

반사실 조건문적 설명 이론에 대한 평가와 맥락 의존 대조 설명의 조건에 대한 인식적 해명*

김 준 성*

선우환 교수(2020)는 저서 ‘때문에’에서 법칙적 설명 모형과 확률적 설명 모형의 문제를 해결할 설명 이론으로 반사실 조건문적 설명 이론을 제시한다. 다른 한편으로, 설명의 수용가능성 기준을 논의하면서 맥락 의존적 대조 설명을 위한 조건으로 설명항과 피설명항의 통상성, 비통상성, 기대 가능성을 제시한다. 이 글에서 필자는 첫째, 반사실 조건문적 설명 이론이 인과의 필연성과 법칙의 필연성이란 양상적 의존관계에 토대한 설명 이론과 다르지 않음을 확인한다. 둘째, 우도 법칙으로 해명된 설명력으로 설명의 수용가능성 기준을 어떻게 형식화할 수 있는지를 제시하고 배경 맥락의 정보에서 인과성이 중요함을 보여준다. 셋째, ‘불구하고’의 인과 사례, 원인의 상반된 인과 경로 사례를 토대로, 배경 맥락에 상대한 설명에서 인과에 관한 양상성 정보가 중요함을 보여준다. 따라서 반사실 조건문적 설명 이론은, 설명의 맥락 의존성과 설명의 양상 의존성을 결합한 것이고, 그 결합은 설명의 새로운 이론이기 보다는 설명을 위하여 반사실적 조건문의 유용성을 보여주는 것임을 확인한다.

【주요어】 대조, 반사실 조건문, 법칙, 배경 맥락, 설명, 양상, 우도, 인과

* 이 논문을 세심히 검토하고 개선의 방향을 제안하여 주신 심사위원님들께 감사의 인사를 올린다. 심사위원님들의 도움으로 논문의 문제들을 개선할 수 있었다. 이 논문은 2021년도 명지대학교 일반교원연구비 지원 사업에 의하여 연구되었음.

** 명지대학교 철학과 교수, jkim30@mju.ac.kr

1. 들어가는 말

최근에 인공지능의 분야에서 인과는 중요하고 유용한 개념의 하나가 되었다. 예를 들어 퍼얼(Pearl 2000)은 인과 모형에 인과에 대한 반사실적 해명을 적극적으로 도입한다. 다른 한편으로, 인과 모형 이론이 반사실적 관계에 대한 형이상학적, 인식적 문제에 어떻게 기여할 수 있는지를 보여주하고자 한다. 이처럼 인과와 반사실적 관계가 주목을 받는 상황에서 선우환 교수의 저서 ‘때문에’(2020)가 갖는 의미는 크다.

‘때문에’에 대한 저자의 다양한 논의는 설명, 인과, 반사실적 조건문 각각에 대해 그리고 이들의 관계에 대해 새로운 관점을 제시하려는 데에서 주목할 만하다. 본 필자는, 선우환 교수의 때문에 이론이 다음 이유로 연구와 논의의 초점이 되어야 한다고 본다. 첫째, 설명의 두 관점을 다시 환기해 준다. 설명이 참인 진술들의 관계라는 존재론적 관점과 설명은 우리의 이해를 돕는다는 인식론적 관점이다. 여기서 인식론적 관점은 설명이 우리의 배경 맥락에 따라 달라진다는 설명의 화용론적 특성이다. 둘째, 설명은 인과적 설명, 기능적 설명 등 다양한 유형의 설명을 일반화하고 포괄할 수 있는 반사실 조건문적 설명의 이론을 제시한다. 셋째, 반사실 조건문적 설명의 이론을 토대로, 인과에 대한 문제와 주요 철학적 인과 이론을 면밀히 분석하고 평가한다.

설명, 인과, 법칙, 반사실적 조건문 모두를 포괄하는 선우환 교수의 이론은 그 논의의 내용에서 매우 풍부하다. 그럼에도 다른 한편으로는, 저자(선우환 교수)의 반사실 조건문적 설명이, 설명의 고전적 모형들에 문제를 제기하고 이들 문제를 해결하려는 기존의 논의와 어떤 큰 차이가 있는지 의문을 갖게 한다. 또한 인과적 필연성, 법칙적 필연성을 해명하는 기존의 반사실적 조건문 이론들과 비교하여 무엇이 다른지 의문을 갖게 한다. 저자의 때문에 이론은, 때문에 관계를 반사실적 조건문으로 해명한 것이다. 여기서 저자는 반사실적 관계는 설명의 때문에 대한 반사실적 이론이 되기를 의도한다. 저자는 때문에 개념이 인과 개념보다 더 근본적이므로 때문에 개념을 규정하는 데에 인과 개념을 사용해서는 안 된다고 주장한다(선우환 2020, 39). 저자는, 저자의 설

명에 대한 반사실적 조건문이 인과에 대한 반사실적 조건문적 이론에 토대하여 얻어진 설명 이론이 아님을 주장한다. 또한 반사실 조건문적 의존 조건이 인과관계에 대한 조건으로서보다는 설명의 유관성에 대한 조건으로서 더 적합하다고 주장한다. 그런 점에서 저자의 반사실 조건문적 설명 이론이 루이스(Lewis 1986)의 인과적 설명 이론과 다름을 강조한다(선우환 2020, 117).

선우환 교수(2020, 6장)는 대조 설명의 수용가능성 기준을 논의하면서 맥락 의존적 대조 설명을 위한 조건을 제시한다. 그 조건으로 설명항과 피설명항의 통상성, 비통상성, 기대 가능성을 제시한다. 이들 기준이 맥락 의존적 설명의 일반화를 의도하는 것이면 그 의미는 크다. 그럼에도 필자는 이들 기준에 대한 더 엄밀하고 형식화된 일반화가 필요하다고 본다. 필자는 그 일반화를 설명력(explanatory power)과 그것에 대한 정량적 평가(측정)에서 찾고자 한다. 설명력에 대한 정량적 평가는 통상성, 비통상성, 기대가능성 기준에 대한 형식화가 될 수 있다. 설명력에 대한 정량적 평가로 두 가지가 제시되어 왔다. 베이지안(Bayesian) 이론과 우도(Likelihood) 이론이다. 이 글에서 필자는 우도 이론의 우도 법칙과 최대 우도 원리로 해명된 설명력이 맥락 의존적 대조 설명의 통상성, 비통상성, 기대 가능성을 어떻게 정량적으로 형식화(일반화)할 수 있는지를 소개하겠다. 이 형식화가 선우환 교수의 반사실 조건문적 설명 이론에 대해 갖는 의미를 세 가지로 볼 것이다. 첫째, 맥락 의존성의 의미와 역할을 더 구체적으로 보여준다. 둘째, 설명항에 대한 선택과 설명력을 정도로 평가하는 데에 배경 맥락의 인과적 정보가 중요하다. 셋째, 저자의 반사실 조건문적 설명의 유관성과 그 설명의 참이 되는 여부는 인과적 필연성, 법칙적 필연성을 해명하는 반사실적 의존 관계에 토대하고 있다. 따라서 반사실 조건문적 설명 이론은, 양상 정보의 맥락에 상대한 인과적 설명이나 법칙적 설명의 다름 아니다.

이 글의 구조는 다음과 같다. 2장에서는 선우환 교수가 제시하는 반사실 조건문적 설명 이론만의 차별성이 있는지를 살펴보겠다. 3장에서는 대조 설명이 배경맥락에 의존하는 관계가 정확히 무엇인지에 대해 논의하겠다. 4장에서는 선우환 교수가 제시한 세 가지 사례를 중심으

로 맥락 의존 대조 설명의 적합성(수용가능성) 조건들을 검토하겠다. 5장에서는 설명력에 대한 정량적 평가로 4장에서 논의한 조건들을 형식화하기 위해 우도를 소개하겠다. 6장에서는 4장에서 논의한 사례 중 첫 번째 사례를 중심으로 맥락 의존 대조 설명의 조건들과 그 설명의 인식적 특성을 우도 법칙으로 해명하겠다. 배경 맥락의 정보에서 인과성이 중요함을 보여주겠다. 7장에서는 ‘불구하고’ 인과 사건에 대하여 선우환 교수의 때문에 이론이 어떻게 응답할지를 묻겠다. ‘불구하고’의 인과 사례, 원인의 상반된 인과 경로 사례를 토대로, 배경 맥락에 상대한 설명에서 인과에 관한 양상성 정보가 중요함을 보여주겠다. 따라서 반사실 조건문적 설명 이론은, 설명의 맥락 의존성과 설명의 양상 의존성을 결합한 것이고, 그 결합은 설명의 새로운 이론이기 보다는 설명을 위하여 반사실적 조건문의 유용성을 보여주는 것임을 확인한다.

2. 반사실 조건문적 설명 이론에 대한 평가

선우환 교수는 반사실 조건문으로 해명된 설명의 유관성을 다음과 같은 일반화로 제시한다(선우환 2020, 78).

- (C) ‘P이기 때문에 Q이다’가 참이다. iff
 (1) P는 참이다. (2) Q는 참이다. (3) $\sim P$ 였다면 $\sim Q$ 였을 것이다.

