의 각 명제는 (4.4)의 각 명제로 변형된다.

¹²⁾ 표기상의 복잡함을 줄이기 위해서 ' $p'(E) = b$ '를 H_E 라고 하자. 그럼, (4.4)와 확률 계산 규칙에 의해서 다음이 성립한다.

$$(4.5) \quad \frac{p'(E)p'(F)}{p'(E \wedge F)} = 1 + p(\neg E | p'(E) = b) \left[\frac{p(F | \neg E \wedge p'(E) = b)}{p(F | E \wedge p'(E) = b)} - 1 \right] > 1.$$

따라서 다음이 성립한다.

$$(4.6) \quad \frac{p'(E \wedge F)}{p'(F)} = p'(E | F) < p'(E).$$

이는 흥미로운 결과이다. (2.5)에 따르면, 철수가 양말의 색에 대한 어떤 경험도 하기 전에 E 와 F 는 확률적으로 독립적이었다. 그리고 2절에서 보여준 것처럼, 그런 확률적 독립성은 JC를 이용한 믿음 갱신을 통해선 보존된다. 하지만 위의 결과가 보여주듯이 HC를 이용한 믿음 갱신을 통해선 E 와 F 사이의 확률적 독립성은 더 이상 보존되지 않으며, 더 나아가 F 가 E 에 대한 신념도의 획득 과정의 신빙성을 의심하게 만드는 믿음 기반 약화 명제인 경우엔 믿음 갱신 이후 E 와 F 는 부정적 상관관계를 가지게 된다. 따라서 이 결

$$\begin{aligned} (i) \quad & \frac{p'(E)p'(F)}{p'(E \wedge F)} \\ &= \frac{p(E|H_E)[p(E|H_E)p(F|E \wedge H_E) + p(\neg E|H_E)p(F|\neg E \wedge H_E)]}{p(E|H_E)p(F|E \wedge H_E)} \\ &= p(E|H_E) + p(\neg E|H_E) \frac{p(F|\neg E \wedge H_E)}{p(F|E \wedge H_E)} \\ &= 1 - p(\neg E|H_E) + p(\neg E|H_E) \frac{p(F|\neg E \wedge H_E)}{p(F|E \wedge H_E)} \\ &= 1 + p(\neg E|H_E) \left[\frac{p(F|\neg E \wedge H_E)}{p(F|E \wedge H_E)} - 1 \right]. \end{aligned}$$

한편 (4.2)에 의해서 다음이 성립한다.

$$(ii) \quad p(F|\neg E \wedge H_E) / p(F|E \wedge H_E) > 1.$$

이에 (i)과 (ii), 그리고 $p(\neg E|H_E) > 0$ 에 의해서 (4.5)가 성립한다.

진술	신념도	진술	신념도
$E \wedge F \wedge p'(E) = 0.8 \wedge p''(F) = 1$	0.005	$\neg E \wedge F \wedge p'(E) = 0.8 \wedge p''(F) = 1$	0.020
$E \wedge F \wedge p'(E) = 0.8 \wedge p''(F) \neq 1$	0.005	$\neg E \wedge F \wedge p'(E) = 0.8 \wedge p''(F) \neq 1$	0.020
$E \wedge F \wedge p'(E) \neq 0.8 \wedge p''(F) = 1$	0.005	$\neg E \wedge F \wedge p'(E) \neq 0.8 \wedge p''(F) = 1$	0.005
$E \wedge F \wedge p'(E) \neq 0.8 \wedge p''(F) \neq 1$	0.005	$\neg E \wedge F \wedge p'(E) \neq 0.8 \wedge p''(F) \neq 1$	0.035
$E \wedge \neg F \wedge p'(E) = 0.8 \wedge p''(F) = 1$	0.000	$\neg E \wedge \neg F \wedge p'(E) = 0.8 \wedge p''(F) = 1$	0.000
$E \wedge \neg F \wedge p'(E) = 0.8 \wedge p''(F) \neq 1$	0.150	$\neg E \wedge \neg F \wedge p'(E) = 0.8 \wedge p''(F) \neq 1$	0.000
$E \wedge \neg F \wedge p'(E) \neq 0.8 \wedge p''(F) = 1$	0.000	$\neg E \wedge \neg F \wedge p'(E) \neq 0.8 \wedge p''(F) = 1$	0.000
$E \wedge \neg F \wedge p'(E) \neq 0.8 \wedge p''(F) \neq 1$	0.030	$\neg E \wedge \neg F \wedge p'(E) \neq 0.8 \wedge p''(F) \neq 1$	0.720
소계	0.2	소계	0.8

과는 (b)가 말하는 바이며, 와이즈버그가 언급한 ‘고정성은 E와 F 사이의 부정적 상관관계가 도입되는 것을 막는다.’는 것이 더 이상 HC를 이용한 믿음 갱신에선 성립하지 않는다는 것을 보인 것이라고 할 수 있다.