저자에 따르면, (3)은 P가 Q에 설명적으로 유관할 조건을 보여주는 것이다. Q의 참이 P의 참에 의존하기 때문에 설명적 유관성이 성립하는 것이다. 그리고 그 의존은 (인과적 필연성, 법칙적 필연성에 관한) 양상적 의존 관계이다. 저자가 제시한 반사실 조건문으로 해명한 설명의 유관성이 반사실 조건문적 설명 이론으로 부를만한 특별함이나 차별성을 갖는지 보자.

일반적으로, 반사실적 조건문 이론의 형이상학적, 의미론적 문제에 도 불구하고, 반사실적 조건문은 법칙의 필연성이나 인과의 필연성을

해명하고 이해하는 데에 참으로 유용하다. 선우환 교수(2020, 7장, 9장)가 자세히 검토하고 논의하였던 인과와 설명에 대한 이론들은, 반사실적 조건문을 이론의 유용한 도구로 활용하고 있다. 예를 들어, 새먼(Salmon 1984)의 인과 과정 이론, 우드워드(Woodward 2005)의 조종(개입) 인과와 설명 이론에서 반사실적 조건문은 유의미한 역할을 한다. 필자는¹⁾ 반사실적 관계가 설명의 유관성을 보여주고 설명 모형에서 유의미한 역할을 한다는 데에는 동의한다. 그러나 때문에 이론과 관련하여 설명의 반사실적 관계에 대한 선우환 교수의 입장이 다음 중 어떤 것인지는 필자에게 분명하지 않다. 첫째, 반사실적 조건문이 설명에서 중요한 역할을 한다는 것인가? 둘째, 반사실적 조건문이 하나의 설명 모형이 된다는 것인가? 셋째, 어떠한 설명이든 반사실적 조건문으로 해명될 수 있으므로, 반사실적 조건문이 설명에 있어서 아르키메데스의 점과 같은 개념적 토대가 된다는 것인가?

첫째 질문과 관련하여, 반사실적 조건문이 설명의 유관성을 해명하는 데에 중요한 역할을 한다면 그 역할은 무엇 때문에 가능한지 물을 수 있다. 저자(선우환 2020, 78)도 언급하였듯이, 양상적 관계로서 반사실적 의존 관계 때문일 것이다. 그렇다면 저자의 반사실 조건문적 설명 이론이 인과와 법칙에 대한 반사실적 조건문 이론과 어떤 차별성을 갖는 것인지는 분명하지 않다. 둘째 질문과 관련하여, 반사실적 조건문 자체만으로 하나의 설명 모형이 된다면, 그 아이디어는 매우 대담한 것으로 보인다. 설명은 존재론적 특성과 인식적 특성 모두를 동반하고 있는데, 이들 특성 모두를 만족할 모형을 설계하는 것은 여전히 쉽지 않을 것 같다. 저자가 논의하였듯이, 반사실적 조건문의 참인 여부가 맥락 의존적임을 토대로 설명의 이들 두 특성 모두를 보여줄 수 있을 것 같다. 그렇다면 마찬가지로 반사실 조건문적 설명 이론이 기존의 반사실적 조건문에 대한 논의와 무엇이 다른지 불분명하다. 셋째 질문과 관련하여, 반사실적 조건문이 설명에 있어서 개념적 토대가 된다는 것은, 앞서 보았던 이유들과 관련하여 수용하기 쉽지 않을 것

1) 여기서 ‘필자’는 이 논문의 필자이다. ‘저자’는 『때문에』를 저술한 선우환 교수를 가리킨다.

같다. 반사실적 조건문은 설명적 관계를 표현하는 데에 도구적 유용성을 갖는다. 그런데 그 유용성은 반사실적 조건문을 유의미한 설명으로 만들 인과성, 법칙성 개념에 토대하기 때문에 가능한 것이다.

선우환 교수는 설명이 맥락 의존적이라는 특성에 주목하고 그 특성을 설명의 존재론적 특성과 함께 반사실적 조건문의 맥락 의존적 특성을 강조한다(선우환 2020, 77-83). 그런 점에서 반사실적 조건문이 설명의 이론이 되기를 의도할 수 있다. 그런데 반사실적 조건문이 설명의 유관성을 보여줄 수 있는 것은, 반사실적 조건문 형식 그 자체 때문이 아니라 반사실적 조건문을 참으로 만들어 주는 인과의 필연성, 법칙의 필연성에 관한 양상적 의존 관계 때문일 것이다. 이 내용은 바로 앞에서 언급하였듯이 선우환 교수도 인정한 것이다. 예를 들어, 법칙적 진술과 우연적 일반화 진술을 구분하는 기준으로 반사실적 조건문은 중요하다. 반사실적 조건문으로 법칙적 진술을 특징할 수 있는 것은, 그 조건문이 물리적 필연성에 대한 진술이기 때문이다. (물론 물리적 필연성에 대한 해명은 다시 반사실적 조건문의 참에 의존하므로, 양자 간에는 순환성 문제가 있다.) 반사실적 조건문으로 인과에 대한 문장을 특징하는 것은, 그 조건문이 인과적 필연성에 대한 진술이기 때문이다. 때문에를 반사실적 조건문으로 해명하고 반사실적 조건문으로 설명의 유관성을 해명할 수 있는 것은, 반사실적 조건문을 참이 되게 하는 법칙적, 인과적 필연성 때문이다. 다시 말해, 반사실적 조건문으로 설명적 유관성을 해명할 수 있는 것은, 그 조건문이 법칙적, 인과적 필연성이란 양상성에 토대하기 때문일 것이다. 이와 같은 필자의 생각을 선우환 교수가 이미 전제하고 반사실 조건적 설명 이론을 제시한 것이면, 그 설명 이론이 설명의 유관성을 보여주는 데에 어떤 차별성이 있는 것인지 궁금하다.

선우환 교수(2020, 51~63, 79~83)는 험펠의 연역법칙 설명 모형, 귀납 통계 설명 모형에 대한 반례들을 반사실 조건문으로 분석하면서 반사실 조건문으로 설명의 유관성을 해명할 수 있음을 보여준다. 새먼의 통계 유관성 설명 모형에 대한 반례들을 분석하면서 반사실 조건문적 설명으로 확률적 설명의 문제를 해결할 수 있음을 보여준다. 이들 논의를 토대로 반사실 조건문적 설명 이론의 일반화를 제시한다. 저자는 인과로 설

명되지 않는 기능적 설명, 인과관계가 아닌 물리적 현상들에 대한 설명 등이 반사실 조건문으로 해명된다는 의미에서 반사실 조건문적 설명 이론은 포괄적, 일반적이고 중립적임을 강조한다(선우환 2020, 119).

그러나 선우환 교수가 기존 설명 모형들에 대한 반례들을 반사실적 조건문으로 해명하고 그 해명 때문에 반사실 조건문을 설명 이론으로 명명하는 것이, 어떤 특별함을 갖는지 필자에게는 여전히 분명하지 않다. 이들 경우에 반사실적 조건문이 설명이 되는 것은 선우환 교수도 주목하였고 필자도 앞서 이미 반복하여 언급한 인과의 필연성, 법칙의 필연성이란 양상성 때문이다. 험펠의 연역법칙, 귀납통계법칙 설명 모형들에 대한 반례들, 새먼의 통계 유관성 설명에 대한 반례들에 대해 반사실 조건문이 설명의 유관성을 보여준 것은 인과의 일방성, 인과적 유관성 때문이다. 반사실적 조건문은 인과의 이들 특성에 대한 표현일 뿐이다. 비 인과적 설명으로서 기능적 설명, 물리적 설명을 반사실 조건문적 설명으로 해명할 수 있는 것도 자연 법칙의 물리적(결정적) 필연성이나 비 결정적 성향 때문이다.

정리하면, 선우환 교수는 인과의 유관성이나 법칙의 필연성을 전제한 것이고 또한 그렇게 전제해야 반사실 조건문으로 설명의 유관성을 해명할 수 있는 것이다. 다른 한편으로는, 잠시 후에 보겠지만 그 전제 하에서 설명의 맥락 의존성을 고려된다. 그렇다면, 반사실 조건문적 설명 이론이 새로운 설명 이론으로서 갖는 차별성이 무엇인지는 불분명하다.

3. 대조 설명에서 배경 맥락의 역할

선우환 교수(2020, 5장, 6장)는 설명의 맥락 의존성을 구체적으로 논의하고, 맥락 의존적 설명의 조건을 제시한다. 그 논의와 조건은 설명이 맥락 의존적이란 특성이 무엇인지를 잘 보여주고 있다. 설명의 맥락 의존성은 일상에서 우리가 자주 접하고 익숙한 설명의 방식이다. 설명의 맥락 의존성은 설명에 대한 과학철학의 논의에서 하나의 전통을 형성한다. 그 전통의 핵심은 인식이나 대화의 맥락에 상대한 대조(contrastive) 설명이다. 대조 설명은 언

어학에서 발전한 화용론(pragmatics)과 그것의 성과에 토대한 것이다. 선우환 교수가 논의한 반 프라센(van Fraassen 1980), 루이스(Lewis 1986), 립튼(Lipton 1990)뿐 아니라 헨슨(Hanson 1975), 험펠(Hempel 1965), 드레츠키(Dretske 1972), 게덴포스(Gärdenfors 1988, 1990), 애치슈타인(Achinstein 1984), 히치콕(Hitchcock 1996), 소버(Sober 2020) 등 수많은 학자들의 논의는 대조 설명 이론의 전통을 계속 만들어 왔다.

최근에 대조 설명에 대한 분석과 논의는 대조 인과에 토대한다(Schaffer 2005; Northcott 2008). 대조 인과는 사건 C2보다 사건 C1이 사건 E2보다 E1 사건의 원인이 되는 구조이다. 맥락 의존적 대조 설명도 형식적으로는 같은 구조이다. 특정 맥락에서 피설명항(설명되는 문장) Y2보다 피설명항 Y1에 대해 설명항(설명하는 문장) X1이 설명항 X2보다 상대적으로 더 수용 가능하거나 적합하다는 것이다. 물론 모든 대조 설명이 인과적 설명은 아니고, 따라서 대조 설명을 대조 인과와 연결할 필요는 없다. 그러나 대조 설명에 대한 대부분의 최근 논의는 인과적 설명을 전제한다. 그 이유로 여러 가지를 생각할 수 있다. 필자의 생각에 한 가지 이유는, 인과적 설명으로 제한하지 않는다면 맥락 의존적 설명은 매우 관대한 이론이 되어 설명 그 자체가 무의미하거나 성립하기 어렵기 때문일 것이다. 선우환 교수(2020, 42~48)도 논의하였는데, 반 프라센(van Fraassen 1980)은 화용론적 설명이란 맥락 의존 설명 모형으로 설명의 유관성 조건을 제시하지만 그 조건에는 너무 느슨하다는 문제가 있다.²⁾ 선우환 교수(2020, 119)는 반사실 조건문적 설명 이론의 중립성을 주장한다. 맥락 의존 대조 설명이 인과적 설명에 제한되지 않기를 의도하였을 것이다.

특정 맥락에서 설명항 X1을 설명항 X2보다 상대적으로 더 수용 가

2) 키처와 새먼(Kitcher and Salmon 1987)이 반프라센의 설명 이론에 제기한 문제는 반프라센이 제시한 설명의 유관성이 너무도 느슨하다는 데에 있다. 다음 질문을 보자. 케네디 대통령은 왜 1966년 11월 23일에 죽었는가? 반프라센이 제시한 설명의 유관성은 그 질문에 대해 행성들의 위치에 토대한 점성술사의 설명도 허용할 수 있다. 그런 허위 설명적 유관성을 배제할 수 있는 방법은, 인과적 유관성으로 설명의 유관성에 제약을 주는 것이다.

능하거나 적합하게 하는 기준은 무엇인가? 첫째, 배경 맥락 의존성이 무엇인지를 정확히 해명할 필요가 있다. 대조 설명항들을 무엇으로 할지, 여러 대조 설명항 중 한 설명항을 다른 설명항보다 바로 그 설명항으로 결정할지는, 배경 맥락을 얼마나 구체적으로 기술하는지, 다시 말해 배경 맥락에서 어떤 속성을 고려하는지에 따라 달라질 것이다. 배경 맥락은 우리가 갖는 배경 정보나 지식의 전체이다. 예를 들어 인간에 대한 배경 정보에 상대하여 갑돌이의 암 질환에 대하여 설명한다고 하자. 인간에 대한 어떤 배경 정보인가? 인간의 속성에 관한 모든 배경 정보를 상대하는지, 인간의 생물학적 속성에 관한 배경 정보를 상대하는지, 인간의 심리적 속성에 관한 배경 정보 등을 상대하는지에 따라 대조 설명항들의 집합이 달라질 것이고, 따라서 피설명항에 대해 설명의 유관성을 가질 대조 설명항도 달라질 것이다.

설명에 대한 험펠이나 헨슨의 논의에서 볼 수 있듯이, 설명항을 구체화하는 것은 확률 관계에서 어떤 지시 집단을 선택하는지에 관한 지시 집단 문제(the problem of reference class)와 연결되어 있다. 예를 들어 2021년 10월 31일 서울의 기온을 설명하거나 예측할 수 있다. 그 설명이나 예측은 10월 31일 서울의 기온에 대한 과거 10년의 데이터에 근거할 수 있고, 과거 20년의 데이터에 근거할 수도 있다. 또한 한국의 데이터에 근거할 수 있고, 한반도 전체나 동북아시아의 데이터에 근거할 수도 있다. 이들 데이터 중 어떤 데이터에 상대하는지에 따라 2021년 10월 31일 서울의 기온에 대한 설명이나 예측을 달라질 것이다. 그와 같이 대조 설명도 어떤 배경 맥락을 고려하는지에 따라, 특정 배경 맥락을 얼마나 구체적으로 기술하는지에 따라 (다시 말해 배경 맥락의 어떤 속성을 고려하는지에 따라) 달라질 것이다.

둘째, 설명항 X1을 설명항 X2보다 상대적으로 가장 수용 가능하거나 적합하게 하는 것은, 배경 맥락의 특정 속성에 상대할 때 대조될 설명항들 $\{X1, X2\}$, $\{X1, X2, X5\}$, $\{X2, X4, X8, X9\}$, ... 가운데 어떤 집합을 선택하는지에 달려있다. 한 사건을 설명하는 데에 대조 설명항들의 다양한 집합이 양립할 수 있다. 고려할 배경 맥락이 무엇인지에 따라 다수의 대조 설명항 집합 중 무엇을 수용할지도 결정될 것이다.

립톤(Lipton 1990)과 루이스(Lewis 1986)는 피설명항에 연결된 다수의 인과 역사에 토대하여 대조 설명을 제시한다. 결과에 실제로 연결되거나 연결 가능한 다수의 인과 역사에서 무엇을 설명의 원인으로 파악할지는 가변적이다. 그 답은 어떤 대조 집단을 선택하는지에 따라 달라질 것이다. 히치콕(Hitchcock 1990)은 설명에 관련된 대조 설명항의 집단을 관련 전제(Presupposition)로 고려한다. 관련 전제를 대조 설명항들의 선언 관계(disjunction), 예를 들어 $X1 \vee X2$ 로 제시한다. 관련 전제에 상대하여 상대적으로 큰 확률의 변화를 초래하는 설명항이 설명의 역할을 할 수 있음을 보여준다. 그런데 관련 전제로 무엇을 선택할지는 결국 다시 어떤 배경 맥락을 선택하는지에 달려 있다.

정리하면, 배경 맥락의 특정 속성에 상대하여 대조 설명항의 집합(관련 전제)이 결정되었을 때, 그 집합에서 무엇을 그 설명항으로 결정할지가 대조 설명의 과제가 된다. 선우환 교수(2020, 6장)는 대조 설명항의 집합에서 어떤 것을 설명으로 결정할지에 대해 맥락적 대조 설명의 적합성, 또는 수용가능성을 위한 조건들을 제시한다. 다음 4장에서 필자는 선우환 교수가 제시한 맥락 의존적 대조 설명의 적합성, 수용가능성을 위한 조건을 검토하고 평가하겠다.

4. 맥락 의존 대조 설명의 적합성, 또는 수용 가능성에 대한 평가

선우환 교수(2020, 6장)에서, 설명으로서 때문에 문장, 특별히 대조 문장이 맥락 의존적 설명이 되는 데에 필요한 적합성과 수용가능성 조건들을 제시하고 논의한다. 이들 조건을 세 가지 일반화로 제시하는데 이들 일반화만으로 조건들을 이해하기는 쉽지 않다. 이들 조건을 이끌어내기 위하여 선우환 교수가 논의하는 사례들을 살펴보는 것이 이해에 도움이 된다. 선우환 교수가 맥락 의존적 대조 설명으로 의도하는 바도 더 분명히 이해할 수 있다. 이들 조건이 적절한지에 대해서 살펴보자.

선우환 교수(2020, 232)가 소개하는 박 상병의 생존을 보자. 설명을

요구하는 ‘왜’ 질문은 다음과 같다. 총알이 빗발치는 전장에서 박 상병이 왜 3년 간 살아남을 수 있었는가? 이 질문의 답으로 두 가지 설명이 있다. 첫째, 박 상병은 방탄복을 착용하였기 때문에 살아남을 수 있었다. 둘째, 박 상병은 물을 마셨기 때문에 살아남을 수 있었다. 선우환 교수에 따르면, 첫 번째 답이 두 번째 답보다 더 적절하다. 그 이유는 다음과 같다. 이들 두 답을 반사실적 조건으로 표현하였을 때, 두 번째 답이 첫 번째 답보다 격전이 벌어지는 전장에서 통상적이고 기대하기 쉬운 것이 아니라는 데에 (다시 말해, 일반적으로 더 기대할만한 것이 아니라는 데에) 있다. 치열한 전쟁 상황이란 맥락에서, 첫 번째 답은 일반적 기대 정도가 상대적으로 높아 유의미한 설명이 되지만, 두 번째 답은 일반적 기대 정도가 상대적으로 낮아 무의미한 설명이라는 것이다.

그러나 치열한 전쟁이란 맥락에서도 오히려 두 번째 답이 첫 번째 답만큼 또는 그 이상으로 기대 정도가 상대적으로 높은 설명인 것으로 우리는 주장할 수도 있다. 만약 전쟁의 치열한 정도만큼이나 또는 그 이상으로 물 부족 현상이 심각하였다면, 두 번째 답이 오히려 상대적으로 더 기대할만한 설명이 될 수 있다. 주목해야 하는 것은 피설명항에 대한 설명항 자체의 일반적 기대 정도의 높고 낮음 그 자체가 아니다. 것처럼 설명항의 일반적 기대 정도의 높낮이가 결정되는 것은 어떤 배경 맥락을 전제하는지에 달려 있다는 것이다. 물론 물 부족이란 배경 맥락도 선우환 교수가 제시한 기준을 다시 입증한 것으로 볼 수 있다.