그럼, E에 대한 철수의 최종 신념도는 어떻게 되는가? 이는 다음과 같은 계산을 통해서 확인할 수 있다.

$$\begin{aligned}
 (4.7) \quad & p''(E) \\
 & = p'(E|p''(F) = 1) \quad (\because \text{HC와 (2.3)에 의해서}) \\
 & = p(E|p''(F) = 1 \wedge p'(E) = b) \quad (\because \text{HC와 (2.2)에 의해서}) \\
 & = p(E|F \wedge p'(E) = b) \quad (\because \text{(4.3)에 의해서}) \\
 & = p'(E|F) < p'(E) \quad (\because \text{HC, (2.2), (4.6)에 의해서}) \\
 & = p(E|p'(E) = b) = b \quad (\because \text{HC, (2.2), (4.1)에 의해서})
 \end{aligned}$$

즉, JC에 의한 믿음 갱신과 달리, HC를 이용해서 믿음을 갱신하면 (2.4)와 일치하는 결과가 도출된다. 따라서 우리는 HC를 이용한 믿음 갱신은 (2.1)~(2.5)와 양립가능하며, 이에 믿음 기반 약화 증거에

의한 믿음 갱신을 HC는 잘 다룰 수 있다고 결론 내릴 수 있다.

이런 결론은 좀 더 구체적인 예를 통해서도 확인할 수 있다. 철수가 앞 페이지의 <표>와 같은 사전 신념도 분포를 가지고 있다고 하자. 그럼 우리는 철수가 HC를 이용하여 신념도를 갱신했을 때, 그의 신념도들이 어떻게 변할지도 계산할 수 있다. 주요 결과를 나열하면 다음과 같다:

$$p(E) = 0.2;$$

$$p(E|F) = 0.2;$$

$$p'(E) = p(E|p'(E) = 0.8) = 0.8;$$

$$p'(E|F) = p(E|F \wedge p'(E) = 0.8) = 0.2;$$

$$p''(E) = p(E|p'(E) = 0.8 \wedge p''(F) = 1) = 0.2;$$

$$p''(E|F) = p(E|p'(E) = 0.8 \wedge p''(F) = 1 \wedge F) = 0.2.$$

이 사례는 몇 가지 흥미로운 점을 드러낸다. 우선 철수의 E 에 대한 신념도는 양말의 색을 관찰한 이후 0.2에서 0.8로 크게 증가하였지만, 믿음의 기반을 약화하는 증거, 즉 불빛이 왜곡되었다는 증거를 획득하자 원래의 값 0.2로 되돌아갔다. 이런 점은 HC에 의한 믿음 갱신이 믿음 기반 약화 증거가 우리의 신념도에 가하는 영향을 올바르게 반영하고 있다는 것을 보여준다.

한편, E 와 F 사이에 성립하던 확률적 관계에도 변화가 있다. 우선, 관련된 어떤 경험도 하기 전 E 와 F 는 확률적으로 독립적이었다. 이는 와이즈버그가 말한 것과 같이, 실제 양말의 색과 불빛의 왜곡 여부는 서로 무관하다는 것을 반영하고 있다. 하지만, 양말이 붉은 색으로 보이는 경험을 한 이후, E 와 F 사이에 성립했던 확률적 독립성은 더 이상 성립하지 않게 된다. 좀 더 정확하게 말하자면, 경험에 의해 E 에 대한 신념도가 0.8로 수정된 경우에 E 와 F 는 부정적 상관관계를 형성하게 된다. (즉

$p'(EF) = 0.2 < 0.8 = p'(E)$ 가 성립한다.) 이런 점 역시 와이즈버그가 바라던 결과이다. 와이즈버그는 믿음 갱신 과정 중 부정적 상관관계가 도입되는 순간이 있어야 한다고 주장했다. 문제는 JC와 SC를 이용하면 그러한 부정적 상관관계가 도입되는 길이 원천적으로 막혀 있다는 것이었다. 하지만, 위 예는 HC를 이용하면 자연스레 부정적 상관관계가 도입되는 것을 확인할 수 있다.

흥미롭게도 위 사례는 ‘불빛이 실제로 왜곡되었다는 정보를 획득하게 되면 E 와 F 는 다시 확률적으로 독립적이게 된다.’는 것도 보여준다. 즉 $p''(EF) = p''(E)$ 가 성립한다. 혹자는 이 점을 들어 HC에 의한 믿음 갱신의 문제점을 지적하려고 할지도 모르겠다. 하지만 이는 문제점이 아니라 오히려 바람직한 현상인 듯 보이기도 한다. 왜냐하면 불빛이 실제로 왜곡되었다는 정보는 우리의 이전 관찰이 잘못되었다는 것을 말해주는 것이며, 이에 우리의 신념도가 원래의 상태로 회복되는 것은 자연스러워 보이기 때문이다. 그렇다면 이 사례가 보여주는 것처럼 원래의 독립성 관계를 회복하는 것도 그리 이상해보이지 않는다.