그런데 여기서 일반적 기대 정도가 높고 낮음이 무엇을 의미하는지 물을 수 있다. 우리의 배경 지식에 비추어 기대 정도가 높고 낮아진다고 답해야 할 것이다. 그렇다면 어떠한 배경 지식인가? 위 사례에서는 인과에 대한 지식을 전제하지 않고 일반적 기대 정도의 높고 낮음을 말하기는 어렵다. 인과의 성립 여부나 인과 연관성의 정도에 따라 기대 정도의 높고 낮음이 결정될 것이다. 전쟁 상황에서 방탄복과 생존의 인과관계, 물 부족 상황에서 물 마시기와 생존의 인과관계가 각 설명항의 수용 여부를 결정하게 한다. 여기서 필자가 주목하는 것은, 인과적 연관성에 따른 일반적 기대 정도의 높고 낮음이 정확히 무엇을 의미하는지에 있다. 예를 들어 기계학습 인공지능의 알고리즘이 기대

정도에 대해 학습하여 결과를 산출하게 할 수 있다. 우리는 다음을 물을 수 있다. 인공지능은 어떻게 그 결과를 산출할 수 있을 것인가? 인공지능은 왜 그 결과를 산출하였는가? 이들 질문의 답은 일반적 기대 정도에 대한 엄밀한 형식화를 필요로 할 것이다. 다음 장에서 그 형식화에 대해 논의하겠다.

선우환 교수(2020, 239)는 성냥에 왜 불이 붙었는가에 대한 두 번째 사례에서 세 가지 답을 설명으로 고려한다. 철수가 성냥을 긁는다. 산소가 주어졌다. 성냥이 말라있다. 선우환 교수에 따르면, 철수가 성냥을 긁는 사건이 그 질문의 가장 적절한 답(설명)이다. 그 이유는 산소가 주어지거나 성냥이 말라있는 것은 늘 있는 통상적(일상적)인 사건이지만, 철수가 성냥을 긁는 것은 상대적으로 비통상적(비일상적)이고 개연성이 낮은 놀라운 사건이기 때문이다. 그런데 철수가 성냥을 긁는 (원인) 사건이 것처럼 특별하려면 그 사건을 유별난 것으로 만들 맥락이 전제되어야 한다. 예를 들어 철수가 생애 처음으로 성냥을 손에 들었다는 배경 맥락이나 다른 성냥은 불량품이지만 철수가 가진 성냥만이 정상적인 제품이라는 배경 맥락이 전제되어야 한다. 또는 그 곳에서 지금까지 한 번도 성냥으로 불이 붙은 경우가 없었다는 배경 맥락이 전제되어야 한다. 산소의 존재나 성냥의 건조함은 통상적(일상적, 일반적)이지만, 철수가 성냥을 긁는 사건이 비통상적(유별난 것)이라는 주장만으로는 충분하지 않다. 철수가 성냥을 긁는 사건이 왜 비통상적인지를 말해 주는 배경 맥락이 전제되어야 한다. 설명항 내용의 통상성, 비통상성의 여부나 정도는 어떤 배경 맥락을 전제하는지에 달려 있다.

선우환 교수(2020, 242)가 소개하는 세 번째 사례로 화학 공장에서 유독 가스가 누출된 사건을 보자. 질문은 다음과 같다. 왜 화학 공장에서 유독 가스가 누출되었는가? 질문의 답은 공장 관리자의 안전 규정 위반, 마을 사람들의 방관, 엘리자베스 여왕의 방관이다. 선우환 교수는 규범적 기대라는 맥락에 비추어 공장 관리자의 안전 규정 위반이 상대적으로 가장 통상적인 사건이기 때문에 적절한 설명이라고 본다. 그리고 통상성이 빈도를 기준으로만 판단되어서는 안 됨을 주장한다. 그러나 규범적 기대라는 기준이 원인 사건의 통상성이나 기대 가능성

을 판단할 ‘일반적’ 조건이 되기는 어렵다. 앞서 보았듯이 맥락의 정보에 대한 구체성과 그 맥락에 상대한 대조 집단의 구성이, 피설명항과 설명항의 통상성이나 기대도를 결정한다. 예를 들어 지금까지 화학 공장의 누출 사고들 대부분은 안전 규정 위반에 있지 않았고, 시설의 노후화에서 발생하였다고 하자. 그 경우에 공장 관리자의 안전 규정 위반은 오히려 통상적이지 않다. 그렇다면 앞서 보았던 세 가지 답 모두, 정도의 차이만 있을 뿐이지 설명으로 적절하지 않을 수 있다.

그런데 세 가지 중 안전 규정 위반이 설명이 될 수 있는 이유는, 그 사건이 얼마나 통상적인지에 있는 것은 아니다. 안전 규정 위반이 공간적으로 가장 근접한 원인 역할을 할 수 있기 때문이다. 다시 말해 가장 근접한, 직접적인 원인 사건이 결과 사건에 가장 큰 영향을 준다는 배경 정보를 전제하기 때문이다. 앞서 보았듯이, 배경 맥락의 어떤 내용이 구체화되는지에 따라 대조 설명항의 후보들 중 하나가 적합한 설명으로 고려되는 것이다. 정리하면, 위 사례에서 일상성 여부는 인과에 대한 정보에 따라 판단되고 그 판단이 설명의 설득력을 결정하므로, 인과를 토대로 인식적 설명이 가능한 것이다.

이번에는 피설명항인 유독 가스 누출 사고에 주목하자. 유독 가스 누출 사고는 빈도가 높은 사고가 아니다. 교통사고처럼 빈도가 높은 사고였다면 여수나 울산 등 공단 도시는 더 이상 사람이 사는 도시의 기능을 하지 못하였을 것이다. 그런 맥락에서 유독 가스 누출은 놀라운 현상이다. 그 놀라운 현상이 왜 발생하였는지를 가장 납득할 수 있게 설명하는 것은, 가장 근접한, 직접적인 원인 사건이 결과 사건에 가장 큰 영향을 준다는 배경 정보를 전제할 때, 안전 규정 위반을 최선의 설명으로 추론할 수 있는 것이다.

선우환 교수가 최선의 대조 설명을 위해 제시한 비정상성, 비통상성, 기대하기 어려움 등은 집단이나 개인에 따라, 다시 말해 배경 맥락에 따라 다르게 이해되는 정성적(질적)인 개념이다. 그런데 이들 개념은 모호하므로, 일반화된 기준이 되려면 엄밀한 형식화가 필요하다. 선우환 교수는 저서에서 이들 조건을 형식화하거나 이들 조건이 얼마나 엄밀한지를 보여주려는 게 목적은 아닐 것이다. 설명의 맥락 의존적 특

성에 주목해야 한다는 것을 환기하는 게 목적일 것이다. 일부 독자들은 그들 조건의 엄밀성을 보여주는 것이 필요한 것인지, 왜 중요한지 물을 수 있다. 그런데 설명을 인식적 이해로 고려해야 한다면 그런 인식적 이해에 대한 엄밀한 일반화를 묻는 것은 자연스럽다.

선우환 교수가 제시한 조건에 대한 엄밀한 형식화를 보여주려면, 필자는 두 가지를 고려해야 한다고 본다. 첫째, 맥락 의존적 설명이란 결국 최선의 설명을 위한 추론이란 인식적 설명, 또는 설명에 대한 인식적 개념으로 귀결된다. 물론 선우환 교수와 필자도 동의하듯이, 설명에서 설명항과 피설명항 모두는 참인 문장들이다. 따라서 설명의 유관성은 객관적인 관계이다. 그럼에도 참인 대조 설명항들 가운데 특정 배경 맥락에 상대하여 가장 적합한 설명항(때문에 문장)을 선택하는 것은 결국 인식의 문제이다. 따라서 맥락 의존적 대조 설명은 맥락에 따라 대조 설명항들 중 최선의 설명항을 추론하는 것으로 귀결될 수밖에 없다.

둘째, 설명항과 설명항의 (비)통상성이나 기대하기 어려움 등 정성적(질적) 개념은, 정량적으로 해명될 필요가 있다. 예를 들어 입증 개념을 구문론적(연역 논리적)으로 해명하는 경우에 까마귀 역설, 그루에메랄드 역설, 험펠의 입증의 조건에 대한 문제가 발생하였다. 이들 문제는 조건부 확률을 이용한 베이지안 이론이나 우도 이론으로 정량적으로 해명되었다. 대조 설명이 최선의 설명을 위한 추론으로 귀결된다면, 이들 이론은 마찬가지로 중요한 역할을 할 수 있다. 필자는 우도 이론으로 해명하는 최선의 설명을 위한 추론이 맥락 의존적 설명을 형식화할 수 있다고 본다. 그 형식화는 선우환 교수가 제시한 정성적 조건들을 일반화하고 정당화하는 역할을 할 수 있을 것이다.

5. 설명력에 대한 우도(likelihood) 평가

필자는 우선, 인과적 설명과 관련하여 설명의 관계와 설명력을 구분하는 데에 주목하고자 한다. 인과적 설명에서 설명의 관계는 사실상 사건 수준의 인과에 대한 해명으로 볼 수 있다. 존재론적, 객관적 개념으

로서 설명이다. 다른 한편으로, 인과적 설명에서 설명력은 인식 주체가 설명으로 얻게 되는 이해의 정도를 해명하는 것으로 볼 수 있다. 설명에 대한 인식론적 개념이다. 결과 사건에 대해 다수의 원인 사건들이 연결될 수 있다, 이들 사건 각각이나 이들 사건의 여러 조합은 그 결과 사건에 대한 설명의 관계에 들어 올 수 있다. 그러나 앞서 보았듯이, 설명의 관계에 들어오는 모든 원인 사건이, 특정 맥락을 전제하는 특정 인식 주체에게 이해로서 설명의 효과를 갖는 것은 아니다. 이해로서 설명 효과의 차이는 설명력의 차이로 해명된다.

인과적 설명의 관계에 대한 해명은 사건 수준의 인과에 대한 해명이 대체할 수 있다. 그렇다면 설명력은 무엇으로 해명하는지를 물을 수 있다. 특정 맥락에서 특정 주체가 결과 사건에 대해 갖는 이해의 정도는 어떻게 해명할지를 물을 수 있다. 선우환 교수가 맥락 의존 대조 설명에서 제시한 통상성, 비통상성, 기대 가능성은 설명력의 정도를 가늠하는 기준이 될 수 있다. 선우환 교수가 제시한 통상성, 비통상성은, 실제(사건) 수준 원인을 해명하는 데에 할편과 히치콕(Halpern & Hitchcock 2010)이 주장한 기본 값(default, normality), 파생 값(deviant)의 구분과 유사하다. 이들의 실제(사건 수준) 인과 이론에 따르면, X는 Y의 우연적 상황(contingency)에서 X가 없었다면 Y는 없었다는 경우에 (다시 말해 Y가 X에 반사실적으로 의존하는 경우에) X는 Y의 원인이 된다. 여기서 우연적 상황이란 실제 세계만큼 통상적인 경우이다.

통상성, 비통상성, 기대 가능성이란 질적 기준을 설명력에 대한 양적 해명으로 보여줄 수 있다면, 이들 기준에 대한 엄밀한 형식화가 가능할 것이다. 기존의 논의된 설명력에 대한 형식화는 베이지안 측정 방식이 중심이다(Schubach and Sprenger 2011; Sprenger and Hartmann 2019). 베이지안 측정 방식에 대한 여러 논의와 논쟁이 있었다. 핵심은 그 측정이 설명력을 측정하는 데에 충분하지 않다는 것이다(Glymour 2015; Sprenger and Hartmann 2019). 첫째, 베이지안 이론의 근본적 문제로, 인식 주체마다 달라지는 사전(prior) 확률을 결정하는 문제이다. 둘째, 베이지안 입증 이론에서 제기된 오래된 증거(old evidence) 문제는 설명력의 측정에도 다시 문제가 될 수 있다. 이미 참으로 확인된 사

실에 대해 설명력의 증가는 없기 때문이다. 셋째, 피설명항에 무관한 진술인데 설명항에 추가되어 설명의 관계에 들어오는 것이다. 이미 설명력의 증가가 있는 경우에, 무관한 설명항은 설명력에 대한 기여가 없음에도 설명의 관계에 들어오게 된다(Sprenger and Hartmann 2019).

설명력에 대한 베이지안 측정 방식의 이들 문제 때문에, 필자는 우도 법칙(the Law of Likelihood)으로 설명력을 측정하기를 고려하겠다. 우도 법칙은 최선의 설명을 위한 추론에 대한 해명에 이용된다. 최선의 설명을 위한 추론의 구조는 간명하다. 일어날 개연성이 낮은, 또는 매우 놀라운 현상이 있다. 그 현상을 덜 놀랍고, 더 개연적인 것으로 설명할 가설들이 있다. 이들 가설 중 그 현상을 가장 덜 놀랍고, 가장 개연적인 것으로 만드는 설명을 찾는 것이 최선의 설명을 위한 추론이다(Peirce 1931). 최선의 설명을 위한 추론은 우도 법칙으로 다음과 같이 형식화할 수 있다.

$\Pr(E | H1) > \Pr(E | H2)$ 이면 E는 H2 보다 H1를 지지한다.

매우 놀라운 현상 E가 관찰되었고 그 현상에 대한 최선을 설명으로서 두 가설 H1, H2가 주어졌다. 이들 두 가설 중 놀라운 현상을 덜 놀랍게 만드는 역할을 하는 것이 그 현상에 대한 최선의 설명이다. 우도의 특성(Sober 1993, 30-33)은 다음과 같이 정리할 수 있다.

첫째, 우도는 주어진 가설을 참으로 가정할 때 증거의 진술이 참일 개연성이다. 앞서 보았던 예를 다시 보자. ‘동전의 앞면이 나올 확률이 1/2이다’는 가설 H가 참일 때 ‘동전을 두 번 던질 때 앞면이 나온다’라는 사실을 증거 E로 고려하자. 이 경우에 E에 대한 H의 우도 $\Pr(E | H)$ 를 묻는다면 누구든지 1/4이라 답할 것이다. 우도는 그렇게 모두에게 특정한 값으로 결정된다는 의미에서 우도는 객관성을 가진다.

둘째, 우도는 증거에 대해 어떤 가설이 상대하는지에 따라 다양한 값을 가질 수 있다. 예를 들어 H*가 ‘양면 모두 앞면으로만 만들어진 동전을 던졌을 때 앞면이 나올 확률 1이다’라는 가설이면 앞서와 같은 증거 E인 ‘동전을 두 번 던질 때 앞면이 나왔다’에 대한 우도 $\Pr(E | H^*)$ 는 1이 될 것이다. E에 대한 우도 $\Pr(E | H) = 1/4$ 과 우도 $\Pr(E |$

$H^*) = 1$ 을 합하면 그 값은 당연히 1을 넘는다. 확률은 모든 가능한 경우에 대한 확률을 합할 때 1이 되는 확률의 공리를 만족해야 한다. 그러나 우도의 합은 얼마든지 1을 넘어설 수 있다. 따라서 우도는 개연성을 표현하지만 확률(probability)과 다르다는 것을 알 수 있다.

셋째, 우도는 입증(confirmation) 관계와 비교할 때 그 특징이 분명하다. 입증은 증거에 비추어 가설의 신빙성을 평가하는 것이지만 우도는 가설에 비추어 증거의 설득력이나 의미(significance)를 평가하는 것이다. 내가 사는 아파트의 위층에서 들리는 소음을 증거 E로 삼아 하나의 가설 H를 추론할 수 있다. 그 가설 H는 아파트 위층에 요정들이 살고 있다는 것이다. 그 증거 E가 주어질 때 이 가설 H에 대한 입증의 정도 $\Pr(H | E)$ 는 매우 낮을 것이다. 반면에 그 가설 H가 참이라고 가정할 때 증거 E에 대한 우도 $\Pr(E | H)$ 는 오히려 높을 수 있다.

넷째, 우도는 베이스주의 추론의 기본 형식인 베이스 정리(Bayes theorem)를 구성하는 필수 요소이다. 베이스 정리는 주관적인 사전(prior) 확률의 관계 $\Pr(H)/\Pr(E)$ 와 우도 $\Pr(E | H)$ 의 곱으로 다음과 같이 제시된다.

$$\Pr(H | E) = [\Pr(H)\Pr(E | H)] / \Pr(E)$$

사전 확률의 존재 가능성이나 사전 확률이 가진 주관적 속성에 회의적인 사람들은 객관적인 관계인 우도만으로 과학적 추론의 형식적 모형을 제시하고자 한다.

우도(likelihood)는 기계학습 인공지능 알고리즘의 설계와 실행에서 중요한 역할을 한다. 특별히 최대 우도 원리(Maximum Likelihood Principle; 이하 MLP)가 중요하다. MLP는 증거에 대한 가설들 중 가장 큰 우도를 갖는 가설을 찾는 방법이다. 기계학습 인공지능의 실현은, 학습을 위해 주어진 데이터를 가장 잘 표현하는 확률 모형을 설계하는 것에 달려있다. 예를 들어 ‘개’에 관한 수많은 이미지 데이터를 학습하여 ‘개’를 판단할 수 있는 알고리즘을 갖게 하는 것이다. 더 정확히 말하면 기계학습 인공지능의 실현은 학습을 위해 주어진 데이터

를 가장 잘 표현하는 확률 모형의 모수(parameter)를 추정하는 것이다. 확률 모형의 모수는, 확률 분포의 모양을 결정하는 함수의 계수이다. 예를 들어 정규 분포는 그 분포가 다양한 중 모양으로 나타날 수 있다. 정규 분포 함수의 계수가 그 모양을 결정하게 된다. 정규 분포에는 평균, 분산에 대한 두 개의 모수가 있다. 이 방식은 피설명항 사건에 대해 그 사건을 가장 설명할 수 있는, 또는 그 사건을 가장 설득력(신빙성, 개연성)이 있게 하는 설명항을 추론하는 것과 같다.