5. 결론을 대신하여

본 논문에서 나는 믿음 기반 약화 증거에 의한 믿음 갱신을 베이즈주의가 다룰 수 없다는 와이즈버그의 주장에 베이즈주의적인 답변을 제시하려고 하였다. 나의 답변의 핵심은 그런 증거에 대해선 제프리 조건화가 아니라 고차 조건화를 이용해야 한다는 것이었다. 논의를 끝내기 전에, 나는 고차 조건화를 믿음 갱신 규칙으로 사용하는 것은 표준적인 베이즈주의로부터의 급진적인 이탈이 아니라는 것을 강조하고 싶다. 왜냐하면 3절에서 설명했듯이 믿음 기반 약화 증거와 같은 고차 증거들을 제외한 증거들에 대해선 제프리

조건화와 고차 조건화가 동일한 결과를 산출할 수 있기 때문이다. 이런 점에서 와이즈버그의 주장에 대한 나의 답변은 베이즈주의자들에게 충분히 환영받을 수 있을 것이라 기대한다.

참고문헌

- 박일호 (2012), “조건화와 고차 믿음 갱신”, 『논리연구』 14, pp. 27-58.
- Bovens, Luc & Hartmann, Stephan (2003), *Bayesian Epistemology*, Oxford University Press.
- Chalmers, David J. (2011), “Revisability and Conceptual Change in “Two Dogmas of Empiricism””, *Journal of Philosophy* 108 (8), pp. 387-415.
- Chandler, Jake & Harrison, Victoria (eds.) (2012), *Probability in the Philosophy of Religion*, Oxford University Press.
- Christensen, David (2010), “Higher-Order Evidence”, *Philosophy and Phenomenological Research* 81 (1), pp. 185-215.
- Earman, John (2000), *Hume's Abject Failure: The Argument Against Miracles*, Oxford University Press.
- Gallow, J. Dmitri (2014), “How to Learn from Theory-Dependent Evidence; or Commutativity and Holism: A Solution for Conditionalizers”, *British Journal for the Philosophy of Science* 65 (3), pp. 493-519.
- Howson, Colin (2013), “Finite additivity, another lottery paradox and conditionalisation”, *Synthese* 191(5), pp. 989-1012.
- Jeffrey, Richard C. (2004), *Subjective Probability: The Real Thing*, Cambridge University Press.
- Maher, Patrick (1993), *Betting on Theories*, Cambridge University Press.
- Olsson, Erik J. (2005), *Against Coherence: Truth, Probability, and Justification*, Oxford University Press.
- Shogenji, Tomoji (2007), “Why does coherence appear truth-conducive?”, *Synthese* 157 (3), pp. 361-372.

- Skyrms, Brian (1980a), *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws*, Yale University Press.
- Skyrms, Brian (1980b), "Higher order degrees of belief". In D. H. Mellor (ed.), *Prospects for Pragmatism*, Cambridge University Press, pp. 109-137.
- van Fraassen, C. (1984), "Belief and the will", *Journal of Philosophy* 81 (5), pp. 235-256.
- Wagner, Carl (2013), "Is Conditioning Really Incompatible with Holism?", *Journal of Philosophical Logic* 42 (2), pp. 409-414.
- Weatherson, Brian (2007), "The Bayesian and the Dogmatist", *Proceedings of the Aristotelian Society* 107 (1pt2), pp. 169-185.
- Weisberg, Jonathan (2009), "Commutativity or holism? a dilemma for conditionalizers", *British Journal for the Philosophy of Science* 60 (4), pp. 793-812.
- Weisberg, Jonathan (2015), "Updating, Undermining, and Independence", *British Journal for the Philosophy of Science* 66 (1), pp. 121-159.
- Wheeler, Gregory (2009), "Focused correlation and confirmation", *British Journal for the Philosophy of Science* 60 (1), pp. 79-100.
- White, Roger (2006), "Problems for dogmatism", *Philosophical Studies* 131 (3), pp. 525-557.

전북대학교 철학과

Chonbuk National University

ipark.phil@gmail.com

ARTICLE ABSTRACTS

Higher Order Conditionalization and Undermining

Ilho Park

This paper aims to respond to Weisberg's claim that the standard Bayesian epistemology cannot model an agent's belief updating that is triggered by some undermining evidence. Our epistemological intuition seems to require that the undermining evidence decreases some particular relevant credences. According to Weisberg, however, such a belief change cannot result from the standard Bayesian belief updating rules—i.e., (Jeffrey) Conditionalization. This is because probabilistic independence between some propositions is preserved under (Jeffrey) Conditionalization on the relevant evidence. Yet I will show in this paper that this conclusion is somewhat hasty. In particular, I will show that there is another version of Conditionalization and that when one updates her credences by means of such a version, the belief updating originated in undermining evidence can be well modeled in the Bayesian framework. Some authors often call the version Higher Order Conditionalization.

Key Words: Simple Conditionalization, Jeffrey Conditionalization, Higher Order Conditionalization, Undermining Evidence, Probabilistic Independence, Rigidity, Belief Updating.