다른 한편으로, MLP에는 주목할 문제가 있다. 주어진 증거 샘플 데이터가 편향된 경우이다. 편향된 데이터를 가장 잘 설명할 가설이 초래할 오류가 무엇인지는 쉽게 예상할 수 있다. 예를 들어 동전을 10회 던졌을 때 앞면만 나온 사건들에 가장 적합한 가설은 동전을 던졌을 때 앞면이 나올 확률이 1이 될 수 있다. 그렇다면 동전의 뒷면이 나올 확률은 0이라는 가설도 따라 나온다. 사건의 무작위성을 고려할 때 그 가설은 오류이다. 이런 오류를 피하기 위한 대안이 베이지안 이론의 베이지 정리에 따라 나오는 사후 확률 극대화 원리(Maximum A posteriori Principle, 이하 MAP)이다. 확률에 대한 주관적 해석에 토대한 베이지안의 MAP와 확률에 대한 객관적 해석에 토대한 빈도주의의 MLP의 차이는 사실상 정도의 차이로 볼 수 있다. 첫째, 베이즈주의자는 증거에 비추어 가설에 대한 정확한 확률 값을 찾고자 한다. 반면에 우도주의자는 주어진 증거의 설득력을 확보할 최선의 가설을 찾고자 한다. 따라서 우도주의는 베이즈주의보다 덜 대담한 이론이거나 약화된 이론으로 볼 수 있다. 둘째, 베이즈 정리에서 우도의 역할이 중요함은 분명하다. 우도가 없는 베이즈 정리는 성립하지 않는다. 그 사실은 베이즈주의가 우도주의에 대해 어떤 의미와 관계를 갖는지 분명히 보여준다. 셋째, 우도 이론의 개척자인 피셔(R. Fisher) 등 우도주의자는 베이즈 정리의 역 조건부 확률에 토대하여 우도를 계산한다(박선용외 2016). 따라서 사전 확률에 관한 정보를 고려한다. 앞서와 마찬가지로 그 사실은 우도주의가 베이즈주의에 대해 어떤 의미와 관계를 갖는지 분명히 보여준다(전영삼 2018).

다음 장에서는 선우환 교수가 소개한 세 가지 사례 중 첫 번째 사례를 중심으로, 선우환 교수가 제시한 대조 설명의 정성적 기준들을 맥

락에 상대한 우도로 해명하겠다.)³⁾

6. 맥락 의존 대조 설명에 대한 인식적 해명과 형식화

전쟁에서 박 상병의 생존을 대조 설명으로 설명하는 데에 우도 법칙과 MLP를 응용하여 보자. 맥락 의존 대조 설명에서 두 가지를 주목해야 한다. 첫째, 대조 설명은 관련 전제에 상대해야 한다. 관련 전제는 대조 설명항 문장들의 집합으로, 그 집합은 우리가 고려하는 맥락에 따라 달라진다. 전쟁터라는 배경 맥락에 따라 관련 전제의 대조 설명항을 논의의 편의를 위해 두 개만 고려하자. <박 상병은 방탄복을 입었다. 박 상병은 물을 마셨다.> ‘박 상병은 방탄복을 입었다’를 P_1 , ‘박 상병은 물을 마셨다’를 P_2 라 하자. 따라서 관련 전제로서 대조 설명항들의 집합은 $\langle P_1, P_2 \rangle$ 이다. 물론 그 외에도 여러 설명항 대조 진술을 고려할 수 있다. 예를 들어 박 상병이 숨을 쉬었다. 박 상병이 전투 식량을 먹었다. 여기서 관련 전제의 설명항들은 일어난 사건에 대한 참인 진술들이다. 최선의 설명을 위한 가설을 찾는다면 가설적 진술도 포함할 수 있다. 그러나 여기서 설명의 관계는 참인 진술들 간의 관계

-
- 3) 기계학습 인공지능에서 MLP는 현재 일반화된 가장 많이 사용하는 방법 중 하나이다. 인공지능의 전 과정을 투명하게 보여줄 설명 가능한 인공지능에 대한 연구가 활발하다. 이들 연구 가운데 MLP가 얼마나 정교하게 적용되는지도 설명 가능한 인공지능을 개발하는 데에 중요한 연구 방향이다. 증거 데이터에 가장 적합한 최선의 설명을 위한 가설을 선택하는 방식은 가설이 최대 우도를 갖게 그 가설을 수정하거나 대안 가설을 선택하는 것이다. 그 수정이나 선택의 방식은 다양할 수 있다. 이들 방식 가운데 배경 맥락의 정보에 따른 방식은 매우 중요하다. 필자는 배경 맥락의 정보를 수용할 변수를 매개 변수로 고려하여, 가설에 대한 변수가 증거 데이터에 대해 갖는 우도를 측정해야 한다고 본다. 앞서 보았듯이 매개 변수는 함수의 특성을 결정하여 주기 때문이다. 필자는 설명 가능한 인공지능, 또는 인공지능 학습과 그 학습의 결과를 설명하는 데에 우도 원리와 대조 설명의 맥락 의존성이 중요한 역할을 할 수 있다고 본다.

라는 객관적 관점을 우선 전제한다. 가능성으로서 가설적 진술은 설명항의 집합에서 고려하지 않겠다. 피설명항은 박 상병의 생존이다. 피설명항에 대해서도 우리는 피설명항 대조 진술들의 집합을 고려할 수 있다. 그런데 박 상병의 생존과 사망만이 우리의 관심이라면 피설명항 대조 진술들의 집합을 고려할 필요는 없다.

둘째, 관련 전제가 주어졌을 때 대조 설명은 배경 맥락을 어떻게 기술하는지에 따라, 다시 말해 배경 맥락의 어떤 속성을 고려하는지에 따라 달라진다. 전쟁터라는 배경 맥락에서 총격전이 치열한 전투 상황이란 속성을 고려할 수 있다. 다른 한편으로 물 부족이 심각한 상황이란 속성을 고려할 수 있다. 배경 맥락의 이들 속성 중 무엇을 고려하는지에 따라 대조 설명항 집합(관련 전제)에서 어떤 설명항이 최선의 설명이 될지는 달라질 것이다. 배경 맥락의 첫 번째 속성(총격전이 치열한 전투 상황)을 B1, 두 번째 속성(물 부족이 심각한 상황)을 B2라 하자.

다음 경우에, 배경 맥락 [B]의 (특정 속성을 사례화한) B1을 상대로 대조 진술 P_1, P_2 가 관련 전제 $\langle P_1 \vee P_2 \rangle$ 로 주어졌을 때, 배경 맥락 B1에 상대하여 P_2 보다 P_1 이 Q를 더 통상적(덜 놀라운 것, 더 기대가 가능한 것)으로 만든다. 그 경우는 다음과 같다.

$$\Pr_{B1}(Q | P_1) > \Pr_{B1}(Q | P_2)$$

다음 경우에, 배경 맥락 [B]의 (특정 속성을 사례화한) B2를 상대로 대조 진술 P_1, P_2 가 관련 전제 $\langle P_1 \vee P_2 \rangle$ 로 주어졌을 때, 배경 맥락 B2에 상대하여 P_1 보다 P_2 가 Q를 더 통상적(덜 놀라운 것, 더 기대가 가능한 것)으로 만든다. 그 경우는 다음과 같다.

$$\Pr_{B2}(Q | P_1) < \Pr_{B2}(Q | P_2)$$

그렇다면, 무엇 때문에 각 배경 맥락 B에서 P_1 이나 P_2 를 Q에 대한 최선의 설명으로 볼 수 있는가이다. 우도 법칙은 증거를 평가하는 데에 있어서, 피설명한 사건(현상)에 대한 최적의 설명항을 평가하는 데에 배경 맥락의 역할이 얼마나 중요한지를 잘 보여주고 있다. 따라서

다음 질문을 제기하는 것은 당연하다. 해당 배경 맥락에서 다른 대조 설명항보다 그 설명항에 더 높은 우도를 부여하는 근거는 무엇인가? 그 근거는 가설로서 설명항을 그래프로 표현할 때 그 그래프가 데이터로서 피설명항에 얼마나 근접하였는지에 달려 있다. 그렇다면 다시 물을 수 있다. 그 피설명항에 특정 설명항이 가장 근접하였다는 것은 무엇을 의미하는가? 피설명항과 설명항 간의 상관성(규칙성)이 크다는 것이다. 상관성이 크다는 것은 무엇을 의미하는가? 그 상관성은 우연적 결과일 수 있고, 공통 원인의 결과일 수도 있고, 상호 간 직접 인과의 결과가 될 수도 있다. 예측의 효율성만이 목적이려면 상관성만을 고려하는 것으로 충분할 수도 있다. 그러나 가설과 데이터 간의 관계가 왜 그렇게 근접한 관계인지를 묻는다면 이해로서 설명이 필요하다. 최대 우도 원리 등 여러 통계적 추론이 그런 설명을 목적으로 한다면, 그 목적은 인과성, 법칙성 등 양상성을 전제하는 것으로 볼 수 있다. 결국 우도를 평가하는 기준이 무엇인지는 다시 인과성 등 양상성에 토대한 일반화 원리에 의존하게 된다.

배경 맥락에서 인과성이 인과적 설명력을 평가하는 데에 왜 중요한지에 대해 다음 예로 더 살펴보자(Eva and Stern 2019). 특정 피임약을 복용하는 것은 혈전증을 초래한다. 다른 한편으로, 임신은 혈전증을 초래한다. 따라서 피임약을 복용하여 임신율이 떨어져서 혈전증이 발생할 확률이 떨어진다. 이처럼 두 가지 상반된 인과 경로가 있다. 첫 번째는 피임약을 복용하는 것이 혈전증을 초래하는 경로이다. 두 번째는 피임약 복용으로 임신이 차단되어 혈전증이 막아지는 경로이다.

두 가지 상반된 인과 경로의 배경 정보에 상대하여 다음 두 가지 배경 정보를 더 고려하여 보자. 첫째, 영희는 피임약을 복용하였고 혈전증에 걸렸다. 그런데 영희가 임신 상태인지에 대해서는 모른다고 하자. 그 배경 정보에 상대한다면 피임약 복용이 임신을 막아 혈전증을 막는다는 것을 고려할 때, 영희의 피임약 미복용보다 영희의 피임약 복용은 그녀가 혈전증에 걸렸다는 것을 놀랍고 비정상적인 것으로 만들 수 있다. 둘째, 이번에는 영희가 결코 임신한 적이 없음을 알게 되었다고 하자. 이처럼 인과에 관한 정보에 상대할 때, 영희의 피임약 복용이 영

회의 혈전증을 초래한 것은 놀랍고, 비통상적인 것으로 만들지 않는다.

때문에 문장이나 때문에 문장에 대한 반사실적 조건문을 평가하는 것이 맥락 의존적이라면, 그 맥락 의존성을 평가하는 데에는 인과성에 관한 정보에 의존하게 되는 것이다. 앞서 2장에서 보았듯이, 선우환 교수는 설명항과 피설명항 간의 반사실 조건문적 설명 유관성이 양상의 의존관계에 토대한다고 주장하였다. 양상의 의존관계는 인과적 필연성, 법칙적 필연성이다. 반사실 조건문은 이들 양상에 대한 표현의 방식일 뿐이다. 그렇다면 반사실 조건문적 설명 이론은, 양상 정보의 맥락에 상대한 인과적 설명이나 법칙적 설명의 다름 아니다. 따라서 반사실 조건문적 설명 이론의 차별성은 무엇인지 여전히 불분명하다. 반사실 조건문적 설명 이론의 차별성은, 설명의 맥락 의존성과 설명의 양상 의존성을 결합에 있는 것인가? 그렇다면 그 결합으로 반사실 조건문적 설명 이론을 설명의 새로운 이론으로 보기는 어렵다. 그 보다는 반사실적 조건문의 유용성을 보여주는 설명의 기존 방식들을 확인하여 준 것이다. 반사실적 조건문의 유용성으로, 설명의 유관성에 인과성이나 법칙성이 중요함을 그리고 설명의 배경 맥락 의존성에 반사실적 조건문의 맥락 의존성이 중요함을 확인하여 준 것으로 볼 수 있다.

7. ‘불구하고(despite)’의 설명력에 대한 인과적 배경 맥락의 역할

선우환 교수(2020, 58-72)는 비결정적 사건에 대한 기존의 확률적 설명 이론을 분석하고 비판한다. 필자는 선우환 교수가 제시한 분석과 비판에 동의하는데, 선우환 교수가 제시한 이유뿐 아니라 그 외에 필자가 주목하는 또 다른 이유가 있다. 선우환 교수(2020, 10장, 360)도 주목하고 논의하였듯이, 비결정적 사건을 해명하는 데에 홀(Hall 2004)이 제시한 인과의 두 개념인 의존(dependence)과 산출(production) 모두가 필요하다. 일반적으로, 비결정적 사건에 대한 설명을 하거나 설명의 모형을 제시하기는 쉽지 않다. 쉽지 않은 이유 중 하나는, ‘때문에’

만으로는 충분한 설명을 할 수 없는 경우가 있기 때문이다.

굿(Good 1961a, 1961b; Eells 1991)이 제시한 유명한 예를 보자. 흙즈는 절벽의 끝에서 아래를 내려 보고 있었는데, 악당 모라이어티가 흙즈를 절벽으로 밀어 죽이고자 하였다. 그 뒤에서 이 상황을 지켜본 흙즈의 동료 왓슨은, 그 절박한 상황에서 모라이어티보다 먼저 흙즈를 밀어 절벽의 바로 밑에 있는 턱에 떨어뜨리는 게 흙즈를 살릴 수 있는 것으로 판단하였다. 그리고 모라이어티보다 먼저 흙즈를 밀었다. 그런데 흙즈는 턱에 떨어졌지만, 턱에서 튕겨 나왔고 절벽에 계속 부딪히면서 죽고 말았다. 왓슨의 밀기는 그렇지 않은 것보다 또는 모라이어티의 밀기보다 흙즈의 생존 확률을 높인다. 그러나 왓슨의 밀기는 흙즈를 사망에 이르게 하였다.

유사한 또 다른 경우로 카트라이트(Cartwright 1979; Eells 1991)가 제시한 사례를 보자. 제초제는 갑돌이의 정원의 잡초를 제거하는 데에 매우 효과적이었고, 제초제를 뿌려서 잡초는 거의 제거되었다. 그러나 살아남은 일부 잡초는 제초제에 대한 내성을 갖게 되었고, 몇 주 후에 더 잘 성장하여 이전보다 더 많은 면적으로 정원을 덮었다. 제초제를 뿌리는 것은 그렇지 않은 것보다 잡초의 생존 확률을 낮춘다. 그러나 제초제를 뿌리는 것은 잡초를 더 잘 생존하게 하였다.

이들 두 경우에서 왓슨이 밀었기 때문에 흙즈가 죽었고, 제초제를 뿌렸기 때문에 잡초가 잘 성장하였다고 말할 수 있는가? 이들 때문에 문장을 반사실적 관계로 분석하면 다음과 같다. 왓슨이 밀지 않았다면 흙즈는 죽지 않았을 것이다. 제초제를 뿌리지 않았다면 잡초는 잘 성장하지 않았을 것이다. 그러나 이들 분석을 우리가 수용하기는 어렵다. 왜냐하면 다음과 같이 말할 수 있기 때문이다. 왓슨이 밀지 않았다면 흙즈는 죽었을 것이다. 제초제를 뿌리지 않았다면 잡초는 잘 성장하였을 것이다. 흙즈와 잡초 사례들은 ‘때문에(because)’로서 설명이 아니라 ‘불구하고(despite)’로서 설명이 필요한 경우이다. 왓슨이 밀었음에도 불구하고 흙즈는 죽었다. 제초제를 뿌렸음에도 불구하고 잡초는 잘 성장하였다. ‘불구하고’의 경우가, 때문에 이론으로 어떻게 분석이 될지는 흥미롭다. 선우환 교수는 홀이 제시한 인과의 두 개념을 수용하면서, 반사실적 관계의 역할을 강조한다. 인과에 대한 해명에 인과의

두 개념 모두가 필요하듯이, 설명의 유관성을 해명하는 데에도 ‘때문에’와 ‘불구하고’ 두 개념 모두가 필요한 것 같다. 특별히 비결정적 사건에 대해서는 더욱 그렇다.

‘불구하고’의 경우에 대해 필자가 주목하는 것은, 인과에 대한 해명에 서와 마찬가지로 설명에 대해서도 단일 개념만으로 이루어진 이론은 없다는 것이다. 다시 말해 반사실적 관계만으로는 설명의 유관성을 해명하기에 충분하지 않다. 물론, 산출이나 인과 과정 또는 그와 유사한 개념 만으로도 충분하지는 않다. 선우환 교수가 인과 과정 이론에 대해 보여준 비판에서도 잘 알 수 있다. 선우환 교수는 홀이 제시한 인과의 두 개념을 논의하고 두 개념 모두의 필요성에 주목한다. 선우환 교수는, 홀이 제시한 의존 개념이 선우환 교수의 반사실적 의존 관계와 차이가 있고, 선우환 교수의 반사실적 조건문 이론이 이들 두 개념 모두를 해명할 수 있는 것으로 주장한다. 홀이 제시한 의존 관계와 선우환 교수의 반사실적 의존 관계의 차이가 그렇게 유의미한 것인지, 인과 과정 개념이 반사실적 관계로 해명될 수 있는지는 분명하지 않다. ‘불구하고’의 경우들은 반사실적 관계만으로 해명하고 설명하기 쉽지 않을 것 같다. ‘불구하고’의 경우를 선우환 교수가 때문에 이론으로 어떻게 분석할지는 흥미롭다.

홈즈의 죽음과 잡초의 생존의 경우는, 설명의 적절한 설명이 인과 과정의 정보에 의존한다는 것을 잘 보여준다. 왓슨의 밀기가 홈즈를 살릴 수 있는 행위임에도 홈즈가 죽은 것은 매우 놀라운 사건이다. 그러나 이들 사건에 관련된 인과 과정을 배경 정보로 상대하였을 때 왓슨의 밀기는 홈즈의 죽음에 대한 설명이 된다. 그 배경 정보를 상대하였을 때 왓슨의 밀기는 홈즈의 죽음을 덜 놀라운 사건으로 납득할 수 있게 한다. 잡초의 생존도 다르지 않다. 체초제를 뿌렸음에도 잡초가 생존한 것은 놀라운 사건이다. 그러나 이들 사건에 관련된 인과 과정을 배경 정보로 상대하였을 때 잡초가 생존한 것은 덜 놀라운 사건으로 납득할 수 있게 한다.

사건을 설명하는 데에 통상성, 비통상성, 기대의 정도는 배경 정보, 특별히 인과에 대한 배경 정보에 상대하며, 이들 배경 정보에서 어떤 속성에 주목하는지에 따라 결정되는 것임을 확인할 수 있다. 우도에

따른 최선의 설명에서도 인과에 대한 배경 정보를 상대하는지에 따라 다수의 대조 설명항 중 어떤 설명항이 피설명항의 개연성(설득력)을 가장 높게 하여, 가장 이해할만한 설명항으로 만드는지 알 수 있다.

새먼의 인과 과정 이론부터 최근 할편과 퍼얼의 인과 모형 이론 등 기존의 주목할 만한 설명 이론은 인과에 의존하는 설명 모형을 제시하였다. 다른 한편으로, 인과 의존형 설명 모형에 대한 비판은 설명의 맥락 의존성으로 설명의 자율성을 강조하였다. 그러나 완전한 자율성으로서 설명은 불가능하다. 만약 그런 설명이 가능하다면 그 설명은 반프라센의 설명 이론이 부딪히는 문제를 다시 만나게 될 것이다. 맥락 의존적 대조 설명은 설명이 상대하는 세계, 정보 등 제약성을 전제하는 설명이다. 그 제약성을 전제할 때 설명의 합리성이 가능하게 된다. 제약성의 핵심에는 인과성이 있고, 인과에 대한 반사실적 관계에 대한 이해를 전제한다. 그 전제에서 설명에 대한 반사실적 조건문이 의미를 가질 수 있게 된다. 선우환 교수는 그 전제가 반사실 조건문적 설명 이론과 일관되며, 따라서 그 설명 이론을 중립적 이론으로 확인할 것이다. 그러나 반사실 조건문적 설명 이론은 인과성에 토대한 양상 의존적 이론이며, 그렇지 않다면 법칙성에 토대한 양상 의존적 이론일 뿐이다. 필자가 논의한 설명의 맥락 의존성이나 반사실 조건문이 맥락 의존적이란 사실도, 인과성이나 법칙성이란 양상성에 의존한다. 그렇다면 반사실 조건문적 설명 이론은 설명의 양상 의존성과 맥락 의존성에 대한 다른 명명에 불과할 수 있다.

8. 나가는 말

필자가 이 글에서 발전시킨 논의들은 선우환 교수의 저서 ‘때문에’에 토대하고 있다. 선우환 교수의 저서는 반사실적 관계가 인과와 설명에 대해 갖는 역할과 중요성을 이해하고 연구하는 데에 매우 큰 도움을 줄 것이다. 또한 선우환 교수의 저서는 여러 분야에서 계속 활발히 논의될 인과와 설명의 문제들을 해결해 나가는 데에 유용한 안내자가 될 것이다.

참고문헌

- 김준성 (2009), 「인과적 설명에서 대조(contrast)의 역할과 의미」, 『철학』, 101권, pp. 109-137.
- _____ (2011), 「설계 논증에서 관찰 선택 효과에 대한 논쟁: 확률적 추론의 관점에서」, 『과학철학』, 14권 2호, pp. 1-45.
- _____ (2012), 「사항(four-placed) 관계로서 대조 인과에 대한 평가 (I): 사항 대조 관계에 대한 노스콧(Northcott 2008)의 정의를 중심으로」, 『철학사상』, 46권, pp. 283-325.
- _____ (2021), 「‘때문에’에 대한 반사실적 관계가 설명과 인과에서 갖는 의미」, 『2021년 한국과학철학회 정기학술대회』, 9회 아시아-태평양 과학철학자대회.
- 박선용, 윤형석 (2016), 「우도원리에 대한 분석과 그에 따른 교육적 시사점에 대한 연구」, 『수학교육』, 55권 2호, pp. 193-208.
- 선우환 (2020), 『때문에 - ‘때문에’의 의미에 대한 철학적 연구』, 아카넷.
- 전영삼 (2018), 「우도와 증거-가설 생성의 메커니즘」, 『과학철학』 21권 2호, pp. 1-45.
- Achinstein, P. (1984), “The pragmatic character of explanation”, *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1984(2): pp. 275-92.
- Cartwright, N. (1979), “Causal Laws and Effective Strategies”, *Nous* 13(4): pp. 419-437.
- Dretske, F. I. (1972), “Contrastive Statements”, *The Philosophical Review* 81(4): pp. 411-37.
- Eells, E. (1991), *Probabilistic Causality* Vol. 1, Cambridge University Press.
- Eva, B. and Stern, R. (2019), “Causal Explanatory Power”, *The British Journal for the Philosophy of Science* 70(4): pp. 1029-50.
- Gärdenfors, P. (1988), *Knowledge in Flux: Modeling the Dynamics of Epistemic States*, The MIT Press.

- _____ (1990), “An Epistemic Analysis of Explanations and Causal Beliefs”, *Topoi* 9(2): pp. 109-24.
- Glymour, C. (2015), “Probability and the Explanatory Virtues”, *The British Journal for the Philosophy of Science* 66(3): pp. 591-604.
- Good, I. J. (1961a), “A Causal Calculus (I)”, *The British Journal for the Philosophy of Science* 11(44): pp. 305-18.
- _____ (1961b), “A Causal Calculus (II)”, *The British Journal for the Philosophy of Science* 12(45): 43-51.
- Hall, N. (2004), “Two Concepts of Causation”, in Collins, J., Hall, N. and Paul, L. (eds.), *Causation and Counterfactuals*, pp. 225-76.
- Halpern, J. Y. and Hitchcock, C. (2010), “Actual Causation and the Art of Modeling”, in Dechter, R., Geffner, H. and Halpern, J. Y. (eds.), *Causality, Probability, and Heuristics: A Tribute to Judea Pearl*, London: College Publications, pp. 383-406.
- Hanson, B. (1975). “*Explanations – Of What?*” (manuscript, Department of Philosophy, Stanford University).
- Hempel, C. G. (1965), *Aspects of Scientific Explanation*, New York: The Free Press.
- Hitchcock, C. R. (1996), “The Role of Contrast in Causal and Explanatory Claims”, *Synthese* 107(3): pp. 395-419.
- Kitcher, P., and Salmon, W. (1987), “Van Fraassen on Explanation”, *The Journal of Philosophy*, 84(6): pp. 315-30.
- Lewis, D. (1986). “*Causal Explanation.*” in *Philosophical Papers Vol. 2*, Oxford: Oxford University Press, pp. 214-40.
- Lipton, P. (1990), “Contrastive Explanation”, *Royal Institute of Philosophy Supplements* 27, pp. 247-66.
- Mackie, J. L. (1965), “Causes and Conditions”, *American Philosophical Quarterly* 2(4): pp. 245-64.
- Northcott, R. (2008), “Causation and Contrast Classes”, *Philosophical Studies* 139(1): pp. 111-23.

- Peirce, C. S. (1931), *The Collected Papers of Charles Sanders Peirce*, Vol. 1-4, Harvard University Press.
- Putnam, H. (1975). "Philosophy and Our Mental Life", in *Mind, Language and Reality: Philosophical Papers*, Vol. 2, Cambridge University Press, pp. 291-303.
- Salmon, W. C. (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton University Press.
- Schaffer, J. (2005), "Contrastive Causation", *The Philosophical Review* 114(3): pp. 327-58.
- Sober, E. (2018), *The Design Argument*, Cambridge University Press.
- _____ (2018), *Philosophy of Biology 2nd Edition*, Routledge.
- _____ (2020), "A Theory of Contrastive Causal Explanation and Its Implications Concerning the Explanatoriness of Deterministic and Probabilistic Hypotheses", *European Journal for Philosophy of Science* 10(3): pp. 1-15.
- Schubach, J. N. and Sprenger, J. (2011), "The Logic of Explanatory Power", *Philosophy of Science* 78(1): pp. 105-27.
- Sprenger, J. and Hartmann, S. (2019), *Bayesian Philosophy of Science*, Oxford University Press.
- Van Fraassen, B. C. (1980), *The Scientific Image*, Oxford University Press.
- Woodward, J. (2005), *Making Things Happen: A Theory of Causal Explanation*, Oxford University Press.

논문 투고일	2022. 02. 02
심사 완료일	2022. 03. 04
게재 확정일	2022. 03. 21

Evaluating the Counterfactual Conditional Theory of Explanation and an Epistemic Assessment of the Conditions for Contrastive Explanation Dependent on Background Contexts

Joosung Kim

Hwan Sunwoo(2020) puts forth the counterfactual conditional theory of explanation that is intended to meet the problems with the nomological models of explanation and the probabilistic (statistical relevance) model of explanation. He also puts forth the conditions for contrastive explanation dependent on background contexts while discussing the acceptability criterion of explanation. The conditions are normality, anormality, and expectability. In this paper, I first show that the counterfactual conditional theory of explanation is not different from the received theories of explanation founded on the modal dependency relation such as causal necessity and lawful necessity. Second, I explicate the acceptability criterion of explanation while assessing explanatory power in terms of the law of likelihood. Third, while discussing the ‘despite’ causal cases and the dual causal paths. I show that the background information about modality such as causality and lawful necessity leads us to choose the best one among different contrastive explanations. I conclude that the counterfactual theory of explanation is a conjunction of the dependency of explanation on modality and the dependency of explanation on background contexts. The counterfactual theory of explanation rather explicates the versatility of counterfactual conditional than is a new theory of explanation.

Key words: background contexts, causality, contrast, counterfactual conditional, explanation, likelihood, modality

